

**INTRODUCTION TO
INSTRUMENTATION
AND
MEASUREMENTS**
Third Edition

**INTRODUCTION TO
INSTRUMENTATION
AND
MEASUREMENTS**
Third Edition

Robert B. Northrop



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

MATLAB® is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB® software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB® software.

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2014 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Version Date: 20140415

International Standard Book Number-13: 978-1-4665-9679-5 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>

*I dedicate this third edition to my wife and daughters, Adelaide, Anne,
Kate, and Victoria, who have always believed in my writing.*

Contents

Preface.....	xix
Author.....	xxv
1. Measurement Systems.....	1
1.1 Introduction.....	1
1.2 Measurement System Architecture.....	1
1.2.1 Sensor Dynamics.....	3
1.2.2 Overview of Signal Conditioning.....	7
1.3 Errors in Measurements.....	7
1.4 Standards Used in Measurements.....	14
1.4.1 Introduction.....	14
1.4.2 Electrical Standards.....	14
1.4.2.1 Volt.....	15
1.4.2.2 Resistance.....	19
1.4.2.3 Current and Charge.....	25
1.4.2.4 Capacitance.....	28
1.4.2.5 Inductance.....	31
1.4.3 Time and Frequency Standards.....	32
1.4.4 Physical Standards.....	33
1.4.4.1 Mass.....	33
1.4.4.2 Length.....	35
1.4.4.3 Temperature.....	35
1.4.4.4 Uncertainties in the SI Base Units.....	35
1.5 Chapter Summary.....	36
Problems.....	36
2. Analog Signal Conditioning in Instrumentation.....	41
2.1 Introduction.....	41
2.2 Differential Amplifiers.....	41
2.2.1 Introduction.....	41
2.2.2 Analysis of Differential Amplifiers.....	42
2.2.3 Common-Mode Rejection Ratio.....	43
2.2.4 Measurement of CMRR, A_D , and A_C	44
2.2.5 Effect of Source Resistance Asymmetry on CMRR.....	44
2.3 Operational Amplifiers.....	47
2.3.1 Types of Op-Amps.....	48
2.3.2 Basic Broadband Amplifier Design Using Op-Amps.....	50
2.3.2.1 Noninverting Amplifier.....	50
2.3.2.2 Inverting Amplifier and Summer.....	52
2.3.3 Current Feedback Op-Amps.....	53
2.4 Analog Active Filter Applications Using Conventional Op-Amps.....	58
2.4.1 Introduction.....	58

2.4.2	Analog Active Filter Architectures	59
2.4.2.1	Controlled-Source Active Filters	59
2.4.2.2	Biquad Active Filters	62
2.4.2.3	Generalized Impedance Converter Active Filters	65
2.4.2.4	High-Order Active Filters	69
2.4.3	Operational Amplifier Integrators and Differentiators.....	69
2.4.4	Summary.....	71
2.5	Instrumentation Amplifiers.....	71
2.5.1	Instrumentation Amplifiers That Can Be Made from Op-Amps	73
2.5.2	Isolation Amplifiers	75
2.5.3	Autozero Amplifiers	77
2.5.4	Absolute Isolation	79
2.5.5	Summary.....	79
2.6	Nonlinear Analog Signal Processing by Op-Amps and by Special Function Modules	81
2.6.1	Introduction.....	81
2.6.2	Precision Absolute Value Circuits	83
2.6.3	Multifunction Converters.....	85
2.6.4	True RMS-to-DC Converters.....	86
2.6.5	Square-Root Circuits and Dividers	88
2.6.6	Peak Detectors and Track-and-Hold Circuits	90
2.6.7	Log Ratio and Trigonometric ICs	94
2.6.8	Summary.....	95
2.7	Charge Amplifiers.....	96
2.7.1	Introduction.....	96
2.7.2	Charge Amplifiers Used with Piezoelectric Transducers.....	96
2.7.3	Charge Amplifier as an Integrating Coulombmeter	98
2.8	Phase-Sensitive Rectifiers	100
2.8.1	Introduction.....	100
2.8.2	Double-Sideband, Suppressed-Carrier Modulation	100
2.8.3	Demodulation of DSBSCM Signals by Analog Multiplier.....	101
2.8.4	Other PSR Designs.....	102
2.8.5	Lock-In Amplifier.....	102
2.8.5.1	Introduction	102
2.8.5.2	Calculation of the SNR Improvement Using a Lock-In Amplifier	107
2.8.5.3	Summary	110
2.8.6	Signal Averaging to Improve SNR of Evoked Transient Signals	111
2.8.6.1	Introduction	111
2.8.6.2	Analysis of SNR Improvement by Averaging.....	113
2.9	Chapter Summary.....	116
	Problems.....	116
3.	Noise and Coherent Interference in Measurements	125
3.1	Introduction	125
3.2	Descriptions of Random Noise in Circuits	125
3.2.1	Probability Density Functions	126

3.2.2	Power Density Spectrum	127
3.2.3	Sources of Noise in Signal Conditioning Systems	130
3.2.3.1	Noise from Resistors	130
3.2.3.2	Two-Source Noise Model for Active Devices	133
3.2.3.3	Noise in JFETs	134
3.2.3.4	Noise in BJTs	135
3.3	Propagation of Gaussian Noise through Linear Filters	137
3.4	Broadband Noise Factor and Noise Figure of Amplifiers	138
3.5	Spot Noise Factor and Figure	140
3.6	Transformer Optimization of Amplifier F_{spot} and Output SNR	143
3.7	Cascaded Noisy Amplifiers	144
3.8	Examples of Calculations of the Noise-Limited Resolution of Certain Signal Conditioning Systems	145
3.8.1	Calculation of the Minimum Resolvable AC Input Voltage to a Noisy, Inverting Op-Amp Amplifier	145
3.8.2	Calculation of the Minimum Resolvable Direct Current in White and $1/f$ Noise	147
3.8.3	Calculation of the Minimum Resolvable AC Input Signal to Obtain a Specified Output SNR in a Transformer-Coupled, Tuned Amplifier	148
3.8.4	Calculation of the Smallest $\Delta R/R$ in a Wheatstone Bridge to Give a Specified SNR _{out}	149
3.8.5	Determination of the Conditions for Maximum Output SNR Given a Simple Inverting Op-Amp Amplifier with Known e_{na} and i_{na}	151
3.9	Modern, Low-Noise Amplifiers for Use in Instrumentation Signal-Conditioning Systems	151
3.10	Coherent Interference and Its Minimization	153
3.10.1	Sources of Coherent Interference	153
3.10.1.1	Direct Electrostatic Coupling of Coherent Interference	154
3.10.1.2	Direct Magnetic Induction of Coherent Interference	155
3.10.1.3	Ground Loops	157
3.10.2	Cures for Coherent Interference	158
3.10.2.1	Powerline Low-Pass Filters	158
3.10.2.2	Transient Voltage Suppressors	159
3.10.2.3	Coherent Interference Induced in Coaxial Cables by Magnetic Coupling	164
3.10.2.4	Single Grounding of Coax Shields to Prevent Ground-Loop Interference	165
3.10.2.5	Use of a Longitudinal Choke or Neutralizing Transformer to Attenuate Common-Mode Coherent Interference	166
3.10.2.6	Experimental Verification of Cabling and Grounding Schemes to Achieve Minimum Noise Pickup	168
3.10.2.7	Circuit Grounding	170
3.10.2.8	Ferrite Beads and Feedthrough Capacitors	170
3.10.2.9	Interruption of Ground Loops by the Use of Isolation Transformers and Photooptic Couplers	171

3.10.2.10	Photooptic Couplers	172
3.10.2.11	Use of Guarding and Shielding to Reduce Common-Mode Coherent Interference	173
3.10.3	Summary of Techniques for Coherent Noise Reduction	177
3.11	Chapter Summary.....	179
	Problems.....	180
4.	DC Null Methods of Measurement	187
4.1	Introduction	187
4.2	Wheatstone Bridge Analysis	189
4.3	Kelvin Bridge	190
4.4	Anderson Constant Current Loop.....	193
4.4.1	Introduction.....	193
4.4.2	Anderson Loop Applied to Groups of Sensors	195
4.4.3	Conclusion	201
4.5	Potentiometers.....	201
4.6	Chapter Summary.....	202
	Problems.....	203
5.	AC Null Measurements	209
5.1	Introduction	209
5.2	Components: Inductor Equivalent Circuits.....	209
5.3	Components: Capacitor Equivalent Circuits.....	211
5.4	AC Operation of Wheatstone Bridges	213
5.5	AC Bridges	214
5.5.1	Bridges Used to Measure Capacitance.....	215
5.5.1.1	Resistance Ratio Bridge.....	215
5.5.1.2	Schering Bridge.....	216
5.5.1.3	Parallel C Bridge	217
5.5.1.4	De Sauty Bridge	218
5.5.1.5	Wien Bridge	219
5.5.1.6	Commutated Capacitor Bridge	221
5.5.2	Bridges Used to Measure Inductance and Mutual Inductance.....	222
5.5.2.1	Maxwell Bridge.....	223
5.5.2.2	Parallel Inductance Bridge	224
5.5.2.3	Hay Bridge.....	224
5.5.2.4	Owen Bridge.....	225
5.5.2.5	Anderson Bridge.....	226
5.5.2.6	Heaviside Mutual Inductance Bridge.....	227
5.5.2.7	Heydweiller Mutual Inductance "Bridge"	228
5.5.3	Null Method of Measuring Transistor Small-Signal Transconductance and Feedback Capacitance	229
5.6	Chapter Summary.....	232
	Problems.....	232
6.	Survey of Sensor Mechanisms	237
6.1	Introduction	237
6.2	Categories of Sensor Mechanisms.....	237
6.2.1	Resistive Sensors.....	237

6.2.1.1	Resistive Temperature Sensors.....	238
6.2.1.2	Resistive Strain Gauges	241
6.2.1.3	Photoconductors.....	242
6.2.1.4	Conductive Relative Humidity Sensors.....	246
6.2.1.5	Direct Resistance Change Used to Sense Position or Angle	248
6.2.1.6	Sensors Based on the Giant Magnetoresistive Effect.....	250
6.2.1.7	Anisotropic Magnetoresistance	253
6.2.2	Voltage-Generating Sensors	260
6.2.2.1	Thermocouples and Thermopiles	260
6.2.2.2	Photovoltaic Cells.....	263
6.2.2.3	Piezoelectric Transducers.....	266
6.2.2.4	Pyroelectric Sensors.....	270
6.2.3	Sensors Whose Voltage Output Is Proportional to $d\Phi/dt$	274
6.2.3.1	Variable Reluctance Phonograph Pickup.....	275
6.2.3.2	Electrodynamic Accelerometer	276
6.2.3.3	Linear Velocity Sensors	277
6.2.4	Sensors Whose Output EMF Depends on the Interaction of a Magnetic Field with Moving Charges.....	278
6.2.4.1	Faraday Effect Flowmeters	278
6.2.4.2	Hall Effect Sensors	281
6.2.5	Sensors Based on Variable Magnetic Coupling.....	284
6.2.5.1	LVDT	284
6.2.5.2	Synchros and Resolvers.....	286
6.2.6	Variable Capacitance Sensors.....	288
6.3	Fiber-Optic Sensors.....	291
6.3.1	Magneto-Optic Current Sensors.....	292
6.3.2	Means of Measuring the Optical Rotation of the Linearly Polarized Light Output of Certain Optical Sensors	295
6.3.3	Fiber-Optic Mechanosensors	306
6.4	Photomultiplier Tubes and Related Photoelectron Multiplication Devices	311
6.4.1	Introduction.....	311
6.4.2	Operation of PMTs.....	313
6.4.3	Single-Channel Photomultiplier.....	316
6.4.4	Microchannel-Plate Photomultipliers	317
6.4.5	Summary.....	319
6.5	Ionizing Radiation Sensors.....	319
6.5.1	Introduction.....	319
6.5.2	Geiger-Muller Tubes	320
6.5.3	Solid-State, Crystal Radiation Sensors.....	322
6.5.4	Scintillation Counters.....	326
6.6	Electrochemical Sensors	327
6.6.1	Introduction.....	327
6.6.2	pH- and Ion-Specific Electrodes	328
6.6.3	Polarographic Electrodes	330
6.6.4	Fuel Cell Electrodes	332
6.7	Mechano-Optical Sensors.....	332
6.7.1	Introduction.....	332
6.7.2	Optical Coding Disks.....	333

6.7.3	Sagnac Effect Sensing of Angular Velocity.....	334
6.7.4	Laser Doppler Velocimetry	338
6.8	Chapter Summary.....	345
	Problems.....	345
7.	Applications of Sensors to Physical Measurements	361
7.1	Introduction.....	361
7.2	Measurement of Angular Acceleration, Velocity, and Displacement.....	361
7.2.1	Angular Acceleration Measurement.....	361
7.2.1.1	Angular Acceleration Measurement with a Constrained Mechanical Gyro	362
7.2.1.2	Simple Inertia Wheel–Spring–Dashpot Angular Accelerometer	363
7.2.1.3	Servoangular Accelerometer	364
7.2.2	Angular Velocity Measurement with Mechanical Gyroscopes	365
7.2.2.1	Mechanical Rate Gyro	369
7.2.2.2	Sagnac-Effect Fiber-Optic Gyroscopes.....	372
7.2.2.3	Vibrating Mass Rate Gyro.....	374
7.2.2.4	Humphrey Air Jet Rate Gyro.....	377
7.2.3	Angular Velocity Measurement with Tachometers.....	381
7.2.4	Angular Position Measurement with Gyroscopes.....	386
7.2.5	Angular Position Measurement with Clinometers.....	391
7.2.6	Angular Position Measurement of Shafts	394
7.3	Measurement of Linear Acceleration, Velocity, Displacement, and Position	396
7.3.1	Accelerometers	396
7.3.1.1	Basic Newtonian Accelerometer	396
7.3.1.2	Servoaccelerometer Design.....	397
7.3.1.3	Piezoelectric Accelerometers	398
7.3.2	Linear Velocity Measurement.....	400
7.3.2.1	Means of Measurement of Fluid Velocity.....	401
7.3.2.2	Other Means of Measuring Fluid Velocity and Flow	404
7.3.3	Measurement of Linear Position.....	415
7.3.3.1	Introduction	415
7.3.3.2	GPS Background.....	420
7.3.3.3	Use of the Global Positioning System to Locate Objects in Three Dimensions	421
7.3.3.4	GPS System Development.....	423
7.3.3.5	GPS Signals	425
7.3.3.6	GPS Hardware.....	426
7.3.3.7	GPS Tag Applications in Biology	430
7.3.3.8	GPS Summary.....	430
7.3.4	Use of Optical Interferometry to Measure Δx	431
7.3.4.1	Introduction	431
7.3.4.2	Measurement of Tympanal Membrane Displacement by Fiber-Optic Fizeau Interferometer.....	438
7.3.4.3	Measurement of Skin Vibration by Optical Interferometry...	439
7.3.4.4	Two-Frequency, Heterodyne Interferometry	442
7.3.4.5	Fabry–Pérot Interferometer.....	444
7.3.4.6	Discussion	446

7.3.5	Phase-Lock Velocity and Range Sensing with Ultrasound or EM Radiation	446
7.4	Measurement of Force and Torque	454
7.4.1	Load Cells for Force Measurement	454
7.4.2	Torque Measurements	458
7.5	Pressure Measurements	463
7.5.1	Introduction	463
7.5.2	High-Pressure Sensors	463
7.5.3	Low-Pressure Sensors	472
7.6	Introduction to Substance Detection and Measurement Using Photons.....	479
7.6.1	Introduction	479
7.6.2	Dispersive Spectrophotometry	482
7.6.3	Nondispersive Spectrophotometry	488
7.6.3.1	Use of Nondispersive Spectrophotometry to Quantify CO ₂ Gas in Air	494
7.6.3.2	Nondispersive Photoacoustic Spectrophotometry	494
7.6.4	Fourier Transform IR Spectrophotometry	495
7.6.5	Single- and Multiple-Bubble Sonoluminescence.....	501
7.6.5.1	Introduction	501
7.6.5.2	Single-Bubble Sonoluminescence	501
7.6.5.3	Multiple-Bubble Sonoluminescence	504
7.6.5.4	Summary	504
7.7	Other Means of Substance Detection	505
7.7.1	Mass Spectrometry	505
7.7.2	Gas Chromatography	510
7.7.3	Surface Plasmon Resonance	516
7.8	Temperature Measurements.....	522
7.8.1	Temperature Standards.....	522
7.8.2	Some Common Means of Temperature Measurement.....	523
7.8.2.1	Mechanical Temperature Sensors.....	523
7.8.2.2	Electrical and Electronic Temperature Sensors	527
7.9	Chapter Summary.....	537
	Problems.....	538
8.	Basic Electrical Measurements	545
8.1	Introduction	545
8.2	DC Voltage Measurements	545
8.2.1	Types of Electromechanical DC Voltmeters.....	547
8.2.2	Electronic DC Voltmeters	555
8.3	Measurement of Static Electric Fields and the Potential of Charged Surfaces.....	559
8.4	DC Measurements	568
8.4.1	Electromechanical DC Ammeters.....	570
8.4.2	Electronic DC Ammeters.....	571
8.4.2.1	Error Analysis of the Shunt Picoammeter	572
8.4.2.2	Error Analysis of the Feedback Picoammeter.....	574
8.4.3	Magneto-Optic Current Sensors.....	576
8.5	AC Voltage Measurements.....	576
8.5.1	Electromechanical AC Voltmeters.....	577
8.5.2	Analog Electronic AC Voltmeters.....	581

8.5.2.1	AC Amplifier–Rectifier AC Voltmeters	582
8.5.2.2	Peak-Reading Electronic AC Voltmeters.....	583
8.5.2.3	True RMS AC Voltmeter of the Feedback Type	586
8.5.2.4	True RMS AC Voltmeters Using the Direct Conversion Approach.....	588
8.5.3	Measurements of Amplifier Noise Voltages, Noise Factor, and Figure	588
8.6	AC Measurements.....	596
8.6.1	Electromechanical (Analog) AC Ammeters.....	597
8.6.2	Electronic and Magneto-Optical AC Ammeters.....	599
8.7	Magnetic Field Measurements	601
8.8	Phase Measurements	611
8.8.1	Analog Phase Measurements.....	613
8.8.2	Digital Phase Detectors.....	616
8.9	Measurements of Frequency and Period (Time)	618
8.10	Measurement of Resistance, Capacitance, and Inductance	624
8.10.1	Resistance Measurements.....	624
8.10.2	Capacitance Measurements.....	630
8.10.2.1	Q Meter Used for Capacitance Measurement	630
8.10.2.2	Capacitance Measurement by Q/V	634
8.10.3	Inductance Measurements.....	639
8.10.3.1	Voltmeter Method of Estimating Inductance.....	639
8.10.3.2	Use of the Q Meter to Measure Inductance and Q	640
8.11	Vector Impedance Meters	641
8.12	Chapter Summary.....	643
	Problems.....	643
9.	Digital Interfaces in Measurement Systems.....	651
9.1	Introduction	651
9.2	Sampling Theorem.....	652
9.3	Quantization Noise.....	655
9.4	Dithering	658
9.5	DACs	661
9.6	Hold Operation.....	665
9.7	ADCs.....	667
9.7.1	Successive-Approximation ADCs.....	667
9.7.2	Tracking or Servo ADCs	669
9.7.3	Dual-Slope Integrating ADCs	671
9.7.4	Flash (Parallel) ADCs	673
9.7.5	Dynamic Range, Floating Point ADCs	675
9.7.6	Delta–Sigma ADCs	676
9.7.7	Data Acquisition Cards for PCs.....	681
9.8	IEEE-488 Instrumentation Bus (GPIB).....	682
9.8.1	Summary of the GPIB Bus Structure	682
9.8.2	GPIB Operation	684
9.9	Serial Data Communications Links	685
9.9.1	RS-232C and D Interfaces	686
9.9.2	TRS-422, RS-423, and RS-485 Interfaces	688

9.9.3	Universal Serial Bus.....	689
9.9.3.1	Introduction	689
9.9.3.2	How USB Works: An Overview	691
9.9.3.3	Summary	693
9.10	CAMAC (IEEE-583) Modular Instrumentation Standard and the VXI Modular Instrumentation Architecture	693
9.11	How Transmission Lines Affect the Transfer of Digital Data	696
9.11.1	Transmission Line Model	696
9.11.2	Reflections on an Improperly Terminated, Lossless Transmission Line	698
9.12	Data Transmission on Fiber-Optic Cables.....	703
9.12.1	FOC Basics	703
9.12.2	Semiconductor Sources and Detectors Used in FOC Data Transmission.....	711
9.12.3	FOC Systems.....	713
9.13	Virtual Instruments	714
9.13.1	Introduction.....	714
9.13.2	PXI Systems.....	716
9.13.3	Summary.....	716
9.14	Chapter Summary.....	717
	Problems.....	718
10.	Introduction to Digital Signal Conditioning in Instrumentation	725
10.1	Introduction.....	725
10.2	Digital Filters and the z-Transform.....	725
10.3	Some Simple DSP Algorithms.....	731
10.4	Discrete and Fast Fourier Transforms and Their Applications.....	740
10.4.1	Use of Data Windows to Improve Spectral Resolution	741
10.4.2	Use of the DFT to Characterize Random Signals and Noise.....	744
10.4.3	Fast Fourier Transform.....	747
10.5	Digital Routines for Interpolating Discrete Data	749
10.5.1	Estimating Missing Data at the Sampling Instants	752
10.6	Chapter Summary.....	754
	Problems.....	754
11.	Solid-State Chemical Microsensors and Wireless Instrumentation.....	757
11.1	Introduction	757
11.2	Chemical Microsensor Designs	758
11.2.1	Silicon-Based Chemical Microsensors.....	758
11.2.1.1	Schottky Diode-Based Microsensors.....	758
11.2.1.2	MOS ChemFET.....	759
11.2.2	Metal Oxide Gas Sensors.....	763
11.2.2.1	Taguchi Gas Sensors	763
11.2.2.2	Modern Metal Oxide Gas Sensors	764
11.2.2.3	CO ₂ Microsensor Based on Tin Oxide Doped with Copper Oxide.....	767
11.3	Electronic Noses.....	768
11.3.1	Introduction.....	768
11.3.2	Rationale for Artificial Noses.....	768

11.3.3	Sensors for E-Nose Applications	769
11.3.4	Feature Extraction in E-Noses.....	781
11.3.5	Some Commercially Available E-Noses	784
11.4	Radio ICs for Wireless Data Transmission	785
11.4.1	EM Spectrum.....	785
11.4.2	Some UHF Radio Chips.....	786
11.4.3	Antennas for Radio Chips	787
11.4.3.1	Introduction	787
11.4.3.2	Fractal Antennas	788
11.5	Wireless Patient Monitoring Systems	793
11.5.1	Introduction.....	793
11.5.2	RF Communications Protocols Used in WDX and WPM.....	794
11.5.3	Commercial WPM Systems	795
11.6	Power Sources for Wireless Sensors, Effectors, and WDX.....	796
11.7	Absorbable Electronic Circuit Implants.....	799
11.8	Chapter Summary.....	801
	Problems.....	802
12.	Introduction to Mechanical Microsensors	805
12.1	Introduction to Microelectromechanical Systems	805
12.1.1	Nanoelectromechanical Systems.....	805
12.2	MEM Accelerometer and Pressure Sensor ICs	806
12.2.1	MEM Linear Accelerometers.....	806
12.2.2	MEM Pressure Sensors	809
12.3	MEM Rate Gyros	810
12.3.1	Introduction.....	810
12.3.2	Oscillating Disk Rate Gyro.....	810
12.3.3	Vibrating Mass MEM Gyros.....	816
12.3.4	Quartz Tuning Fork MEM Rate Gyro	817
12.3.5	Applications of MEM Gyros.....	817
12.4	Cantilever-Like MEMS and NEMS and Their Applications	819
12.4.1	Introduction.....	819
12.4.2	Applications of Cantilever-Based MEMS Sensors	819
12.4.3	Design of Cantilever MEMS and NEMS	821
12.4.3.1	Cantilever Materials and Fabrication.....	821
12.4.3.2	Cantilever Readout Methods.....	821
12.4.3.3	Static MEM Cantilever Deflection Sensing with Piezoresistors	822
12.4.3.4	Cantilever Beam Resonance Frequency.....	824
12.4.3.5	Excitation of Mechanical Oscillation of MEM Cantilevers.....	825
12.4.3.6	Vibrations in a Beam Clamped at Both Ends.....	826
12.4.3.7	Vibrations in a Circularly Clamped Thin Plate	826
12.4.4	Quartz Tuning Fork Biosensor.....	827
12.4.5	Liquid Viscosity Measurement Using Piezoceramic Resonators	827
12.5	Chapter Summary.....	829
	Problems.....	829

13. Examples of the Design of Measurement Systems	835
13.1 Introduction.....	835
13.2 Self-Nulling, Microdegree Resolution Polarimeter to Measure Glucose in Bioreactors	835
13.3 Design of a System to Detect, Measure, and Locate Partial Discharges in High-Voltage Coaxial Power Cables	845
13.4 Design of a Closed-Loop, Constant-Phase, Pulsed Laser Ranging System and Velocimeter.....	852
13.5 Design of Capacitive Sensors for the Detection of Hidden Objects	858
13.5.1 Introduction.....	858
13.5.2 Self-Balancing Circuits Used to Measure $\Delta C/C_0$	860
13.5.3 Summary.....	864
13.6 Chapter Summary.....	864
Glossary	867
References	881

Preface

Purpose

This text is intended to be used in a classroom course for engineers and scientists. It covers the theory, science, and art of modern instrumentation and measurements (I&M). There is more than enough material to support two semesters' work. Thus, the instructor has the option of choosing those topics and the depth of coverage that suit his or her interests and curriculum. Because of its breadth, *Introduction to Instrumentation and Measurements, Third Edition*, will also be useful as a reference for the practicing engineer and scientist interested in I&M.

Why have a classroom course in I&M? In the United States, over the past three or four decades, many electrical engineering departments have discontinued classroom courses on the theory and practice of I&M. In the past decade, we have also seen the swift development of new and exciting means of measurement using new technologies, the adoption of new standards, and, concurrently, the lack of development of a coherent educational base to support their understanding and use. Using an instrument in the laboratory is not the same as understanding the physical and electronic principles underlying its design, and its functional capabilities and limitations. Clearly, there is now more than ever a need for classroom experience in the new I&M that will give students the necessary technical background to use and design sensors, signal conditioning systems, and I&M systems. I feel that this text supports that need.

This text was written based on my 40 years of experience in teaching a classroom course (EE 230) on electrical instrumentation to juniors and seniors in the Electrical and Computer Engineering Department at the University of Connecticut, Storrs. Obviously, in more than 40 years, we have seen the field of I&M evolve with the rest of electrical engineering technology. Because of the rapid pace of technical development, it generally has been difficult to find an up-to-date text for our electrical instrumentation course. After years of frustration in trying to match a text to course content, I decided to write one that would not only encompass the *traditional* aspects of I&M, but also include material on such topics as modern integrated circuit (IC) and photonic sensors, micro-electromechanical (MEM) and nano-electromechanical (NEM) sensors, chemical and radiation sensors, signal conditioning, noise, data interfaces, basic digital signal processing (DSP), etc.

Reader Background

The reader is assumed to have taken core EE curriculum courses or their equivalents. He or she should be skilled in basic, linear circuit theory, that is, should have mastered Thevenin's and Norton's theorems, Kirchoff's laws, superposition, dependent sources, and ideal op-amps, and should know how to describe DC and AC steady-state circuits in terms of linear loop and

node equations. An *introductory systems course* should have given him or her familiarity with both time- and frequency-domain methods of describing linear dynamic systems characterized by ordinary, linear, differential, or difference equations, including state space, Fourier, Laplace and z -transforms, transfer functions, steady-state frequency response of systems, as well as Bode plots. From physics or an *EE course in electromagnetics*, the reader should have a basic knowledge of electric and magnetic fields, inductance, capacitance, reluctance, transformers, etc. There should also be some familiarity with electromagnetic waves, Maxwell's equations, transmission lines, and polarization. From a *first course in electronics*, there should be basic knowledge of bipolar junction transistors (BJTs), junction field-effect transistors (JFETs), diodes, and photodiodes and their simple linear circuit models.

Scope of the Text

A major feature of this book is its breadth of coverage. Throughout the text, a high level of mathematical analytical detail is maintained. It is not a *picture book*; we have assumed that readers have already had contact with basic electrical instruments, including oscilloscopes and meters in their introductory EE and physics labs.

In the following paragraphs, we give an overview of the contents:

Chapter 1, *Measurement Systems*, is an introductory chapter, which illustrates typical measurement system architecture and describes sensor dynamics, signal conditioning, and data display and storage. Errors in measurements are discussed, including the meaning of accuracy and precision, limiting error, etc. The recent (1990) quantum standards adopted for the volt and the ohm are described, as well as other modern electrical and physical standards.

In Chapter 2, *Analog Signal Conditioning in Instrumentation*, we describe, largely at the systems level, means of conditioning the analog outputs of various sensors. Op-amps, differential, instrumentation, autozero, and isolation amplifiers are covered. Applications of op-amps in active filters, differential instrumentation amplifiers, charge amplifiers, phase-sensitive rectifiers, etc., are shown. We also give practical consideration to errors caused by offset voltage, bias currents, input impedance, slew rate and gain \times bandwidth product, etc. There is also a section on nonlinear signal processing with op-amps.

Noise and Coherent Interference in Measurements are treated in depth in Chapter 3. A heuristic yet rigorous approach is used in which we define and use one-sided, noise voltage and current power density spectra to describe the effect of noise in instruments and measurement systems. Noise factor and figure are defined, and output signal-to-noise ratios are used to evaluate measurement system resolution. Examples are given of calculations of the noise-limited resolution of the quantity under measurement (QUM). Techniques are shown for the minimization of coherent interference.

The traditional topics of *DC Null Methods of Measurement* and *AC Null Measurements* are covered in Chapter 4 and Chapter 5, respectively. Wheatstone and Kelvin bridges and potentiometers are described in Chapter 4, and the major AC bridges used to measure inductance, Q , capacitance, and D are treated in Chapter 5. Material in this chapter includes a description and analysis of the *Anderson current loop* method of reading sensor outputs.

A *Survey of Sensor Mechanisms* is presented in Chapter 6. This is a large and substantive chapter covering a broad range of sensor mechanisms and types. Of special note is the introduction of certain fiber-optic and electro-optic sensors, as well as selected chemical

and ionizing radiation sensors. The Sagnac effect is introduced, and the basic fiber-optic gyro is described. New material in Chapter 6 includes a description and analysis of sensors based on the giant magnetoresistive effect (GMR) and the anisotropic magnetoresistive (AMR) effect. Pyroelectric IR sensors are also introduced. Means of measuring the rotation of linearly polarized light is presented. A substantive section on photomultiplier tubes and channel-plate photomultipliers is also provided.

Chapter 7, *Applications of Sensors to Physical Measurements*, presents a detailed analysis of mechanical gyroscopes, clinometers, and accelerometers. It covers the Doppler effect in ultrasonic velocimetry and laser Doppler velocimetry. It also provides an introductory section on the global positioning system (GPS), a section on optical interferometry, and an extensive introduction to spectrophotometry, sonoluminescence, and surface plasmon resonance used for substance quantification. The measurement of force, pressure, and torque is also covered.

In Chapter 8, *Basic Electrical Measurements*, the classic means of measuring electrical quantities are presented, as well as newer methods such as Faraday magneto-optic ammeters and Hall effect gaussmeters and wattmeters. Electronic means of measuring stored charge and static electric fields are also described.

Digital Interfaces in Measurement Systems are covered in Chapter 9. It begins with a description of the sampling theorem, aliasing and quantization. The traditional topics of hold circuits, digital-to-analog convertors (DACs), and many types of ADCs are covered. The chapter also deals with data buses. New material includes a section on dithering as a means to reduce quantization noise, a section on delta-sigma ADCs, and a section on the ubiquitous universal serial bus (USB). Virtual instruments and PCI eXtensions for Instrumentation (PXI) systems are also described.

Chapter 10, *Introduction to Digital Signal Conditioning in Instrumentation*, was written to acquaint the reader with this specialized field because digitized, measured data is processed and stored on computers in modern instrumentation practice. The z-transform and its use in describing filtering operations on discrete, digitized data in the frequency domain are introduced. Examples of FIR and IIR digital filters are given, including numerical integration and differentiation routines, viewed both in the time and frequency domain. The discrete and fast Fourier transforms are covered, and the effect of data windows on spectral resolution is discussed. Finally, the use of splines in interpolating discrete data sequences and in estimating missing data points is described.

An all-new Chapter 11, *Solid-State Chemical Microsensors and Wireless Instrumentation*, has been written to address these contemporary topics. Modern tin oxide gas sensors are described, as well as ChemFETs, ISFETs, and Schottky-diode-based chemical microsensors. Electronic noses (E-noses) are introduced. Wireless data transmission (WDX) protocols are described along with certain radio chips and their antennas, including a new section describing broadband, space-saving, fractal antennas. Energy-harvesting ICs and supercapacitors are described, as well as absorbable electronic circuits for temporary implants in living systems.

Chapter 12, *Introduction to Mechanical Microsensors*, is another all-new chapter covering mechanical microsensors (MEMS and NEMS). It covers micromachined electromechanical accelerometer designs, several MEM rate gyro designs, and cantilever-based MEMS. Resonant cantilevers are shown to make effective chemisensors.

In Chapter 13, *Examples of the Design of Measurement Systems*, four examples of complex measurement systems developed by my students and myself are given to illustrate design philosophy: (1) a self-nulling microdegree polarimeter to measure glucose concentration; (2) a system to detect and locate partial discharges on underground, high-voltage power cables; (3) the design of a laser velocity and distance measuring system; and (4) the design of capacitance sensors to detect hidden objects.

Home Problems

Chapters 1 through 12 are followed by problems taken from my extensive classroom experience in teaching courses in I&M at the University of Connecticut. The problems are doable; they are student tested. A home problem solutions manual is available from the publisher.

Glossary

This book has a comprehensive glossary covering the acronyms and abbreviations used in the broad field of I&M. It also describes and defines many terms used in the text.

References and Bibliography

The references cited encompass a wide time span, from the 1950s to the present. There are many recent entries of review articles and specialized texts that should lead the reader interested in pursuing a specialized area of I&M further into a particular field.

Index

A complete index allows the reader to access topics featured and not featured in the contents.

Features

Every chapter in this book has been revised to reflect modern technology. Traditional material has been retained, however. Two new chapters have been added containing contemporary material. They expand the scope of the text to include geophysical, chemical, micromechanical, and photonic instrumentation. Some of this unique new material includes the following:

- The Anderson current loop technology for conditioning the outputs of remote resistive and capacitive sensors is found in Chapter 4.
- The design of optical polarimeters and their application to polarization-responding sensors.
- Photonic measurements with photomultipliers and channel-plate photon sensors.

- The Sagnac effect and fiber-optic gyroscopes are introduced as a sensitive means of measuring angular velocity.
- Vibrating mass and vibrating disk rate gyros are introduced. The novel Humphrey air jet gyro is analyzed. Traditional pendulum as well as fluid-filled clinometers are described.
- The global positioning system (GPS) and the various modifications to improve its accuracy are described.
- Substance detection using photons is introduced in Chapter 7. Dispersive, nondispersive, and Fourier transform spectroscopy are described, as well as sonoluminescence and surface plasmon resonance.
- Chapter 9 on digital interfaces has new sections on dithering, delta-sigma ADCs, data acquisition cards, the USB, and virtual instruments and PXI systems.
- An all-new Chapter 11 describes solid-state chemisensors, including tin oxide gas sensors, Schottky diode chemisensors, chemFETs, ISFETs, and *E*-noses and introduces radio ICs for use in WDX. Broadband, compact, fractal antennas are also described.
- A new Chapter 12 covers micromachined IC accelerometers and rate gyros and describes resonant MEM cantilever chemisensors.

MATLAB® is a registered trademark of The MathWorks, Inc. For product information, please contact:

The MathWorks, Inc.
3 Apple Hill Drive
Natick, MA 01760-2098 USA
Tel: 508-647-7000
Fax: 508-647-7001
E-mail: info@mathworks.com
Web: www.mathworks.com

Author

Robert B. Northrop, PhD, was born in White Plains, New York, in 1935. After graduating from Staples High School in Westport, Connecticut, he majored in electrical engineering (EE) at MIT, graduating with a bachelor's degree in 1956. At the University of Connecticut, he earned his master's degree in electrical and systems engineering in 1958. As the result of a long-standing interest in physiology, he enrolled in a PhD program at UCONN in physiology, doing research on the neuromuscular physiology of molluscan catch muscles. He received his PhD in 1964.

In 1963, he rejoined the UCONN EE Department as a lecturer and was hired as an assistant professor of EE in 1964. In collaboration with his PhD advisor, Dr. Edward G. Boettiger, Dr. Northrop secured a five-year training grant in 1965 from NIGMS (NIH), and started one of the first, interdisciplinary, biomedical engineering graduate training programs in New England. UCONN currently awards MS and PhD degrees in this field of study, as well as BS degrees in engineering under the BME area of concentration.

Throughout his career, Dr. Northrop's research interests have been broad and interdisciplinary and have been centered on biomedical engineering and physiology. He has done research (sponsored by the US Air Force Office of Scientific Research [AFOSR]) on the neurophysiology of insect and frog vision and devised theoretical models for visual neural signal processing. He also did sponsored research on electrofishing and developed, in collaboration with Northeast Utilities, effective working systems for fish guidance and control in hydroelectric plant waterways on the Connecticut River at Holyoke, Massachusetts, using underwater electric fields.

Still another area of Dr. Northrop's sponsored research (by NIH) has been in the design and simulation of nonlinear, adaptive digital controllers to regulate in vivo drug concentrations or physiological parameters, such as pain, blood pressure, or blood glucose in diabetics. An outgrowth of this research led to his development of mathematical models for the dynamics of the human immune system, which were used to investigate theoretical therapies for autoimmune diseases, cancer, and HIV infection.

Biomedical instrumentation has also been an active research area for Dr. Northrop and his graduate students: An NIH grant supported studies on the use of the ocular pulse to detect obstructions in the carotid arteries. Minute pulsations of the cornea from arterial circulation in the eyeball were sensed using a no-touch, phase-locked, ultrasound technique. Ocular pulse waveforms were shown to be related to cerebral blood flow in rabbits and humans.

More recently, Dr. Northrop addressed the problem of noninvasive blood glucose measurement for diabetics. Starting with a Phase I SBIR grant, he developed a means of estimating blood glucose by reflecting a beam of polarized light off the front surface of the lens of the eye and measuring the very small optical rotation resulting from glucose in the aqueous humor, which in turn is proportional to blood glucose. As an offshoot of techniques developed in micropolarimetry, he developed a magnetic sample chamber for glucose measurement in biotechnology applications. The water solvent was used as the Faraday optical medium.

He has written numerous papers in peer-reviewed journals, and twelve textbooks: *Analog Electronic Circuits* (Addison Wesley, 1990), and the following books published by CRC Press: *Introduction to Instrumentation and Measurements* (1997), *Endogenous and Exogenous Regulation*

and Control of Physiological Systems (2000), *Dynamic Modeling of Neuro-Sensory Systems* (2001), *Noninvasive Instrumentation and Measurements in Medical Diagnosis* (2002), *Analysis and Application of Analog Electronic Circuits in Biomedical Engineering* (2004), *Introduction to Instrumentation and Measurements—2nd edition* (2005), *Introduction to Molecular Biology, Genomics & Proteomics for Biomedical Engineers* (with Anne N. Connor) (2008), *Signals and Systems Analysis in Biomedical Engineering—2nd edition* (2010), *Introduction to Complexity and Complex Systems* (2011), *Analysis and Application of Analog Electronic Circuits in Biomedical Engineering—2nd edition* (2012), and *Ecological Sustainability: Understanding Complex Issues* (with Anne N. Connor) (2013).

Dr. Northrop is a member of Sigma Xi, Phi Kappa Phi, Eta Kappa Nu, and Tau Beta Pi and a founding fellow, Connecticut Academy of Engineers (2003).

His current research interest lies in complex systems and sustainability.

Dr. Northrop was on the Electrical & Systems Engineering faculty at UCONN until his retirement in June 1997. Throughout this time, he was director of the Biomedical Engineering Graduate Program. As emeritus professor, he teaches graduate courses in biomedical engineering, writes texts, sails, and travels. He lives in Chaplin, Connecticut, with his wife and a smooth fox terrier.

1

Measurement Systems

1.1 Introduction

In this introductory chapter, we will examine the architecture of *typical*, modern measurement systems and discuss how noise, calibration errors, sensor dynamic response, and nonlinearity can affect the *accuracy*, *precision*, and *resolution* of measurements. We will also discuss the contemporary physical and electrical standards used by the National Institute of Standards and Technology (US NIST, formerly the US National Bureau of Standards [NBS]) and discuss how these standards are used to create secondary standards used for practical calibration of measurement systems.

Measurement systems are traditionally used to measure physical and electrical quantities, such as mass, temperature, force, pressure, velocity, angular velocity, acceleration, capacitance, and voltage. However, they also can be designed to locate things or events, such as the epicenter of an earthquake, employees in a building, partial discharges (PDs) in a high-voltage (HV) power cable, or a land mine. Often a measurement system is called upon to discriminate and count objects, such as red blood cells, fish of a certain size swimming past a check point, or bottles of beer on a conveyor. Also, a measurement system may have the task of measuring the concentrations of chemical substances, such as various ions in solution such as H^+ (pH), Na^+ , K^+ , Cl^- , or SO_4^- , the concentrations of antibodies or other biological molecules in solution, or the concentrations of certain explosive or toxic gasses such as CH_4 or carbon monoxide. A measurement system is often made part of a control or regulatory system. The old saying that “if you can’t measure it, you can’t control it” is certainly a valid axiom for the control engineer as well as for the instrumentation engineer.

The reader should realize that the field of instrumentation and measurements (I&M) is rapidly changing, and new standards, sensors, and measurement systems are continually being devised and described in the journal literature. The *IEEE Transactions on Instrumentation and Measurement*, the *Review of Scientific Instruments*, the *IEEE Transactions on Biomedical Engineering*, and the *Journal of Scientific Instruments* are four of the more important periodicals dealing with the design of new measurement systems, instruments, and standards.

1.2 Measurement System Architecture

Figure 1.1 illustrates the block diagram of a typical measurement system. The quantity under measurement (QUM) is converted to a useful form, such as a voltage, current, or physical displacement by an input *transducer* or *sensor*. (Note the distinction between a sensor and

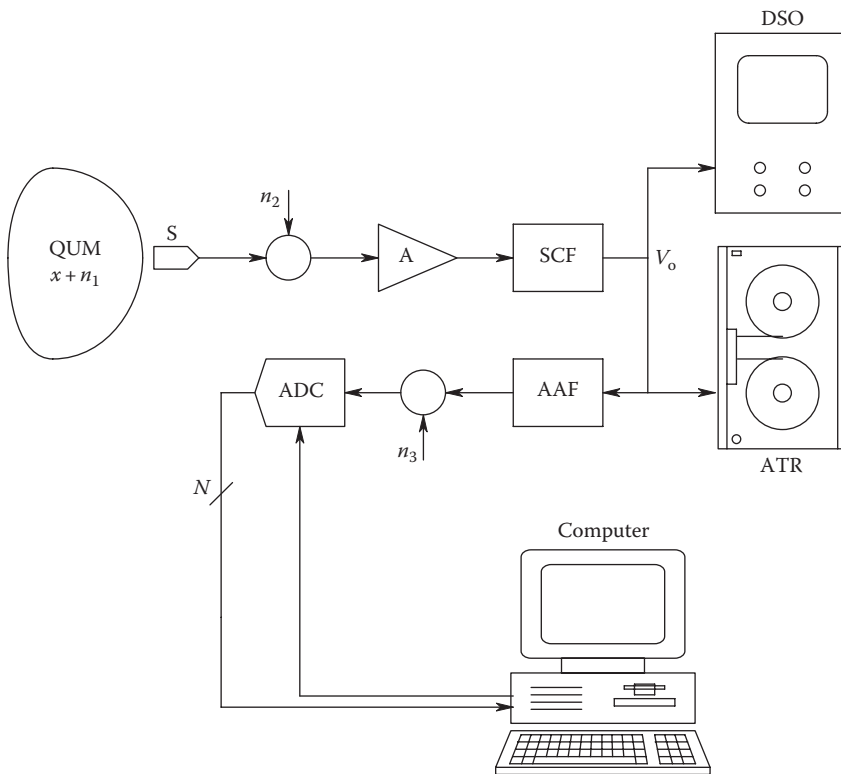


FIGURE 1.1

A generalized measurement system. *Note:* QUM = x = quantity under measurement. S = sensor. A = amplifier. SCF = signal conditioning filter (analog). DSO = digital storage oscilloscope. ATR = analog tape recorder. AAF = anti-aliasing filter (analog). n_1 = noise accompanying the QUM. n_2 = equivalent noise from electronics. n_3 = equivalent quantization noise. ADC = analog-to-digital converter. The digital computer can further filter and store digitally V_o , the conditioned sensor output: $V_o = x + n_1 + n_2$.

a transducer: both devices are sensors; however, transducers have the property of being able to convert an input signal to some analog quantity.) An example of a transducer is a *piezoelectric crystal*. A mechanical displacement (or, equivalently, a force or pressure) applied to the crystal produces an output voltage; an input voltage to the crystal produces a mechanical displacement. Another example of a transducer is a *loudspeaker*: a current through the voice coil causes cone displacement; the rate of mechanical displacement generates a proportional current output because the voice coil moves in a magnetic field.

The output of the sensor, here assumed to be a voltage, is amplified and filtered by a signal conditioning subsystem. The purposes of the signal conditioning subsystem are to amplify, give a low or matched output impedance, and improve the signal-to-noise ratio (SNR) of the analog signal proportional to the QUM. The conditioned analog signal, V_o , can be distributed to various display and recording devices. V_o can be displayed on an analog or digital oscilloscope or a strip-chart recorder. It can also be recorded on a magnetic tape recorder, or it can be low-pass filtered to prevent aliasing and then periodically converted to digital words by an *analog-to-digital converter* (ADC). As you will see, the sensor output can be wirelessly communicated (from sensor S to amplifier A) to the rest of the measurement system by modulated radio frequency or light beam.

The ADC output, D_N , is generally made an input to a digital computer through an appropriate A/D interface. Once in the computer environment, the sampled and digitized QUM can be further (digitally) filtered and processed and stored on a magnetic hard and/or floppy disk or on an optical disk in digital form. Often the sensing instrument contains an ADC and is connected to a computer by a special instrumentation bus such as the USB, GPIB, or IEEE-488 bus (Hewlett-Packard [HP] calls the GPIB the HPIB). The General Purpose Instrumentation Bus (GPIB) is organized so that many GPIB-compatible sensing instruments can be managed by one master (host) computer. More will be said about the GPIB in Chapter 9.

Note that there are three major sources of *noise* in the *typical* measurement system shown in Figure 1.1: noise sensed along with the QUM (environmental noise), noise associated with the electronic signal conditioning system (referred to its input), and equivalent noise generated in the analog-to-digital conversion process (quantization noise). These noise sources generally provide a limit to measurement system resolution and to its accuracy.

Many sensors provide the rate-limiting element in a measurement system. That is, the response speed of the sensor is much slower than any of the other elements in the system. As a result of sensor dynamics, a measurement system can require a minimum settling time before a stable measurement can be made when there is a step change in the QUM. That is, a *settling time* is required for the sensor to reach a steady-state (SS) output.

1.2.1 Sensor Dynamics

Sensor dynamics generally fall into one of three categories: *first-order low-pass*, *second-order low-pass*, and *band-pass*. A general example of a first-order, low-pass sensor is shown in Figure 1.2A. Here, the QUM is $x(t)$, and the sensor output is a voltage, $v_x(t)$. The ordinary differential equation (ODE) relating v_x and x is

$$\dot{v}_x(t) = -av_x(t) + Kx(t) \tag{1.1a}$$

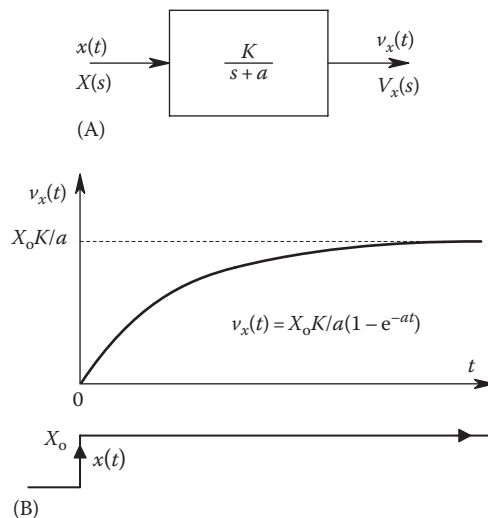


FIGURE 1.2

(A) Transfer function of a sensor with first-order, low-pass dynamics in Laplace format. $x(t)$ is the QUM; $X(s)$ is its Laplace transform. $v_x(t)$ is the sensor's output voltage, $V_x(s)$ is its Laplace transform, where a is the natural frequency in r/s of the sensor. (B) Time domain step response of the first-order sensor.

or

$$\dot{v}_x(t)\tau = -v_x(t) + K\tau x(t), \quad (1.1b)$$

where

a is the natural frequency of the sensor in r/s

$\tau = 1/a$ is its time constant in s

The solution of the ODE previously given for a step input, $x(t) = X_o U(t)$, gives

$$v_x(t) = \frac{X_o K}{a} (1 - e^{-at}) = X_o K \tau \left[1 - \exp\left(\frac{-t}{\tau}\right) \right], \quad t \geq 0. \quad (1.2)$$

The sensor step response is plotted in Figure 1.2B. Note that the response is within 5% of the SS value for $t = 3\tau$ and within 2% of SS for $t = 4\tau$. The transfer function for the first-order, low-pass sensor is found by Laplace transforming the ODE, Equation 1.1a,

$$\frac{V_x(s)}{X(s)} = \frac{K}{s+a} = H(s), \quad (1.3)$$

and its SS, sinusoidal frequency response is found by letting $s = j\omega$ in Equation 1.3.

Second-order sensor dynamics fall into one of three categories, depending on the location of the roots of the characteristic equation of the ODE. These categories are *underdamped* (with complex-conjugate roots in the left-half s -plane), *critically damped* (with two, equal real roots on the negative real axis in the s -plane), and *overdamped* (with unequal real roots on the negative real axis in the s -plane). In all cases, the real parts of the ODE's roots are negative, as the sensor is assumed stable. The second-order ODEs for the three damping conditions are as follows:

$$\ddot{v}_x = -\dot{v}_x(2\zeta\omega_n) - v_x(\omega_n^2) + Kx(t) \quad (\text{underdamped: } 0 \leq \zeta < 1.) \quad (1.4a)$$

$$\ddot{v}_x = -\dot{v}_x(2a) - v_x a^2 + Kx(t) \quad (\text{critically damped: } \zeta = 1.) \quad (1.4b)$$

$$\ddot{v}_x = -\dot{v}_x(a+b) - v_x ab + Kx(t) \quad (\text{overdamped: } \zeta > 1.) \quad (1.4c)$$

In the underdamped case, Equation 1.4a, ζ is the *damping factor* and ω_n is the undamped natural frequency in r/s. For an underdamped sensor, $0 \leq \zeta < 1$, and the roots (poles) of the ODE are located in the s -plane at $s = -\omega_n \pm j\omega_n \sqrt{1-\zeta^2}$. The critically damped sensor has two roots at $s = -\omega_n$, and the overdamped sensor's ODE has roots at $s = -a$ and $s = -b$.

These conditions are shown in Figure 1.3. Using Laplace transforms, we can find the step response of the second-order sensor for the three preceding conditions. The step response for the underdamped case is

$$v_x(t) = \frac{KX_o}{\omega_n^2} \left\{ 1 - \frac{1}{\sqrt{1-\zeta^2}} \exp(-\zeta\omega_n t) \sin \left[\omega_n \sqrt{(1-\zeta^2)} t + \phi \right] \right\}, \quad (1.5)$$

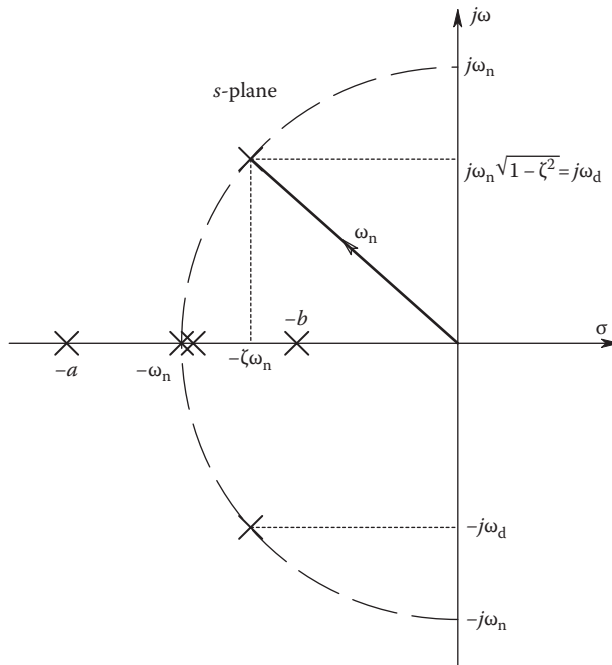


FIGURE 1.3

s-plane pole positions of a second-order, low-pass sensor's transfer function. For overdamping ($\zeta > 1$), the poles are real at $s = -a$ and $-b$. For critical damping ($\zeta = 1$), the poles are both at $s = -\omega_n$, and for underdamping ($0 < \zeta < 1$), the two poles are complex conjugate at $s = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2}$.

where

$$\phi = \tan^{-1}\left(\frac{\sqrt{1-\zeta^2}}{\zeta}\right). \tag{1.6}$$

The underdamped, second-order system step response is illustrated in Figure 1.4A. The step response of the critically damped sensor ($\zeta = 1$) is shown in Figure 1.4B and is

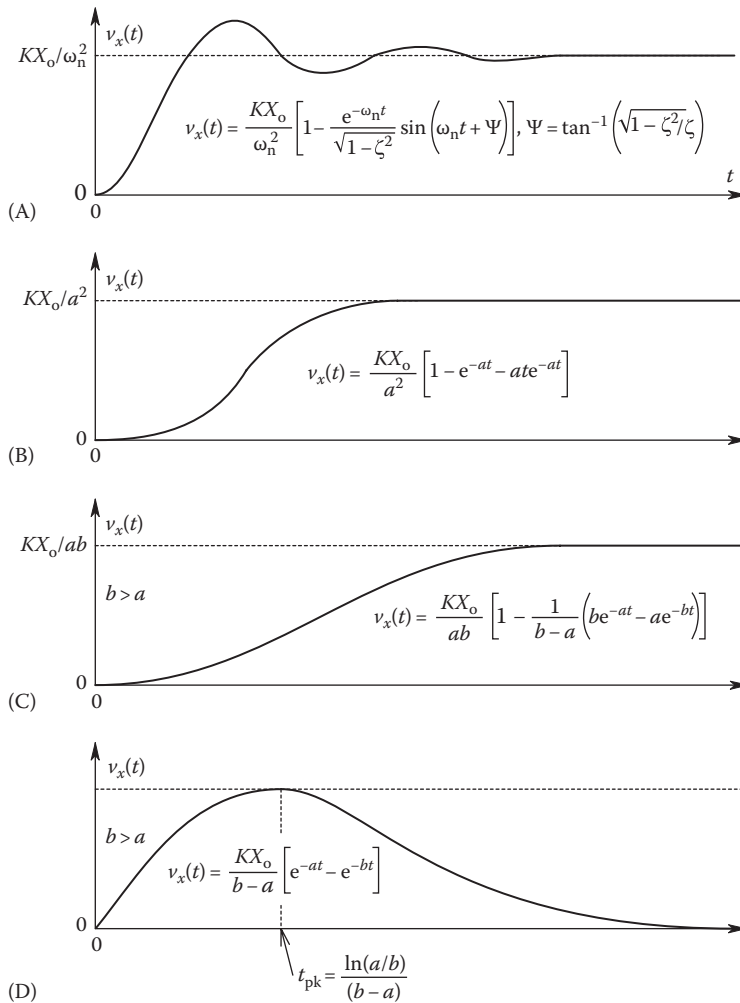
$$v_x(t) = \frac{KX_o}{a^2}(1 - e^{-at} - ate^{-at}). \tag{1.7}$$

Finally, the step response of an overdamped sensor having two real poles (at $s = -a$ and $s = -b$) is shown in Figure 1.4C and can be written as

$$v_x(t) = \frac{KX_o}{ab}\left[1 - \frac{1}{b-a}(be^{-at} - ae^{-bt})\right], \quad b > a. \tag{1.8}$$

While the three types of low-pass sensor dynamics described earlier are most commonly encountered, sensors exist, which do not respond to a constant (DC) QUM. The QUM must in fact be time varying to produce a sensor output. Such sensors are said to have a band-pass frequency response characteristic and can be described by the ODE:

$$\ddot{v}_x = -\dot{v}_x(a+b) - v_x(ab) + K\dot{x}, \quad b > a. \tag{1.9}$$

**FIGURE 1.4**

Step response for different types of second-order sensors: (A) Underdamped ($0 < \zeta < 1$) with complex-conjugate poles. (B) Critically damped ($\zeta = 1$). (C) Overdamped ($\zeta > 1$) with real poles at $s = -a$ and $-b$. (D) Overdamped, band-pass sensor (one zero at $s = 0$).

The corresponding transfer function for a band-pass sensor with two real poles is

$$\frac{V_x(s)}{X(s)} = \frac{Ks}{(s+b)(s+a)}. \quad (1.10)$$

The output voltage step response of a band-pass sensor rises from zero to a peak and then goes to zero in the SS; it is illustrated in Figure 1.4D and is given by

$$v_x(t) = \frac{KX_o}{b-a} (e^{-at} - e^{-bt}). \quad (1.11)$$

A piezoelectric crystal pressure transducer is an example of band-pass sensor.

1.2.2 Overview of Signal Conditioning

The voltages or currents obtained directly at the output of a sensor are generally low level and contain additive noise and coherent interference picked up from the environment of the QUM and from the sensor itself. Sometimes the measurement process or the sensor introduces a nonlinear distortion of the QUM (as in the case of *hot wire anemometers*), which must be linearized. The analog signal conditioning module following the sensor thus must amplify the sensor output voltage and perform linear filtering on it in order to improve the SNR. Such amplification and filtering is usually performed by a low-noise instrumentation amplifier followed by op-amp (OA) active filters. Compensation for inherent nonlinearities, including corrections for temperature changes, is most easily done digitally, with a computer. Thus, the sensor output, after analog conditioning, is converted to digital form by an ADC and read into the computer as a binary number. If the digitization (sampling) is done on a periodic basis, it is common practice to use the analog signal conditioning system to low-pass filter (LPF) the input to the ADC so that the ADC input voltage contains no significant energy having frequencies in excess of one-half of the sampling frequency. This low-pass filtering before sampling is called *anti-aliasing filtering* and is necessary for accurate digital signal processing (DSP). Sampling and anti-aliasing filters are discussed in detail in Chapter 10.

Digital signal conditioning can also be used to effectively remove coherent interference accompanying the QUM and to calculate functions of the sampled signal such as its root-mean-square (RMS) value, its autocorrelation function, or its root power spectrum. It is now common practice to store digitized measurement records on floppy or hard magnetic disks, on optical disks, or on video magnetic tape in digital form.

1.3 Errors in Measurements

Errors in measurements can arise from many causes; there are remedies for some types of errors, but others haunt us as intrinsic properties of the measurement system under use, and often can be mitigated by system redesign.

A. *Gross errors* in a measurement can arise from such human mistakes as follows:

1. Reading the instrument before it has reached its SS. Such a premature reading produces a *dynamic error*.
2. Not eliminating parallax when reading an analog meter scale and incorrect interpolation between analog meter scale markings.
3. Mistakes in recording measured data and in calculating a derived measurand.
4. Misuse of the instrument. A simple example of this source of error is when a 10 V full-scale analog voltmeter of typical sensitivity (20,000 Ω /V) is connected to a voltage source having a high Thevenin resistance (e.g., 100,000 Ω). Significant voltage is dropped across the internal (Thevenin) resistor.

B. *System errors* can arise from such factors as follows:

1. The instrument is not calibrated and has an offset; that is, its sensitivity is off, and it needs zeroing. Loss of calibration and zero can occur because of long-term component value changes with aging or changes associated with temperature rise.

2. Reading uncertainty from the presence of *random noise*. This noise can accompany the measurand and can arise from the signal conditioning electronics in the system.
 - a. External noise from outside is called *environmental noise* and often can be reduced by appropriate electric and magnetic shielding and proper grounding and guarding practices. Environmental noise is often coherent in nature and can come from poorly shielded radio-frequency sources, such as computers or radio stations. Powerline frequency electric and magnetic fields can also be troublesome, as can be ignition noise.
 - b. The significant, internally generated random noise may be shown to arise in the first stage of an instrument's signal conditioning amplifier. Some of this noise comes from resistors (Johnson or thermal noise), some comes from the active elements (transistors) in the headstage, and some comes from the quantization or roundoff inherent in the operation of ADCs in modern digital instruments.
3. *Slow or long-term drift* in the system can destroy the certainty of static measurements, that is, measurements of measurands, which are not varying in time. Drifts can cause slow changes of system sensitivity and/or zero. Drift may arise as the result of a slow temperature change as a system *warms up*. Many parameter values can change with temperature, such as capacitance, resistance, inductance, the electromotive force (EMF) of standard cells, and the avalanche voltage of pn junctions. Humidity changes can also affect circuit parameter values, including resistance and capacitance. Good system design involves temperature regulation of the system and the use of low or zero temperature coefficient (tempco), sealed components. In integrators such as charge amplifiers, output drift can also be caused by the integration of the OA's DC bias current.

Drift or system offset can also arise from DC static charges affecting analog electronic circuits. In some cases, the face of an analog meter can become charged with static electricity, which attracts or repels the meter pointer, causing an error.

In the following, we discuss the concepts of *accuracy*, *precision*, *resolution*, *limiting error (LE)*, and various error statistics. To begin our quantitative discussion of errors in measurements, we define the *error* in the *n*th measurement as

$$\varepsilon_n \equiv X_n - Y_n \quad (1.12)$$

and

$$\% \varepsilon \equiv \left| \frac{\varepsilon_n}{Y_n} \right| 100, \quad (1.13)$$

where

Y_n is the actual, true, defined, or calculated value of the QUM
 X_n is the *n*th *measured value* of the QUM

Philosophically, the use of Y_n in these may present an ontological problem because one can argue that the *true* value can never be known, as it is the result of a nonideal measurement process. In which case, we might take Y_n to be defined by a high-resolution, primary standard. Nevertheless, error as given in Equation 1.12 is generally an accepted definition. In some cases, the absolute value signs in the $\% \epsilon$ expression may be omitted, and the $\%$ error can be negative.

The *accuracy*, A_n , of the n th measurement is defined as

$$A_n \equiv 1 - \left| \frac{Y_n - X_n}{Y_n} \right|. \quad (1.14)$$

The percent accuracy is simply

$$\% A_n \equiv 100 - \% \epsilon = A \times 100. \quad (1.15)$$

The *precision* of the n th measurement is defined as

$$P_n \equiv 1 - \left| \frac{X_n - \bar{X}}{\bar{X}} \right|. \quad (1.16)$$

\bar{X} is defined as the *sample mean* of N measurements:

$$\bar{X} \equiv \frac{1}{N} \sum_{n=1}^N X_n. \quad (1.17)$$

Low noise accompanying the QUM and in the signal conditioning system is required as a necessary (but not sufficient) condition for *precision measurements*. Precision measurements also require a measurement system capable of resolving very small changes in the QUM, say one part in 10^7 . Consequently, many precision, high-resolution measurement systems rely on null methods, such as used with AC and DC bridges and potentiometers. The high resolution inherent in null measurements comes from the fact that the bridge components themselves are known to high precision, and the null sensor is capable of resolving very small voltage or current differences around null. The null condition equation(s) allows calculation of the QUM to high precision, given high-accuracy components (parameters).

Accurate measurements require the use of a precision measurement system, which is *calibrated* against a certified, accurate standard.

Other statistics used to describe the quality of a series of measurements are the *deviation*, d_n ; the *average deviation*, D_N ; and the *standard deviation*, S_N , defined as follows:

$$d_n \equiv X_n - \bar{X}. \quad (1.18)$$

$$D_N \equiv \frac{1}{N} \sum_{n=1}^N d_n. \quad (1.19)$$

$$S_N \equiv \frac{1}{N} \sum_{n=1}^N d_n^2 = \sigma_x. \quad (1.20)$$

The *variance* of X is simply S_N^2 and can be shown to be equal to

$$S_N^2 \equiv \frac{1}{N} \sum_{n=1}^N X_n^2 - (\bar{X})^2 = \sigma_x^2. \quad (1.21)$$

If we assume that samples of the QUM, X_n , obey a normal or Gaussian probability density function (PDF), then the probable error, pe , in any one observation is defined such that there is a 50% probability that X_n lies between $\bar{X} - pe$ and $\bar{X} + pe$. From the normal PDF, it is easy to show that $pe = 0.6745\sigma_x$ to satisfy the preceding condition.

A noisy set of measurements will have a small X/σ_x , which may be considered to be an RMS SNR. If we increase the total number of measurements, N , the SNR can be shown to improve by a factor of \sqrt{N} .

The *resolution* in measuring the QUM is related to the precision of the measurement and is basically the smallest unit of the QUM, which can be reliably detected. Resolution can be expressed statistically.

Analog-indicating instruments with scales covering about 90° of arc provide a fundamental limitation to measurement system accuracy, even when null measurements are performed. Large, mirror-scale analog meters with knife-edge pointers can usually be read to no better than 0.2%, or 2 parts/thousand. Oscilloscopes generally offer slightly poorer resolution. As a result of this physical limitation to resolution, direct-reading instruments with digital outputs are used when accuracies better than 0.2% are needed. Null instruments have taps and verniers to permit high resolution.

LE is an important parameter used in specifying instrument *accuracy*. The LE, or guarantee error, is given by manufacturers to define the outer bounds or *worst-case* expected error. For example, a certain voltmeter may be specified as having an accuracy of 2% of its full-scale reading. If on the 100 V scale, the meter reads 75 V, the LE in this reading is $(2/75) \times 100 = 2.67\%$.

In many cases, such as in the determination of a resistor value by Wheatstone bridge, the QUM must be *calculated* from a formula in which various system parameters, each having a specified accuracy, appear. Thus, we must derive a formula for the determination of the LE in the value of the calculated QUM. Let the QUM be a function of N variables, that is,

$$Q = f(X_1, X_2, \dots, X_N). \quad (1.22)$$

Let us assume each variable, X_j , is in error by $\pm\Delta X_j$. Hence, the calculated QUM will be *noisy* and will be given by

$$\hat{Q} = f(X_1 \pm \Delta X_1, X_2 \pm \Delta X_2, \dots, X_N \pm \Delta X_N). \quad (1.23)$$

\hat{Q} can be expanded into a Taylor's series for N variables. For one variable,

$$f(X \pm \Delta X) = f(X) + \frac{df}{dX} \frac{\Delta X}{1!} + \frac{d^2 f}{dX^2} \frac{(\Delta X)^2}{2!} + \dots + \frac{d^{n-1} f}{dX^{n-1}} \frac{(\Delta X)^{n-1}}{(n-1)!} + R_n. \quad (1.24)$$

Hence, for the N variable case,

$$\begin{aligned}\hat{Q} = f(X_1, X_2, \dots, X_N) &+ \left\{ \frac{\partial f}{\partial X_1} \Delta X_1 + \frac{\partial f}{\partial X_2} \Delta X_2 + \dots + \frac{\partial f}{\partial X_N} \Delta X_N \right\} \\ &+ \frac{1}{2!} \left\{ \frac{\partial^2 f}{\partial X_1^2} (\Delta X_1)^2 + \frac{\partial^2 f}{\partial X_2^2} (\Delta X_2)^2 + \dots + \frac{\partial^2 f}{\partial X_N^2} (\Delta X_N)^2 \right\} + \dots \\ &+ \frac{1}{3!} \left\{ \frac{\partial^3 f}{\partial X_1^3} (\Delta X_1)^3 + \dots \right\} + \dots.\end{aligned}\quad (1.25)$$

The second and higher derivative terms are assumed to be numerically negligible. Thus, the maximum or worst-case uncertainty in Q can be finally approximated by

$$\Delta Q_{\text{MAX}} = |Q - \hat{Q}| = \sum_{j=1}^N \left| \frac{\partial f}{\partial X_j} \Delta X_j \right|. \quad (1.26)$$

As an example of the use of Equation 1.26 for worst-case uncertainty, let us examine the LE in the calculation of the DC power in a resistor:

$$P = I^2 R \quad (1.27)$$

so

$$\Delta P_{\text{MAX}} = 2IR\Delta I + I^2\Delta R \quad (1.28)$$

and

$$\frac{\Delta P_{\text{MAX}}}{P} = 2 \left| \frac{\Delta I}{I} \right| + \left| \frac{\Delta R}{R} \right|. \quad (1.29)$$

Thus, if the LE in R is 0.1%, the 0–10 A ammeter has 1% of full-scale accuracy, the resistor value is 100 Ω and the ammeter reads 8 A, and the nominal power dissipated in the resistor is 6400 W, then the LE in the power measurement is

$$\frac{\Delta P_{\text{MAX}}}{P} = 2 \frac{0.1}{8} + 0.001 = 0.026 \quad \text{or} \quad 2.6\%. \quad (1.30)$$

As a second example, consider finding the LE in calculating the Nernst potential for potassium ions across a semipermeable membrane, given by the following equation:

$$E_K = \frac{RT}{\mathfrak{F}} \ln \left(\frac{C_o}{C_i} \right), \quad (1.31)$$

where

R is the gas constant

T is the Kelvin temperature

\mathfrak{F} is the Faraday number

C_o is the potassium ion concentration outside a semipermeable membrane

C_i is the potassium concentration inside a compartment bounded by the membrane

To find the LE in calculating E_K , we differentiate Equation 1.31 according to Equation 1.26:

$$\Delta E_K = \left| \frac{R}{\mathfrak{F}} \ln \left(\frac{C_o}{C_i} \right) \Delta T \right| + \left| \frac{RT}{\mathfrak{F}} \frac{\Delta C_o}{C_o} \right| + \left| \frac{RT}{\mathfrak{F}} \frac{\Delta C_i}{C_i} \right|. \quad (1.32)$$

Now if we divide Equation 1.32 by Equation 1.31, we find

$$\frac{\Delta E_{K_{MAX}}}{E_K} = \left| \frac{\Delta T}{T} \right| + \left| \frac{\Delta C_o}{C_o} \frac{1}{\ln(C_o/C_i)} \right| + \left| \frac{\Delta C_i}{C_i} \frac{1}{\ln(C_o/C_i)} \right|. \quad (1.33)$$

Let the temperature at 300 K be known to $\pm 0.1^\circ$ and C_o be 100 mM, C_i be 10 mM, and the concentrations be known to $\pm 2\%$. The LE in E_K is thus found to be

$$\frac{\Delta E_{K_{MAX}}}{E_K} = \frac{0.1}{300} + \frac{0.02}{2.3} + \frac{0.02}{2.3} = 1.77 \times 10^{-2} \quad \text{or} \quad 1.77\%. \quad (1.34)$$

In conclusion, we have seen that Taylor's series approach to the calculation of LE for a derived (calculated) measurand provides us with a systematic means of finding the *worst-case error* in that quantity. In actuality, the uncertainties in the parameters may fortuitously combine algebraically to give a lower degree of error. The details of this combination of errors are unknown to us; however, so we generally take the LE as the outer bound (worst case) in evaluating the accuracy of a measurement.

Least mean square (LMS) linear graph fitting, or linear regression, is a technique used to obtain an optimum linear fit of the form $y = mx + b$ to a noisy set of N measurement points, $\{Y_k, X_k\}$, $k = 1, \dots, N$ (such a fit is illustrated for a typical case in Figure 1.5). For example, we

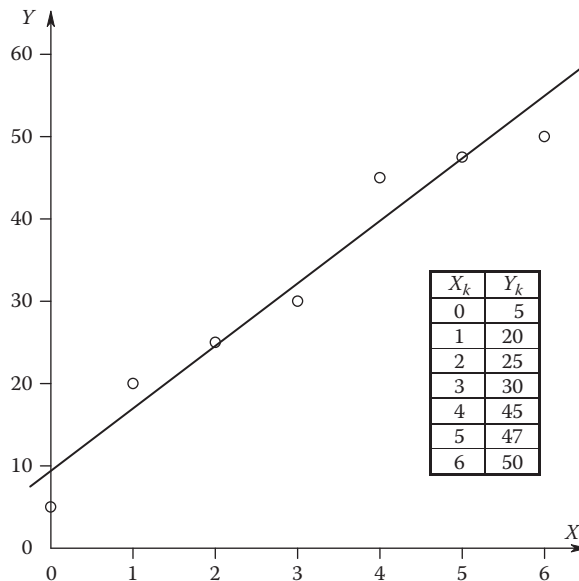


FIGURE 1.5

Illustration of LMS error fitting of a number of data points with a linear regression line.

may wish to measure the optical rotation of polarized light, Y_k , as it passes through sugar solutions of various concentrations, X_k . In the optimum, LMS error fitting of $y = mx + b$ to the data points, we wish to choose m and b to minimize the function σ_y^2 :

$$\sigma_y^2 = \frac{1}{N} \sum_{k=1}^N \{[mX_k + b] - Y_k\}^2. \quad (1.35)$$

Equation 1.35 can be expanded to give

$$\sigma_y^2 = \frac{1}{N} \sum_{k=1}^N \{m^2 X_k^2 + b^2 + Y_k^2 + 2mX_k b - 2Y_k [mX_k + b]\}. \quad (1.36)$$

In order to minimize σ_y^2 , we set the derivatives equal to zero:

$$\frac{\partial \sigma_y^2}{\partial b} = 0 = 2Nb + 2m \sum_{k=1}^N X_k - 2 \sum_{k=1}^N Y_k. \quad (1.37a)$$

$$\frac{\partial \sigma_y^2}{\partial m} = 0 = 2m \sum_{k=1}^N X_k^2 + 2b \sum_{k=1}^N X_k - 2 \sum_{k=1}^N X_k Y_k. \quad (1.37b)$$

Equations 1.37a and b can be written as two simultaneous linear equations in m and b and solved by conventional methods to obtain the following:

$$b = \frac{\bar{Y}\bar{X}^2 - \bar{X}R_{xy}(0)}{\sigma_x^2} \quad (1.38a)$$

$$m = \frac{R_{xy}(0) - \bar{X}\bar{Y}}{\sigma_x^2} \quad (1.38b)$$

for the y -intercept and the slope of the LMS linear fit, respectively. $R_{xy}(0)$ is the *cross correlogram* function evaluated at $\tau = 0$, given by

$$R_{xy}(0) = \frac{1}{N} \sum_{k=1}^N X_k Y_k. \quad (1.39)$$

The goodness of the fit of $y = mx + b$ to the data set, $\{Y_k, X_k\}$, is given by the correlation coefficient for the LMS fit, r , defined as

$$r \equiv \frac{[R_{xy}(0) - \bar{X}\bar{Y}]}{\sigma_x \sigma_y}, \quad 0 \leq r \leq 1. \quad (1.40)$$

r^2 is called the *coefficient of determination* of the linear regression fit. Obviously, $r = 1$ indicates a perfect fit, that is, all $Y_k = mX_k + b$.

It should be pointed out that not all measurands are linear functions of a single independent variable and the preceding analysis represents one of the simpler cases in statistical data analysis.

1.4 Standards Used in Measurements

1.4.1 Introduction

One concern of everyone who has occasion to make a measurement is, is the instrument calibrated? As we have seen earlier, *calibration* is necessary, along with *precision*, to enable *accurate measurements* to be made. Calibration implies observing the instrument's performance when measuring an appropriate *standard*.

In 1960, the *11th General Conference on Weights and Measures* (aka *Conférence Générale des Poids et Mesures—CGPM*) adopted the International System of Units, or SI (*Système International d'Unités*), based on seven *base units* (used for primary standards). The SI base units are given in Table 1.1 (SI 2006).

Table 1.2 shows the derived units computed using the one or more of the seven SI base units.

Major changes have taken place in the late 1980s in the definitions of the base standards, and on January 1, 1990, several of these new standards were adopted by the international community. In the United States, the responsibility for maintaining primary and secondary standards lies with the US NIST (NIST 2013). NIST also actively seeks to establish new and more accurate standards and means of transferring their accuracy when calibrating instruments.

1.4.2 Electrical Standards

A standard is a physical representation of the QUM whose true value is known with great accuracy. Standards can be classified as

1. International standards
2. Primary standards
3. Secondary standards
4. Working standards

International standards are defined by international agreement and are kept at the International Bureau of Weights and Measures (BIPM) in Sèvres, France. An example of an international standard is the *kilogram mass*. (More will be said about the mass standard in the following.) International standards are not available on a daily basis for calibration or comparison.

TABLE 1.1

The Seven SI Base Units

Base Unit Name	Unit Symbol	Quantity	Quantity Symbol	Dimension Symbol
Ampere	A	Electric current	I	I
Candela	cd	Luminous intensity	I_v	J
Kelvin	K	Thermodynamic temperature	T	Θ
Kilogram	kg	Mass	m	M
Meter	m	Length	l, x, r	L
Mole	mol	Amount of chemical substance	n	N
Second	s	Time	t	T

Note: Physical standards are used for their determination.

TABLE 1.2

Units Derived from SI Base Units

Derived Unit Name	Symbol	Quantity	How Derived	Dimension Symbol
Becquerel	Bq	Radioactivity (decays/time)	1/s	T ⁻¹
Coulomb	Q	Electric charge	A s	T I
Degree Celsius	°C	Temperature relative to 273.15 K	K	Θ
Farad	F	Electric capacitance	Q/V	M ⁻¹ L ⁻² T ⁴ I ²
Gray	Gy	Absorbed dose of ionizing radiation	J/kg	L ² T ⁻²
Henry	H	Inductance, mutual inductance	Vs/A = Wb/A	ML ² T ⁻² I ⁻²
Hertz	Hz	Frequency	1/s	T ⁻¹
Joule	J	Work, energy, heat	Nm = CV = Ws	ML ² T ⁻²
Katal	kat	Catalytic activity	mol/s	NT ⁻¹
Lumen	lm	Luminous flux	cd sr	J
Lux	lx	Illuminance	lm/m ²	JL ⁻²
Newton	N	Force, weight	kg m/s ²	MLT ⁻²
Ohm	Ω	Electric resistance, impedance, reactance	V/A	ML ² T ⁻³ I ⁻²
Pascal	Pa	Pressure, stress	N/m ²	ML ⁻¹ T ⁻²
Radian	rad	Angle	m/m	Dimensionless
Siemens	S	Electrical conductance	1/Ω = A/V	M ⁻¹ L ⁻² T ³ I ²
Sievert	Sv	Equivalent dose of ionizing radiation	J/jg	L ² T ⁻²
Steradian	sr	Solid angle	m ² /m ²	Dimensionless
Tesla	T	Magnetic fields strength	Vs/m ² = Wb/m ² = N/(Am)	MT ⁻² I ⁻¹
Volt	V	Voltage, electrical potential difference, EMF	W/A = J/C	ML ² T ⁻³ I ⁻¹
Watt	W	Electrical power, power, radiant flux	J/s = VA	ML ² T ⁻³
Weber	Wb	Magnetic flux	J/A	ML ² T ⁻² I ⁻¹

Primary standards are maintained in national standards laboratories in countries around the world. Primary standards, representing some of the fundamental physical and electrical units, as well as some derived quantities, are independently measured and calibrated at the various national laboratories and compared against each other. This process leads to grand or world average figures for the standards. Primary standards are used continually, but generally do not leave the national standards labs.

Secondary standards are reference standards that are initially calibrated from primary standards and then used in industry and research labs on a daily basis to calibrate their *working standards*, which are used on a daily basis to check and calibrate working laboratory instruments.

1.4.2.1 Volt

The defined international standard for the volt that was in effect from 1908 to 1990 is based on the open-circuit EMF of the saturated *Weston standard cell*, first developed in 1892 by Edward Weston. The Weston cell is an electrochemical battery that consists of two half-cell electrodes

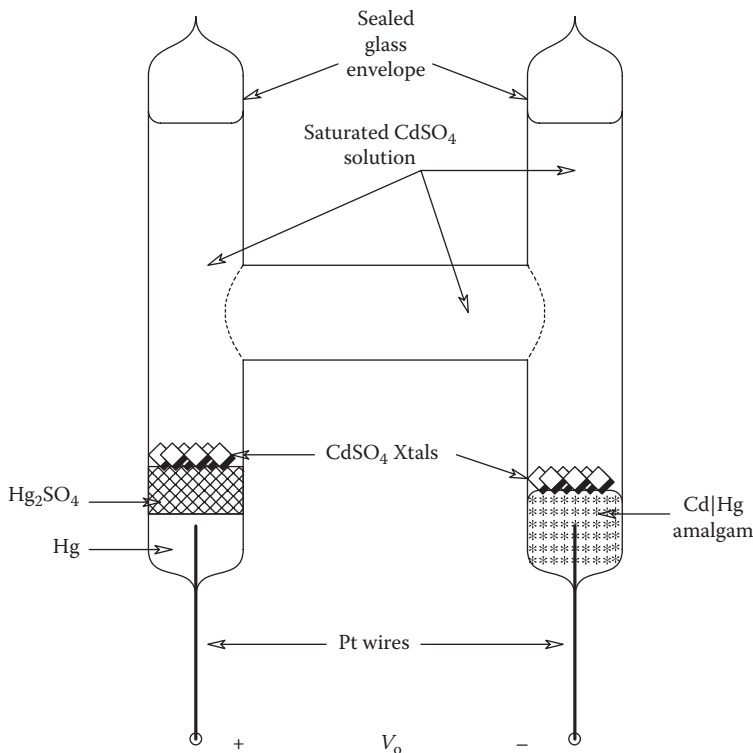


FIGURE 1.6

Diagram of a saturated Weston standard cell battery.

in a sealed, glass *H*-tube (see Figure 1.6). In the *normal* Weston standard cell, the electrolyte is a saturated aqueous solution of cadmium sulfate (CdSO_4). The saturated CdSO_4 electrolyte gives the EMF of the normal cell a relatively high tempco, about $-40 \mu\text{V}/^\circ\text{C}$. However, this form of Weston cell shows higher, long-term, output EMF stability than do cells with unsaturated electrolyte. The EMF of the saturated Weston cell drifts about $1 \mu\text{V}/\text{year}$, and the useful life of a well-treated normal cell is 10–20 years (Helfrick and Cooper 1990).

Under nearly open-circuit conditions, the EMF of the Weston normal cell in absolute volts is given by

$$E(\Delta T) = E_{20} - 4.6 \times 10^{-5}(\Delta T) - 9.5 \times 10^{-7}(\Delta T)^2 - 1 \times 10^{-8}(\Delta T)^3, \quad (1.41)$$

where

E_{20} is the normal cell EMF at exactly 20°C ($E_{20} \equiv 1.01858$ absolute volts)

$\Delta T = (T^\circ\text{C} - 20)$

One volt was defined as $1/1.01858$ times the EMF of a normal Weston cell at exactly 20°C . Because of their high temperature sensitivity, saturated Weston cells are kept in temperature-regulated baths in which the temperature is kept constant to within $\pm 0.000010^\circ$.

The present international standard for the volt was adopted on January 1, 1990, and is based on a quantum-effect phenomenon that takes place at liquid helium temperatures, the *Josephson effect*. One form of a *Josephson junction* (JJ) consists of two thin films of superconducting lead separated by a 1 nm thick layer of lead oxide (Taylor 1990). Another JJ

design uses superconducting niobium separated by a thin film of Al_2O_3 . Such JJs can be fabricated into series arrays of JJs using integrated circuit (IC) technology (Pöpel et al. 1991). Pöpel et al. reported on the performance of IC arrays with 2,000 and 20,160 JJs, which were useful as standards in the range of 1 and 10 V DC, respectively.

A single, superconducting JJ has the unique property such that when it is irradiated with microwave energy in the frequency range from 9 to 100 GHz and biased with a DC current, a precisely known DC voltage appears across the DC portion of the JJ having a stepwise volt-ampere curve, as shown in Figure 1.7. The voltage of each stable step, E_J , is given by

$$E_J = \frac{nf}{(2e/h)} \text{ V}, \tag{1.42}$$

where

- n is the step number
- f is the microwave frequency in GHz
- e is the electron charge
- h is Planck's constant

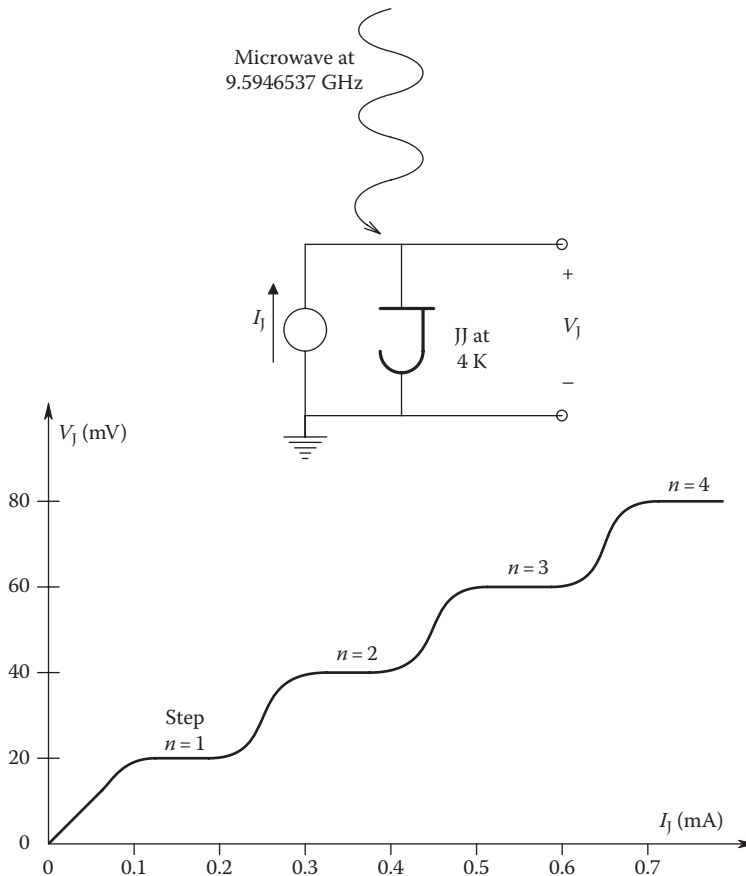


FIGURE 1.7

Volt-ampere curve for a typical JJ voltage standard. The microwave frequency was 9.5946537 GHz. V_J is approximately 1.9 mV at the 97th step, with $I_J = 24$ mA at 4.2 K. (Drawn from data given by Ibuka, M. et al., *IEEE Trans. Instrum. Meas.*, 32(1), 276, 1983.)

The quantity $(2e/h)$ is known as K_J , the *Josephson constant*. The *Comité International des Poids et Mesures* (CIPM) established a universal standard value for the Josephson constant, $K_{J-90} = 483,597.9$ GHz/V on January 1, 1990. For a typical microwave frequency of 94 GHz, derived from a *Gunn oscillator*, the step voltage is seen to be $194.4 \mu\text{V}$. Exactly 1 V can be defined as the EMF of a JJ when it is irradiated with electromagnetic (EM) radiation at a frequency of 483,597.9 GHz. Of course, such a frequency is a mathematical convenience; the JJ would not work when irradiated with 620 nm light, which has that frequency.

When using a JJ array as a primary (SI) standard, the microwave frequency is seen to be quite critical. Thus, the Gunn oscillator is phase-locked to an ultrastable frequency source, such as a cesium beam clock. The frequency source that has been used for several designs of JJ array voltage standards has been a 50–60 mW, 93–95 GHz Gunn oscillator.

Figure 1.8 gives a simple schematic of a JJ array (after Endo et al. 1983), which was used to calibrate other standard DC voltage sources and DC potentiometers. One of the problems in operating a JJ array is knowing which quantum step, n , the JJs are all on. One possible way of solving this problem is to obtain a stable array output voltage, then vary the

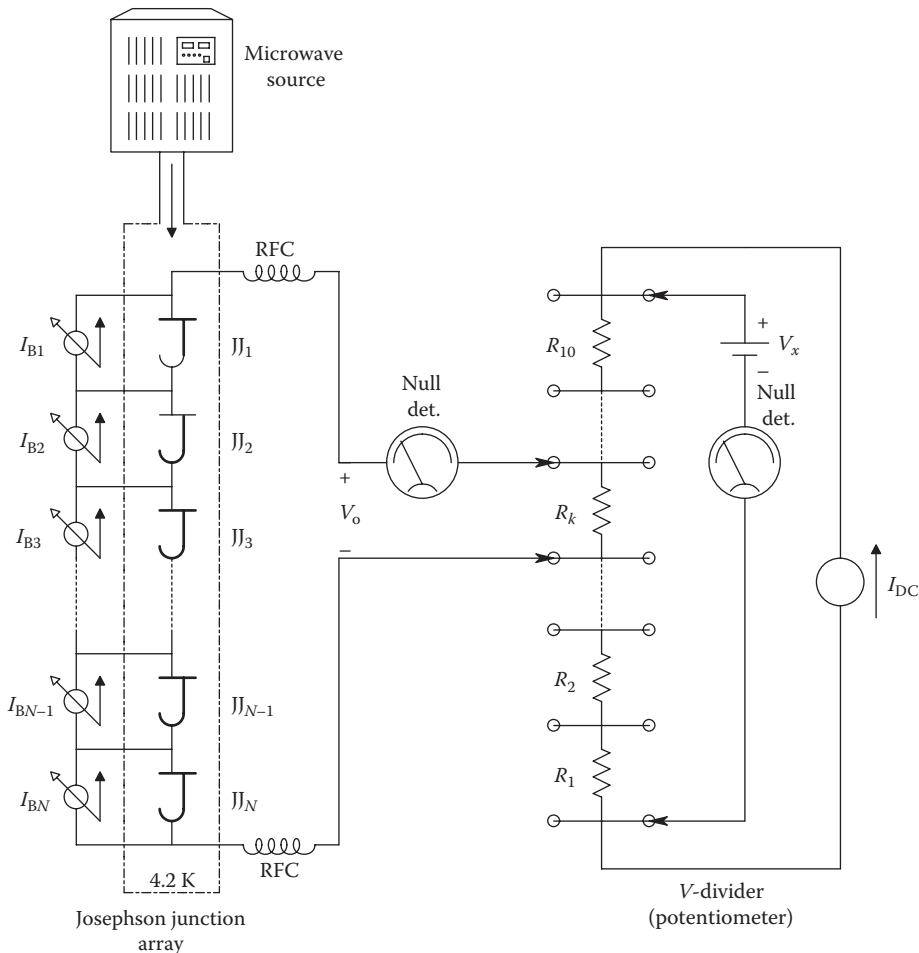


FIGURE 1.8 Simplified schematic diagram of a precision potentiometer used to measure unknown EMFs. The potentiometer is calibrated with a standard voltage from a JJ array by adjusting the potentiometer current, I_{DC} .

DC bias current until the array jumps to the next step level, and then adjust the microwave frequency down to f_2 until the same output voltage is obtained. Using this protocol, we can solve for n :

$$nf_1 = (n+1)f_2 \quad (1.43)$$

so

$$n = \text{INT} \left[\frac{(f_1 - f_2)}{f_2} \right]. \quad (1.44)$$

One can also simply count the steps as the bias current is slowly increased under conditions of constant microwave frequency and power.

The output of an array of N JJs in series all having the same DC bias current and subject to the same microwave frequency and energy is

$$V_o = NE_J(n, f). \quad (1.45)$$

Although the microwave frequency is the critical input parameter to a JJ array voltage standard, the microwave voltage and DC current must be kept within certain bounds for stable operation of the JJ array as a quantized voltage source. Hamilton et al. (1991) have found relations for critical values of the DC current density, J_C ; the length, l ; and width, W , of a JJ for maximum stability of the voltage steps (to prevent noise-induced jumps between steps). They also give an empirical relation for the current range for the n th step, under optimum conditions. Hamilton et al. (1991) have designed JJ arrays having stable operation at 24 GHz and producing up to 1.2 V.

Most of the Josephson array voltage standards, which have been described in the literature (see, e.g., a number of papers on JJ standards in the April 1991 *IEEE Transactions on Instrumentation and Measurement*, 40(2)), have voltage uncertainties around 2×10^{-8} . Reymann (1991) described a JJ array system that is used to measure Weston normal cell EMFs to better than 1 nV accuracy. A problem with the operation of JJ array voltage standards is that the noise-induced instability of the voltage steps leads to jumps to unknown or uncertain step numbers. It is apparent that the JJ array provides an accurate, flexible, DC voltage standard, which can be easily adapted for the calibration of secondary standards and DC voltmeters.

Primary Josephson voltage standards are now commercially available. PREMA[®] Semiconductor, Mainz, Germany, (www.prema.com) offers DC voltage standard systems based on their JVS-7001 and JVS-7010 JJ chips. The chip specifications are given in Table 1.3.

The Sandia National Laboratories' Primary Standards Lab has also developed a 10 V JJ primary standard and lists its accuracy as ± 0.017 ppm. See www.sandia.gov/psl/Dc_factSheet2008_Final.pdf (accessed April 8, 2013).

1.4.2.2 Resistance

In 1884, an international congress in Paris adopted a reproducible standard for the *legal ohm* consisting of a column of ultrapure mercury 106 cm in length and 1 mm² in cross-sectional area, measured at 0°C. In 1893, the Chicago Congress specified the length of

TABLE 1.3

Specifications of PREMA® JJ Primary Voltage Standard Chips

Parameter	1 V Chip	10 V Chip
Chip area	19 × 5 mm ²	25 × 11 mm ²
Number of JJs	2,400	13,920
Operating temperature	4.2 K	4.4 K
Operating frequency	70–75 GHz	70–75 GHz
Lead and bond wire resistance	<3 Ω	<3 Ω
Technology	Refractory all niobium full wafer process	Refractory all niobium full wafer process

mercury column of the *international ohm* to be 106.3 cm and specified the weight of the mercury to be 14.4521 g. The mercury ohm proved to be a poor standard. Mercury has a high tempco, requiring close thermoregulation of the standard in an ice bath. Slight impurities in the mercury also contribute to errors in certainty, and variations in the design of the end terminals caused errors in the resistance. According to Stout (1950), another major cause of error in the design of the international ohm was nonuniformity in the diameter of the glass tubing. It was found that a 30 ppm change in the measured resistance could result from an undulation in the axis of the glass tube too small to be detected by eye.

In the 1920s, work was done on defining an *absolute ohm* as a derived quantity based on the fundamental units of length, mass, and time. The absolute ohm was determined through the use of a rotating commutator or conductor whose speed must be known, giving the time dependency. The determination also used either a self- or mutual inductance, whose value was determined by calculations involving the dimensions. Not surprisingly, the absolute ohm's value was found to differ significantly from that of the international ohm. One absolute ohm was equal to 0.9995052 international ohms. The absolute ohm and other absolute electrical units were formally adopted on January 1, 1948.

After 1892, working standard, wire-wound, one ohm resistors made of the alloy *manganin* were developed and in 1897 were calibrated in terms of the mercury ohm as defined in 1893 by the Chicago Congress. The US NBS adopted the Rosa design, one ohm manganin, wire-wound resistor in 1910 as the international ohm. In 1931, an improved, standard resistor design by J. L. Thomas was adopted at NBS. The cross section of a Thomas-type, one ohm standard resistor is shown in Figure 1.9. This resistor uses a double-walled construction, with the resistance wire-wound on the inner cylinder, and the space between the cylinders is filled with dry air and sealed. The center cylinder is filled with oil for temperature stabilization and control. In the Rosa-type resistor, the resistance wire windings are in direct contact with dry oil in a sealed can. Rosa resistors are available in decade sizes, typically 1, 10, 100, 1k, and 10k ohms.

In 1966, HP developed an improved standard resistor design, their 11100 series, which achieved superior temperature stability and precision through the use of new materials for insulation, resistance coil support, and the resistance wire. HP used the alloy Evanohm for the coil (75% nickel, 20% chromium, 2.5% aluminum, 2.5% copper). According to HP, Evanohm has an α -tempco of 0 to +2 ppm/°C, a high resistivity, and a relatively flat resistance vs. temperature curve, and it is insensitive to moisture. The Evanohm wire coils of the HP 11100 resistors are suspended in sealed, dry oil and supported by an inner polyester coil form and an outer form, rather than a metal form such as used in the Rosa or Thomas resistors. HP claimed the following specifications for its 11100 series resistors: Limit of error at 25°C and 100 mW power dissipation is ± 6 ppm (0.0006%) with reference to the

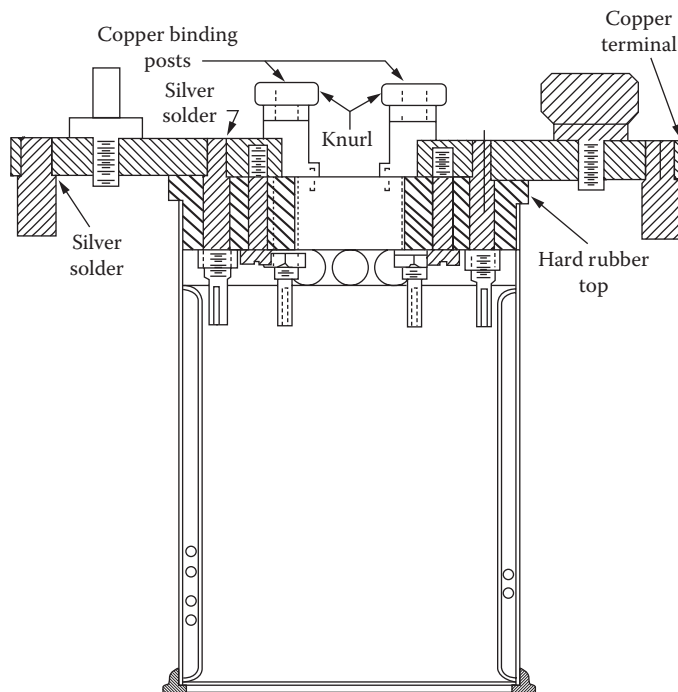


FIGURE 1.9

Cross section of a Thomas-type standard resistor. (From Stout, M.B., *Basic Electrical Measurements*, 2nd edn., Prentice-Hall Inc., Englewood Cliffs, NJ, 1960, Fig. 4.2. With permission.)

legal ohm (1966) maintained by NBS. Rated tempco is ± 4 ppm. Drift per year is < 20 ppm (100 k Ω). Power rating is 100 mW. Internal trim resistance permits adjustment of ± 25 ppm. Connections are NBS-type, 4-terminal, oxygen-free copper, nickel-rhodium plated.

Note that the HP 11100 standard resistors can be set to within ± 0.15 ppm of their nominal (true) resistance, but the 1966 NBS primary standard has an uncertainty of ± 6 ppm.

At the present time, the Fluke 742 series, 4-terminal, working standard resistors are available in values from 1 Ω to 19 M Ω . The Fluke 742 series resistors are used at ambient room temperatures, from 18°C to 28°C, with resistance changes ranging from ± 1.5 (10k) to ± 4 (19M) ppm, showing excellent temperature stability—no oil baths are required. Calibration uncertainty at 23°C ranges from ± 1.0 ppm for 1–100 Ω sizes to ± 10.0 ppm (10 M Ω) and ± 20.0 ppm (19 M Ω). Stability over 1 year ranges from ± 8.0 ppm (1–10 Ω) to ± 10.0 ppm (19 M Ω). (See the Fluke website: <http://us.flukecal.com/products/electrical-calibration/electrical-standards/742a-resistance-standards-0> [accessed April 8, 2013].)

Electro Scientific Industries (ESI) also makes transfer standard resistors. The ESI model SR1060 resistance transfer system consists of six transfer standards in decades from 1 Ω to 100 k Ω . Each decade standard is composed of 12 nominally equal resistors matched to within ± 10 ppm. Each decade standard can be configured as 10 resistors in series, 10 resistors in parallel, or 9 of the 10 resistors in series/parallel. By making a 1:1 comparison with the 10th resistor, it is possible to resolve a series/parallel value to better than ± 1 ppm. All of the standards except the 100 k Ω standard are immersed in an oil bath for temperature stability. The ESI SR1060 resistance transfer standard system is initially calibrated to ± 10 ppm traceable to the NIST primary standard. Long-term stability is ± 35 ppm for 1 year, and tempcos are ± 15 ppm/°C for 1 Ω and ± 5 ppm for 100 Ω to 100 k Ω . (See www.esi.com/Products.aspx [accessed April 8, 2013].)

On January 1, 1990, the international I&M community adopted a standard for the ohm based on the *quantum Hall effect* (QHE). This definition of the ohm, like the JJ definition for the volt, is based on fundamental physical constants, rather than an artifact (such as a column of mercury). The QHE was first described by the noted physicists von Klitzing et al. (1980). The operation of a basic Hall effect sensor at room temperature is described in Section 6.2.4.2. The ordinary Hall sensor is a four-terminal device; one pair of terminals is used to inject current into the thin, doped semiconductor bar, which is ordinarily perpendicular to a magnetic field, \mathbf{B} . The second pair of terminals, as shown in Figure 6.22, is used to pick off the Hall EMF. It can be shown that the Hall EMF is given by

$$E_H = \frac{R_H B_y I_x}{h}, \quad (1.46)$$

where

- B_y is the B field component in the y -direction, orthogonal to the current density vector, \mathbf{J}_x
- I_x is the injected current
- h is the thickness of the doped semiconductor bar in the y -direction
- R_H is the Hall coefficient

$R_H = -1/qn$ for n -doped semiconductor and $1/qp$ for p -doped semiconductor. q is the magnitude of the electron charge in coulombs, n is the electron donor doping density, and p is the hole donor doping density.

The Hall sensor was seen to be useful in measuring magnetic fields and, indirectly, electric power. As will be illustrated in the following, the properties of the Hall sensor change markedly as its temperature approaches 0 K, and it is subject to strong, DC magnetic fields.

Figure 1.10 illustrates an isometric view of a quantum Hall resistor (QHR). Modern QHRs are generally fabricated from GaAs/AlGaAs heterostructures. The channel width may range from 150 to 250 μm , and the distance between lateral voltage contacts on the same side may range from 300 to 600 μm (Piquemal et al. 1991). The thickness of the aluminum-doped conducting channel may range from 10 to 100 nm (von Klitzing 1986). The QHR has six or eight terminals: two for current, two for voltage, and two or four for measuring body resistivity or voltage in the x -direction. Under normal operation, the QHR is operated at a constant, DC current bias, I_x . I_x values ranging from 10 to 50 μA are typically used. A very strong, DC magnetic field, B_y , on the order of 4–15 T, is generated by a superconducting coil in close proximity to the QHR.

A QHR is generally operated at temperatures ranging from 0.3 to 2 K, with 1.20 K being commonly used. As can be seen from Figure 1.11A, as the magnetic field is varied under conditions of constant DC bias current, the Hall resistance assumes a series of steps of precisely known values, regardless of the exact B_y , I_x , or temperature and material of the QHR semiconductor. The quantum Hall resistance, R_{Hv} , has been shown by von Klitzing (1986) to be given by

$$R_H(j) = \frac{E_H}{I_x} = \frac{R_{K-90}}{j} = \frac{h}{q^2 j} \Omega, \quad (1.47)$$

where

- h is Planck's constant
- q is the electron charge
- j is the step number (integer)
- R_{K-90} is the *von Klitzing constant* (universally adopted as 25,812.807 Ω on January 1, 1990)
- E_H is the measured DC Hall voltage

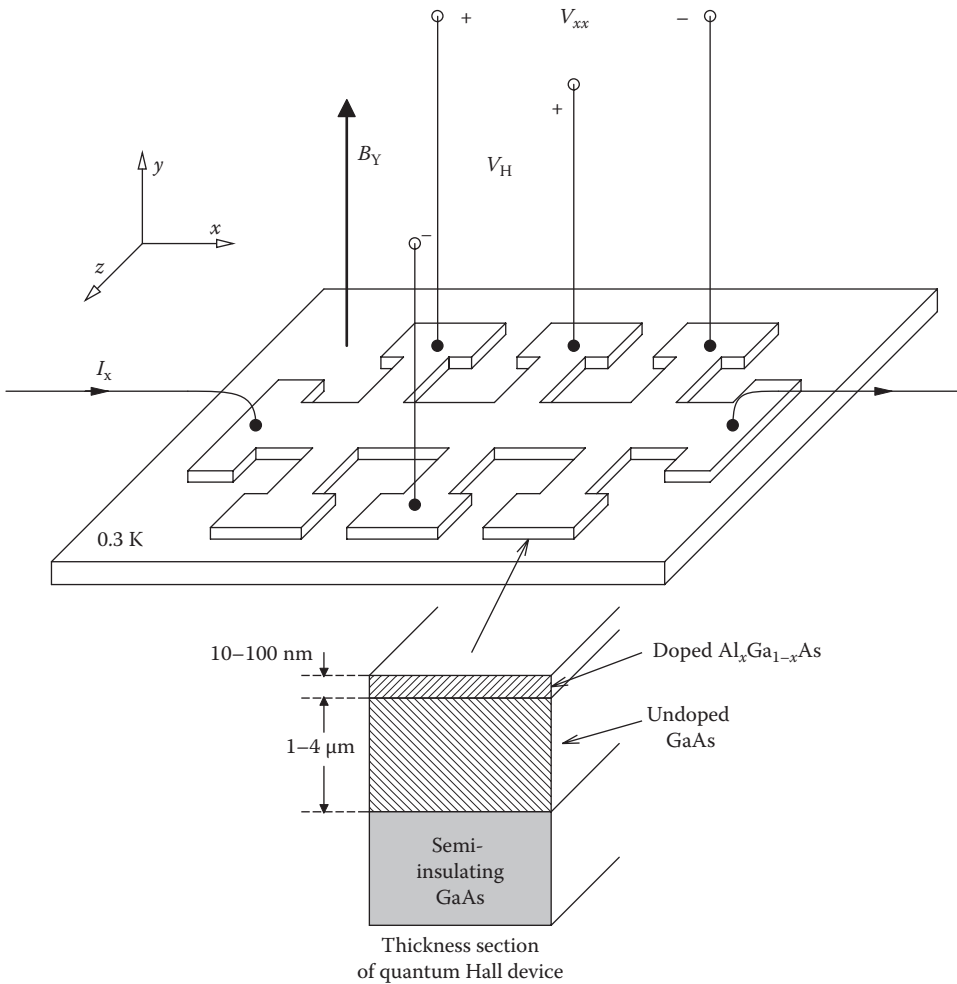


FIGURE 1.10

Isometric view of a quantum Hall device. (Adapted from von Klitzing, K., *The quantum Hall effect*, in *The Physics of the Two-Dimensional Electron Gas*, J.T. Devreese and F.M. Perters, eds., Plenum Press, New York, 1986.) V_H is the Hall EMF; V_{xx} is the voltage drop due to the DC current, I_x ; and B_y is the strong, DC magnetic field. A gallium–aluminum–arsenide heterostructure was used. Operating temperature is typically 0.3 K.

For all QHR devices, the quantum Hall resistances are as follows:

$$R_H(1) = 25,812.807 \, \Omega, \quad R_H(2) = 12,906.400 \, \Omega,$$

$$R_H(3) = 8,604.2690 \, \Omega, \quad R_H(4) = 6,453.2018 \, \Omega,$$

$$R_H(5) = 5,162.5614 \, \Omega, \quad R_H(6) = 4,302.1345 \, \Omega.$$

The value of $R_H(1)$ is defined as one von Klitzing.

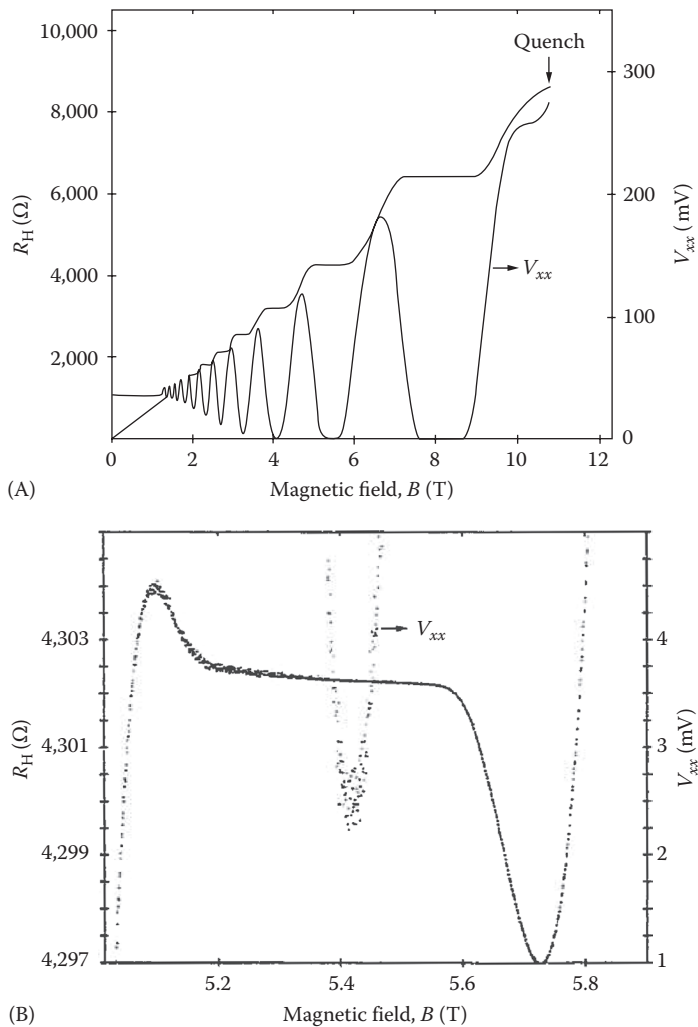


FIGURE 1.11

(A) Plot of the quantum Hall resistance, R_H , and V_{xx} vs. B_y . Note that there are plateaus of $R_H = E_H/I_x$ over a range of B_y on which R_H is precisely defined. In this case, $T = 0.3$ K and $I_x = 50$ μ A. (B) Detail of the $i = 6$ quantum Hall step. Note that there is a distinct minimum of V_{xx} at the flattest part of the R_H curve. Conditions are the same as in Figure 1.11a. (From Jaeger, K.B. et al., *IEEE Trans. Instrum. Meas.*, 40(2), 256, 1991. With permission.)

Also plotted in Figure 1.11A is the voltage drop in the x -direction. Note that V_{xx} has minima that occur at the centers of the $E_H(B)$ or $R_H(j, B)$ steps. This is shown more clearly in Figure 1.11B in which the value of V_{xx} shown as B_y is varied over the range producing the $R_H(4)$ step (Jaeger et al. 1991). Figure 1.12 illustrates the schematic of a simple series circuit, which is used to compare the QHR with a secondary standard, using Ohm's law and precision DC nanovoltmeters and DC picoammeters.

Comparisons between quantum Hall resistance measurements made in various standards laboratories around the world suggest that the QHR method of defining the ohm has an uncertainty of less than one part in 10^8 .

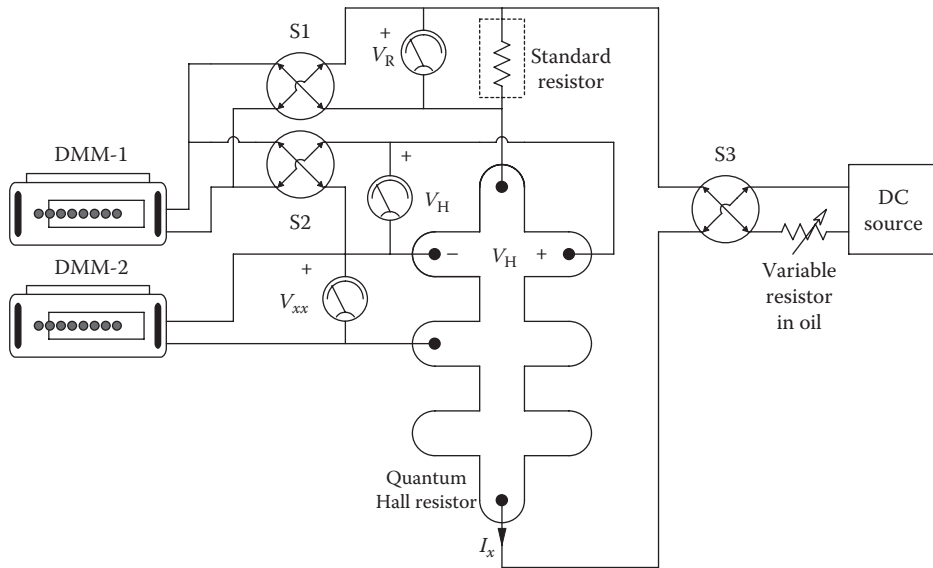


FIGURE 1.12

Schematic of a circuit used to compare a QHR with a secondary standard resistor. (From Jaeger, K.B. et al., *IEEE Trans. Instrum. Meas.*, 40(2), 256, 1991. With permission.)

1.4.2.3 Current and Charge

In a metallic conductor, such as a copper wire, the moving charge is electrons, whose average velocity is in the opposite direction to defined (in the United States) current flow. In semiconductors, two kinds of mobile charges are involved: holes and electrons. In *p*-semiconductor material, the intrinsic semiconductor is doped with electron-accepting atoms, and the majority charge carrier is holes. In *n*-material, electron donor atoms provide electrons as the majority carrier. In wet (chemical and biochemical) materials, the carriers are generally ions, although electrons are involved at metal/metal salt electrode interfaces.

In the SI system of measurements, the *ampere* is defined to be the basic unit of current, and current is considered to be a fundamental unit of measurement. Before the SI protocols were adopted, current was considered to be a derived quantity, namely, 1 C of charge per second flowing past a perpendicular plane through a conductor. *Charge* was considered to be a fundamental quantity in the older, MKS system. In the SI system, charge is derived and has the units of current times time. One coulomb is the charge transferred when one ampere flows for exactly one second. Charge can also be measured by using the relationship that the voltage on a capacitor is equal to the charge divided by the capacitance, or that the change in voltage on a capacitor is equal to the change in charge divided by the capacitance.

Early definitions of the ampere made use of the electrolytic reduction of silver ions from a standard silver nitrate solution at a platinum cathode. Metallic silver is electroplated out on the cathode at a rate proportional to the electric (DC) current flowing in the external circuit. After a known time, the cathode is removed from the electroplating cell, dried, and weighed. Since the gram-molecular weight of silver is known, as is Avogadro's number (cf. Glossary), the accumulated weight of silver on the cathode is proportional to the total charge transferred in the circuit over the known time. At the London Conference of 1908,

the international ampere was defined as that (DC) current that deposits silver at the rate of 0.0011800 g/s from a standard AgNO_3 solution. This definition was based on a knowledge of the gram-molecular weight of silver (107.868) and Avogadro's number, N_A , which is the number of molecules (hence Ag^+ ions) in one gram-molecular weight of a substance. The value of N_A is presently accepted by NIST to be $6.0221415 \pm 0.0000010 \times 10^{23} \text{ mol}^{-1}$ (Fox and Hill 2007). Avogadro's number was recalculated for ^{28}Si enriched to 99.985% by Fox (2012) to be $6.02214129(27) \times 10^{23}$.

Because of the lack of precision and repeatability of the silver electroplating method, the international ampere was superseded in 1948 by the *absolute ampere*. The determination of the absolute ampere is made using a current balance, which weighs the force between two current-carrying coils. *By definition*, the SI absolute ampere is the constant (DC) current, which, if maintained in two, straight, parallel conductors of infinite length and negligible circular cross section placed 1 m apart, in vacuo, will produce between these conductors a force of $2 \times 10^{-7} \text{ N/m}$ length. If the same current is traveling in the same direction in the two, parallel conductors, it may be shown the force exerted between the conductors is attractive and is given by

$$F = \frac{\mu_0 I^2}{2\pi d} \text{ N} \quad (1.48)$$

where

d is the separation of the conductors in m

μ_0 is defined as $4\pi \times 10^{-7} \text{ Wb/A m}$

Such a theoretical definition is unrealizable in practice.

Working standards to measure the absolute ampere generally involve a current balance, in which the force between a moveable coil or coils and a set of fixed coils or a permanent magnet is measured by a weighing balance. This type of determination of the ampere requires accurate measurement of force. The value of the mass used and g , the Earth's gravitational acceleration at the location of the balance, must be accurately known.

Kibble (1983) described the design and development of a moving-coil apparatus for determination of the ampere at the National Physics Laboratory (United Kingdom). A large permanent magnet with maximum flux density of about 0.7 T is used. Two, series-connected, rectangular coils, each of 3362 turns, are mounted one above the other on an armature that moves vertically in the magnet's air gap. The armature is attached by a vertical rod to the balance arm and passes coaxially through a force-producing solenoid coil at the bottom of the balance case. The current in the solenoid is controlled by a servo system so it can generate axial forces to cause the measurement coils to move up and down through the magnetic field at a constant velocity, $u = 0.002 \text{ m/s}$. This motion generates an EMF in the series measurement coils whose magnitude is given by $V = K_v u \text{ V}$. The value of K_v depends on B in the magnet's air gap, the number of turns, and other geometrical factors. In the second phase of measuring the ampere, a DC current, I , is passed through the series measurement coils, which are positioned at the center of the B field, where the determination of K_v was made. A *Lorentz force* is generated, given by $F = K_F I$. $F(I)$ is measured with the balance. The value of I is determined from the fact that in the MKS (SI) system, K_v is equal to K_F . If the relation $V = K_v u$ is divided by $F = K_F I$, we can obtain a relation equating electrical and mechanical power:

$$VI = Fu = Mgu. \quad (1.49)$$

Now we assume that the generated electrical power is dissipated in a resistor R such that $V/R = I$. Thus,

$$VI = I^2R = Mgu. \tag{1.50}$$

Solving for I , we find that

$$I = \sqrt{\frac{Mgu}{R}} \text{ SI amperes.} \tag{1.51}$$

While under development, Kibble et al. reported that their moving-coil current system for the determination of the ampere had ± 1 ppm reproducibility and that they eventually expected to obtain 0.1 ppm accuracy.

A system utilizing the magnetic levitation of a superconducting mass to determine the SI ampere was proposed by Kibble (1983). This system is shown schematically in Figure 1.13. A current source, I , causes current to flow in a superconducting coil surrounding a superconducting mass. Induced current in the mass produces a magnetic field that reacts with the field of the coil to produce an upward force until the mass reaches a position Y_1 where the downward gravitational force, Mg , equals the upward levitation force, $F_I = I(d\Phi/dy)$. Now if the mass is given a vertical velocity, $u_y = dy/dt$, around Y_1 , a voltage, $V = d\Phi/dt$, will be induced in the superconducting coil, across the current source. We also note that $d\Phi/dt = (d\Phi/dy)(dy/dt)$. Now from these relations, it is easy to see that $Mg/I = d\Phi/dy = V/u_y$. Solving this relation for I gives

$$I = \frac{Mgu_y}{V}. \tag{1.52}$$

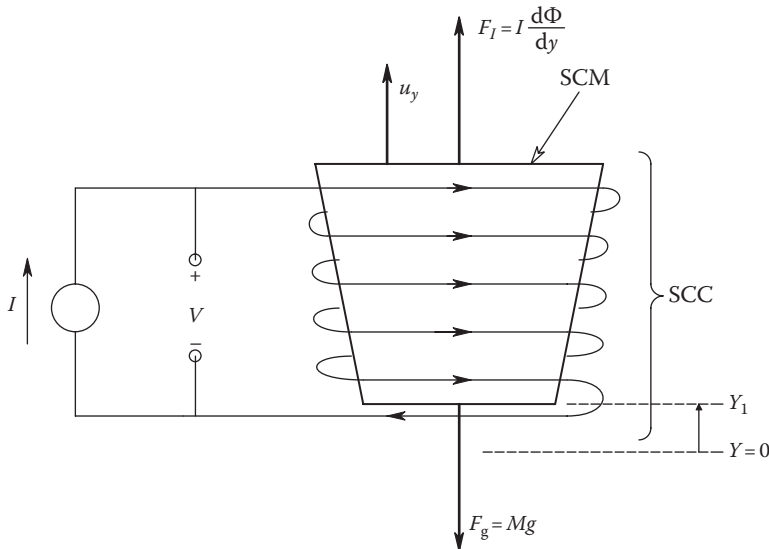


FIGURE 1.13 Diagram of a superconducting mass balance means of measuring current proposed by Kibble 1983. *Note:* SCC, superconducting current-carrying coil; SCM, superconducting mass, M . Means of inducing constant velocity, u_y , not shown. The induced voltage, V , is measured potentiometrically and is related to the current by Equation 1.52.

Kibble's proposed method of measuring DC current thus requires accurate knowledge of the mass, M , and the local acceleration of gravity, g . A constant, linear velocity of the mass, u_y , around its equilibrium displacement, Y_1 , must be generated, either by mechanical means or by electrostatic forces. A *no-touch* means of measuring u_y could be designed using laser interferometry. Lastly, measurement of V can be done potentiometrically by comparison with a Josephson voltage standard.

The reader should be aware that the bases for electrical and physical standards are constantly under development and are changing with the advances in quantum physics, laser technology, etc. We expect to see a trend in metrology away from complex, expensive, ultraprecision measurement systems toward simpler, more elegant solutions for the determination of standard quantities. Perhaps we should consider making the quantum Hall ohm and the Josephson volt basic SI quantities, rather than the ampere.

Is charge a more easily measured quantity than current? Should we return to the earlier MKS system where charge was a basic quantity, rather than current, which is now used in the present SI definitions? There are no simple answers to these questions, and the points of debate change constantly with evolving technology.

1.4.2.4 Capacitance

There are several ways of defining the basic unit of capacitance, the farad. A capacitor is said to have a capacitance of one farad if the potential across its plates is exactly 1 V when it has been charged by 1 C of electric charge. This definition of the farad requires an accurate voltage measurement in which a negligible charge is exchanged with the capacitor and accurate measurement of the total charge put into the capacitor. The latter measurement is most difficult, since it involves precision integration of the capacitor's charging current.

Another way of measuring capacitance is through the use of a commutated-capacitor bridge (cf. Section 5.5.1.6 and Figures 5.9 and 5.10). This technique requires a precision DC null detector, accurate resistances, and a switching clock with an accurately known period, T . Under the assumption that $R_4 C_x \ll T/2 \ll R_2 C_x$, we show in Section 5.5.1.6 that at null (set by R_1),

$$C_x = \frac{T(R_3 + R_1)}{4R_2(R_3 - R_1)}. \quad (1.53)$$

Here, we assume that the switch has zero closed resistance and infinite open resistance, the waveforms at V_2 are exactly rectangular and triangular, and the clock duty cycle is exactly 50%. In practice, the duty cycle can be adjusted to be very close to 50%, taking into consideration the switching waveform rise and fall times and the time delays in operating the metal oxide semiconductor (MOS) switch.

Probably the most accurate means of measuring capacitance is by comparison of the capacitor to be measured with a *calculable capacitor*. Calculable capacitors are based on the electrostatics theorem developed by Thompson and Lampard (1956). A calculable capacitor is typically made from four, parallel, hollow metal cylinders arranged in a square, with a movable shield device, which effectively regulates the length of the active (exposed) cylinders. Capacitance is measured between opposite cylinders, in vacuo. Calculable capacitors have been made with capacitances ranging from 0.1 to 1.0 pF.

The Thompson and Lampard theorem states:

Let the closed curve S be the cross section of a conducting cylindrical shell, which cross section has one axis of symmetry AC but is otherwise arbitrary. Further, let this shell be divided into four parts by two planes at right angles, the line of intersection of the planes being parallel to the generators of the cylinder, and one of the planes containing the symmetry axis AC . Then the direct capacitance, per unit length of the cylinder, between opposing parts of the shell (e.g., $\alpha\beta$ to $\gamma\delta$) due to the field inside (or outside) the shell, is a constant:

$$C_o = \frac{\ln(2)}{4\pi^2} \text{ e.s.u.} = 0.0175576 \text{ e.s.u./m.} \quad (1.54)$$

Note that 1 pF of capacitance in SI units is equal to 0.898797 e.s.u. capacitance units. Thus, it is easy to show that $C_o = 1.953549$ pF/m (or aF/ μm). A consequence of this theorem is that the absolute cross-sectional area of the closed cylindrical shell is not important. What is important is that the sides be perfectly parallel and that the structure has perfect symmetry. Figure 1.14A illustrates the geometry discussed in the theorem stated earlier, and Figure 1.14B illustrates the shielded, four-tube, cross-sectional geometry of a practical calculable capacitor (Igarishi et al. 1968). The length of the calculable capacitor at the US NIST is measured interferometrically using a stabilized HeNe laser. Overall uncertainty in the measurement of the NIST farad, based on the NIST calculable capacitor, is given as 0.014 ppm (Shields et al. 1989).

Calculable capacitor geometry has not been limited to four-tube devices. Delahaye et al. (1987) used five tubes in a pentagonal cluster, each 75.5 mm in diameter and 450 mm long. A cylindrical central screen is displaced axially by a stepping motor, and its position is measured with a laser interferometer. The calculable capacitor of Delahaye et al. was used to establish a standard for the farad and for the quantum Hall resistance $R_H(2)$. The total one standard deviation uncertainty in the measurement of $R_H(2)$ was given as 2.2×10^{-7} .

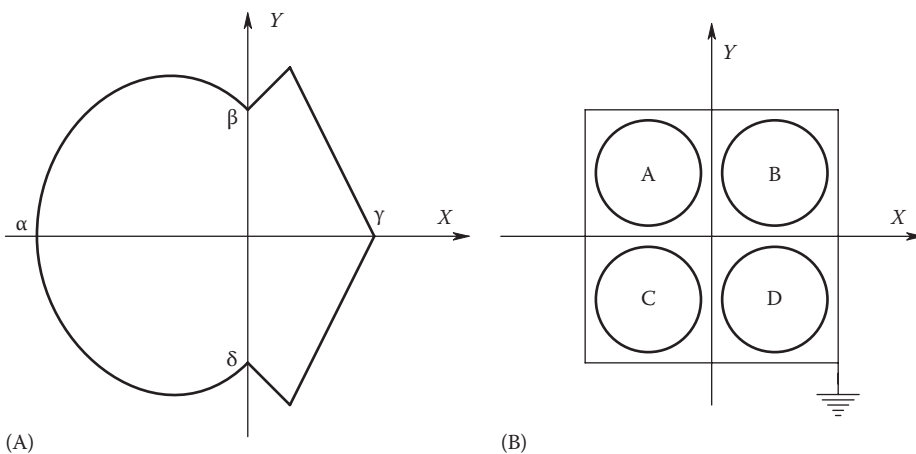


FIGURE 1.14

(A) Cross section of a symmetrical conducting tube relevant to the Thompson–Lampard theorem. The T–L theorem defines the capacitance between sections $\alpha\beta$ and $\gamma\delta$ on the tube as equal to $\ln[2/(4\pi^2)]$ e.s.u./m, regardless of the absolute cross-sectional area of the tube. (B) Cross section of a practical, 4-tube, shielded calculable capacitor, such as that built by Igarishi et al. (1968). Tube diameters were about 1 cm.

Calculable capacitors have been used as standards by others to calibrate the quantum Hall ohm, inductors, and transfer standard capacitors through the use of precision AC bridges (Dahake et al. 1983, Shida et al. 1989).

Secondary and transfer standard capacitors are generally on the order of 1–10 pF and are made with air, quartz (silica), or mica dielectrics. Such capacitors are generally made as three-terminal devices, so the capacitance from either terminal to the grounded case can be compensated for. Air dielectric capacitors of the guard-well geometry (see Figure 1.15) have been constructed with values down to 1 fF. A 0.1 pF capacitor of this design has an accuracy of 0.1% (Stout 1960). For secondary standards with larger values, hermetically sealed, silvered mica capacitors offer long-term stability and low tempcos. For example, a General Radio (GenRad)-type 1409, 1 μ F standard silver mica capacitor has its value guaranteed to $\pm 0.05\%$, and its tempco is 35 ± 10 ppm/ $^{\circ}$ C. Silver mica capacitors have very low D values at 1 kHz. Most bridges used to transfer calibration from calculable capacitors to working and transfer standard capacitors are of the ratio transformer type, operated at frequencies of 2,500 and 5,000 r/s (Delahaye et al. 1987) and at 10,000 r/s (Igarishi et al. 1968, Dahake et al. 1983). (In February 2000, IET Labs Inc., Roslyn Heights, NY, acquired the GenRad product line from QuadTech. They sell certain GenRad products under the logo iET—GenRad.)

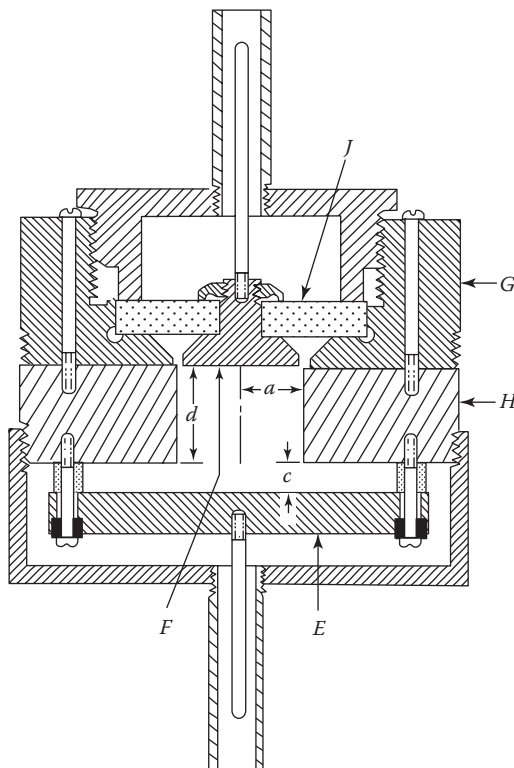


FIGURE 1.15

Cross section through a secondary standard, air dielectric, guard-well capacitor. *Note:* F and E are insulated electrodes. J is the borosilicate glass dielectric. G , H , K , and L are grounded shield conductor pieces. Note that this capacitor is a three-terminal device; both top and bottom electrodes have capacitances to ground, as well as to each other. (From Stout, M.B., *Basic Electrical Measurements*, 2nd edn., Prentice-Hall Inc., Englewood Cliffs, NJ, 1960. With permission.)

1.4.2.5 Inductance

The primary inductance standard is a derived quantity, being measured using an AC bridge circuit using standard resistors and capacitors. We have seen how the quantum Hall ohm can be used to calibrate working standard resistors and the calculable capacitor may be used to establish working standard capacitors. By using a bridge, such as the Maxwell (cf. Section 5.5.2.1), having three accurate resistors and an accurate capacitor, unknown inductances can be measured to a fair degree of accuracy.

Fixed-value, standard inductors are generally wound in solenoidal form on glass or ceramic forms having spiral grooves cut into their surfaces to precisely space the windings. Low-value (μH) inductors have single windings; however, millihenry and larger inductors may have groups of series-connected windings set into evenly spaced slots in a ceramic core. Inductors with solenoidal windings are subject to the external influences of magnetic materials and time-varying magnetic fields, and thus care must be taken when using them not to place them on or near steel tables or cabinets and to keep them well away from other current-carrying solenoidal coils. Modern secondary standard inductors, such as the venerable GenRad model 1491 series of switchable decade inductors, are wound on ferrite *toroidal cores*, from which there is little flux leakage and small influence from nearby magnetic materials or time-varying magnetic fields. Another design that has been used for secondary standard inductors with fixed values in the millihenry range and larger is the use of two, series-connected coils wound on semicircular, *D-shaped* forms, mounted so that one coil can be rotated with respect to the other so that the area of overlap of the enclosed areas of each coil can be adjusted. This rotation allows variation of the coils' mutual inductance and provides a means of adjusting the series coils' inductance in accordance with the following relation:

$$L = L_1 + L_2 + 2M, \quad (1.55)$$

assuming that the mutual inductance is aiding. The same principle of varying the mutual inductance of two series coils has also been used to make continuously variable inductors (Brooks inductometer). The Brooks inductometer's calibration is $\pm 0.3\%$ of its maximum value (Stout 1960).

A problem in the construction of all standard inductors is that their inductance depends on their dimensions, so any small change in dimensions caused by temperature changes can lead to errors. Also, as you will see in Section 5.1, inductors, however made, are fraught with stray capacitance between their windings and between layers and groups of windings. The wire from which they are wound also has a finite resistance. As the operating frequency of an inductor increases, the capacitive effects become more dominant and can cause significant errors in the apparent inductance. Thus, all standard inductors are generally calibrated at a convenient frequency such as 1 kHz and are designed to be used at a constant temperature, such as 25°C . Inductors wound on laminated iron cores or sintered ferrite cores also have the additional limitations of core losses due to hysteresis and eddy currents and core saturation at magnetizing force (ampere turns).

It may certainly be said that of all the electrical quantities, natural inductance is the least pure, being always associated, at a given frequency, with a series resistance and distributed, interwinding capacitance. Inductor equivalent circuits are described in Section 5.1. At superconducting temperatures, for air-core coils of certain metals, the coil resistance disappears, but the interwinding capacitance remains.

Another way of obtaining a very pure inductance, ranging in values from microhenries to kilohenries in the audio frequency range, is by means of the generalized impedance

converter (GIC) active circuit element (see Section 2.4.2.3). One end of the GIC circuit must be grounded. It is shown that the GIC's input impedance is given by

$$Z_{11} = \frac{Z_1 Z_3 Z_5}{Z_2 Z_4}. \quad (1.56)$$

Thus, if the circuit element for Z_2 is made a precisely known capacitor of high quality, such as a silver mica secondary standard, the remaining GIC impedance elements are resistors calibrated from the quantum Hall ohm, and the OAs are high-quality ICs having large-gain-bandwidth products, high input impedances, and low noise, we will see a virtually lossless inductance with value given by

$$L_{\text{eq}} = \frac{R_1 C_2 R_3 R_5}{R_4} \text{ H}. \quad (1.57)$$

It should be pointed out that the GIC circuit can also be used to rescale the size of capacitors. In this case, $Z_{11} = 1/\omega C_{\text{eq}}$. It is easy to see that if Z_1 is a capacitor, then

$$C_{\text{eq}} = \frac{C_1 R_2 R_4}{R_3 R_5} \text{ F}. \quad (1.58)$$

In using the GIC circuit, we must be careful to avoid OA voltage or current saturation and to operate the circuit at frequencies well below the OAs' f_T .

1.4.3 Time and Frequency Standards

The basic SI and MKS unit of time is the second. The basic unit of frequency is the *hertz*, or cycle per second of a periodic waveform. Obviously, time and frequency are related, in a measurement sense. The earliest definitions of the second were based on the Earth's rotational velocity, which is now known to vary over the period of a year due to seasonal changes in the Earth's moment of inertia due to the buildup and melting of polar ice caps, etc. The *second* was first defined as 1/86,400 of the length of a mean solar day. The *mean solar day* is the average time of 365 consecutive days. Refinements in the measurement of the period of the Earth's rotation led to the definitions for *universal time* (UT). In a search for a more precise definition for the second, astronomers developed the units of *ephemeris time* (ET). The ephemeris second was defined as 1/31,556,925.9747 of the tropical year (Helfrick and Cooper 1990). The ET standard was impractical as it took several years to determine the length of the tropical year and it required precise sightings of the positions of the sun and the moon.

The *primary time/frequency standard* in current use is the cesium 133 atomic beam clock. A simplified diagram of the cesium clock is shown in Figure 1.16. This standard is discussed in detail in Chapter 2. To summarize the cesium clock's properties, it oscillates at 9.192631770 GHz, having an effective Q of 2×10^8 . Coordinated universal time (UTC), whose basic unit is the second, is defined as 9,192,631,770 periods of the cesium 133 beam oscillator. This international standard was adopted in October 1967 and is more accurate than any clock calibrated by astronomical measurements (Oliver and Cage 1971).

The venerable HP/Agilent model 5061B (Opt 004) cesium beam frequency standard has a long-term stability of $\pm 2 \times 10^{-12}$ over the life of the cesium beam tube. Accuracy is also $\pm 2 \times 10^{-12}$. The HP5051B cesium clock has sinusoidal outputs at 10, 5, 1, and 0.1 MHz. HP cesium clocks are used to calibrate and synchronize the LORAN-C and GPS radio

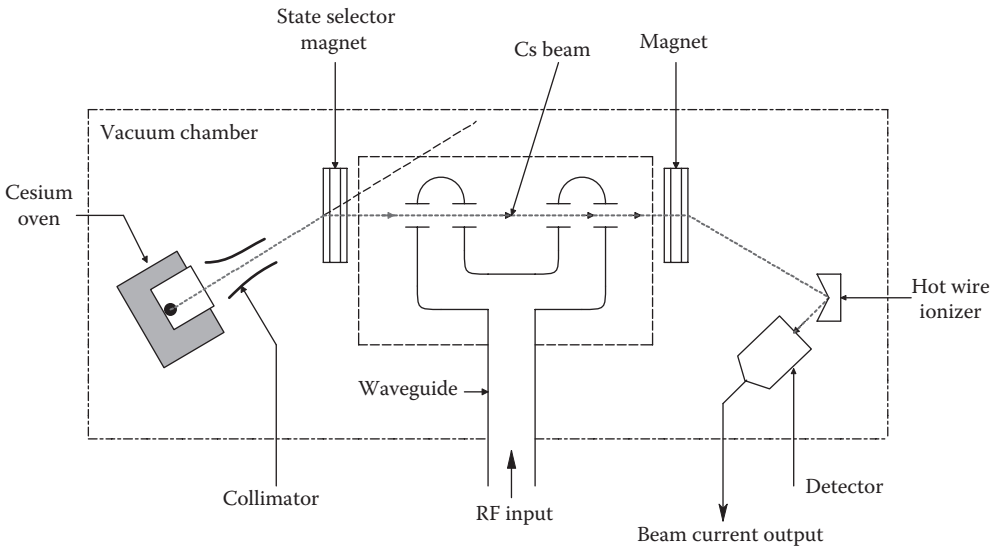


FIGURE 1.16
Simplified schematic diagram of a cesium ion clock.

navigation systems. Cesium (Cs) beam clocks are ordinarily used to adjust secondary rubidium and quartz oscillators used as secondary standards for frequency or period determination. (In 2005, HP/Agilent sold their cesium standard timekeeping business to Symmetricom® [www.symmetricom.com].)

As of 2011, the National Physical Laboratory (United Kingdom) improved its NPL-CsF2 cesium fountain clock so that it had an evaluated frequency uncertainty of 2.3×10^{-16} . With this frequency uncertainty, the NPL-CsF2 is expected to neither lose nor gain a second in more than 138 million years (Lia et al. 2011).

Rubidium frequency standards are second in the hierarchy of accuracy. Similar to the operation of a cesium beam clock, the atomic resonance of a rubidium vapor cell is used to synchronize a quartz crystal oscillator in a frequency-lock loop. The long-term stability of the rubidium vapor oscillator is $\pm 1 \times 10^{-11}$ /month. It, too, has outputs at 5, 1, and 0.1 MHz.

1.4.4 Physical Standards

The physical standards discussed in the following include mass, length, volume, and temperature. We have seen earlier that the present trend is to redefine standard quantities in terms of the (presumably) universal constants of quantum physics, rather than to use physical artifacts (e.g., the present use of the JJ to define the volt instead of the Weston normal cell).

1.4.4.1 Mass

The present (SI) unit of mass is the kilogram (kg). It is presently defined in terms of the standard kilogram artifact, which has been kept at the *Bureau International des Poids et Mesures* (BIPM) at Sèvres, France, for over 100 years. Physically, the standard kilogram is a cylinder of platinum—10% iridium. Comparison of a working standard kilogram mass with the primary standard must be made by means of a precision balance. The current *gold standard* balance is the NBS-2, single-arm, two-knife balance, which has a resolution of one microgram or one part in 10^9 (Quinn 1991). What has been discovered over the years

is that the international prototype kilogram and its copies systematically lose mass when they are cleaned and washed in preparation for weighing. On the other hand, the masses tend to increase with time following cleaning and washing. These changes are on the order of tens of micrograms (Quinn 1991). The long-term stability of the standard kilogram appears to be on the order of 5 $\mu\text{g}/\text{year}$. This figure is based on the assumption that international prototype and its copies are not drifting in mass by more than 10 times the rate that they are drifting apart from each other at about 0.5 $\mu\text{g}/\text{year}$.

A number of workers are considering ways to replace the present artifact standard kilogram at BIPM with a mass standard based on the fundamental, physical constants, which include the *electron charge* (q), *Planck's constant* (h), the *speed of light in vacuo* (c), and the *permeability of vacuum* (μ_0). Other constants that can be used are the *von Klitzing constant* (R_{K-90}), the *Josephson constant* (K_{J-90}), the *fine structure constant* (α), the *Faraday number* (\mathfrak{F}), and *Avogadro's number* (N_A).

Taylor (1991) argued that Avogadro's number, which is the number of molecules in a gram-molecular weight, or mole, of some pure element (X) can be one determining factor, along with the molar mass of the element (M_X) in kg/mol . For example, $M(^{12}\text{C})$ is equal to 0.012 kg , by definition (^{12}C is the common isotope of the element carbon). Thus,

$$n_X = \frac{N_A}{M_X}, \quad (1.59)$$

where n_X is the number of free X s at rest required to make the mass of an international kilogram prototype.

There are a number of interesting ways to determine Avogadro's number. Taylor (1991) proposed that N_A can be found from the x-ray crystal density of a pure, grown silicon crystal from the following relation:

$$N_A = \frac{M(\text{Si})}{\left[\delta(\text{Si}) (d_{220} \sqrt{8})^3 / 8 \right]}, \quad (1.60)$$

where

$M(\text{Si})$ is the mean molar mass of silicon in kg/mol

$\delta(\text{Si})$ is the crystal density in kg/m^3

d_{220} is the 2,2,0 silicon lattice spacing in m

$M(\text{Si})$, $\delta(\text{Si})$, and d_{220} must be measured.

In a practical determination of N_A , the ratio of the three, natural occurring isotopes (^{28}Si , ^{29}Si , ^{30}Si) in a crystal must be measured. Another way to find N_A indirectly is to use the following relation:

$$N_A = \frac{K_{J-90} R_{K-90} \mathfrak{F}}{2}, \quad (1.61)$$

where

$K_{J-90} = (2q/h)$ is the Josephson constant

$R_{K-90} = (h/q^2)$ is the von Klitzing constant

$\mathfrak{F} = 96,486.7$ is the Faraday number (the coulomb charge of Avogadro's number of electrons)

Perhaps in the near future, progress will be made in replacing the standard kilogram artifact with a standard kilogram defined by fundamental constants. Uncertainty in determining such a new standard may be on the order of 1×10^{-8} .

1.4.4.2 Length

The SI standard unit of length is the meter. Formerly, the meter was defined as one ten millionth of the arc distance of a meridian passing from the north pole through Paris to the equator. It was represented by an artifact consisting of a platinum–iridium bar with lines scribed on it one meter apart, kept at constant temperature at the BIPM near Paris. In 1960, the SI meter was redefined in terms of the wavelength of monochromatic light. One meter was exactly 1,650,763.73 wavelengths in vacuum of the orange radiation corresponding to the transition between the levels $2p_{10}$ and $5d_5$ of the Krypton-86 atom. However, this definition of the meter by wavelength assumes the permanence and invariance of the atomic energy levels, Planck's constant, and the speed of light. These quantities are basic to the relation for the following wavelength:

$$\lambda = \frac{c}{\nu} = \frac{ch}{E} = \frac{ch}{E_2 - E_1}. \quad (1.62)$$

To eliminate the requirement for h and ΔE , the invariant speed of light, c , was used in 1983 to redefine the standard meter as the distance light travels in free space in $1/299,729,458$ s (3.3363420 ns). Frequency-stabilized laser sources are used to realize this definition (Sirohi and Kothiyal 1991).

Gauge blocks are used as secondary standards for length. These are generally platinum alloys or stainless steel cubes or blocks whose dimensions are established interferometrically to within a fraction of a wavelength of light and whose faces are polished to optical smoothness. Gauge blocks are used to calibrate micrometers and to set up precision mechanical and optical systems.

1.4.4.3 Temperature

The SI unit of temperature is the Kelvin. The standard reference temperature is defined by the *triple point of water*, at which pressure and temperature are adjusted so that ice, water, and water vapor simultaneously exist in a closed vessel. The triple point of pure water occurs at $+0.0098^\circ\text{C}$ (273.16 K) and 4.58 mmHg pressure. The Kelvin was defined in 1967 at the 13th CGPM to be the unit of thermodynamic temperature, equal to $1/273.16$ of the thermodynamic temperature of the triple point of water. Other primary, fixed temperature points are also used for temperature calibration: the boiling point of O_2 (-182.97°C), the boiling point of sulfur (444.6°C), the freezing point of silver (960.8°C), and the freezing point of gold (1063°C), all at atmospheric pressure. Absolute zero is at 0 K, or -273.15°C . Obviously, Celsius and Kelvin degrees are the same *size*.

Because so many physical and chemical phenomena are strong functions of temperature, there are many ways to measure temperature, as will be seen in Chapters 6 and 7.

1.4.4.4 Uncertainties in the SI Base Units

The approximate uncertainties in the quantification of the *fundamental* SI units determined by international treaty signed in 1960 are given in Table 1.4.

As shown in Table 1.4, all other physical and electrical units are derived from the basic SI units (L , T , M , I , etc.). For example, *charge*, Q , has the dimensions of IT . *Force* in newtons, F , has the dimensions of MLT^{-2} ; this follows from Newton's $F = ma$. The *volt* (unit of potential difference) has the dimension of joule/coulomb; in SI units, the volt is $ML^2T^{-3}I^{-1}$. The MKS units of the volt are $ML^2T^{-2}Q^{-1}$. *Current density*, J , in amperes/meter has the SI dimensions of IL^{-1} ; the MKS dimensions of J are $QT^{-1}L^{-1}$. Finally, the SI dimensions of the *ohm* are $ML^2T^{-3}I^{-2}$.

TABLE 1.4

Uncertainties in the Basic SI Units

Quantity	Unit	Approx. Uncertainty
Length	Meter (m) {L}	3×10^{-11}
Time	Second (s) {T}	1×10^{-13}
Mass	Kilogram (kg) {M}	5×10^{-9}
Electric current	Ampere (A) {I}	1×10^{-6}
Temperature	Kelvin (K) {Θ}	2.5×10^{-4}
Luminous intensity	Candela (cd)	1.5×10^{-2}
Amount of substance	Mole (mol)	TBD

1.5 Chapter Summary

In this introductory chapter, we have introduced the architecture of typical measurement systems and have shown that they contain not only sensors, signal conditioning operations, and data display, storage, and retrieval options but also sources of noise, which can limit measurement precision. A measurement error is shown to occur if sensor dynamics are not considered and the sensor is read before it reaches an SS output, given a transient change in the value of the QUM.

The types of errors in measurements are considered. Elementary statistics are discussed and the important concept of LE in a measurement is introduced and derived; examples are given for calculating LE.

Standards in electrical and physical measurements are described. The important SI standards for voltage (the JJ), resistance (the QHR), and capacitance (the Thompson–Lampard theorem) are discussed in detail. The reader is advised that primary standards are constantly evolving with science and technology and accuracies of one part in 10^7 are routine; many systems exceed this accuracy.

The following chapter addresses the important topic of analog signal conditioning applied to the analog output voltages or currents from I&M system input sensors. Chapter 2 also treats the important topics of phase-sensitive rectifiers (PSRs) and signal averaging to improve output SNR.

Problems

- 1.1** One of the earliest means devised for measuring moderate pressure differences was the U tube, liquid-filled manometer (see Figure P1.1). Liquids used are typically mercury or colored water. The dynamics of a manometer can be described by a second-order, linear ODE:

$$\Delta PA = \rho g h A + \left(\frac{\dot{h}}{2} \right) D (2\pi r L) + \left(\frac{\ddot{h}}{2} \right) (\rho L A).$$

From Figure P1.1, we see that L is the total length of liquid in the U tube, A is the area of liquid exposed to pressures P_1 and P_2 , $A = \pi r^2$, and $r =$ tube lumen radius. $\Delta P = P_1 - P_2$,

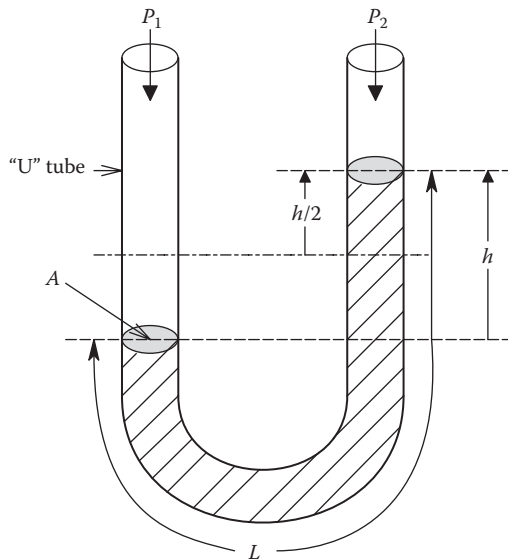


FIGURE P1.1

ρ = liquid density in kg/m^3 , D = viscous damping constant, g = acceleration of gravity in m/s^2 , and h = height difference between the two sides of the manometer.

- A. Derive a transfer function for the manometer's dynamic response, $\frac{h}{\Delta P}(s)$, in time-constant form.
 - B. Give expressions for the manometer's undamped natural frequency, ω_n , and its damping factor, ζ .
 - C. Calculate the numerical value for ω_n and D , given $\zeta = 0.4$. Assume water at 20°C is the liquid. Let $L = 1 \text{ m}$, $A = 3.142 \times 10^{-6} \text{ m}^2$.
- 1.2 An accelerometer has critically damped, second-order, low-pass dynamics with the transfer function:

$$\frac{V_o}{\ddot{X}}(s) = \frac{K_A}{(\tau s + 1)^2},$$

where

$$\tau = 1 \text{ s}$$

$$K_A = 0.001 \text{ V/(m/s}^2\text{)}$$

Given a step input of acceleration at $t = 0$, find how long it will take $v_o(t)$ to reach the following: (A) 95% of its SS value, (B) 99% of its SS value, and (C) 99.9% of its SS value.

- 1.3 A piezoelectric pressure sensor and its signal conditioning amplifier have a transfer function approximated by

$$\frac{V_o}{P}(s) = \frac{-10s}{s100 + 1}.$$

The input is a 100 psi pressure step. (A) Sketch and dimension $v_o(t)$. (B) Find the peak $v_o(t)$. (C) Find the time for $v_o(t)$ to drop to 99.5% of its peak value.

- 1.4 An electronic temperature sensor converts temperature in degrees Kelvin to output microamperes. When it is immersed in water, its transfer function is

$$\frac{I_o}{T}(s) = \frac{1.5 \times 10^{-2}}{(s + 0.3)(s + 0.05)} \mu\text{A}/\text{K}.$$

- A. Sketch and dimension $i_o(t)$; when at time $t = 0$, the sensor is taken out of a glass of 20°C water and immersed in a glass of 55°C water.
- B. How long will it take $i_o(t)$ to reach 99.8% of its SS value?
- 1.5 The pressure sensor of Problem 1.3 is given an input: $p(t) = 100[1 + \sin(2\pi ft)]$ psi for a long time. Make a Bode plot of $\text{dB} = 20 \log_{10} |V_o/P(jf)|$ vs. f (log scale).
- 1.6 Calculate the net DC open-circuit output voltage of a series array of $N = 550$ JJs at 4.2 K, given a series direct current of 24.00 mA and a microwave irradiation frequency of $f_o = 9.0646279$ GHz. Operation is on the $n = 97$ th step. Give the output voltage to six significant figures.
- 1.7 A QHR and a JJ array are put in series so that the direct current, I , passes through the QHR and also passes through the JJ array (see Figure P1.7). The B field is adjusted on the QHR so that it is on its $k = 7$ th step. I_x is set at 150 μA , and the JJ array is operated so all of the JJs are on step $n = 1$. The microwave frequency irradiating the JJ array is adjusted to $f_o = 13.374664$ GHz to make $V_{\text{QH}} = V_{\text{JJ}}$.
- A. Under the condition of $V_{\text{QH}} = V_{\text{JJ}}$, find a general expression for I_x in terms of q , h , n , k , f_o , and N .
- B. Find the numerical value of the number of JJs, N , in the JJ array required to meet the aforementioned conditions.
- C. Find the numerical value of $V_{\text{QH}} = V_{\text{JJ}}$ to six significant figures. Note that the electron charge $q = 1.60217733 \times 10^{-19}$ Cb and Planck's constant $h = 6.6260755 \times 10^{-34}$ J s.
- 1.8 Use *MicroCap* or *PSPICE* to simulate the frequency response of a GIC-synthesized 1.000 H inductor. The GIC schematic is shown in Figure P1.8. Note that the accuracy

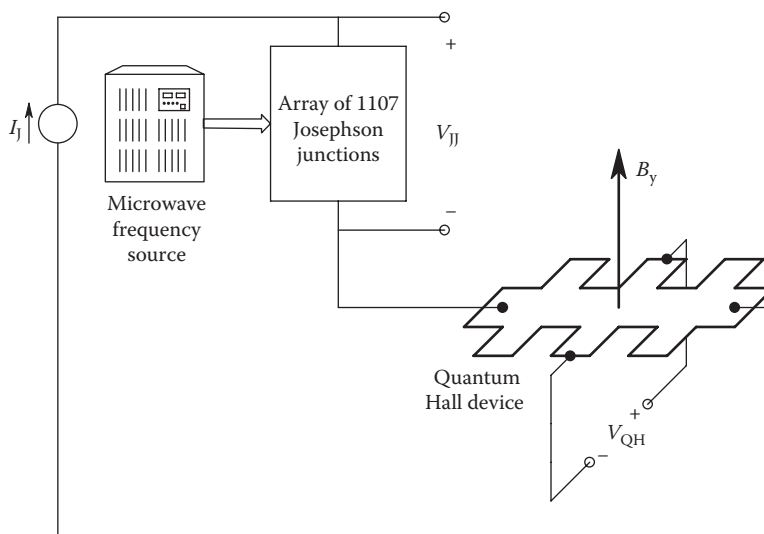


FIGURE P1.7

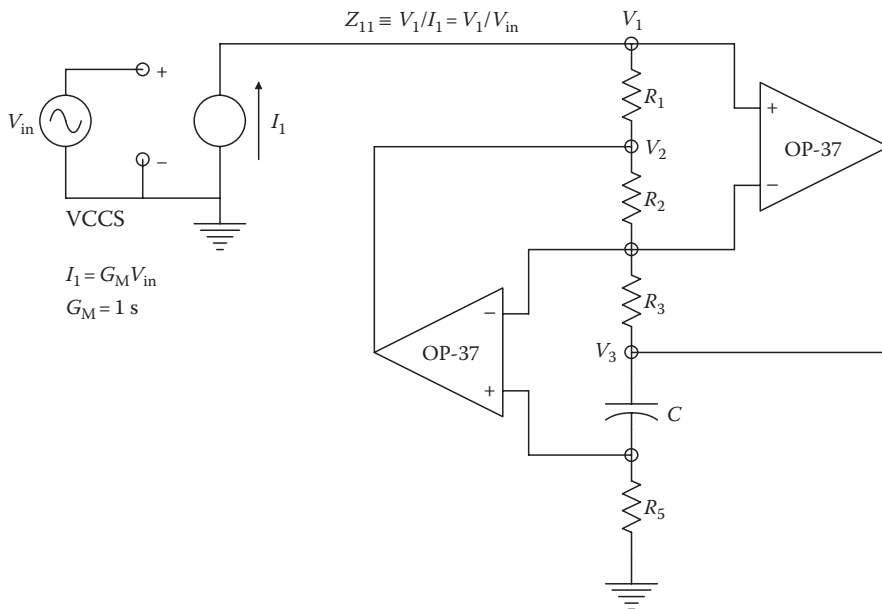


FIGURE P1.8

of the simulated inductor will degrade at high frequencies because of the finite gain \times bandwidth and slew rate of the nonideal (real) OAs used. You are to use three $10,000\ \Omega$ and one $1,000\ \Omega$ resistors, a $0.001\ \mu\text{F}$ silver mica capacitor, and two OP-37 OAs. The OA parameters are as follows: $R_{in} = 6\ \text{M}\Omega$, $A_{OL} = 1.8\text{E}6$, $R_{out} = 70\ \Omega$, $f_1 = 44.4\ \text{Hz}$, $f_2 = 3.7\text{E}7\ \text{Hz}$, $\eta = 2.0\text{E}7\ \text{V/s}$, $V_{OS} = 10\ \mu\text{V}$, $I_B = 10\ \text{nA}$, $I_{OS} = 7.0\text{E}7$, ($f_T = 80\ \text{MHz}$). The VCCS is used to make simulation with MicroCap easier; note that $Z_{11} \equiv V_1/I_1$.

- A. Make Bode and magnitude and phase plots of $Z_{11}(f)$. Indicate the range of frequencies over which the GIC inductor simulation is valid.
 - B. Find the frequency where the phase of $Z_{11}(f)$ departs from 90° by $\pm 0.1^\circ$. What is the equivalent resistance that appears in series with the simulated inductor at this frequency?
 - C. Give the simulated inductor's Q at this frequency.
- 1.9 Find the LE in the midfrequency L_{eq} synthesized by the GIC circuit of Figure P1.8. The inductance can be shown to be $L_{eq} = C_4 R_1 R_3 R_5 / R_2$. C_4 is known to 0.05%, R_2 to 0.01%, and R_1 , R_3 , and R_5 to 0.002%.

2

Analog Signal Conditioning in Instrumentation

2.1 Introduction

Practically all instrumentation systems require some type of analog signal conditioning between the analog input sensor and the processing, data display, and storage subsystems. In its simplest form, analog signal conditioning can be voltage amplification with a change in impedance level between the conditioning amplifier's input and output.

Analog signal conditioning may also involve linear filtering in the frequency domain, such as band-pass filtering to improve SNR at the amplifier's output. In other cases, the analog input to the signal-conditioning system may be processed nonlinearly. For example, depending on system requirements, the output of the analog signal conditioner may be made proportional to the square root of the input, to the RMS value of the input, to the logarithm of the input, to the cosine of the input, etc.

Analog signal conditioning is often accomplished by use of operational amplifiers, as well as special IAs, isolation amplifiers (IsoAs), analog multipliers, and/or dedicated, nonlinear, analog signal processing ICs. In some modern instrumentation systems, the amplified and filtered signal from the sensor is directly digitized, and the nonlinear operations are performed numerically on the samples by the host DSP computer.

In the following sections of this chapter, we examine the properties of the analog IC systems used in analog signal conditioning in instrumentation systems, beginning with the differential amplifier (DA), which has wide use in the headstages of nearly all types of OAs, IAs, IsoAs, analog multipliers, oscilloscopes, etc.

2.2 Differential Amplifiers

2.2.1 Introduction

DAs are widely used as input stages in a wide variety of amplifier types including OAs, analog comparators, analog multipliers, IAs, oscilloscope vertical and horizontal amplifiers, and certain specialized ICs. There are two important reasons for their wide use in signal-conditioning system designs: The first and foremost is the ability of the DA to respond to the difference in the input signals ($V_i - V_i'$) and to discriminate against noise or interference common to both inputs. Such a common-mode (CM) input voltage that is desired to be eliminated is often a DC level, hum, or other coherent interference. Another reason for using a DA headstage is that it inherently discriminates against changes in the amplifier's DC power supply voltages.

2.2.2 Analysis of Differential Amplifiers

Figure 2.1 illustrates the most general form of DA. Note that including ground, it is a four-port circuit. Most practical DAs have only a single-ended output, V_o , and are thus three-port circuits. The deflection amplifier circuits of most analog oscilloscopes are an exception to this rule, preserving the differential circuit architecture from input to the output to the cathode ray tube (CRT) deflection plates.

In the analysis of DAs, it is expedient to define *difference-mode (DM)* and *CM* input and output signals:

$$V_{id} \equiv \frac{V_i - V_i'}{2} \text{ (DM input),} \quad (2.1)$$

$$V_{ic} \equiv \frac{V_i + V_i'}{2} \text{ (CM input),} \quad (2.2)$$

$$V_{od} \equiv \frac{V_o - V_o'}{2} \text{ (DM output),} \quad (2.3)$$

$$V_{oc} \equiv \frac{V_o + V_o'}{2} \text{ (CM output).} \quad (2.4)$$

The input/output (I/O) relationships for a general DA can be expressed in terms of the *Middlebrook equations*:

$$V_{od} = A_{dd}V_{id} + A_{dc}V_{ic}, \quad (2.5)$$

$$V_{oc} = A_{cd}V_{id} + A_{cc}V_{ic}. \quad (2.6)$$

The four gains of the voltage-controlled voltage sources (VCVSs) in Equations 2.5 and 2.6 are written in vector format to stress that they are functions of frequency ($j\omega$).

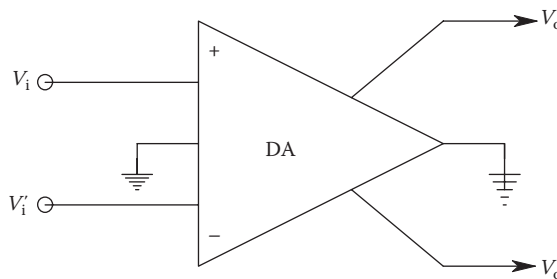


FIGURE 2.1

Symbol for a generalized DA: note that it is a four-port circuit (V_i , V_i' , V_o , and V_o').

It is easy to show that the single-ended outputs are given by

$$V_o = (A_{dd} + A_{cd})V_{id} + (A_{dc} + A_{cc})V_{ic}, \quad (2.7)$$

$$V'_o = (A_{cd} - A_{dd})V_{id} + (A_{cc} - A_{dc})V_{ic}. \quad (2.8)$$

The *single-ended output*, V_o , can also be written in terms of the actual DA's inputs:

$$V_o = \frac{A_{dd} + A_{cd} + A_{dc} + A_{cc}}{2} V_i + \frac{A_{cc} + A_{dc} - A_{dd} - A_{cd}}{2} V'_i. \quad (2.9)$$

Often it is expedient to write

$$V_o = A_D V_{id} + A_C V_{ic}, \quad (2.10)$$

where A_D and A_C are given by the corresponding terms in Equation 2.7. In an *ideal DA*, $A_{cc} = A_{dc} = A_{cd} = 0$, so in this case, $A_C = 0$ and $A_D = A_{dd}$ in Equation 2.10 given earlier.

2.2.3 Common-Mode Rejection Ratio

The common-mode rejection ratio (CMRR) is an important figure of merit for DAs, usually given in decibels. It describes how well a real DA's behavior approaches that of an ideal DA. The scalar CMRR, measured at some frequency, is defined as

$$\text{CMRR} \equiv \frac{|V_{ic}| \text{ to give a certain output magnitude}}{|V_{id}| \text{ to give the same output}}. \quad (2.11)$$

Using the definition for the aforementioned CMRR and Equation 2.7, we find

$$\text{CMRR} = \left| \frac{A_{dd} + A_{cd}}{A_{cc} + A_{dc}} \right|. \quad (2.12)$$

If the DA circuit is perfectly symmetrical, then $A_{dc} = A_{cd} = 0$, and the CMRR magnitude at a given frequency $f = \omega/2\pi$ Hz reduces to

$$\text{CMRR}(f) = \left| \frac{\mathbf{A}_{dd}}{\mathbf{A}_{cc}} \right| = \left| \frac{\mathbf{A}_D}{\mathbf{A}_C} \right|. \quad (2.13)$$

State-of-the-art instrumentation and IsoAs may have CMRRs of 120 dB or more ($\geq 10^6$), under ideal conditions. Generally, CMRR *decreases with frequency*, and as we will see, an imbalance in the Thevenin source resistances at the input of a DA can lead to either a loss or an increase in CMRR over the manufacturer's specified value.

2.2.4 Measurement of CMRR, A_D , and A_C

Often to save money, an engineer puts together a differential IA from OA *building blocks*. (The design of such amplifiers is covered in Section 2.2.2.) In order to measure the DA's CMRR, we make use of Equation 2.11. The two inputs are shorted together and a sinusoidal signal of frequency f and amplitude V_{s1} is applied to them, generating a CM input signal, V_{s1c} , which is adjusted to give an output, V_o . Next, the negative input is grounded, so $V_i' = 0$. The sinusoidal signal source of frequency f and amplitude V_{s2} is now connected to the + input terminal of the DA. This procedure is necessary because of the difficulty in generating a perfectly balanced DM input signal. The single-ended input generates both DM and CM signals: $V_{id} = V_{s2}/2$, and $V_{ic} = V_{s2}/2$. V_{s2} is adjusted to make $V_{o2} = V_o$ of the CM case. Thus, from Equation 2.10,

$$A_C = \frac{V_o}{V_{s1}}, \quad (2.14)$$

and also, considering the single-ended input, V_{s2} , we have

$$V_o = \frac{A_D V_{s2}}{2} + \frac{A_C V_{s2}}{2}. \quad (2.15)$$

If Equation 2.14 for A_C is substituted into Equation 2.15, we may solve for the DM gain, A_D :

$$A_D = V_o \left(\frac{2}{V_{s2}} - \frac{1}{V_{s1}} \right). \quad (2.16)$$

Hence, the experimentally found CMRR can be expressed as Equation 2.17 using Relations 2.14, 2.16, and 2.13:

$$\text{CMRR} = \frac{A_D}{A_C} = \left(\frac{2V_{s1}}{V_{s2} - 1} \right). \quad (2.17)$$

If the amplifier's scalar A_D is very large, a precision attenuator may be needed to reduce the output of the signal source to an appropriately small V_{s2} .

2.2.5 Effect of Source Resistance Asymmetry on CMRR

Not only does CMRR decrease with input signal frequency (largely due to an increase of A_C with frequency), but it is also severely affected by unbalance in the Thevenin resistance (or impedance) of the sources connected to the DA's input leads. Figure 2.2 illustrates the DC equivalent input circuit of a typical DA. Manufacturers typically specify a CM input resistance, R_{ic} , measured from one input lead to ground under CM excitation and a DM input resistance, R_{id} , measured under DM excitation from either input to ground.

By Ohm's law, the DM current into the noninverting input node is just

$$I_d = \frac{2V_{id}}{R_1} + \frac{V_{id}}{R_{ic}}, \quad (2.18)$$

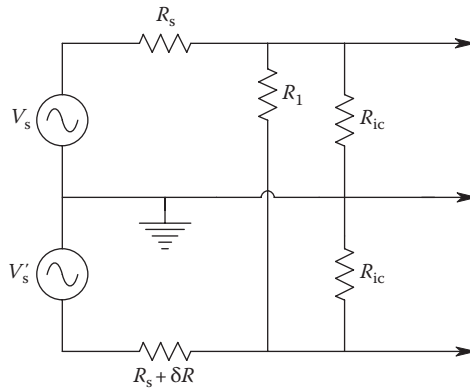


FIGURE 2.2
DA input circuit at DC, showing source resistance unbalance.

from which we can write

$$\frac{I_d}{V_{Id}} = \frac{1}{R_{Id}} = \left(\frac{2}{R_1} + \frac{1}{R_{ic}} \right) \tag{2.19}$$

Solving for the shunting resistance in Equation 2.19, we get

$$R_1 = \frac{2R_{id}R_{ic}}{(R_{ic} - R_{id})} \tag{2.20}$$

In many DAs, $R_{ic} > R_{id}$. If $R_{ic} = R_{id}$, then from Equation 2.20, $R_1 = \infty$.

Let us assume that $R_{ic} = R_{id}$. Thus, R_1 may be eliminated from Figure 2.2, which illustrates two Thevenin sources driving the DA through unequal source resistances, R_s and $R_s + \Delta R$. Using superposition and the definitions in Equations 2.1 and 2.2, it is possible to show that a purely CM excitation, V_{sc} , produces an unwanted DM component at the DA's input terminals, sic:

$$\frac{V_{id}}{V_{sc}} = \frac{R_{ic}\Delta R}{2(R_{ic} + R_s)^2} \tag{2.21}$$

Also, V_{sc} produces a large CM component; the ΔR term is numerically negligible:

$$\frac{V_{ic}}{V_{sc}} = \frac{R_{ic}}{(R_{ic} + R_s)} \tag{2.22}$$

For purely DM excitation in V_s , we can also show that

$$\frac{V_{id}}{V_{sd}} = \frac{R_{ic}}{(R_{ic} + R_s)}' \tag{2.23}$$

and

$$\frac{V_{ic}}{V_{sd}} = \frac{R_{ic}\Delta R}{2(R_{ic} + R_s)^2}. \quad (2.24)$$

In order to find the CMRR of the circuit of Figure 2.2, we will use Equation 2.10 for V_o and the definition for CMRR, Relation 2.11. Thus,

$$V_o = \frac{A_D V_{sc} R_{ic} \Delta R}{2(R_{ic} + R_s)^2} + \frac{A_C V_{sc} R_{ic}}{(R_{ic} + R_s)}, \quad (2.25)$$

and

$$V_o = \frac{A_D V_{sd} R_{ic}}{(R_{ic} + R_s)} + \frac{A_C V_{sd} R_{ic} \Delta R}{2(R_{ic} + R_s)^2}. \quad (2.26)$$

After some algebra, we find that the circuit's CMRR, $CMRR_{sys}$, is given by

$$CMRR_{sys} = \frac{\left[A_D + A_C \Delta R / 2(R_{ic} + R_s) \right]}{\left[A_D \Delta R / 2(R_{ic} + R_s) + A_C \right]}. \quad (2.27)$$

Equation 2.27 may be reduced to the hyperbolic relation:

$$CMRR_{sys} = \frac{\left[(A_D/A_C) + \Delta R / 2(R_{ic} + R_s) \right]}{\left[(A_D/A_C) \Delta R / 2(R_{ic} + R_s) + 1 \right]}, \quad (2.28)$$

which can be approximated by

$$CMRR_{sys} \cong \frac{CMRR_A}{\left[CMRR_A \Delta R / 2(R_{ic} + R_s) + 1 \right]} \quad (2.29)$$

in which the manufacturer-specified $CMRR = CMRR_A = A_D/A_C$ and $CMRR_A \gg \Delta R / 2(R_{ic} + R_s)$. A plot of $CMRR_{sys}$ vs. $\Delta R/R_s$ is shown in Figure 2.3. Note that when the Thevenin source resistances are matched, $CMRR_{sys} = CMRR_A$. Also, when

$$\frac{\Delta R}{R_s} = \frac{-2(R_{ic}/R_s + 1)}{CMRR_A}, \quad (2.30)$$

$CMRR_{sys} \rightarrow \infty$. This implies that a judicious addition of an external resistance in series with one input lead or the other to introduce a $\Delta R > 0$ may be used to increase the effective CMRR of the system. For example, if $R_{ic} = 100 \text{ M}\Omega$, $R_s = 10 \text{ k}\Omega$, and $CMRR_A = 100 \text{ dB}$, then $\Delta R/R_s = -0.2$ to give ∞ $CMRR_{sys}$. Since it is generally not possible to reduce R'_s , it is easier to externally add a similar δR to R_s .

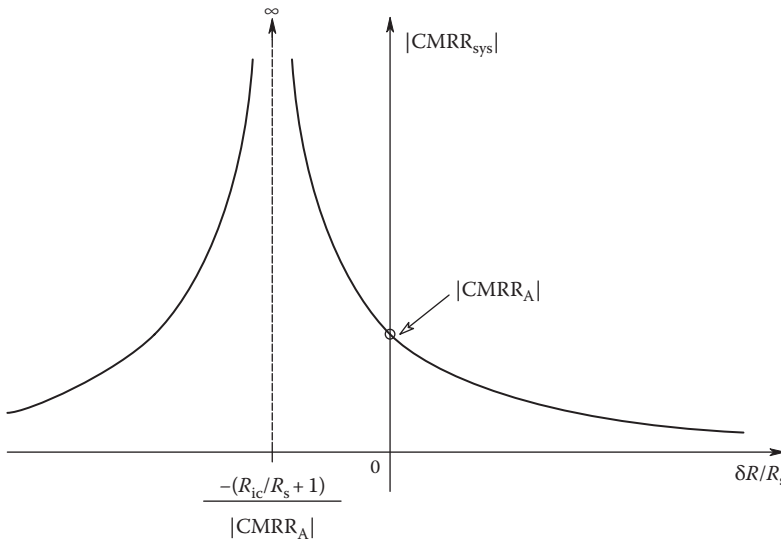


FIGURE 2.3
DA CMRR magnitude vs. fractional unbalance in source resistance, $\Delta R_s/R_s$.

Again, we stress that an amplifier’s $CMRR_A$ is a decreasing function of frequency because of the frequency dependence of the gains, A_D and A_C . Also, the AC equivalent input circuit of a DA contains capacitances in parallel with R_1 , R_{ic} , and R'_{ic} , and the source impedances often contain a reactive, frequency-dependent component. Thus, in practice, in a given range of frequencies, $CMRR_{sys}$ can often be maximized by the ΔR method but seldom can be drastically increased because of reactive unbalances in the input circuit.

2.3 Operational Amplifiers

In this section, we review the systems properties of IC operational amplifiers of various types and show the many ways engineers use them for signal conditioning in the design of instrumentation systems.

A typical OA is a DA with a single output. It has a very high DC gain, K_{vo} , and high CMRR. Because OAs are generally used with large amounts of negative feedback, the open-loop transfer function of most OAs is designed to be of the form

$$A_D = \frac{V_o}{V_i - V'_i} = \frac{K_{vo}}{[(\tau_1 s + 1)(\tau_2 s + 1)]} \tag{2.31}$$

The second break frequency, $f_2 = 1/(2\pi\tau_2)$, generally is made to occur so that $|A_D(jf_2)| \ll 1$. This ensures closed-loop stability over a wide range of feedback conditions. The break frequency, $f_1 = 1/(2\pi\tau_1)$, occurs at a relatively low value. However, a critical parameter governing

the closed-loop, high-frequency (HF) response of an OA circuit is the small-signal (SS) gain \times bandwidth product, which for most OAs can be approximated by

$$\text{GBWP} \cong \frac{K_{vo}}{2\pi\tau_1} \text{ Hz.} \quad (2.32)$$

The symbol, f_T , is the unity gain or 0 dB frequency of the open-loop OA. f_T is approximately equal to the gain \times bandwidth product. The larger the GBWP, the higher the -3 dB cutoff frequency that can be made in a broadband amplifier made with an OA.

Another manufacturer-specified parameter that determines the large-signal dynamic behavior of an OA is its *slew rate*, η . η is defined as the largest possible attainable rate of change of the output voltage, and its units are generally given in volts/microsecond. OA slew rates range from less than 1 V/ μ s in old models such as the venerable LM741 to 3000 V/ μ s for the Comlinear CLC300A, which is intended to condition video-frequency signals. As with the case of GBWP or f_T , the cost of OAs increases with increasing slew rate. If an OA signal-conditioner circuit is driven by an HF *and* high-amplitude sinusoidal source such that $|dV_o/dt|$ exceeds the amplifier's slew rate, then the output will appear to be a triangle wave with rounded (sinusoidal) peaks. When it occurs, slew rate limiting generates harmonic distortion in V_o . A prudent designer will always check to see that the maximum expected $|dV_o/dt|$ in an OA circuit does not exceed η .

While the parameters K_{vo} , f_T , η , f_1 , and f_2 govern an OA's dynamic behavior, manufacturers also give specifications on an OA's output resistance (R_o), CM and DM input resistances, DC bias current (I_B), DC offset voltage (V_{OS}), temperature coefficients for I_B and V_{OS} , short-circuit voltage noise root power spectrum [$e_{na}(f)$], and equivalent input noise current root power spectrum [$i_{na}(f)$], to mention the most important ones.

2.3.1 Types of Op-Amps

OAs may be categorized by characteristics and applications. As you will see in Chapter 3, low-noise OAs are best used in the headstages of analog signal-conditioning systems, because their noise characteristics determine the overall noise performance of the system. When great DC stability and freedom from thermally caused DC drift are required, then designers often specify *chopper-stabilized op-amps* (CSOAs), provided that other criteria for dynamic response and noise are met.

Electrometer op-amps (EOAs) are used when input bias currents of 10^{-13} A or less and input resistances of greater than 10^{13} Ω are required, other conditions being met.

Fast OAs are defined here as those having slew rates in excess of 25 V/ μ s and SS GBWPs of 75 MHz or more. Most OAs are designed to operate with ± 15 or ± 12 V DC supplies and can source about ± 10 mA to a load. Modern OAs have been designed to operate in portable instrumentation and communications equipment with supply voltages of ± 3.2 , $+5$ V (single-ended), etc.

We define *power and HV OAs* as those which can source in excess of 10 mA to a load or which can operate with HV supplies of greater than ± 15 V (typically ± 40 to ± 150 V). Some power OAs can deliver as much as 30 A at 68 V (Apex PA03). In instrumentation systems, power OAs can be used to drive strip-chart recorder pen motors or other EM and piezoelectric transducers. It is possible for an OA to fall in two or more of the categories described earlier; for example, the Apex PA84 has a GBWP of 75 MHz, a slew rate of 200 V/ μ s, a peak output voltage of 143 V, and a peak output current of 40 mA. Thus, it falls in the HV and high-power category as well as the fast category. See Apex (2012) for more information on power OA applications.

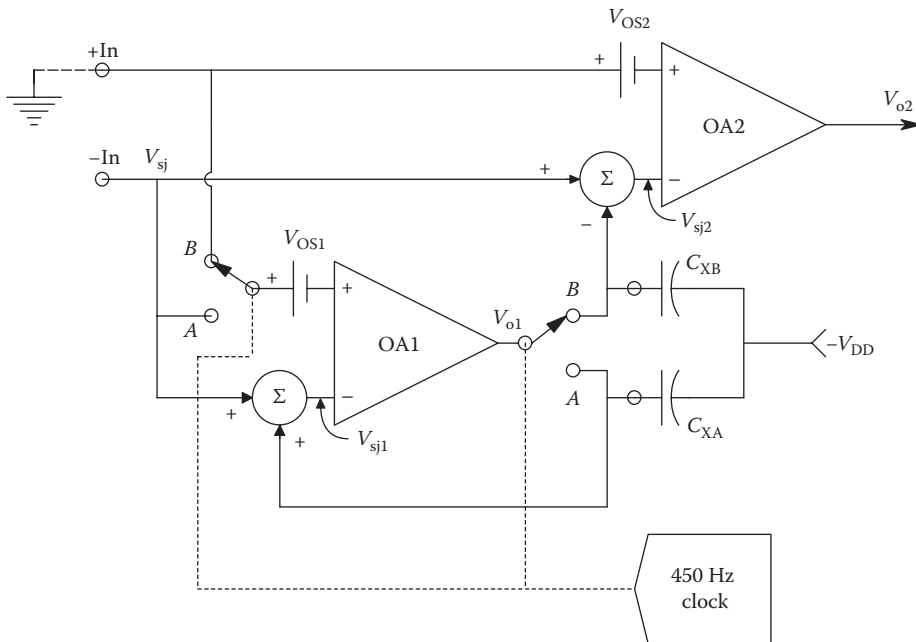


FIGURE 2.4
Simplified circuit of a Texas Instruments’ TLC2652 CSOA.

CSOAs are used in applications where stable, low-drift, DC amplification is required over a long time and under varying conditions of environmental temperature. Applications include weighing (electronic scales), electrochemical applications, and static photometric measurements. CSOAs use a special, internal feedback architecture to continuously minimize the DC input offset voltage, V_{OS} , to compensate for V_{OS} drift due to temperature changes or from power supply drift. Figure 2.4 illustrates the basic internal circuit architecture of a Texas Instruments’ TLC2652 series CSOA (TI 2005). Note that there are two OAs and two SPDT MOS switches. The MOS switches are commutated synchronously by an internal 450 Hz clock or, in some versions, can be operated up to 1 kHz by an external clock. Not shown are the filters and circuits required to eliminate switching glitches and to shape amplifier bandwidths. The capacitors C_{XA} and C_{XB} are low-loss dielectric, 0.1–1 μF .

When the switches are in position *A*, the input to OA1 is short-circuited, and it operates as a unity-gain follower. Thus, its output is its DC offset voltage, V_{OS1} , which is also the incremental voltage on capacitor C_{XA} . The incremental voltage on C_{XB} is also V_{OS1} , which is subtracted from V_{OS2} of OA2. When the switches are in position *B*, $V_{sj2} = (1 + K_1)V_{sj}$, and the two amplifiers are essentially in series, V_{OS1} is canceled out, and $V_{CXB} = K_1V_{sj}$. V_{OS2} is negligible because $K_1 \gg 1$.

Some key parameters for the TLC2652 series of CSOAs are as follows: $V_{OS} = 0.5 \mu\text{V}$, $V_{OS \text{ tempco}} = 3 \text{ nV}/^\circ\text{C}$, $I_B = 4 \text{ pA}$, DC $A_{vo} = 150 \text{ dB}$, GBWP $f_T = 1.9 \text{ MHz}$, the open-loop break frequency $f_b \cong 0.06 \text{ Hz}$, and $\eta = 3.1 \text{ V}/\mu\text{s}$.

A final type of OA that is widely encountered in preliminary circuit design practice, but unfortunately is not commercially available, is the *ideal op-amp* (IOA). The IOA assumption is used in pencil-and-paper design calculations. The IOA has the following properties: *infinite* differential gain, CMRR, input resistance, slew rate and GBWP, *zero* noise, bias currents, offset voltage, and output resistance. Probably the most important of these

assumptions is that $A_D = \infty$. In order to realize a finite output voltage under this condition, $(V_i - V_i') = 0$ or V_i must equal V_i' . In summary, an IOA is a differential VCVS with infinite gain and frequency response.

The designer of an analog signal-conditioning system for an instrumentation system faces a bewildering array of OA types and specifications. In the following sections, we try to give some rationale for choosing one OA over another. In making such choices, we are generally balancing meeting all minimum system performance specifications with dollar cost.

2.3.2 Basic Broadband Amplifier Design Using Op-Amps

In many instances, the output signal of a sensor is too small to effectively record, observe, or store. The signal is assumed to be a voltage or current. Thus, the designer has the task of amplifying the weak signal while introducing as little noise or distortion as possible. Often impedance levels are a concern. For example, the transducer output may be represented by a Thevenin equivalent circuit in which the Thevenin (series) impedance may not be negligible compared to a signal-conditioning amplifier's input impedance. Thus, the design problem to be resolved includes the selection of an OA type and circuit, which will minimize loading of the sensor's output.

Another problem in designing high-gain, broadband signal conditioners with OAs is the gain \times bandwidth trade-off. With simple OA amplifiers, the system's gain \times bandwidth product is approximately equal to the OA's GBWP. This implies that to realize a signal-conditioning gain of 1000 with a bandwidth of 100 kHz using low-noise OAs with $f_T = 10$ MHz, the strategy would be to cascade two amplifier circuits with their gains set to 31.62 (overall gain of 1000). The bandwidth (-3 dB point) of each stage is just $107/31.62 = 316.2$ kHz. It can be shown that when N identical stages with transfer function

$$A_v(s) = \frac{K_{vo}}{\tau s + 1} \quad (2.33)$$

are cascaded, then the -3 dB frequency of the cascade, f_c , is given by the following relation:

$$f_c = \left(\frac{1}{2\pi\tau} \right) (2^{1/N} - 1)^{1/2}. \quad (2.34)$$

Thus, in the preceding example, $f_c = 203.5$ kHz, and the frequency response design criterion is well met. If a single OA circuit were used with gain of 1000, then f_c would obviously be 10^4 Hz.

2.3.2.1 Noninverting Amplifier

Figure 2.5A illustrates the common, single-ended, noninverting, OA DC amplifier configuration. Because of the OA's high K_{vo} , $V_i \cong V_i'$ and the input impedance seen by the source circuit is R_{ic} , the OA's CM input resistance (cf. Figure 2.2).

In a modern, FET input OA, such as the Texas Instruments' TL071, R_{ic} can be on the order of $10^{12} \Omega$. Thus, in this case, unless the Thevenin source resistance approaches $10^8 \Omega$, loading effects and offset due to DC bias current (about 30 pA) will be negligible.

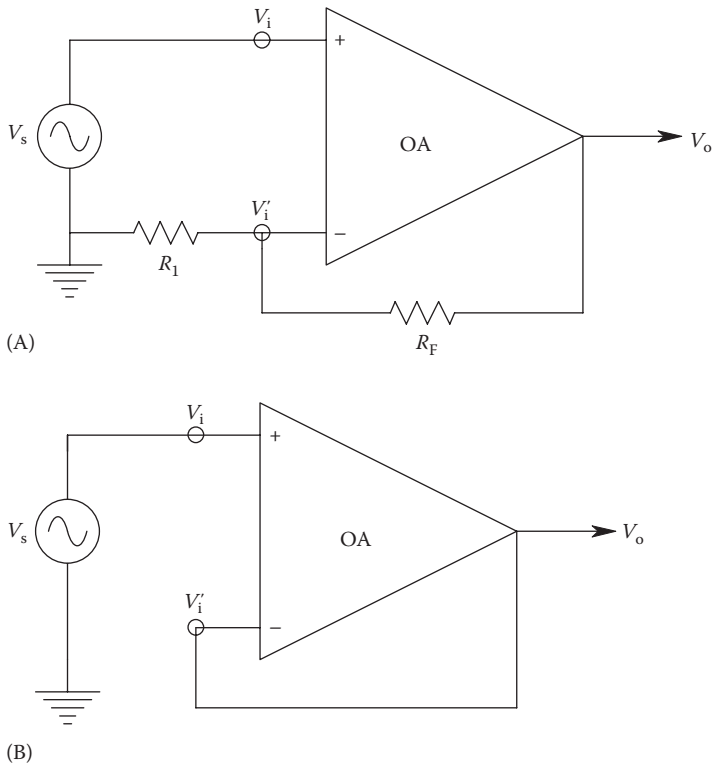


FIGURE 2.5
 (A) A noninverting OA circuit with gain set by R_1 and R_F . (B) Unity-gain buffer OA circuit.

To simplify calculations, we will treat R_{ic} at the V'_i node as infinite and consider the OA's output resistance to be negligible compared to R_F . The amplifier's closed-loop transfer function can be found by noting that

$$V_o = \frac{(V_s - V'_i) K_{vo}}{\tau s + 1} \tag{2.35}$$

and

$$V'_i = \frac{V_o R_1}{R_1 + R_F} = \beta V_o. \tag{2.36}$$

If Equation 2.36 is substituted into Equation 2.35, we can solve for V_o/V_s :

$$\frac{V_o}{V_s} = \frac{K_{vo}/(\tau s + 1)}{1 + \beta K_{vo}/(\tau s + 1)}, \tag{2.37}$$

which reduces to

$$\frac{V_o}{V_s} = \frac{1/\beta}{s(\tau/\beta K_{vo}) + 1}, \quad (2.38)$$

if we assume that βK_{vo} is $\gg 1$, which is generally valid.

The gain \times bandwidth product for this amplifier is found from Equation 2.37 to be

$$\text{GBWP} = \frac{K_{vo}}{1 + \beta K_{vo}} \frac{1 + \beta K_{vo}}{2\pi\tau} = \frac{K_{vo}}{2\pi\tau} \text{ Hz}, \quad (2.39)$$

which is the same as for the OA by itself.

If R_1 is made infinite and R_F is made a short circuit, the noninverting follower circuit of Figure 2.5B becomes a unity-gain voltage follower. Unity-gain followers are often used as output buffers to achieve impedance isolation.

2.3.2.2 Inverting Amplifier and Summer

Figure 2.6 illustrates the common inverting amplifier configuration. Here we have assumed two inputs: V_{s1} and V_{s2} . The amplifier output can be found by superposition. The output voltage can be written:

$$V_o = \frac{(0 - V_i') K_{vo}}{\tau s + 1}. \quad (2.40)$$

V_i' is found by solving the node equation

$$(V_i' - V_{s1})G_1 + (V_i' - V_{s2})G_2 + (V_i' - V_o)G_F = 0. \quad (2.41)$$

When V_i' is substituted into Equation 2.40, we may solve for V_o/V_{s1} , letting $V_{s2} = 0$:

$$V_o = \frac{-V_{s1} (G_1 K_{vo} / \sum G) / (1 + K_{vo} G_F / \sum G)}{s \left[\tau / (1 + K_{vo} G_F / \sum G) \right] + 1}, \quad (2.42)$$

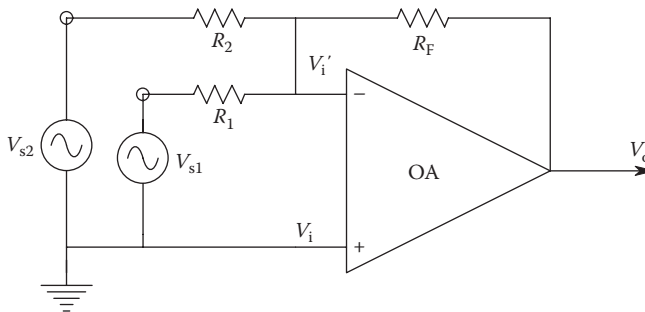


FIGURE 2.6
Inverting, summing OA configuration.

where $\sum G \equiv G_1 + G_2 + G_F$. When we assume that $K_{vo}G_F/\sum G \gg 1$, we can write

$$\frac{V_o}{V_{s1}} = \frac{-(G_1/G_F)}{s\left(\tau \sum G/K_{vo}G_F\right) + 1}. \tag{2.43}$$

In general, the gain for the k th inverting input is, by superposition,

$$\frac{V_o}{V_{sk}} = \frac{-(G_k/G_F)}{s\left(\tau \sum G/K_{vo}G_F\right) + 1}. \tag{2.44}$$

The GBWP for the k th inverting input is easily seen to be

$$\text{GBWP}_k = \left(\frac{K_{vo}}{2\pi\tau} \frac{G_k}{\sum G} \right) \text{Hz}. \tag{2.45}$$

Note that in the inverting amplifier case, the amplifier GBWP, given by Equation 2.45, is not equal to the OA's GBWP, except for very large G_k .

2.3.3 Current Feedback Op-Amps

Current feedback op-amps (CFOAs) constitute a relatively new class of OA that offers certain design advantages in terms of independent control of gain and bandwidth, and they generally have very high slew rates. The internal system architecture of a CFOA connected as a simple noninverting amplifier is shown in Figure 2.7. A nearly ideal,

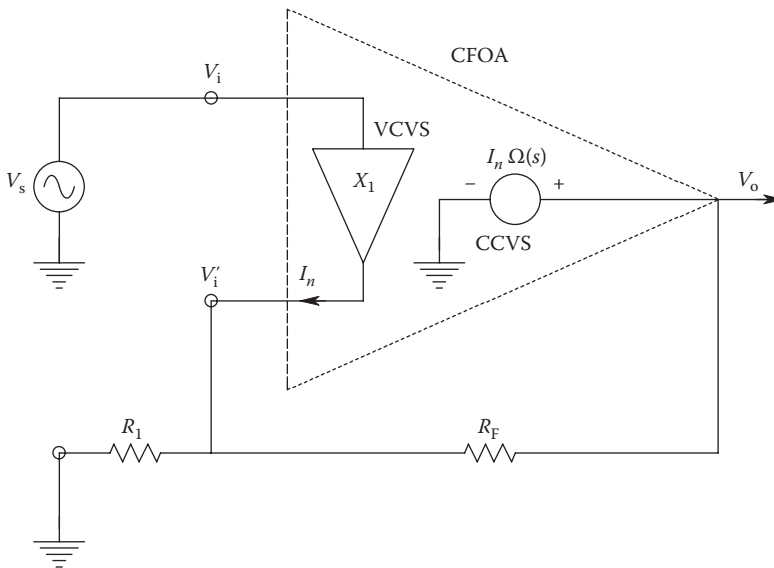


FIGURE 2.7
A noninverting amplifier circuit using a CFOA.

unity-gain, VCVS replaces the high-impedance input circuit of a conventional OA. Thus, the noninverting (V_i) terminal has a very high input impedance, but the V_i' node is the output of the VCVS and therefore presents a very low input (Thevenin) impedance, in the order of 50Ω . The output current of the VCVS, I_{nr} , is the input for a current-controlled voltage source (CCVS), Ω_o , which determines the CFOA's output voltage, V_o . The CCVS is designed so that

$$\Omega(s) = \frac{\Omega_o}{\tau s + 1} \Omega. \quad (2.46)$$

Hence,

$$V_o = I_n \left(\frac{\Omega_o}{\tau s + 1} \right) V. \quad (2.47)$$

The control current, I_{nr} , may be found from Kirchhoff's current law applied to the V_i' node (here, we neglect the CCVS's output resistance):

$$I_n = V_s G_1 + (V_s - V_o) G_F. \quad (2.48)$$

If Equation 2.48 for I_n is substituted into Equation 2.47, we may solve for the transfer function for the noninverting CFOA amplifier:

$$\frac{V_o}{V_s} = \frac{(G_1 + G_F) \Omega_o / (1 + G_F \Omega_o)}{s [\tau / (1 + G_F \Omega_o)] + 1}. \quad (2.49)$$

If we assume that $G_F \Omega_o \gg 1$, then the transfer function reduces to

$$\frac{V_o}{V_s} = \frac{1 + R_F/R_1}{s(\tau R_F/\Omega_o) + 1}. \quad (2.50)$$

The circuit's DC gain, A_{VO} , is just

$$A_{VO} = \frac{R_1 + R_F}{R_1}. \quad (2.51)$$

The circuit's GBWP is

$$\text{GBWP} = \frac{\Omega_o (R_F + R_1)}{2\pi\tau R_F R_1} = \frac{\Omega_o A_{VO}}{2\pi\tau R_F} \text{ Hz}. \quad (2.52)$$

The -3 dB frequency of the circuit is given by the following equation:

$$f_b = \frac{\Omega_o + R_F}{2\pi\tau R_F} \approx \frac{\Omega_o}{2\pi\tau R_F} \text{ Hz}. \quad (2.53)$$

Note that the break frequency, f_b , is dependent on R_F alone. Thus, the DC gain can be set with R_1 , holding the break frequency constant with R_F . Note that with a conventional OA noninverting amplifier, the break frequency varies inversely with the circuit's DC gain.

Although CFOAs offer advantages in terms of high output slew rate and f_T and the absence of the customary trade-off between gain and bandwidth found with conventional OAs, they are not as flexible in terms of stability and the ability to synthesize transfer functions.

First, let us examine the transfer function of a CFOA used in the CFOA inverting configuration, shown in Figure 2.8. Here, the positive voltage input (V_i) node is tied to ground, causing the output of the unity-gain buffer to be zero. Neglecting R_o , $V_i = 0$, I_n is given simply by

$$I_n = -(V_s G_1 + V_o G_F). \tag{2.54}$$

If this expression for I_n is substituted into Equation 2.47 for V_o , we can solve for the inverting amplifier's transfer function, assuming that $\Omega_o G_F \gg 1$:

$$\frac{V_o}{V_s} = \frac{-R_F/R_1}{s(\tau R_F/\Omega_o) + 1}. \tag{2.55}$$

Note that for the inverter, the DC gain is simply $-R_F/R_1$; the -3 dB break frequency, f_b , is $\Omega_o/(2\pi\tau R_F)$ Hz; and the GBWP is $\Omega_o/(2\pi\tau R_1)$ Hz. In using a conventional OA, it is a simple

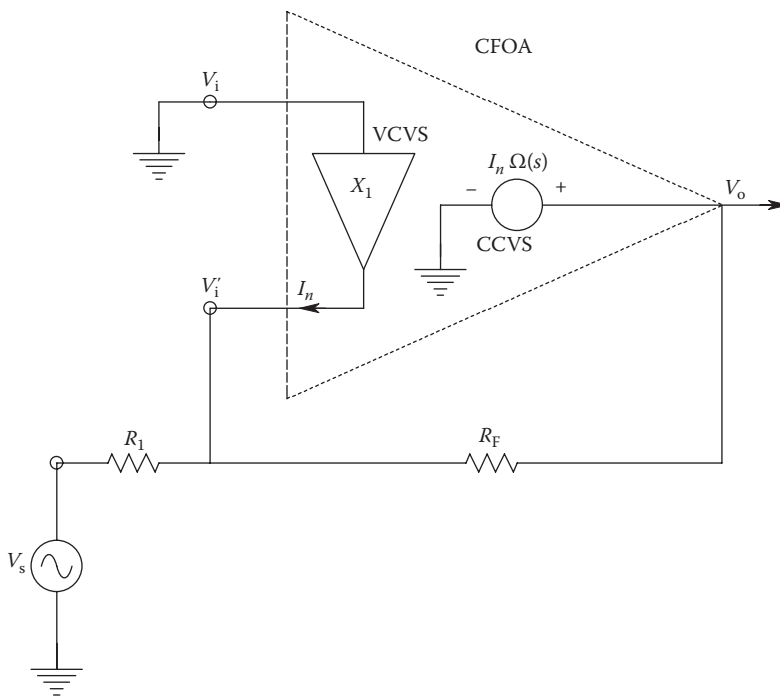


FIGURE 2.8
An inverting amplifier circuit using a CFOA.

extension of the gain relation to replace R_1 and/or R_F with capacitors or series or parallel RC circuits; stability is generally preserved. In the case of CFOAs, however, the designer must avoid applying direct capacitive feedback between the V_o and V_i' nodes; some manufacturers provide an internal R_F to avoid this problem.

To examine what happens when a capacitor is connected in place of R_F in the inverting CFOA amplifier in Figure 2.8, we note that the current I_n can be written:

$$I_n = -(V_s G_1 + V_o s C). \tag{2.56}$$

When Equation 2.56 for I_n is substituted into Equation 2.47 for V_o , we can derive the following transfer function:

$$\frac{V_o}{V_s} = \frac{-\Omega_o G_1}{s(\tau + C\Omega_o) + 1}. \tag{2.57}$$

If we set $R_1 = 1 \text{ M}\Omega$, $C = 1 \text{ }\mu\text{F}$, and $\Omega_o = 10^8 \text{ }\Omega$, the transfer function (2.57) becomes

$$\frac{V_o}{V_s} = \frac{-100}{s100 + 1}. \tag{2.58}$$

This is certainly not an effective integrator; it is best characterized as an inverting LPF with a break frequency of 10^{-2} r/s .

A noninverting integrator can be made from CFOAs, as shown in Figure 2.9. If we let the CFOA's $\tau_s \rightarrow 0$, the integrator's gain may be shown to be approximately

$$\frac{V_o}{V_s} \cong \frac{\Omega_{o1}/R_{o1}}{sRC(R_F\Omega_{o1}/R_{o1}R_1) + 1}. \tag{2.59}$$

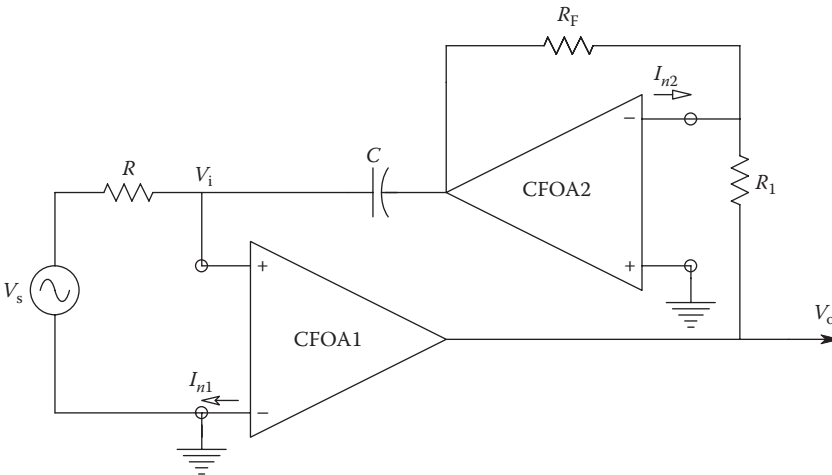


FIGURE 2.9
A noninverting integrator using two CFOAs.

Note that a Thevenin output resistance, R_{o1} , must be assumed for CFOA1's unity-gain buffer to permit analysis in this case. At frequencies above the break frequency, f_b , the integrator's gain becomes

$$\frac{V_o}{V_s} = \frac{R_1}{sRCR_F} \tag{2.60}$$

The break frequency for this integrator is quite low, about 5×10^{-7} r/s for $\Omega_o = 10^8 \Omega$, $C = 1 \mu\text{F}$, $R = 1 \text{ M}\Omega$, $R_F = R_1 = 10^3 \Omega$, and $R_{o1} = 50 \Omega$.

As a means to synthesize simple active filters (AFs) using CFOAs, consider the general circuit shown in Figure 2.10. Here, we have replaced the integrator's R and C in Figure 2.9 with admittances Y_1 and Y_2 , respectively. Using the same assumptions we made earlier in the analysis of the integrator, we find

$$\frac{V_o}{V_s} = \frac{Y_1 \Omega_1 G_{o1}}{Y_1 + Y_2 (1 + \Omega_1 \Omega_2 G_{o1} G_A / G_F \Omega_2)} \tag{2.61}$$

If we let $\Omega_1 = \Omega_2$ and $R_F > R_A$ and assume $\Omega_1 G_{o1} \gg 1$, the gain expression Equation 2.61 reduces to

$$\frac{V_o}{V_s} \cong \frac{Z_2 R_A}{Z_1 R_F} \tag{2.62}$$

(A more detailed analysis would include the real poles of the CFOAs in Equation 2.61.) Note that using this circuit architecture, it is possible to synthesize a simple, two-pole, one-zero band-pass filter (BPF) if we let $Z_1 = R_1 + 1/j\omega C_1$ and $Y_2 = G_2 + j\omega C_2$.

In summary, we see that in simple inverting and noninverting configurations, CFOAs offer HF response and slew rates. However, care must be exercised in using them in circuits having reactive elements. The basic inverting gain relationship of conventional IOAs, $V_o/V_s = -Z_F/Z_{IN}$, does not apply to CFOAs. Special circuit architecture must be used in many cases to realize even simple transfer functions, such as integration or BP filtering.

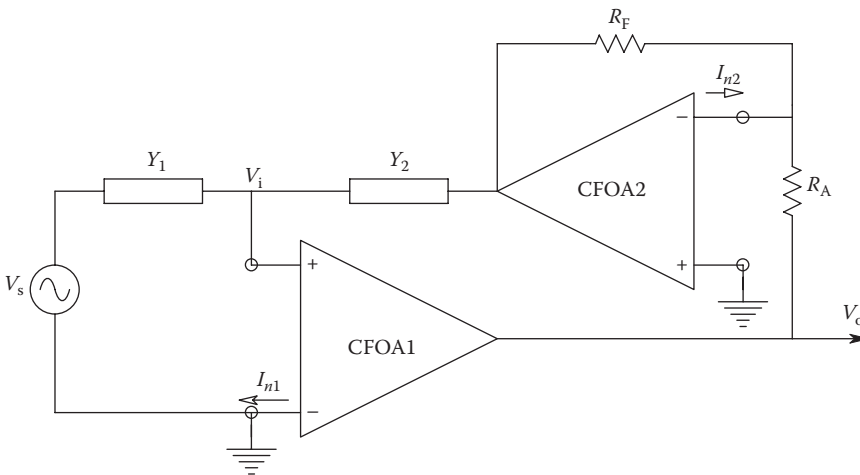


FIGURE 2.10
A general, noninverting feedback amplifier circuit using two CFOAs.

2.4 Analog Active Filter Applications Using Conventional Op-Amps

2.4.1 Introduction

In the previous sections of this chapter, we have stressed the use of OAs to make single-ended, broadband amplifiers. There are many instances in instrumentation systems where it is also necessary to filter a time-varying signal by passing it through a linear, frequency-dependent gain device. Such filtering is often used to improve SNR by attenuating those frequencies containing noise, but which do not contain significant signal information. Coherent interference may also be reduced or effectively eliminated by filtering. Finally, before ADC (periodic sampling), analog signals must be low-pass filtered to prevent errors in subsequent digital processing caused by *aliasing*. (Sampling and aliasing are discussed in detail in Chapter 9.) Low-pass filtering is also used to smooth or average the outputs of nonlinear circuits, such as rectifiers, PSRs, and true RMS converters.

Analog AFs fall into several broad categories dependent on the shape of the filter's passband. Figure 2.11 illustrates the major categories we will consider in the following sections. These include high-pass, broad BP, narrow BP (high Q), low-pass, notch (band-reject), and all-pass (used for phase compensation; the transfer function magnitude remains constant) active filters.

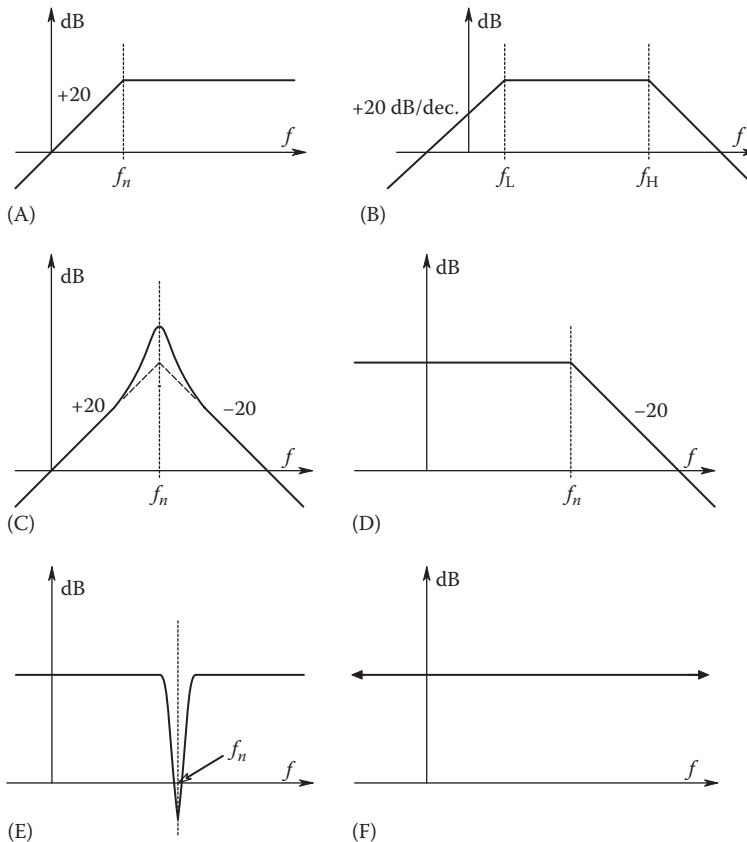


FIGURE 2.11

Bode asymptotes of common types of filters: (A) HPF. (B) Broad BP. (C) Narrow (tuned) BP. (D) LPF. (E) Band-reject or notch. (F) All-pass (flat amplitude passband, phase is frequency dependent).

Our treatment of analog AFs is by no means exhaustive; it is provided as a review of AF circuits that find common use for analog signal conditioning in instrumentation systems. We will not go into the design details of high-order AFs in this text (pole and zero placement). The reader should appreciate that the analysis and design of analog AFs is an extensive subject and books and chapters of books have been written on the subject. Many references on analog AF design are included in the chapter bibliography.

2.4.2 Analog Active Filter Architectures

The reader should be aware that there are many AF circuit architectures that can be used to realize a given transfer function. Some circuits offer the advantage of few components or minimum sensitivity to OA gain changes, while others offer design flexibility in terms of being able to independently set one or more of the filter’s parameters (midband gain, break frequency, damping factor or Q) by adjusting one, unique R or C for each parameter. In this section, we stress certain AF architectures that allow easy design. To expedite analysis, we assume all of the OAs are of conventional design and are ideal.

2.4.2.1 Controlled-Source Active Filters

This type of filter can be used to realize BP, low-pass, and high-pass quadratic transfer functions. A single OA is used as a low- or unity-gain VCVS with four or five resistive or capacitive feedback elements. Figure 2.12 illustrates the well-known *Sallen and Key LPF*. To derive this filter’s transfer function, we observe that the OA behaves as a VCVS with gain $K_v = 1 + R_F/R_A$. Thus,

$$V_2 = \frac{V_1(1/sC_2)}{R_2 + 1/sC_2} = \frac{V_1}{1 + sR_2C_2}. \tag{2.63}$$

We write a node equation on the V_1 node:

$$V_1 \left(G_1 + \frac{1}{R_2 + 1/sC_2} + sC_1 \right) - sC_1V_o = V_sG_1. \tag{2.64}$$

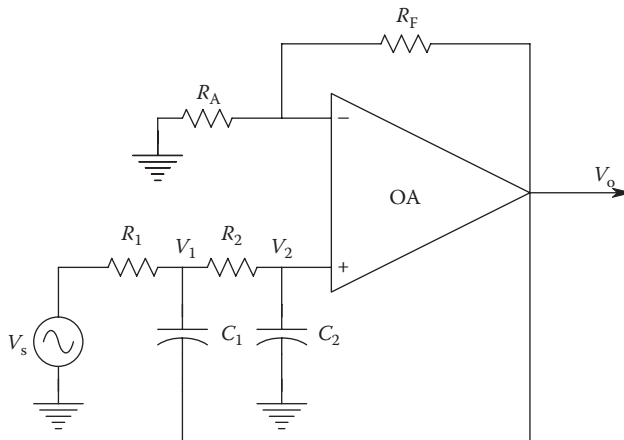


FIGURE 2.12
A Sallen and Key quadratic LPF.

Equation 2.63 can be solved for V_1 and substituted into Equation 2.64, allowing us to solve for V_o in terms of V_s . The Sallen and Key LPF's transfer function can thus be written as

$$\frac{V_o}{V_s} = \frac{K_v}{s^2 C_1 C_2 R_1 R_2 + s [C_2 (R_1 + R_2) + R_1 C_1 (1 - K_v)] + 1}. \quad (2.65)$$

If K_v is made unity, then the transfer function becomes

$$\frac{V_o}{V_s} = \frac{1}{s^2 C_1 C_2 R_1 R_2 + s C_2 (R_1 + R_2) + 1}. \quad (2.66)$$

The transfer function's denominator has the *standard* quadratic form $(s^2/\omega_n^2 + s^2\zeta/\omega_n + 1)$. The undamped natural frequency, ω_n , for the Sallen and Key LPF is

$$\omega_n = \frac{1}{C_1 C_2 R_1 R_2} \text{ r/s}. \quad (2.67)$$

The damping factor, ζ , can be written as

$$\zeta = \frac{C_2 (R_1 + R_2)}{2 C_1 C_2 R_1 R_2}. \quad (2.68)$$

If R_1 is made equal to R_2 , then the damping factor reduces to the simple expression

$$\zeta = \sqrt{\frac{C_2}{C_1}}. \quad (2.69)$$

Conveniently, the LPF's damping factor is set by the square root of the ratio of the capacitor values. Simultaneous adjustment of $R_1 = R_2$ with a ganged variable resistor can set ω_n independently from the damping factor.

Figure 2.13 illustrates a *Sallen and Key high-pass filter (HPF)*. Analysis of this circuit follows that for the aforementioned LPF. If K_v is set to unity and $C_1 = C_2 = C$, then it is easy to show that the transfer function is

$$\frac{V_o}{V_s} = \frac{s^2 C_2 R_1 R_2}{s^2 C_1 C_2 R_1 R_2 + s R_1 2C + 1}. \quad (2.70)$$

The undamped natural frequency, ω_n , of the HPF is also given by Equation 2.67. The damping factor is given by

$$\zeta = \frac{R_1}{2} \sqrt{\frac{1}{C^2 R_1 R_2}} = \sqrt{\frac{R_1}{R_2}}. \quad (2.71)$$

It is also possible to make a *controlled-source BPF*, as shown in Figure 2.14. Although a general analysis is interesting, it is more useful to realize this circuit with $K_v = 1$ and $C_1 = C_2 = C$. The BPF's transfer function is

$$\frac{V_o}{V_s} = \frac{s [CG_1 / (G_2 G_1 + G_2 G_3)]}{s^2 [C^2 / (G_2 G_1 + G_2 G_3)] + s [C (G_2 + 2G_1 + G_3) / (G_2 G_1 + G_2 G_3)] + 1}. \quad (2.72a)$$

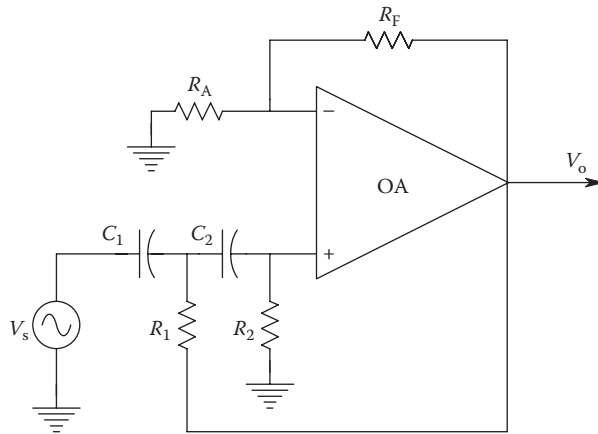


FIGURE 2.13
A Sallen and Key quadratic HPF.

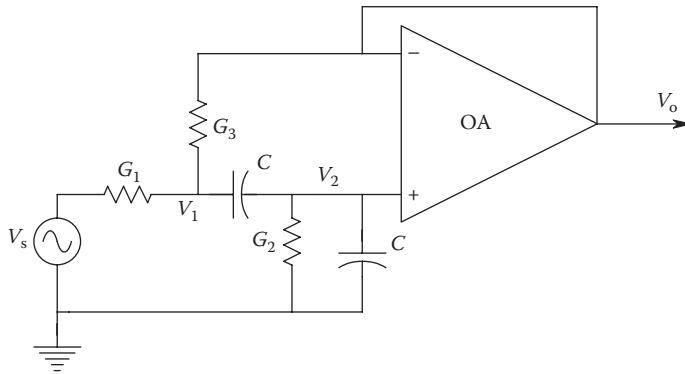


FIGURE 2.14
A controlled-source, quadratic BPF.

Here,

$$\omega_n = \frac{\sqrt{G_2(G_1 + G_3)}}{C} \text{ r/s.} \tag{2.72b}$$

The gain at resonance is

$$A_{Vpk} = \frac{G_1}{G_2 + 2G_1 + G_3}, \tag{2.73}$$

and Q of the tuned circuit is given by

$$Q = \frac{\sqrt{G_2(G_1 + G_3)}}{G_2 + 2G_1 + G_3} = \frac{1}{2\xi}. \tag{2.74}$$

Note that the peak response, A_v , and Q are independent of C ; $C_1 = C_2 = C$ can be used to set ω_n at constant Q and A_v .

2.4.2.2 Biquad Active Filters

Another AF architecture with which it is extremely easy to meet design criteria is the *biquad* AF. There are two versions of the biquad, but Figure 2.15 shows the more versatile, the two-loop biquad. This circuit can act as an HPF, sharply tuned BP, or LPF, depending on which output is selected. It also allows *independent adjustment* of the filter's peak gain, break frequency, and damping factor or *Q*. All-pass and band-reject (notch) filters can also be realized using the basic biquad filter with the addition of another OA.

First, we will examine the LPF transfer function of the biquad AF in Figure 2.15. Note that the subunits of this filter are two inverting summers and two integrators. The LPF's transfer function can be found from an application of Mason's rule:

$$\frac{V_5}{V_s} = \frac{(-R/R_1)(-R/R_2)(-1/sCR_3)(-1/CR_3)}{1 + [(-1)(-R/R_2)(-1/CR_3) + (-1)(-1/sCR_3)(-1/sCR_3)]}. \tag{2.75}$$

This transfer can be simplified to standard quadratic (time-constant) form:

$$\frac{V_o}{V_s} = \frac{R^2/(R_1R_2)}{s^2R_3^2C^2 + (sR_3CR/R_2) + 1}. \tag{2.76}$$

Inspection of Equation 2.76 shows that the low-frequency (LF) gain of this LPF is

$$A_{VLOW+} = \frac{R^2}{R_1R_2}. \tag{2.77}$$

The natural frequency is

$$\omega_n = \frac{1}{CR_3} \text{ r/s}, \tag{2.78}$$

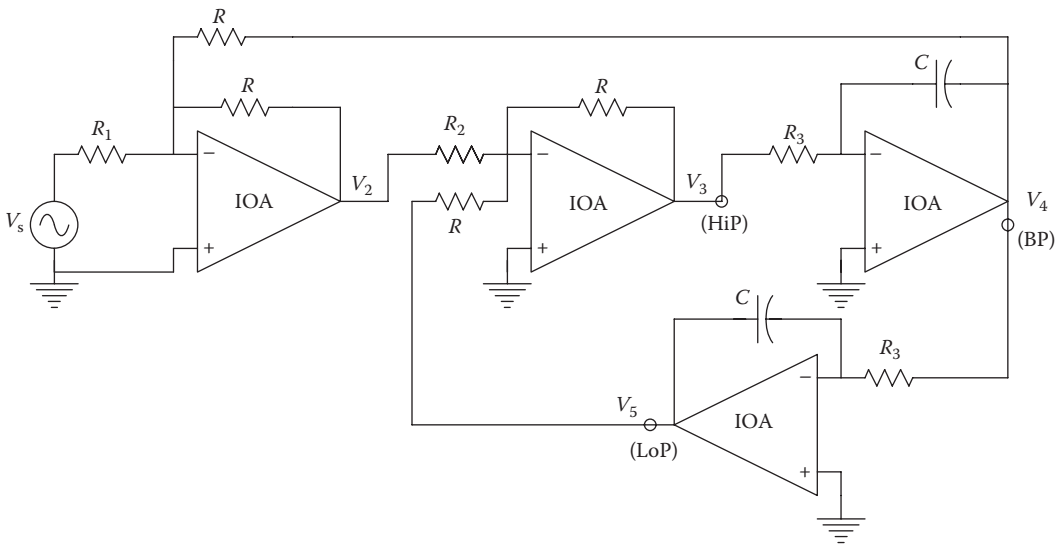


FIGURE 2.15
A two-loop, biquad AF.

and the damping factor is

$$\zeta = \frac{R}{2R_2}. \tag{2.79}$$

The same biquad filter can be used as a *tuned BPF*. In this case, the output is taken from the V_4 node. The transfer function again can be found by the application of Mason’s rule:

$$\frac{V_4}{V_s} = \frac{-sR_3CR^2/(R_1R_2)}{s^2R_3^2C^2 + (sR_3CR/R_2) + 1} \tag{2.80}$$

The BPF has a peak response at the natural frequency given by Equation 2.78. The peak gain is

$$A_{Vpk} = -\frac{R}{R_1}. \tag{2.81}$$

The BPF’s Q , a measure of the sharpness of its tuning, is given by

$$Q = \frac{1}{2\xi} = \frac{R_2}{R}. \tag{2.82}$$

Finally, the biquad AF of Figure 2.15 can produce a *quadratic high-pass characteristic*. Here, the output is taken from the V_3 node. The transfer function can be shown to be

$$\frac{V_3}{V_s} = \frac{s^2R_3^2C^2R^2/(R_1R_2)}{s^2R_3^2C^2 + (sR_3CR/R_2) + 1}. \tag{2.83}$$

The HPF’s natural frequency and damping factor are the same as for the LPF. Its HF gain is the same as the LF gain of the LPF.

It is also possible through the use of a fourth OA inverting summer to generate a *notch filter* that will sharply attenuate a selected frequency and pass all other frequencies. Such notch filters often find application in eliminating unwanted coherent interference, such as powerline frequency hum. The basic operation performed by a notch filter is illustrated in Figure 2.16. Simply stated, the output of a BPF is subtracted from its input. The BPF’s

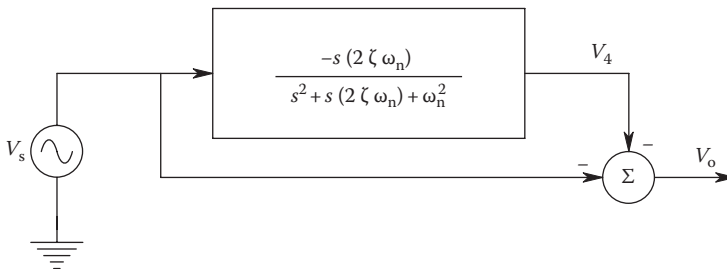


FIGURE 2.16 Block diagram for a biquad notch filter. V_4 is the inverting BP output of the two-loop biquad filter of Figure 2.15.

peak gain is made unity. As a result of this operation, the notch filter's transfer function has a pair of conjugate zeros on the $j\omega$ axis in the s -plane at $\pm j\omega_n$ and the usual complex-conjugate, quadratic pole pair at $s = \pm j\omega_n$. Its transfer function is given by the following equation:

$$\frac{V_6}{V_s} = \frac{(s^2/\omega_n^2) + 1}{(s^2/\omega_n^2) + s(1/Q\omega_n) + 1} \tag{2.84}$$

A biquad realization of a notch filter is shown in Figure 2.17. It is left as an exercise to derive its transfer function (note that R_1 should equal R to obtain the form of Equation 2.84).

As a final example of the versatility of the biquad AF architecture, we examine the all-pass filter design shown in Figure 2.18. All-pass filters have a flat frequency magnitude response; only the phase varies with frequency. All-pass filters are used to modify the phase response of electronic signal-conditioning systems without introducing frequency-dependent gain changes. Here, the BP output, V_4 , of the biquad of Figure 2.15 is summed with the filter input, V_s . The all-pass filter output, V_6 , is easily seen to be

$$V_6 = -R_6 \left(\frac{V_s}{R_5} + \frac{V_4}{R_4} \right) \tag{2.85}$$

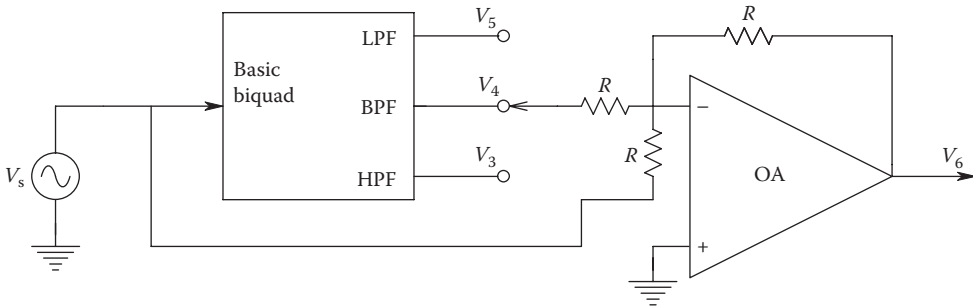


FIGURE 2.17
Realization of a biquad notch filter using the basic two-loop biquad of Figure 2.15.

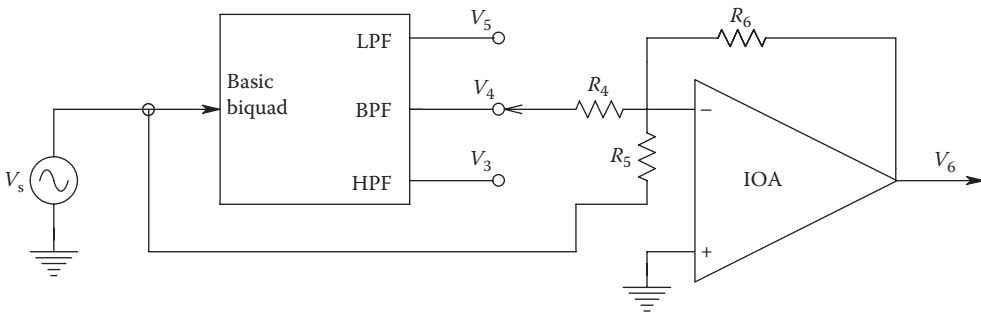


FIGURE 2.18
Realization of a biquad all-pass filter. See text for special relations for R_4 , R_5 , and R_6 .

When the relation for V_4/V_s , Equation 2.80, is substituted into Equation 2.85, we can show that

$$\frac{V_6}{V_s} = -\frac{(R_5/R_6)\left[s^2R_3^2C^2 + (sR_3CR/R_2) + 1\right] - \left[(R_6/R_4)sR_3CR^2/R_1R_2\right]}{s^2R_3^2C^2 + (sR_3CR/R_2) + 1}. \quad (2.86)$$

This relation may be reduced to the standard all-pass filter format given by Equation 2.87 if we let $R_1 = R$, $R_5 = 2R_4$, and $R_6 = R_5$:

$$\frac{V_6}{V_s} = -\frac{s^2R_3^2C^2 - (sR_3CR/R_2) + 1}{s^2R_3^2C^2 + (sR_3CR/R_2) + 1}. \quad (2.87)$$

In summary, note that the frequency response magnitude function for the all-pass filter is unity over all frequencies; only the phase varies. When $s = j\omega$ in Equation 2.87, the phase of the frequency response function is given by

$$\phi = -180^\circ - 2 \tan^{-1}\left(\frac{1 - \omega^2R_3^2C^2}{\omega CR_3R/R_2}\right). \quad (2.88)$$

2.4.2.3 Generalized Impedance Converter Active Filters

GIC circuits, using two OAs and five, two-terminal elements (either Rs or Cs), can be used to synthesize impedances to ground of the form $Z(s) = sX$ or $Z(s) = 1/s^2D$. These impedances in turn can be combined with other Rs, Cs, and OAs to generate various quadratic transfer functions, such as those discussed for the previously shown biquad AF architecture. Figure 2.19 illustrates the basic GIC architecture. Analysis of the GIC's driving point impedance is made easier if we assume that the OAs are ideal. Under this condition, we can write that the input impedance of the GIC to ground as

$$Z_{11} = \frac{V_1}{I_1} = \frac{V_1}{(V_1 - V_2)/Z_1}. \quad (2.89)$$

Also, from the IOA assumption and Ohm's law, we have

$$I_2 = \frac{V_2 - V_1}{Z_2} = I_3 = \frac{V_1 - V_3}{Z_3}, \quad (2.90)$$

$$I_4 = \frac{V_3 - V_1}{Z_4} = I_5 = \frac{V_1}{Z_5}. \quad (2.91)$$

From the three preceding equations, we can show

$$V_3 = V_1\left(1 + \frac{Z_4}{Z_5}\right), \quad (2.92)$$

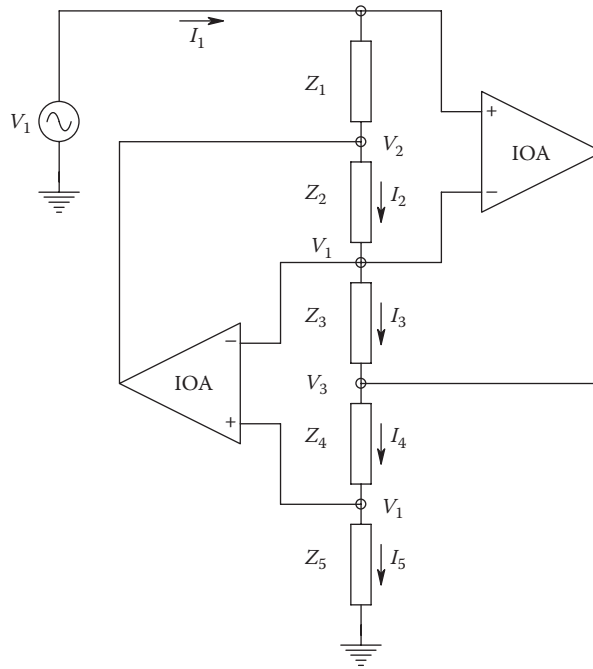


FIGURE 2.19
The basic GIC circuit architecture.

and

$$\mathbf{V}_2 = \mathbf{V}_1 \left(\frac{\mathbf{Z}_3 \mathbf{Z}_5 - \mathbf{Z}_2 \mathbf{Z}_4}{\mathbf{Z}_3 \mathbf{Z}_5} \right) \quad (2.93)$$

The GIC driving point impedance can be shown to be

$$\mathbf{Z}_{11}(s) = \frac{\mathbf{V}_1}{\mathbf{I}_1} = \frac{\mathbf{Z}_1 \mathbf{Z}_3 \mathbf{Z}_5}{\mathbf{Z}_2 \mathbf{Z}_4} \quad (2.94)$$

If we let $\mathbf{Z}_2 = 1/sC_2$ (a capacitor) and the other \mathbf{Z}_s be resistors, then \mathbf{Z}_{11} is in the form of a high- Q inductive reactance:

$$\mathbf{Z}_{11} = s \left(\frac{C_2 R_1 R_3 R_5}{R_4} \right) \quad (2.95)$$

where the equivalent inductance is given by

$$L_{\text{eq}} = \left(\frac{C_2 R_1 R_3 R_5}{R_4} \right) \text{H} \quad (2.96)$$

If both Z_1 and Z_5 are made capacitors, then the GIC's input impedance becomes a *frequency-dependent negative resistance* (FDNR) element:

$$Z_{11} = \frac{-R_3}{s^2 C_1 C_5 R_2 R_4} = \frac{-1}{s^2 D} \tag{2.97}$$

Thus, the GIC *D element* value is given by

$$D = \frac{C_1 C_5 R_2 R_4}{R_3} \tag{2.98}$$

A number of AF designs based on the inductive and $D Z_{11}$ forms have been devised. BP, all-pass, notch, low-pass, and high-pass designs are possible. Examples of these designs may be found in Chapter 4 of the text by Franco (1988). We illustrate a two-pole BPF in Figure 2.20 and a FDNR-based, two-pole LPF in Figure 2.21.

The transfer function for the BPF can be shown to be

$$\frac{V_o}{V_s} = \frac{sL_{eq}/R}{s^2 CL_{eq} + (sL_{eq}/R) + 1} \tag{2.99}$$

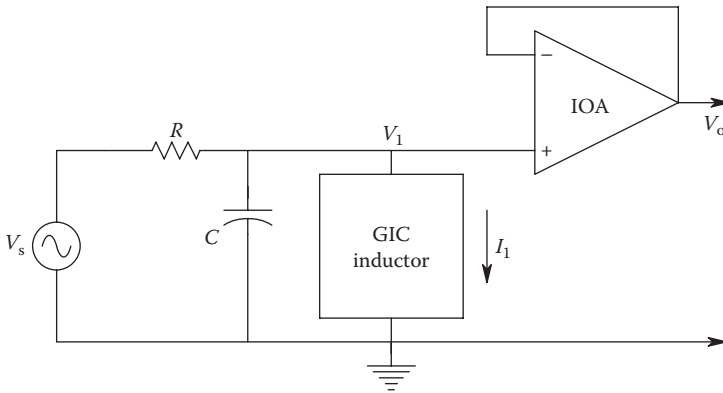


FIGURE 2.20
A tuned, quadratic BPF using a GIC as a (synthesized) low-loss inductor.

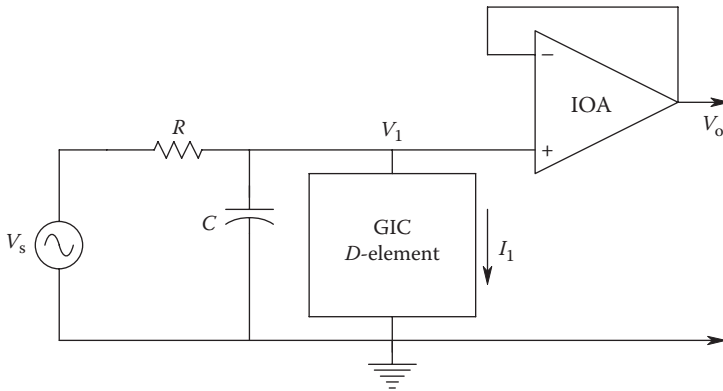


FIGURE 2.21
A quadratic LPF realized with a GIC FDNR element.

where L_{eq} is given by Equation 2.96. The LPF's transfer function is given by Equation 2.100:

$$\frac{V_o}{V_s} = \frac{1}{s^2RD + sRC + 1} \tag{2.100}$$

where D is given by the aforementioned Equation 2.98. Figure 2.22 shows a GIC-derived all-pass filter and Figure 2.23 a GIC-derived notch filter. Derivation of their transfer functions is left as exercises at the end of this chapter.

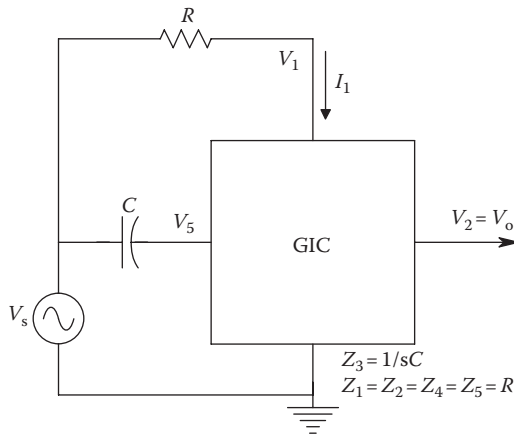


FIGURE 2.22
A GIC-derived all-pass filter.

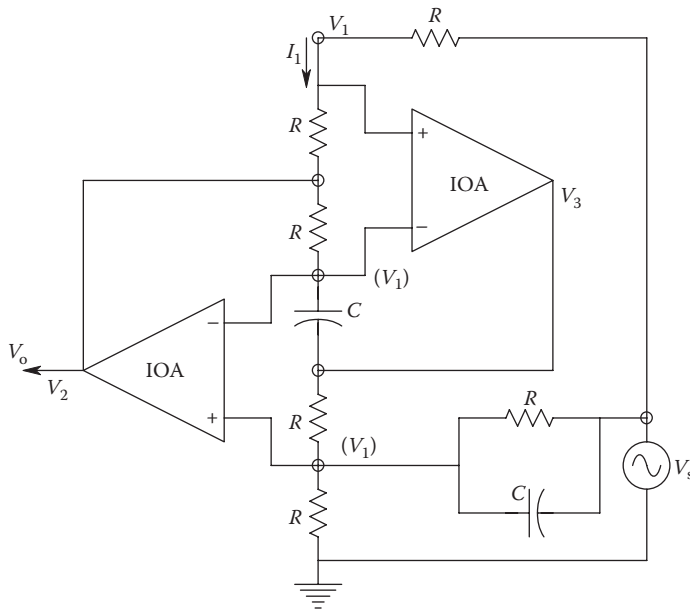


FIGURE 2.23
A GIC-derived notch filter.

2.4.2.4 High-Order Active Filters

Often the simple two-pole (quadratic) AF designs, which we have described earlier, do not meet the signal-conditioning requirements of an instrumentation system. In these cases, quadratic filters may be cascaded to create high-order filters having four or more poles. Such high-order filters are described according to how they perform in the time and frequency domains. In the following, we list the more commonly used types of high-order AFs:

Butterworth filters: The frequency response magnitude for this type of high-order AF has no ripple in the passband. Its phase vs. frequency response is nonlinear, so that the signal propagation time through the filter is not constant over frequency. There is considerable output overshoot in the time domain response of Butterworth filters of order greater than $n = 4$ to transient inputs. The slope of the stopband attenuation is $-20n$ dB/decade.

Chebyshev filters: In the frequency domain, the Chebyshev filter has no ripple in its stopband but has a specified passband ripple. Generally, a Chebyshev LPF having the same attenuation slope in the stopband as a comparable Butterworth LPF requires a lower order (fewer poles).

Elliptic or Cauer filters: In order to attain a higher attenuation (for a given order) in the transition band of frequencies (between passband and stopband), the elliptic filter allows ripple in the frequency response magnitude in both the pass- and stopbands. High-order elliptic filters have oscillatory transient responses.

Bessel or Thompson filters: In order to achieve a transient response that is free from overshoot and ringing, Bessel filters trade off steep attenuation slope in the transition band for linear phase response. For any order, n , Bessel filters make good LPFs where faithful time domain signal reproduction is required. They are used in medical systems such as ECGs.

2.4.3 Operational Amplifier Integrators and Differentiators

Often in conditioning analog measurement records, we wish to either integrate or differentiate these signals. Simple *analog integration* of an input signal can present two practical problems: One is that if the signal contains a DC or average level, the integrator will eventually saturate. That is, the output voltage of the OA is bounded at levels less than its power supply voltage. The second problem with a practical analog integrator is that even with its input grounded, it will integrate its own DC bias current and offset voltage. Using Figure 2.24 and assuming an otherwise ideal OA, it is easy to show that the integrator's output due to its bias current and offset voltage is given by

$$V_o(t) = \frac{-tI_B}{C_F} + V_{OS} + \frac{tV_{OS}}{RC_F}. \quad (2.101)$$

Here, we assume zero initial conditions until $t = 0$. Clearly, $v_o(t)$ will linearly approach the + or - saturation voltage of the OA. The use of EOA or autozero OAs can minimize analog integrator drift, however. Note that digital integration can generally eliminate most of the bias current and drift problems.

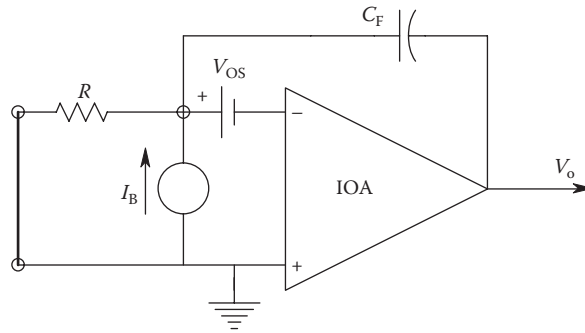


FIGURE 2.24

An OA integrator circuit showing sources of DC drift error. The magnitudes and signs of the DC bias current, I_B , and the offset voltage, V_{OS} , depend on the particular type of OA used and its temperature.

Analog differentiation also presents practical problems. In *OA differentiators*, bias current and drift present little problem. However, a real OA's finite bandwidth can give HF transient problems. Refer to Figure 2.25. At the summing junction, we can write the node equation:

$$V_i' (G + sC) - V_o G = V_i sC. \tag{2.102}$$

The OA's output is given by

$$\frac{V_o}{V_i'} = \frac{-K_V}{\tau_A s + 1}. \tag{2.103}$$

If we combine Equations 2.101 and 2.102, we can show

$$\frac{V_o}{V_i} = \frac{-[K_V / (K_V + 1)]sRC}{[s^2 \tau_A RC / (K_V + 1)] + [s(\tau_A + RC) / (K_V + 1)] + 1}. \tag{2.104}$$

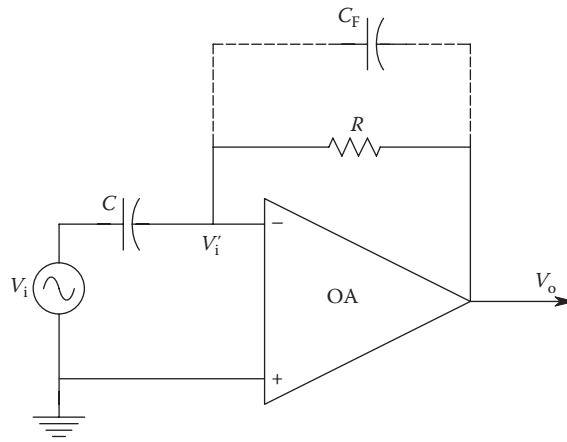


FIGURE 2.25

A basic OA differentiator. C_F can be used to damp the HF response.

At low input frequencies,

$$\frac{V_o}{V_i} \cong -sRC. \quad (2.105)$$

At high input frequencies, the quadratic denominator of the transfer function, Equation 2.104, is significant. The break frequency of the transfer function is

$$\omega_n = \sqrt{\frac{K_V + 1}{\tau_A RC}} \text{ r/s}. \quad (2.106)$$

The damping factor of the analog differentiator can be shown to be

$$\zeta = \frac{\tau_A + RC}{2(K_V + 1)} \sqrt{\frac{K_V + 1}{\tau_A RC}}. \quad (2.107)$$

To see where the problem arises, let us substitute typical numerical values for the circuit: $R = 1 \text{ M}\Omega$, $C = 1 \text{ }\mu\text{F}$, $\tau_A = 0.001 \text{ s}$, $K_V = 10^6$. Now we calculate $\omega_n = 3.16 \times 10^4 \text{ r/s}$, and the damping factor is $\zeta = 1.597 \times 10^{-2}$, which is quite low. Thus, any transients in V_i will excite a poorly damped, HF, sinusoidal transient at the output of the differentiator. This behavior can present a problem if the differentiator input is not bandwidth limited. If the differentiator's input is not bandwidth limited, we must limit the bandwidth of the circuit by placing a small capacitor in parallel with R . It is left as an exercise to compute the transfer function of the practical differentiator with a parallel RC feedback circuit.

2.4.4 Summary

In the preceding subsections, we have reviewed the architectures of three commonly used, quadratic, AF *building blocks* in the design of analog signal-conditioning systems. There are many other AF architectures that filter designers also use. However, lack of space prohibits their description here. The interested reader should consult the chapter bibliography.

2.5 Instrumentation Amplifiers

As defined by current practice, IAs are generally described as direct-coupled, low-noise, DAs having very high input impedances, high CMRR, user-settable gain from 0 to 60 dB, and bandwidth from DC to the hundreds of kHz, depending on design and operating gain. Manufacturers offer IAs as ICs. IAs are often used at the output of a transducer, such as a bridge circuit, as the first stage in an analog signal-conditioning system. As headstages, IAs should have low-noise characteristics in noise-critical applications. We illustrate in the following two IA circuits that can be made from low-noise OAs in order to obtain custom low-noise IA designs.

Table 2.1 as follows summarizes the systems characteristics of some commercially available IAs. Note that IA bandwidths are not generally very broad, generally covering the audio spectrum, with some exceptions. Their slew rates are also modest. IAs are generally characterized by high CMRRs and high input impedances, however.

TABLE 2.1
Selected Specifications of Selected Instrumentation Amplifiers

IA Mfr. and Model	CMRR at Gain dB	R_{in}/C_{in} PF	Small-Signal kHz BW at Gain		Slew Rate V/ μ s	ϵ_{na} at 1 kHz at Gain: nV RMS/ $\sqrt{\text{Hz}}$	i_{na} pA/ $\sqrt{\text{Hz}}$	V_{os}/Tempco μ V/ $^{\circ}\text{C}$	I_{Br} nA	I_B Tempco
			Gain	Gain						
PMI AMP-01	100 at 1	20 G Ω , CM	570 at 1	4.5 at G = 10	540 at 1	0.15 at G = 10 ³	0.3 μ V/ $^{\circ}\text{C}$	3	40 pA/ $^{\circ}\text{C}$	
	120 at 10	1 G Ω , DM	100 at 10		59 at 10					
	130 at 10 ² -10 ³		82 at 100		10 at 100					
Maxim MAX4462	120	2 G Ω , CM	26 at 1000	0.5 at G = 1	38	—	—	1 pA	—	
		2 G Ω , DM	2500 at 1	0.5 at G = 10						
			25 at 100	0.25 at G = 10 ²						
AD8221	90 at 1	100 G Ω /2, CM and DM	825 at 1	2 at G = 1	8	40 fA/ $\sqrt{\text{Hz}}$	45 μ V/0.3 μ V/ $^{\circ}\text{C}$	0.2	1 pA/ $^{\circ}\text{C}$	
	110 at 10		562 at 10	2.5 at G = 10-100						
	130 at 100		100 at 100							
AD621B	140 at 1000		14.7 at 10 ³							
	110 at 10	10 G Ω /2	800 at 10	1.2	13 at 10	100 fA/ $\sqrt{\text{Hz}}$	50 μ V/0.6 μ V/ $^{\circ}\text{C}$	0.5	1.5 pA/ $^{\circ}\text{C}$	
	130 at 100		200 at 100	3.5	9 at 100	4 fA/ $\sqrt{\text{Hz}}$		4 pA	—	
MN2310	ca. 95 dB	600 M Ω /12, DM	1000		1.65					
		150 M Ω /48, CM								
		10 ¹⁰ Ω /2, DM	1300 at 1	4	8 at 1000	0.3	\pm 10 \pm 100/G μ V	\pm 2	\pm 30 pA/ $^{\circ}\text{C}$	
BB2128	106 at 10	10 ¹¹ Ω /9, CM	700 at 10							
	125 at 100		200 at 100							
	130 at 1000		20 at 1000							
BB INA118	90 at 1	10 ¹⁰ Ω /1, DM	800 at 1	0.9 at G = 10	10 at 1000	0.3	\pm 10 \pm 50/G μ V/ \pm 0.2 \pm 2/G μ V/ $^{\circ}\text{C}$	\pm 1	\pm 40 pA/ $^{\circ}\text{C}$	
	110 at 10	10 ¹⁰ Ω /4, CM	500 at 10							
	120 at 100		70 at 100							
BB PGA206/207	125 at 1000		7 at 1000							
	92 at 1	10 ¹³ Ω /1, DM	5000 at 1	25 at G = 1-10	18 at 8	1.5 fA/ $\sqrt{\text{Hz}}$ at G = 8	\pm 0.5 mV/ \pm 2 μ V/ $^{\circ}\text{C}$	2 pA	0.4 pA/ $^{\circ}\text{C}$	
	96 at 2	10 ¹² Ω /4, CM	4000 at 2							
	100 at 4 or 5		1300 at 4,5							
	100 at 8 or 10		600 at 8,10							

Note: BB, Burr-Brown; AD, Analog Devices; MN, Micro-Networks; PMI, Precision Monolithics.

2.5.1 Instrumentation Amplifiers That Can Be Made from Op-Amps

Figure 2.26 illustrates the well-known, three-OA IA circuit. The practical range of the differential gain of this circuit ranges from 1 to 1000 by adjusting the value of R_1 . Note that OA3 is set up as a DA using matched resistors: R_3 and R'_3 and R_4 and R'_4 . Assuming OA3 is ideal, it is easy to show that its output is given by

$$V_o = (V_4 - V_3) \left(\frac{R_4}{R_3} \right). \tag{2.108}$$

This design also assumes that the resistors R_4 and R'_4 and R_3 and R'_3 are perfectly matched. The DA circuit of OA3, although having an apparent high CMRR and differential gain, is unsuitable as an IA because of its unequal and relatively low input resistances. V_3 sees an input impedance of R_3 to the summing junction of OA3, which is at virtual ground, while V_4 sees an input impedance of $R_3 + R_4$. OAs 1 and 2 form a symmetrical circuit that can be analyzed for the cases of pure CM input signals and pure DM input signals. (Note that any combination of V_s and V'_s can be broken down into CM and DM input components to OA1 and OA2 and the responses to these components can be summed at the output by superposition.)

For purely CM inputs, $V_s = V'_s$, and symmetry considerations lead to the conclusions that the summing junction potentials of OA1 and OA2 are identical; hence, there is no current flowing in R_1 . (Here, we have assumed that OA1 and OA2 are IOAs and that $R_2 = R'_2$) Because no current flows in R_1 under CM input excitation, we may remove it from the circuit (i.e., replace it with an open circuit). Thus, OA1 and OA2 are seen to be unity-gain followers under CM excitation. Thus, their outputs V_3 and V_4 are seen to be equal and equal to the CM input, V_{sc} . The following DA circuit of OA3 subtracts V_3 from V_4 , producing, ideally, a $V_o = 0$ for CM inputs. Because the OAs are not ideal and have finite CMRRs, gains, bandwidths, and input and output impedances and because resistor matching may not be perfect, there will be a small V_o for a pure CM input.

In the case of a purely DM input, $V'_s = -V_s$, and the summing junctions of OA1 and OA2 are at $-V_s$ and V_s , respectively. Hence, a current of $2V_s/R_1$ amps flows in R_1 . Furthermore,

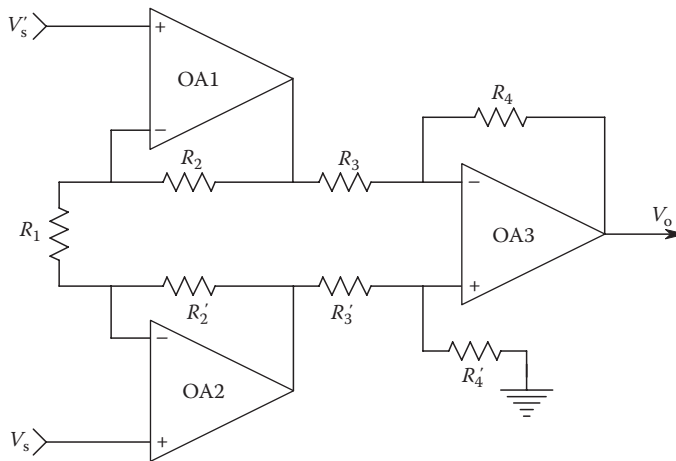


FIGURE 2.26
Schematic of a three-OA IA.

consideration of Kirchhoff's voltage law leads to the conclusion that the center of R_1 (i.e., $R_1/2 \Omega$ from either end) is at zero volt potential with respect to ground and thus may be tied to ground without disturbing the circuit. Now the DM OA1 and OA2 circuits are seen to consist of noninverting amplifiers with feedback resistor R_2 and resistance $R_1/2$ to ground from the summing junction. Hence, both the DM OA1 and OA2 circuits can easily be shown to have the gains

$$\frac{V_3}{V'_s} = \frac{V_4}{V_s} = 1 + \frac{2R_2}{R_1}. \quad (2.109)$$

Thus, the output for DM inputs is just

$$V_o = \left(1 + \frac{2R_2}{R_1}\right) \left(\frac{R_4}{R_3}\right) (V_s - V'_s). \quad (2.110)$$

The three-OA IA configuration is used in some commercial IC IA designs. It provides the designer with some flexibility if special circuit behavior is desired, such as low noise or wide bandwidth. Resistor matching is critical for obtaining a high CMRR. Often a resistor, such as R_4 , can be made variable to tweak the CMRR to a maximum.

Figure 2.27 illustrates a *two-OA IA design*. Again, two matched OAs must be used, and resistors, R , must be closely matched to obtain a high CMRR. This circuit may be analyzed using superposition and the IOA assumption. First, we assume that $V'_s > 0$ and $V_s \rightarrow 0$. Thus, $V_3 = V'_s$ and $V_5 = 0$. Now we write node equations on the V_3 and the V_5 nodes:

$$V'_s(2G + G_F) - V_4G - V_5G_F = 0, \quad (2.111)$$

$$V_5(2G + G_F) - V_3G_F - V_oG - V_4G = 0. \quad (2.112)$$

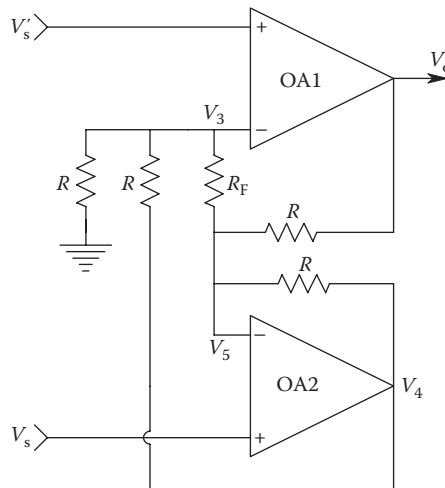


FIGURE 2.27
A two-OA IA circuit.

Simultaneous solutions of Equations 2.111 and 2.112 give us

$$V_o = -V'_s 2 \left(1 + \frac{R}{R_F} \right). \quad (2.113)$$

In a similar manner, we set $V_s > 0$ and $V'_s \rightarrow 0$ and note that $V_5 = V_s$ and $V_3 = 0$. Node equations are written on the V_3 and V_5 nodes:

$$V_3(2G + G_F) - V_4G - V_5G_F = 0, \quad (2.114)$$

$$V_5(2G + G_F) - V_3G_F - V_oG - V_4G = 0. \quad (2.115)$$

Solution of these equations gives us

$$V_o = V_s 2 \left(1 + \frac{R}{R_F} \right). \quad (2.116)$$

Hence, by superposition, the two-OA IA's output can be written as

$$V_o = (V_s - V'_s) 2 \left(1 + \frac{R}{R_F} \right). \quad (2.117)$$

Note that for both IA designs previously given, the inputs are made directly to the non-inverting, high-input-impedance inputs of OAs. Hence, the input impedance of these IAs will essentially be that of the OA. If FET input OAs are used, then input resistances in excess of $10^{12} \Omega$ can be attained with very low bias currents.

2.5.2 Isolation Amplifiers

IsoAs are a special class of IA that find extensive application in biomedical instrumentation, specifically, the recording of low-level bioelectric signals from human and animal subjects. Their use is generally dictated by electrical safety considerations. Their use is also called for when the input signal is found in the presence of a high common-mode voltage (CMV) and also in the situation where ground-loop currents can introduce errors in the signals under measurement. IsoAs have an extremely low conductance between the reference ground of the output signal and the differential headstage's ground and input terminals. The DC conductive isolation between the headstage ground and the output stage ground is generally accomplished by a magnetic transformer coupling or by optocoupling techniques.

In the Analog Devices family of IsoAs, the output of the input DA is coupled to the output stage of the IsoA by an isolation transformer. The operating power for the input DA is also transformer-coupled from an HF power converter oscillator, which derives its power from the nonisolated or normal supply for the IsoA. The block diagram of an Analog Devices AD295 precision hybrid IA is shown in Figure 2.28. Note that in this design, both the input DA and the output circuit have isolated (internal) power supplies.

To illustrate the differences between ordinary IAs and IsoAs, we review some of the key specifications for the now discontinued Analog Devices Model AD294A medical IsoA.

The overall gain is the product of G_{IN} and G_{OUT} .

$$G_{IN} = (1 + 10^5/R_G), \quad R_G \geq 1 \text{ k}\Omega, \quad G_{INMAX} = 100, \quad G_{OUT} = (1 + R_A/R_B), \quad 1 < G_{OUT} < 10.$$

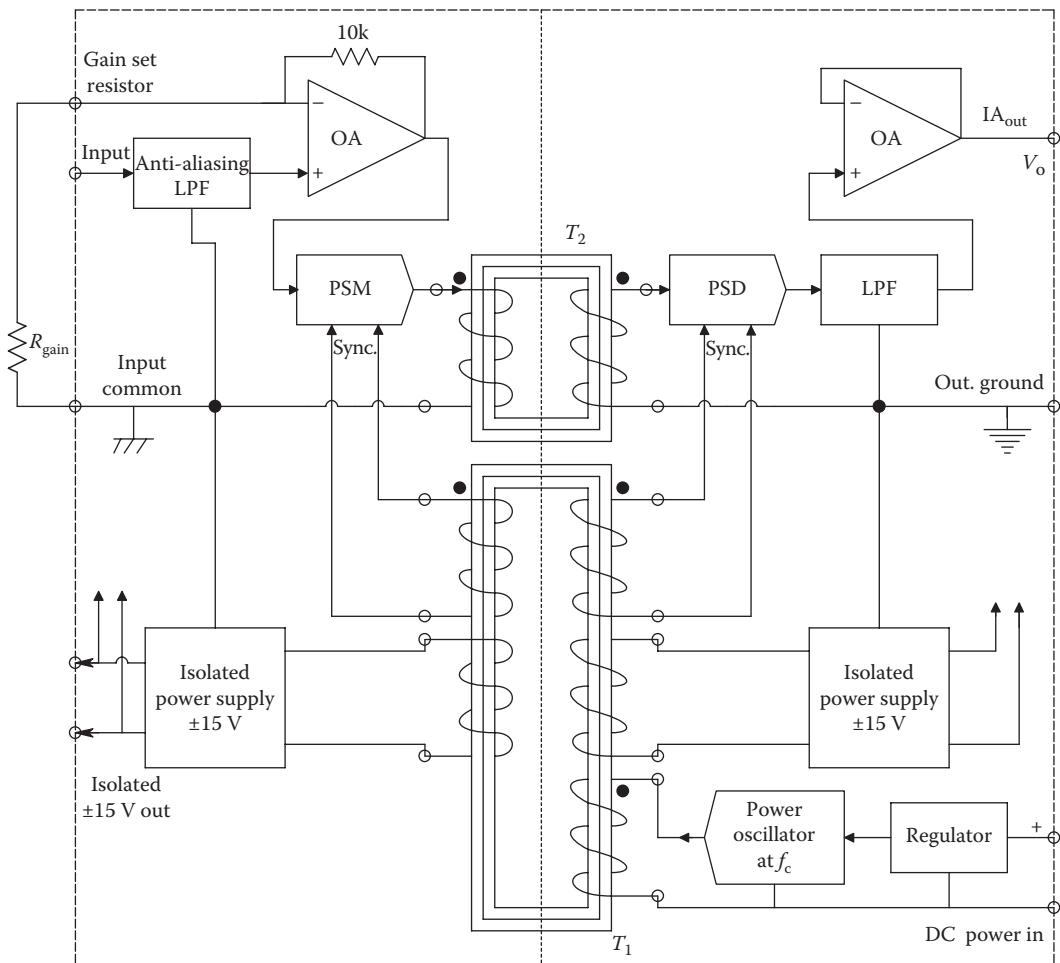


FIGURE 2.28

A simplified block diagram of an Analog Devices AD295 precision IsoA. (Figure used with permission from Analog Devices.)

Hence, the overall gain may be 1000. The CMRR of the AD294A is 100 dB, not exceptionally high for an IA. The maximum continuous CMV is ± 3500 Vpk. The amplifier is rated to stand a ± 8000 Vpk, 10 ms pulse at the rate of 1 pulse every 10 s. The input impedance is $10^8 \Omega \parallel 150$ pF. The input (DC) bias current, I_B , is 2 nA. The SS bandwidth is 2.5 kHz (gain 1–100 V/V). The slew rate is 9.1 V/ μ s. The maximum leakage current between headstage and output stage is 2 μ A RMS, given 240 V RMS 60 Hz applied to the input. The AD294A was typically used to record electrocardiographic or electromyographic signals using body surface electrodes.

The Intronic Model IA294 medical IsoA is also suitable for ECG recording and EEG signal conditioning. This IsoA also uses the two-transformer architecture to couple power into the differential headstage and signal out. Its specifications are summarized in the following.

The IA296's gain is fixed at 10. The CMRR is 160 dB with a 5 k Ω source resistance imbalance. The maximum safe DC CM input voltage is ± 5000 V. The CM input impedance

is $10^{11} \Omega \parallel 10 \text{ pF}$; the DM input impedance is $3 \times 10^8 \Omega \parallel 2.2 \text{ pF}$. The input bias current is $\pm 200 \text{ pA}$; the I_b tempco is $\pm 5 \text{ pA}/^\circ\text{C}$. The -3 dB bandwidth is 1 kHz . Input noise with a single-ended input with $R_s = 5 \text{ k}\Omega$ is $0.3 \mu\text{V}$ in a 10 Hz to 1 kHz bandwidth (about $9.5 \text{ nV}/\sqrt{\text{Hz}}$); the current noise is about $0.13 \text{ pA}/\sqrt{\text{Hz}}$ in a 0.05 Hz to 1 kHz bandwidth. Maximum leakage current due to component failure is $10 \mu\text{A}$. (Intronics Power Inc.: www.intronicspower.com/products/pdf/ia294.pdf [accessed February 2, 2014].) Because of its low noise, the Intronics IA294 is well suited for measurements of low-level biological signals such as fetal ECGs, EEGs, and evoked brain potentials.

Finally, as an example of a photooptically coupled IsoA, we examine the Burr-Brown Model 3652 IA. (BB-3652A data sheet: www.ti.com/lit/ds/sbos129/sbos129.pdf [accessed February 3, 2014].) (Note Texas Instruments bought BB in September 2000.) Because this IA does not use a built-in transformer to couple power to the input DA, an isolated DC/DC converter must be used, such as the Burr-Brown Model 722. The BB722 supply uses a 900 kHz oscillator to drive a transformer that is coupled to two independent rectifiers and filters. It can supply $\pm 5 \text{ V DC}$ to $\pm 16 \text{ V DC}$ at about 16 mA to each of the four outputs. It has isolation test ratings of $\pm 8000 \text{ Vpk}$ for 5 s and $\pm 3500 \text{ Vpk}$ continuous between inputs and outputs. The leakage current maximum for this power supply is $1 \mu\text{A}$ for 240 V RMS at 60 Hz , insuring medical safety. The Model 3652 IsoA has the following specifications when used with the BB722 DC/DC power supply.

Gain is settable from 1 to 1000. The CMRR is 80 dB . The BB3652 can block a CMV of $\pm 2000 \text{ V DC}$ and a pulse of $\pm 5000 \text{ V}$ for 10 ms . The input resistance is given as $10^{11} \Omega$, CM and DM. The isolation impedance between input and output is $10^{12} \Omega \parallel 1.8 \text{ pF}$. The input bias current is 10 pA . The offset voltage tempco is $25 \mu\text{V}/^\circ\text{C}$. The -3 dB bandwidth is 15 kHz at all gains. The short-circuit voltage noise is $5 \mu\text{V RMS}$ measured over a 10 Hz – 10 kHz bandwidth. No figure is given for the input current noise. Maximum leakage current is $0.35 \mu\text{A}$.

2.5.3 Autozero Amplifiers

Yet another type of IA merits attention. The commutating autozero (CAZ) amplifier employs a rather unique design to obtain an almost negligible DC input offset voltage drift. The Intersil ICL7605/7606 CAZ amplifiers make use of two, novel, switched capacitor networks to: (1) Convert a true differential input to a single-ended signal that is the actual input to the CAZ amplifier; and (2) The second pair of switched capacitors is used in conjunction with two matched OAs to cancel out each OA's offset voltage. This process is illustrated in Figure 2.29A and B. During the *A* clock cycle, C_2 is forced to charge up to $-V_{OS2}$, while C_1 charged to $-V_{OS1}$ is switched in series with the input, V_1 , and V_{OS1} to cancel V_{OS1} . In the *B* position, C_1 is charged up to $-V_{OS1}$, and C_2 now charged to $-V_{OS2}$ is put in series with V_{OS2} and V_1 to cancel V_{OS2} , etc.

The maximum CMRR is 104 dB , and the amplifier bandwidth is 10 Hz . The offset voltage tempco is an amazing $10 \text{ nV}/^\circ\text{C}$. The commutation clock frequency of the Intersil CAZ amplifier can range from 160 to 2560 Hz . Overall gain can be set between 1 and 1000. The ICL7605/7606 CAZ amplifier is intended for use with DC measurement applications such as strain gauge bridges used in electronic scales.

National Semiconductor (NS) offered the LMC669 autozero module that can be used with any OA configured as a single-ended inverting summer or as a noninverting amplifier. (Note that NS was bought by Texas Instruments 9/23/11.) Figure 2.30A and B illustrates

two applications of the LMC669 autozero circuit. Note that the autozero module does not limit the dynamic performance of the OA circuit to which it is attached; the amplifier retains the same DC gain and SS bandwidth and slew rate. The effective offset voltage drift of any OA using the autozero is $100 \text{ nV}/^\circ\text{C}$. The maximum offset voltage using the autozero is $\pm 5 \mu\text{V}$. The autozero's bias currents are 5 pA (these must be added to the OA's I_B); its clock frequency can be set from 100 Hz to 100 kHz . The NS LMC669 autozero module was very useful to compensate for DC drift in circuits using special-purpose OAs, such as EOAs, which normally have large offset voltages ($200 \mu\text{V}$) and large V_{OS} tempcos ($5 \mu\text{V}/^\circ\text{C}$). The autozero does contribute to an increase in the circuit's equivalent input voltage noise. However, the choice of sampling rate and step size and the use of low-pass filtering in the feedback path can minimize this effect.

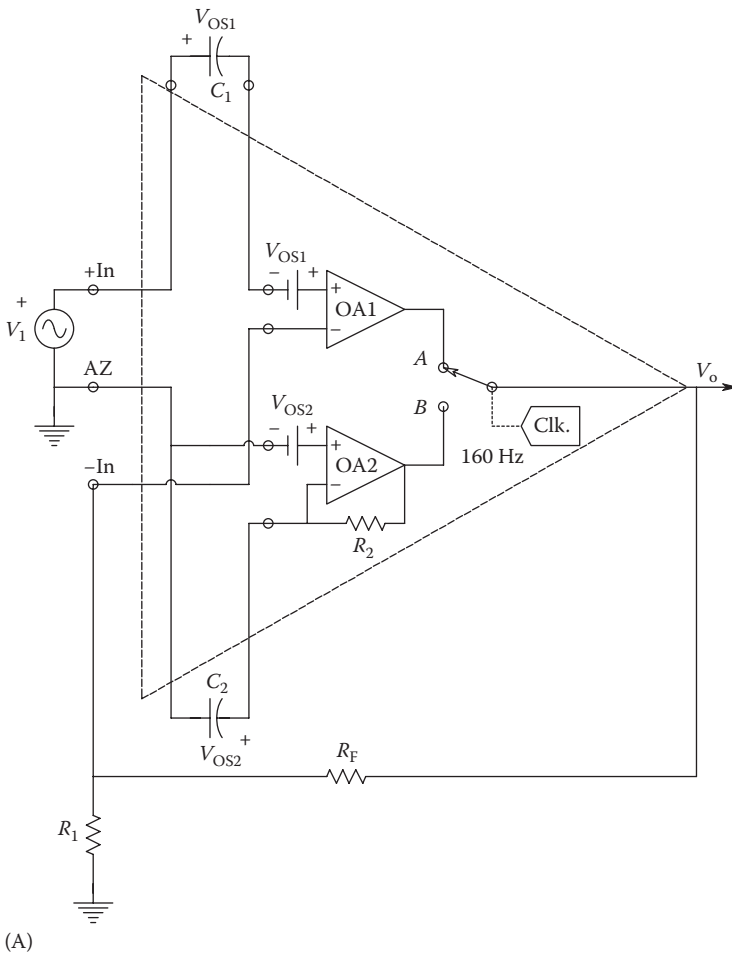


FIGURE 2.29

(A) Schematic of the innards of an Intersil ICL7605/7606 CAZ amplifier. In this (A) switching cycle, OA1 is connected to the output. OA2 is connected such that its DC offset voltage charges up C_2 . C_1 is switched in series with the input node of OA1 so that the voltage on C_1 cancels V_{OS1} .

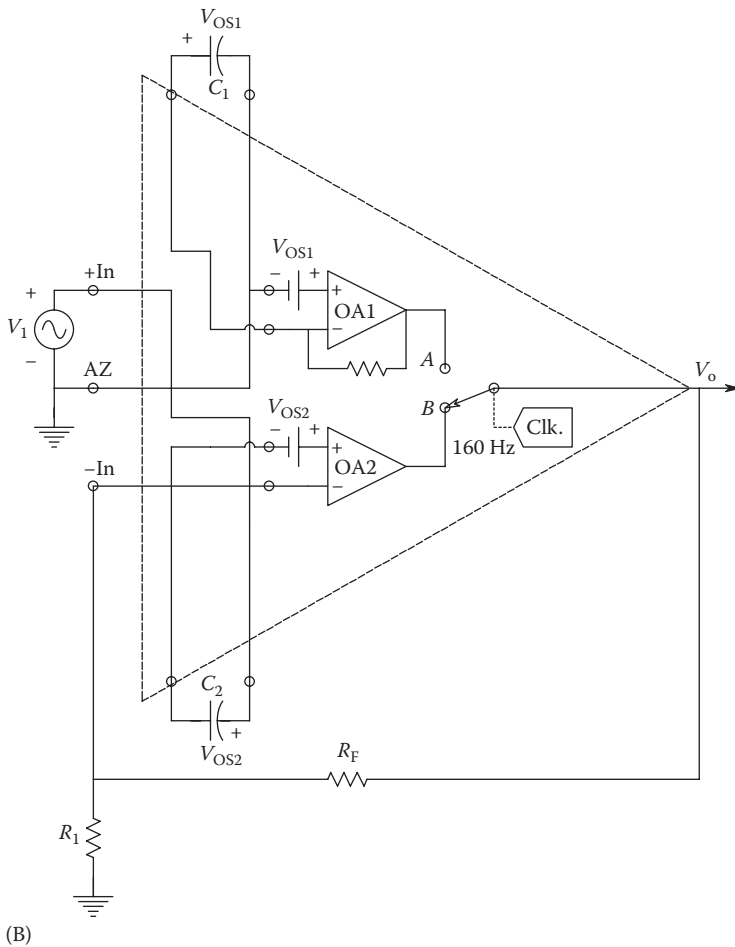


FIGURE 2.29 (continued)

(B) In the (B) switching cycle, OA2 is connected in series with C_2 , so $V_{OS2(k)}$ of OA2 subtracts from $V_{OS2(k-1)}$ on C_2 .

2.5.4 Absolute Isolation

One scheme that can produce *absolute isolation* between the transducer, headstage, and system under test environment and the signal conditioning, digital interface, and computer is the fiber-optic cable (FOC). The transducer and headstage are powered by batteries or an autonomous energy-harvesting system. The headstage analog output modulates the intensity of a LED projecting light down the FOC. Several modulation schemes might be used, including pulse frequency modulation (FM), sigma-delta modulation, or double-sideband, suppressed-carrier modulation (DSBSCM). A photodiode responds to the modulated light signal on the near end of the FOC. Its output is directed to a suitable demodulator, and the signal is recovered. The FOC provides essentially zero conductance between the system under test and the signal conditioner.

2.5.5 Summary

In this section, we have examined some examples of IAs. IAs are generally defined as DAs with high input impedances, high CMRR, and programmable gain, usually set by

choice of one or two resistors. Their bandwidths generally depend on the gain setting and roughly follow the constant gain \times bandwidth rule. Typical unity-gain bandwidths are in the hundreds of kHz, although some are now available in the 3 MHz range. Special isolation IAs and autozeroing amplifiers are also available.

2.6 Nonlinear Analog Signal Processing by Op-Amps and by Special Function Modules

2.6.1 Introduction

Perhaps the best way to define what we mean by nonlinear signal processing is to first describe the properties of a *linear system (LS)*. LS outputs can be found, in general, by the operation of *real convolution*. Real convolution can be derived from the general property of LSs, *superposition*: the operation of real convolution can be expressed by the well-known definite integral, sic:

$$y_1(t) = x_1 \otimes h = \int_{-\infty}^{\infty} x_1(\sigma)h(t-\sigma) d\sigma, \quad (2.118)$$

where

- $y_1(t)$ is the system output, given an input
- $x_1(t)$ to a LS with an impulse response, $h(t)$
- \otimes denotes the operation of real convolution

The property of *superposition* is illustrated by Equation 2.119. We note that

$$y_3 = (a_1y_1 + a_2y_2), \quad (2.119)$$

given input $x_3 = (a_1x_1 + a_2x_2)$.

LS outputs can be *scaled*, that is, if $x_2 = ax_1$, then

$$y_2 = a(x_1 \otimes h). \quad (2.120)$$

Finally, LSs obey the property of *shift invariance*. If the input x is delayed t_o seconds, then the output is also shifted:

$$y(t-t_o) = x(t-t_o) \otimes h. \quad (2.121)$$

Nonlinear systems do not obey the properties of LSs discussed previously. Nonlinear systems can be dynamic, static, or a combination of both. A static nonlinear system is one in which the output, y , is some nonlinear function of the input, x . For example, $y = \sin(x)$, $y = \operatorname{sgn}(x)$, $y = a + bx + cx^2$, and $y = \operatorname{abs}(x)$ describe nonlinear, static relations between input x and output y . $y = mx + b$ is a *linear* static function of x .

A *dynamic nonlinearity* is one described by a set of first-order, nonlinear ODEs. An example of such a system is the well-known Volterra equations describing predator–prey relations in a simple ecosystem. These are

$$\dot{x}_1 = k_1x_1 - k_3x_1x_2, \quad (2.122)$$

$$\dot{x}_2 = k_2x_2 - k_4x_1x_2. \quad (2.123)$$

The Volterra system responds to initial conditions on x_1 and x_2 . That is, a solution to the Volterra equations is run as an initial value problem (IVP).

Another example of a dynamic nonlinear system is illustrated in Figure 2.31. A linear, parallel LC circuit is connected to a negative resistance element whose current is determined by the following relation:

$$i_{nl} = -av + bv^3. \quad (2.124)$$

Hence, by Kirchhoff's current law, we can write

$$0 = \ddot{v} - \dot{v} \left(\frac{1-3bv^2}{a} \right) \left(\frac{a}{C} \right) + \frac{v}{LC}. \quad (2.125)$$

This relation can be written as two first-order ODEs:

$$\dot{x}_1 = \frac{-x_2}{LC} + x_1 \left(\frac{1-3bx_2^2}{a} \right) \left(\frac{a}{C} \right), \quad (2.126)$$

$$\dot{x}_2 = x_1, \quad (2.127)$$

where, obviously, $x_2 = v$. Equation 2.125 is the well-known *van der Pol equation*.

Often nonlinear systems can be modeled by separating the linear dynamics from a no-memory nonlinearity, such as $y = \tanh(x)$, which can either precede or follow the linear block or lie between two linear blocks, depending on circumstances.

There are several important nonlinear analog signal processing operations that are used in instrumentation. These include, but are not limited to, *precision full-wave rectification* or

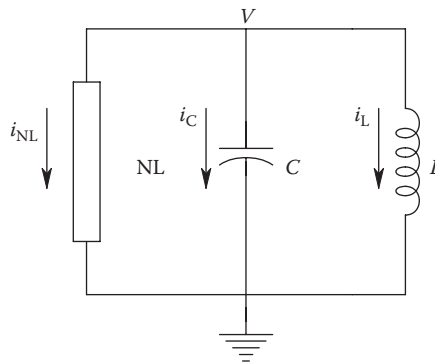


FIGURE 2.31

A simple nonlinear RLC circuit that can be used to illustrate the van der Pol equation for nonlinear oscillations.

absval circuits, true RMS-to-DC conversion circuits, peak detection circuits, sample-and-hold or track-and-hold (T&H) circuits, square-root circuits, special function modules that give an analog output proportional to $x(y/z)^m$, log ratio modules, and trigonometric function modules.

We will review the applications of each of the nonlinear signal processing circuits before discussing its architecture and design details. Certain nonlinear operations may easily be realized with OA circuits; for others, it is more expedient to use off-the-shelf, special-purpose microcircuits. We cite examples of these microcircuits, but for practical reasons, it is not possible to list all of a given type available.

2.6.2 Precision Absolute Value Circuits

As the name implies, this type of nonlinear circuit takes the *absval* of the input waveform. Thus, the output approximates

$$y(t) = k \text{ abs}[x(t)] = k|x(t)|, \tag{2.128}$$

where k is a gain or scaling constant. Precision *absval* circuits can be used for *transfer standards*. That is, they perform nearly ideal, full-wave rectification of symmetrical sinusoidal signals, which, after low-pass filtering, appear as a DC voltage proportional to the peak or RMS value of the input sine wave. This DC voltage can be generally measured with greater accuracy than the corresponding AC voltage, especially when the peak value is less than 1 V. Precision *absval* circuits have been used in adaptive analog AFs, such as the well-known Dolby B™ audio noise suppression system. Here, the *absval*/LPF system output is proportional to the signal power in a certain range of high frequencies and is used to adjust the filter's HF response. Precision *absval* circuits can also be used to precede square-law circuits used to square time-variable signals. When this is done, only half of a parabolic nonlinearity needs to be used because

$$y = [\text{abs}(x)]^2 = x^2. \tag{2.129}$$

Many precision *absval* circuits using OAs exist (see, e.g., Sections 5.2 through 5.4 in Graeme 1974); we will examine several of the more useful designs here. Figure 2.32 illustrates the

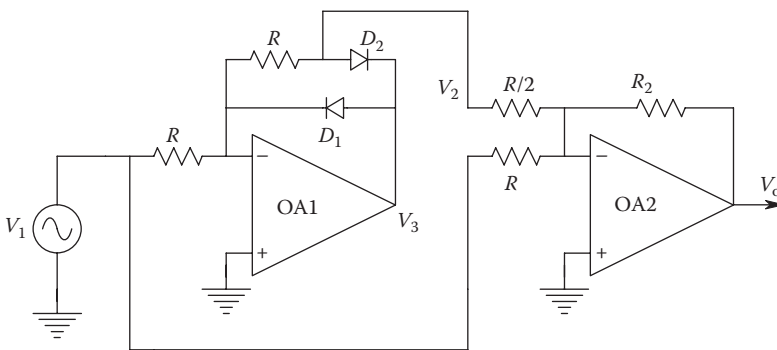


FIGURE 2.32
A precision, operational, full-wave rectifier or *absval* circuit.

circuit of a commonly used absval operator that uses two OAs. Its input resistance is just $R/2$, and its output is ideally given by

$$V_o = +\left(\frac{R_2}{R}\right)V_1| \tag{2.130}$$

However, at low levels of V_1 , some tweaking of the $R/2$ resistor may be required to obtain equal and opposite slopes of the absval curve.

Another absval circuit having high input resistance is shown in Figure 2.33. In this circuit, correct resistor values are important. The parameter n is the rectifier gain and must be >1 . A small capacitor, C , is used to improve the HF response of the circuit. R_1 and R_3 can be made variable in order to adjust the absval operation to be symmetrical (i.e., have slopes of $+n$ or $-n$).

Finally, Figure 2.34 illustrates an absval circuit having a differential input. An IA with high CMRR and differential gain K_D is used to condition the floating absval voltage

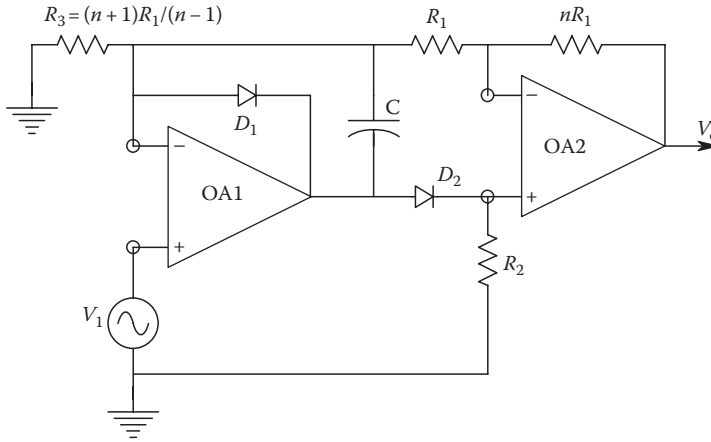


FIGURE 2.33
A precision absval circuit having high input impedance.

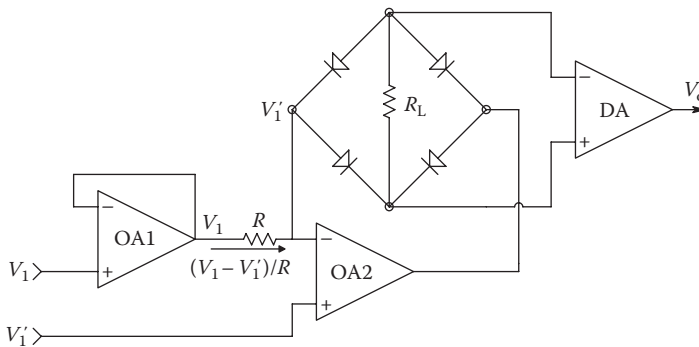


FIGURE 2.34
A precision absval circuit having a high-impedance, differential input.

across R_L . OA1 is used as a follower; OA2 forces a unidirectional current through R_L . The output of this circuit can easily be shown to be

$$V_o = K_D \left(\frac{R_L}{R} \right) |V_1 - V_1'|. \tag{2.131}$$

In describing the aforementioned absval circuits, we have neglected second-order errors caused by nonzero offset voltages and bias currents. Also neglected are the effects of finite OA dynamic response and diode capacitance. If an absval circuit is to be used at high audio frequencies or above, then the designer must choose the OAs with care and use HF switching diodes in the design. The designer should also simulate the HF behavior of the absval circuit in the time domain using an electronic circuit analysis program (ECAP), such as one of the versions of SPICE or MicroCap™ in order to verify that it will meet specifications.

2.6.3 Multifunction Converters

Multifunction converters are LSI analog ICs that perform the operation

$$V_o = K V_Y \left(\frac{V_Z}{V_X} \right)^m. \tag{2.132}$$

The input voltages can range from 0 to 10.0 V and from DC to 400 or 500 kHz. The exponent, m , can range from 0.2 to 5.0. Commercial examples of multifunction converters are the Burr-Brown Model 4302 and the Analog Devices AD538 Analog Computational Unit (cf. Analog Devices 2013).

The multifunction converter is a useful IC that can be used as a one-quadrant multiplier, squarer, square rooter, or exponentiator, depending on the value of m and the accessory OA circuits used. The multifunction converter can also be used to produce an output voltage proportional to the sine, cosine, or arctangent of an input voltage using an external OA. In addition, using implicit analog computational techniques involving feedback, the true RMS of an input signal may be computed (discussed in the following section), as well as real-time vector magnitude, as shown in Figure 2.35, where

$$V_o = \sqrt{V_1^2 + V_2^2}. \tag{2.133}$$

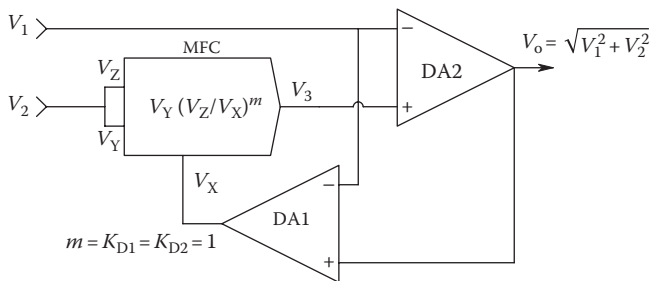


FIGURE 2.35

A multifunction converter used in a real-time Pythagorean system where $V_o = \sqrt{V_1^2 + V_2^2}$, when $m = K_{D1} = K_{D2} = 1$.

2.6.4 True RMS-to-DC Converters

True RMS converters are widely used in the quantitative amplitude measurement of noise and nonsinusoidal signals. Electrical noise may arise from resistors, devices (e.g., transistors), capacitive or magnetic pickup from environmental noise sources such as sparking motor commutators, or slip-ring contacts. Acoustical noise is also important to measure, for example, noise on a factory floor and mechanical vibrations of bearings. Often it is useful to measure the RMS value of noise-like bioelectric waveforms, such as muscle action potentials (EMGs) in order to diagnose problems with a patient’s neuromuscular system. RMS EMGs can also be used to actuate prosthetic limbs and robotic manipulators. Still another use for true RMS converters is in the design of analog adaptive filters, where the RMS voltage measured in some frequency band is used to adjust some feature of a filter’s overall frequency response.

The operations on a signal done by an RMS converter are, in order of operation, squaring, smoothing or averaging by low-pass filtering, and then square rooting. The direct method of computing the analog true RMS of a time-varying signal is shown in Figure 2.36. Another way to find the RMS value of a time-varying signal is to use the implicit method illustrated in Figure 2.37. Here, a multifunction module is used with feedback of the DC output voltage to force square rooting to occur: $m = 1$ in this application. Yet another implicit method of computing the true RMS value of a signal is the feedback thermocouple

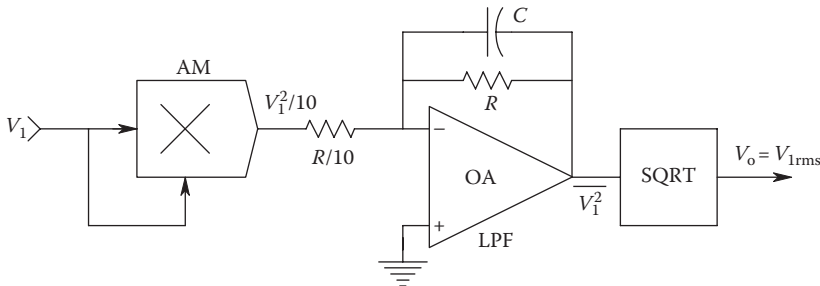


FIGURE 2.36 Circuit for the generation of the RMS value of an AC signal. The output V_o is a DC voltage.

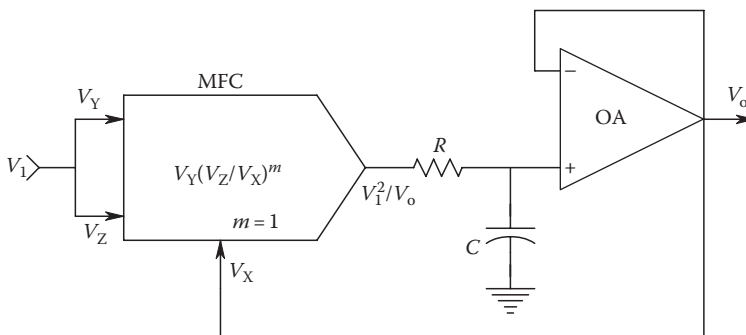


FIGURE 2.37 In this RMS converter, R and C form an LPF that performs the *mean* operation. A multifunction converter is used in a feedback circuit.

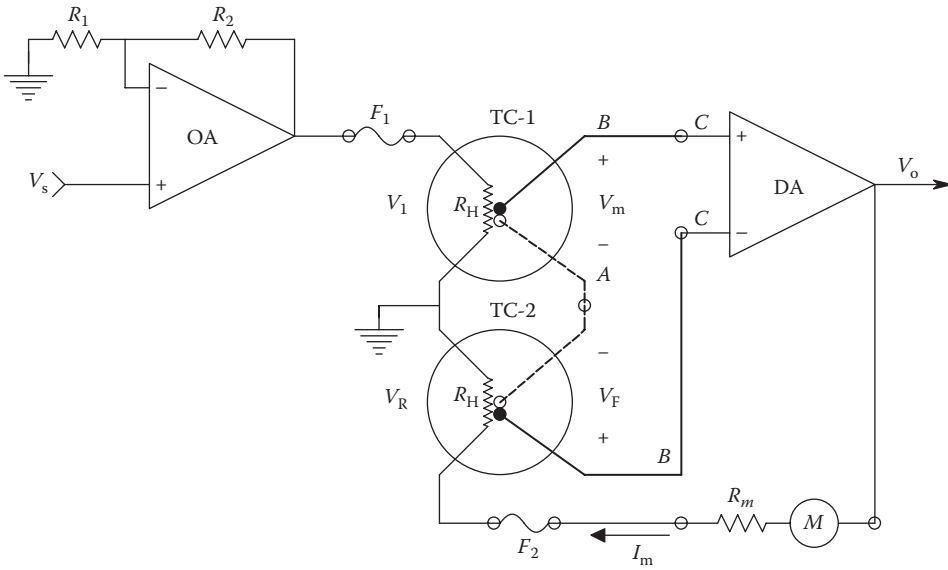


FIGURE 2.38

Schematic of a simple, feedback-type, true RMS voltmeter using a pair of matched vacuum thermocouples. Fuses F_1 and F_2 protect the delicate thermocouple heaters. I_m is a DC.

method shown in Figure 2.38. The thermocouple method is also used in the design of true RMS voltmeters. In the feedback thermocouple circuit, we may write

$$V_{m(\text{DC})} = K_T \overline{(V_1^2)}, \tag{2.134a}$$

$$V_{F(\text{DC})} = K_T \overline{(V_R^2)} = K_T I_m^2 R_H^2. \tag{2.134b}$$

Now the DC output current of the DA is just

$$I_m = \frac{K_D (V_m - V_F)}{R} = \frac{K_D K_T (V_1^2 - I_m^2 R_H^2)}{R}, \tag{2.135}$$

where

R is the sum of the thermocouple heater resistance

R_H , the microammeter resistance, R_m , and the DA's output resistance, R_o

Previously given solution of the quadratic equation in I_m yields

$$I_m = \frac{-R}{2K_D K_T R_H^2} \pm \left(\frac{1}{2} \right) \sqrt{\frac{R^2}{K_D^2 K_T^2 R_H^4} + \frac{4(V_1^2)}{R_H^2}}. \tag{2.136}$$

The exact solution for I_m reduces to

$$I_m = \frac{V_{1(\text{RMS})}}{R_H}, \tag{2.137}$$

for the condition

$$\overline{V_1^2} \gg \frac{R_2}{4K_D^2 K_I^2 R_H^4}, \quad (2.138)$$

which is easily met. If an analog microammeter is not used to read the system output, then it is easy to see that the DA's output voltage is given by

$$V_o = \frac{V_{1(\text{RMS})} R}{R_H}. \quad (2.139)$$

Several specialized true RMS-to-DC converter ICs have been developed. The Burr-Brown Model 4341 RMS-to-DC converter operates by first taking the absval of the input signal and then forming a voltage proportional to twice the log of the absval. An antilog transistor is used in a feedback circuit. The converter's DC output voltage, V_o , is fed back to the base of the antilogging transistor, so that its collector current is given by

$$i_c = k \log^{-1}(\log V_1^2 - \log V_o) = k \left(\frac{V_1^2}{V_o} \right). \quad (2.140)$$

$i_c(t)$ is averaged by an OA, current-to-voltage LPF to form V_o . It is easy to see that V_o is proportional to the true RMS value of V_1 . The BB 4341 true RMS converter of course works with DC inputs and is down 1% in frequency response at 80 kHz and -3 dB at 450 kHz. To be useful at low frequencies, the output must not contain ripple. Hence, the capacitor in the LPF (averaging) must be made very large. For example, for negligible output ripple given a 1 V, 10 Hz input sine wave, the filter capacitor should be over 100 μF .

Analog Devices also offers several IC, true RMS-to-DC converters having basically the same internal circuit architecture as the BB 4341. For example, the frequency response of the high-accuracy AD637 extends to 8 MHz. Its maximum output is 7 V RMS, and its accuracy is $5 \text{ mV} \pm 0.2\%$ of reading, with 0.02% nonlinearity for 0–2 V RMS input.

2.6.5 Square-Root Circuits and Dividers

The *square root* of a voltage signal may be taken in several ways. Obviously, a multifunction converter with $m = 0.5$ can be used. Another implicit method of square rooting is shown in Figure 2.39. Here, an analog multiplier is used to square the output and feed it back to the summing junction. The output voltage can be found from the following node equation, assuming the summing junction is at $V_1 \geq 0$:

$$\frac{V_1}{R_1} = \frac{V_o^2}{10R_F}. \quad (2.141)$$

This leads to the desired result:

$$V_o = \sqrt{V_1 \left(\frac{10R_F}{R_1} \right)}. \quad (2.142)$$

Analog division may also be accomplished by an implicit feedback circuit, one of which is shown in Figure 2.40. Here, V_1 and V_2 must be >0 . The node equation at the summing junction is, assuming an IOA,

$$\frac{V_1}{R_1} = \frac{V_2 V_o}{10R_F}. \quad (2.143)$$

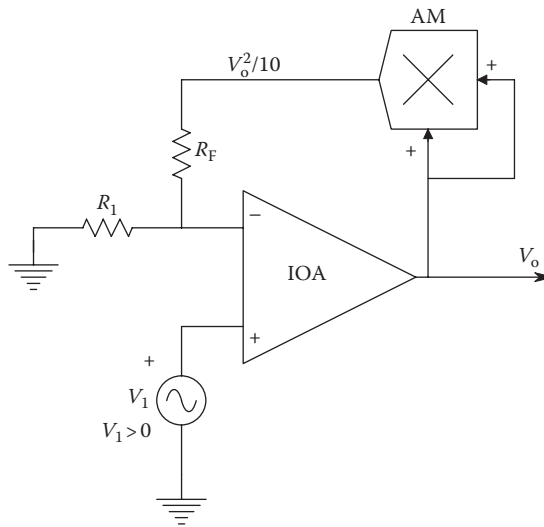


FIGURE 2.39
A feedback, square-rooting circuit. V_1 must be nonnegative. An analog multiplier IC is used.

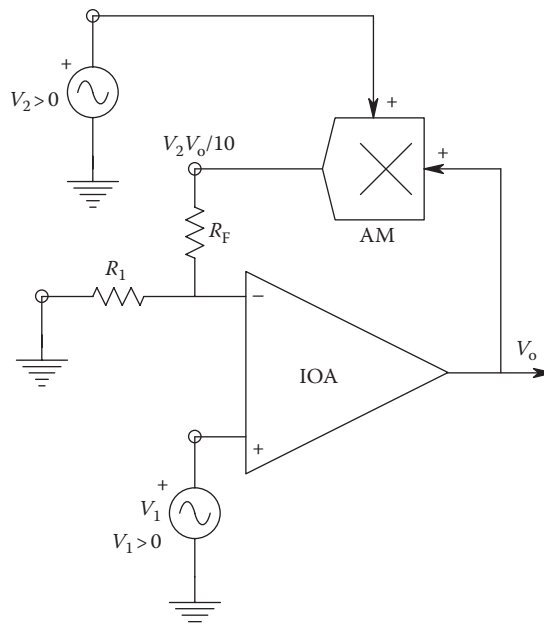


FIGURE 2.40
The circuit of Figure 2.39 is modified to make an analog divider. Again, V_1 must be nonnegative, and V_2 must be positive and greater than $V_{2MIN} = (V_{IMIN}/V_{OSat})(10R_F/R_1)$ to prevent OA output saturation and also less than the supply voltage of the OA (e.g., +15 V) to avoid damaging the OA's input transistors.

Hence,

$$V_o = \left(\frac{V_1}{V_2} \right) \left(\frac{10R_F}{R_1} \right) \tag{2.144}$$

Of course, a multifunction module with $m = 1$ will also give the same result.

2.6.6 Peak Detectors and Track-and-Hold Circuits

These circuits are closely related and will be treated together. Peak detectors follow an irregular waveform until it reaches a maximum (or minimum, if a negative peak detector is used), then hold that value at its output until an even greater maximum comes along, then hold that value, etc. Peak detectors are used in peak-reading voltmeters and in applications where measurement of the peak value of a waveform has significance (as opposed to the RMS value or rectified *average value*); an example of a waveform where the maximum and minimum values are important is arterial blood pressure (BP) recorded during surgery. The peak value is the systolic pressure, and the minimum value is the diastolic pressure. Figure 2.41 illustrates a two-OA peak detector of conventional design. Note that OA2 is used as a voltage follower. It should be an FET input type with a high input resistance and low bias current. Before a peak is reached in V_1 , D_1 is reverse-biased by the forward drop on D_2 and $V_3 = V_C = V_o = V_1$. After the peak in V_1 is reached, $V_1(t)$ decreases, D_2 is reverse-biased, and D_1 is forward-biased so that OA1 does not saturate. The output voltage remains at V_{1pk} , assuming zero capacitor leakage current, diode D_2 reverse current, and OA bias current. In practice, these currents will cause V_C to slowly drift. The reset switch, SW1, is generally a metal oxide semiconductor field effect transistor (MOSFET) switch that is used to reset V_o to zero when required.

The addition of a third diode and a resistor R_2 can improve the performance of the peak detector by eliminating the effective leakage current through D_2 . Reference to Figure 2.42

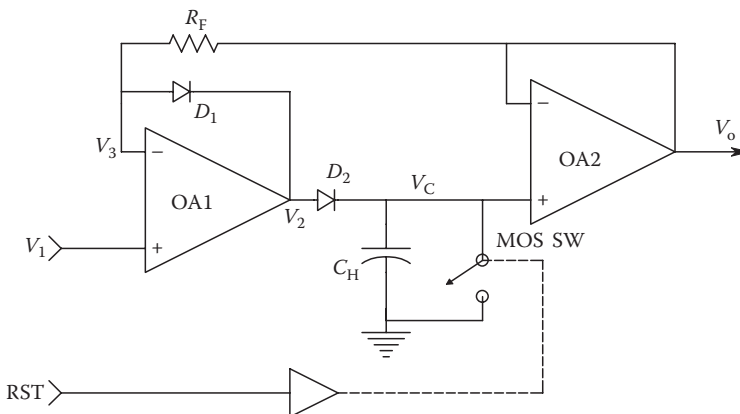


FIGURE 2.41
A MOSFET switch, SW1, is used to reset the two-OA peak detector.

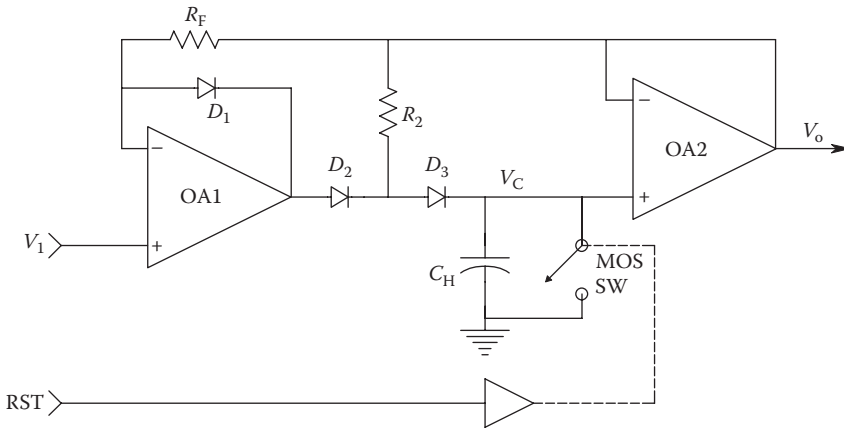


FIGURE 2.42
An improved peak detector design.

shows that D_3 is held at zero bias by R_2 , which pulls D_3 's anode to $V_o = V_C$ after the peak is reached. The overall rate of change in the peak detector's output voltage is thus practically

$$\frac{dV_o}{dt} = \frac{I_B}{C_H} \text{ V/s.} \tag{2.145}$$

There is a trade-off in determining the proper size of C_H . A large, low-leakage C_H will minimize drift but is expensive. Also, too large a C_H will not allow V_C to follow V_1 below the peak due to current saturation in OA1. Typical C_H values range from 1 nF to 0.1 μF . A good section on the design of peak detectors can be found in Franco (1988).

T&H (also called sample-and-hold) circuits are important components of many ADC systems. In operation, the output of a T&H circuit follows the conditioned analog input signal until the hold command is given. After a short delay, the T&H output assumes a constant level, during which the A/D conversion takes place. Constancy is required because certain ADC designs compare the analog input signal to an analog signal from a digital-to-analog converter (DAC) in the ADC whose output is determined from the output digital *word*. If the ADC's analog input changes during the conversion process, error will result. Ideally, the digital output of an ADC is the result of an impulse modulation where the analog input signal is multiplied by a periodic train of unit impulses. The output of the impulse modulator is a periodic sequence of numbers whose values represent the values of the analog input signal only at sampling instants. A T&H circuit allows a closer approximation of true impulse modulation to take place.

Figure 2.43 illustrates a basic T&H circuit. When the FET sampling switch is closed, the system is in the tracking mode. OA2 acts as a unity-gain buffer so $V_o = V_C = V_1$ as long as OA slew rates are not exceeded and the input signal frequencies lie well below the OAs' gain \times bandwidth products. C_H provides a capacitive load for OA1 and thus must be able to provide enough output current to satisfy $[C_H(dV_1/dt)_{MAX}]$. When the FET switch opens, the charge on C_H is trapped there, so V_o is held at V_C . However, things are not that simple; V_C changes slowly due to charge leaking off C_H through its leakage resistance and charge

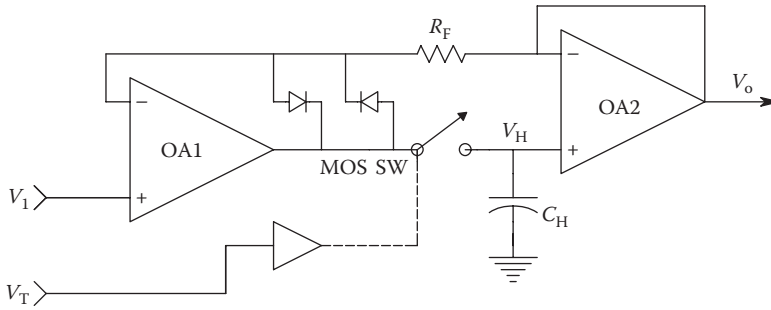


FIGURE 2.43
A basic, unity-gain, T&H circuit.

transfer to or from OA2 due to its DC bias current, I_B . This slow drift in V_C is called droop, and its slope can be either positive or negative, depending on the sign of I_B . In addition to the droop, there is also a feedthrough error caused by the fact that the FET switch does not have an infinite off resistance. When the FET switch is switched off, there is also an unwanted transfer of charge from the FET's gate to C_H through the FET's gate-to-drain capacitance. This charge can produce an unwanted *pedestal voltage*, which is given approximately by

$$\delta V_o = \left(\frac{C_{gd}}{C_H} \right) (V_{OB} - V_o), \quad (2.146)$$

where

V_{OB} is the off level bias voltage on the FET gate

V_o is of course the value of V_1 at the time the hold command is given

Thus, the pedestal voltage depends on the input voltage and, as numerical calculations reveal, can prove to be quite an objectionable source of T&H circuit error (see, e.g., Franco 1988, Sec. 7.11). A number of schemes have been devised to minimize pedestal error. One is to inject an equal and opposite switching charge into C_H by means of an auxiliary active circuit. Another is illustrated in the circuit architecture of the Burr-Brown SHC803BM T&H, shown in Figure 2.44. In this 1993 T&H system, a capacitor $C'_H = C_H$ was used to compensate for pedestal and droop at the same time, using the differential gain of OA2. Capacitor leakage, OA2's DC bias currents, and FET switch charge injection on switching cause equal voltage errors on both capacitors, which are subtracted out. C_H is operated in the integrator mode in the Burr-Brown design. In the hold mode, both SW1 and SW2 are off, and hence the charge on C_H is trapped, and V_o remains constant. In the tracking mode, we assume the on resistance of SW1 is 0Ω . Thus, V_o follows V_1 with the transfer function

$$\frac{V_o}{V_1} = \frac{-1}{1000C_H s + 1}. \quad (2.147)$$

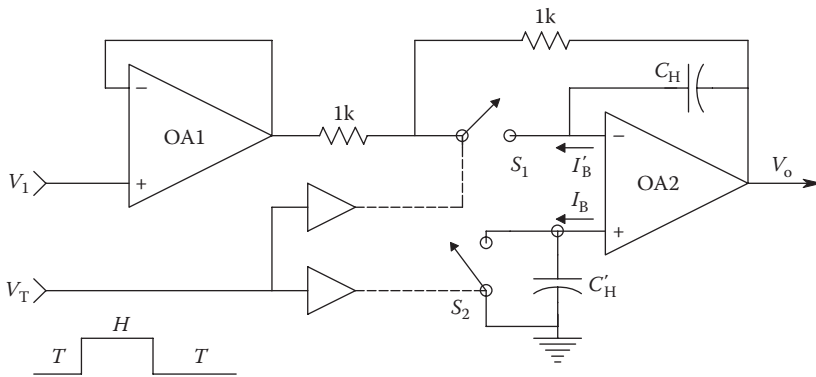


FIGURE 2.44 Basic architecture of the Burr-Brown SHC803BM T&H circuit. S_2 and capacitor C'_H are used for cancellation of the pedestal voltage.

If C_H is 100 pF, then the -3 dB frequency is 1.6 MHz. In general, because C_H appears in the feedback loop of OA2, the tracking-mode frequency response of the integrator type of T&H is lower than that attainable from the grounded C_H configuration shown in Figure 2.43. Figure 2.45 illustrates the T&H output errors discussed earlier. Illustrated are droop, signal feedthrough due to finite FET switch-off resistance, pedestal, and several important operating times specified by manufacturers (see caption).

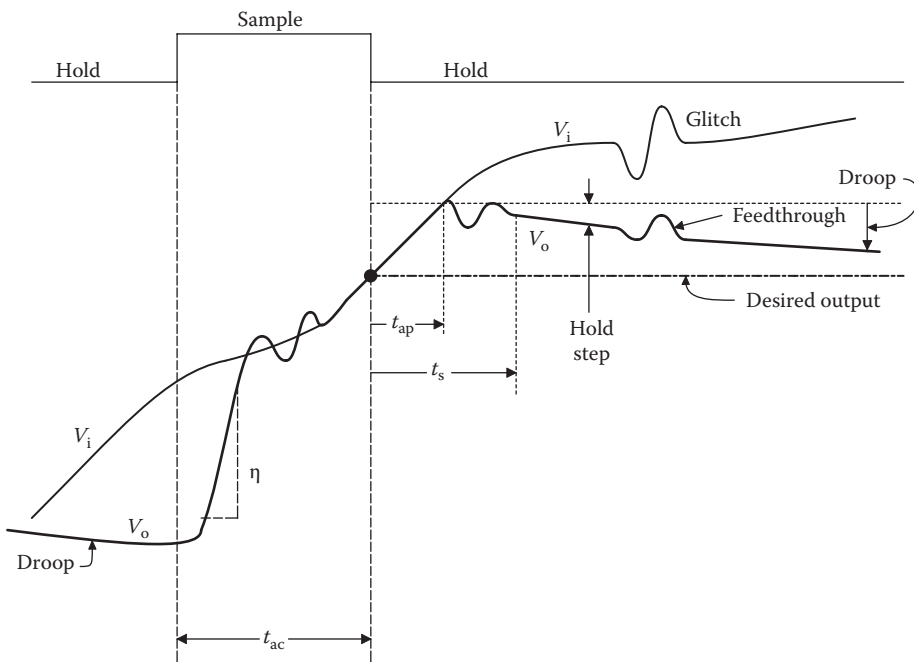


FIGURE 2.45 Waveforms seen in the operation of a typical T&H circuit. (From Franco, S., *Design with Operational Amplifiers and Integrated Circuits*, McGraw-Hill Book Co., New York, 1988. With permission.)

2.6.7 Log Ratio and Trigonometric ICs

Another class of nonlinear analog signal processing circuit that should be considered is the log ratio IC. The analog output voltage of this IC is given by

$$V_o = K \log_{10} \left(\frac{I_1}{I_2} \right) \tag{2.148}$$

Typically, K is set to some convenient integer value such as 1, 2, 3, or 5. Note that the inputs are currents that may vary over a six-decade range (e.g., 1 nA–1 mA). This nonlinear circuit module finds applications in spectrophotometry (analytical chemistry) where parameters such as absorbance or optical density need to be simply calculated and displayed. Also, if I_1 and I_2 are made to be DC proportional to the RMS value of LSS’ sinusoidal output and input, respectively, the log ratio converter output is proportional to the frequency response magnitude (Bode log magnitude) in dB. Of course if I_2 is fixed, then we have a simple log converter.

The Burr-Brown LOG100 is a commercial log ratio converter; Figure 2.46 shows the internal circuit of this IC. Analog devices offered a similar unit, the AD757 (now obsolete), which will accept either current or voltage inputs. Note that log ratio modules are normally used with DC inputs. Their frequency responses are a function of the input signal amplitude, ranging from about 100 Hz for nanoamp peak input currents to over 40 kHz for milliamp peak input currents.

Often, the system’s nonlinear signal processing requirements demand a broadband logarithmic nonlinearity, which can operate on video-frequency signals. The Analog Devices AD640 is an example of an IC that can produce an output of the form

$$V_o = V_y \log \left(\frac{V_1}{V_{sc}} \right) \tag{2.149}$$

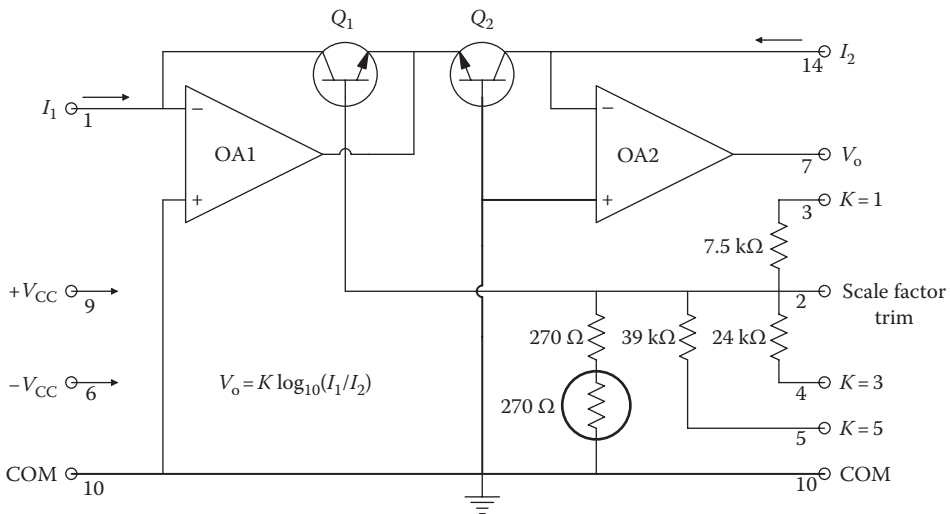


FIGURE 2.46 Simplified schematic of the Burr-Brown LOG100 log ratio converter. (Courtesy of Burr-Brown, Tucson, AZ.)

over a 46 dB conversion range and an 80 MHz bandwidth. Here, V_{sc} is the intercept scaling voltage and V_y is the slope voltage, both constants. Two cascaded AD640s are claimed to have a 95 dB dynamic range in a 20 Hz–100 kHz bandwidth; a 70 dB range is claimed in a 50–150 MHz BP. The IC is direct-coupled, however, and will work from DC through the audio spectrum.

The Analog Devices AD639 trigonometric converter is another specialized, nonlinear, analog processing module that has applications such as generation of a sine wave from a triangle wave at frequencies up to 1.5 MHz, coordinate conversion and vector resolution, design of quadrature and variable phase oscillators, and imaging and scanning linearization circuits. The nonlinear transfer function of the AD639 is

$$V_o = U \frac{\sin(V_{x1} - V_{x2})}{\sin(V_{y1} - V_{y2})}. \quad (2.150)$$

The scaling factor for the differential voltage inputs is $50^\circ/\text{V}$. The IC has on board a 1.8 V DC bandgap voltage reference that can be used to preset either angle argument to 90° . The output of the AD639 is scaled by the constant, U , which can be set between 0 and +15 V by either external or internal means. A wide variety of trig functions can be modeled: $\sin(x)$, $\cos(x)$, $\tan(x)$, $\text{cosec}(x)$, $\sec(x)$, and $\cotan(x)$, as well as certain useful inverse trig functions. For the arctangent,

$$V_o = K\theta = K \tan^{-1} \left(\frac{V_{z2} - V_{z1}}{V_{u1} - V_{u2}} \right), \quad (2.151)$$

where $K = 20 \text{ mV}/^\circ$. With minor changes in the external circuit architecture, the AD639 will also calculate the arcsine and arctangent functions with the same ratio of voltage differences argument as the preceding arctangent function. With the exception of sine wave generation, it would appear that this versatile IC is best used with inputs that are DC or at best in the low audio-frequency range.

2.6.8 Summary

We have begun this section by discussing some of the properties of linear and nonlinear analog systems. There are many uses for nonlinear analog signal processing ICs. The present trend is to sample and digitize the input signal and then use a computer to calculate the nonlinear operation and then realize the result in analog form by DAC. However, there are instances where the instrumentation system cost will be less expensive if an all-analog approach is taken and use is made of a specialized analog function module. Computers, while versatile, can be expensive. Obviously, there is a trade-off between accuracy, cost, and convenience in choosing between an all-analog vs. an A/D and a computer vs. a D/A type of system for nonlinear signal processing in an instrumentation system. The instrumentation system designer should be aware that all-analog signal processing ICs are available that may compete with the digital approach in terms of function, accuracy, and cost.

2.7 Charge Amplifiers

2.7.1 Introduction

Charge amplifiers make use of electrometer (FET input) OAs that have ultralow DC bias currents at their inputs and ultrahigh input resistances. They find primary application in isolating and conditioning the outputs of piezoelectric transducers used to measure pressure and force transients. They are also used to measure the charge in coulombs stored on capacitors and, indirectly, electric field strength.

2.7.2 Charge Amplifiers Used with Piezoelectric Transducers

In Chapter 6, we show that the equivalent circuit of a piezoelectric pressure transducer operated well below its mechanical resonant frequency is given by the simple parallel current source, conductance, and capacitance, shown in Figure 2.47. Note that d is a constant that depends on the piezoelectric material and how it is cut relative to the crystal's axes to form the transducer and that, in general, current is the rate of transfer of charge through a plane through which it is passing. By US convention, current is conventionally taken as positive in the direction of motion of positive charges, such as holes in semiconductors, or positive ions in solutions or plasma discharges. Hence, in a conductor, such as copper, where charge is carried by electrons, the current direction is opposite to the direction of electron motion.

Figure 2.48 illustrates an IOA used as a charge amplifier. The OA is assumed to have infinite gain, frequency response and input resistance, and zero bias current, offset voltage, noise, and output resistance. Under these IOA assumptions, it is easy to see that because the summing junction is at 0 V, all of the current, $\dot{P}d$, flows into the feedback capacitor, C_F . Hence, V_o is proportional to $\dot{P}(t)$ and is given by

$$V_o(t) = \frac{\dot{P}(t)d}{C_F}. \tag{2.152}$$

This result of the ideal case is seen to become less simple when we face reality and note that the EOA has a noninfinite gain approximated by

$$K_V = \frac{K_{V_o}}{\tau s + 1} (V_i - V_i'). \tag{2.153}$$

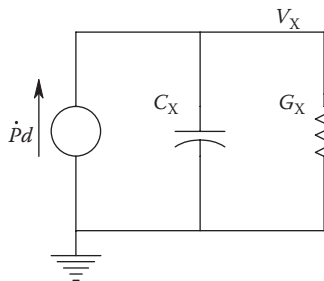
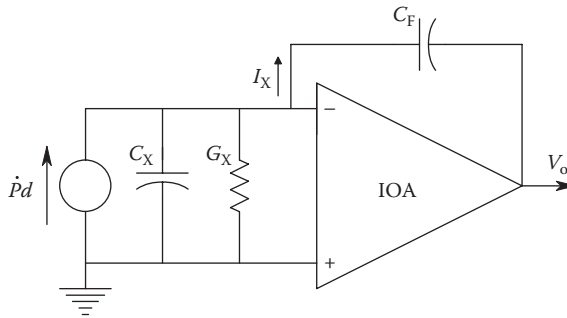


FIGURE 2.47

Equivalent circuit of a piezoelectric crystal responding to pressure on its active surface at frequencies well below its mechanical resonance frequency.


FIGURE 2.48

A charge amplifier circuit used to condition the output of a piezoelectric crystal sensor.

Also, we assume that there is a small but finite conductance, G_F , in parallel with C_F . G_F is in the order of 10^{-10} Siemens or smaller. Now if we write the node equation at the summing junction and substitute the preceding gain expression to eliminate V_i' , we can show that the charge amplifier's transfer function is given by (approximately)

$$\frac{V_o}{P} = \frac{-s(d/G_F)}{s^2[\tau(C_T + C_F)/K_{vo}G_F] + s[\tau(G_T + G_F) + K_{vo}C_F]/K_{vo}G_F + 1}. \quad (2.154)$$

This is the transfer function of a BP system. The LF pole is at about

$$f_{\text{low}} \approx \frac{1}{2\pi C_F R_F} \text{ Hz}. \quad (2.155)$$

The midband gain is

$$A_{V_{\text{mid}}} = \frac{-K_{vo}d}{\tau(G_T + G_F) + K_{vo}C_F} \cong \frac{-d}{C_F}. \quad (2.156)$$

The HF pole is at

$$f_{\text{hi}} = \frac{K_{vo}C_F}{2\pi\tau(C_T + C_F)} \text{ Hz}. \quad (2.157)$$

Note that $K_{vo}/(2\pi\tau)$ is simply the SS gain \times bandwidth product of the EOA. Figure 2.49 shows the circuit of the nonideal charge amplifier. Here, C_T is the total capacitance to ground that includes the piezoelectric crystal's capacitance, the capacitance of the cable connecting the transducer to the charge amp, and the charge amp's input capacitance. Typically C_T might be about 400 pF. G_T is the total shunt conductance to ground at the OA's summing junction. G_T includes the crystal's leakage conductance, the cable's leakage conductance, and the input conductance of the OA. Typically, G_T might be about 10^{-13} S. If, in addition, we take C_F to be 2×10^{-9} F, G_F to be 10^{-10} S, $K_{vo} = 5 \times 10^4$, $d = 2 \times 10^{-12}$ C/psi for quartz, and the OA's gain \times bandwidth product to be 350 kHz, the system ends up having a midband gain of 1 mV/psi, an HF pole at 1.5×10^5 Hz, and an LF pole at 8×10^{-3} Hz.

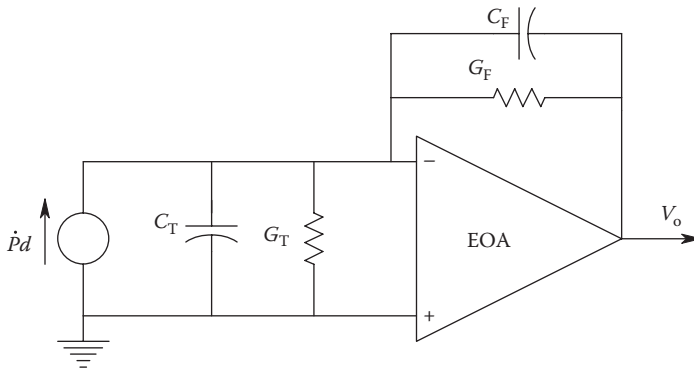


FIGURE 2.49
Circuit of a practical charge amplifier and crystal sensor.

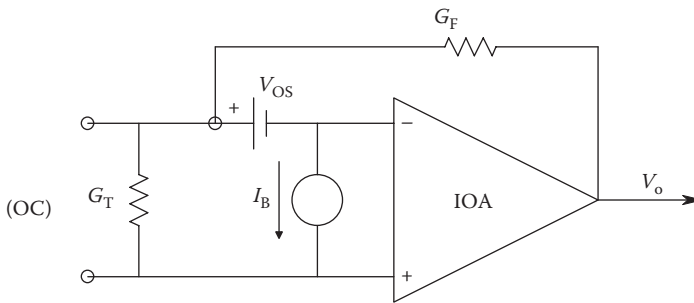


FIGURE 2.50
The DC equivalent circuit of a practical charge amplifier. V_{OS} = DC offset voltage, I_B = DC input bias current.

Note that the LF pole is set by C_F/G_F and is independent of the input circuit. Thus, the piezoelectric transducer charge amplifier is not a system that can respond to static (DC) pressures; it must be used to measure transient pressure changes. In addition to the LF limitations imposed on the transfer function by nonideal EOA behavior, there are also DC errors at the output due to the OA’s bias current and DC offset voltage. Figure 2.50 shows the DC equivalent circuit of the charge amplifier. Note that the capacitors drop out from the circuit at DC because they carry no current. It is easy to see that the DC error in the output is

$$V_{oDC} = I_B R_F + V_{OS} (1 + G_T/G_F). \tag{2.158}$$

Note that I_B and V_{OS} can have either sign. Typical values of V_{OS} are about $\pm 200 \mu\text{V}$, and I_B values are around 75 fA ($75 \times 10^{-15} \text{ A}$).

2.7.3 Charge Amplifier as an Integrating Coulombmeter

A DC charge of Q_1 coulombs is stored in a capacitor, C_1 . The capacitor can be a real capacitor or the capacitance of a charged object such as the fuselage of a helicopter or a human body. We wish to measure Q_1 . The voltage across C_1 is given by the well-known relation, $V_1 = Q_1/C_1$.

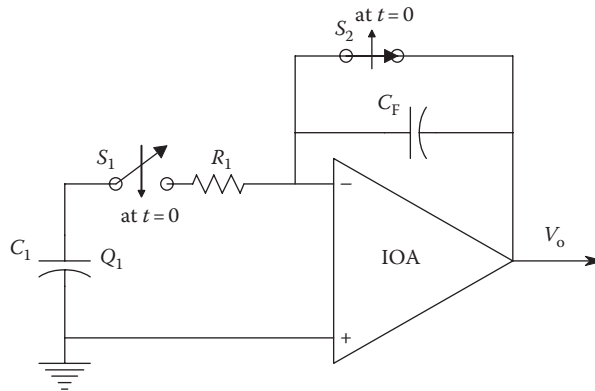


FIGURE 2.51
An integrating coulombmeter circuit. An EOA is generally used.

In Figure 2.51, C_1 is connected to the summing junction of an EOA and at the same time the short is removed from the feedback capacitor, C_F , allowing it to charge from the OA's output to a value that maintains the summing junction at ≈ 0 V. Assuming the OA to be ideal, the output voltage of the OA, after three or four R_1C_1 time constants, is easily shown to be

$$V_{o(SS)} = \frac{-Q_1}{C_F} \text{ dc V.} \tag{2.159}$$

Hence, the system can measure the charge on a small capacitor, providing the peak current $i_{1PK} = Q_1/R_1C_1$ does not exceed the current sourcing or sinking capability of the OA.

In practice, this circuit is subject to DC errors caused by a practical OA's input bias current and offset voltage. The output error caused by I_B and V_{OS} can be found by superposition and is given by

$$\delta V_o = T \left(\frac{I_B}{C_F} \right) + V_{OS} \left(1 + \frac{C_1}{C_F} \right) V, \tag{2.160}$$

where T is the time elapsed between opening switch 1 and closing switch 2 and reading the SS V_o . In practice, δV_o can be made quite small by judicious choice of circuit parameters.

In this section, we have presented two examples of applications of the OA charge amplifier configuration. The basic charge amplifier consists of an IOA with an ideal feedback capacitor to integrate charge transferred to the summing junction. In practice, we use electrometer amplifiers that have ultrahigh input resistances and finite gains, bandwidths, offset voltages, and bias currents. Because a pure C_F will integrate the bias current of the summing junction, a large resistance, R_F , must be put in parallel with C_F to give a fixed DC error proportional to I_B . The parallel combination of R_F and C_F sets the LF pole of a BP frequency response for the piezoelectric transducer. In the midband range of frequencies, the output voltage is proportional to the pressure on the transducer, rather than its derivative.

2.8 Phase-Sensitive Rectifiers

2.8.1 Introduction

PSRs, also called *phase-sensitive demodulators* or *synchronous rectifiers*, are an important class of analog circuit widely used in instrumentation, control, and communications systems. The PSR is used to convert the output of certain ac, carrier-operated sensors to an LF signal proportional to the QUM. One example of the use of PSRs is to condition the output of a resistive Wheatstone bridge using strain gauges to measure mechanical strain, pressure, or force. The strain gauge bridge excitation is generally an audio-frequency sine wave. The bridge output is a sine wave of the same frequency with amplitude proportional to the strain, etc. The phase of the bridge output is 0° with respect to the excitation phase if $\Delta X > 0$ and is -180° with respect to the excitation if $\Delta X < 0$. The PSR uses the bridge excitation voltage to sense this phase change and to produce a rectified signal that has amplitude proportional to ΔX and the sign of ΔX .

2.8.2 Double-Sideband, Suppressed-Carrier Modulation

The outputs of many sensor systems have information encoded on them in what is called double-sideband, suppressed-carrier modulation (DSBSCM). A DSBSCM signal is formed by multiplying the ac carrier by the modulating signal. To illustrate this process, we examine a Wheatstone bridge with one variable element and AC excitation, as shown in Figure 2.52. It is easy to show (see Chapter 4) that the bridge output is given by

$$V'_o(t) = K_D [V_b \sin(\omega_c t)] \left[\frac{X}{X+P} - \frac{N}{M+N} \right]. \tag{2.161}$$

Now we let the resistance X vary in time so

$$X(t) = X_o + \Delta X(t). \tag{2.162}$$

Furthermore, we balance the bridge and make it maximally sensitive. This requires that

$$M = N = P = X_o. \tag{2.163}$$

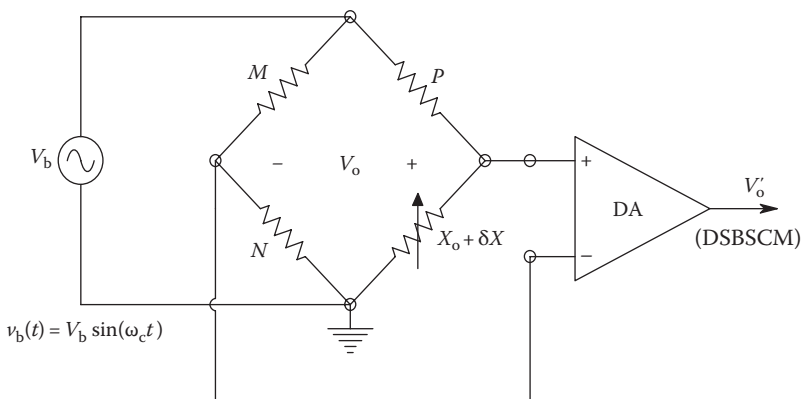


FIGURE 2.52
A Wheatstone bridge circuit with AC excitation generates a DSBSCM output.

Thus, around null,

$$V'_o(t) = \Delta X(t) K_D V_b \left(\frac{1}{4X_o} \right) \sin(\omega_c t). \tag{2.164}$$

Note that we have the product of $\Delta X(t)$ and the excitation sinusoid at the bridge output. If $\Delta X(t)$ is sinusoidal with a $\omega_m \ll \omega_c$, we can write, using the well-known trig identity, $[\sin(\alpha) \sin(\beta)] = [\cos(\alpha - \beta) - \cos(\alpha + \beta)]/2$:

$$V'_o(t) = K_D V_b \left[\frac{\Delta X}{8X_o} \right] \left[\cos((\omega_c - \omega_m)t) - \cos((\omega_c + \omega_m)t) \right]. \tag{2.165}$$

Thus, from Equation 2.165, we see that the conditioned bridge output, $V'_o(t)$, consists of two sidebands and no carrier frequency term, hence the designation DSBSCM.

Another example of a transducer that produces a DSBSCM output is the *linear variable differential transformer* (LVDT), a device that is used to sense small, linear mechanical displacements. Some LVDTs come with a built-in PSR and provide a DC output directly proportional to core displacement; others must be used with an external PSR. (LVDTs are discussed in detail in Chapter 6.)

2.8.3 Demodulation of DSBSCM Signals by Analog Multiplier

Figure 2.53 illustrates one of the simplest ways to demodulate a DSBSCM signal and recover $\Delta X(t)$. An analog multiplier IC is used to multiply a reference signal that is in-phase with the transducer's AC excitation by the transducer's DSBSCM output. This product can be written as

$$V_z(t) = \frac{\Delta X K_D V_b V_r}{(10)8X_o} \left\{ \sin(\omega_c t) \cos[(\omega_c - \omega_m)t] - \sin(\omega_c t) \cos[(\omega_c + \omega_m)t] \right\}. \tag{2.166}$$

The bracketed term can again be expanded by trig identity:

$$V_z(t) = \frac{\Delta X K_D V_b V_r}{(10)16X_o} \left\{ \sin[(2\omega_c - \omega_m)t] + \sin(\omega_m t) - \sin[(2\omega_c + \omega_m)t] - \sin(-\omega_m t) \right\}. \tag{2.167}$$

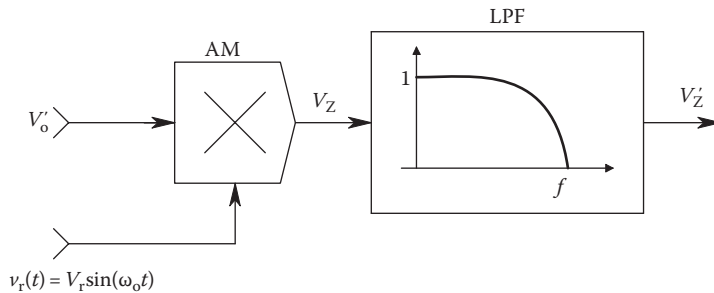


FIGURE 2.53
An analog multiplier followed by an LPF is used to demodulate a DSBSCM signal.

$V_z(t)$ is now passed through a unity-gain LPF that removes the two HF terms. The output of the LPF is the desired, demodulated signal, multiplied by a scale factor:

$$V'_z(t) = \frac{K_D V_b V_r}{80 X_o} [\Delta X \sin(\omega_m t)]. \quad (2.168)$$

2.8.4 Other PSR Designs

Figure 2.54A illustrates a simple PSR that can be made from two OAs and a bidirectional MOS switch operated by a logical square wave derived from the sinusoidal excitation source by first conditioning it with a comparator used as a zero-crossing detector and then passing the TTL waveform through two one-shot multivibrators (e.g., a 74LS123) to adjust the reference signal's phase. Because the MOS switch closes at the same time in the reference signal's cycle, when the DSBSCM signal changes phase 180°, the rectifier's output changes sign, producing a negative output. An LPF is used to smooth the full-wave rectified output of the switching circuit. The relevant waveforms are shown in Figure 2.54B.

PSR ICs are available from a variety of manufacturers and include the Analog Devices AD630 balanced modulator/demodulator, the NS LM1596 balanced modulator/demodulator, and the Philips Semiconductor (formerly Signetics) Model NE/SE5521 LVDT signal conditioner, to mention a few.

2.8.5 Lock-In Amplifier

2.8.5.1 Introduction

The lock-in amplifier (LIA) can be used to condition and demodulate DSBSCM signals and recover the modulating signal under conditions of abominable SNRs, where the RMS noise voltage can exceed that of the modulated carrier by over 60 dB. Noise is reduced in the operation of an LIA by three mechanisms: (1) BP filtering around the carrier frequency, (2) rejection of coherent interference at frequencies other than that of the carrier, and (3) low-pass filtering at the PSR's output. Carrier frequencies are best chosen to lie between 10 and 40 Hz and 200–10⁵ Hz. Carrier frequencies lying too close to the 60 Hz powerline frequency and its first two harmonics are avoided because they may *leak* through the LPF following the PSR. Very low carrier frequencies are also avoided as they may place LIA operation in the range of 1/*f* noise from the analog electronic circuits.

A block diagram of an LIA is shown in Figure 2.55, along with some pertinent LIA waveforms in Figure 2.56. We see that the basic LIA performs four basic operations on the DSB-modulated carrier with additive, broadband noise: the incoming DSBSCM signal plus noise is generally conditioned by a BP amplifier whose center frequency is the same as that of the carrier. The BP signal conditioning improves the SNR of the modulated carrier that next passes through a PSR, usually of the switching type shown in Figure 2.54A. The PSR output is next conditioned by a DC pass LPF, which generally has an adjustable cutoff frequency and attenuation slope. Some DC postamplification by a low-noise, low-drift amplifier may precede the LPF, if required. A reference signal directly derived from the carrier is used to operate the PSR.

For optimum performance of the PSR, the phase of the reference signal must be adjusted to be exactly in-phase with the carrier in the DSBSCM input. To illustrate this principle of LIA operation in the noise-free case, we let the modulating signal be a DC voltage:

$$v_m(t) = V_m. \quad (2.169)$$

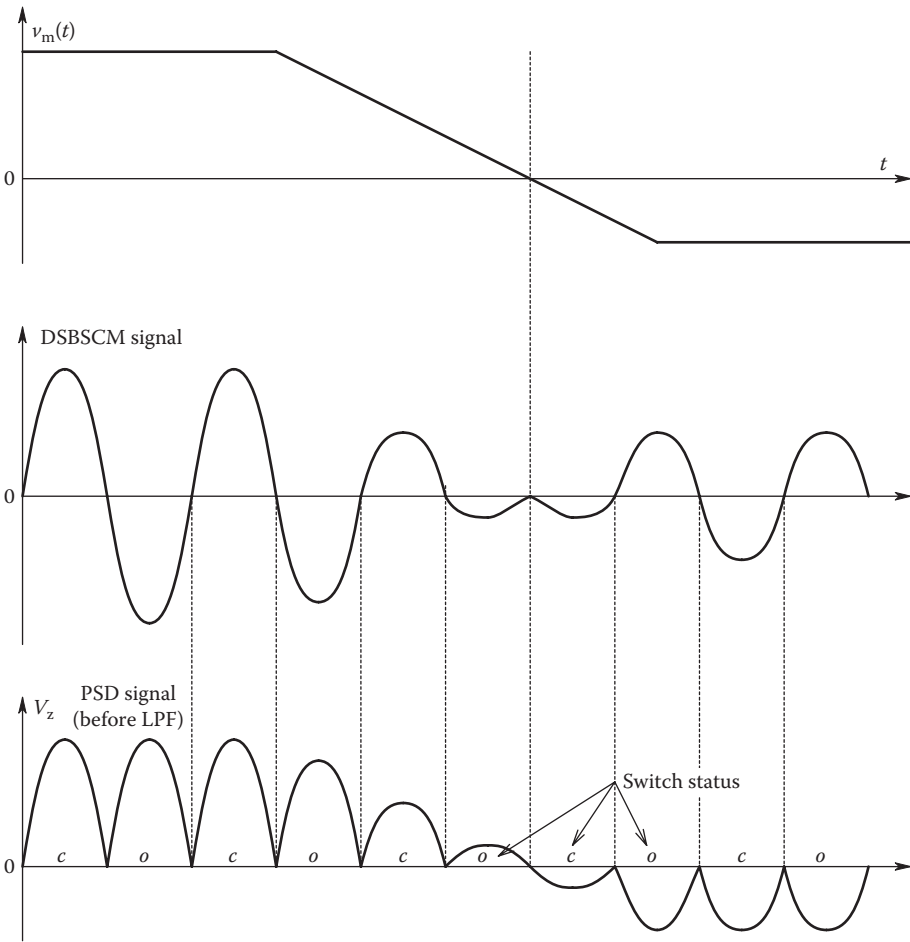
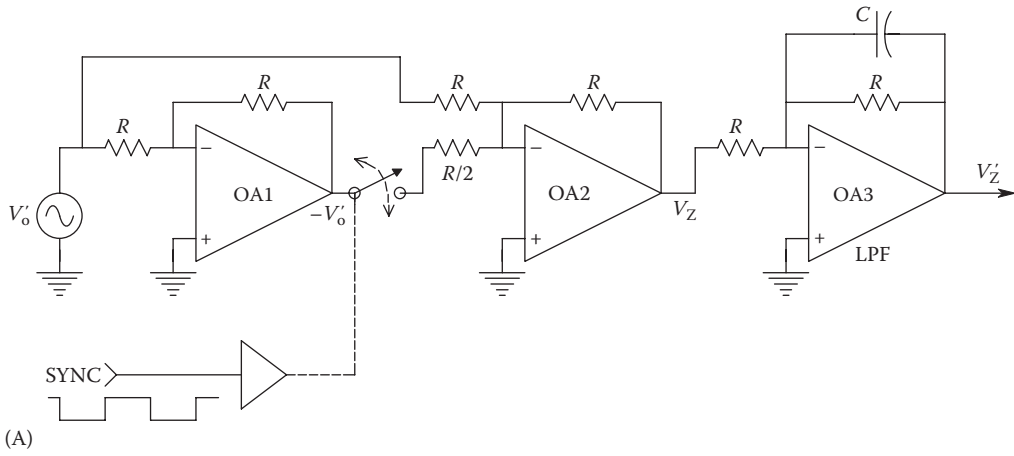


FIGURE 2.54

(A) A simple PSR and LPF circuit using OAs and a MOS switch. (B) Waveforms in the PSR/LPF circuit.

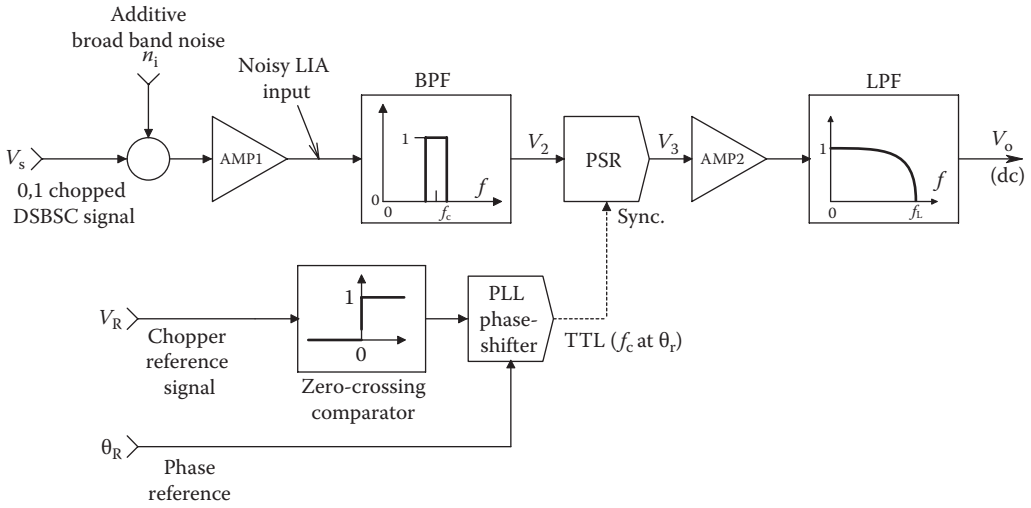


FIGURE 2.55

Block diagram of an LIA. Input is a 0,1 chopped, noisy signal, such as the output of a photosensor. Amplification and BP filtering at the chopper frequency converts the input signal into a noisy AC signal, which is then rectified by the PSR. The LF modulating signal is seen at the LIA output. The LPF bandwidth is adjusted to pass the desired modulating signal with the least amount of noise.

The carrier and reference signal are sinusoids, respectively,

$$v_c(t) = V_c \sin(\omega_c t + \theta_c), \tag{2.170}$$

$$v_r(t) = V_r \sin(\omega_c t + \theta_r). \tag{2.171}$$

The DSBSM signal is

$$v_{mc}(t) = V_m V_c \sin(\omega_c t + \theta_c). \tag{2.172}$$

Demodulation, as we have shown earlier, requires the multiplication of the modulated carrier with the reference signal. A scale factor K_D is introduced. Thus, the demodulated signal before low-pass filtering can be written as

$$v_d(t) = K_D V_m V_c V_r \sin(\omega_c t + \theta_c) \sin(\omega_c t + \theta_r). \tag{2.173}$$

After the $\sin(\alpha) \sin(\beta)$ trig identity is applied, we have

$$v_d(t) = \left(\frac{K_D V_m V_c V_r}{2} \right) \left[\cos(\theta_c - \theta_r) - \cos(2\omega_c t + \theta_c + \theta_r) \right]. \tag{2.174}$$

After low-pass filtering, the DC output of the LIA can be written as

$$V_o = V_m \left[\frac{K_D K_F V_c V_r}{2} \right] \cos(\theta_c - \theta_r), \tag{2.175}$$

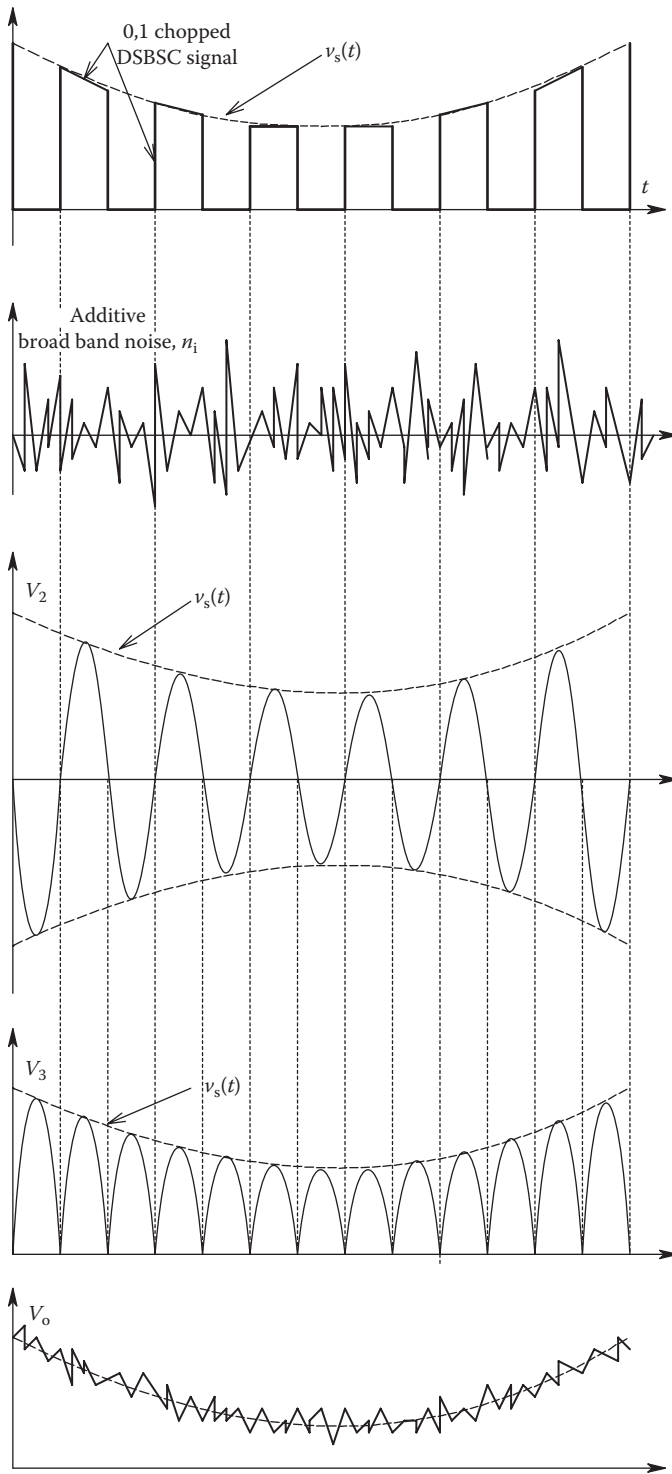


FIGURE 2.56
Waveforms in the LIA of Figure 2.55.

where K_F is the DC gain of the LPF. Note that the output is maximum when the phase of the reference signal and the carrier are equal, that is, there is zero phase shift between the carrier and reference signal. Also, for the LIA to remain in calibration, K_D , K_F , V_c , and V_r must remain constant and known. Most modern LIAs have adjustable phase shifters in the reference channel to permit output sensitivity maximization; others perform this task automatically using phase-lock circuitry (cf. McDonald and Northrop 1993). Figure 2.57A illustrates the generalized block diagram of a two-phase, vector-tracking LIA devised by McDonald and Northrop. Figure 2.57B shows a functional block diagram of the two-phase vector-tracking LIA. This LIA was found experimentally to operate robustly, tracking the input signal's phase buried in 60 dB of noise, over a $\pm 360^\circ$ phase difference between the reference waveform and the signal in noise

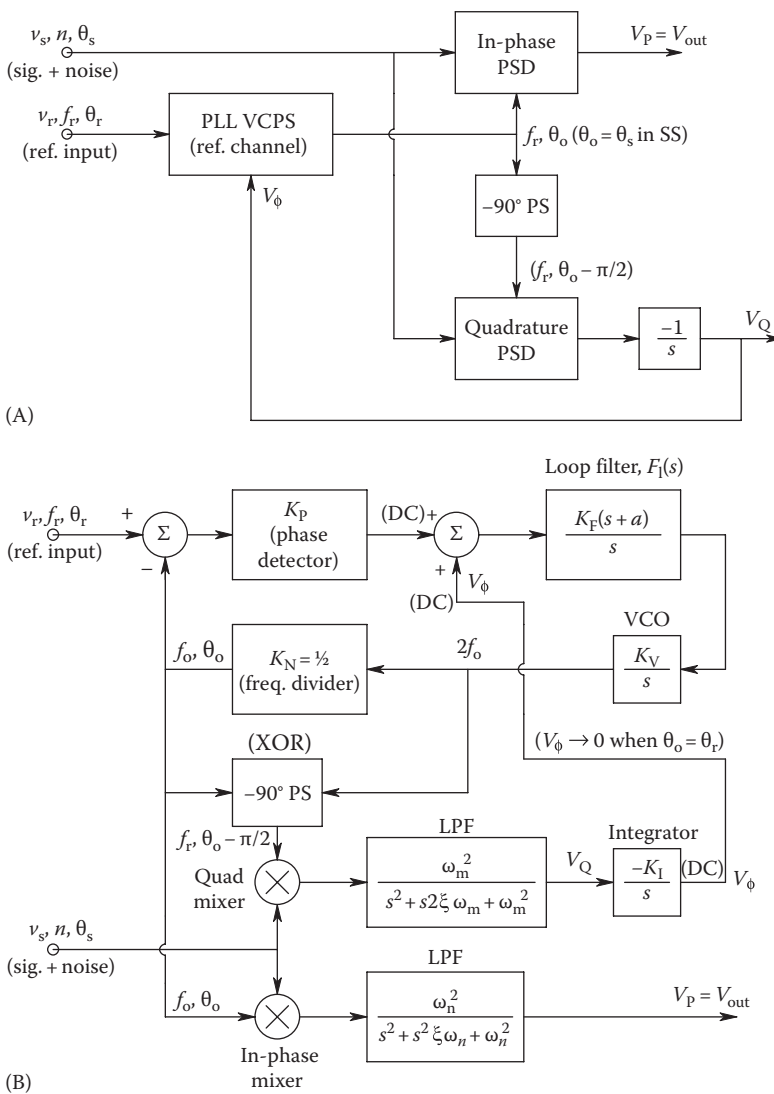


FIGURE 2.57 (A) Block diagram of McDonald's vector-tracking LIA. (B) Functional block diagram of McDonald's vector LIA.

being detected. An integrator in the quadrature negative feedback loop ensures that the quadrature component of the signal is maintained in the SS at 0 V—thus continuously nulling the phase angle between the reference signal and the signal being recovered, thus maximizing the LIA output SNR.

2.8.5.2 Calculation of the SNR Improvement Using a Lock-In Amplifier

The major application of an LIA is to measure coherent, double-sideband, suppressed-carrier (DSBSC) signals, generally at audio frequencies, which are *buried in noise*. More precisely, the input SNR is very low, often as low as -60 dB. There are two, basic architectures used for LIAs. One uses an analog multiplier for signal detection followed by an LPF; this system is shown in Figure 2.58A. A second LIA architecture, shown in Figure 2.58B, uses a synchronous rectifier followed by an LPF to measure the amplitude of a coherent signal of constant frequency that is present in an overwhelming amount of noise. Assume the input to the LIA is

$$x(t) = n_1(t) + V_s \cos(\omega_0 t). \tag{2.176}$$

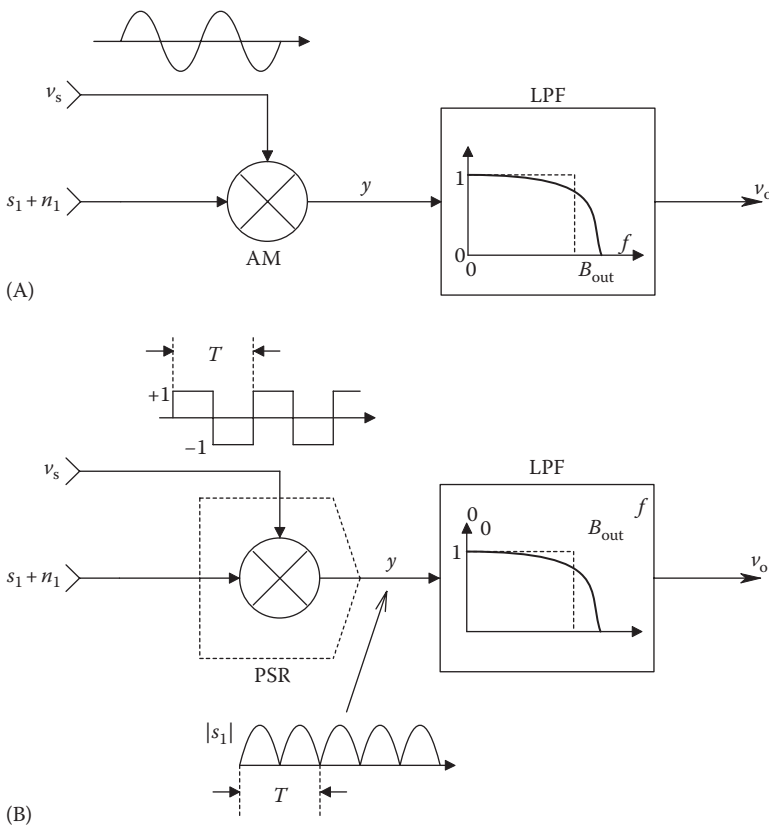


FIGURE 2.58 Two means of demodulating DSBSCM signals used in LIAs. (A) Demodulation by analog multiplier and LPF. (B) Demodulation by synchronous rectifier and LPF.

Assume the noise is zero-mean, Gaussian white noise with one-sided power density spectrum (PDS), $S_n(f) = \eta$ msV/Hz. The mean squared (MS) input SNR is

$$\text{SNR}_{\text{in}} = \frac{V_s^2/2}{\eta B_{\text{in}}}, \quad (2.177)$$

where B_{in} is the Hz bandwidth over which the input noise is measured.

In the analog multiplier form of the LIA, $x(t)$ is multiplied by an in-phase sync signal, $v_s(t) = A \cos(\omega_0 t)$. The multiplier output is thus

$$\begin{aligned} y(t) &= v_s(n_i + s_i) = A \cos(\omega_0 t) [n_i(t) + V_s \cos(\omega_0 t)] = An_i(t) \cos(\omega_0 t) + AV_s \cos^2(\omega_0 t) \\ y(t) &= An_i(t) \cos(\omega_0 t) + \left(\frac{AV_s}{2}\right) [1 + \cos(2\omega_0 t)]. \end{aligned} \quad (2.178)$$

Note that the one-sided power spectrum of the first term in Equation 2.178 is the noise PDS shifted up to be centered at $\omega = \omega_0$. Because $S_n(f)$ is white, shifting the center of $S_n(f)$ from 0 to $f_0 = \omega_0/2\pi$ does not change the noise output of the LIA's output BPF. Thus, the MS output noise voltage from the LIA is $v_{\text{on}}^2 = A^2\eta(1/4\tau)$ msV, where $(1/4\tau)$ can be shown to be the equivalent noise Hz bandwidth of the simple output RC LPF and τ is the output LPF's time constant. The output LPF removes the $2\omega_0$ frequency term in the right-hand term of Equation 2.178, leaving a DC output proportional to the peak value of the sinusoidal signal. Clearly, $v_{\text{OS}} = AV_s/2$. Thus, the MS SNR at the LIA output is

$$\text{SNR}_{\text{out}} = \frac{A^2V_s^2/4}{A^2\eta(1/4\tau)} = V_s^2\tau/\eta. \quad (2.179)$$

The noise figure, F , is defined as the ratio of SNR_{in} to SNR_{out} . For an amplifier, F is desired to be as small as possible. In a noise-free amplifier (the limiting case), $F = 1$. It is easy to show that F for this kind of LIA is

$$F_{\text{LIA}} = \frac{B_{\text{out}}}{B_{\text{in}}}, \quad (2.180)$$

where

$B_{\text{out}} = 1/4\tau$ is the Hz noise bandwidth of the LIA's output LPF

B_{in} is the Hz noise bandwidth over which the input noise is measured

Since the LIA's desired output is a DC signal proportional to the peak sinusoidal input signal, B_{out} can be on the order of single Hz, and F_{LIA} can be $\ll 1$.

In the second model for LIA noise, consider the synchronous rectifier LIA architecture of Figure 2.58B. Now the sinusoidal signal is full-wave rectified by passing it through an absval function that models the synchronous rectifier. The full-wave rectified, sinusoidal input signal can be shown to have the Fourier series (Aseltine 1958):

$$y(t) = |v_s(t)| = V_s \left(\frac{2}{\pi}\right) \left[1 - \left(\frac{2}{3}\right) \cos(2\omega_0 t) - \left(\frac{2}{15}\right) \cos(4\omega_0 t) - \dots \right]. \quad (2.181)$$

Clearly, the average or DC value of $y(t)$ is $V_s(2/\pi)$.

Curiously, the zero-mean white noise *is not* conditioned by the absval function of the synchronous rectifier. This is because the synchronous rectifier is really created by effectively multiplying $x(t) = n_i(t) + V_s \cos(\omega_0 t)$ by an even ± 1 square wave having the same frequency as, and in-phase with, $V_s \cos(\omega_0 t)$. This square wave can be represented by the Fourier series:

$$\text{SQWV}(t) = \sum_{n=1}^{\infty} a_n \cos(n\omega_0 t), \quad (2.182)$$

where

$$\omega_0 = 2\pi/T$$

$$a_n = \left(\frac{2}{T}\right) \int_{-T/2}^{T/2} f(t) \cos(n\omega_0 t) dt = \left(\frac{4}{n\pi}\right) \sin\left(\frac{n\pi}{2}\right) \text{ (nonzero for } n \text{ odd)}. \quad (2.183)$$

For a unit cosine (even) square wave that can be written in expanded form:

$$\text{SQWV}(t) = \left(\frac{4}{\pi}\right) \left[\cos(\omega_0 t) - \left(\frac{1}{3}\right) \cos(3\omega_0 t) + \left(\frac{1}{5}\right) \cos(5\omega_0 t) - \left(\frac{1}{7}\right) \cos(7\omega_0 t) + \dots \right]. \quad (2.184)$$

The input Gaussian noise is multiplied by SQWV(t):

$$y_n(t) = n(t) \times \left(\frac{4}{\pi}\right) \left[\cos(\omega_0 t) - \left(\frac{1}{3}\right) \cos(3\omega_0 t) + \left(\frac{1}{5}\right) \cos(5\omega_0 t) - \left(\frac{1}{7}\right) \cos(7\omega_0 t) + \dots \right]. \quad (2.185)$$

The white, one-sided noise PDS before passing through the LIA's output LPF can be shown to be

$$S_n(f) = \eta \left(\frac{4}{\pi}\right)^2 \left[1 - \left(\frac{1}{3}\right) + \left(\frac{1}{5}\right) - \left(\frac{1}{7}\right) + \left(\frac{1}{9}\right) - \left(\frac{1}{11}\right) + \left(\frac{1}{13}\right) - \dots \right]^2 = \eta \text{ msV/Hz}. \quad (2.186)$$

This white noise is conditioned by the output LPF, so the output MS SNR is

$$\text{SNR}_{\text{out}} = \frac{V_s^2 (2/\pi)^2}{\eta B_{\text{out}}}. \quad (2.187)$$

The MS SNR at the input to the LIA is

$$\text{SNR}_{\text{in}} = \frac{V_s^2/2}{\eta B_{\text{in}}}. \quad (2.188)$$

Thus, F for this LIA architecture is found to be

$$F = \frac{\text{SNR}_{\text{in}}}{\text{SNR}_{\text{out}}} = \frac{\pi}{2^3} \frac{B_{\text{out}}}{B_{\text{in}}} = 1.23 \frac{B_{\text{out}}}{B_{\text{in}}}, \quad (2.189)$$

which can be made quite small by adjusting the ratio of B_{out} to B_{in} .

As an example of signal recovery by an LIA, consider an LIA of the analog multiplier architecture. Let the signal to be measured be $v_s(t) = 20 \times 10^{-9} \cos(2\pi 50 \times 10^3 t)$. The Gaussian white noise present with the signal has a root power spectrum of $\sqrt{\eta} = 4n \text{ V RMS}/\sqrt{\text{Hz}}$.

The input amplifier has a voltage gain of $K_v = 10^4$ with a 100 kHz noise and signal bandwidth. Thus, the raw input MS SNR is

$$\text{SNR}_{\text{in1}} = \frac{(20 \times 10^{-9})^2 / 2}{(4 \times 10^{-9})^2 \times 10^5} = \frac{2 \times 10^{-16} \text{ msV}}{1.6 \times 10^{-12} \text{ msV}} = 1.25 \times 10^{-4}, \text{ or } -39 \text{ dB.} \quad (2.190)$$

If the sinusoidal signal is conditioned by an (ideal) unity-gain LPF with a $Q = 50 = \text{center freq}/\text{bandwidth}$, the noise bandwidth $B = 50 \times 10^3 / 50 = 1 \text{ kHz}$. The MS SNR at the output of the ideal BPF is still poor:

$$\text{SNR}_{\text{filt}} = \frac{2 \times 10^{-16} \text{ msV}}{(4 \times 10^{-9})^2 \times 10^3} = 1.25 \times 10^{-2}, \text{ or } -19 \text{ dB.} \quad (2.191)$$

Now the signal and noise are amplified and passed directly into the LIA, and the LIA's output LPF has a noise bandwidth of 0.125 Hz. The output DC component of the signal is $K_v V_s / 2 \text{ V}$. Thus, the MS DC output signal is

$$\overline{v_{\text{so}}^2} = \frac{K_v^2 V_s^2}{4} = K_v^2 \times 1 \times 10^{-16} \text{ msV.}$$

The MS noise output is

$$\overline{v_{\text{no}}^2} = K_v^2 (4 \times 10^{-9})^2 \times 0.125 \text{ msV.}$$

Thus, the MS $\text{SNR}_{\text{out}} = 50$, or +17 dB.

The costs we pay for using an LIA are that the AC input signal is reduced to a proportional DC level at the LIA output and the output LPF means that the LIA does not reach a new SS output, given a step change in V_s , until about 3 time constants of the LPF (about 3 s in the preceding example). The benefit of using an LIA is that coherent signals buried in up to 60 dB noise can be measured.

Many applications of LIAs lie in the area of electro-optical or photonics instrumentation. In these systems, the DSBSCM signal is derived by optically chopping a light beam. The chopper wheel generates a square wave, 0,1 carrier, rather than a sinusoidal carrier. An LED and photodiode on the chopper can generate the reference signal, also a square wave. The LIA's PSR demodulates square wave DSBSCM signals by the same means that is used with a sinusoidal carrier.

Commercially available LIAs are multipurpose instruments and, as such, are expensive. For a dedicated LIA application, one can assemble the basic LIA system components from low-noise OAs and a commercial PSR IC, such as the AD630 or LM1596. The reference signal is converted to a TTL wave by a zero-crossing comparator. The TTL reference phase can be adjusted in a carrier frequency-independent manner by using a phase-lock loop as a phase shifter (Northrop 1990). Such a dedicated LIA can be built for less than about US\$150, exclusive of power supplies.

2.8.5.3 Summary

The PSR is a circuit widely used in instrumentation systems to demodulate DSBSC signals. We have presented a description of the DSBSCM signal and how it is created and have shown two simple circuits that can be used to demodulate it. We have also mentioned a few of the commercially available ICs suitable for PSR use.

Probably the most important instrument based on the PSR is the LIA, which can be used to recover the signals from DSBSM carriers severely contaminated with noise, even noise in the signal passband. We algebraically illustrate the LIA demodulation process and show the subsystems of an LIA.

2.8.6 Signal Averaging to Improve SNR of Evoked Transient Signals

2.8.6.1 Introduction

As we have seen, the SNR of a signal recorded with additive noise is an important figure of merit, which characterizes the expected resolution of the signal. SNRs are typically given at the input to a signal-conditioning system and also at its output. SNR can be expressed as a positive, real number or in decibels (dB). The SNR can be calculated from the RMS (or peak) signal voltage divided by the RMS noise voltage or noise standard deviation. The noise bandwidth should be specified. The SNR can also be formed from the mean-squared signal voltage divided by the mean-squared noise voltage (noise variance). If the MS SNR is computed, the $\text{SNR}_{(\text{dB})} = 10 \log_{10}(\text{MS SNR})$; otherwise, it is $\text{SNR}(\text{dB}) = 20 \log_{10}(\text{RMS SNR})$.

Signal averaging is widely used in both experimental and clinical electrophysiology and neurology in order to extract a repetitive, evoked, quasi-deterministic, electrophysiological transient response buried in noise. One example of the type of signal being extracted is the evoked cortical response, which can be recorded directly from the surface of the brain (or from the scalp) by an electrode pair, while the subject is given a repetitive, periodic, sensory stimulus, such as a tone burst or flash of light. Every time the stimulus is given, a *hardwired*, electrophysiological transient voltage, $s_j(t)$, lasting several hundred ms is produced in the brain. When viewed directly on an oscilloscope or recorder, each individual evoked cortical response can be invisible to the eye because of the accompanying noise. Signal averaging can also be used to characterize low SNR, sonar targets; every time the target is *pinged*, the return is averaged.

Signal averaging is generally used to extract the evoked potential, $s(t)$, from the accompanying noise. Signal averaging is also used to extract evoked cortical magnetic field transients recorded with superconducting quantum interference device (SQUID) sensors (Northrop 2002) and to extract multifocal electroretinogram (ERG) signals obtained when testing the competence of macular cones in the retina of the eye. A small spot of light illuminating only a few cones is repetitively flashed on the macular retina. ERG averaging is done over N flashes for a given spot position in order to extract the local ERG flash response; then the spot is moved to a new, known, position on the macula; and the process is repeated until the 2-D, macular, ERG response of one eye is mapped (Northrop 2002).

Signal averaging is *ensemble averaging*; following each identical, periodic stimulus, the response can be written for the j th stimulus:

$$x_j(t) = s_j(t) + n_j(t), \quad 0 \leq j \leq N, \quad (2.192)$$

where

$s_j(t)$ is the j th evoked transient response

$n_j(t)$ is the j th noise following the stimulus

t is the local time origin taken as 0 when the j th stimulus is given

The noise is assumed to be, in general, nonstationary, that is, its statistics are affected by the stimulus. We assume that the noise has zero mean, however, regardless of time following any stimulus. That is, $E\{n(t)\} = 0, 0 \leq t < T_i$. T_i is the interstimulus interval. Also, to be general, we assume that the evoked response varies from stimulus to stimulus. That is, $s_j(t)$ is not exactly the same as $s_{j+1}(t)$.

We assume that each $x_j(t)$ is sampled and digitized beginning with each stimulus; the sampling period is T_s , and M samples are taken following each stimulus. Thus, there are N sets of sampled $x_j(k)$, $0 \leq k \leq (M - 1)$; also, $(M - 1)T_s = T_D < T_i$. T_D is the total length of analog $x_j(t)$ digitized following each input stimulus (data epoch length).

When the j th stimulus is given, the $x(k)$ th value is summed in the k th data register with the preceding $x(k)$ values. At the end of an experimental run, we have $[x_1(k) + x_2(k) + x_3(k) + \dots + x_N(k)]$ in the k th data register. Figure 2.59 illustrates the organization of a signal averaging system. The two analog circuits enclosed in dotted lines are generally on a plug-in card. The averaging controller and register memory either are on a dedicated plug-in card or are performed by computer CPU and RAM. Early signal averagers were stand-alone, dedicated instruments. Modern signal averaging typically uses a PC or laptop computer with a dual-channel A/D interface card to handle the trigger event that

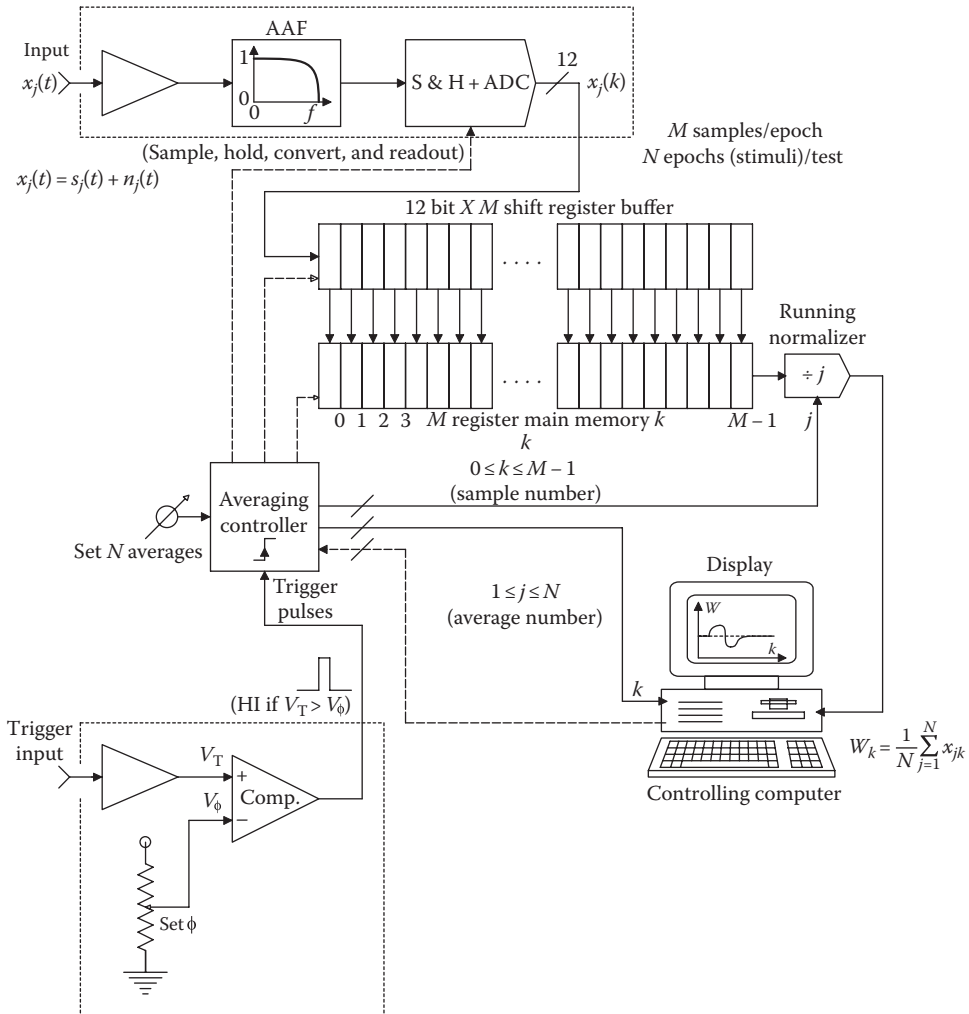


FIGURE 2.59 Organizational block diagram of a synchronous signal averager. The trigger input and signal digitizer (inside dotted lines) are typically on an accessory card; the registers and averaging controller typically exist as software operations inside the computer.

initiates sampling and data conversion of the evoked transient plus noise, $x_j(t) = s_j(t) + n_j(t)$. Modern averagers give a running display in which the main register contents are continually divided by the running j value as j goes from 1 to N .

2.8.6.2 Analysis of SNR Improvement by Averaging

The average contents of the k th register after N epochs are sampled can be written formally:

$$\overline{x(k)}_N = \frac{1}{N} \sum_{j=1}^N s_j(k) + \frac{1}{N} \sum_{j=1}^N n_j(k), \quad 0 \leq k \leq (M-1), \quad (2.193)$$

where the left-hand summation is the *signal sample mean* at $t = kT_s$ after N stimuli and the right-hand summation is the *noise sample mean* at $t = kT_s$ after N stimuli.

We have previously shown that the *variance of the sample mean* is a statistical measure of its noisiness. In general, the larger the N is, the smaller the noise variance will be. The variance of the sample mean of $x(k)$ is written as

$$\begin{aligned} \text{Var}\{\overline{x(k)}_N\} &= \mathbf{E}\left\{\left[\frac{1}{N} \sum_{j=1}^N x_j(k)\right]^2\right\} - \langle x(k) \rangle^2, \\ &\downarrow \\ \text{Var}\{\overline{x(k)}_N\} &= \mathbf{E}\left\{\left[\frac{1}{N} \sum_{j=1}^N x_j(k)\right]\left[\frac{1}{N} \sum_{i=1}^N x_i(k)\right]\right\} - \langle x(k) \rangle^2, \\ &\downarrow \\ \text{Var}\{\overline{x(k)}_N\} &= \frac{1}{N^2} \sum_{j=1}^N \sum_{i=1}^N \mathbf{E}\{x_j(k)x_i(k)\} - \langle x(k) \rangle^2, \\ &\downarrow \\ &\quad (N \text{ terms, } j=i) \quad (N^2 - N \text{ terms}) \\ \text{Var}\{\overline{x(k)}_N\} &= \frac{1}{N^2} \sum_{j=1}^N \mathbf{E}\{x_j^2(k)\} - \frac{1}{N^2} \sum_{j=1}^N \sum_{\substack{i=1 \\ j \neq i}}^N \mathbf{E}\{x_j(k)x_i(k)\} - \langle x(k) \rangle^2 \quad (2.194) \end{aligned}$$

Now for the N squared terms,

$$\begin{aligned} \mathbf{E}\{x_j^2(k)\} &= \mathbf{E}\{[s_j(k) + n_j(k)]^2\} = \mathbf{E}\{s_j^2(k)\} + 2\mathbf{E}\{s_j(k)\}\mathbf{E}\{n_j(k)\} + \mathbf{E}\{n_j^2(k)\}, \\ &= \sigma_s^2(k) + s(k)^2 + \sigma_n^2(k). \end{aligned} \quad (2.195)$$

For the $(N^2 - N)$ terms with unlike indices (i.e., $i \neq j$),

$$\begin{aligned} \mathbf{E}\{x_j(k)x_i(k)\} &= \mathbf{E}\{[s_j(k) + n_j(k)][s_i(k) + n_i(k)]\} \\ &= \mathbf{E}\{s_j(k)s_i(k)\} + \mathbf{E}\{n_j(k)n_i(k)\} + \mathbf{E}\{s_j(k)n_i(k) + s_i(k)n_j(k)\}. \end{aligned} \quad (2.196)$$

Several important assumptions are generally applied to Equation 2.196: First, noise and signal are uncorrelated and statistically independent. This means that $\mathbf{E}\{s, n\} = \mathbf{E}\{s\}\mathbf{E}\{n\} = \mathbf{E}\{s\} \times 0 = 0$. Also, we assume that noise samples taken at or more than $t = T$ seconds apart will be uncorrelated; these assumptions lead to $\mathbf{E}\{n_j(k) n_i(k)\} = \mathbf{E}\{n_j(k)\} \mathbf{E}\{n_i(k)\} \rightarrow 0$. So, $\mathbf{E}\{x_j(k) x_i(k)\} = \mathbf{E}\{s_j(k) s_i(k)\}$. We also assume that $s_j(k)$ and $s_i(k)$ taken at or more than T seconds apart are independent. So finally we can write

$$\mathbf{E}\{x_j(k) x_i(k)\} = \mathbf{E}\{s_j(k) s_i(k)\} = \mathbf{E}\{s_j(k)\} \mathbf{E}\{s_i(k)\} = \overline{s(k)}^2. \quad (2.197)$$

Now, putting all the terms together,

$$\text{Var}\{\overline{x(k)}_N\} = \frac{1}{N^2} \left\{ N \left[\sigma_s^2(k) + \overline{s(k)}^2 + \sigma_n^2(k) \right] + (N^2 - N) \overline{s(k)}^2 \right\} - \overline{s(k)}^2, \quad (2.198)$$

↓

$$\text{Var}\{\overline{x(k)}_N\} = \frac{\sigma_s^2(k) + \sigma_n^2(k)}{N}. \quad (2.199)$$

The variance of the sample mean for the k th data sample following a stimulus is a measure of the noisiness of the averaging process. The variance of the averaged signal, x , is seen to decrease as $1/N$, where N is the total number of stimuli given and of the responses averaged.

Another measure of the effectiveness of signal averaging is the *noise factor*, $F \equiv (S_{\text{in}}/N_{\text{in}})/(S_o/N_o)$, where S_{in} is the *mean-squared input signal* to the averaging process, N_{in} is the *MS input noise*, S_o is the *MS output signal*, and N_o is the *MS output noise*. The noise factor is normally used as a figure of merit for amplifiers; amplifiers generally add noise to the input signal and noise, so the output SNR (SNR_o) is less than the input SNR_{*i*}, so for a nonideal amplifier, $F > 1$. For an ideal, noiseless amplifier, $F = 1$. The exception to this behavior is in signal averaging, where the averaging process generally produces an output SNR_{*o*} > the input SNR_{*i*}, making $F < 1$. Note that the *noise figure* of a signal-conditioning system is defined as

$$\text{NF} \equiv 10 \log_{10}(F) \text{ dB}. \quad (2.200)$$

From the previous calculations on the averaging process,

$$S_{\text{in}}(k) = \mathbf{E}\{s_j^2(k)\} = \sigma_s^2(k) + \overline{s(k)}^2, \quad (2.201)$$

$$N_{\text{in}}(k) = \mathbf{E}\{n_j^2(k)\} = \sigma_n^2(k), \quad (2.202)$$

$$S_o(k) = \mathbf{E}\{s_o^2(k)\} = \frac{\sigma_s^2(k)}{N} + \overline{s(k)}^2, \quad (2.203)$$

$$N_o(k) = \frac{\sigma_n^2(k) + \sigma_s^2(k)}{N}. \quad (2.204)$$

These terms can be put together to calculate the noise factor of the k th sample in the averaging process:

$$F(k) = \frac{\left[\sigma_s^2(k) + \overline{s(k)^2} \right] \left[1 + \sigma_s^2(k) / \sigma_n^2(k) \right]}{\left[\sigma_s^2(k) + N \overline{s(k)^2} \right]}. \quad (2.205)$$

Note that if the evoked transient is exactly the same for each stimulus, $\sigma_s^2(k) \rightarrow 0$, and $F = 1/N$.

The reader should appreciate that this is an idealized situation; in practice, a constant level of noise, σ_a^2 , is present on the output of the signal averager. This noise comes from the signal-conditioning amplifiers, the quantization accompanying ADC, and arithmetic roundoff. The averager MS output noise can thus be written:

$$N_o = \frac{\sigma_n^2(k) + \sigma_s^2(k)}{N} + \sigma_a^2 \text{ mean-squared volts.} \quad (2.206)$$

And as before, the MS signal is

$$S_o(k) = \frac{\sigma_s^2(k)}{N} + \overline{s(k)^2}. \quad (2.207)$$

Thus, the averager MS output SNR_o is

$$\text{SNR}_o(k) = \frac{\sigma_s^2(k) + N \overline{s(k)^2}}{\sigma_s^2(k) + \sigma_n^2(k) + N \sigma_a^2}. \quad (2.208)$$

And if the evoked response is deterministic, $\sigma_s^2(k) \rightarrow 0$, and we have

$$\text{SNR}_o(k) = \frac{N \overline{s(k)^2}}{\sigma_n^2(k) + N \sigma_a^2}. \quad (2.209)$$

Note that if the number, N , of stimuli and responses averaged becomes very large, then

$$\text{SNR}_o(k) \rightarrow \frac{\overline{s(k)^2}}{\sigma_a^2}, \quad (2.210)$$

and also in the limit,

$$F(k) \rightarrow \frac{\sigma_a^2}{\sigma_n^2(k)}. \quad (2.211)$$

Signal averaging can recover a good estimate of $s(k)$ even when $\sigma_n(k)$ is 60 dB larger (10^3 times) than $s(k)$.

From Equation 2.209, we can find the N required to give a specified SNR_o, given $\sigma_n^2(k)$, $\sigma_a^2(k)$, and $\overline{s(k)^2}$.

2.9 Chapter Summary

Chapter 2 has dealt with the analysis and design of analog signal-conditioning circuits. Our treatment has been subdivided to include both linear and nonlinear signal processing systems. IC designs are stressed, and examples of commercially available ICs are given. This chapter was opened in Section 2.2 with a basic treatment of DAs that are widely used in the front ends of OAs and IAs to reduce CM interference.

Section 2.3 deals with the ubiquitous operational amplifier amp that finds many applications in linear and nonlinear signal processing. Basic design types of OAs are reviewed, and examples of their applications in linear signal processing are discussed. In Section 2.4, we review basic quadratic analog AF designs using conventional OAs. Although many, many AF designs exist, we have focused on three of the more popular and easy to use designs: controlled-source AFs (including the well-known Sallen and Key designs), biquad AFs, and GIC-based AF designs.

The design, characteristics, and uses of IAs, including IsoAs and autozero amplifiers, are described in Section 2.5.

Section 2.6 deals with nonlinear analog signal processing. In it, we cite examples of commercially available function modules, and nonlinear circuits that can be built with OAs are presented. Considered are precision absval circuits, multifunction converters, true RMS converters, square-root and divider circuits (other than multifunction converters), peak detectors, T&H circuits, log ratio circuits, and a generalized trigonometric system.

Two examples of the use of the OA charge amplifier configuration are given in Section 2.7: conditioning the output of piezoelectric crystal pressure transducers and measuring charge on a capacitance by integration.

The demodulation of DSBSC signals generated by such systems as Wheatstone bridges with AC excitation and LVDTs is considered in Section 2.8 on PSRs. This section also introduces the LIA, its uses, and signal detection limits set by noise.

Home problems at the end of this chapter illustrate the design of amplifiers, filters, and signal-conditioning systems. Use should be made of computer-aided circuit analysis in reaching problem solutions.

The following Chapter 3 treats the all-important topics of noise and coherent interference arising in the environment, sensors, analog signal-conditioning circuits, and digital interfaces and DSP functions. Noise is shown to be an important limitation to I&M system resolution, precision, and accuracy. Low-noise signal-conditioning systems are described.

Problems

- 2.1 An electrometer charge amplifier is used to find the inductance of a coil. The coil's DC resistance is measured to be $R_L = 10 \Omega$. An SS DC of 10 mA is passed through the coil. At $t = 0$, the switch connects the coil to the OA's summing junction. C_F is initially uncharged so $v_o(0) = 0$ V. See Figure P2.1.
- Sketch and dimension $i_L(t)$ for $t \geq 0$.
 - Sketch and dimension $v_o(t)$ for $t \geq 0$. Give an expression for L in terms of the SS v_o .

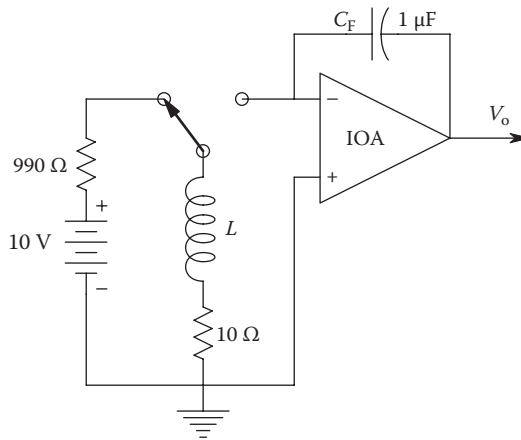


FIGURE P2.1

- 2.2 A. Derive the transfer function of the *controlled-source* AF shown in Figure P2.2 in time-constant form.
 B. Give expressions for the peak gain, the frequency of the peak gain, and its Q . Assume the OA is ideal.
- 2.3 See Figure P2.3. *Design a 60 Hz notch filter* using a two-loop biquad BPF and a fifth OA summer. The passband gain of the notch filter is to be +20 dB, and the BPF's Q is to be 50. The high -3 dB frequency of the filter is to be above 100 kHz. Specify the OAs and all R s and C s in your design. Verify your design by a *SPICE* or *MicroCap* simulation. What is the simulated filter's attenuation at 60 Hz?
- 2.4 We wish to pass a constant current through a SnO_2 carbon monoxide sensor. The presence of CO causes a drop in the sensor's resistance, hence a drop in V_2 . A Howland current source is to be used, shown in Figure P2.4. Define the following: $\alpha \equiv R_2/(R_1 + R_2)$, $\beta \equiv R_1/(R_1 + R_2)$, $\gamma \equiv R_3/(R_3 + R_4)$. Assume an IOA. Using circuit analysis, find the conditions for α , β , and γ and the various resistors that make I_L completely independent of the sensor's resistance, R_s . Give an expression for the circuit's transconductance, $G_M = I_L/V_1$.

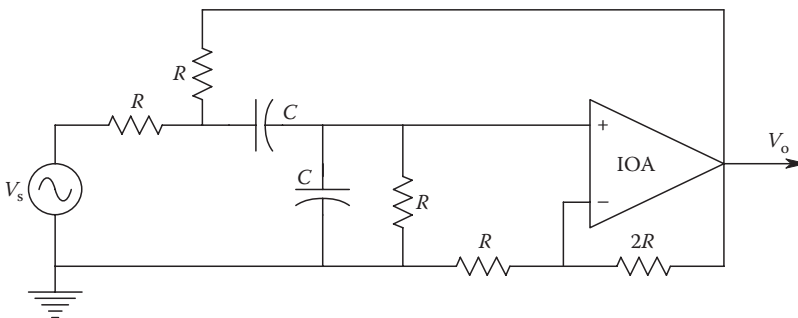


FIGURE P2.2

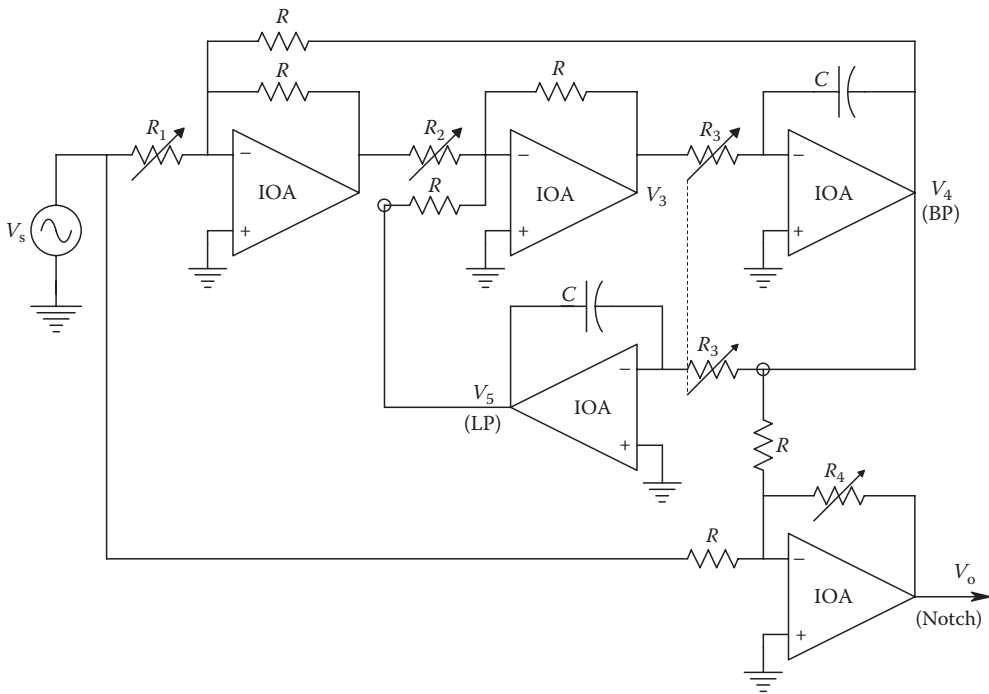


FIGURE P2.3

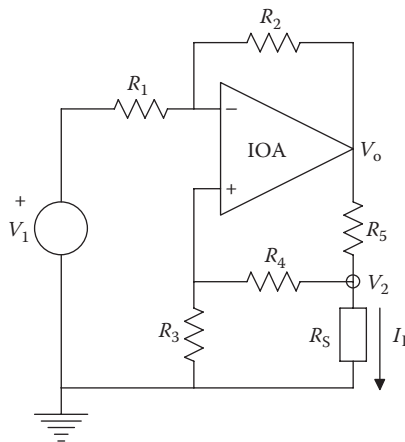


FIGURE P2.4

2.5 The circuit of Figure P2.5 is a Deboo noninverting integrator. Assume the OA is ideal and the resistors are exactly matched, that is, $R_1 = R'_1$, $R_2 = R'_2$.

- A. Derive an expression for $\frac{V_o}{V_s}(s)$.
- B. Now let $V_s = 0$. Let a DC bias current, I_B , enter the v_i node and a DC bias current, I'_B , enter the v'_i node. Find an expression for $V_o(t)$, assuming the bias currents are applied as steps at $t = 0$ and $v_i(0) = 0$.

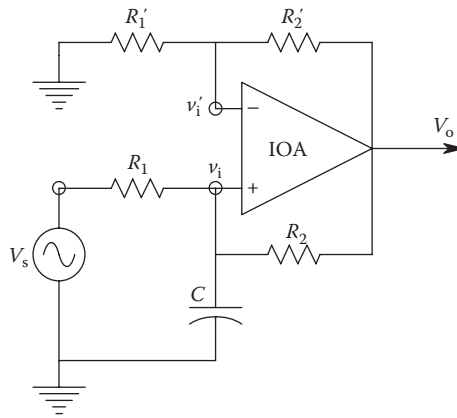


FIGURE P2.5

- 2.6 Two IOAs and a single diode are used to perform a nonlinear operation on the input signal. See the circuit of Figure P2.6. Find an expression for $V_o = f(V_s, R, R')$.
- 2.7 The circuit of Figure P2.7 is another *notch filter* design.
- Design the filter to have its notch at 60 Hz. That is, find numerical values for R_1, R_2, R_3 , and C . Give the DC gain of the filter and the damping factor of its denominator.
 - Simulate the filter's frequency response; use OP27 OA models in *SPICE* or *MicroCap*. From the Bode magnitude plot, find the notch's $Q = -3$ dB bandwidth/60.
- 2.8 The circuit in Figure P2.8 is a single-loop biquad AF. Assume the OAs are ideal.
- Derive an expression for $\frac{V_o}{V_s}(s)$ in time-constant form. Sketch and dimension the Bode magnitude plot for this transfer function.
 - Give expressions for this filter's ω_n, Q , and gain at $\omega = \omega_n$. Does this version of the biquad AF have the design flexibility of the 4-OA, two-loop biquad AF?

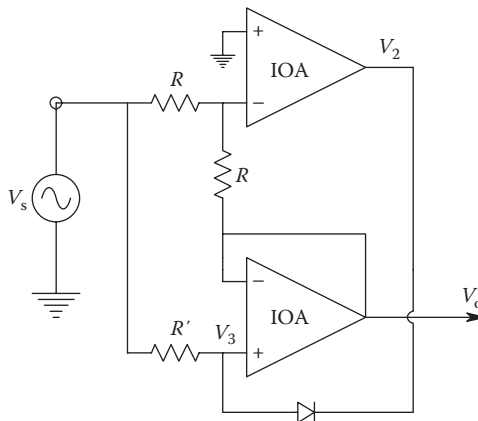


FIGURE P2.6

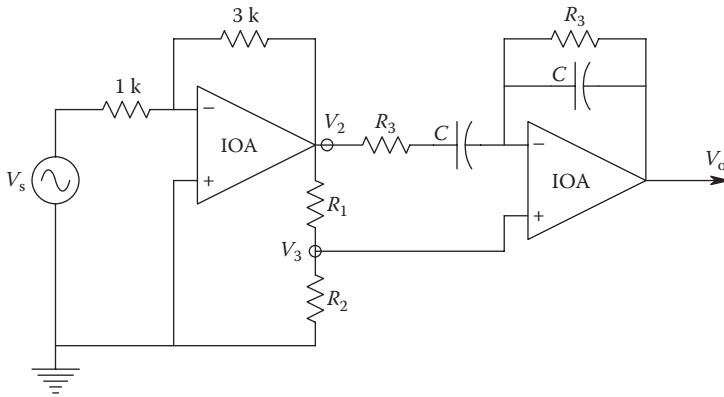


FIGURE P2.7

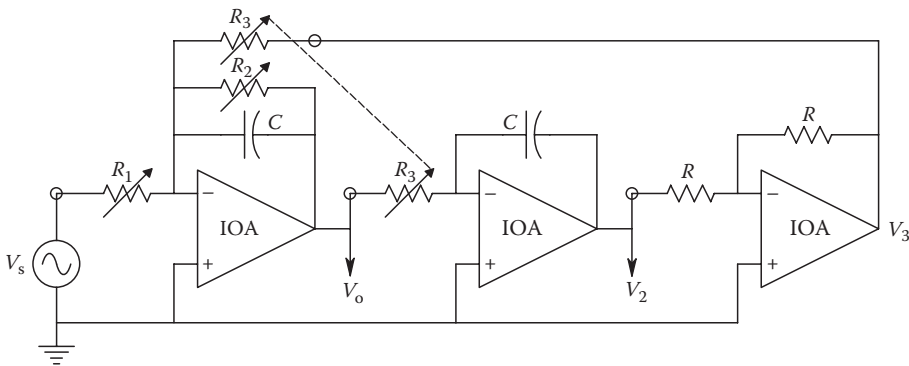


FIGURE P2.8

- 2.9 Find expressions for the gains of the seven IOA circuits shown in Figure P2.9.
- 2.10 The circuit of Figure P2.10 uses two OAs to linearize a Wheatstone bridge. Note that voltage V_3 hence V_2 are at virtual ground. Write node equations for V_2 and V_3 and use them to find $V_o = f(\Delta R, R, R_2, R_3, V_B)$.
- 2.11 The CMRR of a certain DA is 120 dB at 60 Hz. When a source, $v_s(t) = 0.004 \sin(377t)$, is connected to the + input and the - input is grounded, the output voltage is 1.0 V RMS. Find numerical values for A_D and A_C in the gain equation: $v_o = A_D v_{1d} + A_C v_{1c}$.
- 2.12 The input voltages to a differential IA are $v_1(t) = 0.002 \sin(2\pi 400t)$ and $v'_1(t) = -0.002 \sin(2\pi 400t) + 2 \sin(377t)$. The amplifier's output is given by $v_o = A_D v_{1d} + A_C v_{1c}$, where $A_D = 10^3$ and $A_C = 10^{-2}$.
 - A. What is the amplifier's CMRR in dB?
 - B. Write an expression for $v_o(t)$.

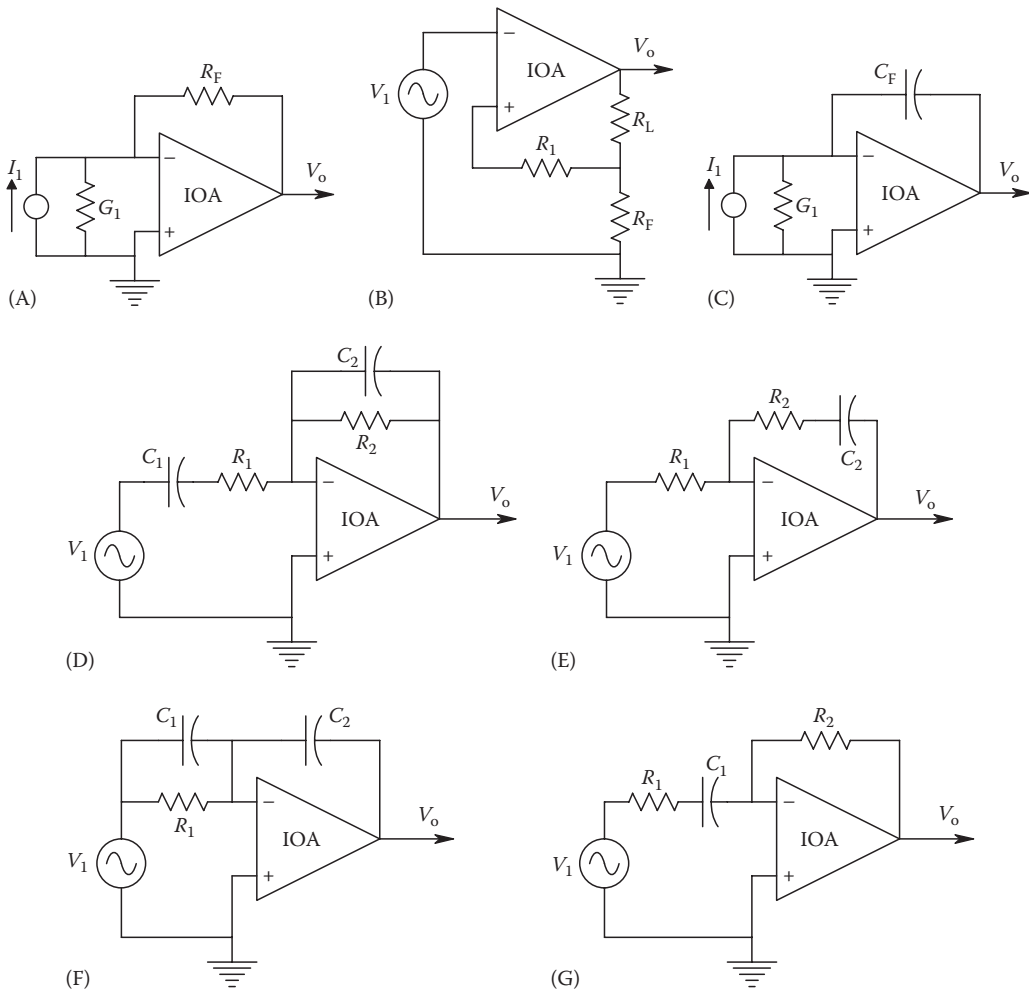


FIGURE P2.9

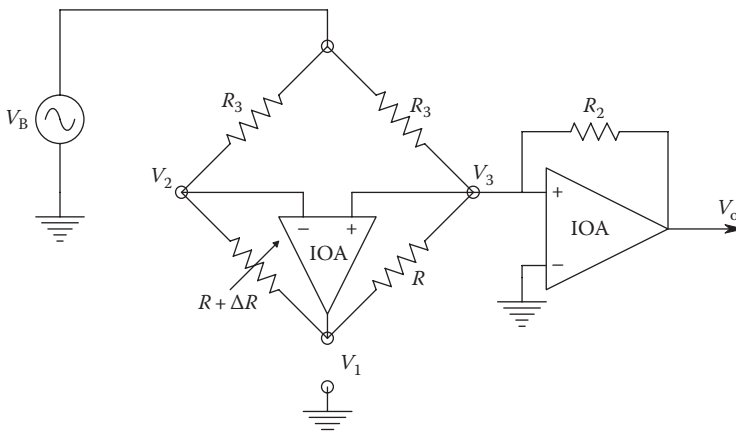


FIGURE P2.10

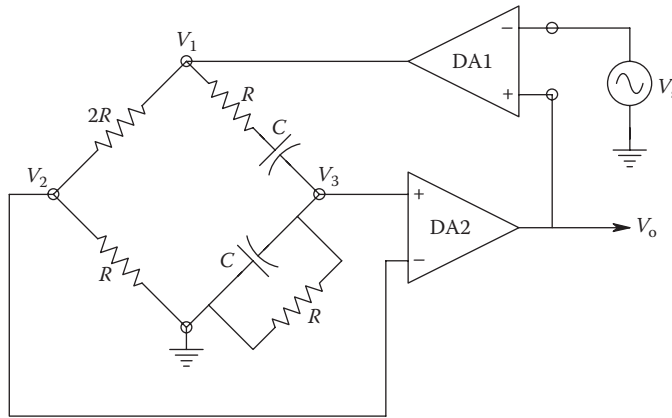


FIGURE P2.13

2.13 A Wien bridge is used to make a notch filter at 60 Hz. In Figure P2.13, DA1 has $A_D = K_D$, $A_C = 0$, and DA2 has $A_D = 1$, $A_C = 0$. (The DAs are not OAs.)

- A. Find an expression for the frequency response of the filter, $\frac{V_o}{V_s}(j\omega)$, in time-constant form.
- B. Pick R and C so the notch frequency is exactly 60 Hz.
- C. Give the filter's DC gain, A_{v0} , ω_n , and the Q of the denominator (note that the Q of an underdamped quadratic pole pair is $1/2\xi$).

2.14 Design a quadratic, GIC BPF for geophysical applications to condition a seismometer output having $f_n = 1$ Hz and $Q = 10$. Use low-noise OAs. Verify your design with a MicroCap or PSPICE simulation. Note all R s must be between 10^3 and $10^6 \Omega$ and capacitors must lie between 3 pF and 1 μ F.

2.15 Figure P2.15 illustrates a voltage-tuned AF. An analog multiplier is used as a variable gain element.

- A. Derive an expression for $\frac{V_o}{V_1}(j\omega, V_c)$ in time-constant form.
- B. Make a dimensioned Bode plot for the filter for $V_c = 0.1, 1.0,$ and 10 V. Give expressions for the DC gain and break frequency.

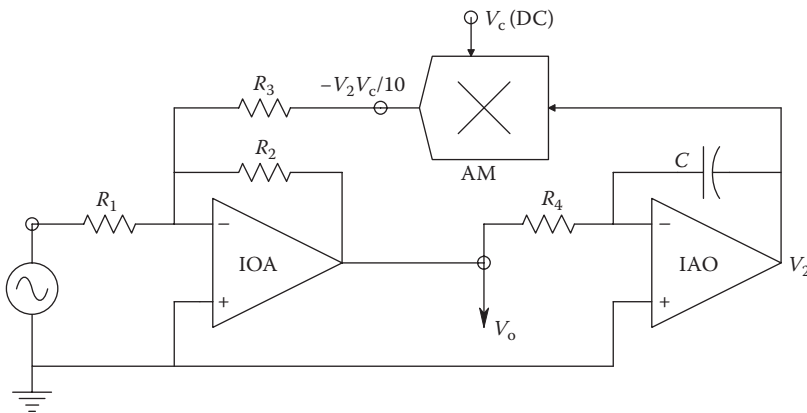


FIGURE P2.15

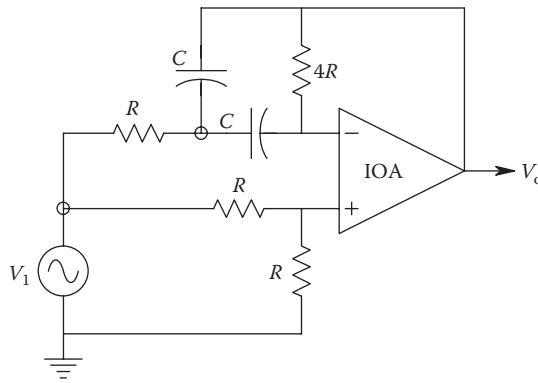


FIGURE P2.16

- 2.16 Derive an expression for the frequency response function of the AF shown in Figure P2.16 in time-constant form. Give expressions for the filter's ω_n , DC gain, and damping factors of the numerator and denominator. What kind of filter is this?
- 2.17 Find an algebraic expression for $V_o = f(V_1, V_2)$ in the nonlinear analog circuit, Figure P2.17. Assume IOAs. Why is the inverter IOA2 required?
- 2.18 A comparator and an OA are used to create a symmetrical hysteresis voltage transfer function, as shown in Figure P2.18. An ideal comparator and an IOA are shown in the circuit. The ideal comparator's transfer characteristic is shown. Assume R_1 and $R_2 \gg R_3$.

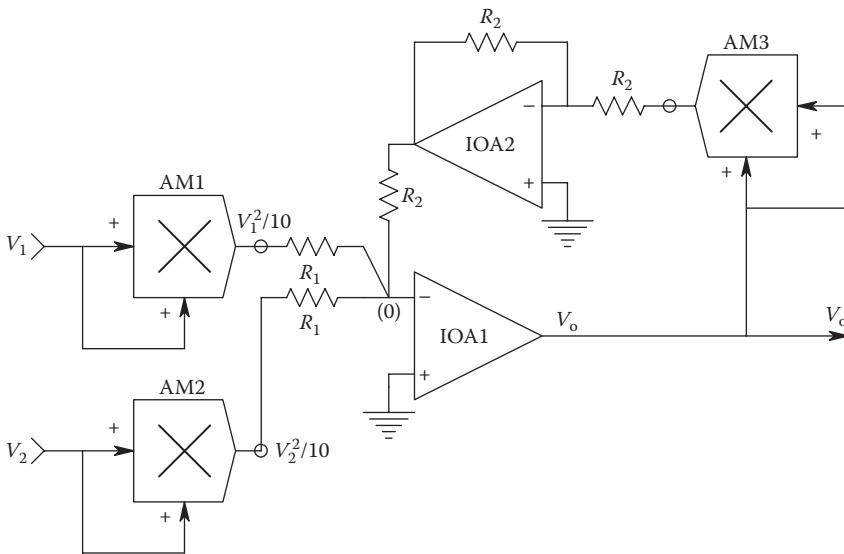


FIGURE P2.17

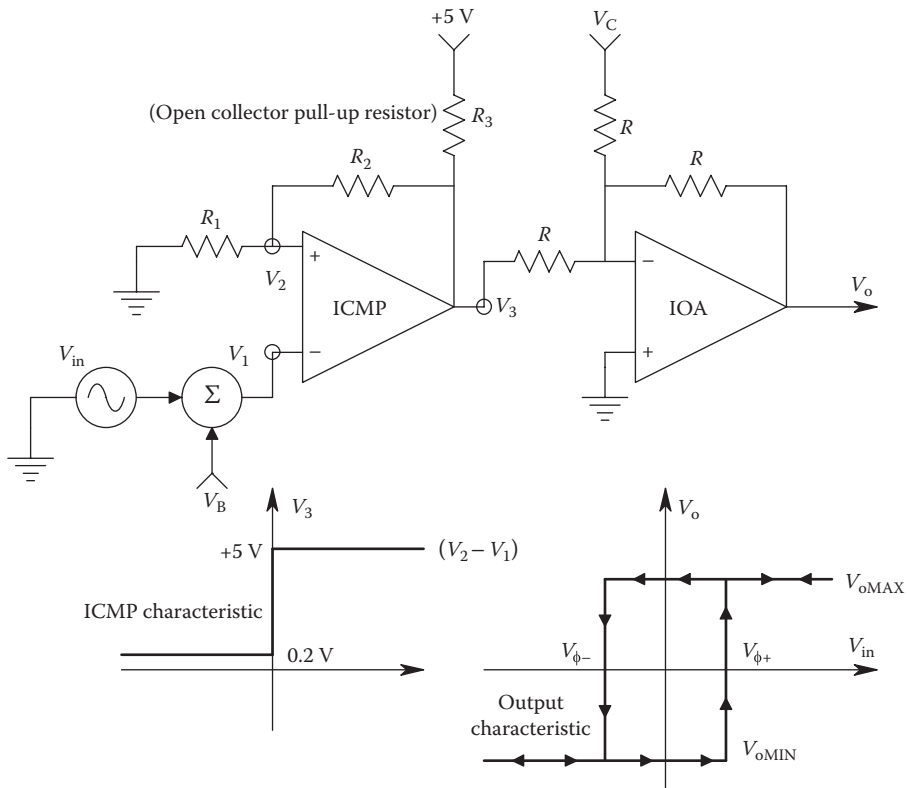


FIGURE P2.18

- Assume $V_B = 0$. Plot and dimension $V_3 = f(V_{in}, R_1, R_2)$ for the ICMP. Note that it has hysteresis. Let $R_1 = R_2 = 100 \text{ k}\Omega$.
- Plot and dimension $V_o = f(V_{in}, R_1, R_2)$. Let $V_B = V_C = 0$, $R = 10 \text{ k}\Omega$.
- The symmetrical output hysteresis characteristic is shown in Figure P2.18. Find numerical values of V_B and V_C required to give symmetry. Give numerical values for $V_{\phi+}$, $V_{\phi-}$, V_{oMAX} , and V_{oMIN} .

3

Noise and Coherent Interference in Measurements

3.1 Introduction

Because of the differences in the sources and the means of reduction of random *noise* vs. *coherent interference*, we will treat them separately in subsections of this chapter. Both noise and interference provide a major limitation to the precision of measurements and the threshold detectability of the QUM. Thus, a knowledge of how to design low-noise instrumentation systems is an important skill that a practicing instrumentation engineer should learn.

First, let us examine the distinction between noise and coherent interference. Noise is considered to arise in a circuit or measurement system from completely random phenomena. Any physical quantity can be *noisy*, but we will generally restrict ourselves to the consideration of noise voltages and currents in this chapter. Such noise voltages and currents will be considered to have zero means, meaning that they have no additive DC components. The unwanted DC components are best considered to be drift or offset, and their reductions are more appropriately treated in a text on DC amplifier design or OA applications.

Coherent interference, as the name suggests, generally has its origins in periodic, man-made phenomena, such as powerline frequency electric and magnetic fields, radio-frequency sources such as radio and television station broadcast antennas, certain poorly shielded computer equipment, spark discharge phenomena such as automotive ignitions and motor brushes and commutators, and inductive switching transients such as SCR motor speed controls. As you will see, minimization of coherent interference is often *arty* and may involve changing the grounding circuits for a system, shielding with magnetic and/or electric conducting materials, filtering, using isolation transformers and ferrite beads, etc.

Minimizing the impact of incoherent noise in a measurement system often involves a prudent choice of low-noise components, certain basic electronic design principles, and filtering. The incoherent noise that we will consider is random noise from within the measurement system; coherent noise usually enters a system from without.

3.2 Descriptions of Random Noise in Circuits

In the following discussions, we assume that the noise is *stationary*. That is, the physical phenomena giving rise to the noise are assumed not to change with time. When stationarity is assumed for a noise, averages over time are equivalent to ensemble averages.

An example of a *nonstationary noise* source is a resistor, which, at time $t = 0$, begins to dissipate power so that its temperature slowly rises, as does its resistance, hence its thermal noise voltage.

Several statistical methods for describing random noise exist. These include, but are not limited to, the *PDF*, the *cross-* and *autocorrelation functions* and their Fourier transforms (FTs), the *cross-* and *auto-power density spectra*, and the *root PDS*, used to characterize amplifier equivalent input noise.

3.2.1 Probability Density Functions

The PDF is a mathematical model based on the amplitude statistics of a stationary noise waveform, $n(t)$, and not how $n(t)$ varies in time. For a random sample of $n(t)$, the PDF of $n(t)$ is defined as

$$p(x) \equiv \frac{\text{Probability that } x < n \leq (x + dx)}{dx}, \quad (3.1)$$

where

x is a specific value of n taken at some time t

dx is a differential increment in x

The PDF is the mathematical basis for many formal derivations and proofs in statistics and probability theory. The PDF has the following properties:

$$\int_{-\infty}^v p(x) dx = \text{Prob}[x \leq v], \quad (3.2)$$

$$\int_{v_2}^{v_1} p(x) dx = \text{Prob}[v_1 < n \leq v_2], \quad (3.3)$$

$$\int_{-\infty}^{\infty} p(x) dx = 1 = \text{Prob}[x \leq \infty]. \quad (3.4)$$

Several PDFs are widely used to describe the amplitude characteristics of electrical and electronic circuit noise. These include the following:

$$p(x) = \left(\frac{1}{\sigma_x \sqrt{2\pi}} \right) \exp \left[-\frac{(x - \bar{x})^2}{2\sigma_x^2} \right] \text{ Gaussian or normal PDF}, \quad (3.5)$$

$$\left. \begin{array}{l} p(x) = 1/2a, \quad \text{for } -a \leq x \leq a, \quad \text{and} \\ p(x) = 0, \quad \text{for } |x| > a \end{array} \right\} \text{ Rectangular PDF}, \quad (3.6)$$

$$p(x) = \left(\frac{x}{\alpha^2} \right) \exp \left[-\frac{x^2}{2\alpha^2} \right], \quad x \geq 0 \text{ Rayleigh PDF}. \quad (3.7)$$

Under most conditions, we assume that the random noise arising in electronic circuits is modeled by a *Gaussian* PDF. Many mathematical benefits follow this approximation; for example, the output $y(t)$ of an LS is Gaussian with variance σ_y^2 given the input $x(t)$ to be Gaussian with variance σ_x^2 . If Gaussian noise passes through a non-LS, the output PDF in general will not be Gaussian.

Our concerns with noise in instrumentation systems will be focused in the frequency domain. Here, we will be concerned with the PDS and the root power spectrum, as treated in the following texts. We will generally assume the noise PDFs are Gaussian.

3.2.2 Power Density Spectrum

We will use a heuristic approach to illustrate the meaning of the PDS of noise.

Formally, the two-sided PDS is the FT of the *autocorrelation function*, defined in the following by

$$R_{nn}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T n(t) n(t + \tau) dt. \quad (3.8)$$

The two-sided PDS is an even function and can be written as the FT of the autocorrelation function:

$$\Phi_{nn}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R_{nn}(\tau) e^{-j\omega\tau} d\tau. \quad (3.9)$$

Because $R_{nn}(\tau)$ is an even function, its FT, $\Phi_{nn}(\omega)$, is also an even function, which when stated mathematically means

$$\Phi_{nn}(\omega) = \Phi_{nn}(-\omega). \quad (3.10)$$

In the following discussion, we will consider the *one-sided PDS*, $S_n(f)$, which is related to the two-sided PDS by

$$S_n(f) = 2\Phi_{nn}(2\pi f) \quad \text{for } f \geq 0, \quad (3.11)$$

and

$$S_n(f) = 0 \quad \text{for } f < 0. \quad (3.12)$$

To see how we may experimentally find the one-sided PDS, examine the system shown in Figure 3.1. Here, a noise voltage is the input to an ideal LPF with adjustable cutoff frequency, f_c . The random output of the ideal LPF is measured with a broadband, true RMS, AC voltmeter. We begin with $f_c = 0$ and systematically increase f_c , each time recording the square of the RMS meter reading, or mean-squared output voltage, v_{on}^2 , of the filter. As f_c is increased, v_{on}^2 increases monotonically, as shown in Figure 3.2. Because of the finite bandwidth of the noise source, v_{on}^2 eventually reaches an upper limit that is the total mean-squared noise voltage of the noise source. The plot of $v_{on}^2(f_c)$ vs. f_c is called the *cumulative mean-squared noise characteristic* of the noise source. Its units are mean-squared volts, in this example. A simple interpretation of the one-sided, noise PDS is that it is the derivative, or slope, of the mean-squared noise characteristic curve. Mathematically stated, this is

$$S_n(f) = \frac{dv_{on}^2(f)}{df} \text{ Mean squared volts/Hertz.} \quad (3.13)$$

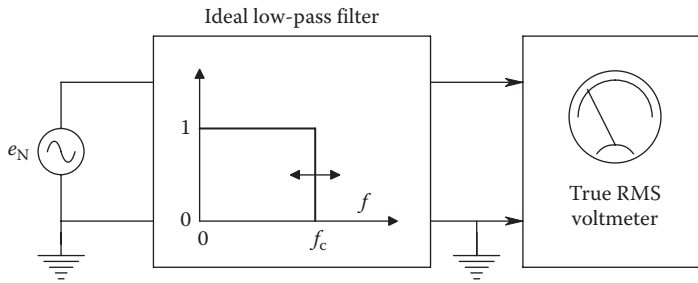


FIGURE 3.1 System used to calculate the one-sided PDS of the noise, e_N .

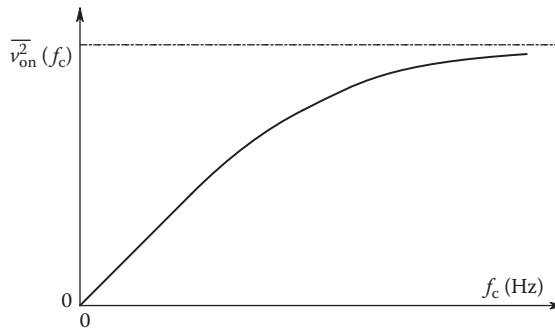


FIGURE 3.2 Mean-squared output noise $\overline{v_{on}^2}$ of the system of Figure 3.1 as a function of the ideal LPF's cutoff Hz bandwidth, f_c .

A plot of a typical noise PDS is shown in Figure 3.3. Note that a practical PDS generally drops off to zero at high frequencies.

A question that is sometimes asked by those first encountering the PDS concept is: Why is it called a *PDS*? The power concept has its origin in the consideration of noise in communication systems and has little relevance in the consideration of noise in instrumentation systems. One way to rationalize the power term is to consider a noise voltage source with

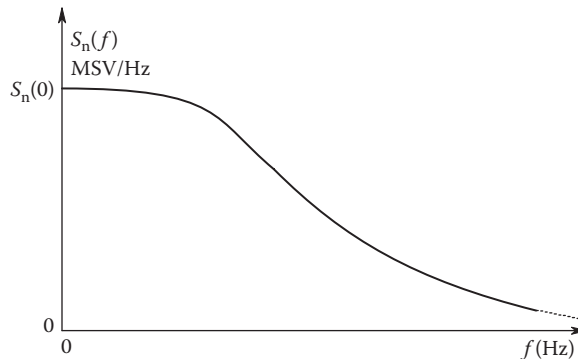


FIGURE 3.3 A one-sided, noise PDS, derived by differentiation of the plot of $\overline{v_{on}^2}(f_c)$ with respect to f_c . Its units are mean-squared volts per Hz.

a 1 Ω load. In this case, the average power dissipated in the resistor in watts is simply the total mean-squared noise voltage.

From our heuristic definition of the PDS earlier, we see that the total mean-squared voltage in the noise voltage source can be written as

$$\overline{v_{\text{on}}^2} = \int_0^\infty S_n(f) df. \tag{3.14}$$

The mean-squared voltage in the frequency interval (f_1, f_2) is found by

$$\overline{v_{\text{on}}^2}(f_1, f_2) = \int_{f_1}^{f_2} S_n(f) df. \tag{3.15}$$

($f_2 - f_1$) is also called the noise bandwidth, B.

Often, noise is specified or described using *root power density spectra*, which is simply a plot of the square root of $S_n(f)$ vs. f and which has the units of RMS volts (or other units) per root Hz.

Special (ideal) PDSs are used to model or approximate portions of real PDSs. These include the *white noise PDS* and the *one-over-f PDS*. A white noise PDS is shown in Figure 3.4. Note that this PDS is flat; this implies that

$$\int_0^\infty S_{\text{nw}}(f) df = \infty, \tag{3.16}$$

which is clearly not realistic. A one-sided 1/f PDS is shown in Figure 3.5. The 1/f spectrum is often used to approximate the LF behavior of real PDS. Physical processes that generate 1/f-like noise include surface phenomena associated with electrochemical electrodes, carbon composition resistors carrying DC (metallic resistors are substantially free of 1/f noise), and surface imperfections affecting emission and diffusion phenomena in semiconductor devices. The presence of 1/f noise can present problems in the electronic signal conditioning systems used for low-level, LF, and DC measurements, such as found in biomedical and geophysical applications.

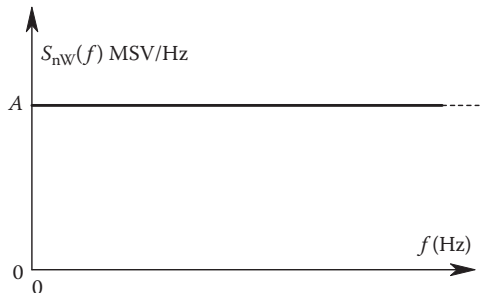


FIGURE 3.4
A white noise PDS.

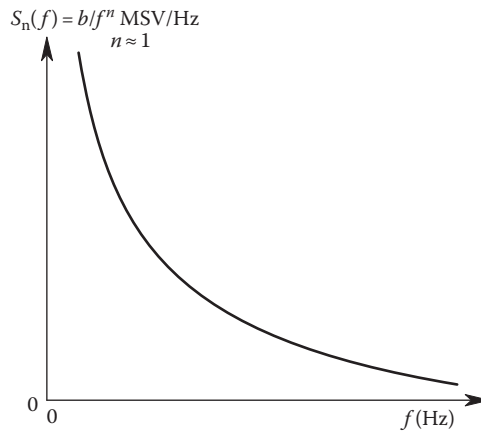


FIGURE 3.5
A $1/f$ PDS.

3.2.3 Sources of Noise in Signal Conditioning Systems

Sources of noise in signal conditioning systems can be separated into two major categories: noise from passive resistors and noise from active circuit elements such as bipolar junction transistors (BJTs), field effect transistors, and vacuum tubes. In most cases, the Gaussian assumption for noise amplitude PDFs is valid, and the noise generated can generally be assumed to be white over a major portion of the spectra.

3.2.3.1 Noise from Resistors

From statistical mechanics, it can be shown that any pure resistance at a temperature T K will have a zero mean noise voltage associated with it. This noise voltage appears in series with the (noiseless) resistor to form a Thevenin equivalent circuit. From DC to radio frequencies where the resistor's capacitance to ground and its lead inductance can no longer be neglected, the resistor's noise is well modeled by a Gaussian white noise source. Noise from resistors is called *thermal* or *Johnson noise*. Its one-sided PDS is given by

$$S_n(f) = 4kTR \text{ mean squared volts/Hz (MSV/Hz)}, \quad 0 \leq f \leq \infty, \quad (3.17)$$

where

k is Boltzmann's constant (1.38×10^{-23})

T is in Kelvin

R is in ohms

In a given bandwidth, the mean-squared noise from a resistor can be written as

$$v_{\text{on}}^2(B) = \int_{f_1}^{f_2} S_n(f) df = 4kTR(f_2 - f_1) = 4kTRB \text{ mean squared volts}, \quad (3.18)$$

where B is the equivalent Hz noise bandwidth, $B = (f_2 - f_1)$.

A Norton equivalent of the Johnson noise source from a resistor can be formed by assuming that a white noise current source with PDS

$$S_{ni}(f) = 4kTG \text{ MS Amps/Hz} \tag{3.19}$$

is in parallel with a noiseless conductance, $G = 1/R$.

The Johnson noise from several resistors connected in a network may be combined into a single noise voltage source in series with a single, noiseless, equivalent resistor. Figure 3.6 illustrates some of these reductions for two-terminal circuits.

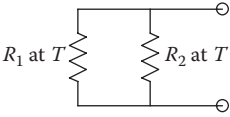
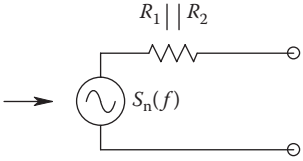
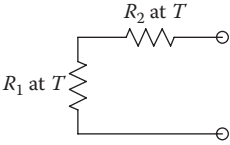
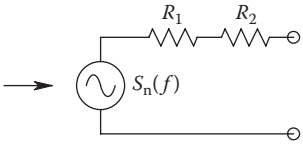
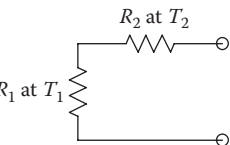
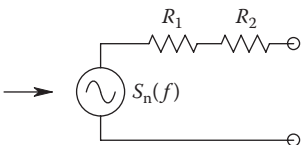
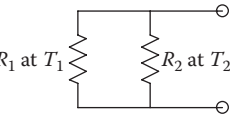
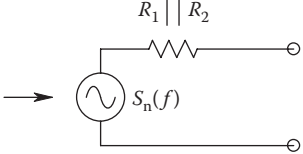
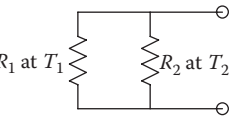
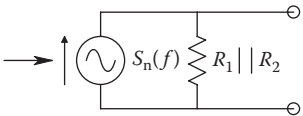
Circuit	Noise equivalent circuit with noiseless R_s	Power density spectrum
		$S_n(f) = 4kT[R_1R_2/(R_1 + R_2)]$
		$S_n(f) = 4kT(R_1 + R_2)$
		$S_n(f) = 4k(T_1R_1 + T_2R_2)$
		$S_n(f) = 4kT_1R_1[R_2/(R_1+R_2)]^2 + 4kT_2R_2[R_1/(R_1+R_2)]^2$
		$S_n(f) = 4kT_1G_1 + 4kT_2G_2 \text{ msA/Hz}$

FIGURE 3.6 Means of representing the thermal or Johnson noise from pairs of resistors using Thevenin and Norton equivalent circuits. Noises are always added as mean-squared quantities.

It has been observed that when DC (or average) current is passed through a resistor, the basic Johnson noise PDS is augmented by the addition of a $1/f$ spectrum:

$$S_n(f) = 4kTR + \frac{AI^2}{f} \text{ MSV/Hz}, \tag{3.20}$$

where

I is the average or DC component of current through the resistor

A is a constant that depends on the material from which the resistor is constructed (e.g., carbon composition, resistance wire, metal film)

An important parameter for resistors carrying average current is the crossover frequency, f_c , where the $1/f$ PDS equals the PDS of the Johnson noise. This is easily seen to be

$$f_c = \frac{AI^2}{4kTR} \text{ Hz}. \tag{3.21}$$

It is possible to show that the f_c of a noisy resistor can be reduced by using a resistor of the same type but having a higher wattage or power dissipation rating. As an example of this principle, consider the circuit of Figure 3.7 in which a single resistor of R ohms, carrying a DC I , is replaced by *nine* resistors of resistance R connected in series-parallel to give a net resistance R that can dissipate nine times the power of a single resistor R . The noise PDS in any one of the nine resistors is seen to be

$$S'_n(f) = 4kTR + \frac{A(I/3)^2}{f} \text{ MSV/Hz}. \tag{3.22}$$

Each of the nine PDSs given by Equation 3.22 earlier contributes to the net PDS seen at the terminals of the composite 9 W resistor. Each resistor's equivalent noise voltage

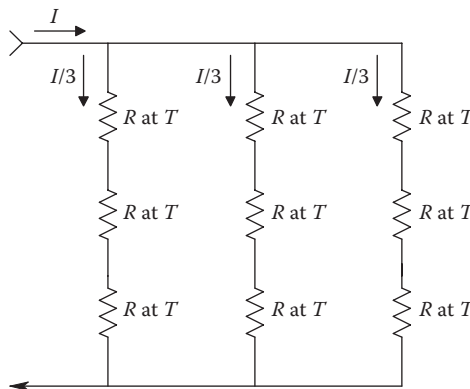


FIGURE 3.7

Equivalent circuit giving a ninefold increase in resistor wattage rating while reducing the net resistance's $1/f$ noise by nine. The thermal noise remains the same.

source *sees* a voltage divider formed by the other eight resistors in the composite resistor. The attenuation of each of the nine voltage dividers is given by

$$\frac{3R/2}{3R/2+3R} = \frac{1}{3}. \tag{3.23}$$

The net voltage PDS at the composite resistor’s terminals may be found by superposition of PDSs. It is

$$S_{n(9)}(f) = \sum_{j=1}^9 \left[4kTR + \frac{A(I/3)^2}{f} \right] \left(\frac{1}{3} \right)^2 = 4kTR + \frac{AI^2}{9f} \text{ MSV/Hz}. \tag{3.24}$$

Thus, the composite 9 W resistor enjoys a ninefold reduction in the $1/f$ noise spectral power because the DC density through each element is reduced by $1/3$. The Johnson noise PDS remains the same, however. It is safe to generalize that the use of high-wattage resistors of a given type and resistance will result in reduced $1/f$ noise generation when the resistor carries DC (average). Note that wire-wound resistors generally have smaller A factor values than do carbon composition or metal oxide resistors.

3.2.3.2 Two-Source Noise Model for Active Devices

Noise arising in JFETs, BJTs, and complicated IC amplifiers is generally described by the *two-source input model*. The total noise observed at the output of an amplifier, given that its input terminals are short-circuited, is accounted for by defining an *equivalent short-circuited input noise voltage*, e_{na} , which replaces *all* internal noise sources affecting the amplifier output under short-circuited input conditions. The amplifier, shown in Figure 3.8, is now considered noiseless. e_{na} is generally specified by manufacturers for low-noise discrete transistors and IC amplifiers.

e_{na} is a root PDS, that is, it is a function of frequency, and has the units of RMS volts per root Hz (RMS V/ $\sqrt{\text{Hz}}$). Figure 3.9 illustrates a plot of a typical $e_{na}(f)$ vs. f for a low-noise JFET. Also shown in Figure 3.9 is a plot of the equivalent input current noise, $i_{na}(f)$ vs. f , for the same device.

In addition to the equivalent short-circuited input noise voltage, the modeling of the net noise characteristics of amplifiers requires the inclusion of an equivalent input current

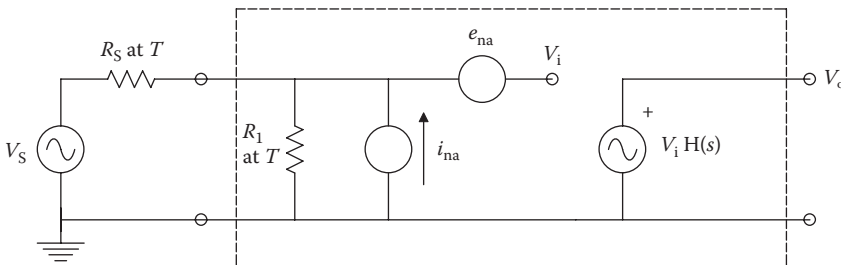


FIGURE 3.8

Equivalent two-source model for a noisy amplifier: e_{na} is the equivalent input short-circuited voltage noise root PDS. Its units are RMS volts per root Hz. i_{na} is the equivalent input noise current root PDS. Its units are RMS amps per root Hz. e_{na} and i_{na} are generally functions of frequency.

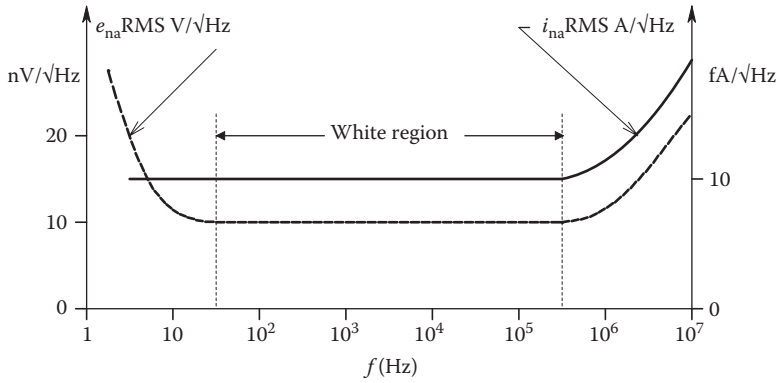


FIGURE 3.9

Plots of typical $e_{na}(f)$ and $i_{na}(f)$ for a JFET input amplifier. Note units.

noise source, i_{na} , as shown in Figure 3.8. $i_{na}(f)$ is the root PDS of the input equivalent noise current; its units are RMS amps per root Hz (RMS A/ $\sqrt{\text{Hz}}$). Note that both $e_{na}(f)$ and $i_{na}(f)$ have flat, midfrequency portions that invite approximation by white noise sources. At high frequencies, both equivalent noise root PD spectra slope upward. $e_{na}(f)$ for discrete JFETs, MOSFETs, BJTs, and IC amplifiers generally shows a $1/f$ region at low frequencies.

3.2.3.3 Noise in JFETs

Certain selected, discrete JFETs are used in the design of low-noise headstages for signal conditioning in measurement systems. Some JFETs give exceptional low-noise performance in the audio and sub-audio-frequency regions of the spectrum, while others excel in the video and radio-frequency end of the spectrum giving them applications for RF oscillators, mixers, and tuned amplifiers.

Van der Ziel (1974) has shown that the theoretical thermal noise generated in the conducting channel of a JFET can be approximated by an equivalent short-circuited input noise with PDS given by

$$e_{na}^2 = \frac{4kT}{g_m} = \left(\frac{4kT}{g_{m0}} \right) \sqrt{\frac{I_{DSS}}{I_{DQ}}} \text{MSV/Hz}, \tag{3.25}$$

where g_{m0} is the FET’s SS transconductance measured when $V_{GS} = 0$ and $I_D = I_{DSS}$. I_{DSS} is the DC drain current measured for $V_{GS} = 0$ and $V_{DS} > V_P$, and I_{DQ} is the quiescent DC drain current at the FET’s quiescent operating point where $V_{GS} = V_{GSQ}$.

In reality, due to the presence of $1/f$ noise, the theoretical short-circuited input voltage PDS expression can be modified, sic:

$$e_{na}^2(f) = \left(\frac{4kT}{g_m} \right) \left(1 + \frac{f_c}{f^n} \right) \text{MSV/Hz}. \tag{3.26}$$

The exponent n has the range $1 < n < 1.5$ and is device and lot determined. n is usually set equal to one for algebraic simplicity. The origins of the $1/f^n$ effect in JFETs are poorly understood. Note that e_{na} given by Equation 3.26 is temperature dependent. Heatsinking or actively

cooling the JFET will reduce e_{na} . The parameter f_c used in Equation 3.26 is the corner frequency of the $1/f$ noise spectrum. Depending on the device, it can range from below 10 Hz to above 1 kHz. f_c is generally quite high in RF and frequency JFETs because in this type of transistor, $e_{na}(f)$ dips to around 2 nV RMS/ $\sqrt{\text{Hz}}$ in the 10^5 – 10^7 Hz region, which is quite good.

Reverse-biased JFET gates have a DC leakage current, I_{GL} , which produces *shot noise* that is superimposed on the leakage current. This noise component in I_{GL} is primarily due to the random occurrence of charge carriers that have enough energy to cross the reverse-biased gate-channel diode junction. The PDF of the gate current shot noise is Gaussian, and its one-sided PDS is given by

$$i_{na}^2 = 2qI_{GL} \text{ MSA/Hz}, \tag{3.27}$$

where

$$q = 1.602 \times 10^{-19} \text{ C (electron charge)}$$

I_{GL} is the DC gate leakage current in amperes

I_{GL} is typically about 2 pA; hence, i_{na} is about 1.8 fA RMS/ $\sqrt{\text{Hz}}$ in the flat midrange of $i_{na}(f)$. $i_{na}(f)$, like $e_{na}(f)$, shows a $1/f$ characteristic at LFs. The PDS, $i_{na}^2(f)$, can be modeled by

$$i_{na}^2(f) = 2qI_{GL} \left(1 + \frac{f}{f_{ic}} \right) \text{ MSA/Hz}. \tag{3.28}$$

where f_{ic} is the current noise corner frequency.

For some JFET numbers, the measured $e_{na}(f)$ and $i_{na}(f)$ have been found to be greater than the predicted theoretical values; in other cases, they have been found to be less. No doubt the causes for these discrepancies lie in the oversimplifications used in their derivations.

3.2.3.4 Noise in BJTs

The values of e_{na} and i_{na} associated with BJT amplifiers depend strongly on the device's quiescent operating (Q) point. This is because there are shot noise components superimposed on the quiescent base and collector currents. The SS model of a grounded-emitter BJT amplifier is shown in Figure 3.10B. In this circuit, we assume negligible noise from R_L , the voltage-controlled current source (VCCS), $g_m v_{be}$, and the SS base input resistance, r_x . The shot noise sources are

$$i_{nb}^2 = 2qI_{BQ} \text{ MSA/Hz}, \tag{3.29}$$

$$i_{nc}^2 = 2q\beta I_{BQ} \text{ MSA/Hz}. \tag{3.30}$$

In this example, it is algebraically simpler to not find the equivalent input e_{na} and i_{na} . Rather, we work directly with the two white shot noise sources in the midfrequency, hybrid- π , SS model. It can be shown (Northrop 1990) that the total output noise voltage PDS is given by

$$S_{No} = 4kTR_L + 2q \left(\frac{I_{BQ}}{\beta} \right) (\beta R_L)^2 + \frac{4kTR_s'(\beta R_L)^2 + 2qI_{BQ}R_s'^2(\beta R_L)^2}{(V_T/I_{BQ} + r_x)^2} \text{ MSV/Hz}, \tag{3.31}$$

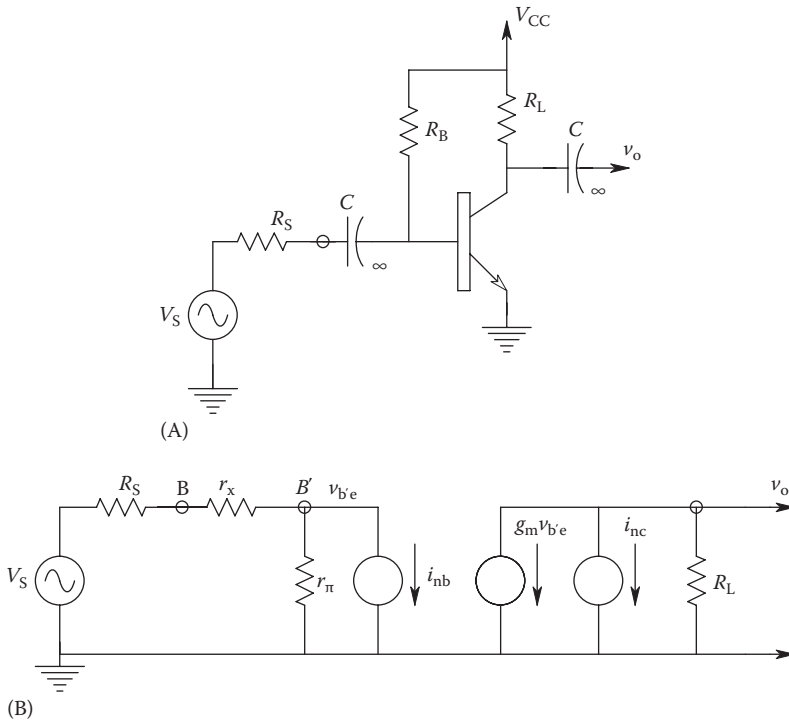


FIGURE 3.10

(A) A simple, noisy BJT amplifier. Capacitances are assumed to have negligible reactances in the operating range of frequencies. (B) Midfrequency, SS, equivalent circuit of the BJT amplifier. It is assumed that $R_B \gg r_x$, r_π and R_L are assumed to be noiseless, and $R'_s = r_x + R_s$ makes thermal noise. i_{nb} and i_{nc} are white shot noise current sources.

where it is clear that r_π is approximated by V_T/I_{BQ} , $R'_s = r_x + R_s$, and we neglected the Johnson noise from R_L because it is numerically small compared to the other terms. It is easy to show that the mean-squared output signal can be written as

$$v_{os}^2 = \frac{v_s^2 (\beta R_L)^2}{(V_T / I_{BQ} + r_x)^2} \text{MSV}. \tag{3.32}$$

Thus, the MS SNR at the amplifier output can be written as

$$\text{SNR}_o = \frac{v_s^2 / B}{4kTR'_s + 2qI_{BQ}R_s'^2 + 2q(I_{BQ} / \beta)(V_T / I_{BQ} + R'_s)^2}, \tag{3.33}$$

where B is the specified, equivalent (Hz) noise bandwidth for the system.

The SNR_o given by Equation 3.33 has a maximum for some nonnegative $I_{BQ\text{MAX}}$. The $I_{BQ\text{MAX}}$ that will give this maximum can be found by differentiating the denominator of Equation 3.33 and setting the derivative equal to zero. This gives

$$I_{BQ\text{MAX}} = \frac{V_T}{(R'_s \sqrt{\beta + 1})} \text{A}. \tag{3.34}$$

What we should remember from the earlier exercise is that the best noise performance for linear, SS, BJT amplifiers is a function of quiescent biasing conditions (Q-point). Often, these conditions have to be found by trial and error when working at high frequencies. While individual BJT amplifiers may best be modeled for noise analysis with the two shot noise current sources as we did earlier, it is more customary when describing complex, BJT IC amplifier noise performance to use the more general and more easily used e_{na} and i_{na} , two-source model.

3.3 Propagation of Gaussian Noise through Linear Filters

In a formal, rigorous treatment of noise in LSs, it is possible to show that the PDF of the output of an LS is Gaussian, given a Gaussian noise input. In addition, the PDS of the system's output noise is given by

$$S_y(f) = S_x(f) |H(2\pi fj)|^2 \text{ MSV / Hz.} \tag{3.35}$$

This is the scalar product of the positive real input PDS and the magnitude squared of the system's frequency response function. This relation can be extended to include two or more cascaded systems, as shown in Figure 3.11:

$$S_y(f) = S_x(f) |H(2\pi jf)|^2 |G(2\pi jf)|^2 \text{ MSV / Hz} \tag{3.36}$$

or

$$S_y(f) = S_x(f) |H(2\pi jf)G(2\pi jf)|^2 \text{ MSV / Hz.} \tag{3.37}$$

If white noise with a PDS $S_x(f) = \eta$ MSV/Hz is the input to an LS, then the output PDS is simply

$$S_y(f) = \eta |H(2\pi jf)|^2 \text{ MSV / Hz.} \tag{3.38}$$

The total mean-squared output noise of the system is

$$\overline{v_{on}^2} = \int_0^\infty S_y(f) df = \eta \int_0^\infty |H(2\pi jf)|^2 df \text{ MSV.} \tag{3.39}$$

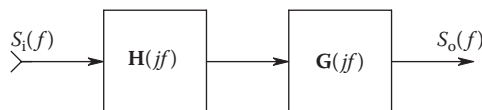


FIGURE 3.11
Propagation of noise through two cascaded LSs.

TABLE 3.1

Noise Gain² × Bandwidth Products for Some Common System Frequency Response Functions

Frequency Response Function, $H(j\omega)$	Gain ² Hz Bandwidth
$\frac{K_V}{j\omega\tau + 1}$	$K_V^2 [1/4\tau]$
$\frac{K_V}{(j\omega\tau_1 + 1)(j\omega\tau_2 + 1)}$	$K_V^2 [1/(4(\tau_1 + \tau_2))]$
$\frac{K_V}{(1 - \omega^2/\omega_n^2) + j\omega(2\xi/\omega_n)}$	$K_V^2 [\omega_n / 8\xi]$
$\frac{K_V j\omega(2\xi/\omega_n)}{(1 - \omega^2/\omega_n^2) + j\omega(2\xi/\omega_n)}$	$K_V^2 [\omega_n \xi] = K_V^2 [\omega_n / 2Q]$
$\frac{K_V j\omega\tau_1}{(j\omega\tau_1 + 1)(j\omega\tau_2 + 1)}, \tau_1 > \tau_2$	$K_V^2 \left[\frac{\tau_1}{4\tau_2(\tau_1 + \tau_2)} \right]$

The right-hand integral of Equation 3.39 may be shown to be, for transfer functions with one more finite pole than zeros, the product of the transfer function’s LF or midband gain squared times the filter’s equivalent Hz noise bandwidth. The filter’s *gain squared × bandwidth* product is thus given by

$$\text{GAIN}^2 \text{ BW} = \int_0^\infty |\mathbf{H}(2\pi jf)|^2 df. \tag{3.40}$$

Gain² × bandwidth integrals have been evaluated using complex variable theory for a number of transfer functions (see James et al. 1947). In Table 3.1, we give the gain² × bandwidth integrals for five common amplifier frequency response functions. Note that the equivalent noise bandwidths (in brackets in each case) are in Hz, not rad/s. Also note the absence of 2π factors in these expressions. Gain² × bandwidth integrals are used to estimate the total MS output noise from amplifiers having (approximate) white noise input sources and are thus useful in calculating output SNRs.

3.4 Broadband Noise Factor and Noise Figure of Amplifiers

An amplifier’s *noise factor* is defined as the ratio of the mean-squared SNR at the amplifier’s input to the SNR at the amplifier’s output. Since a real amplifier is noisy and adds noise to the signal as well as amplifying it, the output SNR (SNR_O) is always less than the input SNR (SNR_I); hence, the *noise factor* is always greater than one for a noisy amplifier:

$$F \equiv \frac{\text{SNR}_I}{\text{SNR}_O}. \tag{3.41}$$

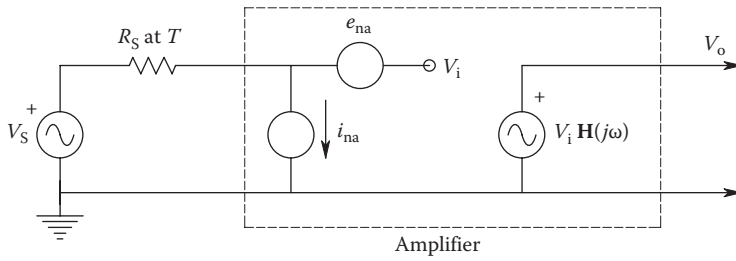


FIGURE 3.12
A minimum noisy amplifier model relevant to the calculation of amplifier noise factor and noise figure.

The *noise figure* is defined as

$$NF \equiv 10 \log_{10}(F) \tag{3.42}$$

when the SNRs are in terms of mean-squared quantities.

Figure 3.12 illustrates a minimum model for a noisy amplifier. Here, we assume that the spectra of e_{na} and i_{na} are white and that $R_1 \gg R_s$. The MS input signal is $S_i = v_s^2$, and the MS output signal is $S_o = K_v^2 v_s^2$, where K_v^2 is the amplifier's midband gain. The MS input noise is simply that associated with v_s (here zero) plus the Johnson noise from the source resistance, R_s , in a specified Hz noise bandwidth, B.

It is written as

$$N_i = 4kTR_s B \text{ MSV}. \tag{3.43}$$

The mean-squared noise at the amplifier's output, N_o , is composed of three components: one from the R_s

Johnson noise and two from the equivalent noise sources. N_o can be written as

$$\begin{aligned} N_o &= \left(4kTR_s + e_{na}^2 + i_{na}^2 R_s^2\right) \int_0^\infty |H(2\pi jf)|^2 df \\ &= \left(4kTR_s + e_{na}^2 + i_{na}^2 R_s^2\right) K_v^2 B \text{ MSV}. \end{aligned} \tag{3.44}$$

Using the definition for F, we find that the noise factor for the simple noisy amplifier model can be written as

$$F = 1 + \frac{e_{na}^2 + i_{na}^2 R_s^2}{4kTR_s}. \tag{3.45}$$

Note that this expression for F contains no bandwidth terms; they cancel out. When the NF is given for an amplifier, R_s must be specified, as well as the Hz bandwidth B over which the noise is measured. The temperature should also be specified, although common practice usually sets T at 300 K.

For practical amplifiers, NF and F are functions of frequency because e_{na} and i_{na} are functions of frequency (see Figure 3.9). For a given R_s , F tends to rise at LFs due to the $1/f$

components in the equivalent input noise sources. F also increases at HFs, again, due to the HF increases in e_{na} and i_{na} . Often, we are interested in the noise performance of an amplifier in either LFs or HFs where NF and F are not minimum. To examine the detailed noise performance of an amplifier at LFs and HFs, we use the *spot noise figure*, discussed in the following texts.

3.5 Spot Noise Factor and Figure

Spot noise measurements are made through a narrow BPF in order to evaluate an amplifier's noise performance in a certain narrow frequency range, particularly where $e_{na}(f)$ and $i_{na}(f)$ are not constant, such as in the $1/f$ range. Figure 3.13A illustrates a set of spot noise figure contours for a commercial, low-noise preamplifier having applications at audio-frequencies. Figure 3.13B shows the spot NF contours when an ideal transformer with $n = 100$ is used to couple the Thevenin source to the amplifier.

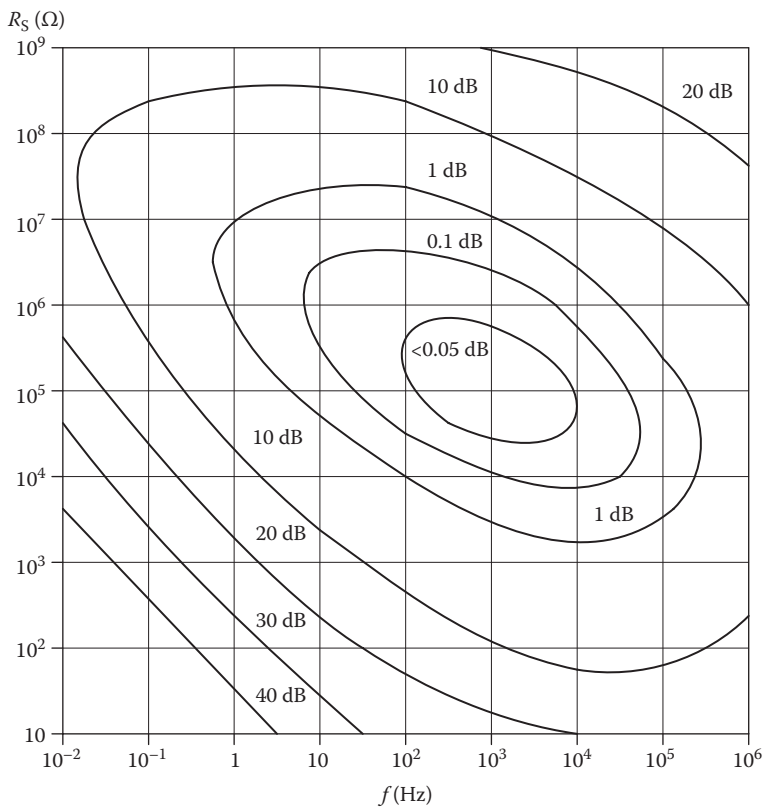


FIGURE 3.13

(A) Spot noise figure contours for a typical amplifier with a direct input having a Thevenin source resistance R_S at temperature T K.

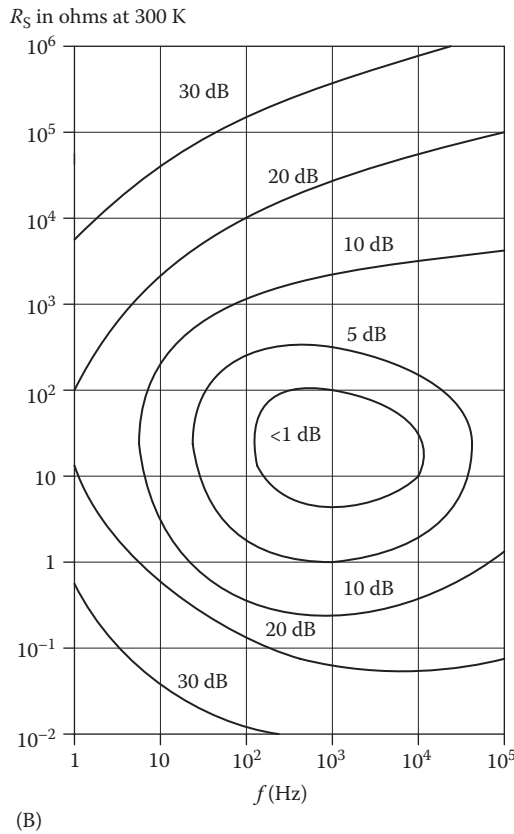


FIGURE 3.13 (continued)

(B) Spot noise figure contours for the same amplifier using a 1:100 turns ratio input transformer to couple the Thevenin source to the amplifier. Note the shift of the minimum contour to a new, lower range of R_s .

A system for determining an amplifier’s spot noise figure is shown in Figure 3.14. An adjustable white noise source is used at the amplifier’s input. Note that the output resistance of the white noise source plus some external resistance must add up to R_s , the specified Thevenin equivalent input resistance. The system is used as follows: *First*, the BPF is set to the desired center frequency and the white noise generator output is set to $e_N = 0$. We assume that the total mean-squared noise at the system output under these conditions can be modeled as

$$N_o(f_c) = [4kTR_s + e_{na}^2(f_c) + i_{na}^2(f_c)R_s^2] K_V^2 B_F \text{ MSV}, \tag{3.46}$$

where

$e_{na}(f_c)$ is the value of e_{na} at the BPF center frequency, f_c

B_F is the equivalent noise bandwidth of the BPF

K_V is the combined gain of the amplifier under measurement at f_c , the BPF at f_c , and the post-amplifier. K_V can be written as

$$K_V = |\mathbf{H}(2\pi jf_c)| K_F K_A. \tag{3.47}$$

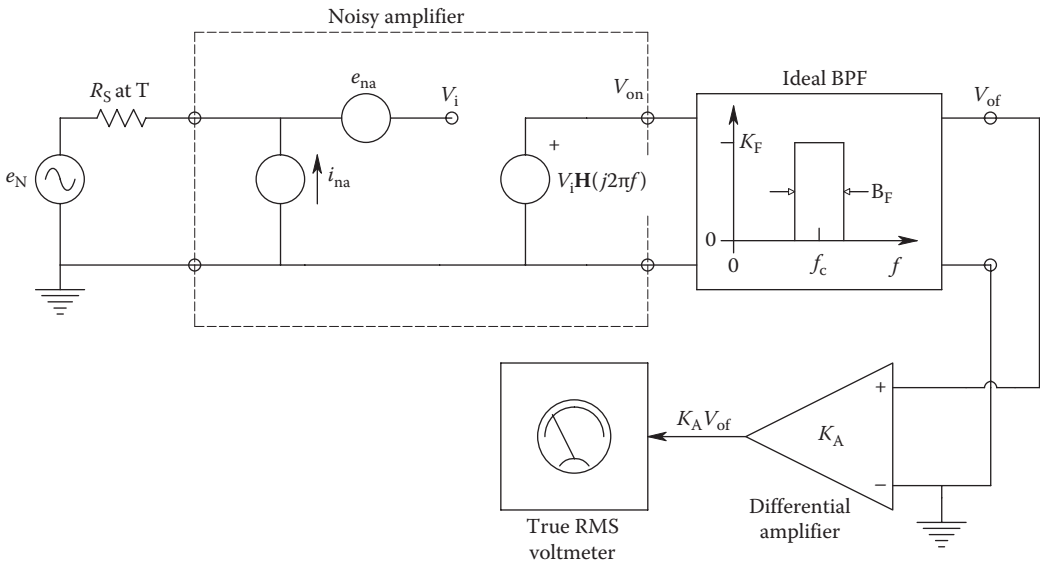


FIGURE 3.14
A system for measuring the spot noise figure of an amplifier.

Second, the white noise source is made nonzero and adjusted so that the true RMS meter reads $\sqrt{2}$ higher than in the first case with $e_N = 0$. The MS output voltage can now be written as

$$\begin{aligned}
 N'_o(f_c) &= 2N_o(f_c) = 2[4kTR_s + e_{na}^2(f_c) + i_{na}^2(f_c)R_s^2]K_V^2B_F \\
 &= [e_N^2 + 4kTR_s + e_{na}^2(f_c) + i_{na}^2(f_c)R_s^2]K_V^2B_F
 \end{aligned}
 \tag{3.48}$$

Under this condition, it is evident that

$$e_N^2 = 4kTR_s + e_{na}^2(f_c) + i_{na}^2(f_c)R_s^2.
 \tag{3.49}$$

Hence,

$$[e_{na}^2(f_c) + i_{na}^2(f_c)R_s^2] = [e_N^2 - 4kTR_s].
 \tag{3.50}$$

If the left-hand side of Equation 3.50 is substituted into Equation 3.45 for the noise factor, F, we find

$$F_{spot} = \frac{e_N^2}{(4kTR_s)}.
 \tag{3.51}$$

Note that this expression for the spot noise factor does not contain specific terms for the measurement center frequency, f_c , or the BPF's Hz noise bandwidth, B_F . However, these

parameters are always specified when giving F_{spot} for an amplifier. F_{spot} is actually calculated by setting f_c and R_s , then determining the e_N^2 value that doubles the mean-squared output noise. This value of e_N^2 is then divided by the calculated white noise spectrum from the resistor R_s to give a value for F_{spot} at a given frequency and R_s .

It is also possible to measure F_{spot} using a sinusoidal source of frequency f_c instead of the calibrated white noise source, e_N . See the home problems for this chapter for a detailed treatment of this method.

3.6 Transformer Optimization of Amplifier F_{spot} and Output SNR

Figure 3.13A illustrates the fact that for a given set of internal biasing conditions, a given amplifier will have an optimum operating region where NF_{spot} is a minimum in (R_s, f_c) space. In some instances, the input sensor to which the amplifier is connected has an R_s that is far smaller than the R_s giving the lowest NF_{spot} on the amplifier's spot NF contours. Consequently, the system (i.e., sensor and amplifier) is not operating to give either the lowest NF or the highest output SNR. One way of improving the output SNR is to couple the input sensor to the amplifier through a low-noise, low-loss transformer, as shown in Figure 3.15. Such coupling, of course, presumes that the signal coming from the transducer is AC and not DC, for obvious reasons. A practical transformer is a BP device, which loses efficiency at LFs and HFs, limiting the range of frequencies over which output SNR can be maximized.

The output MS \overline{SNR} can be calculated for the circuit of Figure 3.15 as follows: the MS input signal is simply v_s^2 . In the case of a sinusoidal input, it is well known that $v_s^2 = V_s^2/2$ MSV. The MS signal at the output is

$$S_o = \overline{v_s^2} n^2 K_v^2, \tag{3.52}$$

where

n is the transformer's secondary-to-primary turns ratio

K_v is the amplifier's midband gain

The transformer is assumed to *ideal* (lossless and noiseless). In practice, transformer windings have finite resistance; hence, they make Johnson noise, and their magnetic cores

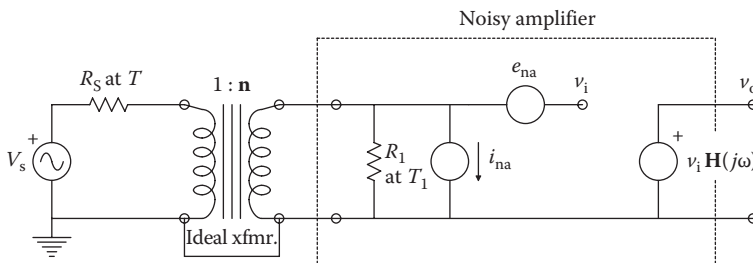


FIGURE 3.15 Use of an ideal input transformer to improve the output SNR of a noisy amplifier when R_s is small but finite.

contribute *Barkhausen noise* to their noisiness. The ideal transformer also has infinite frequency response. From this latter assumption, it is easy to show that the amplifier *sees* a transformed Thevenin equivalent circuit of the input transducer having an open-circuit voltage of $\mathbf{n} v_s(t)$ and a Thevenin resistance of $\mathbf{n}^2 R_s$. Thus, the mean-squared output noise can be written as

$$N_o = \left[\mathbf{n}^2 4kTR_s + e_{na}^2 + i_{na}^2 (\mathbf{n}^2 R_s)^2 \right] K_V^2 B \text{ MSV} \quad (3.53)$$

and the output SNR is

$$\text{SNR}_o = \frac{\overline{v_s^2} / B}{4kTR_s + e_{na}^2 / \mathbf{n}^2 + i_{na}^2 \mathbf{n}^2 R_s^2}. \quad (3.54)$$

The SNR_o clearly has a maximum with respect to \mathbf{n} . If the denominator of Equation 3.54 is differentiated with respect to \mathbf{n}^2 and set equal to zero, we find that an optimum turns ratio, \mathbf{n}_o , exists and is given by

$$\mathbf{n}_o = \sqrt{\frac{e_{na}}{i_{na} R_s}}. \quad (3.55)$$

If the noiseless (ideal) transformer is given the turns ratio of \mathbf{n}_o , then it is easy to show that the maximum output SNR is given by

$$\text{SNR}_{o\text{MAX}} = \frac{\overline{v_s^2} / B}{4kTR_s + 2e_{na} i_{na} R_s}. \quad (3.56)$$

The general effect of transformer SNR maximization on the system's noise figure contours is to shift the locus of minimum NF_{spot} to a lower range of R_s ; there is no obvious shift along the f_c axis. Also, the minimum NF_{spot} is higher with the transformer. This is because a practical transformer is noisy, as discussed earlier (see Figure 3.13B).

As a rule of thumb, use of a transformer to improve output SNR and reduce NF_{spot} is justified if $(e_{na}^2 + i_{na}^2 R_s^2) > 20 e_{na} i_{na} R_s$ in the range of frequencies of interest.

3.7 Cascaded Noisy Amplifiers

In this section, we examine a rule for the design of low-noise signal conditioning systems. Figure 3.16 illustrates three cascaded noisy amplifiers. The i_{na} terms are assumed to be negligible because $i_{na}^2 R_s^2 \ll e_{na}^2$ in each amplifier. We may write the MS output signal voltage as

$$S_{\text{out}} = v_s^2 K_{V(1)}^2 K_{V(2)}^2 K_{V(3)}^2 \text{ MSV}. \quad (3.57)$$

The MS output noise voltage can be written as

$$N_{\text{out}} = (4kTR_s + e_{na1}^2) K_{V(1)}^2 K_{V(2)}^2 K_{V(3)}^2 B + e_{na2}^2 K_{V(2)}^2 K_{V(3)}^2 B + e_{na3}^2 K_{V(3)}^2 B. \quad (3.58)$$

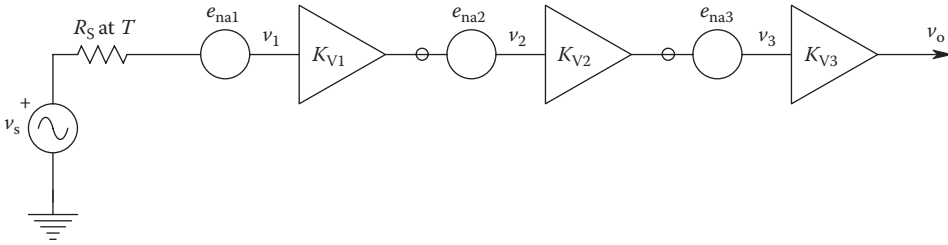


FIGURE 3.16
Three cascaded stages of noisy, but otherwise ideal, voltage amplifiers.

The three-amplifier system’s MS output SNR is thus

$$SNR_{out} = \frac{\overline{v_s^2}/B}{4kTR_s + e_{na1}^2 + e_{na2}^2/K_{V(1)}^2 + e_{na3}^2/K_{V(1)}^2K_{V(2)}^2}. \tag{3.59}$$

Note that the output SNR is set essentially by the noise characteristics of the input amplifier (headstage), as long as $K_{V(1)} > 5$, that is, the two right-hand terms in the denominator of the SNR_{out} expression are not numerically significant.

The point to remember from this section is that when it is necessary to design a low-noise, signal conditioning system, use the lowest noise amplifier in the headstage position and give it a gain of at least five.

3.8 Examples of Calculations of the Noise-Limited Resolution of Certain Signal Conditioning Systems

In designing instrumentation systems, it is often necessary to be able to predict the noise-limited, threshold resolvable measurand or, alternately, what input measurand level will produce a given system output SNR. In this section, and in the chapter home problems, we present some examples of situations in which a noise-limited, minimum input signal level is found.

3.8.1 Calculation of the Minimum Resolvable AC Input Voltage to a Noisy, Inverting Op-Amp Amplifier

Figure 3.17 illustrates a simple inverting OA circuit. The noisy OA is modeled as a frequency-compensated OA with a white, equivalent, short-circuit input noise, e_{na} . i_{na} is not included in the model because $i_{na} R_F \ll e_{na}$. We assume the input signal to be a sinusoidal voltage with peak value V_s and frequency f_s . The mean-squared output signal voltage is

$$S_{out} = \left(\frac{V_s^2}{2} \right) \left(\frac{R_F}{R_1} \right)^2 MSV. \tag{3.60}$$

The total mean-squared output noise is

$$N_{out} = 4kTR_1 \left(\frac{R_F}{R_1} \right)^2 B + 4kTR_F B + e_{na}^2 \left(1 + \frac{R_F}{R_1} \right)^2 B MSV. \tag{3.61}$$

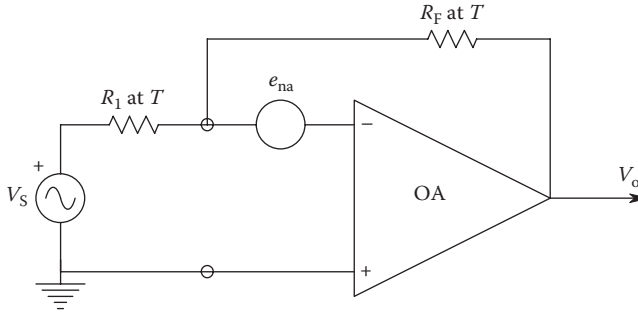


FIGURE 3.17
A noisy inverting OA amplifier.

The equivalent Hz noise bandwidth, B , may be calculated as follows: A frequency-compensated OA operated as a simple inverting amplifier has been shown (Northrop 1990) to have a frequency response given by

$$\frac{V_o}{V_s} = \mathbf{A}_V(j\omega) = \frac{-A_{vo}R_F / [R_F + R_1(1 + A_{vo})]}{j\omega \frac{\tau_a(R_F + R_1)}{R_F + R_1(1 + A_{vo})} + 1}. \quad (3.62)$$

In this expression, A_{vo} is the open-loop OA's DC gain, and τ_a is its open-loop time constant. τ_a may be expressed in terms of the OA's SS, unity-gain frequency, f_T :

$$\tau_a = \frac{A_{vo}}{2\pi f_T} \text{ s}. \quad (3.63)$$

Hence, the closed-loop, inverting amplifier has a time constant that can be expressed approximately by

$$\tau_{CL} = \frac{(1 + R_F/R_1)}{2\pi f_T} \text{ s}. \quad (3.64)$$

Thus, the equivalent Hz noise bandwidth for the amplifier of Figure 3.17 may be found using Table 3.1. It is

$$B = \frac{1}{4\tau_{CL}} = \frac{\pi f_T}{2(1 + R_F/R_1)} \text{ Hz}. \quad (3.65)$$

Thus, the output MS SNR can be written as

$$\text{SNR}_{\text{out}} = \frac{(V_s^2/2)(R_F/R_1)^2 2(1 + R_F/R_1)}{\left[4kTR_1(R_F/R_1)^2 + 4kTR_F + e_{na}^2(1 + R_F/R_1)^2\right] \pi f_T}. \quad (3.66)$$

In this SNR calculation, we assume that the frequency of the input signal, f_s , is much less than the closed-loop amplifier's break frequency, $1/(2\pi\tau_{CL})$ Hz, so that there is no attenuation of the output signal due to the LPF characteristic of the amplifier.

Using Equation 3.66 earlier, we now calculate the peak value of v_s that will give an MS SNR of unity, given the following typical system parameters: $R_F = 10^5$, $R_1 = 10^3$, $f_T = 12$ MHz, $4kT = 1.66 \times 10^{-20}$, and $e_{na} = 10$ nV RMS/ $\sqrt{\text{Hz}}$. Calculations yield $V_s = 6.65$ μV , peak. This input value is relatively high because of the broadband nature of the amplifier's output noise ($B = 5.94 \times 10^4$ Hz).

3.8.2 Calculation of the Minimum Resolvable Direct Current in White and 1/f Noise

In this example, a very small DC is to be measured, using an EOA transresistor circuit, shown in Figure 3.18. The DC source is represented by a Norton equivalent circuit, having current source, I_s , and conductance, G_s . The thermal noise from G_s , R_F , and i_{na} is assumed to have white spectra. e_{na} , however, is assumed to have a 1/f component:

$$e_{na}^2(f) = e_{na0}^2 + \frac{b}{f} \text{MSV/Hz.} \tag{3.67}$$

The PDS of the amplifier's output noise can be written as

$$S_{no}(f) = \left(e_{na0}^2 + \frac{b}{f} \right) (1 + R_F G_s)^2 + [4kT(G_F + G_s) + i_{na}^2] R_F^2 \tag{3.68}$$

This noise is, of course, broadband and is conditioned by a noiseless, unity-gain, quadratic LPF with transfer function

$$F(s) = \frac{1}{s^2/\omega_n^2 + s(2\zeta)/\omega_n + 1}. \tag{3.69}$$

Reference to No. 3 in Table 3.1 shows that the quadratic LPF has an equivalent noise bandwidth of $B = \omega_n/(8\zeta)$ Hz. The total MS output noise voltage can thus be found by integrating the white noise terms of $S_{no}(f)$ from 0 to $\omega_n/(8\zeta)$. Integration of the b/f term between the same limits leads to the perplexing result of infinity! To avoid this problem, we must impose a practical LF limit for the b/f integral. One arbitrary but effective solution to this

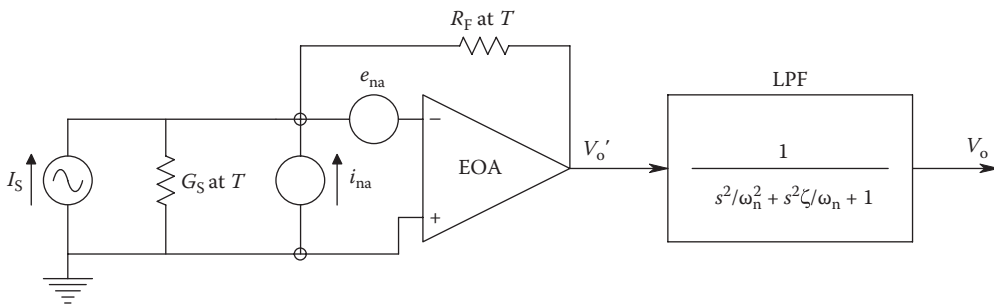


FIGURE 3.18 A noisy operational transresistor circuit followed by a noiseless LPF, used to condition small currents.

problem is to take the LF limit to be the reciprocal of the time it takes to confidently read the output voltage from the LPF with an analog or digital voltmeter. Let us assume this takes 4 s, so the LF limit is 0.25 Hz. Thus,

$$N_o = \left\{ (e_{na0}^2)(1 + R_F G_s)^2 + [4kT(G_F + G_s) + i_{na}^2] R_F^2 \right\} \left[\frac{\omega_n}{8\zeta} \right] + b \ln \left(\frac{4\omega_n}{8\zeta} \right) (1 + R_F G_s)^2 \text{ MSV.} \quad (3.70)$$

The MS, DC output signal is $S_o = I_s^2 R_F^2 \text{ MSV}$.

Now let us find the I_s to give an output MS SNR of 4. We assume the following parameters for an AD549 EOA: $4kT = 1.66 \times 10^{-20}$ at 300 K, $\omega_n/(8\zeta) = 2 \text{ Hz}$, $R_F = 10^{10} \Omega$, $G_s = 10^{-7} \text{ S}$, $e_{na0} = 35 \text{ nV RMS}/\sqrt{\text{Hz}}$, $i_{na} = 0.2 \text{ fA RMS}/\sqrt{\text{Hz}}$, $b = 1.36 \times 10^{-17} \text{ MSV}$. Evaluating the noise terms in Equation 3.70, we find that the dominant term comes from the Johnson noise in R_F and R_s ; this is $3.3233 \times 10^{-7} \text{ MSV}$. The MS output noise from the b/f term in e_{na}^2 is $2.8336 \times 10^{-11} \text{ MSV}$, and the total MS output noise from e_{na0}^2 and i_{na}^2 is $8.2455 \times 10^{-12} \text{ MSV}$. When we set the MS output SNR equal to four, we find that the minimum $I_s = 0.1153 \text{ DC microamps}$, and the DC voltage at the system output is 1.153 mV .

3.8.3 Calculation of the Minimum Resolvable AC Input Signal to Obtain a Specified Output SNR in a Transformer-Coupled, Tuned Amplifier

Figure 3.19 illustrates a signal conditioning system used to amplify the 10 kHz, sinusoidal voltage output of a low-impedance sensor. A transformer (here assumed to be ideal and noiseless) is used to maximize the output SNR. A noiseless, quadratic BPF, tuned to 10 kHz, follows the low-noise OA input stage. The peak gain of the tuned filter is unity. Looking toward the source through the ideal transformer, the OA summing junction sees a Thévenin equivalent circuit consisting of an open-circuit voltage, $v_{oc}(t) = n v_s(t)$, and an equivalent resistance of $R_{TH} = n^2 R_s$. The mean-squared output voltage is thus

$$S_{out} = \left(\frac{V_s^2}{2} \right) \left[\frac{R_F^2}{(nR_s)^2} \right] \text{ MSV.} \quad (3.71)$$

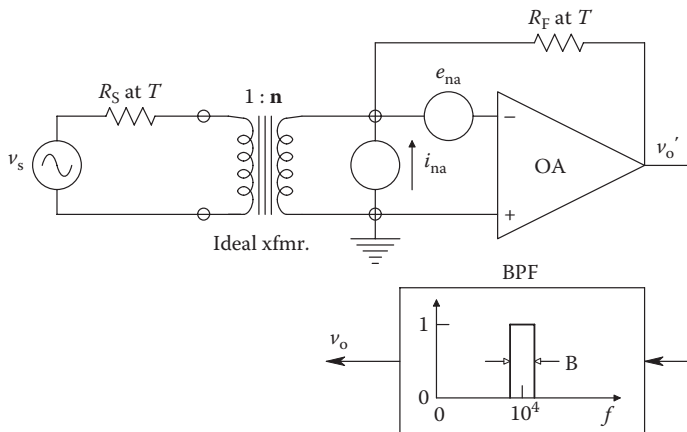


FIGURE 3.19 Use of an input transformer to improve the output SNR when conditioning a low-level sinusoidal input voltage from a low resistance source.

All sources of noise in the circuit are assumed to be white. The noise bandwidth is determined by the BPF and is found in Table 3.1 to be $B = \omega_n \zeta = \omega_n / 2Q$ Hz. The total MS output noise can thus be written as

$$N_{\text{out}} = \left[\frac{4kTR_s R_F^2}{(\mathbf{n}R_s)^2} + e_{\text{na}}^2 \left(1 + \frac{R_F}{\mathbf{n}^2 R_s} \right) + i_{\text{na}}^2 R_F^2 + 4kTR_F \right] \left(\frac{\omega_n}{2Q} \right) \text{MSV}. \tag{3.72}$$

Therefore, the MS output SNR is

$$\text{SRN}_{\text{out}} = \frac{(V_s^2/2)/B}{4kTR_s + e_{\text{na}}^2 (\mathbf{n}R_s/R_F + 1/\mathbf{n})^2 + i_{\text{na}}^2 R_s^2 \mathbf{n}^2 + 4kTG_F R_s^2 \mathbf{n}^2}. \tag{3.73}$$

The denominator of Equation 3.73 has a minimum for some nonnegative \mathbf{n}_o ; hence, the SNR has a maximum for that \mathbf{n}_o . To find \mathbf{n}_o , we differentiate the denominator with respect to \mathbf{n}^2 , set the derivative equal to zero, and then solve for \mathbf{n}_o . This yields

$$\mathbf{n}_o = \frac{\sqrt{e_{\text{na}}}/\sqrt{R_s}}{\left[e_{\text{na}}^2/R_F^2 + 4kTG_F + i_{\text{na}}^2 \right]^{1/4}}. \tag{3.74}$$

We now examine a numerical solution for \mathbf{n}_o and V_1 to give an SNR_{out} equal to unity. We use the following parameters: $e_{\text{na}} = 3$ nV RMS/ $\sqrt{\text{Hz}}$, $i_{\text{na}} = 0.4$ pA RMS/ $\sqrt{\text{Hz}}$, $R_F = 10^4 \Omega$, $R_s = 10 \Omega$, the BPF $Q = 100$, and $4kT = 1.66 \times 10^{-20}$. \mathbf{n}_o is found to be 14.73. When \mathbf{n}_o is substituted for \mathbf{n} in Equation 3.73, and a numerical solution is obtained, V_s to get unity SNR = 13 peak nV. If the transformer is not used ($\mathbf{n} = 1$), then V_s to give SNR = 1 is 76.0 peak nV.

3.8.4 Calculation of the Smallest $\Delta R/R$ in a Wheatstone Bridge to Give a Specified SNR_{out}

In this example, we calculate the smallest ΔR required to produce an MS output SNR of 10 in the AC-powered, Wheatstone bridge circuit shown in Figure 3.20. Two arms of the bridge are variable, one having increasing resistance while the other's resistance decreases

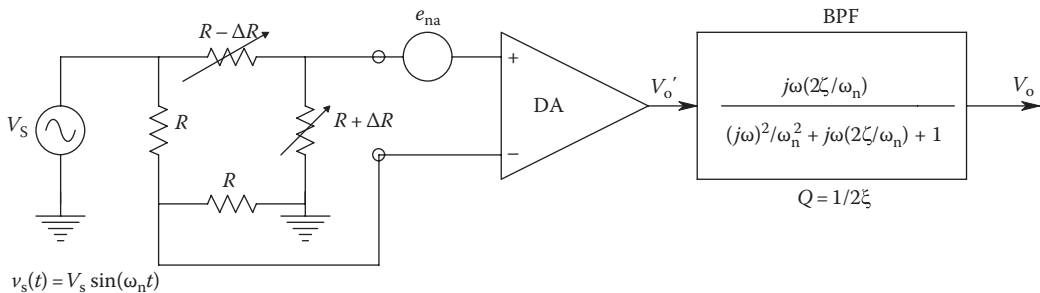


FIGURE 3.20 Use of a quadratic BPF to improve the output SNR and resolution of a Wheatstone bridge with AC excitation.

a corresponding amount. Such a configuration is found in certain unbonded strain gauges used to measure forces. The output of the bridge is amplified by a PMI AMP-01 low-noise, differential IA that is followed by a noiseless, quadratic BPF. The purpose of the BPF is to pass the amplified AC output of the bridge while restricting the bandwidth of the noise, thus achieving an output SNR that is higher than that at the output of the DA.

The signal output of the system is a sine wave whose amplitude and phase is determined by v_s and $\Delta R/R$. It may be found as follows: let the bridge excitation be $v_s(t) = V_s \sin(2\pi 400t)$. The voltage at the left-hand corner of the bridge, v_1' , is easily seen to be $v_s(t)/2$. The voltage at the right-hand corner of the bridge, v_1 , can be written as

$$v_1(t) = \left[\frac{v_s(t)}{2} \right] \left(1 + \frac{\Delta R}{R} \right). \quad (3.75)$$

Hence, the output of the DA is

$$v_o'(t) = K_d \left(\frac{\Delta R}{2R} \right) V_s \sin(2\pi 400t). \quad (3.76)$$

The center frequency of the BPF is 400 Hz, so $v_o(t) = v_o'(t)$. Thus, the mean-squared signal output is

$$S_{\text{out}} = \left(\frac{V_s^2}{2} \right) \left(\frac{K_d^2}{4} \right) \left(\frac{\Delta R}{R} \right)^2 \text{MSV}. \quad (3.77)$$

The MS noise at the system output is assumed to be due to the white thermal noise in the bridge resistors and the equivalent input short-circuit voltage noise, e_{na} , which is assumed to be white. The total MS output noise is found by adding up the MS noise contributions from all the noise sources. The equivalent Hz noise bandwidth is set by the BPF and, from Table 3.1, is $(2\pi f_n/2Q)$ Hz. Thus, the MS output noise voltage can be written as

$$N_{\text{out}} = \left(\frac{4kTR}{2} + \frac{4kTR}{2} + e_{\text{na}}^2 \right) K_d^2 \left(\frac{2\pi f_n}{2Q} \right) \text{MSV}. \quad (3.78)$$

The MS SNR at the output of the system is thus

$$\text{SNR}_{\text{out}} = \frac{(V_s^2/2)(K_d^2/4)(\Delta R/R)^2}{(4kTR + e_{\text{na}}^2)K_d^2(2\pi f_n/2Q)}. \quad (3.79)$$

We now calculate the $\Delta R/R$ required to give a MS output SNR of 10. We let $f_n = 400$ Hz, $Q = 5$, $R = 1000 \Omega$, $K = 1000$, $4kT = 1.66 \times 10^{-20}$, $e_{\text{na}} = 5$ nV RMS/ $\sqrt{\text{Hz}}$, $V_s = 4$ V peak. Using Equation 3.79 earlier, we find $\Delta R/R = 7.23 \times 10^{-8}$, or $\Delta R = 7.23 \times 10^{-5} \Omega$. Using Equation 3.76, we calculate that this ΔR will produce a 400 Hz sinusoidal output with peak value $V_o' = 1.446 \times 10^{-4}$ V.

As can be seen, AC operation of a Wheatstone bridge with a tuned output filter can result in considerable sensitivity. Normally, any slow changes in $\Delta R(t)$ would be extracted with a PSR following the BPF.

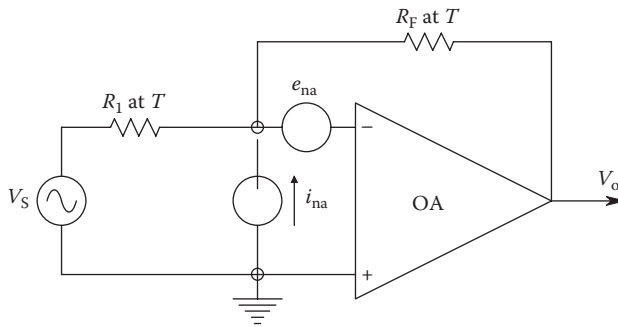


FIGURE 3.21
A noisy inverting OA amplifier.

3.8.5 Determination of the Conditions for Maximum Output SNR Given a Simple Inverting Op-Amp Amplifier with Known e_{na} and i_{na}

Figure 3.21 shows a simple OA inverter circuit having a signal source that lies well within the Hz noise bandwidth, B , for the circuit. Both R_1 and R_F are assumed to make white thermal noise. The MS signal at the output can be written as

$$S_{out} = \overline{v_s^2} \left(\frac{R_F}{R_1} \right)^2 MSV. \tag{3.80}$$

The total MS noise at the output is

$$N_{out} = \left[4kTR_1 \left(\frac{R_F}{R_1} \right)^2 + (i_{na}^2 + 4kTG_F) R_F^2 + e_{na}^2 \left(1 + \frac{R_F}{R_1} \right)^2 \right] B. \tag{3.81}$$

After some algebra, we can write the MS output SNR as

$$SNR_{out} = \frac{\overline{v_s^2}/B}{4kTR_1 (1 + R_1/R_F) + e_{na}^2 (1 + R_1/R_F)^2 + i_{na}^2 R_1^2}. \tag{3.82}$$

Clearly, from Equation 3.82, we see that the SNR is larger for high ratios of R_F/R_1 , that is, for high signal gains. Also, the SNR_{out} is larger for small R_1 . Small R_1 means, of course, that V_s sees a low effective input resistance (i.e., R_1). A practical limit on the gain is set by the necessary closed-loop bandwidth, which will be approximately $f_T(R_1/R_F)$ Hz. Good low-noise design often involves compromise of other design parameters.

3.9 Modern, Low-Noise Amplifiers for Use in Instrumentation Signal-Conditioning Systems

In the past 10 years, as the result of advances that have been made in transistor fabrication technology and in electronic circuit design, we have seen a rise in the number of available, low-noise, IC amplifiers suitable for use in signal conditioning in measurement systems.

TABLE 3.2

Partial Listing of Some Commercially Available, Low-Noise OAs and IAs Having Low Noise Characteristics Suitable for Instrumentation Signal Conditioning Applications

Some Low-Noise Amplifiers							
Amp. Model	Type	e_{na}	i_{na}	Low Freq V_n	R_{in}	f_T (MHz)	A_{v0}
AD8597	OA/BJT	1.1	2.4 pA/ $\sqrt{\text{Hz}}$	76 nVppk	—	10	116 dB
AD743	OA/FET	3.2	6.9 fA/ $\sqrt{\text{Hz}}$	0.38 μV ppk	$10^{10} \Omega$ (DM) $3 \times 10^{11} \Omega$ (CM)	4.5	4×10^3
OP27 (PMI)	OA/BJT	3	0.4 pA/ $\sqrt{\text{Hz}}$	0.08 μV ppk	6 M	5	10^6
OP61AP (PMI)	OA/BJT	3.4	0.8 pA/ $\sqrt{\text{Hz}}$	—	—	200	4×10^5
NE5532	OA/BJT	5.0	0.7 pA/ $\sqrt{\text{Hz}}$	—	300 k Ω	10	2.2×10^3
OPA111 (BB)	OA/FET	6	0.4 fA/ $\sqrt{\text{Hz}}$	1.6 μV ppk	10^{13} (DM) 10^{14} (CM)	2	125 dB
MAX410	OA/BJT	1.5	1.2 pA/ $\sqrt{\text{Hz}}$	—	20 k Ω (DM)	28	122 dB
LT1028	OA/BJT	0.85	1 pA/ $\sqrt{\text{Hz}}$	35 nV ppk	20 M (DM) 300 M (CM)	75	7×10^6
INA163 (BB)	IA/BJT	1.0	0.8 pA/ $\sqrt{\text{Hz}}$	—	60 M Ω (DM) 60 M Ω (CM)	3.4 kHz	$1-10^3$
INA103 (BB)	IA/BJT	1.0	2.0 pA/ $\sqrt{\text{Hz}}$	—	60 M Ω (DM) 60 M Ω (CM)	6	$1-10^2$
AD624	IA/BJT	4	60 pA ppk (0.1–10 Hz)	—	1G 10 pF (DM, CM)	25 MHz	$1-10^3$
MAX2641	UHF Amp	NF = 1.5 dB at 2.45 GHz		—	50 Ω , Rev. isol. = –24 dB	—	13.5 dB at 2.45 GHz
MAX2656	UHF Amp	NF = 1.9 dB at 1.96 GHz		—	50 Ω , Rev. isol. = –28 dB	—	13.5 dB at 1.96 GHz

Source: Northrop, R.B., *Analysis and Application of Analog Electronic Circuits to Biomedical Instrumentation*, 2nd edn., CRC Press, Boca Raton, FL.

Note: *Type* refers to whether the amplifier is an OA or an IA and whether it has a BJT or FET headstage. e_{na} is given in nV RMS/ $\sqrt{\text{Hz}}$ measured at 1 kHz; i_{na} is given either in picoamps (pA) or femtoamps (fA) RMS/ $\sqrt{\text{Hz}}$ measured at 1 kHz. *Low Freq.* V_n is the equivalent, peak-to-peak, short-circuit input noise measured over a standard 0.1 Hz to 10 Hz bandwidth; R_{in} is the input resistance; f_T is the unity-gain bandwidth in MHz, unless followed by (–3), in which case it is the high frequency –3 dB frequency; A_{v0} for OAs is their open-loop, DC gain; the useful gain range is given for IAs. (Manufacturers: AD, Analog Devices; BB, Burr-Brown; MAX, Maxim; PMI, Precision Monolithics; LT, Linear Technologies.)

Low-noise, as used in this section, shall mean amplifiers with e_{na} s less than 10 nV RMS/ $\sqrt{\text{Hz}}$. It is always risky to attempt to list and categorize low-noise amplifiers in a text because of (1) the danger of omitting one or more of the more obscure models and (2) by the time the list is read, new, better low-noise amplifiers will have been put on the market. Facing these risks, we do list in the following texts certain low-noise operational amplifiers and IAs that may have application in the design of low-noise signal conditioning systems.

In specifying the noise performance of IC amplifiers, manufacturers typically give the following information:

1. e_{na} measured in the *flat* portion of its spectrum.
2. The peak-to-peak or RMS output noise measured under short-circuited input conditions, over a specified, LF portion ($1/f$ region) of the $e_{na}^2(f)$ spectrum.
3. The value of i_{na} measured in its flat region. For some amplifiers, the manufacturers also give a plot of the spot noise figure contours vs. R_s and f .

The noise performance of an IC amplifier is largely set by the designer's choice of headstage transistors. Consequently, low-noise amplifiers may be broadly characterized by whether they use low-noise BJTs or FETs in their input circuits. Those using FETs generally have high input resistances (over $10^{12} \Omega$) and low i_{na} s (less than 1 fA RMS/ $\sqrt{\text{Hz}}$). Amplifiers with BJT headstages typically have lower input resistances and higher i_{na} s but, in some cases, exhibit e_{na} s an order of magnitude lower than the e_{na} s from JFET amplifiers. Table 3.2 in the following is a partial listing of the more widely used, low-noise, IC amplifiers; it is subdivided into instrumentation amps (IA) and OAs.

3.10 Coherent Interference and Its Minimization

In this section, we consider the other common form of unwanted input that degrades the resolution and accuracy of measurement systems, that is, periodic or coherent interference. Coherent interference, unlike random noise, has narrowband PDS, often with harmonic peaks at integral multiples of the fundamental frequency. A serious problem is created when the coherent interference spectrum has power in the signal frequency band. In the following sections, we will discuss common sources of coherent interference and various means of minimizing its input to a measurement system. The techniques of minimizing coherent interference often appear arcane to the inexperienced, but ultimately, all such techniques follow the laws of physics and circuit theory.

3.10.1 Sources of Coherent Interference

There are, broadly speaking, four major sources of coherent interference: (1) powerline frequency interference, often coupled to the measurement system by an electric field and/or a magnetic field (obvious sources of powerline interference are unshielded power transformers, poorly filtered AC-to-DC power supplies, and EM field pickup from unshielded, HV 60 Hz powerlines); (2) radio-frequency electromagnetic interference (RFI) coupled into the measurement system from local radio transmitters, poorly shielded computers, and other digital instruments having MHz clock frequencies (RFI may be picked up through the AC powerline, by magnetic induction in internal wiring loops or by magnetic

induction in the cables connecting sensors to signal conditioning systems); (3) of a more transient nature is the periodic noise coupled into a measurement system from sparking motor brushes, gasoline engine ignition systems, or HF powerline transients caused by silicon-controlled rectifier (SCR) and TRIAC switching in motor speed controls and ovens; although hardly periodic, the broadband EM pulses produced by lightning in thunderstorms are often picked up by EMI-sensitive electronic signal conditioning systems); and (4) in a hybrid electronic measurement system, coherent interference often may arise from within the system. Such interference may be the result of poor grounding and/or shielding practices allowing the digital portion of the system to *talk* to the low-level analog front end of the system. Internal coherent interference may also arise through poor decoupling of a common power source that supplies a low-level analog front end and an HF digital circuit.

3.10.1.1 Direct Electrostatic Coupling of Coherent Interference

To illustrate this phenomenon, we assume two conductors lie in close proximity. One conductor is at a sinusoidal potential, $v_1(t) = V_1 \sin(\omega t)$. There is a capacitance C_{12} between the two conductors, and conductor 2, which picks up the interference from conductor 1, has a resistance R_2 in parallel with a capacitance C_2 to ground, as shown in Figure 3.22A.

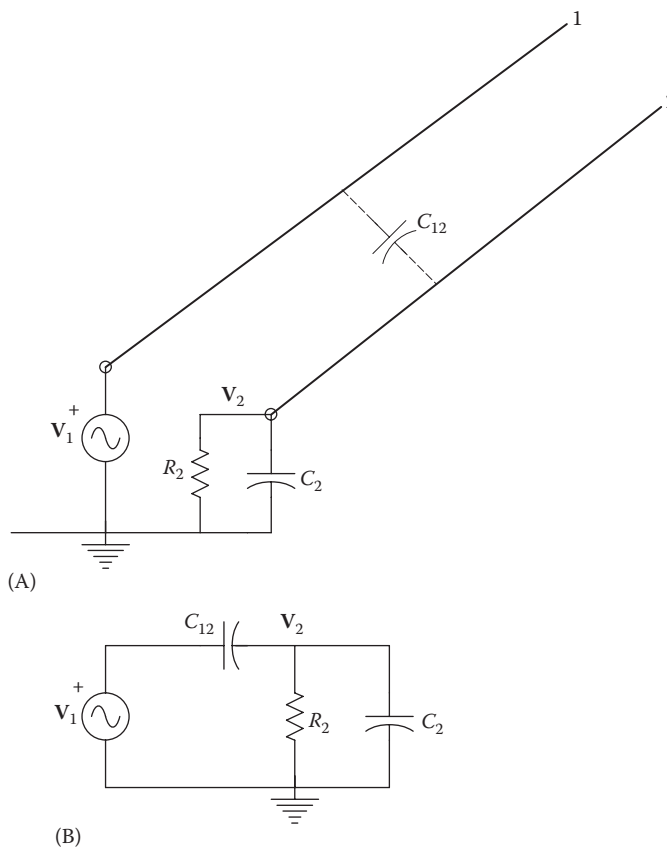


FIGURE 3.22

(A) Circuit modeling electrostatic coupling between two adjacent conductors. A sinusoidal voltage source is assumed. (B) Lumped-parameter equivalent circuit of (A).

The discrete component circuit describing this simple capacitive coupling situation can be redrawn as a simple voltage divider, shown in Figure 3.22B. We may write a frequency response function for the coherent interference induced on conductor 2:

$$\frac{V_2}{V_1}(j\omega) = \frac{j\omega C_{12}R_2}{j\omega(C_{12} + C_2)R_2 + 1}. \quad (3.83)$$

By inspection, we see that this is a simple, high-pass transfer function. C_{12} is typically on the order of single picofarads, C_2 may be in the tens of pF, and R_2 may vary widely, depending on the application. For a specific example, let $C_{12} = 2.5$ pF, $C_2 = 27.5$ pF, and $R_2 = 50 \Omega$. The break frequency is found to be 6.67×10^8 r/s. Hence, at frequencies below the break frequency,

$$V_2 = j\omega C_{12}R_2 V_1. \quad (3.84)$$

From the equations earlier, we gather the following lesson: To minimize V_2/V_1 , C_{12} and R_2 should be made small. C_{12} is minimized by arranging the wiring geometry so that high-level or output signal wires or printed circuit lines are far from sensitive, high-impedance, low-level circuit wiring. Insulated wire crossings, if necessary, should be at right angles. Also, sensitive wires should be shielded, if possible. Resistance R_2 to ground should be as low as possible. Often, the sensor output resistance is several kilohms, so the best protection for the headstage is to shield it and the wire connecting it to the sensor. As you will see in the following, what would appear to be a simple solution to the capacitive pickup of coherent interference by coaxial shielding is not that simple; details of shield grounding determine the effectiveness of the shield.

3.10.1.2 Direct Magnetic Induction of Coherent Interference

Figure 3.23A illustrates the basic mechanism whereby a time-varying magnetic field caused by current in conductor 1 induces an EMF in a second conductor. To model this happening with a lumped-parameter circuit, we assume that a mutual inductance, M , exists between the two conductors and that the conductors can be modeled by a transformer, as shown in Figure 3.23B. The EMF induced in wire 2 is simply

$$V_2 = j\omega M I_1. \quad (3.85)$$

Thus, the current induced in loop 2 is given by Ohm's law to be

$$\frac{I_2}{I_1}(j\omega) = \frac{j\omega M / (R_2 + R'_2)}{[1 + j\omega L_2 / (R_2 + R'_2)]}. \quad (3.86)$$

And the voltage observed across R_2 is simply $I_2 R_2$. L_2 is the very small self-inductance of wire 2. The break frequency of the pole in Equation 3.86 is generally quite large, so for most cases,

$$I_2 = I_1 [j\omega M / (R_2 + R'_2)]. \quad (3.87)$$

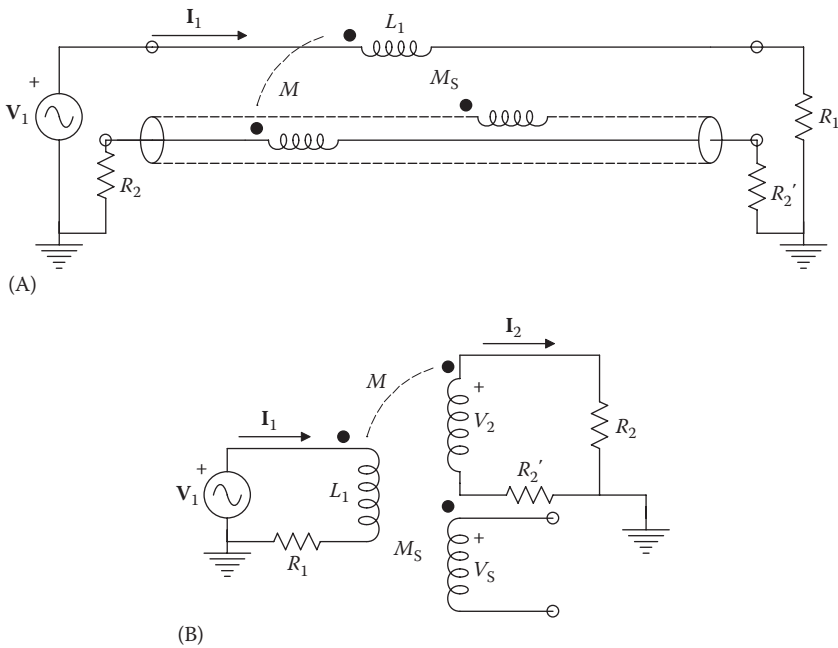


FIGURE 3.23
 (A) A circuit describing magnetic induction of coherent interference. The shield of the cable is not grounded.
 (B) Lumped-parameter equivalent circuit of (A).

In some cases of induced magnetic interference, a time-varying **B** field exists in space in the vicinity of a loop of area *A*, as shown in Figure 3.24. The **B** field can arise from an unshielded power transformer or, at HF, can be the result of a strong, radio-frequency EM field. By Faraday’s law, the EMF induced in the loop can be written as the integral of the vector dot product:

$$E = -\frac{d}{dt} \int \mathbf{B} \cdot d\mathbf{s}. \tag{3.88}$$

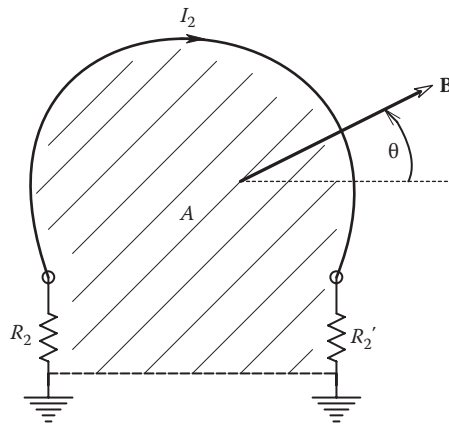


FIGURE 3.24
 Circuit modeling the induction of a coherent interference current, *I*₂, in a conductive loop enclosing a current-carrying conductor.

For the case of a **B** field that varies sinusoidally in time and that intersects the plane of the loop at an angle θ , we may write

$$\mathbf{E} = j\omega B A \cos(\theta), \tag{3.89}$$

and the interference current induced in the loop is simply

$$\mathbf{I}_2 = \frac{j\omega B A \cos(\theta)}{(R_2 + R'_2)}. \tag{3.90}$$

(Here, we have neglected the inductance in the loop.) Clearly, steps taken to reduce **B** and *A* will reduce the pickup of magnetic coherent interference.

3.10.1.3 Ground Loops

Whenever a ground conduction path is forced to carry a power supply return current that has a significant HF component added to the DC, and this ground path is also used in common for a low-level signal ground, it is possible for the AC power component of ground current to develop a small voltage across the small but finite impedance of the ground path. This situation is illustrated in Figure 3.25. There are, of course, causes other than power supply currents that cannot make the ground potentials of two subsystems be isopotential. The fact that two ground points are not at exactly zero potential means that an unwanted ground-loop current can flow in a coaxial cable shield that is grounded at each end to a different chassis or enclosure. The ground-loop current in the shield can induce coherent interference on the center conductor. Figure 3.26 illustrates how a difference in ground potential can add a coherent noise voltage directly to the desired signal. Here, we assume that $R_{IN} \gg R_S \gg R_{C1}, R_{C2}, R_{SH}$. The coherent interference voltage from $V_{GL(OC)}$ appears at the amplifier's input as V_N and may be shown to be equal to

$$V_N = V_{GL(OC)} \frac{R_{C2}}{R_{C2} + R} \frac{R_{IN}}{R_{IN} + R_S}. \tag{3.91}$$

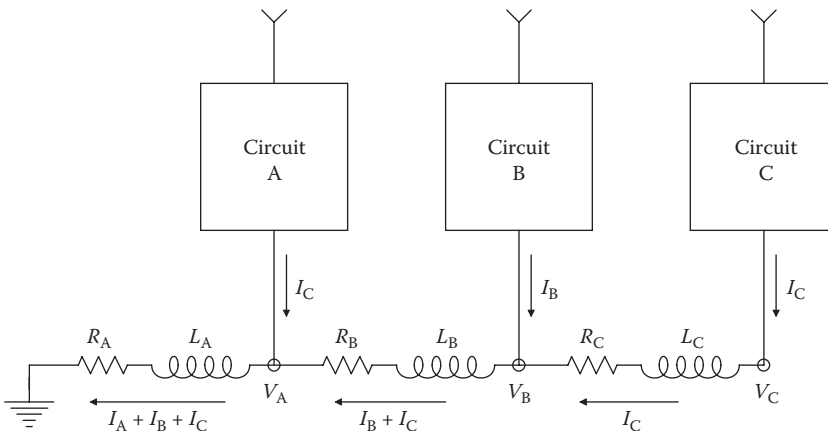
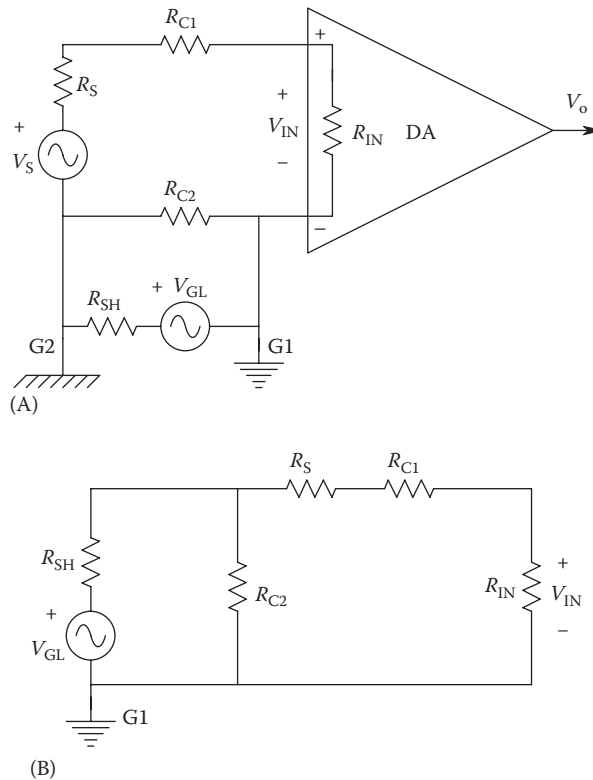


FIGURE 3.25 Circuit illustrating how ground current of one circuit can affect the ground potential of other circuits sharing the ground line. This condition is called a ground loop.

**FIGURE 3.26**

(A) Circuit showing how a ground-loop voltage, V_{GL} , can affect the input to an amplifier. (B) Equivalent circuit illustrating the two voltage dividers acting on the ground-loop voltage.

The numerical value of the product of the two voltage divider terms depends mostly on the values of R_{C2} and R_{SH} and is typically about $1/5$. Thus, we see that an unwanted voltage, $0.2V_{GL(OC)}$, appears in series with the desired signal, V_S , at the amplifier's input as the result of a ground loop and grounding the coax's shield at both ends.

3.10.2 Cures for Coherent Interference

As described earlier, there are many sources of coherent interference; the means to reduce a given type of interference depends on the physical mechanism(s) by which it is coupled into the measurement system. Comprehensive treatments of the causes of, and means for, reducing coherent interference may be found in the texts by Ott (1976) and Barnes (1987).

3.10.2.1 Powerline Low-Pass Filters

In many cases, the cure for coherent interference lies in preventing it from escaping from a coherent noise source such as a computer or digital system, an electronic system with a switching power supply, or a radio transmitter. Obviously, in the latter case, there must be EM energy radiated from an antenna; however, great care must be taken in all the systems mentioned earlier to keep the coherent interference from entering the power mains where it can affect other line-powered instruments. Powerline filters are generally multistage, LC

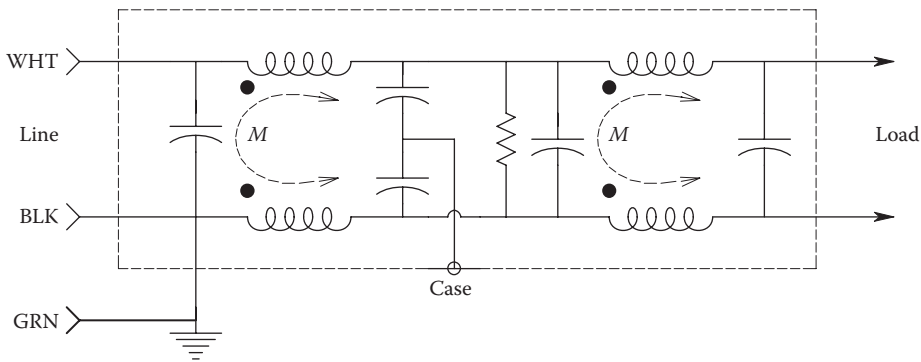


FIGURE 3.27 Schematic of a typical powerline LPF, designed to keep HF coherent interference on the powerline out of a system and to prevent HF interference arising within the system from escaping on to the powerline.

LPFs that attenuate frequencies from about 10 kHz up to about 40 MHz. Maximum attenuation at about 30 MHz ranges from -40 to -65 dB, depending on the number of LC stages in the filter. A typical line filter’s schematic is shown in Figure 3.27. Powerline filters are effective in keeping HF coherent noise from entering a system on the powerline, as well as keeping noise generated internally from escaping from the system.

HF interference on powerlines can be described in two forms: CM and DM interference. In CM interference, the unwanted voltage has the same value on both the black and white wires with respect to ground (green wire). The CM interference voltage can be written as

$$V_{i(\text{CM})} = \frac{V_{iB} + V_{iW}}{2}. \tag{3.92}$$

The DM interference voltage is given by

$$V_{i(\text{DM})} = \frac{V_{iB} - V_{iW}}{2}. \tag{3.93}$$

HF powerline filters are designed to attenuate both $V_{i(\text{CM})}$ and $V_{i(\text{DM})}$. Powerline filters must be used in conjunction with a robust, grounded metal instrument case; the metal case of the powerline filter must have a solid, multipoint electrical connection with the grounded instrument case for best results.

One design trade-off in the application of powerline LPFs is that the capacitors used to attenuate CM noise inject a current into the green wire powerline ground. Generally this current must be kept less than 3.5 mA in nonmedical applications. Hence, the size of the line-to-ground capacitors and thus the amount of CM interference attenuation are limited by this ground current limit.

3.10.2.2 Transient Voltage Suppressors

Transient voltage suppressors are used to prevent HV, spikelike transients occurring on the power mains input to an instrument system from causing physical damage to system power supply components and system circuits or causing damage and anomalous results in associated computer equipment. There are many sources of powerline spikes.

Some spikes are periodic, occurring once or more every power frequency cycle. These spikes are generally of a biphasic nature and their voltages may be as much as 30–50 V above the instantaneous line voltage. They are typically several microseconds in duration and arise from SCR or TRIAC switching of inductive loads attached to the powerline, such as in motor speed controls. Such *SCR spikes* generally pose no safety problem but do constitute a major source of coherent interference if allowed to get into the signal ground path or the power supply outputs. SCR spikes are of course attenuated by line filters but in some cases may be large enough after filtering to pose a problem. Other powerline spikes occur randomly as the result of lightning strikes on powerlines or from infrequent switching of inductive loads attached to the powerline. Figure 3.28 illustrates the statistical occurrence of singular spikes on 120 V, 60 Hz powerline in the United States (Hey and Kram 1978). The large, singular, transient spikes are best reduced by the use of nonlinear circuit elements such as varistors or Zener diodes. Varistors have SS current–voltage curves given by the following relations:

$$I = G_{\text{OFF}}V \quad < V < V_i, \quad (3.94a)$$

$$I = KV^\alpha \quad V > V_i, \quad (3.94b)$$

$$I = G_{\text{ON}}V \quad I > I_m. \quad (3.94c)$$

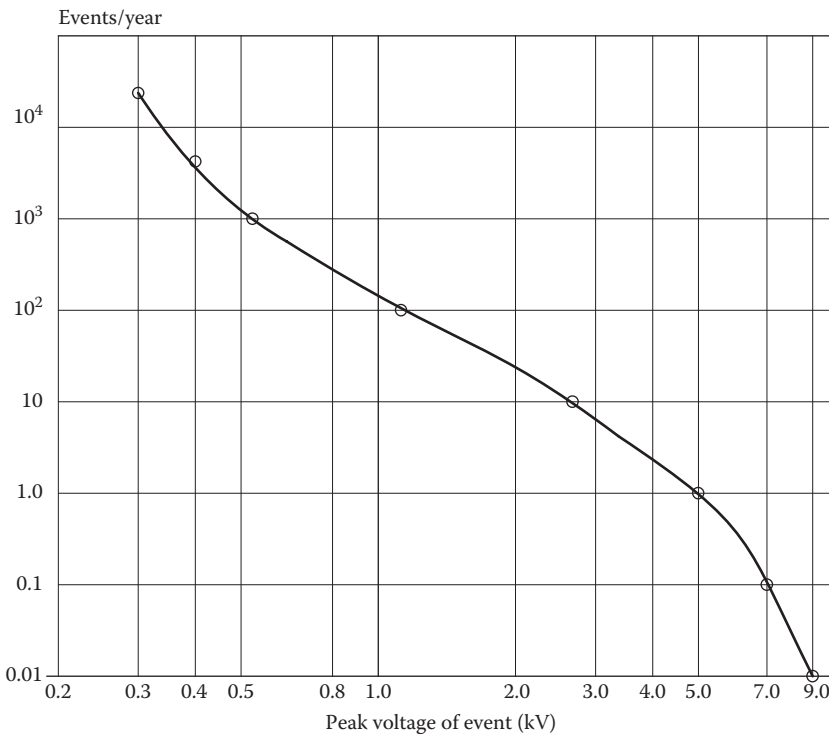
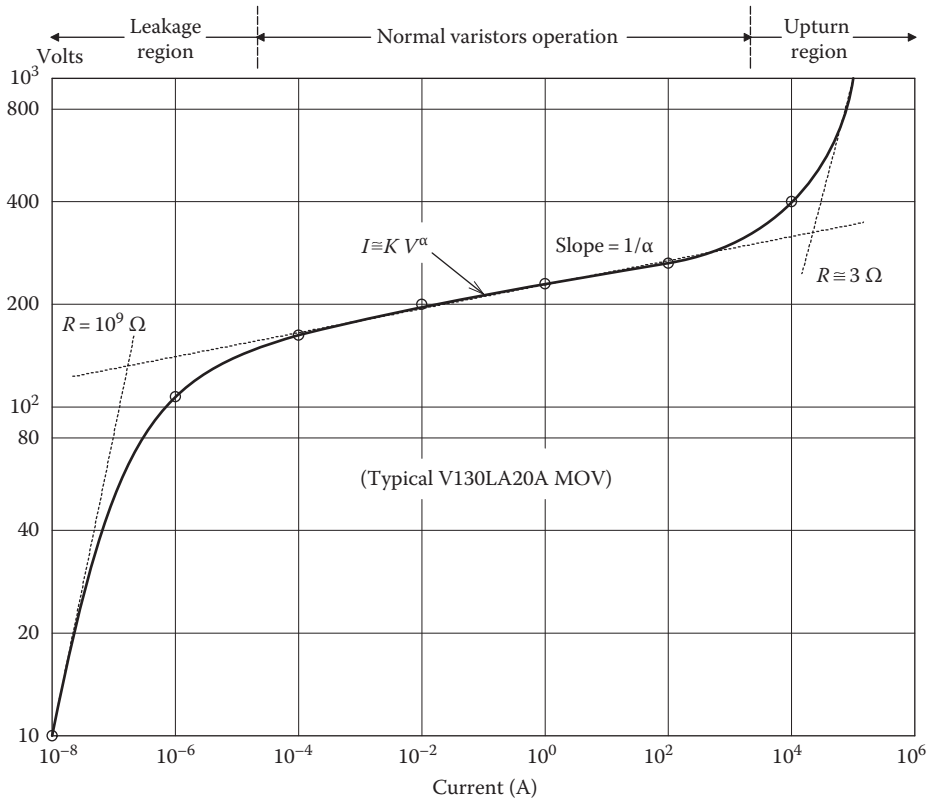


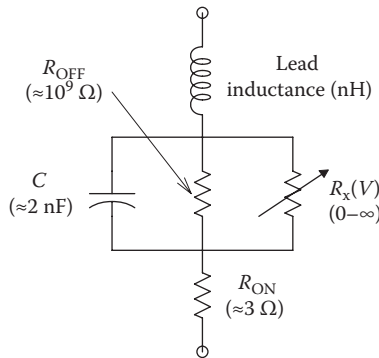
FIGURE 3.28

Statistical frequency of occurrence of spike transients on the 120 V, 60 Hz mains vs. the spike peak voltage. (Adapted from Fig. 1.7 in Hey, J.C. and Kram, W.P., eds., *Transient Voltage Suppressor Manual*, 2nd edn., General Electric Co., Auburn, NY, 1978.)

A typical varistor volt-ampere curve is shown in Figure 3.29A. An equivalent, lumped-parameter circuit for a General Electric (GE) MOV is shown in Figure 3.29B. Note that in a linear (Ohm's law) resistor, $\alpha = 1$. In varistors, α can range from 2 to 30, depending on the application and construction of the device. For efficient transient suppression, we generally want α to be as large as possible. Metal oxide varistors were originally developed by



(A)



(B)

FIGURE 3.29

(A) Volt-ampere (V - A) curve for a GE V130LA20A metal oxide varistor (MOV) spike transient clipper. Device is symmetrical; only first quadrant behavior is shown here. (B) Equivalent lumped-parameter circuit model for the GE MOV device of (A).

Matsushita Electric Co. and are sintered, polycrystalline semiconductor devices. GE manufactures and sells a family of zinc oxide and metal oxide varistors (MOVTM) under license from Matsushita. The α of GE MOVs ranges from 15 to 30 measured over several decades of surge current. GE MOVs are designed to operate at AC voltages from 10 to 1000 V and can handle current surges of up to 25 kA and energy capability of over 600 J in the larger units. Of primary consideration in the design of MOV transient suppression systems is the ability of the MOV device to conduct the transient current that the source can supply under conditions of overvoltage. MOVs are best used in applications where the current rise times (from the powerline) are longer than 0.5 μ s. (See Harnden et al. 1972.)

General Semiconductor Industries, Inc., manufactures and sells the TransZorbTM transient voltage suppressor. This silicon *pn* junction device acts similar to the MOV with the exception that its alpha is higher (35) and its response time is faster than MOV devices; the manufacturer claims 60 ps vs. 1–50 ns for varistors. The TransZorb has a relatively flat volt–ampere curve vs. varistors. See Figure 3.30. TransZorb AC operating voltages range from 5 to 340 peak volts; 5 V devices clamp at 9.6 V; 340 V devices clamp at 548 V.

Figure 3.31 shows a schematic for an extremely simple six-outlet, *power bar* or *line monitor power conditioner* typically used with personal computers and related equipment to protect them from line spike transients. One MOV-type varistor is used to protect each pair of line outlets. The power bar also has a switch, a neon pilot light, and an inexpensive circuit breaker for 15 A. MOVs or TransZorb type devices are also used inside sensitive equipment when the presence of a protective power bar cannot be guaranteed.

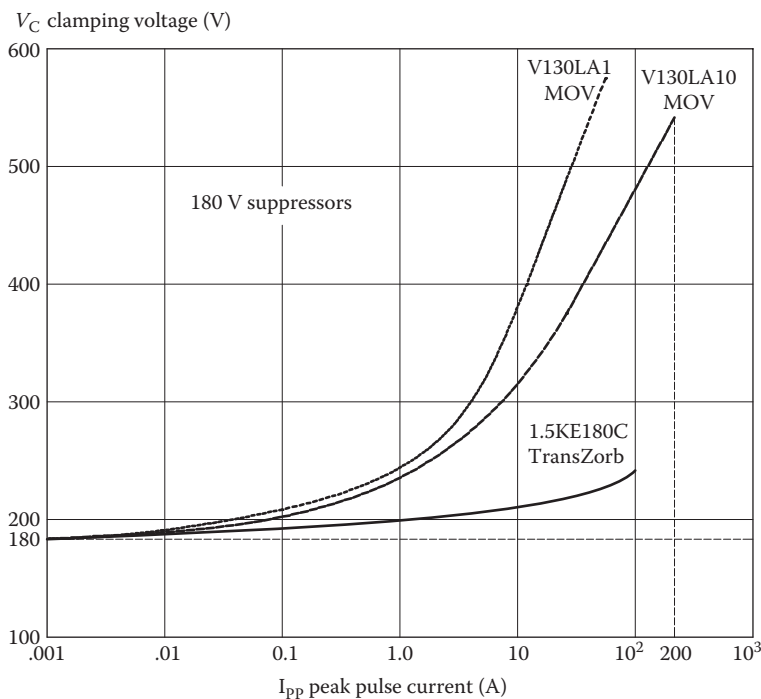


FIGURE 3.30

Comparison of the V - A curves of two GE MOVs and a TransZorb + *pn* junction transient suppressor made by General Semiconductor Industries, Inc., New York, NY. (Courtesy of GSI Inc.)

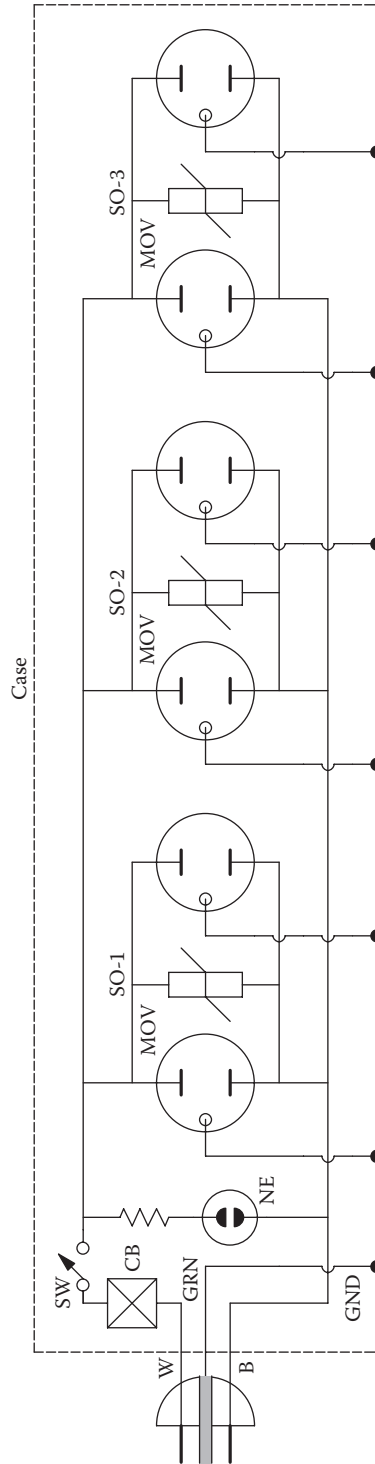


FIGURE 3.31 Schematic of a transient-protected, 6-outlet, *power bar*, such as those used with personal computer systems.

3.10.2.3 Coherent Interference Induced in Coaxial Cables by Magnetic Coupling

Shielded coaxial cables and twinaxial cables are widely used in instrumentation systems to couple sensors to preamplifiers and signal conditioning subsystems together. Unfortunately, coaxial cables can act as magnetic loop antennas under certain grounding conditions. Referring to Figure 3.32A, we see that an AC I_1 in an adjacent wire creates flux lines that link the coaxial cable and its shield, creating a mutual inductance between conductor 1 and the coaxial cable. Both ends of the shield are assumed to be grounded in this case. This system can be represented by the lumped-parameter circuit in Figure 3.32B. The open-circuit coherent interference voltage induced in the center conductor is simply written as

$$V_{2(OC)} = j\omega M_{12} I_1 - j\omega M_{S2} I_S \tag{3.95}$$

Because of the symmetrical coaxial geometry, it may be argued (Ott 1976) that the mutual inductance between the center conductor and the shield, M_{S2} , is simply the self-inductance of the shield, L_S . Also, we assume that $M_{1S} = M_{12}$. Now the current in the shield loop is given by

$$I_S = \frac{j\omega M_{12} I_1}{j\omega L_S + R_S} = \frac{(j\omega M_{12}/R_S) I_1}{j\omega L_S/R_S + 1} \tag{3.96}$$

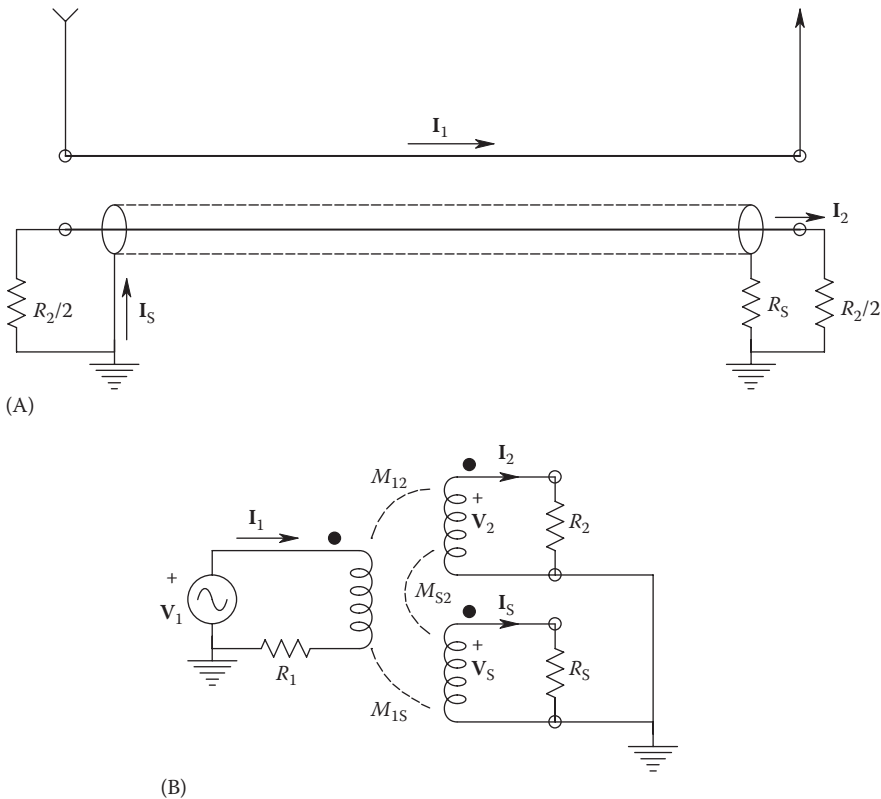


FIGURE 3.32 (A) A current-carrying conductor in proximity to a coaxial cable that has both ends of its shield grounded. This architecture is shown to minimize induced interference in the center conductor and to prevent magnetic interference from the cable. (B) Lumped-parameter equivalent circuit of the coaxial cable and wire shown earlier.

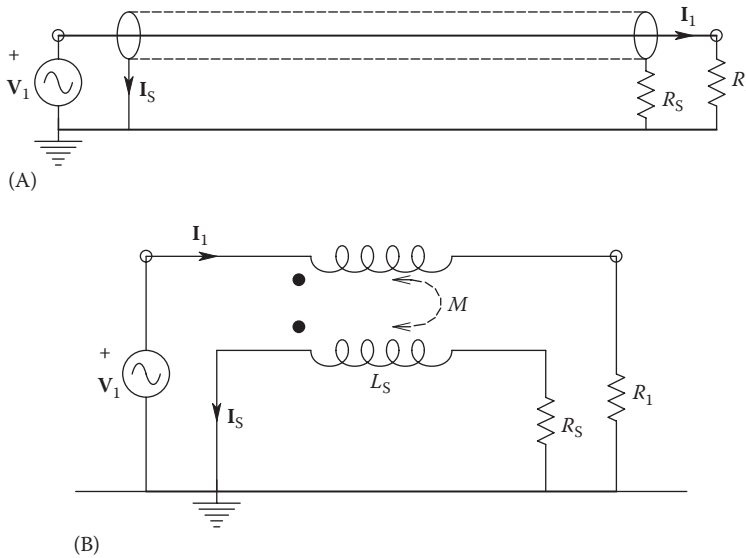


FIGURE 3.33

(A) A coaxial cable carrying AC I_1 on its center conductor. (B) Lumped-parameter equivalent circuit of the coaxial cable, showing the current induced in the shield loop, I_s . At distances greater than several cable diameters from the cable, the magnetic fields from I_1 and I_s effectively sum to zero.

The shield cutoff frequency, $R_s/2\pi L_s$, is generally on the order of 2 kHz for common RG-58C coaxial cable and is lower for other types of coax. If we assume that the radian frequency of the interference lies above the shield's break frequency, we find that $I_s \approx I_1$. Thus, from Equation 3.95, we see that $V_{2(OC)} \approx 0$. The lesson here is that to prevent magnetically induced coherent interference on a coaxially shielded line, both ends of the shield must be securely grounded. As you will see in the following, grounding both ends of the coax's shield can, in some cases, cause interference from a ground loop.

Just as we seek to minimize pickup of magnetically induced coherent interference, it is also important to prevent radiation of magnetic interference from a current-carrying coaxial conductor. This situation is illustrated in Figure 3.33A and B. At frequencies above the shield cutoff frequency, we have

$$I_s = \frac{j\omega L_s I_1}{j\omega L_s + R_s} = \frac{(j\omega L_s/R_s) I_1}{j\omega L_s/R_s + 1} \approx I_1. \quad (3.97)$$

Thus, the induced shield current is equal to the center conductor current and flows in the opposite direction. Hence, outside the coax, the \mathbf{B} fields from I_s and I_1 cancel vectorially, and magnetic interference is not generated.

3.10.2.4 Single Grounding of Coax Shields to Prevent Ground-Loop Interference

In the circuit describing ground-loop interference, Figure 3.34A, we can easily show, using Thévenin's theorem, that the ground-loop EMF, V_{GL} , produces a substantial interference voltage at the amplifier's input, given by

$$V_{CI} = V_{GL} \left[\frac{R_{C2}}{R_G + R_{C2}} \right]. \quad (3.98)$$

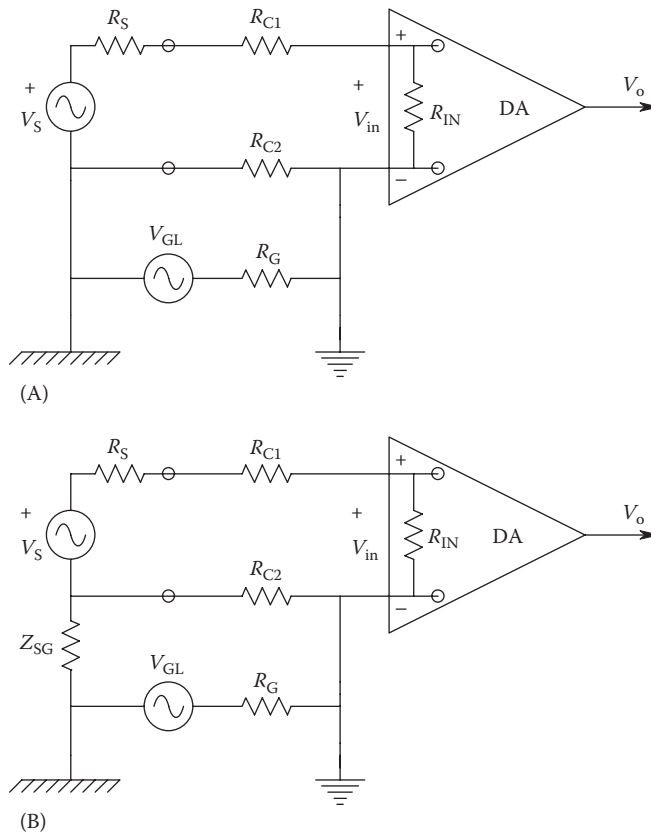


FIGURE 3.34

(A) A DA is used to condition the voltage of a grounded source (V_S , R_S). (B) In this case, the DA conditions the voltage from a floating source (Z_{SG} ohms above ground). The effect of V_{GL} is minimal.

We can easily examine the case where the input source is floating with respect to ground. This case is illustrated in Figure 3.34B. Because there is a very low conductance pathway between $V_{GL(OC)}$ and V_S , there is very little interference from the ground-loop EMF. In this case, the coherent interference voltage at the amplifier input is given by

$$V'_{CI} \approx V_{GL} \left[\frac{R_{C2}}{Z_{SG}} \right]. \tag{3.99}$$

Here it is obvious that $V_{CI} \gg V'_{CI}$, and we have assumed that $R_{IN} \gg R_S \gg (R_{C1}, R_{C2}, \text{ and } R_G)$. In this case, a DA or IA should be used to condition V_S , and a shielded, twisted-pair (*twinax*) cable is used to minimize electrostatic and magnetically induced interference.

3.10.2.5 Use of a Longitudinal Choke or Neutralizing Transformer to Attenuate Common-Mode Coherent Interference

Figure 3.35A illustrates the use of a toroidal ferrite magnetic core (or bead) to form a *longitudinal choke*. A longitudinal choke is used to reduce coherent interference caused by HF, CM interference picked up by wires connecting sensors to signal conditioning modules

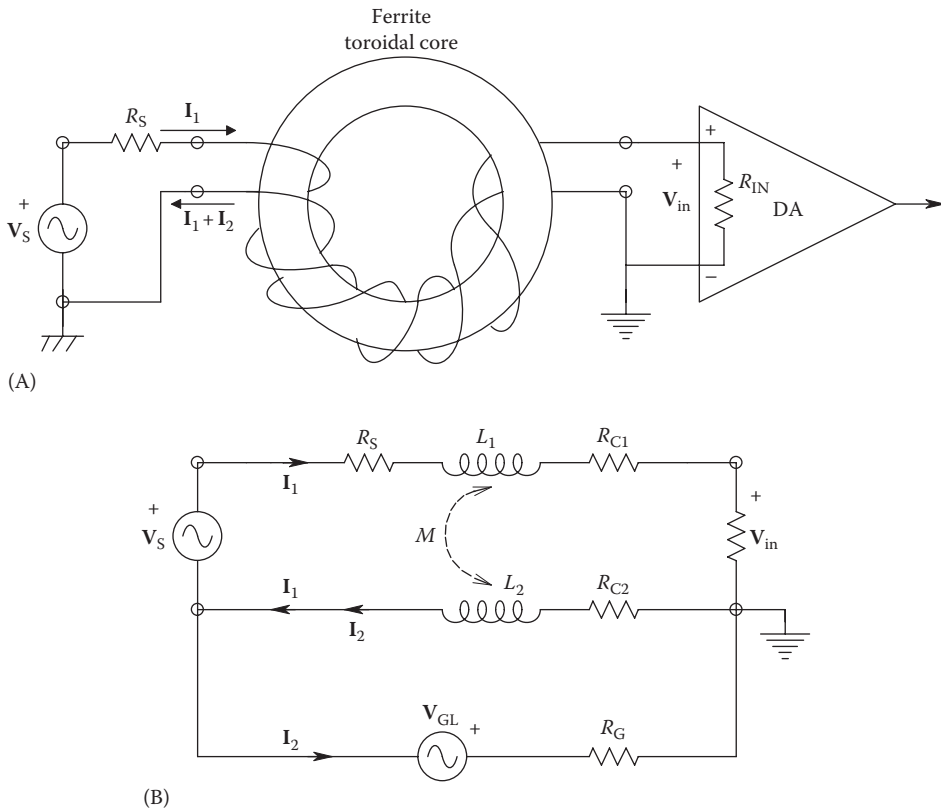


FIGURE 3.35 (A) A toroidal longitudinal choke between the source and the DA. (B) Equivalent circuit of the longitudinal choke.

or connecting subsystems in an instrumentation system. The longitudinal choke may be used with pairs of current-carrying wires, twisted pairs of wires, or coaxial cables. More than one pair of wires or coaxial cable may be wound on the same core without the circuits cross talking. In the simplest form, the wires are passed through the center of a cylindrical ferrite core to form a one-turn longitudinal choke. This geometry is often seen on accessory cables used with PCs. To examine the operation of the longitudinal choke, refer to Figure 3.35B. First, we use superposition to find the signal voltage across R_{IN} of the signal conditioning amplifier. Setting $V_{GL} = 0$, we write two loop equations. We assume that $L_1 = L_2 = M_{12} = M_{21} = L$:

$$V_S = I_1[R_S + R_{IN} + R_{C1} + R_{C2} + j\omega 2L] + I_2[R_{C2} + j\omega L] - j\omega 2MI_1 - j\omega MI_2, \quad (3.100a)$$

$$0 = I_1[R_{C2} + j\omega L] + I_2[R_G + R_{C2} + j\omega L_2] - j\omega MI_1. \quad (3.100b)$$

Canceling terms and using Cramer’s rule to solve the loop equations, we find

$$\Delta = (R_S + R_{IN})(R_G + R_{C2} + j\omega L_2). \quad (3.101)$$

And V_{IN} due to V_S is given by

$$\frac{V_{IN}}{V_S} = \frac{I_1 R_{IN}}{V_S} = \frac{R_{IN}(R_G + R_{C2} + j\omega L_2)}{(R_S + R_{IN})(R_G + R_{C2} + j\omega L_2)} \approx 1. \quad (3.102)$$

Here, we have assumed as before that $R_{IN} \gg R_S \gg R_{C1}, R_{C2}, R_G$. Of great significance is that the signal component at the amplifier input undergoes no HF attenuation due to the longitudinal choke.

Next, we set $V_S = 0$ and consider the amplifier input caused by the ground-loop voltage, V_{GL} . Again, writing loop equations,

$$0 = I_1[R_{IN} + R_S + R_{C1} + R_{C2} + j\omega 2L] + I_2[R_{C2} + j\omega L_2] - j\omega 2M I_1 - j\omega M I_2, \quad (3.103a)$$

$$V_{GL} = I_1[R_{C2} + j\omega L_2] + I_2[R_G + R_{C2} + j\omega L_2] - j\omega M I_1. \quad (3.103b)$$

Canceling terms, we again solve for $V_{IN} = I_1 R_{IN}$ using Cramer's rule:

$$\frac{V_{IN}}{V_{GL}}(j\omega) = \frac{R_{C2}}{R_{C2} + R_G} \frac{1}{[1 + j\omega L_2 / (R_G + R_{C2})]}. \quad (3.104)$$

From Equation 3.104, we see that the interference due to V_{GL} is attenuated and low-pass filtered by the action of the longitudinal choke.

3.10.2.6 Experimental Verification of Cabling and Grounding Schemes to Achieve Minimum Noise Pickup

Ott (1976) described the effectiveness of 12 different cabling schemes in minimizing the pickup of a 50 kHz magnetic field. The unwanted coherent interference was measured across a 1 M Ω resistor. The shielded cables were wound helically (3 turns, 7 in. diameter) inside the helical field-generating coil (10 turns of 20 ga. wire, 9 in. diameter, carrying 0.6 A at 50 kHz). Ten volts peak to peak was impressed across the field-generating coil. From his tests, Ott reported that 0 dB was 0.8 V. Figure 3.36 summarizes his results. We have arranged the cables in the order of increasing magnetic (and electric) shielding effectiveness. Note that the simple coaxial cable with ungrounded shield performed the poorest. The better performing cables had circuits in which the 100 Ω Thevenin source resistance was not tied to ground. Although Ott's evaluation was focused on evaluating the shielding from a magnetic field, his system also produced an electric field. Caution should be used in overgeneralizing his results. Many subtle factors contribute to the pickup of coherent interference, including circuit geometry, the frequency, and the spatial distributions of **E** and **B** fields.

Ott's results show that the use of shielded twinax cables with single grounds on the shield gave the best results (lowest pickup of coherent interference from a predominantly magnetic source). He attributed the -80 dB attenuation shown by the simple coaxial cable to be anomalous and due to the fact that this circuit had a smaller net loop area, A , than the twinax cables. This anomalous result underscores what often appears to be the arcane nature of coherent noise reduction techniques, which can appear to the inexperienced as being *arty*. However, the laws of physics apply to coherent noise reduction practices, as well as other branches of electrical engineering and instrumentation.

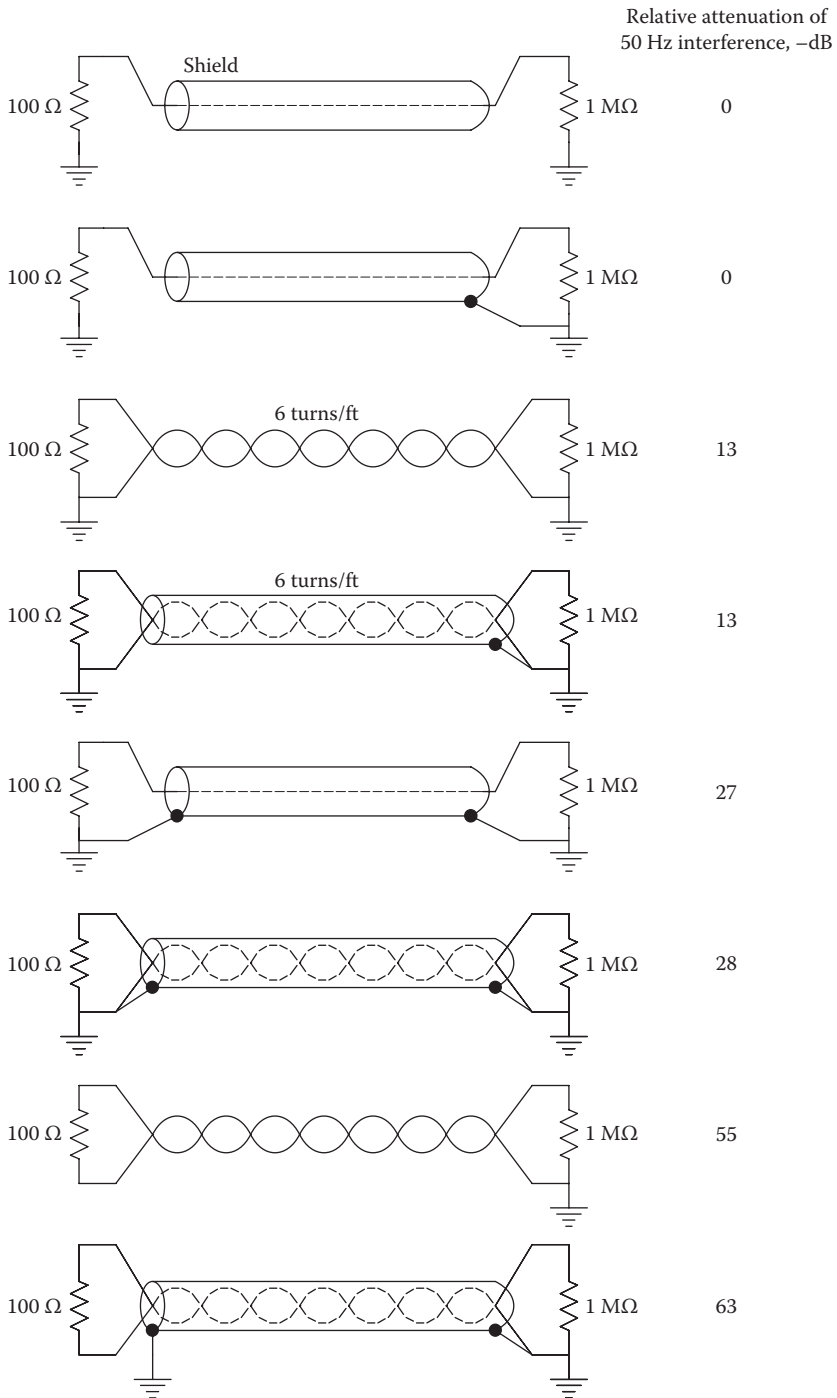


FIGURE 3.36
Relative effectiveness of 12 different cabling circuits subject to a 50 kHz magnetic field.

(continued)

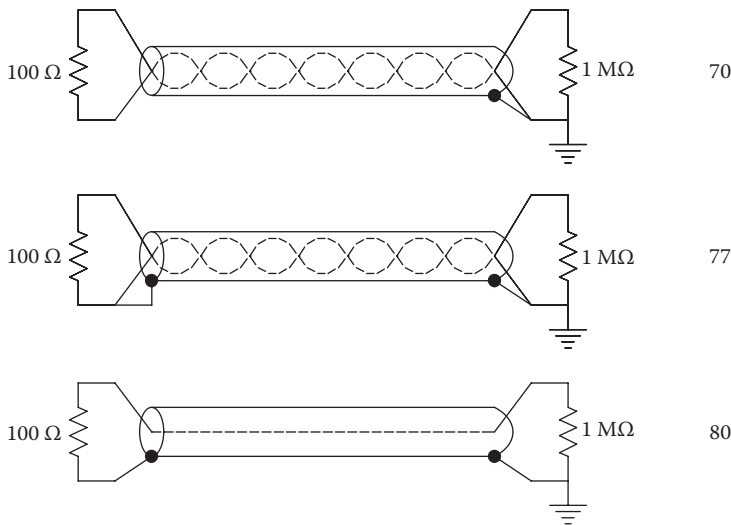


FIGURE 3.36 (continued)

Relative effectiveness of 12 different cabling circuits subject to a 50 kHz magnetic field. (Determined experimentally by Ott, H.W., *Noise Reduction Techniques in Electronic Systems*, John Wiley & Sons, New York, 1976.)

3.10.2.7 Circuit Grounding

Probably the single most important principle of reducing the pickup of coherent noise from internal sources in an electronic system is good grounding practice. Appreciate that typical ground currents contain a large DC component on which rides a coherent (AC) component that is the result of conditioning the desired signal or is from the HF switching of digital (logic) circuits. At HFs, ground wires appear as series $R-L$ circuits, whose impedance increases with frequency. If a circuit card contains both digital circuitry and low-level analog amplifiers, and has a single ground connection, it is almost a certainty that the analog system will pick up the coherent interference from the digital circuits' ground current flowing through the common ground impedance. Obviously, separate ground paths will help this situation. Figure 3.37 illustrates a good grounding and electrostatic shielding architecture. Note that there is a common tie point for all grounds at the power supply common terminal. This terminal may or may not be tied to the metal instrument case that is always tied to the green wire (powerline earth ground). It is important to note that separate grounds are used for low-level analog, high-level analog, digital, and inductive (arcing) circuits such as motors, solenoids, and relays. The ground wire itself should be of heavy gauge; if the + and -15 V DC supplies to a circuit card use #18 wire, then the ground should be #14 stranded wire.

3.10.2.8 Ferrite Beads and Feedthrough Capacitors

An effective way of attenuating unwanted coherent interference on signal and powerlines at frequencies above 1 MHz is through the use of one or more (seldom more than three) ferrimagnetic beads strung on the wire carrying the offending interference. Ferrite beads act as small chokes, making the wire upon which they are strung appear to be a series $R-L$ circuit. Because of their internal losses, some ferrite beads cause their wire to have a nearly constant (resistive) impedance of about 30–75 Ω from <1 to 100 MHz. Other beads are made to appear more inductive, and their Z increases linearly with frequency, going from 20 Ω at 5 MHz to 110 Ω at 100 MHz (Ott, Ch. 5 1976). The *type 64* material bead made by Amidon

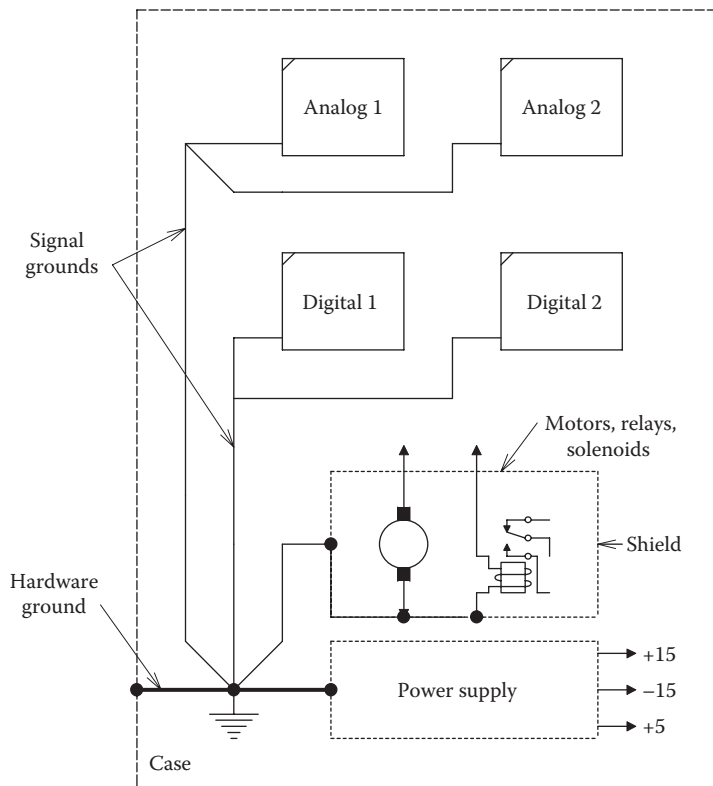


FIGURE 3.37
Good grounding and shielding practices in an electronic system.

Associates, N. Hollywood, CA, is designed to operate above 200 MHz. Ferrite beads make effective, well-damped, second-order LPFs when combined with appropriate bypass capacitors to ground. One widely used bypass capacitor is the feedthrough capacitor, used to decouple RF noisy sources from the power supplies. Typical feedthrough capacitors have cylindrical, threaded metal bodies that are mounted in the metal case or shield wall, and terminals at both ends to permit the DC to pass through. Capacitance from the center through conductor to the case can range from 0.5 to 5 nF, depending on the application.

In some cases, the use of ferrite beads will produce more harm than good; misuse of beads can cause unwanted, HF resonances and oscillations. Properly applied, ferrite beads and feedthrough capacitors can provide a simple and low-cost means of reducing coherent interference and parasitic oscillations.

3.10.2.9 Interruption of Ground Loops by the Use of Isolation Transformers and Photooptic Couplers

Isolation transformers are a simple, add-on means that can often improve measurement system SNR when used between analog signal conditioning subsystems. A typical isolation transformer used with 50 Ω coaxial cables has terminating and source impedances of 50 Ω and a frequency response of 25 Hz–20 MHz. There can be several thousand megohms impedance between the primary and secondary coax shields; hence, the CM EMF has little effect (see Equation 3.99) on the signal.

3.10.2.10 Photoptic Couplers

Photoptic couplers are nonlinear IC devices, primarily used in digital interfaces where it is desired to have complete ground and signal isolation between digital circuits. A typical optocoupler consists of an LED on a chip in optical but not electrically conductive intimacy with an output circuit that consists of a phototransistor or photodiode that is caused to conduct when illuminated by the input LED. The primary use of photoptic couplers is for isolated, high-speed, serial digital data transmission. A simple TTL logic circuit coupled with a photoptic coupler is shown in Figure 3.38A.

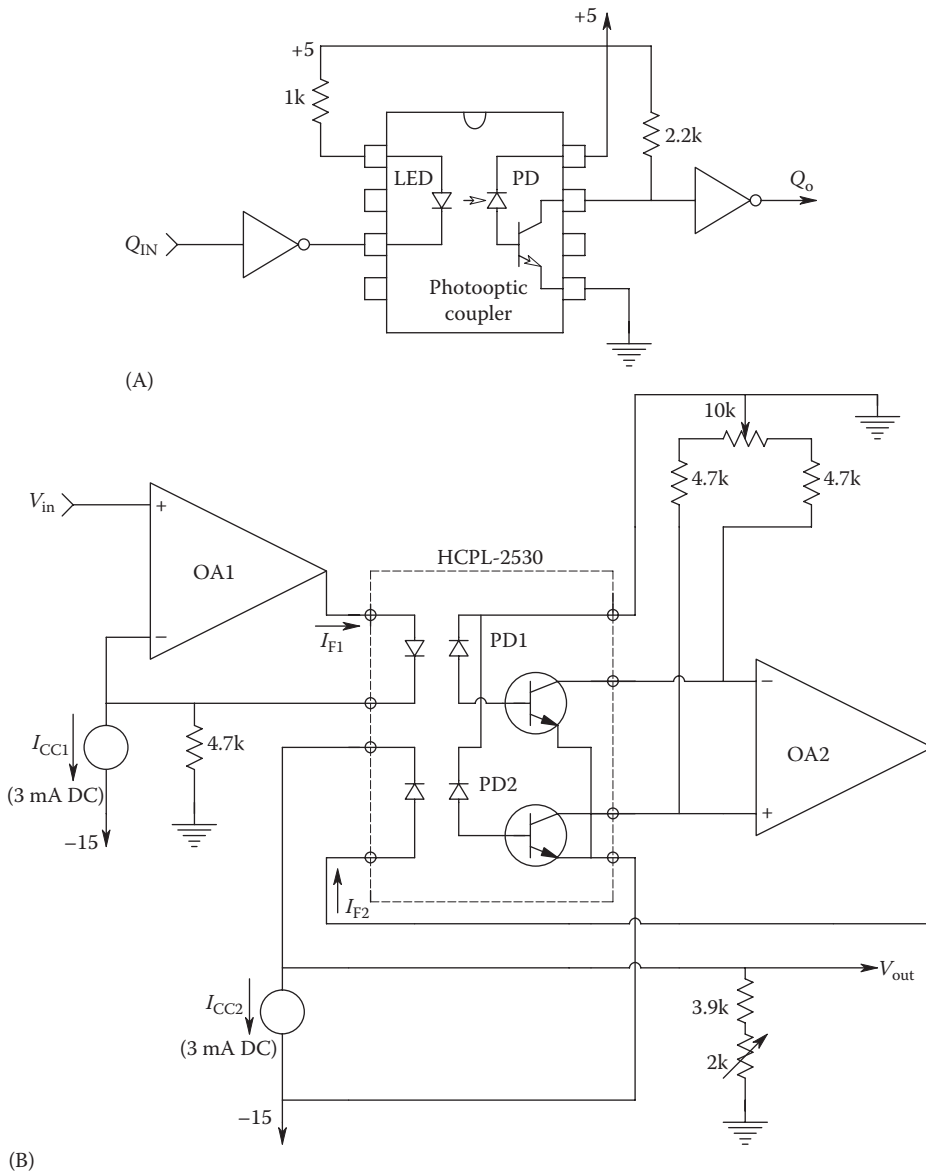


FIGURE 3.38

(A) A photoptic coupler is used to galvanically isolate a TTL digital signal line. (B) Use of a dual optocoupler in a feedback circuit to realize isolated, linear, analog signal transmission.

Photoptic couplers can be modified to give optically isolated, analog signal transmission by DC biasing the photodiode and phototransistor to operate in the middle of their dynamic regions and using negative feedback to reduce nonlinearity. Such linear photoptic coupler circuits can have SS bandwidths from DC to over 1 MHz. A feedback analog photoptic coupler circuit, adapted from a circuit in the HP/Agilent *Optoelectronics Designer's Catalog* (1996), is shown in Figure 3.38B. In this circuit, an Avago Technologies HCPL-2530 dual optocoupler is used. Feedback forces the AC signal current through D2 to track the input signal current in D1, effectively canceling out much of the nonlinearity inherent in the photodiodes and phototransistors.

3.10.2.11 Use of Guarding and Shielding to Reduce Common-Mode Coherent Interference

When extremely low-level signals are being measured, or when very high CM coherent noise is present, as in the case of certain biomedical applications, strain gauge bridge applications, or thermocouple measurements, the use of a guard shield can effectively reduce the amplifier's input capacitance to ground, hence coherent interference. Figure 3.39 illustrates a conventional differential instrumentation amplifier (DIA) connected to a Thevenin source (V_s, R_s), by a shielded, twinaxial cable. The cable shield is tied to the source ground and not the power supply ground. A large, coherent, CMV, V_{GL} , exists between the power ground and the source ground. If no currents flow in the twisted-pair wires, then V_{GL} appears as the CMV at the amplifier's input and is rejected by the amplifier's CMRR; there is no DM component of V_{GL} . However, the input capacitances of the DIA allow currents to flow. Note that there is a capacitance C_{1S} between the HI side of the source and the cable shield and conductor 2. Because of asymmetry in the circuit, a DM component of V_{GL} appears across the amplifier's input. This DM interference component may be calculated: by inspection of the circuit of Figure 3.39, $V_- = V_{GL}$. V_+ is given by the reactive voltage divider:

$$V_+ = V_{GL} \frac{1/j\omega C_{1G}}{[1/j\omega C_{1G} + 1/(j\omega C_{1S} + G_s)]} = V_{GL} \frac{j\omega C_{1S} R_s + 1}{j\omega(C_{1S} + C_{1G}) + 1}. \tag{3.105}$$

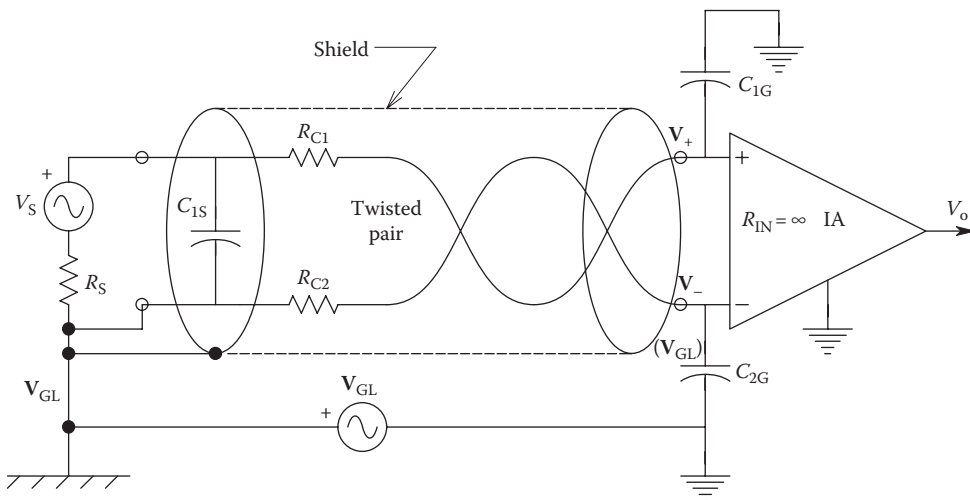


FIGURE 3.39 An IA connected to a Thevenin source by twinax shielded cable.

The phasor differential interference voltage at the DA's input is simply

$$V_d = \frac{V_+ - V_-}{2}. \tag{3.106}$$

Substituting from the relations earlier, we find

$$V_d = -V_{GL}(j\omega C_{1G}R_S) \quad \text{for } \omega \ll \omega_b = \frac{1}{[R_S(C_{1G} + C_{1S})]} \text{ r/s} \tag{3.107a}$$

$$= -\left(\frac{V_{GL}}{2}\right) \frac{C_{1G}}{C_{1G} + C_{1S}} \quad \text{for } \omega \gg \omega_b. \tag{3.107b}$$

For a numerical example, let $R_S = 1 \text{ k}$, $C_{1G} = 20 \text{ pF}$, $C_{1S} = 280 \text{ pF}$, and $\omega = 377$ (60 Hz ground-loop voltage).

Assume $R_{IN} = \infty$ and $R_{C1} = R_{C2} = R_{GL} = 0$. f_b is found to be 531 kHz, and from Equation 3.107a, the magnitude of the DM interference component is

$$V_d = V_{GL}(3.77 \times 10^{-6}). \tag{3.108}$$

This DM interference can seriously degrade output SNR when V_S is in the single microvolt range and V_{GL} is in the hundreds of millivolts. To reduce the interference V_d , we take steps to reduce the input capacitances to ground. This can be accomplished by using an IsoA (see Section 2.4.2) in conjunction with a *guard shield* and external shield, as shown in Figure 3.40. Here, the guard shield is tied to the amplifier's ground and is made to float with respect to *true ground* as a consequence of the IsoA's RF power supply. The current in the ground loop, and hence the magnitude of the DM interference voltage, is reduced at least 20-fold because the capacitance from either amplifier input to true ground (C_{1G} , C_{2G}) is now reduced by the external shield and is on the order of 1 pF. The same voltage divider is used to calculate V_d .

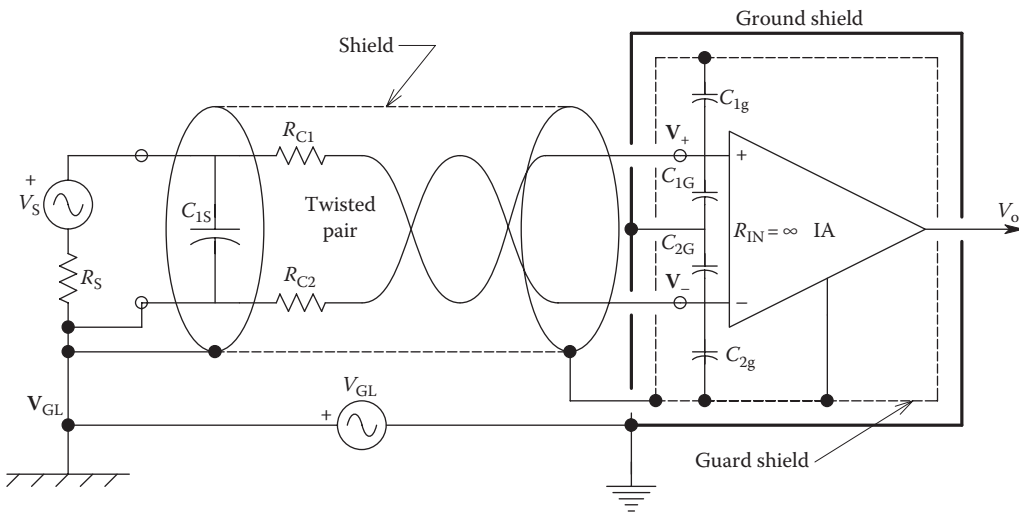


FIGURE 3.40 An IsoA is used with a guard shield and a ground shield to reduce ground-loop interference.

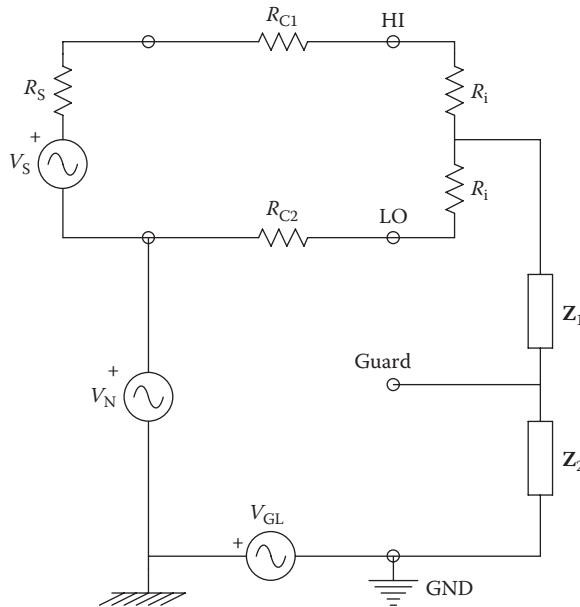


FIGURE 3.41 Equivalent input circuit of a voltmeter with guard and ground terminals. V_{GL} is the ground-loop voltage and V_N is another CM noise source.

Figure 3.41 illustrates a guarded meter’s input circuit. Note that the following analysis of the guarded meter circuit is also valid for a DA circuit in which the common-mode point (CP) is taken to the guard terminal and to the ground through impedances Z_1 and Z_2 , as shown in Figure 3.41. In this circuit, 4 m terminals are of interest: HI, LOW, GUARD, and GROUND. Here, as in the preceding analysis, an unwanted, coherent, CM, ground-loop voltage, V_{GL} , is present between true ground and the remote source ground, as well as some other, undesired, CMV associated with the source, V_N .

Five ways of connecting the guard of the guarded meter will be considered. The analysis of each case is simplified by calculating the input voltage at the meter ($V_{HI} - V_{LO}$) as a function of circuit parameters. In the first case, we will leave the guard terminal open-circuited. We assume that the input resistances of the meter or DA, R_i , are very large—at least $10^8 \Omega$. The resistances of the twisted pair wires, $R_{C1} = R_{C2} = R_C$, are about 0.1Ω . The source resistance is 1000Ω , and Z_1 and Z_2 are typically capacitive reactances of about $10^5 \Omega$. As can be seen from Figure 3.42, the differential input voltage to the meter, V_d , can be considered to be the unbalance voltage of a Wheatstone bridge whose four arms are R_C , $(R_S + R_C)$, R_i , and R_i . The voltage across the bridge, V_B , is given approximately by

$$V_B = (V_{GL} + V_N) \frac{R_i/2}{R_i/2 + Z_1 + Z_2} = (V_{GL} + V_N). \tag{3.109}$$

Again, taking into account the sizes of the resistors in the circuit, it is easy to show that V_d is given by

$$V_d = (V_{GL} + V_N) \left(\frac{-R_S}{R_i} \right) = (V_{GL} + V_N)(-10^{-5}). \tag{3.110}$$

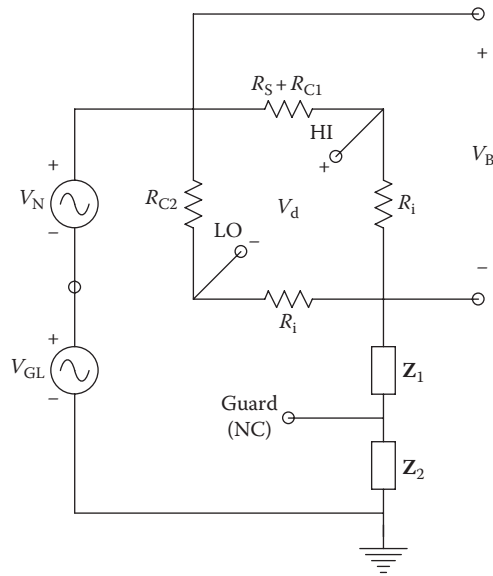


FIGURE 3.42 Equivalent input circuit of the voltmeter when the guard is not used. $V_s = 0$. Note the bridge architecture.

In the second case, we assume that the guard is incorrectly connected to the source’s ground. This connection effectively shorts out V_{GL} ; however, V_N still produces a DM interference voltage at the meter input. The equivalent circuit for this case is shown in Figure 3.43. Again, a bridge circuit is formed with an unbalance voltage, V_d , of

$$V_d = V_N(-10^{-5}). \tag{3.111}$$

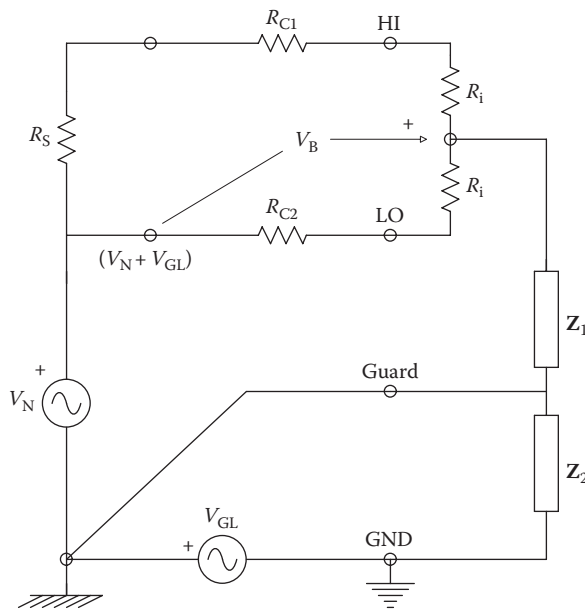


FIGURE 3.43 Equivalent input circuit of the voltmeter when the guard is incorrectly connected to the source ground.

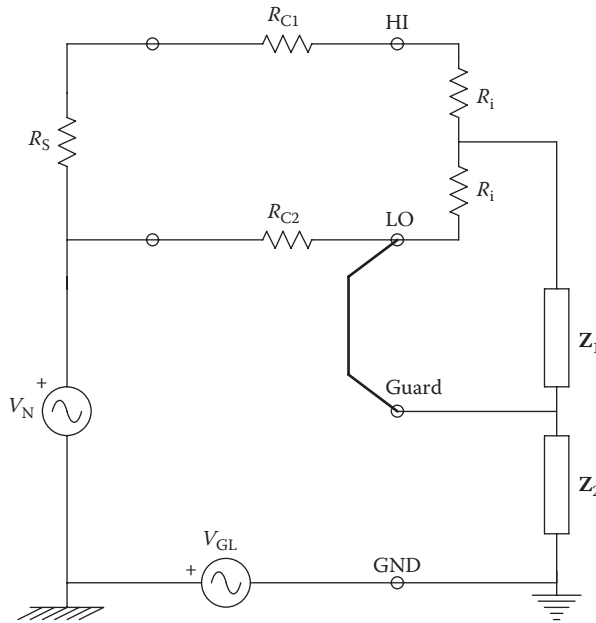


FIGURE 3.44 Equivalent input circuit of the voltmeter when the guard is again incorrectly connected to the LO (-) input terminal.

In the third case, shown in Figure 3.44, the guard is *incorrectly connected* to the meter’s LOW terminal.

Here, it is easy to see that

$$V_d = (V_{GL} + V_N) \left[\frac{1 - Z_2}{Z_2 + R_C} \right] = (V_{GL} + V_N) \left(\frac{R_C}{Z_2} \right) = (V_{GL} + V_N) (10^{-6}). \quad (3.112)$$

Figure 3.45 illustrates the *fourth case* where the guard is *incorrectly connected* to the meter ground, shorting out Z_2 . The differential interference voltage is found from the bridge circuit to be approximately

$$V_d = (V_{GL} + V_N) \left(\frac{-R_S}{R_I} \right) = (V_{GL} + V_N) (-10^{-5}). \quad (3.113)$$

In the fifth example, the guard terminal is *correctly connected* to the source low (i.e., the node between R_{C2} , R_S , and V_N). This situation is shown in Figure 3.46. Clearly in this case, $V_B = V_d = 0$, and there is no DM component at the meter input from the coherent noise voltages.

3.10.3 Summary of Techniques for Coherent Noise Reduction

The practice of the art and science of coherent noise reduction should begin with the elimination of local noise sources. Obviously, these are noise sources that have been identified and that we have authority over. These sources might include digital and computer equipment operating in the same immediate environment or interference caused by sparking motor

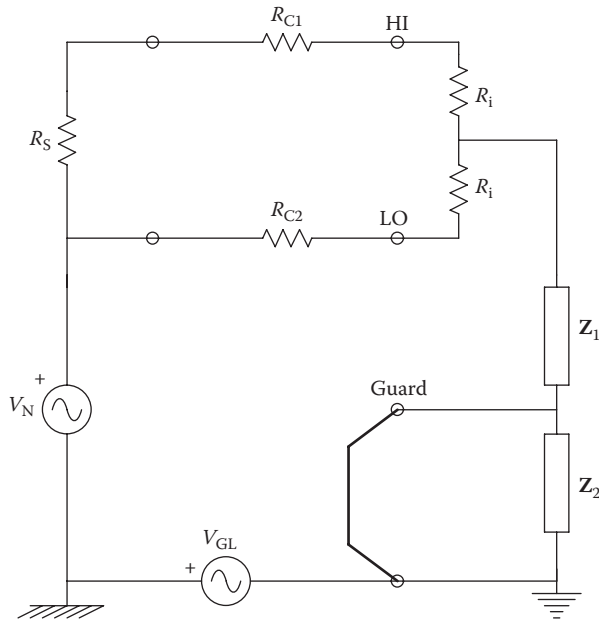


FIGURE 3.45 Equivalent input circuit of the voltmeter when the guard is incorrectly connected to the meter ground.

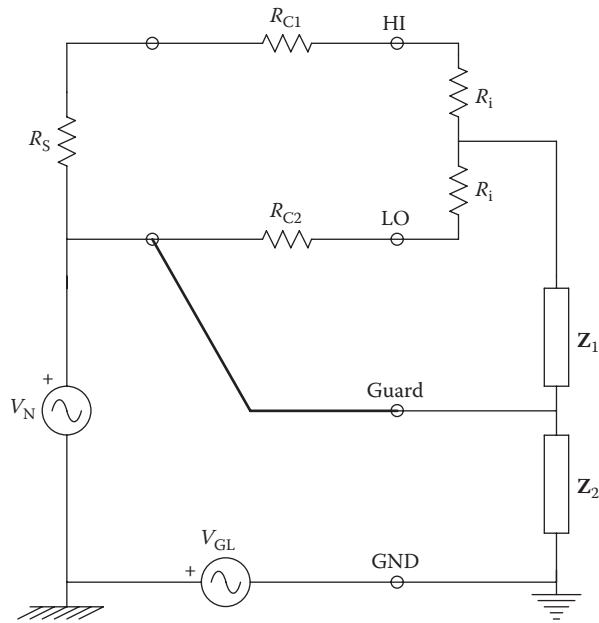


FIGURE 3.46 Equivalent input circuit of the voltmeter when the guard is *correctly* connected to the source low. There is no DM input component to the meter from V_N and V_{GL} .

commutators. The offending noise sources must be enclosed in appropriate, grounded, electrostatic shields. Their power leads must be low-pass filtered or bypassed as they leave the shield; the use of ferrite beads and feedthrough capacitors may be indicated. In addition, power leads should be twisted so there is minimum magnetic field created.

Inductive spikes induced on powerlines can be minimized by using varistors or diodes across switched relay and solenoid coils. Both ends of coaxial cables carrying RF should be grounded to minimize the radiation of magnetic field interference. As we have seen earlier, there are many means of reducing the pickup of coherent interference in low-level analog signal conditioning systems. Lead dress is important; separate low-level leads from wires carrying high currents and voltages. Low-level leads should be twisted and routed near the grounded case. If twinaxial (twisted and shielded) cable is used, *it should be grounded at one end only* and insulated. Alternate grounds and signal leads on ribbon cables and in connectors. High-impedance, low-level leads should be kept as short as possible; this is especially true in wiring OAs. The summing junction should never have an exposed lead of more than 1/2 in. connecting components to it.

Often, the pickup of coherent interference can be traced to the magnetic component of an LF EM field. Such pickup may occur in spite of rigorous electrostatic shielding practice. In this case, a mu metal or other lossy, nonsaturating magnetic shield may be effective, as will be reducing the area and changing the spatial orientation of the pickup loop.

We have seen earlier that ground loops can be a serious source of interference in instrumentation systems. The effect of ground-loop EMFs often can be minimized by the use of photooptical couplers (for digital circuits), isolation transformers (for signals in the 20 Hz–200 kHz range), longitudinal chokes (to attenuate HF CM interference), differential IsoAs, and guarded inputs. Good grounding practice can significantly reduce ground loops. We have seen that a single ground point is the best practice (except for RF circuits). Individual, insulated grounds from subsystems should be brought separately to the common ground point.

Just as separate grounds are important for interference reduction, so is the use of decoupled, low-impedance, DC power supplies with proper power ratings. Dedicated, low-noise, analog power supplies for signal conditioning system headstages can help reduce hum and the cross coupling of interference from other stages.

3.11 Chapter Summary

In this chapter, we have presented a heuristic yet mathematically sound treatment of random noise in signal conditioning systems based on the one-sided PDS. Sources of random noise in active and passive circuits were considered: thermal (Johnson) noise in resistors, shot noise in BJTs, and noise in FETs. The dependence of transistor noise performance on DC operating point was stressed. The standard, two-generator noise source model for amplifiers was introduced, as well as the concepts of noise factor, noise figure, and output SNR as figures of merit with respect to noise.

The propagation of Gaussian noise through linear filters was introduced, and the concept of a filter's $\text{gain}^2 \times \text{bandwidth}$ was applied to finding the MS noise output of signal conditioning systems, given white input noise. We showed how a low-noise transformer can be used to increase output SNR when a signal source with a low source resistance drives the input of a noisy amplifier.

Five, practical, numerical examples of noise calculations in measurement system design are given. Home problems for this chapter also stress calculation of threshold input signal levels to get a specified output SNRs.

We presented in tabular form a list of commonly used, low-noise, IC amplifiers. Here, we noted that the lowest system e_{na} values are generally found in amplifiers having BJT input stages, hence relatively low input resistances and high bias currents, thus large i_{na} values. FET input amplifiers have slightly higher e_{na} values but offer much higher input resistances and lower DC bias currents, hence lower i_{na} values than do BJTs. The wise designer of low-noise signal conditioning equipment thus must make compromises between e_{na} , i_{na} , input impedance, bandwidth, and other performance factors when choosing an IC amplifier for headstage use.

In Section 3.10, we considered coherent interference, its sources, and some of the many ways to minimize its effect in instrumentation systems. Coherent interference can appear at the output of a signal conditioning system as the result of electrostatic or magnetic coupling to input leads. We also show that it can enter an amplifier through the power supply or through improper grounding. We discuss some of the more common means of mitigating coherent interference and analyze equivalent circuits relevant to those means.

Problems

- 3.1 A. Find the numerical value for the RMS noise voltage in a 100 Hz–20 kHz noise bandwidth made by a 1 M Ω resistor at 300 K. Use $4kT = 1.656 \times 10^{-20}$ at 300 K.
 B. The same resistor is shunted by a 23 pF capacitor to ground. Find an expression for $S_{NO}(f)$, the one-sided noise PDS across the RC circuit.
 C. Find the RMS noise voltage across the parallel RC circuit over $0 \leq f \leq \infty$. Note that

$$\int_0^{\infty} \frac{a dx}{a^2 + x^2} = \frac{\pi}{2}.$$

- 3.2 A low-level sinusoidal signal, $v_s(t) = V_s \sin(2\pi f_o t)$, is conditioned by a noisy amplifier with a gain of $K_V = 10^3$. The amplifier is followed by a noiseless LPF. The amplifier has an equivalent, white short-circuit input voltage noise root PDS of $e_{na} = 14 \text{ nV}/\sqrt{\text{Hz}}$. The 1.5 k Ω source resistor also makes thermal white noise at 300 K. $f_o = 100 \text{ Hz}$, $\tau = 0.01 \text{ s}$, $V_s = 10 \text{ }\mu\text{V}$ pk. See Figure P3.2.
 A. Find an expression for the RMS sinusoidal output signal and the V_o node. Evaluate numerically. You should consider the LPF gain at f_o .
 B. Find an expression for the total RMS output noise voltage, v_{onr} , at the V_o node. Use the $\text{gain}^2 \times \text{BW}$ tables.
 C. Find the filter time constant, τ_o , that will maximize the output RMS SNR. Calculate this maximum SNR.
- 3.3 A Wheatstone bridge is given sinusoidal excitation: $v_b = 7.071 \sin(2\pi 400t)$. The bridge resistors are $R = 330 \text{ }\Omega$ at 300 K. An ideal DA with a diff. gain $K_D = 10^3$ is used to condition the bridge output. The DA has a white $e_{na} = 15 \text{ nV}/\sqrt{\text{Hz}}$. Following the DA is

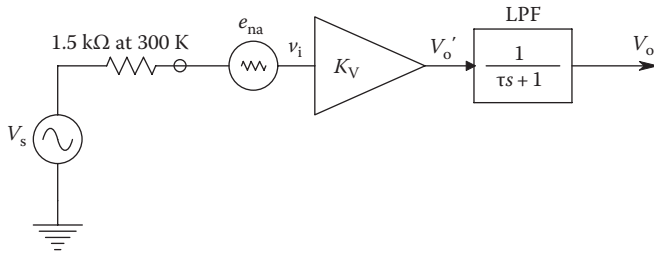


FIGURE P3.2

an ideal, noiseless, BPF with unity gain in the passband centered at 400 Hz and zero gain elsewhere. The passband has width $B = 10$ Hz.

Find the ΔR that will give an RMS sinusoidal output voltage at the V_o node equal to the RMS noise at V_o .

Neglect ΔR in calculating the output noise. See Figure P3.3.

3.4 A noise source has a one-sided PDS given by

$$S_n(f) = \frac{4 \times 10^{-6}}{f^2 + 10^{10}} \text{ msV/Hz}$$

- A. Find the value of the equivalent white noise, root PDS that approximates the LF portion of $S_n(f)$ mentioned earlier.
 - B. Find the total RMS noise voltage in $S_n(f)$.
- 3.5 The short-circuit input noise PDS of a certain JFET is given by $e_{na}^2(f) = (4kT/g_m)(1 + f_c/f)$ msV/Hz, where $T = 27^\circ\text{C}$, $g_m = 0.003$ S, and $f_c = 15$ Hz. Calculate the RMS noise voltage in $e_{na}^2(f)$ for $0.1 \leq f \leq 100$ Hz.
- 3.6 The resistor in a parallel $R-L-C$ circuit makes thermal white noise.
- A. Find an expression for the one-sided, voltage noise PDS at the v_o node.
 - B. Find an expression for the total RMS noise voltage at the v_o node over $0 \leq f \leq \infty$. Use the table of $\text{gain}^2 \times \text{BW}$ integrals.

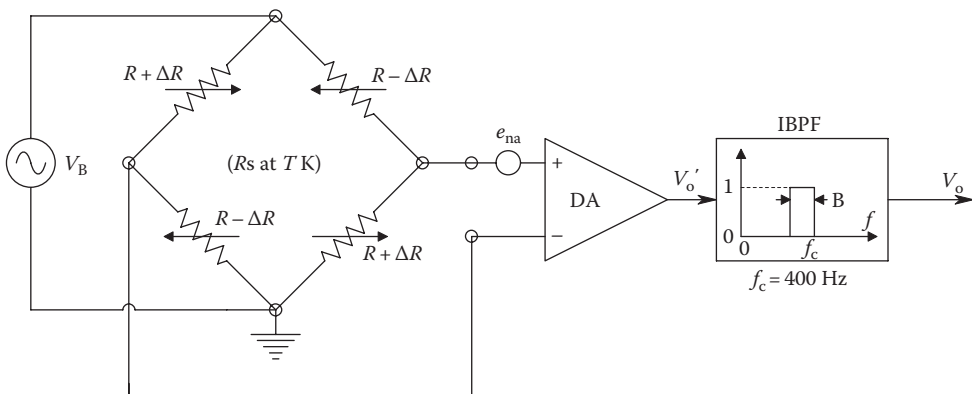


FIGURE P3.3

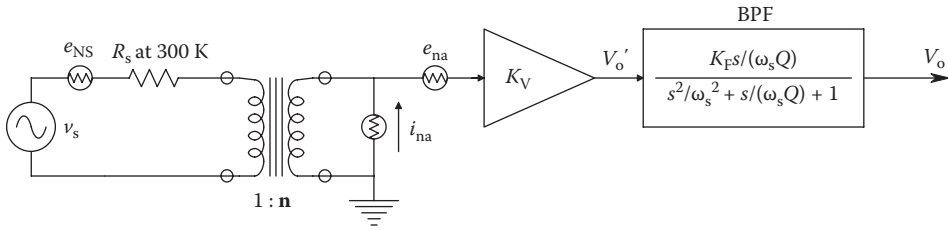


FIGURE P3.7

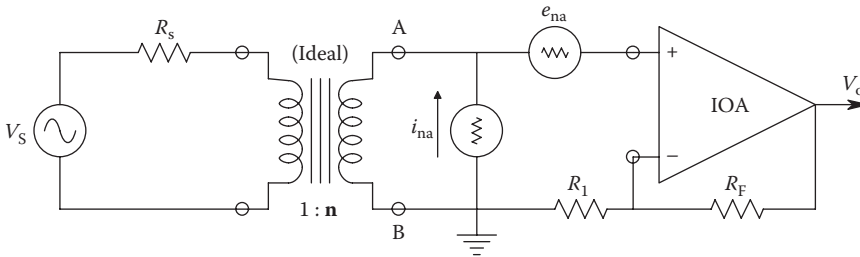


FIGURE P3.8

- 3.7 The system of Figure P3.7 is used to condition a small 100 Hz sinusoidal voltage accompanied by white noise before it is detected by a PSR/LPF. Assume the transformer is ideal and has a 1:n turns ratio. Assume the quadratic BPF is noiseless and is tuned to 100 Hz. A white environmental noise, e_{NS} , is recorded with $v_s(t) = V_s \sin(2\pi 100t)$.
- Give the Thevenin equivalent circuit the amplifier sees looking toward v_s , R_s .
 - Write an expression for the one-sided white noise PDS at the V_o' node, S'_{no} . Note that there are four noise sources in the system. The IA has a gain of $K_V = 10^3$.
 - Write an expression for the mean-squared signal voltage at the V_o node, v_{so}^2 .
 - Write an expression for the mean-squared noise voltage at the V_o node. Note that the gain² \times BW of the quadratic BPF is $K_F^2 [w_s/2Q]$.
 - Write an expression for the MS SNR at the V_o node, SNR_o . Put all n^2 terms in the denominator.
 - Derive an expression for the turns ratio, n_o , that will maximize SNR_o . Note that minimizing the denominator of SNR_o maximizes SNR_o .
 - Give an expression for the maximum SNR_o .
 - Let $e_{na} = 10 \text{ nV RMS}/\sqrt{\text{Hz}}$, $i_{na} = 1 \text{ pA RMS}/\sqrt{\text{Hz}}$, $R_s = 52 \text{ } \Omega$ at 300 K, $K_V = 10^3$, $K_F = 10$, $Q = 10$, $\omega_s = 2\pi 100 \text{ r/s}$, $e_{NS} = 1 \text{ nV RMS}/\sqrt{\text{Hz}}$, $4kT = 1.656 \times 10^{-20}$. All noise sources are white. Calculate n_o to give maxSNR_o . Also calculate the peak V_s to give a $\text{maxSNR}_o = 1.0$ at the V_o node.
- 3.8 In the circuit of Figure P3.8, in order to maximize the SNR_o at V_o , a sinusoidal source with a low R_s is connected to the OA through an ideal transformer. The transformer's turns ratio, n , is adjusted to maximize the SNR_o . The OA is noisy but otherwise ideal. Its $e_{na} = 10 \text{ nV}/\sqrt{\text{Hz}}$, $i_{na} = 1 \text{ pA}/\sqrt{\text{Hz}}$, both white. $R_1 = 1 \text{ k}\Omega$, $R_F = 99 \text{ k}\Omega$, $R_S = 50 \text{ } \Omega$.

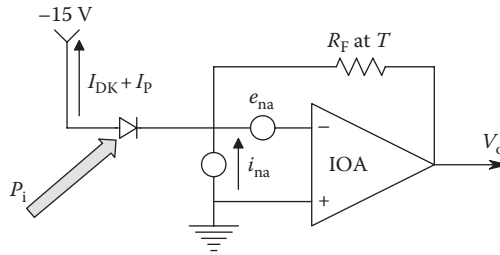


FIGURE P3.9

Assume all three resistors are at 300 K and make thermal white noise. The signal is $v_s(t) = V_s \sin(2\pi ft)$.

- A. Find an expression for the MS output signal voltage, $\overline{v_{os}^2}$.
 - B. Find an expression for the MS output noise voltage, v_{on}^2 . Assume a known noise bandwidth, B Hz, and include the five white noise sources.
 - C. Derive an expression for the transformer turns ratio, n_o , that will maximize the output SNR.
 - D. Find numerical values for n_o and the output maxSNR_o .
- 3.9 A silicon PIN photodiode is used in the biased mode, as shown in Figure P3.9. An AD549 EOA is used. It has $e_{na} = 35 \text{ nV}/\sqrt{\text{Hz}}$ (white), $i_{na} = 0.16 \text{ fA}/\sqrt{\text{Hz}}$ (white), $V_{OS} = 0$, $I_B = 35 \text{ fA}$ DC (inward), $R_{in} \rightarrow \infty$. The photodiode (PD) has $I_{rs} = 0.1 \text{ nA}$, its dark current at 15 V reverse bias is $I_{DK} = I_{rs} + I_{leak} = 4 \text{ nA}$, and its photocurrent sensitivity at 850 nm is $[\eta q \lambda / hc] = 0.342 \text{ A/W}$. Assume: The PD makes only white shot noise with PDS: $i_{sh}^2 = 2q(I_{DK} + I_P) \text{ mA/Hz}$. The Hz noise bandwidth is $B = 1 \text{ Hz}$, $4kT = 1.656 \times 10^{-20}$, $q = 1.602 \times 10^{-19} \text{ Cb}$, $R_F = 10^6 \Omega$ makes white thermal noise, and the reverse-biased PD presents an infinite dynamic resistance to the summing junction.
- A. Find the DC V_o in the dark, V_{oD} .
 - B. Assume an 850 nm input light power of P_i watts. Derive an algebraic expression for the MS SNR_{out} of the amplifier.
 - C. Calculate the P_i value to produce a DC ΔV_o equal to the RMS noise voltage output in the dark, where

$$V_o = V_{oD} + \Delta V_o.$$

- 3.10 Refer to Figure P3.10. When the SNR_{out} maximization transformer is truly ideal, we have shown that the optimum turns ratio is given by $n_o = e_{na} / (i_{na} R_s)$. In this problem, we consider finite winding resistances of the transformer, which is otherwise ideal. The primary resistance is $n_w = 10 \Omega$; the secondary resistance is $10 \times n \Omega$. Other parameters are $R_s = 100 \Omega$, $e_{na} = 10 \text{ nV}/\sqrt{\text{Hz}}$, $i_{na} = 1 \text{ pA}/\sqrt{\text{Hz}}$, $K_V = 10^3$, and $B = 10^4 \text{ Hz}$.
- A. Draw the Thevenin equivalent circuit seen by the amplifier looking into nodes C and D.
 - B. Write an algebraic expression for the amplifier's MS SNR_{out} .
 - C. Put all the n terms in the SNR_{out} expression's denominator, $D(n)$. Note that $D(n)$ has a minimum at some n_o , giving $\text{SNR}_{out}(n_o)$ a maximum. Set $dD(n)/dn = 0$ and solve numerically for the positive real root, n_o . (Note that $dD/dn = 0$ gives a fourth-order algebraic equation in n . Its roots can be found using the MATLAB® roots utility.)

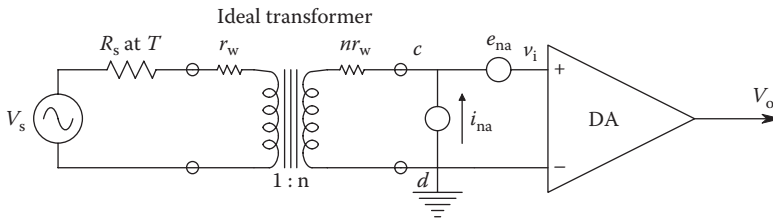


FIGURE P3.10

- 3.11** Gaussian white noise is added to a sine wave of frequency ω_0 . The sum, $x(t)$, is passed through a linear filter described by the ODE: $\dot{v}_o = a v_o + K x(t)$, $x(t) = n(t) + V_s \sin(\omega_0 t)$.
- Find an expression for the SS output MS SNR.
 - Find the value of a that will maximize the output MS SNR.
 - Give an expression for the maximized output MS SNR.
- 3.12** In an attempt to increase the MS SNR at the output of a signal conditioning system, the circuit architecture of Figure P3.12 is proposed. N low-noise preamplifiers, each with voltage gain K_v and the same short-circuit input voltage noise, e_{na} RMS $V/\sqrt{\text{Hz}}$, are connected as shown. The IOA and its gain-determining resistors can be considered to be noiseless. R_s at temperature T K makes Johnson noise. *Note:* $e_{na1}(t) \neq e_{na2}(t)$, etc., and $e_{na1} = e_{na2} = \dots = e_{naN} = e_{na}$ RMS $V/\sqrt{\text{Hz}}$.
- Find an expression for the MS output SNR.
 - Now assume the resistors R also make Johnson noise. Repeat A.
- 3.13** Refer to Figure P3.13. The current output of a PIN photodiode operating in the zero voltage mode is conditioned by an OP37 OA. The input signal is a sinusoidally modulated light power at 640 nm, $P_i(t) = 0.5P_{pk} [1 + \sin(\omega_0 t)]$. Assume: The PD's $I_{rs} = 0.1$ nA, $V_T = 0.0259$ V, $I_p = [\eta q / (h\nu)] P_i$, the PD makes white thermal *and* shot noise currents of $(4kT\sigma_d + 2q I_{pav})$ mA/Hz, the OA has $e_{na} = 3$ nV/ $\sqrt{\text{Hz}}$, $i_{na} = 0.4$ pA/ $\sqrt{\text{Hz}}$, and R_F makes

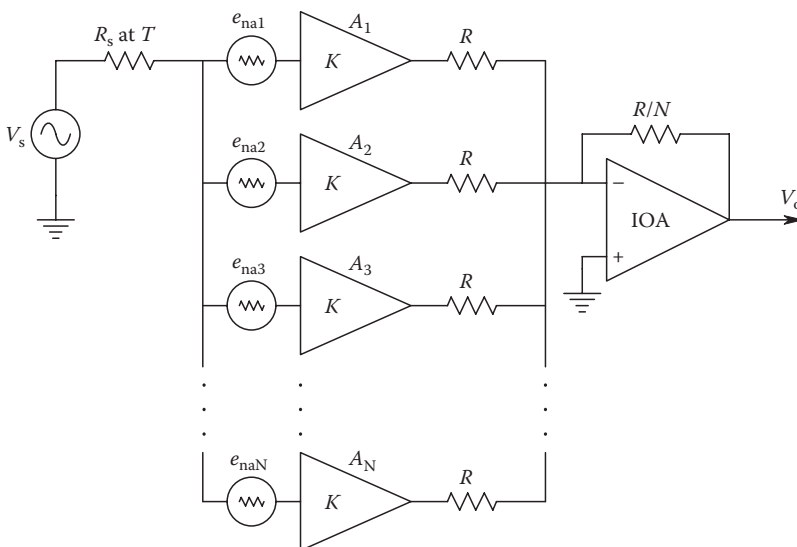


FIGURE P3.12

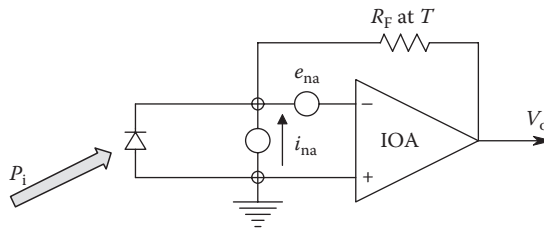


FIGURE P3.13

thermal noise. $4kT = 1.656 \times 10^{-20}$, $R_F = 10^4 \Omega$, and the noise bandwidth is $B = 1$ kHz around $f_o = \omega_o/2\pi = 10$ kHz.

- A. Write an expression for the MS output SNR.
- B. Find a numerical value for P_{pk} that will give unity SNR_{out} .
- C. What P_{pk} will saturate V_o at 12.5 V?

4

DC Null Methods of Measurement

4.1 Introduction

Classical electrical I&M systems (developed in the late nineteenth and twentieth centuries) often made use of null measurement systems. One reason for this was that the DC and AC analog meters available in this era lacked accuracy. Null-based measurements were seen to not rely on meter accuracy, only meter sensitivity, which was available.

DC null techniques are used with *Wheatstone bridges*, *Kelvin bridges*, and *potentiometers* to obtain increased measurement accuracy on DC voltages, currents, and resistances, based on the fact that the human observer can estimate the occurrence of a DC voltage or current null on a null meter with greater precision than he or she can directly read an analog meter scale in millivolts or microamperes. The accuracy of a DC null measurement system is derived from the known, calibrated accuracy of the resistors making up the bridges or potentiometer and has little dependence on the analog null meter. The human eye is quite sensitive in detecting the minute, transient deflections of an analog null meter's pointer under nonnull conditions.

Wheatstone bridges are traditionally used to make precise, accurate resistance measurements (generally in the range of 1– $10^6 \Omega$) or to measure some physical quantity, such as temperature, light intensity, or strain, which causes a known change in resistance. Figure 4.1 illustrates the complete circuit of a DC Wheatstone bridge.

In earlier days, the bridge null at balance was sensed by a sensitive galvanometer, an electromechanical, current-measuring transducer in which the bridge's DC unbalance current is passed through a coil that is delicately suspended in a strong permanent magnetic field. The magnetic flux caused by the small coil current interacts with the permanent magnet's flux and causes a deflection torque on the coil proportional to the current. The magnetic torque acts on the suspension's torsion spring and produces an angular deflection, which was traditionally optically sensed by reflecting a collimated light beam off a small mirror attached to the galvanometer's coil, creating an *optical lever* that greatly amplifies the mirror (and coil) rotation. Needless to say, although sensitive, a mirror galvanometer is not a mechanically robust instrument and consequently is hardly ever used in modern DC null measurement systems. Present day, DC null-sensing devices make use of electronic amplification and display devices; the amplifiers generally have differential, high-impedance ($>10^{12} \Omega$) input stages, so the bridge unbalance is sensed as an open-circuit DC voltage.

Kelvin bridges are null devices used to measure low values of resistance between $10 \mu\Omega$ and 10Ω . The null condition for the Kelvin bridge is described in the following.

Potentiometers, described in more detail in the following, are used to make very precise DC voltage measurements in the range of 1 V or higher. They are basically accurate voltage dividers that allow differential comparison of their known, calibrated DC output voltage with an unknown DC EMF. A simplified potentiometer circuit is shown in Figure 4.2.

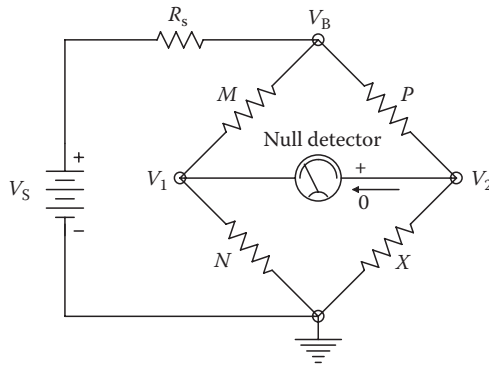


FIGURE 4.1

A basic Wheatstone bridge with DC excitation. The null is detected with a sensitive, center-scale, electronic DC voltmeter that draws negligible current.

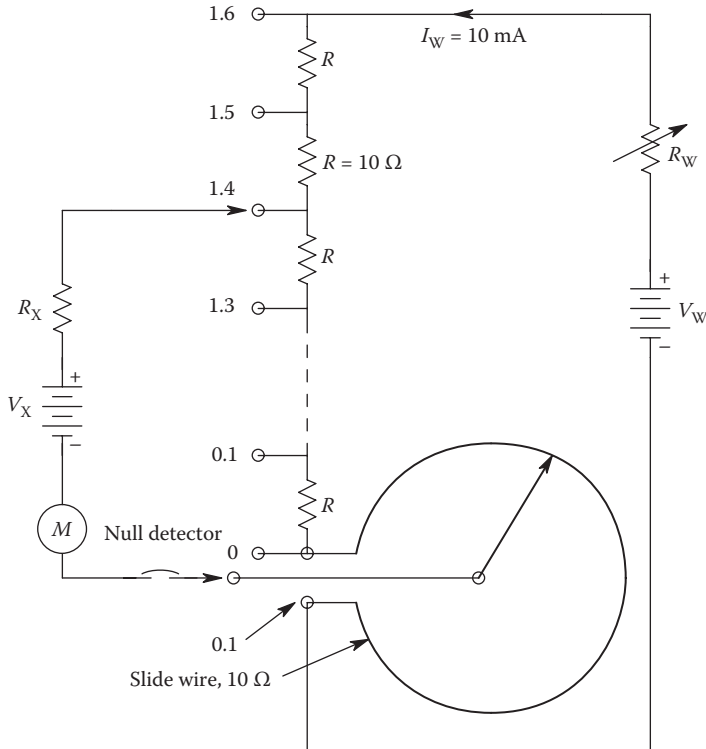


FIGURE 4.2

A basic DC potentiometer circuit. The null is sensed with an electronic null voltmeter. V_X is the unknown voltage to be measured; V_W is the open-circuit voltage of the working battery of the potentiometer.

Traditionally, potentiometers were calibrated with the saturated Weston cell, a precision electrochemical battery of known open-circuit voltage or EMF at a given temperature. The same null detection means is used to compare the calibrated potentiometer's open-circuit voltage output with the unknown EMF as is used with Wheatstone and Kelvin bridges. Modern potentiometers use series Josephson junction voltages as a reference EMF.

4.2 Wheatstone Bridge Analysis

If we assume that the null detector in Figure 4.1 is a high input impedance voltmeter, then in the null condition, $V_o = 0 = V_2 - V_1$. The open-circuit voltages at the corners of the bridge are given by the voltage divider relations:

$$V_1 = \frac{V_B N}{M + N}, \quad (4.1)$$

$$V_2 = \frac{V_B X}{P + X}. \quad (4.2)$$

Hence, at null, we have

$$V_o = V_B \left[\frac{X}{P + X} - \frac{N}{M + N} \right] = V_B \frac{XM - NP}{(P + X)(M + N)} = 0. \quad (4.3)$$

Because at null, $V_o = 0$, we have the well-known relation for the unknown resistor, X , in terms of bridge arm resistances:

$$X = \frac{NP}{M}. \quad (4.4)$$

The LE in determining resistance X at null may be found through application of the LE relation from Chapter 1 to Equation 4.4. In this case, we have

$$\Delta X = \left| \frac{\partial X}{\partial M} \Delta M \right| + \left| \frac{\partial X}{\partial P} \Delta P \right| + \left| \frac{\partial X}{\partial N} \Delta N \right|. \quad (4.5)$$

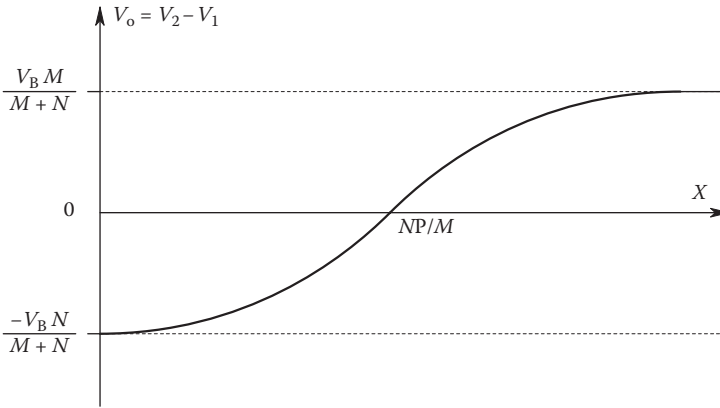
Algebraic substitution yields the simple relationship

$$\frac{\Delta X}{X} = \frac{\Delta M}{M} + \frac{\Delta P}{P} + \frac{\Delta N}{N}. \quad (4.6)$$

Thus, for example, if the resistors M , P , and N are known to 100 ppm (0.01%), then the LE in X is $\Delta X/X = 0.03\%$. Note that the LE is a *worst-case* error. In practice, the quantities of the form $\Delta Q/Q$ may sum with either sign, reducing the actual $\Delta X/X$.

The null detector resolution can also affect the LE in measuring X . Every null detector has a *dead zone* of $\pm \Delta V_o$ below which it is impossible to tell whether V_o is nonzero or not. This dead zone is caused by noise and/or the inability of a human operator to tell whether the physical indicator of the null, such as an analog meter pointer, has moved. To see how the dead zone, ΔV_o , affects the LE in X , we examine the Taylor's series for the Wheatstone bridge output relation:

$$\Delta V_o = \left(\frac{\partial V_o}{\partial X} \right) \Delta X_{\text{MIN}}. \quad (4.7)$$

**FIGURE 4.3**

Output (unbalance) voltage of a Wheatstone bridge as a function of the unknown arm resistance, X .

For the maximum bridge sensitivity case,

$$\left(\frac{\partial V_o}{\partial X} \right) = V_B S_V = \frac{V_B P}{(X + P)^2} \rightarrow \frac{V_B}{4X}. \quad (4.8)$$

Equation 4.8 for $\partial V_o / \partial X$ is substituted into Equation 4.7 and we solve for $\Delta X_{\text{MIN}} / X$:

$$\frac{\Delta X_{\text{MIN}}}{X} = \frac{\Delta V_o}{V_B} \left[\frac{(P + X)^2}{PX} \right] \rightarrow 4 \left(\frac{\Delta V_o}{V_B} \right). \quad (4.9)$$

Equation 4.9 for $\Delta X_{\text{MIN}} / X$ can be added to Equation 4.6 for the LE in X . Generally, this term is numerically much smaller than the $\Delta R / R$ terms in Equation 4.6, demonstrating that the null detector is generally the smallest source of error in measuring resistances by a DC Wheatstone bridge.

Often, we are not interested in measuring X , per se, but rather wish to examine V_o for small changes in X around the value that satisfies Equation 4.4. For this case, we plot Equation 4.3 in Figure 4.3. We define a voltage sensitivity for the bridge in Equation 4.5:

$$S_V \equiv \frac{\partial(V_o / V_b)}{\partial X} = \frac{P}{(X + P)^2}. \quad (4.10)$$

Clearly, S_V is maximum when $P = X$. This leads to the conclusion that the maximum sensitivity Wheatstone bridge is one where $P = X = M = N$. $S_{V\text{MAX}} = 1/4X$ in this case.

4.3 Kelvin Bridge

The Kelvin bridge is a specialized circuit used to measure very low resistances, such as ammeter shunts and motor armatures. The basic circuit for a Kelvin bridge is shown in Figure 4.4. Note that this bridge differs from a Wheatstone bridge by having four resistors

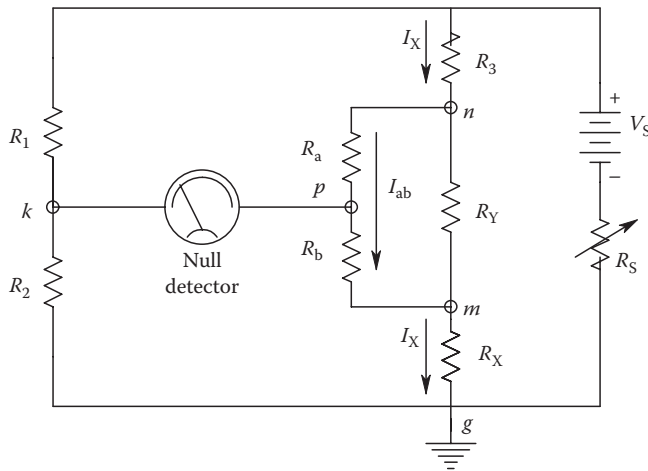


FIGURE 4.4
Circuit and currents in a Kelvin bridge used to measure very low resistances, R_X .

in the right-half circuit. To derive the conditions for null, we note that at null, $V_k = V_p$. V_k can be written by the voltage divider expression

$$V_k = \frac{V_S R_2}{R_1 + R_2}, \tag{4.11}$$

and

$$V_p = V_{pm} + V_m = I_{ab} R_b + I_X R_X, \tag{4.12}$$

where

$$I_{ab} = \frac{I_X R_Y}{R_a + R_b + R_Y} \tag{4.13}$$

and

$$I_X = \frac{V_S}{R_3 + R_X + R_Y (R_a + R_b) / (R_a + R_b + R_Y)}. \tag{4.14}$$

Thus, at null,

$$\frac{V_S R_2}{R_1 + R_2} = \left[\frac{V_S}{R_3 + R_X + R_Y (R_a + R_b) / (R_a + R_b + R_Y)} \right] \left[R_X + \frac{R_b R_Y}{R_a + R_b + R_Y} \right]. \tag{4.15}$$

Now, it turns out that if we let

$$\frac{R_a}{R_b} = \frac{R_1}{R_2} = \alpha \tag{4.16}$$

and substitute α in Equation 4.15, terms cancel, and we arrive at the Kelvin bridge null condition

$$R_x = \frac{R_3}{\alpha} = \frac{R_3 R_2}{R_1} \tag{4.17}$$

Generally, the resistances in the Kelvin bridge are relatively small, so the battery or DC working voltage source, V_s , must be able to source considerable current. A practical version of a Kelvin bridge is shown in Figure 4.5. Note that the low resistance R_3 is set by a tap switch and 0.001Ω resistors, with interpolation resistance, ΔR_3 , being given by a sliding contact on a manganin bar resistor; the ΔR_3 resistance is proportional to the tap position along the bar.

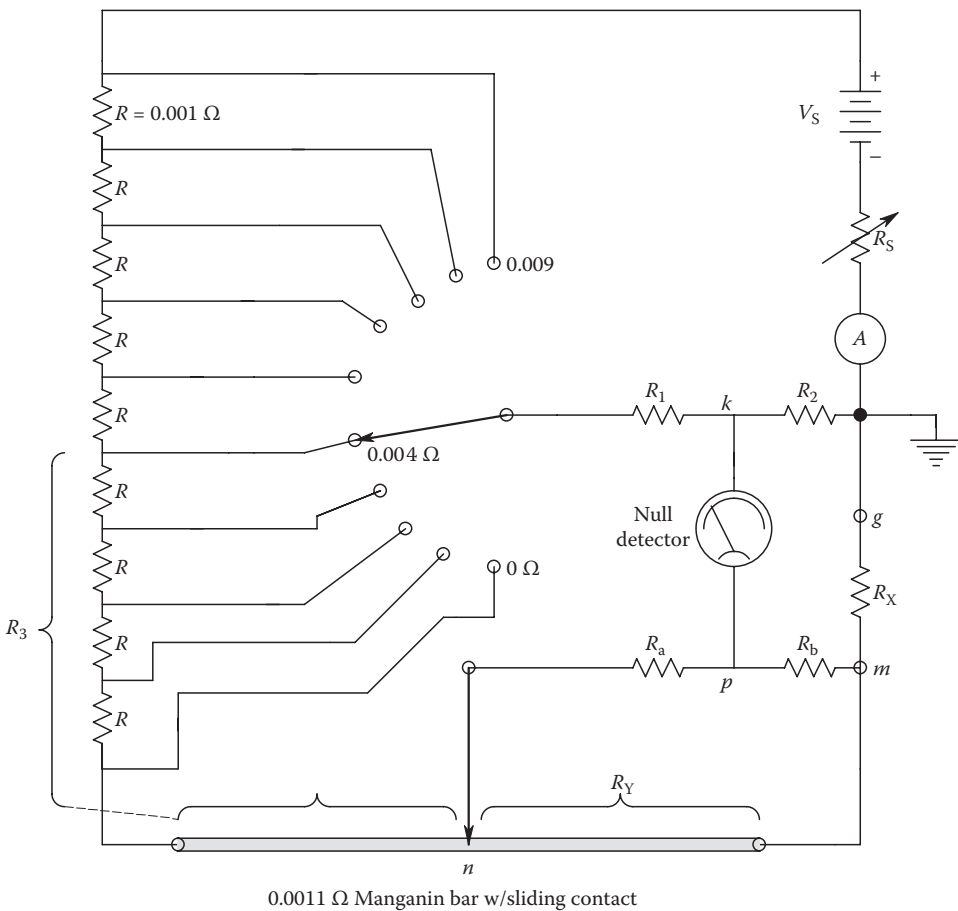


FIGURE 4.5
Practical circuit of a Kelvin bridge.

4.4 Anderson Constant Current Loop

4.4.1 Introduction

The current-loop technology for measuring the ΔR in remote-resistive sensors was developed in 1992 at the NASA Dryden Flight Research Center, Edwards, CA, by engineer Karl F. Anderson. It is our purpose in this section to describe how the *Anderson loop* works and discuss its benefits and shortcomings, comparing it with such circuits as the traditional, voltage-excited, Wheatstone bridge. Anderson received US Patent # 5,731,469, *Constant Current Loop Impedance Measuring System That Is Immune to the Effects of Parasitic Impedances*, in December 1994.

The advantages of the Anderson circuit architecture are quoted from Anderson (1998):

Larger and inherently linear outputs that are individually available from each element in a sensor (e.g., a bridge circuit).

Power dissipation of the sensor is lowered-beneficial in portable and temperature-sensitive applications.

Even random lead wire and connector [resistance] variations are tolerated without a need for expensive transmitters nearby in hostile environments or the removal of difficult-to-estimate systematic errors from the resulting data.

Fewer, smaller and less expensive lead wires are required in tight installations.

Quieter engineering unit readouts are available.

Smarter sensors can be designed to deliver multiaxis and multiparameter outputs from simpler sensing structures.

Stiffer sensors can have twice the frequency response for the same output level.

Refinements in compensation and calibration can be implemented after installation by adjustments conveniently away from the sensor within the signal conditioner.

At the heart of the Anderson loop is a signal-conditioning module called the *dual-differential subtractor* (DDS). A DDS circuit is shown in Figure 4.6. Note that it uses three, high-performance, differential, instrumentation amplifiers. A DDS is a linear, three-port voltage amplifier having two differential inputs and a single-ended output. Assuming ideal DAs (i.e., having ∞ CMRR), the DDS output is simply

$$v_o = K_1 K_2 (v_1 - v_2), \quad (4.18)$$

where

K_1 is the DM gain of the two input DAs

K_2 is the DM gain of the output DA

v_1 and v_2 are DM input signals

As an illustration of a simple Anderson loop system, consider the circuit of Figure 4.6. A remote sensor converts the QUM (e.g., temperature) into a change of resistance, ΔR . A current source (here ideal and referenced to ground) forces a current I_s through a loop

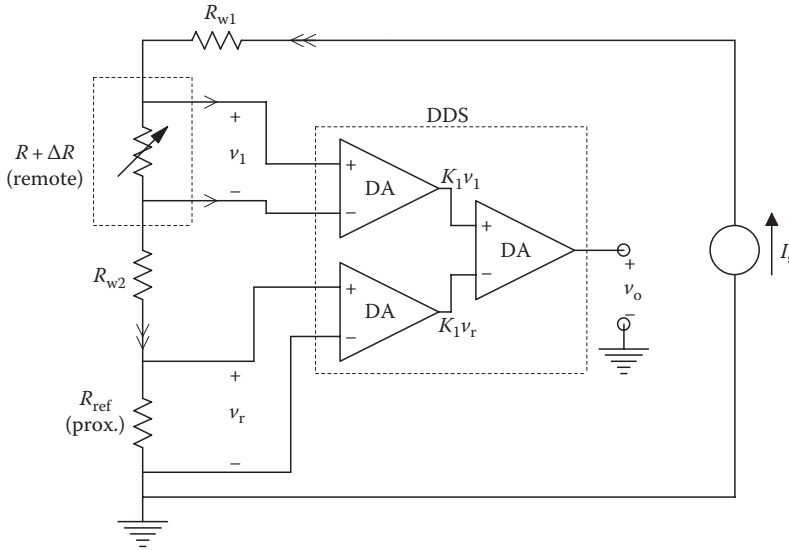


FIGURE 4.6

A simple, single-loop Anderson loop system, showing the DDS used in all Anderson current-loop systems.

that contains the wire resistances (R_{w1} and R_{w2}), the sensor ($R + \Delta R$), and a reference resistance, R_{ref} . The DM signal voltage input to the upper DA is, by Ohm's law, $v_1 = I_s(R + \Delta R)$. The DM voltage input to the lower DA is simply $v_2 = I_s R_{ref}$. Thus, the DDS output is

$$v_o = K_1 K_2 \left[I_s R (1 + \Delta R/R) - I_s R_{ref} \right]. \quad (4.19)$$

Now, if R_{ref} is made equal to R , the output is linearly proportional to ΔR :

$$v_o = K_1 K_2 I_s R \left(\frac{\Delta R}{R} \right). \quad (4.20)$$

Note that the DC CMV for the upper DA is $v_{1CM} = I_s(R_{w2} + R_{ref})$. Because of its very high CMRR, the upper DA has negligible output from v_{1CM} .

In Figure 4.7, we illustrate an Anderson loop applied to a capacitive sensor in which the QUM causes a change in capacitance, ΔC . Obviously, an AC excitation must be used. By inspection, we can write an expression for the output voltage of the DDS:

$$\mathbf{V}_o = \mathbf{I}_s \left[\frac{1}{j\omega(C_o + \Delta C)} - \frac{1}{j\omega C_v} \right]. \quad (4.21)$$

Now, we make $C_v = C_o$ and find for small $\Delta C/C_o$:

$$\mathbf{V}_o = \mathbf{I}_s \left[\frac{1 - \Delta C/C_o}{j\omega C_o} - \frac{1}{j\omega C_o} \right] = \frac{\mathbf{I}_s}{\omega C_o} \left(\frac{\Delta C}{C_o} \right) \angle -270^\circ. \quad (4.22)$$

Note that the magnitude of \mathbf{V}_o is proportional to ΔC , and calibration depends on knowing I_s , its frequency, ω , and C_o .

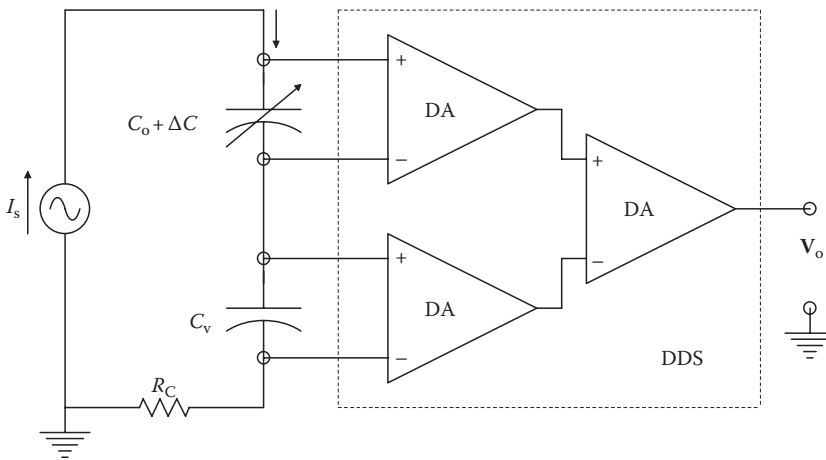


FIGURE 4.7
An Anderson current loop applied to a capacitive sensor (see text for analysis).

4.4.2 Anderson Loop Applied to Groups of Sensors

Figure 4.8 illustrates a dual-differential, resistive sensor in an Anderson loop. The QUM causes the upper sensor’s resistance to increase the same amount that the lower sensor’s resistance decreases. Again, the ubiquitous DDS is used. The DM input voltages are

$$v_1 = I_s(R + \Delta R), \tag{4.23a}$$

$$v_2 = I_s(R - \Delta R). \tag{4.23b}$$

The DDS output voltage is

$$v_o = K_2 [K_1 v_1 - K_1 v_2] = K_1 K_2 I_s R \left(\frac{2\Delta R}{R} \right). \tag{4.24}$$

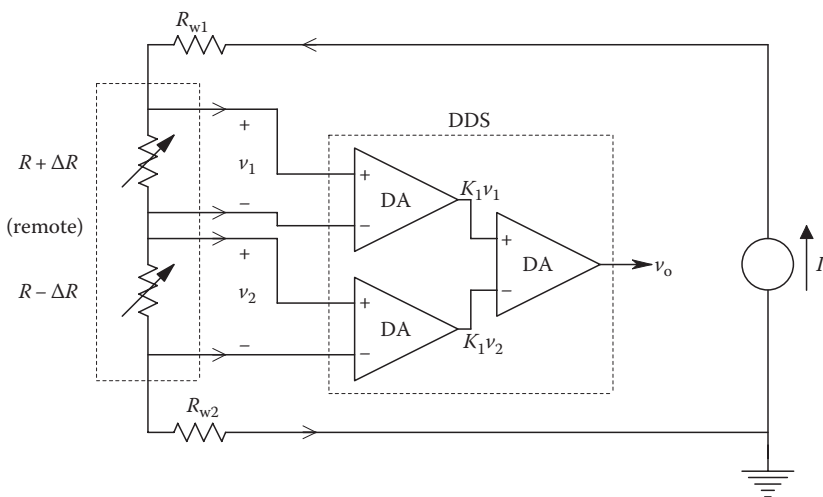


FIGURE 4.8
An Anderson loop DDS system is applied to remote, differential resistive sensors. The four wires from the sensors to the DDS carry negligible current.

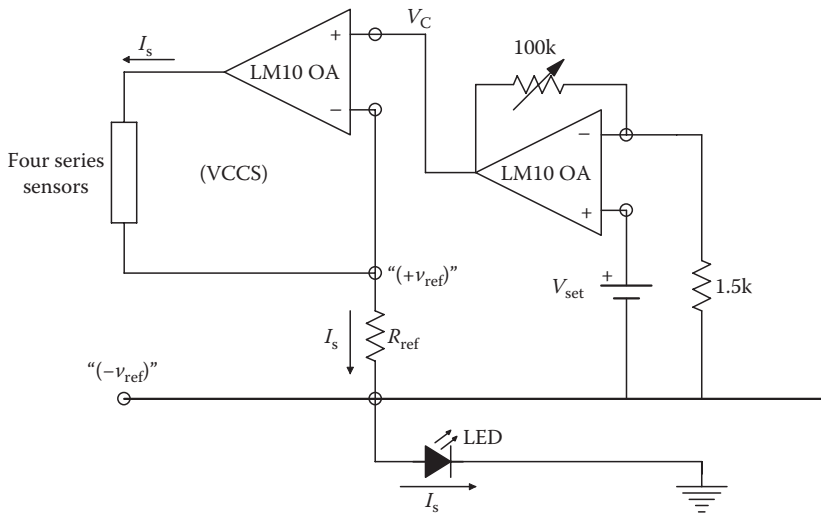


FIGURE 4.9
A stiff current source used by Anderson.

The DC CMVs are

$$v_{1CM} = I_s [(R - \Delta R) + R_{w2}], \tag{4.25a}$$

$$v_{2CM} = I_s R_{w2}. \tag{4.25b}$$

Before addressing the problem of adapting a four-sensor Wheatstone bridge to an Anderson current loop, let us examine the problem of designing a stiff current source for the Anderson loop. Figure 4.9 illustrates a simplified schematic of an OA VCCS such as used by Anderson. Notice that the series sensing elements in the loop are floating. That is, not one is directly conductively coupled to ground. The loop current, I_s , must return to ground through the common (feedback) resistor, R_{ref} , and in this case, also an LED. If we assume the OA is ideal, then the reference voltage $(+v_{ref})$ must equal the control voltage, V_C . Thus, by Ohm’s law, $I_s = V_C/R_{ref}$. To see how *stiff* this current source is, we must examine a more detailed circuit of the VCCS using a simple model for a finite-gain, real OA. Figure 4.10 illustrates such a circuit. Note that the OA’s output resistance without feedback, R_o , is on the order of 25 Ω . Ideally, a VCCS should have an output resistance over 1 M Ω . To calculate the VCCS’s output resistance, we remove R_L and replace it with a small-signal test voltage source, v_T . Also set $V_C = 0$. The VCCS’s output resistance is, by Ohm’s law, simply $R_{out} = v_T/i_T$. Simple circuit analysis reveals that $R_{out} = R_o + R_{ref}(1 + K_D) \cong R_{ref}K_D$. Consider the typical values of $R_{ref} = 300 \Omega$ and $K_D = 10^5$, thus $R_{out} = 30 \text{ M}\Omega$, which is adequate. It is also easy to show that this VCCS has the following transconductance: $G_M = I_s/V_C \cong 1/R_{ref} \text{ S}$. A major drawback of the simple Anderson VCCS is that the load is *floating*, that is, its elements must be insulated from ground.

The schematic of a standard, four-active-arm, voltage-excited Wheatstone bridge is illustrated in Figure 4.11. This bridge configuration is typical of a force or pressure sensor using four unbonded strain gauges (see Figure 6.4). Assume the bridge excitation source and the DA are remote from the bridge. The DA has a very high input impedance (generally over $10^8 \Omega$),

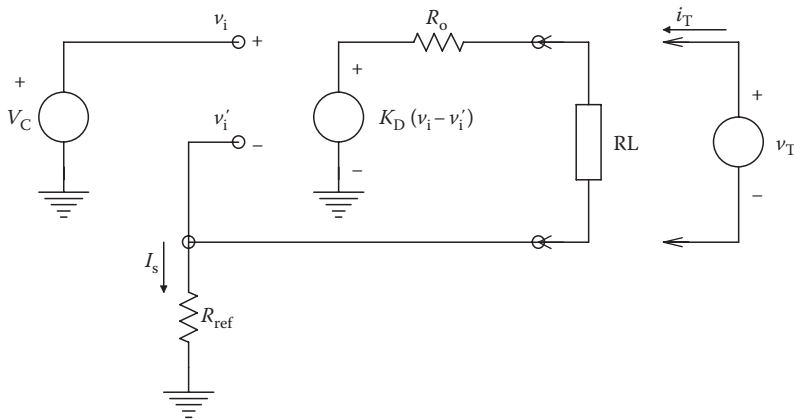


FIGURE 4.10
Circuit model for the current source shown in Figure 4.9.

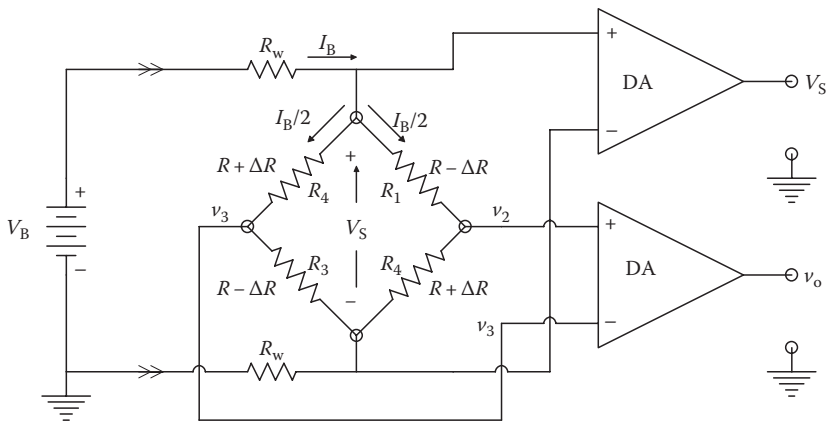


FIGURE 4.11
Schematic of a remote, four-active-arm Wheatstone bridge. Two DAs are used to condition the bridge output, v_o , and the DC voltage across the bridge, V_s .

so it is possible to ignore any small resistance of the wires leading from the bridge to the DA. V_s , the voltage across the bridge, is slightly less than the excitation voltage, V_B , by the voltage dropped across the resistances of the wires supplying the bridge. Elementary circuit theory tells us that $I_B = V_B / (2R_w + R)$, and thus $V_s = V_B R / (2R_w + R)$. The bridge output can easily be found using voltage divider relations:

$$v_o = (v_2 - v_3) = V_s \left[\frac{R + \Delta R}{2R} - \frac{R - \Delta R}{2R} \right] = V_s \frac{\Delta R}{R}. \tag{4.26}$$

The bridge calibration depends on V_s , which in turn is dependent on R_w , which is temperature sensitive.

The uncertainty in V_s can be removed if this voltage is measured by another DA whose output is compared to the desired value and the error used to adjust V_B .

To examine the evolution of the four-active-arm Anderson circuit derived from a four-active-arm Wheatstone bridge, examine the circuit of Figure 4.12A. The top of the

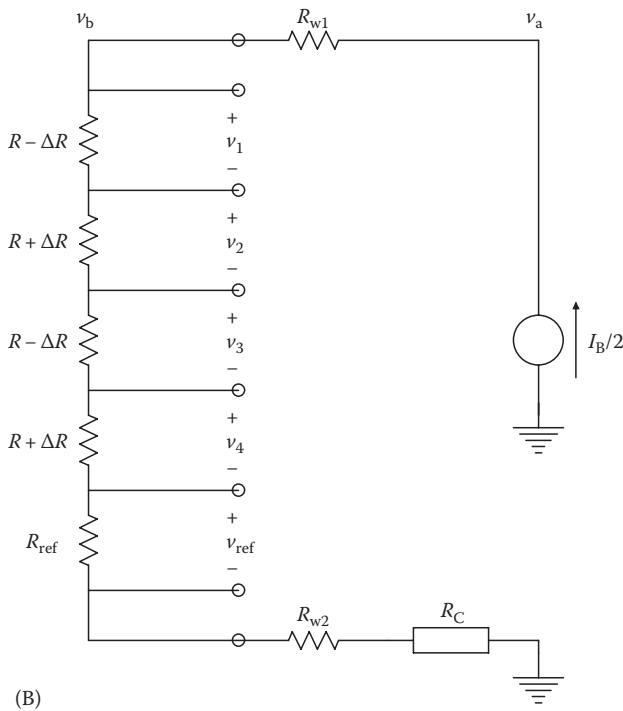
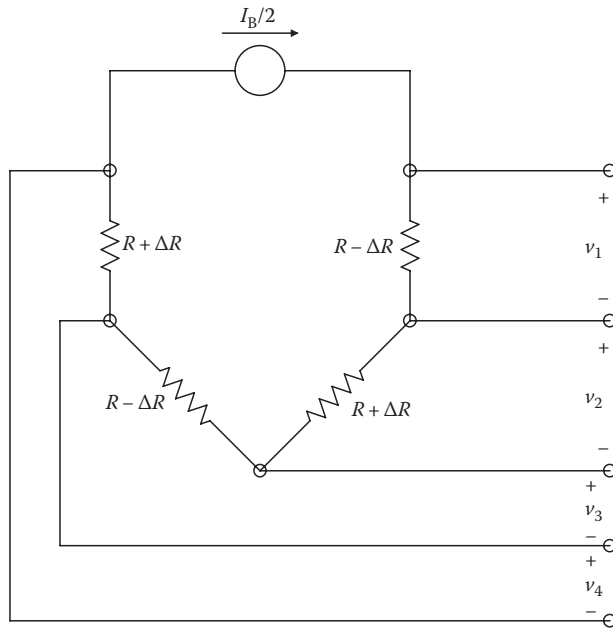


FIGURE 4.12 Steps in *Andersonizing* a remote, four-active-arm Wheatstone bridge. (A) The bridge is powered by a floating DC current source. (B) The bridge circuit is redrawn.

Wheatstone bridge is opened, and a current source of $I_B/2$ is inserted. ($I_B/2$ is used because this is the current in each of the four Wheatstone bridge active elements.) Note that in this circuit, we have lost the ground from the bottom of the bridge. The Anderson circuit is redrawn in Figure 4.12B, illustrating the floating (ungrounded) nature of the Anderson loop. Note that a reference resistance, R_{ref} , is added and also an additional CM resistance, R_C , may be present.

Figure 4.13 illustrates the final *Andersonized* four-active-arm sensor system with the DDSs added. Note that no less than 15 precision DAs are required with high CMRRs! The CM input voltage for the top DA is simply $v_{1CM} = (I_B/2) [3R + \Delta R + R_{ref} + R_C]$, the CMV for the second sensor's DA is $v_{2CM} = (I_B/2) [2R + R_{ref} + R_C]$, etc. The reference voltage is simply $v_{ref} = (I_B/2)R_{ref}$ and the CMV for the four reference input DAs is $v_{rCM} = (I_B/2)R_C$. Assume the DAs are ideal, that is, their input resistances, bandwidths, and CMRRs are infinite, they have zero output resistance, and their differential gains are finite. So for the top DDS,

$$v_1 = \left(\frac{I_B}{2} \right) (R - \Delta R), \quad (4.27)$$

$$v_r = \left(\frac{I_B}{2} \right) R_{ref}, \quad (4.28)$$

$$v_{o1} = K_1 K_2 \left(\frac{I_B}{2} \right) R [1 - \Delta R/R + R_{ref}/R]. \quad (4.29)$$

Similarly, we can show

$$v_{o2} = K_1 K_2 \left(\frac{I_B}{2} \right) R [1 + \Delta R/R + R_{ref}/R] = v_{o4} \quad (4.30)$$

and

$$v_{o3} = v_{o1}. \quad (4.31)$$

The net output of the *Andersonized* four-arm bridge is formed by the output DDS:

$$v_{oB} = K_1^2 K_2^2 I_B (2\Delta R). \quad (4.32)$$

Using the DDS modules, there is no need to make $R_{ref} = R$ to obtain Equation 4.32, but this condition will allow extraction of $(v_{o1} - v_{o4})$ without a DC offset, for example, $v_{o2} = K_1 K_2 (I_B/2) R [\Delta R/R]$.

Note that you do not need to go to the complexity of the circuit of Figure 4.13 to measure $\Delta R/R$. One sensor works, as shown in Figure 4.6. If R_{ref} is made a sensor without an input and placed adjacent to the active element, then thermal compensation will be inherent in the Anderson circuit. Note that a two-active-arm or four-active-arm Wheatstone bridge is inherently thermally stabilized as well.

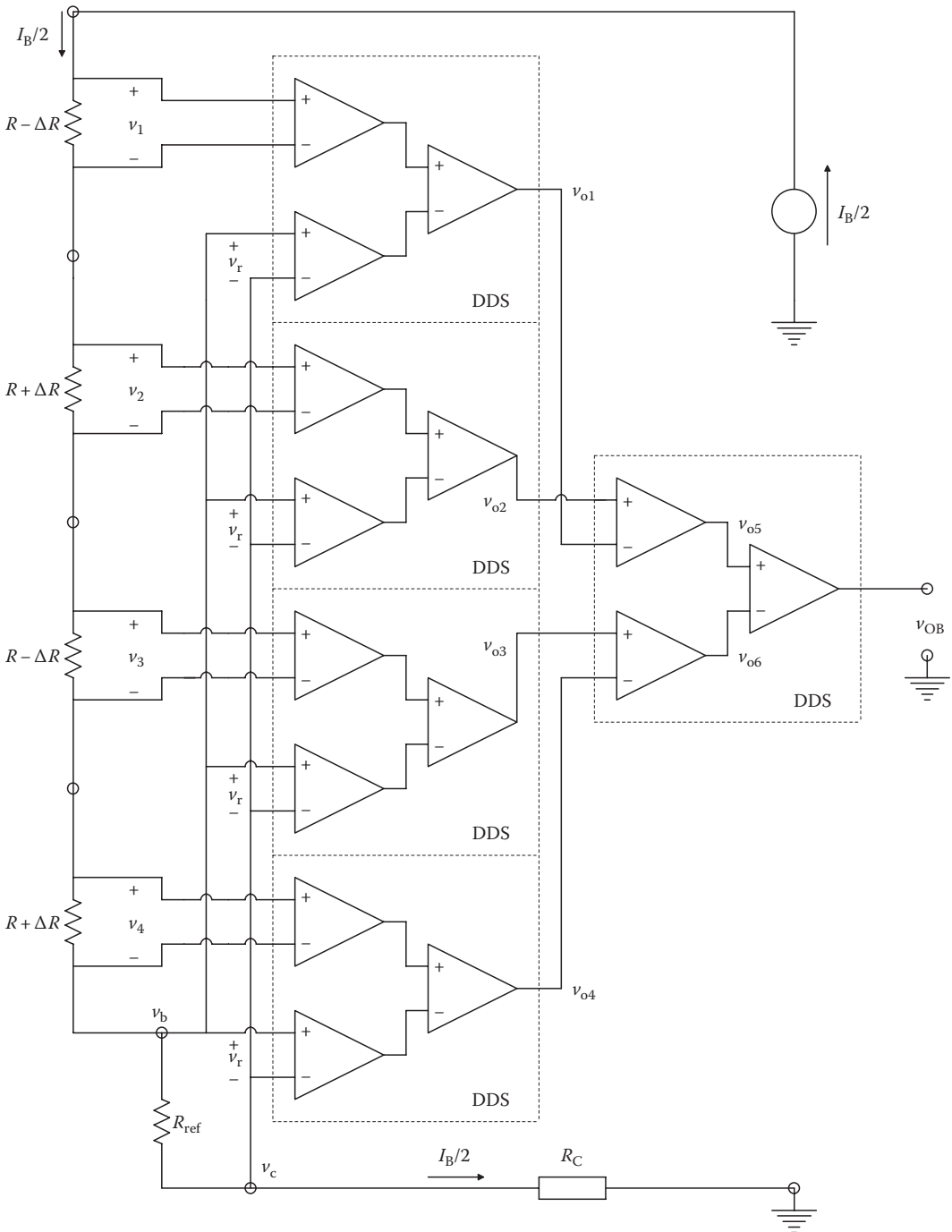


FIGURE 4.13 The final, *Andersonized*, remote, four-active-arm Wheatstone bridge circuit. Note that a total of five DDS elements are required to obtain an output voltage proportional to ΔR . Compare this circuit to the conventional bridge circuit of Figure 4.11.

4.4.3 Conclusion

The effectiveness of the Anderson loop architecture is predicated on the use of the DDS module in which the DAs all have high CMRRs, generally exceeding 120 dB ($>10^6$). However, this requirement is also valid for a conventional Wheatstone bridge supplied by a voltage source. Calibration of the Wheatstone bridge depends on the existence of a known, stable potential difference between the top and bottom of the bridge. Similarly, calibration of an Anderson loop sensor requires the current source (I_s or $I_B/2$) to be stiff, known, and stable. Just as in the case of a conventional Wheatstone bridge, an Anderson loop system can be excited sinusoidally, permitting signal amplification out of the $1/f$ range of amplifier noise and allowing synchronous unbalance detection.

4.5 Potentiometers

Figure 4.14 illustrates the simplified circuit of a potentiometer. Like Wheatstone and Kelvin bridges, potentiometers are DC null instruments. They now generally find applications in standards laboratories. They are used as a precision means of calibrating secondary DC voltage standards. Originally, potentiometers were used with thermocouples to obtain

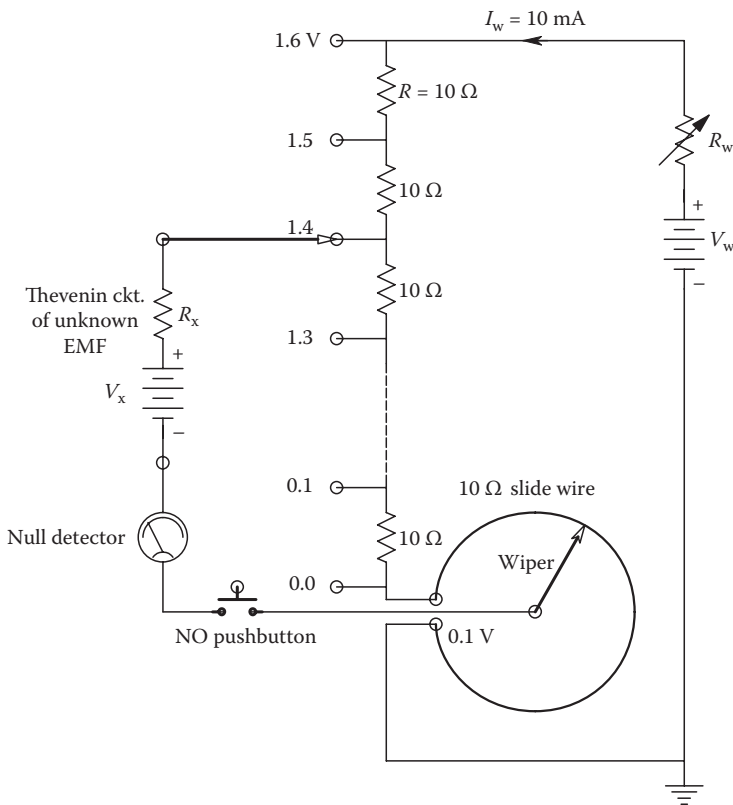


FIGURE 4.14 Circuit of a simple potentiometer used for precision, direct comparative measurements of EMFs in the 0–1.6 V range.

TABLE 4.1

Ranges and Accuracy of the L&N 7556-1 Potentiometer

Range (V)	Accuracy
0–1.61100	$\pm(0.0005\% + 1 \mu\text{V})$
0–0.16110	$\pm(0.0015\% + 0.1 \mu\text{V})$
0–0.01611	$\pm(0.0025\% + 20 \text{ nV})$

precise temperature measurements and in electrochemistry laboratories to measure cell EMFs. Most of these applications are now taken over by precision electronic voltmeters.

For example, the venerable Leeds and Northrup Model 7556-1 potentiometer has three ranges. The accuracy of each range is given in Table 4.1.

To measure HVs, a precision voltage divider called a *volt box* is used. For example, the Leeds and Northrup Model 7594-2 volt box was rated at 0–1500 V, with an accuracy of $\pm 0.0025\%$. It had 25 taps and was rated at $667 \Omega/\text{V}$ on the 1500 V range, quite a lossy device compared to modern voltmeter sensitivities. (*Note:* L&N was acquired by General Signal Corp, Stamford, CT, in 1978. GSC divested of L&N in 1998 to SPX, Charlotte, NC.)

Potentiometers were often calibrated using *Weston standard cells*, an electrochemical battery, which until fairly recently was used to define the volt EMF. The Weston cell was developed in 1892 as an EMF standard. Two types of Weston cells exist, saturated and unsaturated. The more widely used saturated (normal) cell is illustrated in Figure 1.6. A Weston cell is generally used in a thermostatted oil bath to maintain its internal temperature to within 0.01°C . This is necessary because of the rather high EMF tempco of the normal cell, $-40 \mu\text{V}/^\circ\text{C}$. The EMF of the saturated Weston cell is given by

$$\text{EMF} \equiv 1.01830 - 4.06 \times 10^{-5}(T - 20) - 9.5 \times 10^{-7}(T - 20)^2 - 10^{-8}(T - 20)^3. \quad (4.33)$$

(This EMF is in *international volts*; the EMF in *absolute volts* is 1.01864 at 20°C .) Thus, the volt was formerly defined as $1/1.01830$ of the Weston normal cell at 20°C . Weston cells are very delicate and subject to instabilities if their EMFs are subjected to large or sudden temperature changes. They also should never have currents in excess of $100 \mu\text{A}$ drawn from them for any period of time. They also age and should be recalibrated against freshly prepared cells every year or against Josephson junction primary EMF standards.

Calibration of a potentiometer is done by first adjusting the potentiometer's reading to the calculated EMF of the standard cell and then adjusting R_W to obtain a null in the galvanometer or electronic null detector. Next, the unknown EMF is connected to the potentiometer and the null detector, and the calibrated potentiometer's output voltage is adjusted to obtain a null.

4.6 Chapter Summary

In this chapter, we have discussed three basic measurement systems that make use of the principle of nulling or zeroing a DC voltage to obtain precision measurement of a resistance or an EMF: the Wheatstone bridge, the Kelvin bridge, and the potentiometer.

The accuracy attainable in a null measurement technique generally depends on the use of accurate resistors, rather than the accurate measurement of a DC voltage or current, per se.

The Anderson loop technology was introduced and shown to use electronic, constant current sources and a basic three-DA configuration called the DDS in conjunction with sensors whose resistance changes proportional to the QUM. The Anderson loop allows the voltage output to be linear in terms of the ΔR of the sensor. It is shown to be an alternative to the Wheatstone bridge in certain resistive sensor applications.

In Chapter 5, AC null-based measurements are shown to be applicable to AC bridge measurements of R , L , C , and Q .

Problems

- 4.1** Two ideal OAs are used to linearize the output of a single-active-arm *Wheatstone bridge*, as shown in Figure P4.1. The resistance of the variable resistor, R_v , is given by $R_v = R + \Delta R$, where $\Delta R > -R$.
- A. Find an algebraic expression for $V_o = f(V_s, R, R_1, R_2, \Delta R)$. *Hint:* Write node equations for V_1 and V_2 ; note that by the ideal OA assumption, $V_1 = V_2 = 0$.
- B. Let $V_s = 5 \text{ V}$ and $R = R_1 = R_2 = 10^3 \text{ } \Omega$. Find the permissible range in R_v so that neither OA's output reaches $\pm 12 \text{ V}$ (saturation).
- 4.2** A DC Wheatstone bridge contains a resistance temperature detector (RTD) whose resistance around 25°C is approximated by $R(T) = R_{25}[1 + \alpha(T - 25)]$. The smallest ΔV_o the detector can resolve is 100 nV . $V_s = 3.1 \text{ V}$, $R = R_{25} = 300 \text{ } \Omega$, and $\alpha = 0.004$. Find the just-resolvable ΔT under these conditions (see Figure P4.2).
- 4.3** A DSBSCM signal is generated in many sensor systems. A DSBSCM signal is basically the product of two sine waves with different frequencies. In the Wheatstone bridge circuit of Figure P4.3, two strain gauges are subject to an LF sinusoidal strain such that $\delta R(t) = \Delta R \sin(2t)$. $\Delta R = 0.03 \text{ } \Omega$, $R = 300 \text{ } \Omega$, $v_b(t) = V_B \sin(1000t)$, and $V_B = 5 \text{ Vpk}$. The DA's gain is 10^3 . Find an expression for $v_o(t)$ and show that it can be resolved into the algebraic sum of sum and difference frequencies.

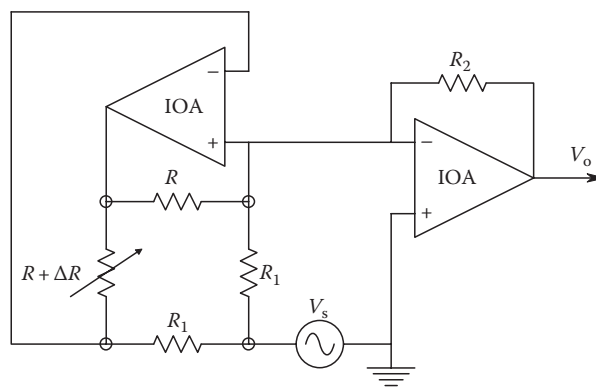


FIGURE P4.1

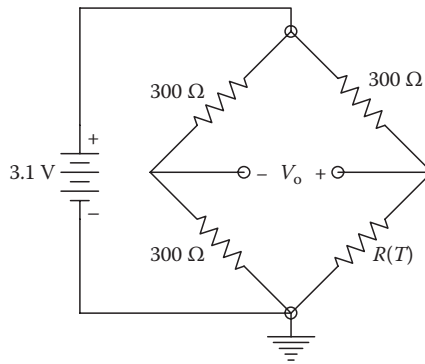


FIGURE P4.2

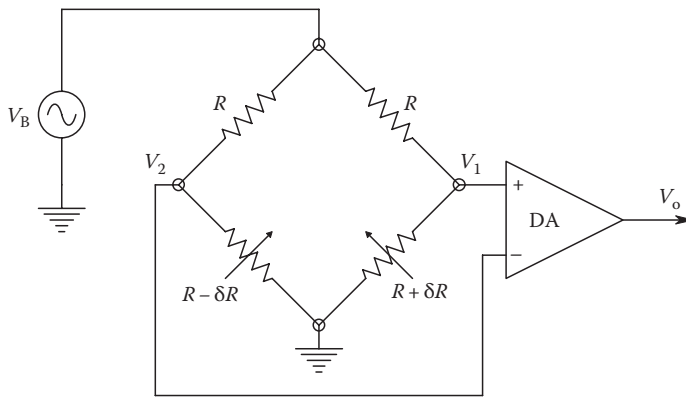


FIGURE P4.3

- 4.4 A Wheatstone bridge is powered with a $1\ \text{V rms}$, $200\ \text{Hz}$ sinusoidal source. Fixed resistors A and B are $1000\ \Omega$, with LEs of $10\ \text{ppm}$. The value of variable resistor, D , is known to be $100\ \text{ppm}$. At null, $D = 100\ \Omega$ (see Figure P4.4).
- Find the nominal value of X .
 - The AC null detector can just resolve $\Delta V_o = 1\ \mu\text{V rms}$. The DA is ideal with gain = 1. Calculate the LE in X , considering the LEs in A , B , D , and V_o .

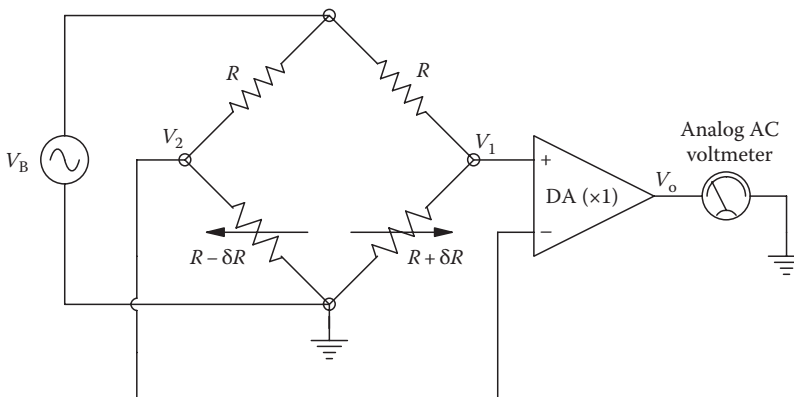


FIGURE P4.4

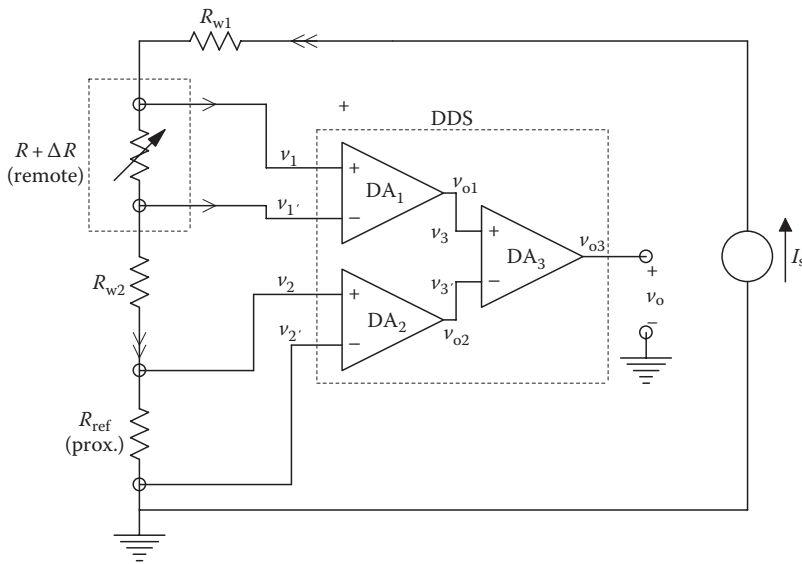


FIGURE P4.5

4.5 Consider Figure P4.5 for the basic Anderson DDS circuit. Assume that the identical input DA's outputs are described by the following equations:

$$V_{o1} = A_{D1}v_{1d} + A_{C1}v_{1c}, \quad \text{CMRR}_1 \equiv \frac{A_{D1}}{A_{C1}},$$

$$V_{o2} = A_{D2}v_{2d} + A_{C2}v_{2c}, \quad \text{CMRR}_2 \equiv \frac{A_{D2}}{A_{C2}}.$$

Let $A_{D1} = A_{D2}$, $A_{C1} = A_{C2}$, so $\text{CMRR}_1 = \text{CMRR}_2$. The output subtractor DA's output is described by

$$v_o = v_{o3} = A_{D3}v_{3d} + A_{C3}v_{2d}, \quad \text{CMRR}_3 \equiv \frac{A_{D3}}{A_{C3}}.$$

Note that $A_{C1} = A_{D1}/\text{CMRR}_1$, etc., and R_{ref} is made equal to R_o of the sensor.

- Write expressions for v_1 , v_1' , v_2 , and v_2' in terms of circuit parameters and ΔR .
 - Find expressions for the DA outputs, v_{o1} and v_{o2} .
 - Find an expression for $v_o = v_{o3}$. Give $v_o = v_o'$ when all of the CMRRs $\rightarrow \infty$.
 - Give an expression for the Anderson DDS circuit's voltage error, $v_e = v_o - v_o'$, for finite CMRR_1 and CMRR_3 .
- 4.6 Figure P4.6 illustrates the schematic of a potentiometer. R_w is used to adjust the working current, I_w , in the loop to exactly 50.000 mA when setting the pot up with a standard cell. The standard cell's EMF is defined as 1.01830 V at 20°C. A separate, high-resolution, standard cell slide-wire is used to initially set I_w . The EMF slide-wire has 100 divisions over its range and can be read to 0.1 division. The maximum range of E_x that can be read is 1.7 V.

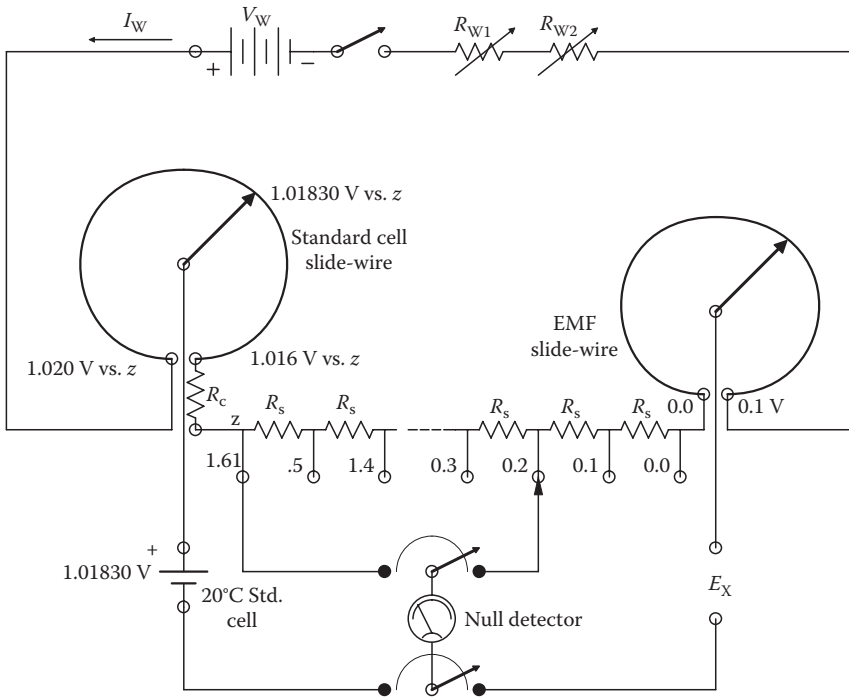


FIGURE P4.6

- A. Find the smallest ΔE_X that can be read.
 - B. Give ohm values for each of the 16 fixed resistors, R_s , and the EMF slide-wire's resistance, the standard cell's slide-wire resistance, and R_c .
 - C. Show the polarity of the E_X terminals.
- 4.7 A Kelvin bridge schematic is shown in Figure P4.7. At null, $V_k = V_p$.
- A. Find expressions for V_k , I_X , and I_{ab} .
 - B. Derive an expression for R_X at null. (Note: it has two terms.)
 - C. At null, $R_1 = 100.0\ \Omega$, $R_2 = 1000.0\ \Omega$, $R_Y = 0.1000\ \Omega$, $R_3 = 0.00377\ \Omega$, $R_b = 99.2\ \Omega$, and $R_a = 1000.6\ \Omega$.
 Calculate R_X using the simple, first term in the equation of part (B). Calculate R_X using the whole expression from (B).
 - D. Calculate I_X assuming $V_B = 1.500\text{ V}$. Find the power dissipated in R_X .
- 4.8 Consider the Anderson DDS used to measure the small ΔC of a capacitive sensor, as shown in text Figure 4.7. Consider the DAs of the DDS to have flat frequency responses (scalar gains) with outputs given by

$$v_{o1} = A_{D1}v_{1d} + A_{C1}v_{1c}, \quad CMRR_1 \equiv \frac{A_{D1}}{A_{C1}},$$

$$v_{o2} = A_{D2}v_{2d} + A_{C2}v_{2c}, \quad CMRR_2 \equiv \frac{A_{D2}}{A_{C2}}.$$

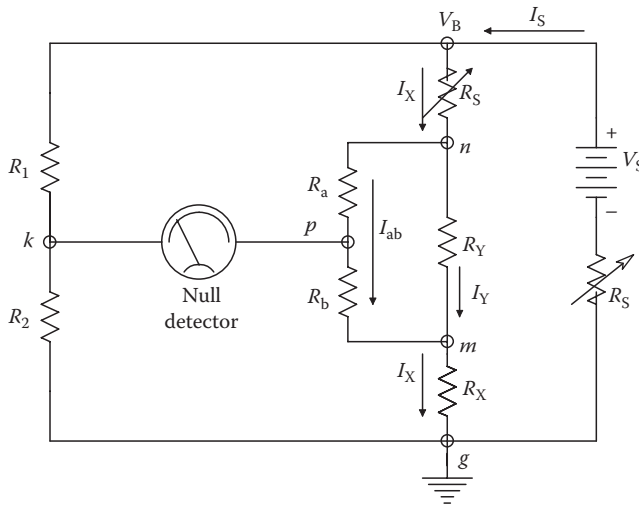


FIGURE P4.7

Let $A_{D1} = A_{D2}$, $A_{C1} = A_{C2}$, so $CMRR_1 = CMRR_2$. The output subtractor DA's output is described by

$$v_o = v_{o3} = A_{D3}v_{3d} + A_{C3}v_{2d}, \quad CMRR_3 \equiv \frac{A_{D3}}{A_{C3}}.$$

Note that $A_{C1} = A_{D1}/CMRR_1$, etc., and C_v is made equal to C_o of the sensor.

- A. Write phasor expressions for V_1 , V'_1 , V_2 , and V'_2 in terms circuit parameters and ΔC . Give I_s zero phase.
- B. Find expressions for the phasor DA outputs, V_{o1} and V_{o2} .
- C. Find an expression for $V_o = V_{o3}$. Give $V_o = V'_o$ when all of the CMRRs $\rightarrow \infty$.
- D. Give an expression for the Anderson DDS circuit's voltage error, $V_\epsilon = V_o - V'_o$, for finite $CMRR_1$ and $CMRR_3$.

5

AC Null Measurements

5.1 Introduction

In this chapter, we examine measurement techniques that make use of sensing an AC null voltage or current. These include, but are not limited to, AC operation of (resistive) Wheatstone bridges, various types of AC bridges used to measure *capacitance*, *capacitor dissipation factor (D)*, *inductance*, *inductor quality factor (Q)*, *mutual inductance*, and the *small-signal transconductance (g_m)* of bipolar junction transistors (BJT) and field-effect transistors (FETs). As in the case of DC null methods, AC methods are used to obtain accurate measurements of component values based on the accuracy of the bridge's components.

It should be obvious that in the real world of circuit components, there is no such thing as a pure or ideal resistor, capacitor, or inductor. All real-world devices have parasitic parameters associated with them. In some cases, these parasitic components may be treated as lumped-parameter circuits, while in other situations they are best described as distributed-parameter networks, similar to transmission lines. For example, depending on the frequency of the AC voltage across a resistor, the resistor may appear as having a pure inductance in series with a resistor in parallel with a capacitor and distributed capacitance to ground along its length. The resistance of a resistor has a value that is an increasing function of frequency due to skin effect at VHF – UHF. The presence of parasitic components at high frequencies makes the operation of bridges and null circuits at high frequencies more subject to errors. To minimize the effects of parasitic components, most laboratory AC bridges operate at or around 1 kHz.

5.2 Components: Inductor Equivalent Circuits

Most practical inductors are made from one or more turns of a conductor such as copper, wound as a solenoid on either an air core or a ferromagnetic core made from ferrite ceramic or laminated iron. If the coil is wound around a ferromagnetic ring or *doughnut*, it is called a toroidal inductor. The use of ferromagnetic cores concentrates the magnetic flux and produces a higher inductance than would be attainable with an air core and the same coil geometry. Because the conductor used to wind an inductor has a finite resistance, the simplest, LF equivalent circuit of a practical inductor is a resistor in series with a pure inductance. At very high frequencies, the small, stray capacitance between adjacent turns of the inductor's coil produces a complex, distributed-parameter, *RLC* circuit. If the coil is wound

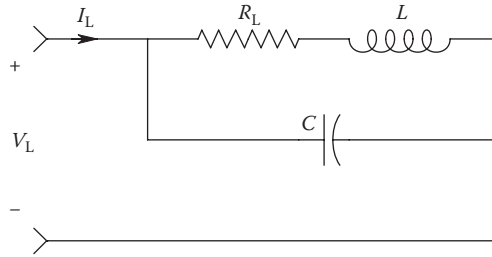


FIGURE 5.1
HF, lumped-parameter, linear circuit model of a practical inductor.

with several layers of turns, capacitance between the layers, as well as between adjacent turns, and between the inner turns to the core produces a very complex equivalent circuit of the inductor at very high frequencies.

One compromise to the problem of representing an inductor realistically at high frequencies is shown in Figure 5.1. Here, the distributed capacitance between windings is modeled by a single, equivalent, lumped capacitor in parallel with the series R - L circuit. This model is reasonably valid in the low audio-frequency range where most AC inductance bridges operate. To characterize and measure inductors at radio frequencies where the lumped model is not valid, we generally use a Q meter, discussed in Section 8.10.3.2.

The *quality factor* or Q of an inductor is defined at a given frequency as the ratio of the inductor's inductive reactance to the real part of its impedance. The higher the Q , the *purser* the inductor and the lower are its losses. High- Q inductors allow us to build more sharply tuned RLC frequency-selective circuits than do low- Q inductors of the same inductance. For a given inductance, high- Q inductors are generally more expensive. Note that Q is a dimensionless function of frequency. For a simple series R - L circuit model, the Q is given by

$$Q \equiv \frac{X_L}{R_L} = \frac{\omega L}{R_L}. \quad (5.1)$$

The Q of the RLC inductor model of Figure 5.1 may be found by finding its impedance:

$$\mathbf{Z}_L(j\omega) = \frac{j\omega L + R}{1 + j\omega RC + (j\omega)^2 LC} = \frac{R}{(1 - \omega^2 LC) + (\omega RC)^2} + j \frac{\omega L [(1 - \omega^2 LC) - CR^2/L]}{(1 - \omega^2 LC) + (\omega RC)^2}. \quad (5.2)$$

From the basic definition of Q given in Equation 5.1, and the impedance given by Equation 5.2, we can write the series model inductor's Q as

$$Q_s(\omega) = \frac{\omega L [(1 - \omega^2 LC) - CR^2/L]}{R}. \quad (5.3)$$

Interestingly, $Q_s(\omega)$ rises to a maximum at $\omega_p = \omega_o/\sqrt{3}$ and then drops to zero at ω_o , the resonance frequency of the lumped RLC model of the inductor. At the inductor's resonance frequency, $Z_L(j\omega_o) = \text{Real}$. From Equation 5.2, this resonance is seen to occur when

$$\omega = \omega_o = \sqrt{\frac{1}{LC} - \frac{R^2}{L^2}} \text{ r/s, and } \mathcal{Q}(\omega_o) = 0. \quad (5.4)$$

Although the lumped RLC model for the inductor is crude, its behavior does mimic that observed in real inductors at high frequencies. Their Q does increase to a maximum and then decreases to zero as the frequency is further increased. Some practical inductors exhibit multiple peaks in their $Q(\omega)$ curves at high frequencies due to the distributed nature of the stray capacitances. At 1 kHz, the frequency used in most AC bridges used to measure inductors, we are operating well to the left of the peak of the $Q(\omega)$ curve, and capacitive effects are generally second order.

5.3 Components: Capacitor Equivalent Circuits

There are many physical types of capacitors having a wide variety of geometries and using many different types of dielectrics, all depending on the selected application. The basic capacitor consists of a pair of parallel metal plates separated by a dielectric (insulating substance), which can be vacuum, air, sulfur hexafluoride, oil, oil-impregnated paper, glass, mica, metal oxides, or various plastics or ceramics. Each dielectric has its unique properties of DC leakage, dielectric constant, losses at high frequencies, temperature coefficient, etc. At high frequencies, capacitors can be modeled by a pure capacitance surrounded by parasitic inductors due to leads and resistances modeling dielectric losses and leakage. Figure 5.2 illustrates a general, lumped-parameter, HF equivalent circuit for a nonpolarized (nonelectrolytic) capacitor. As in the case of the inductor's equivalent circuit, we may write an expression for the impedance of the model:

$$Z_C(j\omega) = R_S + j\omega L + \frac{R_P(1/j\omega C)}{R_P + 1/j\omega C}. \quad (5.5)$$

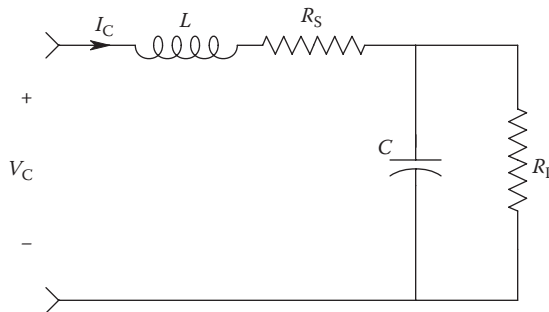


FIGURE 5.2
HF, lumped-parameter, linear circuit model for a practical capacitor.

$Z_C(j\omega)$ may be put into the form, $Z_C = Re + jIm$, using complex algebra:

$$Z_C(j\omega) = \left[R_S + \frac{R_P}{1 + (\omega R_P C)^2} \right] + j \frac{(\omega L - \omega R_P^2 C + \omega^3 R_P^2 L C^2)}{1 + (\omega R_P C)^2}. \quad (5.6)$$

This relation implies that the complex RLC circuit model for the capacitor at high frequencies can be reduced to a simple, equivalent model with a resistor whose value is given by the real term of Equation 5.6. The real term is in series with a reactance given by the second term in Equation 5.6. It is possible to find an expression for the equivalent series capacitance, $C_{s(EQ)}$. From the second term in Equation 5.6, we can write

$$X_C = \frac{+1}{\omega C_{S(EQ)}} = \frac{\omega(L - R_P^2 C) + \omega^3 R_P^2 C^2 L}{1 + (\omega R_P C)^2}, \quad (5.7)$$

from which it is easy to find

$$C_{S(EQ)} = \frac{1 + (\omega R_P C)^2}{\omega^2 [(R_P^2 C - L) - \omega^2 R_P^2 C^2 L]}. \quad (5.8)$$

At low frequencies, where L is negligible, the capacitive reactance reduces to $-1/\omega C$. Note that the series HF capacitor model exhibits resonance at a frequency where the reactance magnitude goes to zero. The resonant frequency of the capacitor is easily found to be

$$\omega_r = \sqrt{\frac{1}{LC} - \frac{1}{(R_P C)^2}} \text{ r/s.} \quad (5.9)$$

Thus, the reactance of the equivalent circuit appears as $-1/\omega C$ at low frequencies. The reactance magnitude decreases with ω until it reaches zero at ω_0 and then at $\omega \gg \omega_0$ appears inductive as $+\omega L$.

Most laboratory bridges do not measure capacitive reactance; rather, they are calibrated in capacitance units (e.g., microfarads) and a *dissipation factor*, D . The D_S is defined as the ratio of the *series equivalent resistance* to the capacitive reactance. In algebraic terms,

$$D_S \equiv \frac{R_{S(EQ)}}{X_{S(EQ)}} = \omega C_{S(EQ)} R_{S(EQ)}. \quad (5.10)$$

Note that at any fixed frequency, a *series R-C* circuit, such as we have been discussing, has an equivalent *parallel R-C* circuit, which has the same impedance. We may also define a *dissipation factor* for the *parallel equivalent circuit*. It is the ratio of the capacitive reactance to the equivalent parallel resistance. In algebraic terms,

$$D_P \equiv \frac{X_{P(EQ)}}{R_{P(EQ)}} = \frac{1}{\omega C_{P(EQ)} R_{P(EQ)}}. \quad (5.11)$$

By equating the impedances or conductances of the series and parallel R - C models, it is possible to derive relations relating one circuit to the other:

$$R_{P(\text{EQ})} = R_{S(\text{EQ})} \left(\frac{D_S^2 + 1}{D_S^2} \right) \quad (5.12)$$

$$C_{P(\text{EQ})} = \frac{C_{S(\text{EQ})}}{1 + D_S^2} \quad (5.13)$$

$$R_{S(\text{EQ})} = R_{P(\text{EQ})} \left(\frac{D_P^2}{D_P^2 + 1} \right) \quad (5.14)$$

$$C_{S(\text{EQ})} = C_{P(\text{EQ})} (1 + D_P^2). \quad (5.15)$$

As you will see, certain capacitance bridge configurations make the assumption that the capacitance under measurement is represented by the series equivalent circuit model, while others use the parallel equivalent circuit. Since at a given frequency the impedances of the two circuits are by definition equal, the reason for using one model or the other lies in the practical derivation of the bridge balance equations in a form that allows two bridge elements to each be uniquely calibrated in C_S and D_S or in C_P and D_P .

5.4 AC Operation of Wheatstone Bridges

In this mode of operation, an AC excitation signal is used, usually ranging from 100 to 1000 Hz. The bridge arms are resistors, and the effects of stray capacitance to ground and inductance are generally negligible. Balance conditions are generally the same as for a DC Wheatstone bridge. However, one advantage of the AC excitation is that an electronic null detector with a lower uncertainty voltage, ΔV_o , can be used. This is probably not important, because bridge accuracy in measuring the unknown resistor, X , is largely due to the accuracy to which arms M , N , and P are known, as we demonstrated in the previous chapter. However, when the bridge is used in the voltage output mode, as with strain gauges, greater threshold sensitivity in measuring small changes in X are realized because amplification of V_o can be done with a low-noise, narrowband, AC amplifier working above the $1/f$ noise region.

Following amplification, the AC output voltage is generally converted into a DC signal proportional to ΔX through the use of a PSR. (PSRs are discussed in detail in Section 2.7.)

Another application for AC Wheatstone bridge operation is where the passage of DC through X will alter the magnitude of X . An example of this phenomenon is when we wish to measure the resistance of electrochemical electrodes (such as used in electrocardiography) attached to the body. The passage of DC causes polarization of the electrodes due to ion migration in the DC electric field. Polarization increases electrode impedance. The use of AC generally avoids this phenomenon.

5.5 AC Bridges

The design, analysis, and application of AC bridges is large field with a rich, old, literature. Most of the texts dealing with this topic in detail were written over 50 years ago. Great emphasis was placed on how an AC bridge approached its null. Vector (circle) diagrams were used to illustrate how the complex V_o behaves due to the interaction of the two variable bridge elements in reaching a null. In the following sections, we do not consider the details of how nulls are approached. Rather, we summarize the conditions at null and discuss the applications of each type of bridge. Note that most modern AC bridges are designed to work at 1000 Hz, although other frequencies may be used for special applications.

Figure 5.3 illustrates a *general* AC bridge, in which the arms are impedances having real and imaginary parts. In general, the AC bridge output (unbalance) voltage can be written as a complex (vector) equation:

$$V_o = V_s \left[\frac{Z_X}{Z_X + Z_P} - \frac{Z_N}{Z_N + Z_M} \right]. \tag{5.16}$$

From Equation 5.16, we find that at null, where $V_o = 0$, we can write the vector equation:

$$Z_X Z_M = Z_N Z_P, \tag{5.17}$$

which leads to the vector equation for Z_X in polar form, assuming bridge null:

$$Z_X = |Z_X| \angle \theta_X = \frac{Z_N Z_P}{Z_M} \angle \theta_N + \theta_P - \theta_M. \tag{5.18}$$

Thus, finding unique expressions for the unknown L and Q , or C and D , requires solving vector equations of the form earlier. Generally, this is done by equating the real terms on both sides of Equation 5.17 and, independently, the imaginary terms.

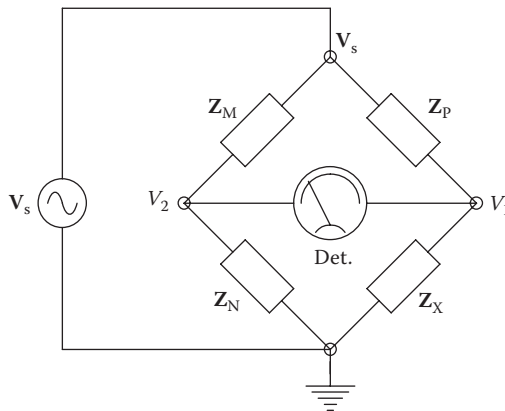


FIGURE 5.3 General configuration for an AC bridge. The null detector (Det.) is a sensitive AC voltmeter assumed to have infinite input resistance.

It can be shown for all, conventional, four-arm, AC bridges that the same conditions for null exist if the null detector is exchanged for the AC source; Equation 5.17 still applies. We present in the following texts a summary of bridge designs used to measure capacitance, inductance, and mutual inductance using the null method.

5.5.1 Bridges Used to Measure Capacitance

In this section, we examine the designs of bridges useful for the measurement of capacitance. Capacitance-measuring bridges can be subdivided into those designs suitable for low-loss (low D) capacitors and those giving best results for lossy (high- D) capacitors.

5.5.1.1 Resistance Ratio Bridge

The resistance ratio bridge is shown in Figure 5.4. It is best used to measure capacitors with low D_s , and it uses the series equivalent capacitor model. It was used in the venerable, GenRad Model 1650A bridge. (GenRad was acquired by Teradyne in 2001.)

At null, $V_o = 0$, and we may write

$$R_3 (R_{XS} + 1/j\omega C_{XS}) = R_2 (R_4 + 1/j\omega C_4). \quad (5.19)$$

Equating the real terms, we find

$$R_{XS} = \frac{R_2 R_4}{R_3}. \quad (5.20)$$

Equating the imaginary terms, we obtain

$$C_{XS} = \frac{R_3 C_4}{R_2}. \quad (5.21)$$

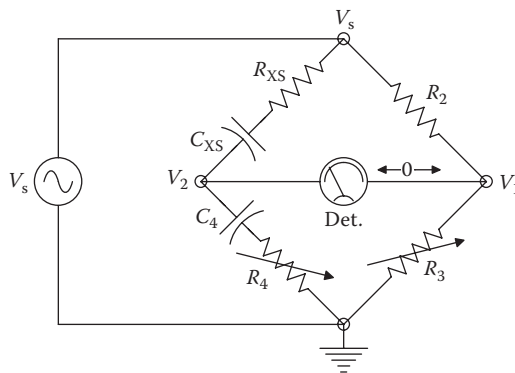


FIGURE 5.4
The resistance ratio bridge, used to measure capacitors with low D_s .

The dissipation factor for the series capacitor model was shown to be $D_s = \omega C_{XS} R_{XS}$. If we substitute the relations for C_{XS} and R_{XS} into D_s , we obtain

$$D_s = \omega C_4 R_4. \quad (5.22)$$

Thus, we see that R_4 can be uniquely calibrated in low D_s , and R_3 should be calibrated in C_{XS} units.

5.5.1.2 Schering Bridge

The Schering bridge is useful for measuring capacitors with high losses (high D_s). In finding the balance conditions for this bridge, it is expedient to use the parallel R - C equivalent circuit, as shown in Figure 5.5. At null, we can write, as before,

$$\mathbf{Z}_X \mathbf{Z}_3 = \mathbf{Z}_2 \mathbf{Z}_4 \quad (5.23)$$

or

$$\mathbf{Z}_3 \mathbf{Y}_2 = \mathbf{Y}_X \mathbf{Z}_4. \quad (5.24)$$

Thus,

$$(R_3 + 1/j\omega C_3) j\omega C_2 = (G_{XP} + j\omega C_{XP}) R_4. \quad (5.25)$$

Now by equating real terms and imaginary terms, we obtain the conditions at balance:

$$C_{XP} = \frac{R_3 C_2}{R_4}, \quad (5.26)$$

$$R_{XP} = \frac{R_4 C_3}{C_2}, \quad (5.27)$$

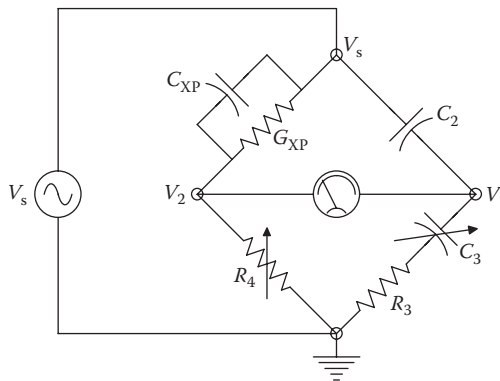


FIGURE 5.5
The Schering bridge, used to measure capacitors with high D_p .

$$D_p = \frac{1}{\omega C_{XP} R_{XP}} = \omega C_3 R_3. \quad (5.28)$$

Notice that to obtain uniqueness in finding C_{XP} and D_p , C_3 can be calibrated in D_p and R_4 can be calibrated in C_{XP} . The other components are fixed for a given range of C_{XP} .

5.5.1.3 Parallel C Bridge

This bridge design, shown in Figure 5.6, was also used in the GR 1650A bridge to measure high D (lossy) capacitors ($0.1 < D_p < 50$). At null, we can write $R_2 Z_4 = R_3 Z_X$. This leads to

$$R_2 (G_{XP} + j\omega C_{XP}) = R_3 (G_4 + j\omega C_4), \quad (5.29)$$

from which we can easily find

$$C_{XP} = \frac{C_4 R_3}{R_2}, \quad (5.30)$$

$$R_{XP} = \frac{R_2 R_4}{R_3}, \quad (5.31)$$

$$D_p = \frac{1}{\omega C_{XP} R_{XP}} = \frac{1}{\omega C_4 R_4}. \quad (5.32)$$

From the earlier equations, it is easy to see that R_4 should be calibrated in D_p and R_3 or R_2 can give C_{XP} .

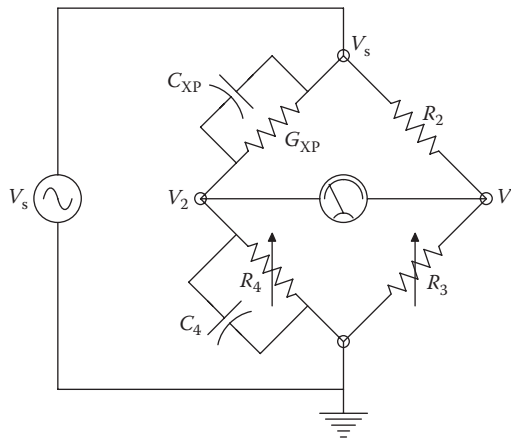


FIGURE 5.6

A parallel C bridge, used to measure capacitors with $0.1 < D_p < 50$.

5.5.1.4 De Sauty Bridge

The De Sauty bridge, shown in Figure 5.7, is a deceptively simple bridge that is often used to produce an output voltage that is proportional to a small change, δC , in one of the capacitors, rather than to measure capacitance. The δC may be caused by a variety of physical phenomena, such as the deflection of a diaphragm due to a pressure difference across it, a change in capacitor plate separation due to a change in material thickness, and a change in capacitor plate separation due to applied force.

Because the output voltage is a nonlinear function of large capacitance changes, the De Sauty bridge is usually operated so that $\delta C/C_0 \ll 1$, so output will be linearly proportional to the QUM. Also, the resistors R_1 and R_2 are made equal to R , and null is achieved by setting $C_4 = C_3 = C$. Now, the output of the bridge as a function of R , C , ω , and δC can be easily written as

$$\mathbf{V}_o = \mathbf{V}_b \frac{j\omega RC(\delta C/C)}{(1 + j\omega RC)(1 + j\omega R(C + \delta C))}. \quad (5.33)$$

From Equation 5.33, we see that three approximate relations for \mathbf{V}_o can be written, depending on the operating frequency. *First*, we let $\omega \gg 1/RC$. The transfer function reduces to

$$\frac{\mathbf{V}_o}{\mathbf{V}_b} = \left(\frac{\delta C}{C} \right) \left[1 - \frac{\delta C}{C} \right] \left(\frac{1}{j\omega RC} \right). \quad (5.34)$$

Generally, the second-order term can be neglected. However, a square-law nonlinear distortion will occur for $\delta C/C \rightarrow 0.1$.

Second, we let the bridge be excited at an LF so that $\omega \ll 1/RC$. Under this condition, we find

$$\frac{\mathbf{V}_o}{\mathbf{V}_b} = \left(\frac{\delta C}{C} \right) (j\omega RC). \quad (5.35)$$

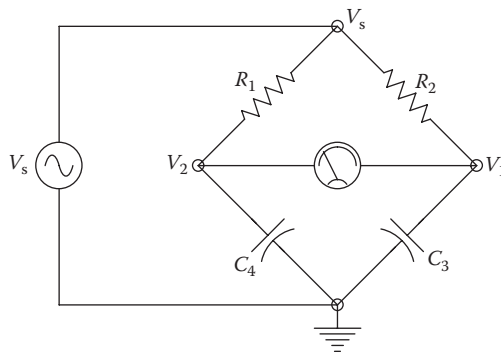


FIGURE 5.7

The De Sauty bridge, used to measure small changes in capacitance: $C_4 = C$; $C_3 = C + \Delta C$; $R_1 = R_2 = R$, and $\omega = 1/RC$.

Third, if $\omega = 1/RC$ (the tuned bridge condition), the output can be written as

$$\frac{V_o}{V_b} = \frac{\delta C}{2C} \text{ (Real)}. \tag{5.36}$$

Maximum sensitivity generally occurs for the third case where $\omega = 1/RC$. In this case, the just detectable δC can be estimated by using the series expansion:

$$\Delta V_o = \frac{\partial V_o}{\partial C} \delta C = \frac{V_b \delta C}{2C}. \tag{5.37}$$

For example, the just detectable δC is found by assuming that the bridge detector AC voltmeter resolution is $\Delta V_o = 0.1 \mu\text{V}$, and the bridge excitation, V_b , is 5 V. If we let $C = 100 \text{ pF}$, then δC_{MIN} is given by

$$\delta C_{\text{MIN}} = \frac{0.1 \times 10^{-6} \times 2 \times 100 \times 10^{-12}}{5} = 4 \times 10^{-6} \text{ pF} = 4 \text{ aF}. \tag{5.38}$$

This is an incredible theoretical sensitivity, generally not reachable in practice because of stray capacitances associated with the bridge arms, detector noise, etc.

5.5.1.5 Wien Bridge

The Wien bridge, illustrated in Figure 5.8, is generally not used to measure capacitors because of the complexity of its solution at null. The Wien bridge is a frequency-dependent null network and, as such, finds application in the design of tuned BP and band-reject (notch) filters and also oscillators. From Figure 5.8, we can write the general form of the bridge balance equation at null:

$$Z_1 Z_3 = Z_2 Z_X. \tag{5.39}$$

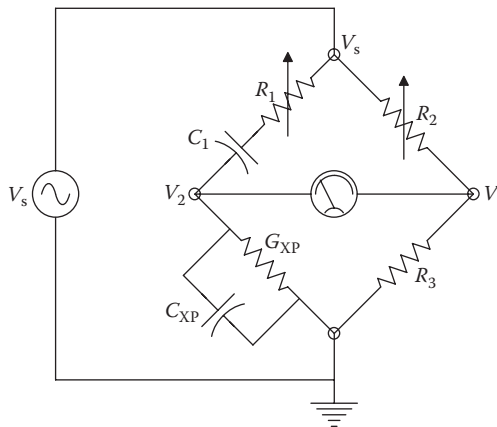


FIGURE 5.8 The ubiquitous Wien bridge; it is frequency dependent. Here it is used to measure capacitance with $D_p \ll 0.1$. R_2 can be calibrated in C_X ; R_1 in D_{PX} .

This can be put as

$$\mathbf{Z}_1 \mathbf{Y}_X = \mathbf{Z}_2 \mathbf{Y}_3 \quad \text{or} \quad (R_1 + 1/j\omega C_1)(G_X + j\omega C_X) = R_2 G_3. \quad (5.40)$$

To find the null conditions, we must solve the simultaneous equations derived from Equation 5.40 by setting real terms to be equal and imaginary terms to be equal:

$$G_X R_1 + C_X / C_1 = R_2 G_3, \quad (5.41)$$

$$-G_X / \omega C_1 + \omega C_X R_1 = 0. \quad (5.42)$$

Their solution leads to

$$C_X = \frac{(R_2 / R_3) C_1}{(\omega^2 C_1^2 R_1^2 + 1)} = \frac{(R_2 / R_3) C_1}{(D_{1S}^2 + 1)} \rightarrow \left(\frac{R_2}{R_3} \right) C_1, \quad (5.43)$$

$$G_X = \frac{(\omega R_1 R_2 / R_3)(\omega C_1^2)}{\omega^2 C_1^2 R_1^2 + 1}, \quad (5.44)$$

$$D_{XP} = \frac{1}{\omega C_X R_X} = \omega R_1 C_1 = D_{1S}. \quad (5.45)$$

Providing $D_{XP} = D_{1S} \ll 0.1$, we can uniquely calibrate R_2 in C_X units and R_1 in D_{XP} units.

The use of the Wien bridge as a frequency-sensitive null network can be demonstrated by writing its transfer function:

$$\frac{\mathbf{V}_o}{\mathbf{V}_b} = \frac{1/(G_X + j\omega C_X)}{1/(G_X + j\omega C_X) + R_1 + 1/j\omega C_1} - \frac{R_3}{R_3 + R_2}, \quad (5.46)$$

which reduces to

$$\frac{\mathbf{V}_o}{\mathbf{V}_b} = \frac{j\omega C_1 R_X}{1 + j\omega(C_1 R_X + C_1 R_1 + C_X R_X) + (j\omega)^2 C_1 R_1 C_X R_X} - \frac{R_3}{R_3 + R_2}. \quad (5.47)$$

Now, if we tune the source frequency to

$$\omega \rightarrow \omega_0 = \left(\frac{1}{C_1 R_1 C_X R_X} \right)^{1/2} \text{r/s}, \quad (5.48)$$

the transfer function becomes

$$\frac{\mathbf{V}_o}{\mathbf{V}_b} = \frac{C_1 R_X}{C_1 R_X + C_1 R_1 + C_X R_X} - \frac{R_3}{R_3 + R_2}. \quad (5.49)$$

This real transfer function goes to zero (null) at $\omega = \omega_0$ if we let

$$R_1 = R_X = R, \tag{5.50}$$

$$C_1 = C_X = C, \tag{5.51}$$

$$R_2 = 2R_3 = R. \tag{5.52}$$

Then we have

$$\frac{V_o}{V_b} = \frac{1}{3} - \frac{1}{3} = 0 \text{ (Null)}. \tag{5.53}$$

As was mentioned earlier, the frequency-dependent null of the Wien bridge can be exploited to create tuned filters and oscillators useful in instrumentation and control systems.

5.5.1.6 Commutated Capacitor Bridge

A commutated capacitor bridge is shown in Figure 5.9. Unlike a conventional AC bridge, this bridge uses a DC V_s . Such bridges do not allow measurement of capacitor D. However, they are often used for high-accuracy, LF applications. A MOS switch is driven by a square wave clock with period T . For $T/2$ s, the capacitor is connected to node "a" and charges toward V_s through R_2 . For an alternate $T/2$ s, C_X is allowed to discharge

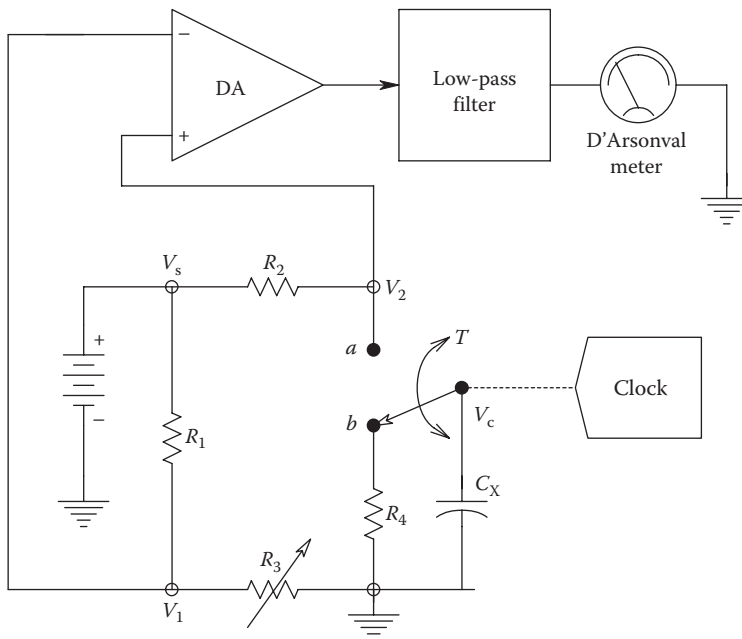


FIGURE 5.9

A commutated capacitor bridge. Unlike other AC bridges, its null detector is a DC meter responding to the average null voltage, $\overline{V_2 - V_1}$.

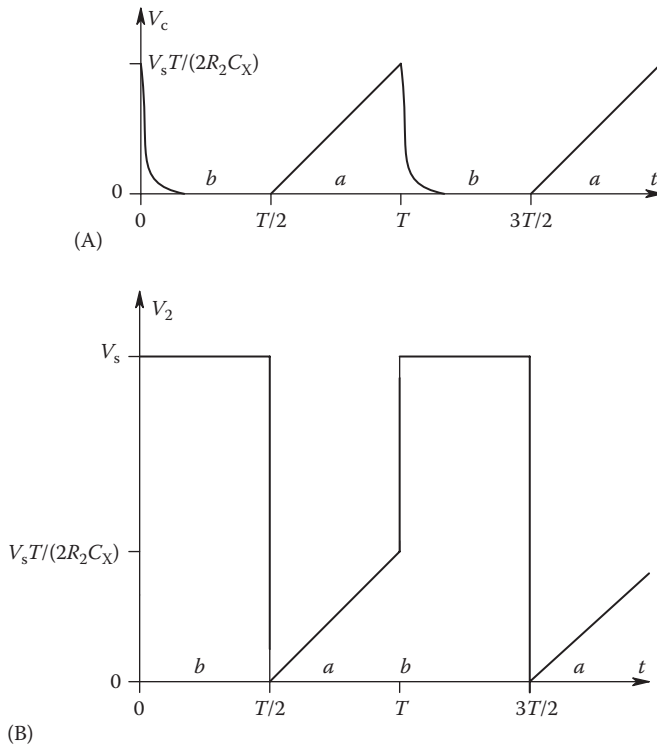


FIGURE 5.10
 (A) Voltage waveform across C_x in the commutated capacitor bridge, $v_c(t)$. Note $RC_x \ll R_2 C_x$. (B) The voltage, $v_2(t)$.

through R_4 to ground. The waveforms at nodes “a” and “ V_c ” are shown in Figure 5.10A and B, respectively. We assume that the clock period and resistors R_2 and R_4 are chosen so $R_4 C_x \ll T/2$ and $R_2 C_x \gg T/2$. We also assume that the null detector responds to the average of $(V_2 - V_1)$. Hence, at null, we find that

$$\frac{V_s R_3}{R_1 + R_3} = \left(\frac{V_s}{2} \right) \left[1 + T/4R_2 C_x \right], \tag{5.54}$$

from which we can solve for C_x

$$C_x = \frac{T(R_3 + R_1)}{4R_2(R_3 - R_1)}, \tag{5.55}$$

where, obviously, $R_3 > R_1$. R_3 or R_2 can be varied to obtain null and for a given range setting (fixed values of T, R_2 , and R_1) can be calibrated according to Equation 5.55 earlier.

5.5.2 Bridges Used to Measure Inductance and Mutual Inductance

As in the case of bridges used for measuring capacitance, the inductance bridges can be subdivided into those optimal, in terms of reaching null, for measuring high- Q inductors

and those best suited for the measurement of low- Q inductors. In addition, we describe in the following texts the design of two specialized bridges used to measure the mutual inductance of power and audio-frequency transformers.

5.5.2.1 Maxwell Bridge

As the name implies, this is a well-known bridge that is used to measure low- Q inductors having Q_s in the range of 0.02–10. The Maxwell bridge was used in the GenRad Model 1650A multipurpose bridge; its circuit is shown in Figure 5.11. At null, we can write

$$\mathbf{Z}_X = \mathbf{Z}_2 \mathbf{Y}_3 \mathbf{Z}_4 \quad (5.56)$$

or

$$R_X + j\omega L_X = R_2 (G_3 + j\omega C_3) R_4. \quad (5.57)$$

From Equation 5.54, we easily find

$$L_X = C_3 R_2 R_4, \quad (5.58)$$

$$R_X = \frac{R_2 R_4}{R_3}, \quad (5.59)$$

$$Q_X = \frac{\omega L_X}{R_X} = \omega C_3 R_3. \quad (5.60)$$

Thus, R_4 is calibrated in inductance units, and R_3 's scale reads Q . In the GR 1650A bridge, R_4 can be varied from 0 to 11.7 k Ω , and R_3 can be set from 0 to 16 k Ω .

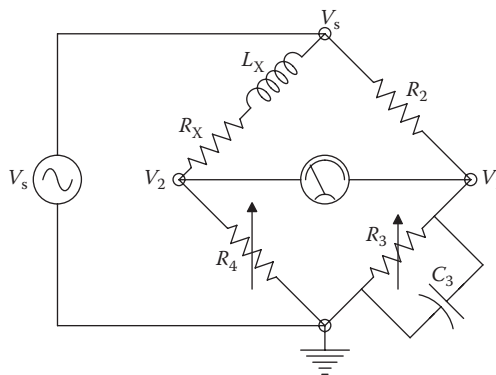


FIGURE 5.11

The Maxwell bridge, used to measure low- Q_s inductors ($0.02 < Q_s < 10$).

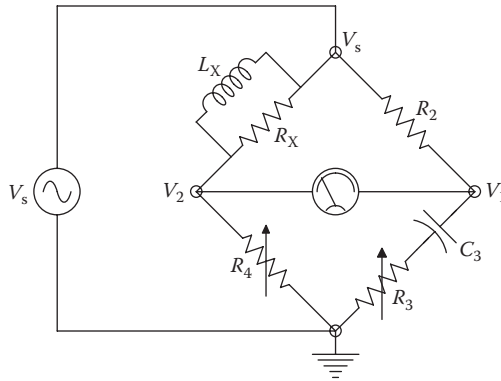


FIGURE 5.12
The GenRad parallel inductance model bridge for high- Q_p inductors ($1 < Q_p < \infty$). R_3 is calibrated in Q_p ; R_4 in L_{pX} .

5.5.2.2 Parallel Inductance Bridge

This bridge, shown in Figure 5.12, is also used in the GR 1650A bridge to measure high- Q inductors ($1 < Q_p < \infty$). It is somewhat unusual in that it uses a parallel equivalent circuit for the inductor. At a given frequency, any series R - L circuit can be made to be equal in impedance to a parallel R - L circuit. We describe these equivalences in the following. At balance, we can write

$$R_2R_4 = (R_3 + 1/j\omega C_3) \frac{j\omega L_X R_X}{R_X + j\omega L_X}, \tag{5.61}$$

from which we obtain

$$j\omega L_X R_2 R_4 + R_X R_2 R_4 = j\omega L_X R_X R_3 + L_X R_X / C_3. \tag{5.62}$$

Equating real terms and then imaginary terms in Equation 5.62, we finally obtain the expressions for L_{XP} , R_{XP} and Q_p :

$$L_{XP} = C_3 R_2 R_4, \tag{5.63}$$

$$R_{XP} = \frac{R_2 R_4}{R_3}, \tag{5.64}$$

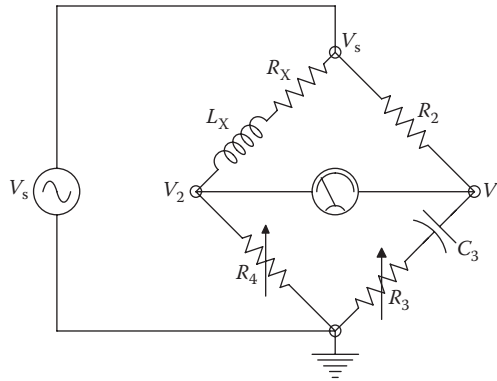
$$Q_p = \frac{R_X}{\omega L_X} = \frac{1}{\omega C_3 R_3}. \tag{5.65}$$

Thus, it is expedient to make R_4 the L dial and R_3 the *high- Q* dial on this bridge.

5.5.2.3 Hay Bridge

The Hay bridge, shown in Figure 5.13, uses the series R - L model for an inductor to measure the inductance and Q of high- Q coils. Assuming null, we find

$$(j\omega L_X + R_X)(R_3 + 1/j\omega C_3) = R_2 R_4. \tag{5.66}$$

**FIGURE 5.13**

The Hay bridge is used to measure high- Q_S inductors. R_3 is calibrated in Q_S and R_4 in L_{SX} .

The balance equation, Equation 5.66, can be broken down to two simultaneous equations:

$$L_X/C_3 + R_X R_3 = R_2 R_4, \quad (5.67)$$

$$\omega L_X R_3 - R_X/\omega C_3 = 0. \quad (5.68)$$

Their solution yields

$$L_X = \frac{R_4 R_2 C_3}{(\omega^2 R_3^2 C_3^2 + 1)} = \frac{R_4 R_2 C_3}{(Q_S^2 + 1)}, \quad (5.69)$$

$$R_X = \frac{\omega R_3 R_2 R_4 (\omega C_3^2)}{(\omega^2 R_3^2 C_3^2 + 1)}, \quad (5.70)$$

$$Q_S = \frac{\omega L_X}{R_X} = \frac{1}{\omega C_3 R_3}. \quad (5.71)$$

Here, R_3 may be calibrated in Q_S and R_4 in inductance units. Inductance calibration is substantially independent of $Q_S(R_3)$ as long as $Q_S > 10$.

5.5.2.4 Owen Bridge

The Owen bridge, shown in Figure 5.14, uses the conventional, series inductance model and is best used on large, low- Q_S inductors. This bridge is somewhat unique in that null may be obtained by varying both elements in the four arm, R_4 , and C_4 . At null, we have

$$\frac{j\omega L_X + R_X}{j\omega C_1} = R_2 (R_4 + 1/j\omega C_4). \quad (5.72)$$

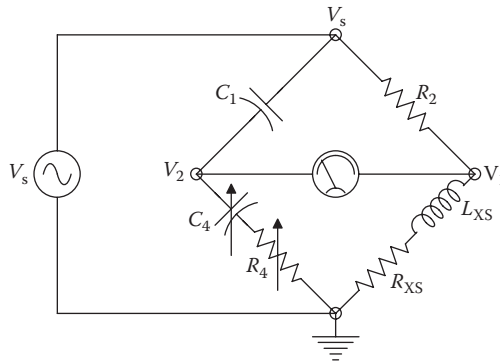


FIGURE 5.14
The Owen bridge is best suited to measure large, low- Q_S inductors. R_4 can be calibrated in inductance, L_{XS} , and C_4 in the series equivalent resistance, R_{XS} .

By equating the real terms and then the imaginary terms in Equation 5.72, it is easy to find

$$R_{XS} = \frac{R_2 C_1}{C_4}, \tag{5.73}$$

$$L_{XS} = R_2 C_1 R_4. \tag{5.74}$$

Hence, R_4 can be uniquely calibrated in inductance and C_4 in resistance (nonlinear scale). A major disadvantage of the Owen bridge is that it requires a precisely calibrated variable capacitor (C_4) and another is that for high- Q , low inductance coils, impractically large C_4 values may be required.

5.5.2.5 Anderson Bridge

The Anderson bridge is illustrated in Figure 5.15. According to Stout (1950), the Anderson bridge gives the best convergence to nulls for low- Q coils. To analyze this bridge, we note

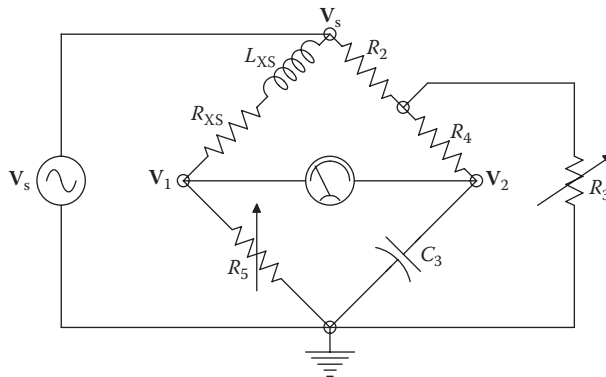


FIGURE 5.15
The Anderson bridge works best with low- Q_S inductors. It unfortunately has relatively complex balance equations.

that at null, $V_o = V_1 - V_2 = 0$, and voltage-divider relations can be used to find V_1 and V_2 . Thus, we have

$$\frac{V_o}{V_s} = \frac{R_5}{R_5 + R_X + j\omega L_X} - \frac{R_3}{R_3 + R_2 + j\omega C[R_3 R_4 + R_3 R_2 + R_4 R_2]}. \quad (5.75)$$

Setting $V_o = 0$, we find

$$R_X = \frac{R_2 R_5}{R_3}, \quad (5.76)$$

$$L_X = CR_5 [R_4 + R_2 + R_4 R_2 / R_3], \quad (5.77)$$

$$Q_X = \omega CR_3 [1 + R_4 / R_2 + R_4 / R_3]. \quad (5.78)$$

It is easily seen that the price we pay in using this bridge, which is easily balanced with low- Q inductors, is a relatively complex set of balance equations. R_5 can be calibrated for L_X , and R_3 used to read Q_X .

5.5.2.6 Heaviside Mutual Inductance Bridge

Before discussing the circuit and balance conditions for the Heaviside mutual inductance bridge, we should review what is meant by the *mutual inductance* of a transformer or a pair of magnetically coupled coils. Figure 5.16 illustrates a simple equivalent circuit for a transformer operating at low frequencies. Under AC SS conditions, we may write the loop equations

$$V_1 = I_1 (R_1 + j\omega L_1) + I_2 (j\omega M), \quad (5.79)$$

$$V_2 = I_1 (j\omega M) + I_2 (R_2 + j\omega L_2). \quad (5.80)$$

M is the mutual inductance in henrys, given a positive sign for the dot convention shown in Figure 5.16.

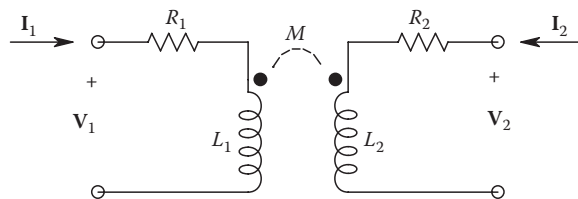


FIGURE 5.16

Circuit model for a transformer with mutual inductance at low frequencies. L_1 and L_2 are the self-inductances of the primary and secondary windings, respectively.

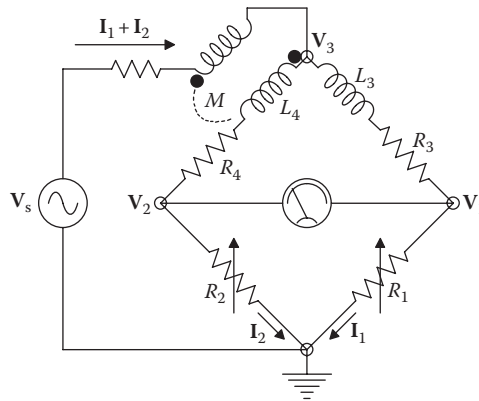


FIGURE 5.17
The Heaviside mutual inductance bridge.

Figure 5.17 illustrates the circuit of a Heaviside bridge. At null, $V_o = V_1 - V_2 = 0$; hence, we can write

$$I_1 R_1 = I_2 R_2 \quad (5.81)$$

and also

$$\mathbf{V}_3 = \mathbf{I}_1 (R_1 + R_3 + j\omega L_3) = \mathbf{I}_2 (R_2 + R_4 + j\omega L_4) + j\omega M (\mathbf{I}_1 + \mathbf{I}_2). \quad (5.82)$$

If we solve for \mathbf{I}_1 in Equation 5.81 and substitute it into Equation 5.82, and then equate real terms and then imaginary terms, we find that at null,

$$R_4 = \frac{R_2 R_3}{R_1} \quad (5.83)$$

and

$$M = \frac{R_2 L_3 - R_1 L_4}{R_1 + R_2}. \quad (5.84)$$

Note that M may have either sign, depending on the dots. The transformer's primary self-inductance, L_4 , must be measured with the secondary open-circuited. There is no need to know the secondary self-inductance or the primary resistance when using the Heaviside mutual inductance bridge.

5.5.2.7 Heydweiller Mutual Inductance "Bridge"

The Heydweiller means of measuring transformer mutual inductance uses a *bridge* with one arm a short circuit (see Figure 5.18). Thus, the conventional method of examining $\mathbf{Z}_1 \mathbf{Z}_3 = \mathbf{Z}_2 \mathbf{Z}_4$ will not work with this bridge! Instead, we note that at null, $\mathbf{V}_2 = \mathbf{V}_1 \rightarrow 0$; thus, by use of KVL,

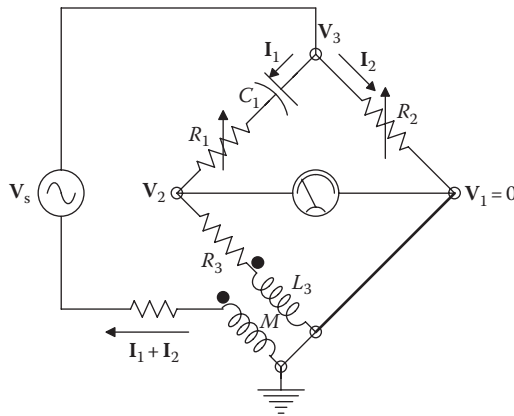


FIGURE 5.18
The Heydweiller mutual inductance null circuit.

$$V_2 = 0 = I_1 (R_3 + j\omega L_3) - j\omega M (I_1 + I_2) \tag{5.85}$$

and also

$$I_2 = \frac{I_1 (R_1 + 1/j\omega C_1)}{R_2} \tag{5.86}$$

If Equation 5.86 is substituted into Equation 5.85 and the sum of the real terms is set equal to zero, we find

$$M = C_1 R_2 R_3 \tag{5.87}$$

Equating the sum of the imaginary terms to zero leads to

$$L_3 = C_1 R_3 (R_1 + R_2) \tag{5.88}$$

Note that R_3 must be known, as it appears in both the expression for M and the primary self-inductance, L_3 . R_1 and R_2 are manipulated to obtain the null.

5.5.3 Null Method of Measuring Transistor Small-Signal Transconductance and Feedback Capacitance

The null circuits discussed in the following are used to measure BJT or FET small-signal transconductance at a given DC quiescent operating point. An audio-frequency AC signal, usually 1 kHz, is used. Measurement of small-signal transconductance is important to verify that transistors are good and to match pairs for like characteristics when building a discrete differential amplifier.

First, we illustrate the small-signal, modified hybrid-pi model for a BJT operating in its linear region (i.e., neither cutoff nor saturated). This model, shown in Figure 5.19,

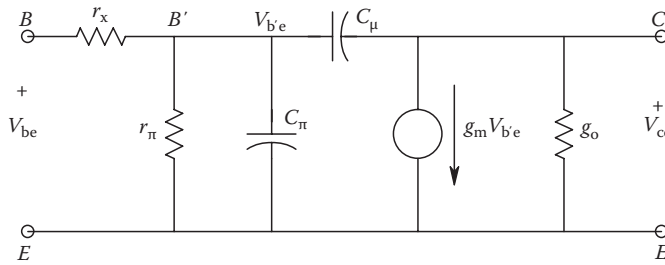


FIGURE 5.19
Simplified, hybrid-pi, HF, small-signal circuit model for BJTs.

is generally accepted to be valid for frequencies up to the BJT's $f_T/3$. The small-signal transconductance, g_m , is defined as

$$g_m \equiv \left(\frac{\partial i_c}{\partial v_{be}} \right), \quad v_{ce} = 0. \tag{5.89}$$

The BJT's g_m may be approximated by

$$g_m \approx \frac{h_{fe}}{r_\pi} = \frac{I_{CQ}}{V_T}. \tag{5.90}$$

The capacitance C_μ is the DC voltage-variable capacitance of the reverse-biased, collector–base junction evaluated at the DC operating point of the BJT. C_π is the capacitance of the forward-biased base–emitter junction, measured at the operating point; it is also a function of the DC voltage across the junction. r_π is the small-signal base input resistance. It is approximately equal to V_T/I_{BQ} . The base input spreading resistance, r_x , is generally less than 100 Ω , and we set it equal to zero to simplify analysis. The output conductance, g_o , is generally very small, but its exact value is not important in determining g_m .

Figure 5.20 illustrates the complete circuit of the transconductance bridge. Resistors R_B and R_C and DC source V_{CC} are used to set up the DC quiescent operating (Q) point of the BJT. A is a milliammeter to read I_{CQ} , and V is a DC voltmeter to read V_{CEQ} . A small, audio-frequency, sinusoidal signal, v_1 , is applied to the transistor's base through a large DC blocking capacitor, C_1 . Simultaneously, the inverted AC signal, $-v_1$, is applied to the BJT's collector. Capacitor C_N and resistor R_1 are adjusted to get a null or minimum AC signal at the v_c node. A node equation for v_c can be written for the small-signal equivalent circuit of Figure 5.21. Note that $v_{be} \cong v_1$. At null,

$$0 = V_c [G_1 + G_c + j\omega(C_\mu + C_N)] - V_1 (j\omega C + G_1) - j\omega C_N (-V_1) + g_m V_1. \tag{5.91}$$

For $V_c \rightarrow 0$, it is evident that

$$g_m = G_1 \tag{5.92}$$

and

$$C_\mu = C_N. \tag{5.93}$$

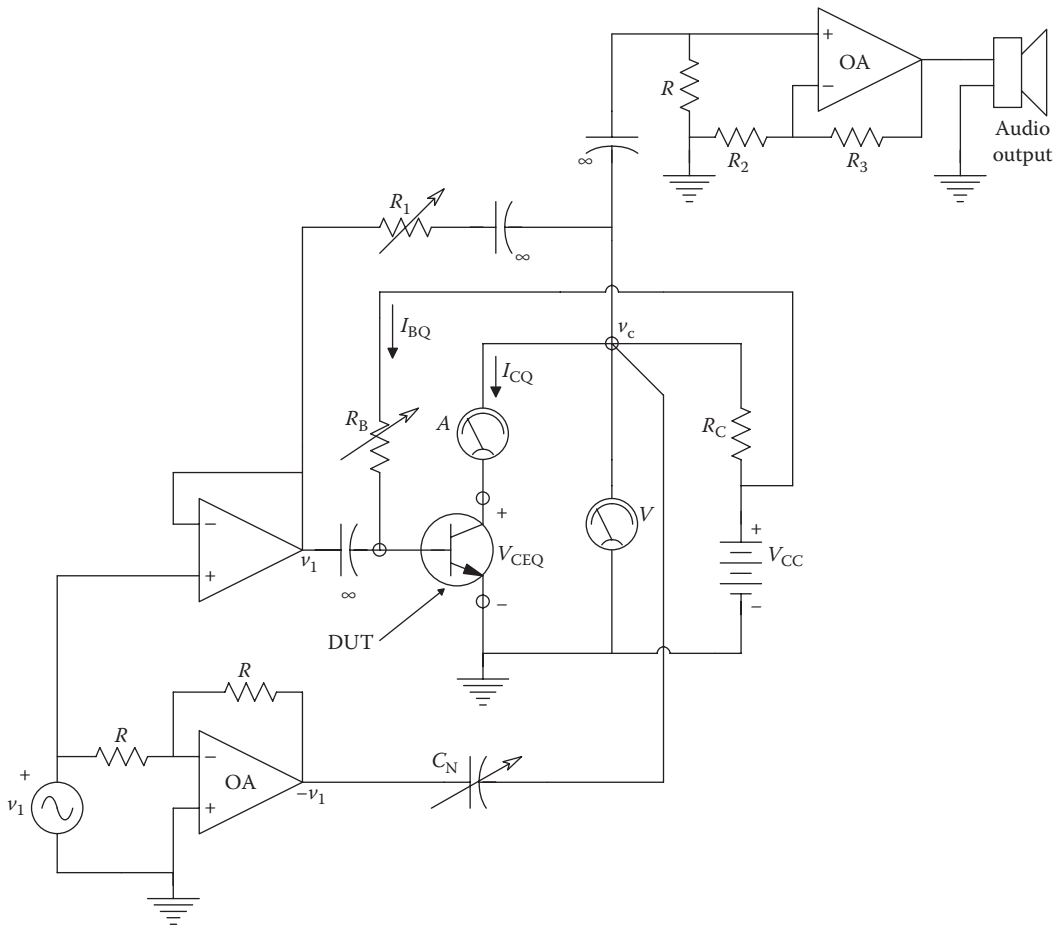


FIGURE 5.20

Circuit of a null system used to measure the small-signal transconductance of BJTs. OAs are ideal operational amplifiers. Capacitors marked “ ∞ ” are large and have negligible reactance at the operating frequency. Meter A is a DC milliammeter used to measure the BJT’s I_{CQ} . Meter V is a DC voltmeter used to measure V_{CEQ} . The null is sensed acoustically with a loudspeaker or headphones.

A similar g_m null circuit can be used to measure the small-signal transconductance of various types of FETs. Note that a JFET’s g_m at the DC operating point can be shown to be given by

$$g_m = \left(\frac{\partial i_d}{\partial v_{gs}} \right)_{v_{ds} \rightarrow 0} = g_{m0} \sqrt{\frac{I_{DQ}}{I_{DSS}}} = g_{m0} (1 - V_{GS}/V_P), \quad (5.94)$$

where

g_{m0} is the small-signal transconductance evaluated for $V_{GS} = 0$

V_P is the pinch-off voltage (for JFETs)

I_{DSS} is the JFET’s DC drain current measured for $V_{GS} = 0$ and $V_{DS} \gg V_{GS} + V_P$

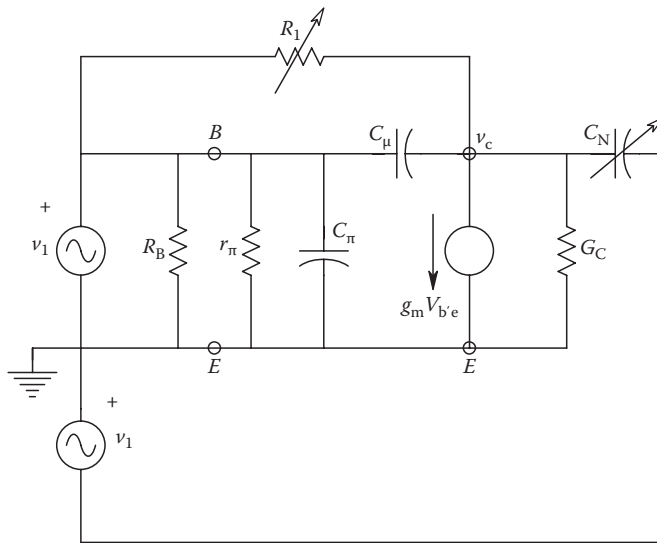


FIGURE 5.21
Small-signal equivalent circuit of the g_m measuring circuit of Figure 5.20.

5.6 Chapter Summary

In this chapter, we have reviewed the more important audio-frequency AC bridges and null systems used for making precision measurements of capacitance, capacitor dissipation factor (D), inductance, inductor Q , mutual inductance, and transistor small-signal transconductance. General conditions for bridge null were derived, and the conditions at null for each circuit have been presented. No attempt has been made to discuss the effects of stray capacitances between bridge elements and the capacitance to ground offered by the two detector terminals. These are generally not significant sources of measurement error for most AC bridges operating at 1 kHz. A common means of compensating for detector capacitance to ground is by the Wagner earth circuit (see Stout 1950, Sections 9 through 11 for details).

Measurement of circuit parameters at high frequencies (including video and radio frequencies) requires special apparatus such as the Q meter or instruments such as the HP/Agilent Model 4191A RF Impedance Analyzer or the Agilent Technologies' Model 4294A RF Impedance Analyzer (40 Hz–110 MHz).

Problems

- 5.1 The bridged- T circuit provides an alternate means to four-arm bridges for the measurement of circuit element parameters. Figure P5.1 illustrates a bridged- T circuit used to measure the parallel model for an inductor. A null at the V_o node is obtained by varying the two capacitors together, ω of the sinusoidal source and R . Assume $V_o = 0$. Find expressions for ω , L_p , R_p , and Q_p in terms of R and C . The expressions for L_p and R_p should not contain terms in R_p and L_p , respectively. In general, $Q_p = R_p/(\omega L_p)$.

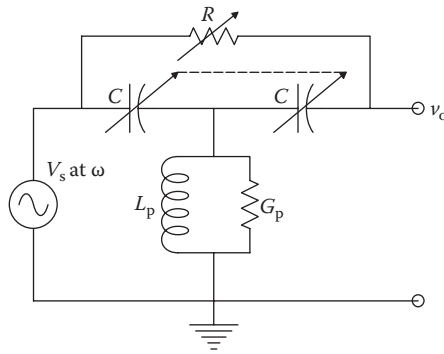


FIGURE P5.1

- 5.2 Repeat Problem 5.1 for the bridged- T circuit of Figure P5.2, except find expressions for ω , L_s , R_s , and Q_s , where $Q_s = \omega L_s / R_s$.
- 5.3 In the bridged- T circuit of Figure P5.3, find ω_0 for null. Also find a numerical value for R_p .
- 5.4 Figure P5.4 illustrates a parallel- T null circuit. Find the frequency ω_0 at which the $V_o \rightarrow 0$.

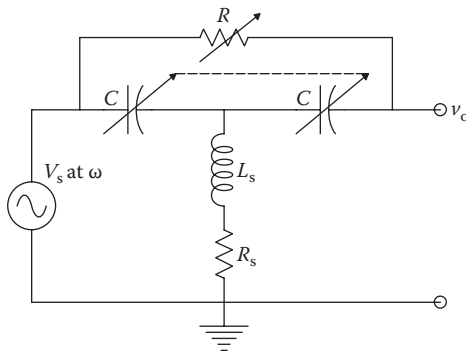


FIGURE P5.2

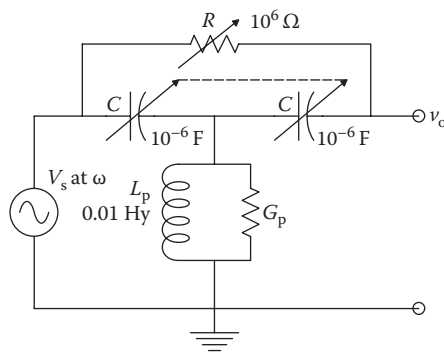


FIGURE P5.3

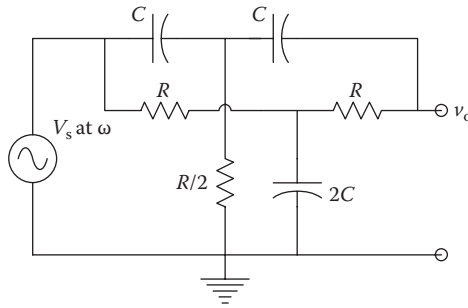


FIGURE P5.4

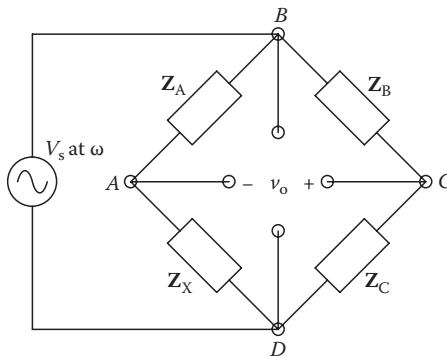


FIGURE P5.5

- 5.5 Prove that the vector conditions for null (bridge balance) are unchanged if you exchange detector and source nodes, as shown in Figure P5.5.
- 5.6 At a given frequency, ω , a series R - L circuit can be exactly replaced by a parallel R - L circuit. Find expressions for L_p and R_p in terms of L_s , R_s , and $Q_s = \omega L_s/R_s$.
- 5.7 Figure P5.7 illustrates a De Sauty bridge in which the capacitors change differentially. The source is $v_s(t) = 5 \sin(\omega t)$. Assume $\omega = 1/RC_o$ r/s. The detector can only resolve $\delta V_o = 5 \mu\text{V}$ pk. Find the smallest $\Delta C/C_o$ resolvable.

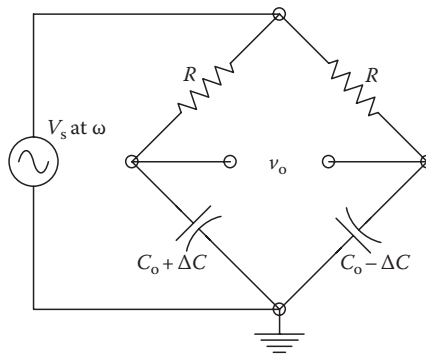


FIGURE P5.7

- 5.8 A four-arm AC bridge operated at 1 kHz is at null. $Z_{AB} = 1500 \Omega$, $Z_{BC} = 1000 \Omega$, $Z_{CD} = Z_X$, $Z_{DA} = 10 + 1/j\omega \cdot 10^{-7} \Omega$. Z_X is a series resistance, R_x , and a reactive element. Find the numerical value of R_x and the value and kind of reactance (L or C).
- 5.9 A four-arm AC bridge is operated at $\omega = 2\pi f = 10^4$ r/s. Node D is grounded. Arm AB is a 10^{-10} F capacitor. Arm BC contains the unknown. Arm CD is a pure $10^4 \Omega$ resistor, and arm DA is a 12.5 k Ω resistor in parallel with a variable capacitor, C_{DA} . Null is obtained with $C_{DA} = 120$ pF. Find the two series elements of Z_{BC} .
- 5.10 A Wien bridge is shown in Figure P5.10. Find f_o of the source and R_x that will give null.
- 5.11 A De Sauty bridge is shown in Figure P5.11:
 A. Find Z_X .
 B. Now Z_{BC} is made 1002Ω , unbalancing the bridge. Find an expression for $v_o(t)$.
- 5.12 A four-arm AC bridge is shown in Figure P5.12. The oscillator frequency, $\omega = 50,000$ r/s. Find Z_X at null. Assume a series, $R_x + jX_x$ model.
- 5.13 A four-arm AC bridge is shown in Figure P5.13. $f = 1000$ Hz. Assuming null, find the constants of arm CD.

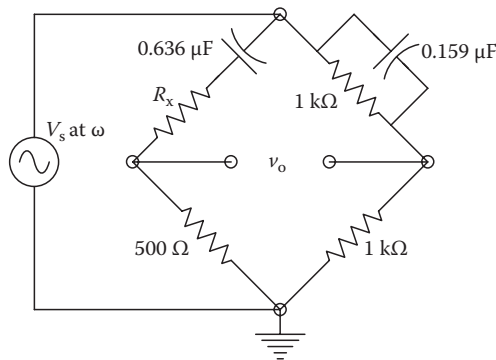


FIGURE P5.10

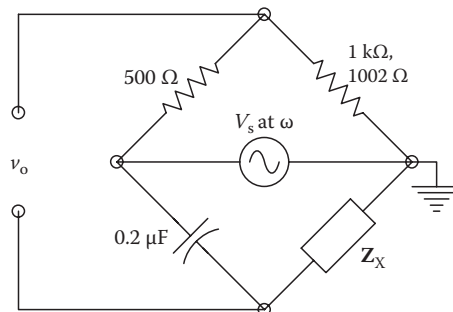


FIGURE P5.11

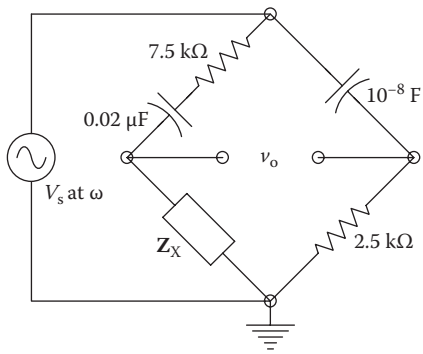


FIGURE P5.12

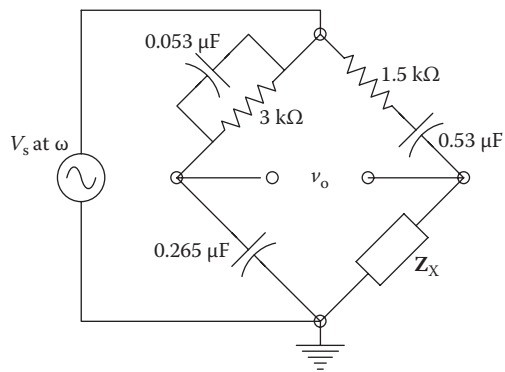


FIGURE P5.13

6

Survey of Sensor Mechanisms

6.1 Introduction

An input *sensor* or *transducer* is a device that permits the conversion of energy from one form to another. It is the first (input) element in an instrumentation/measurement system. For example, it might convert temperature to voltage. A broader definition might substitute *information* for *energy*. Its linearity, range, noise, and dynamic response largely determine the resolution, sensitivity, and bandwidth of the overall system.

A *transducer* is a sensor that in some way obeys reciprocity; for example, its physical input quantity is converted to an output voltage, and if a voltage is applied to the output terminals, the input quantity is generated at the input interface of the transducer. Examples of transducers include piezoelectric crystals used to sense force or pressure. Electrodynamic devices such as loudspeakers, piezoelectric crystals, and d'Arsonval meter movements are transducers. Obviously, not all sensors are transducers.

There are two approaches to categorizing sensors. One way, which we use in Chapter 7, is to group together all those different types of sensors used for a given application, such as the measurement of fluid pressure. The method, which we use in this chapter, is to group sensors by the mechanism by which they work, such as the generation of an open-circuit voltage due to the input QUM or a change in resistance proportional to the QUM.

6.2 Categories of Sensor Mechanisms

In this section, we present in outline form a comprehensive (yet incomplete) list of the categories of mechanisms by which sensors work. In some cases, a physical QUM changes the resistance of a sensor; in other cases, the sensors' capacitance or inductance is affected; other sensors act as generators, changing the QUM to a proportional voltage, current or charge. In the following sections, we elaborate on details of representative sensors and discuss their dynamic ranges, bandwidths, and the auxiliary circuitry needed to produce a useful electrical (analog) output.

6.2.1 Resistive Sensors

The resistance of practically all resistive sensors varies around some baseline or average value, R_0 , as the input quantity varies. Consequently, the most widely used means for converting the change in R due to the input to an output voltage is to include the transducer as

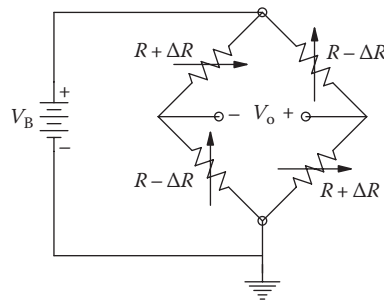


FIGURE 6.1

A Wheatstone bridge with four active arms. Bridge sensitivity is four times that of a bridge with a single variable resistance.

an arm of a Wheatstone bridge (see Section 4.2) or as part of an Anderson constant current loop circuit (see Section 4.4). As long as the ratio $\Delta R/R_0$ is $\ll 1$, the bridge output voltage will be linear with ΔR . As $\Delta R/R_0 \rightarrow 1$, the bridge output voltage vs. $\Delta R/R_0$ tends to saturate; hence, sensitivity and linearity are lost. In some resistive sensors, such as *unbonded strain gauge* force sensors, we have available two resistors whose values *increase* linearly ($R_0 + \Delta R$) with the applied force and two resistors whose values *decrease* linearly ($R_0 - \Delta R$) with the input force. These resistors can be assembled into a four-active-arm Wheatstone bridge that has an output linear in ΔR (see Figure 6.1).

6.2.1.1 Resistive Temperature Sensors

The electrical resistance of all metals and alloys *increases* with temperature. This increase can be modeled by a power series equation of the form

$$R(T) \approx R_0(25^\circ\text{C}) + \alpha(T - 25) + \beta(T - 25)^2. \quad (6.1)$$

Here, 25°C is taken as the reference temperature, T_0 . The tempco of a resistive conductor is defined as α :

$$\alpha \equiv \frac{dR(T)/dT}{R(T)}. \quad (6.2)$$

Table 6.1 gives tempcos and useful temperature ranges for various metals used as resistance thermometers, or RTDs.

RTDs are generally used in a Wheatstone bridge configuration. In order to compensate for lead lengths and thermoelectric EMFs at junctions with the RTD material, extra leads are run to the measurement site, as shown in Figure 6.2.

TABLE 6.1

Properties of Some Conductors Used for RTDs

Material	Tempco (α)	Useful Range ($^\circ\text{C}$)	Relative Resistivity (vs. Cu)
Copper	0.00393	-200 to +260	1
Platinum	0.00390	-200 to +850	6.16
Nickel	0.0067	-80 to +320	4.4
Thermistors	-0.05	-100 to +300	—

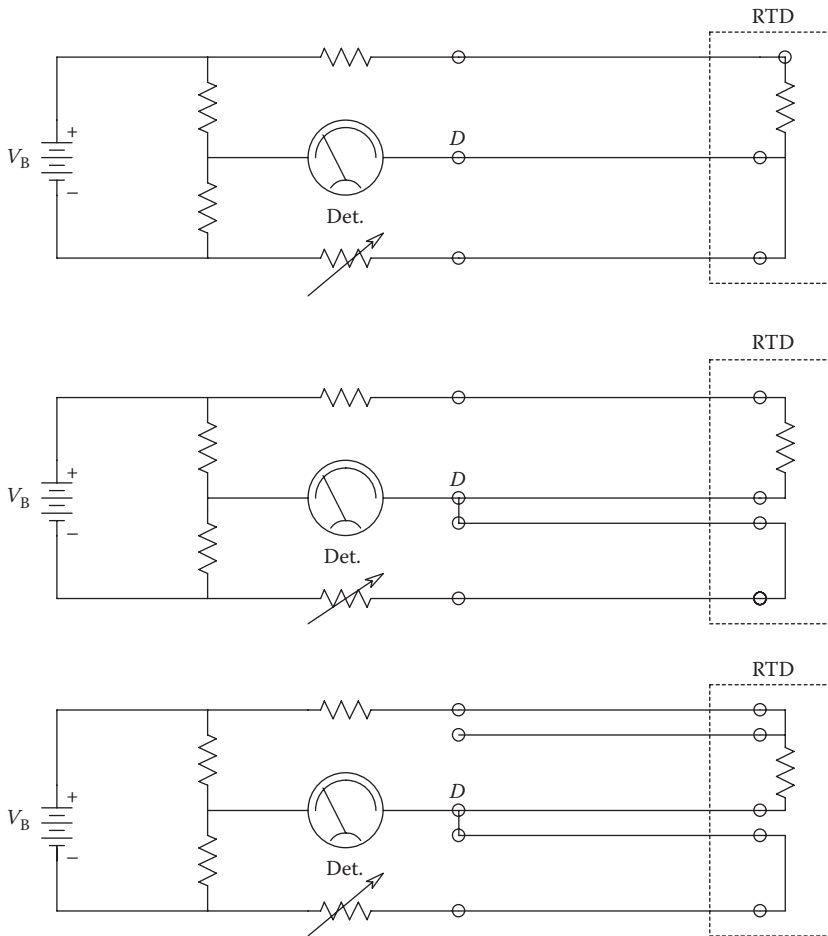


FIGURE 6.2

Three methods for compensating for load resistance in a Wheatstone bridge used with a remote RTD. The RTD is the resistance element in the rectangular block. (From Beckwith, T.G. and Buck, N.L., *Mechanical Measurements*, Addison-Wesley, Reading, MA, 1961. With permission.)

Another type of RTD material that has a much higher tempco than pure metals or metallic alloys is the thermistor, which is made from amorphous semiconductor materials, generally sintered mixtures of oxides, sulfides, and silicates of such elements as Al, C, Co, Cu, Fe, Mg, Mn, Ni, Ti, U, and Zn. The resistance of negative tempco (NTC) thermistors generally follows the rule

$$R(T) = R_o \exp \left[\beta \left(\frac{1}{T} - \frac{1}{T_o} \right) \right] \tag{6.3}$$

where T and T_o are Kelvin temperatures and T_o is customarily taken to be 298°.

The NTC thermistor tempco is easily calculated to be

$$\alpha = \frac{dR/dT}{R} = -\beta \frac{1}{T^2}. \tag{6.4}$$

The tempco evaluated at 298 K with $\beta = 4000$ is -0.045 . When used in a Wheatstone bridge circuit, it is possible to resolve temperature changes of ca. a millidegree Celsius.

If the power dissipation of the thermistor is sufficient to cause it to warm to *above* ambient temperature, then fluid at ambient temperature moving past the thermistor will conduct heat away from it, cooling it and producing an increase in its resistance. Thus, this mode of operation may be used to measure fluid velocity when the fluid temperature is known.

pn junction devices (diodes) can also be used as temperature sensors. For example, the current in a forward-biased diode is given by the well-known approximation

$$i_D = I_{rs} \left[\exp\left(\frac{v_D q}{nkT}\right) - 1 \right] \quad (6.5)$$

where

v_D is the voltage across the diode (positive for forward bias)

n is a constant ranging from 1 to 2

k is the Boltzmann constant

T is the junction temperature in Kelvins

I_{rs} is the reverse saturation current

Experimentally, it has been observed that I_{rs} varies with temperature according to the relation (Millman 1979):

$$I_{rs}(T) \cong \frac{I_{rs}(T_0) 2(T - T_0)}{10} \quad (6.6)$$

where $I_{rs}(T_0)$ is the reverse saturation current measured at temperature T_0 . T_0 is generally taken as 300 K.

The reverse saturation current approximately doubles for each 10°C rise in temperature. In general, the current through a forward-biased *pn* junction diode in a simple, series circuit with a resistor and a battery will increase, and the voltage across the diode will decrease with increasing temperature. Figure 6.3 illustrates a simple Wheatstone bridge circuit that can be used to measure temperature using a diode. In this case, the diode forward voltage decreases approximately by $2 \text{ mV}/^\circ\text{C}$.

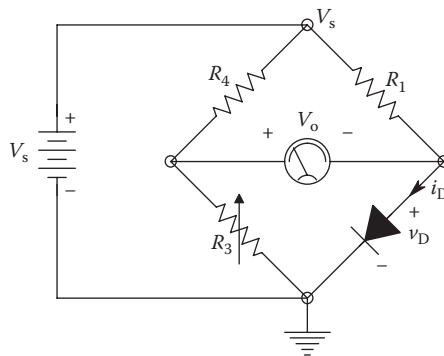


FIGURE 6.3

A forward-biased Si *pn* diode used as a nonlinear RTD in a Wheatstone bridge.

6.2.1.2 Resistive Strain Gauges

Resistive strain gauges fall into two, broad categories: bonded and unbonded gauges. Bonded gauges consist of fine wires or conducting films that are cemented to some structural beam or machine part in which we wish to measure the strain. The minute elongation or compression of the beam in response to a mechanical load causes a length change in the bonded gauge conductors. This length change, in turn, causes a small change in the resistance of the conductors that is usually sensed with a Wheatstone bridge circuit. Unfortunately, strain gauges also respond to temperature changes, a property that must be compensated for in design.

To examine how a strain gauge works, we consider a length of wire, L cm, with a circular cross section of area A cm², and a resistivity of ρ ohm cm. The resistance of this wire (at a given temperature) is given by

$$R = \frac{\rho}{L} \text{ ohms.} \quad (6.7)$$

If the wire is stressed mechanically by a load W Newtons (or pounds), a stress of $s = W/A$ Pascals (or psi) will occur. (A is in m² for the SI (or MKS) units and in.² for the English units.) As the result of this stress, a strain, ϵ , will occur, given by

$$\epsilon = \frac{\Delta L}{L} = \frac{s}{Y} \text{ in./in. or m/m,} \quad (6.8)$$

where Y is the Young's modulus for the material of the wire, generally given in psi in the United States.

Using a Taylor's series expansion on R , we can write

$$\Delta R = \left(\frac{-L}{A^2} \right) \Delta A + \left(\frac{L}{A} \right) \Delta \rho + \left(\frac{\rho}{A} \right) \Delta L \text{ ohms.} \quad (6.9)$$

To obtain the fractional change in R , we divide the equation given earlier with Equation 6.7 for R :

$$\frac{\Delta R}{R} = -\frac{\Delta A}{A} + \frac{\Delta \rho}{\rho} + \frac{\Delta L}{L}. \quad (6.10)$$

We now define the *gauge factor*, GF, as

$$\text{GF} = \frac{\Delta R/R}{\Delta L/L} = \frac{\Delta R/R}{\epsilon} = 1 + \frac{\Delta \rho/\rho}{\Delta L/L} + \frac{\Delta A/A}{\Delta L/L}. \quad (6.11)$$

The right-hand term in Equation 6.11 is *Poisson's ratio*, μ . From Equation 6.11, we can calculate the change in resistance, ΔR , for a given strain, ϵ . *For example*, if we load a 0.01" diameter steel wire with a 1 lb load, this causes a stress of $1/(\pi \times 0.0052) = 1.27 \times 10^4$ psi. The strain caused by this stress is $\epsilon = s/Y = 1.27 \times 10^4 / 3 \times 10^7 = 4.23 \times 10^{-4}$ in./in. or 423 microstrains.

Assume the wire has an unstrained resistance of 220 ohms, and the gauge factor, GF, is 3.2. ΔR is then given by

$$\Delta R = GF \epsilon R = +0.298 \text{ ohms.} \quad (6.12)$$

The output voltage of an initially balanced, equal-arm, 220 ohm Wheatstone bridge with a 6 V excitation, V_s , which includes the wire described earlier as one arm, would be

$$V_o = V_s \frac{\Delta R}{4R} = 6 \frac{0.298}{4(220)} = 2.03 \text{ mV.} \quad (6.13)$$

The outputs of strain gauge bridges generally need amplification. Sensitivity to temperature changes can be reduced by using two matched gauges on one side of the bridge; the lower one is active, the upper one is unstrained and acts as a thermal compensator for the active bottom gauge. Both gauges are at the same temperature. See Beckwith and Buck (1961) for an excellent description of bonded strain gauge applications. A good discussion of the sources of error in bonded strain gauge systems can be found in Lion (1959).

The frequency response of bonded strain gauges depends in large part on the mechanical properties of the structure to which they are bonded. It generally ranges from DC through the audio range.

Unbonded strain gauges are used in several applications: the direct measurement of small forces and the measurement of pressure (pressure acts on a diaphragm or piston to produce a force). Figure 6.4A shows the innards of an unbonded force-sensing strain gauge. There are two pairs of wires under tension. When force is applied in the direction shown, the left-hand pair is strained more, and the right-hand pair less. Due to this asymmetry, one pair of wires' resistance increases, while the other pair's decreases a like amount. The two pairs of wires are connected to form a Wheatstone bridge having inherent temperature compensation, as shown in Figure 6.4B. Typically, the maximum $\Delta R/R$ attainable at full-rated load is about 0.01. The bridge excitation voltage is practically limited, as in the case of bonded gauges, by power dissipation in the wires and the effects of heating on their resistances.

Unbonded strain gauges exhibit a mechanical resonance that limits their HF response. The taut resistance wires and the moving armature form a spring and mass system that typically resonates at some audio frequency. Some unbonded strain gauge transducers are designed to have a full bridge output, that is, two of their four wire elements are resistors that increase like amounts, while two other wire elements decrease in resistance like amounts due to applied force. Zero output balance of a four-arm bridge at zero force is usually done with a potentiometer connected as shown in Figure 6.4B.

6.2.1.3 Photoconductors

Photoconductors (PCs) are materials whose resistance decreases upon illumination with light. (They should not be confused with photodiodes or solar cells, which can produce an EMF or short-circuit current in response to the absorption of light quanta.) Uses of PCs include exposure meters for cameras, light sensors in spectrophotometers, light sensors in a variety of counting systems where an object interrupts a light beam illuminating the PC, systems that sense a decrease of overall ambient illumination and turn on outside lighting (of course, the lighting must not be sensed by the PC), oil and gas burner safety systems,

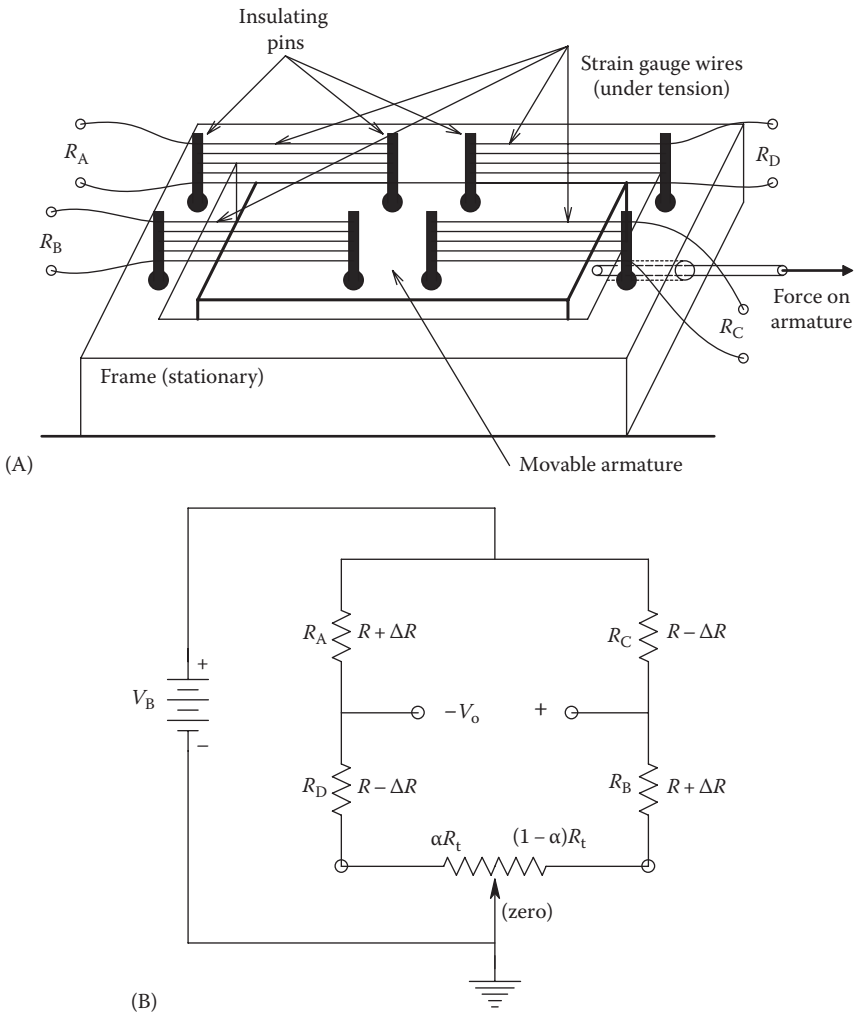


FIGURE 6.4 (A) A four-arm, unbonded strain gauge force sensor. With zero applied force on the armature, the four resistance windings are under equal tension. When a force is applied to the armature in the direction shown, strain increases in R_A and R_B and decreases in R_C and R_D . (B) Schematic diagram of the four-arm, unbonded strain gauge force sensor. R_t is used to obtain an output null at zero or reference force.

and burglar alarms. PCs are also called *light-dependent resistors* or *photoresistors*. Figure 6.5 illustrates the cross-sectional schematic of a typical PC device. Note that it is a thin wafer.

PCs can be made from a number of intrinsic and doped semiconductor materials. Each semiconductor has a distinct spectral response to light, ranging from UV to FIR. Absorption of light quanta with a certain energy range by semiconductor substances produces electron-hole pairs that enter the conduction band and drift in an applied electric field and contribute to the current passing through the semiconductor. Greatest PC sensitivity occurs when the absorbed photons have energies equal to or slightly greater than the semiconductor's *bandgap energy*.

For example, cadmium sulfide (CdS) with an energy gap of 2.42 eV works best in the visible range of wavelengths. Germanium and indium antimonide with energy

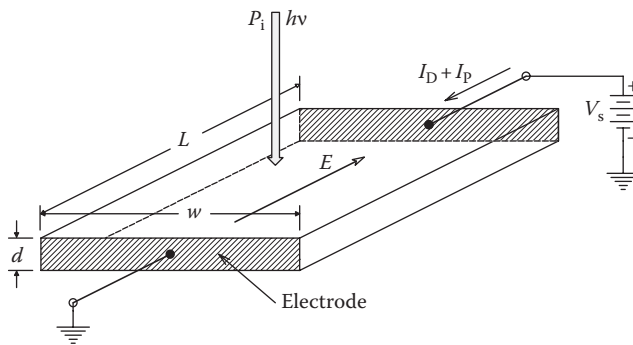


FIGURE 6.5
Diagram of a PC slab.

bandgaps of 0.67 and 0.18 eV, respectively, are used in the near and far infrared (IR) wavelengths, respectively. EG&G-Judson Infrared offered a J15D series of mercury cadmium telluride PCs, which operate in the 2–22 μm band (far IR). These PCs are operated at 77 K and are used for IR spectroscopy, thermography, CO₂ laser detection, and missile guidance. Their response time constants range from 0.1 to 5 μs , depending on the lifetime of electrons in the crystal, which in turn depend on material composition and temperature. HgCdTe photoconductive detectors generate $1/f$ noise, which limits their resolution below 1 kHz.

CdS PCs are often made by depositing a thin layer of the material on an insulating (ceramic) substrate, often in a zigzag pattern to increase the effective length of the conductor's path, hence its resistance.

The frequency response of PCs (in response to modulated light intensity) depends on, among other factors, the intensity and wavelength of the incident light, the Kelvin temperature of the PC, and the amount of metallic impurities in the semiconductor that can trap charge carriers and delay their spontaneous recombination. It can be shown that when the recombination time constant, τ_p , is large (corresponding to a low HF response), the gain of the PC is high. A short time constant, giving good HF response, results in lower *gain* (Yang 1988). Gain is defined by Yang as the ratio of carriers collected by the ohmic contacts to the carriers generated by photons per unit time. The range of frequency response of PCs ranges from over 1 MHz to below the audio range. Germanium and lead sulfide have relatively broadband responses (10's of kHz), while materials such as selenium and cadmium sulfide have bandwidths in the hundreds of Hz.

Table 6.2 lists some of the materials used in PCs, their bandgap energies in eV, and the wavelength of their peak spectral response.

In general, the total current in a PC can be written as follows:

$$I_{\text{PC}} = V_s [G_D + G_P] \quad (6.14)$$

where

V_s is the bias voltage

G_D is the equivalent dark conductance

G_P is the photoconductance

TABLE 6.2
Properties of Some Photoconductors

PC Material (Temp)	Bandgap Energy, eV	Wavelength of Peak Response, λ_c , μm	Rise Time/Fall Time
ZnS	3.60	0.345	
CdS	2.40	0.52	30 ms/10 ms
CdSe	1.80	0.69	15 ms/15 ms
CdTe	1.50	0.83	
Si (intrinsic)	1.12	1.10	1 μs /1 μs
Ge (intrinsic)	0.67	1.85	0.1 μs /0.1 μs
PbS	0.37	3.35	
InAs	0.35	3.54	≈ 1 ns
Te	0.33	7.75	
PbTe	0.30	4.13	
PbSe	0.27	4.58	2 μs
HgCdTe (77 K)		5.0	* /5 μs
InSb (77 K)	0.18	6.90	
GeCu (4 K)		25	
GeBe (3 K)		55	

* not known.

G_p can be shown to be given by (Yang 1988)

$$G_p = \frac{I_p}{V_s} = \frac{q\eta\tau_p(\mu_p + \mu_n)}{L^2} \left[\frac{P_i\lambda}{hc} \right] \text{ S} \tag{6.15}$$

where

- τ_p is the mean lifetime of holes
- $\mu_p = |v_p/E|$ is the hole mobility in $\text{cm}^2/\text{V}\cdot\text{s}$
- $\mu_n = |v_n/E|$ is the electron mobility in $\text{cm}^2/\text{V}\cdot\text{s}$
- v_p and v_n are the mean drift velocities of holes and electrons, respectively
- E is the uniform E-field in the semiconductor
- $q, \eta, P_i, \lambda, h,$ and c have been defined earlier

Note that in general, $\mu_n > \mu_p$. L is the distance between the electrodes.

The expression for G_p is an approximation, valid up to the cutoff wavelength, λ_c . $[P_i\lambda/hc] = \Phi_i$, the incoming number of photons/s on area wL m^2 . The dark conductance of a Si PC, G_D , can be found simply from the room temperature resistivity of Si, ρ , and the geometry of the PC. For example,

$$G_D = \frac{A}{\rho L} = \frac{wd}{\rho L} = \frac{0.2 \text{ cm} \times 0.001 \text{ cm}}{(2.3 \times 10^3 \Omega \text{ cm}) \times 0.02 \text{ cm}} = 4.348 \times 10^{-6} \text{ S}. \tag{6.16}$$

The photoconductance for a Si PC illuminated by 1 μW of 512 nm photons is, from Equation 6.15,

$$G_p = \frac{1.6 \times 10^{-19} \times 0.8 \times 10^{-4} \times (1350 + 450) \frac{\text{cm}^2/\text{V}\cdot\text{s}}{\text{cm}^2}}{10^{-4} \text{ cm}^2} \left[\frac{10^{-6} \times 512 \times 10^{-9} \frac{\text{W}}{\text{J}\cdot\text{s}} \times \text{m}}{6.625 \times 10^{-34} \times 3 \times 10^8 \frac{\text{m}}{\text{s}}} \right] = 5.935 \times 10^{-4} \text{ S}. \tag{6.17}$$

A great advantage of PCs is their unique ability to respond to mid-infrared (MIR) and FIR photons at wavelengths not sensed by PIN PDs or APDs. We note that for various materials, as the bandgap energy *decreases*, λ_c *increases*, and the PC's response time constants *decrease*.

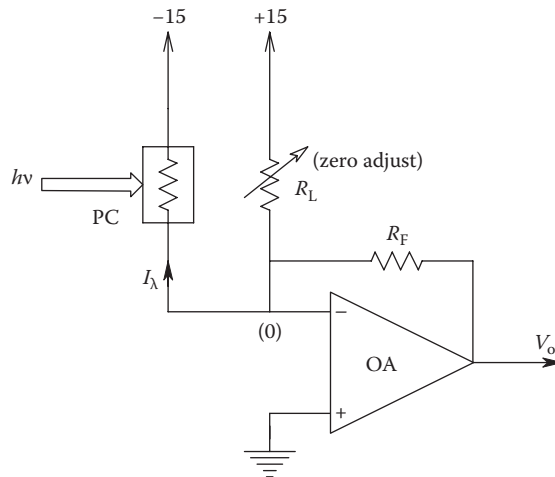


FIGURE 6.6

An OA is used as an operational transresistor to condition the photocurrent from a PC. The OA should have low-DC drift, such as found in a chopper design.

Certain PCs also are unique in being able to convert incident x-ray photons to a conductance change. X-ray sensing PCs are made of *amorphous selenium* (a-Se) (Soltani et al. 1999). An a-Se PC with $E = 10 \text{ V}/\mu\text{m}$ produces ca. 10^3 electron-hole pairs per 50 keV x-ray photon. There is a 50% attenuation of a 50 keV electron beam in $d = 365 \mu\text{m}$ a-Se. Soltani et al. describe a charge-coupled, x-ray photon sensing *array* using a-Se sensors that have superior image resolution to phosphor x-ray sensors.

Figure 6.6 illustrates a simple OA circuit that gives $V_o \propto P_i$. Note that R_C is used to cancel the PC's dark current. A Wheatstone bridge can also be used to convert photoconductance to output voltage, albeit nonlinearly.

Unlike PDs, PCs make thermal noise. It can be shown that the total MS noise current input to the OA's summing junction is

$$\overline{i_{\text{ntot}}^2} = \{4kT[G_D + G_P + G_C + G_F] + i_{\text{na}}^2\} B \text{ MSA}. \quad (6.18)$$

Note that the MS noise increases with input light power because G_P increases with P_i . In addition to the noise current, the OA also has an equivalent, short-circuit input noise voltage, $e_{\text{na}}^2 B \text{ MSV}$.

6.2.1.4 Conductive Relative Humidity Sensors

Humidity sensors respond to the amount of water vapor in the air or other gas. In this section, we will consider only those humidity sensors that change their conductance in response to the partial pressure of water vapor to which they are exposed. In one embodiment, the Dunmore sensor, two, noble metal wires are wound in the form of two, nontouching helices on an insulating bobbin or form. A thin, hygroscopic coating consisting of an ionizable salt covers the wires. As the humidity increases, the salt in the coating reversibly absorbs water vapor and dissociates into ions. The ions conduct electric current, which is an increasing function of the humidity. This sensor's conductance increases with humidity. Because this sensor is an electrochemical system, the use of DC excitation for the bridge and sensor will lead to polarization and possible electrolysis and chemical breakdown of

the hydrated salt. Polarization due to ion migration causes the apparent resistance of the sensor to increase, hence causes erroneous readings. The ionizable salt humidity sensors must be used with low-voltage, LF, AC excitation to prevent these problems.

Lithium chloride, potassium dihydrogen phosphate, and aluminum oxide have been used in various conductive humidity sensor designs (Lion 1959). The sensors using only lithium chloride appear to respond to a limited range of humidity, which is dependent on the LiCl concentration. The sensor's resistance is strongly dependent on temperature, which affects water absorption and dissociation of the salt. At constant temperature, the LiCl sensor's resistance decreases logarithmically over a three- or four-decade range in proportion to the relative humidity.

Figure 6.7A illustrates a wide-range relative humidity sensor described in Lion (1959), attributed to Cutting, Jason, and Wood (1955). An aluminum rod was coated at one end

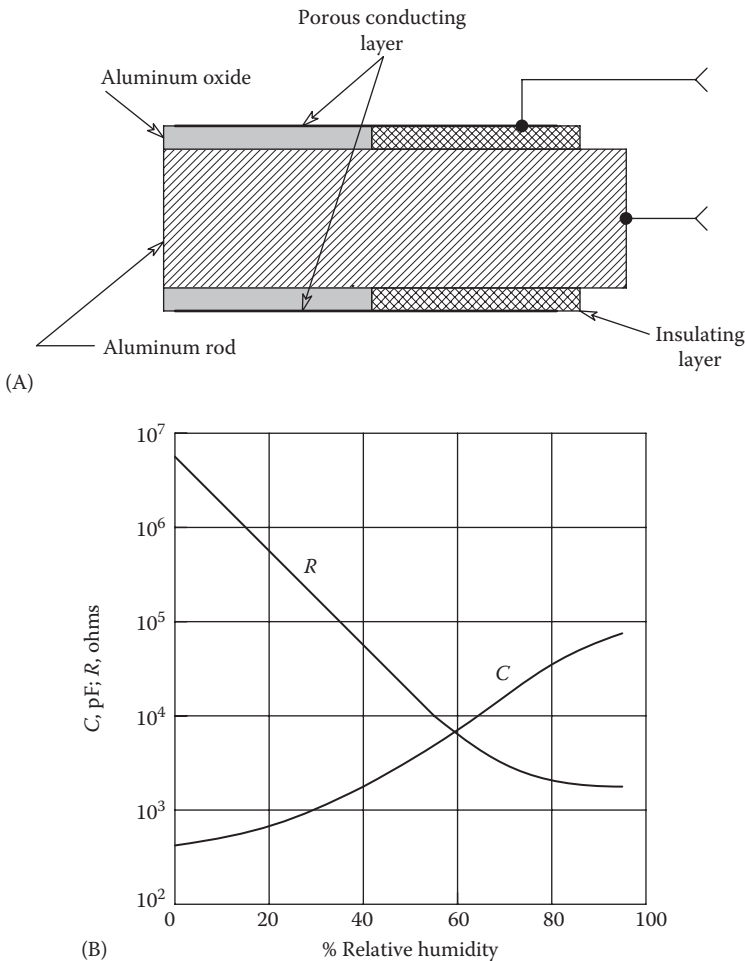


FIGURE 6.7

(A) Cross section through an aluminum oxide humidity sensor. (From Lion, K.S., *Instrumentation in Scientific Research: Electrical Input Transducers*, McGraw-Hill Book Co., Inc., New York, 1959. With permission.) (B) Plot of an aluminum oxide humidity sensor's equivalent parallel resistance and capacitance as a function of relative humidity (RH). (From Lion, K.S., *Instrumentation in Scientific Research: Electrical Input Transducers*, McGraw-Hill Book Co., Inc., New York, 1959, Figure (1-7)8. With permission.)

with an aluminum oxide layer and then a porous conducting layer such as graphite or vapor deposited metal. This sensor is used with AC excitation. Figure 6.7B, adapted from Lion, shows the sensor's resistance varies from about 5 M Ω at 0 RH to about 2 k Ω at 100% RH. Of special interest is the fact that there is a corresponding increase in sensor capacitance from about 500 pF at 0% RH to 10⁵ pF at 100% RH. Hence, an AC, capacitive-sensing bridge can also be used to obtain sensor output, as well as an AC Wheatstone bridge. This humidity sensor is unique in that it is claimed to be temperature insensitive over a -15°C to 80°C range.

Commercial humidity measurement systems and sensors are offered by several companies. Omega Corp. (2013) markets several different humidity sensing transmitters that use a thin-film polymer capacitor sensor (the HX92 family). These sensors run on 6–24 VDC, and the output is either current (4–20 MA for 0%–100% RH; model HX92A) or voltage (0–1 V for 0%–100% RH; model HX92V).

Ohmic Instruments Co. offers a series of conductive humidity sensors based on water vapor absorption by dissociable salts. They offer 10 different Dunmore-type sensors, each with a narrow range of RH sensitivity. Sensor resistances vary from over 10 M Ω to under 1 k Ω over their narrow ranges of RH sensitivity. A broad-range instrument (model WHS 5-99L) covering from 5% to 90% RH uses four, overlapping, narrow-range sensors in a weighted resistor network, and a 60 Hz excitation is used.

Thunder Scientific Corp. has developed and marketed what they call the *Brady array* humidity sensor. From published technical specifications, this sensor is a precise array of semiconductor crystals and interstitial spaces. Physical adsorption of water molecules from water vapor somehow causes an increase in conduction band electrons in the crystals, raising conductivity. Unlike the Dunmore-type sensors that require water vapor adsorption and ionic dissociation, the Brady sensor operates on a purely physical basis and thus has a very rapid response time to step changes in RH (in hundreds of milliseconds) vs. minutes for Dunmore sensors. The simplest circuit for a Brady sensor consists of a 5 VRMS, 1000 Hz voltage source in series with the sensor and a 1 M Ω resistor. The voltage across the 1 M Ω resistor is proportional to RH. Brady sensors work over a range from nearly 0% to 100% RH. They have been shown to work normally in atmospheres of water vapor in air, oxygen, nitrogen, CO₂, hydrogen, natural gas, butane, and vapors from various hydrocarbons such as acetone, trichloroethylene, jet fuel, gasoline, kerosene, and oils. Their working temperature range has been given as -20°C to $+70^{\circ}\text{C}$. Typical accuracy is stated as $\pm 2\%$ of indicated RH, and resolution is 0.1% RH.

Temperature compensation of a Brady sensor is easily accomplished by taking a like sensor and hermetically sealing it so that it is at 0% RH and then using the sealed sensor as an element in the reference side of a Wheatstone bridge (see Figure 6.8). A temperature-compensated Brady array is claimed to drift no more than 1 mV/ $^{\circ}\text{F}$ over the -20°C to $+70^{\circ}\text{C}$ operating range. This is 0.02% of full-scale per degree Fahrenheit.

Calibration of resistive and other kinds of RH sensors is traditionally done by using saturated salt solutions at known temperatures. For example, at 25°C , the RH over a saturated aqueous LiCl(H₂O) solution is 11.3%; over MgCl₂(6H₂O), it is 32.8%; over K₂CO₃, it is 43.2%; over NaBr, it is 57.6%; over NaCl, it is 75.3%; over KCl, it is 84.3%; and over K₂SO₄, it is 97.3%. Generally, as the temperature of the saturated salt solution and the atmosphere over it increases, there is a slight reduction of RH, with the exception of sat. LiCl, which maintains 11.3% from 5°C to 25°C (Groselj 2010).

6.2.1.5 Direct Resistance Change Used to Sense Position or Angle

Direct measurement of shaft angle in applications such as electromechanical, position feedback control systems is often done by coupling the shaft either directly, or through

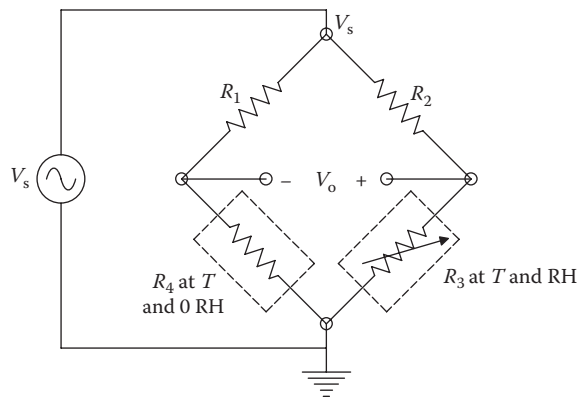


FIGURE 6.8

Two Brady humidity sensors are used in a Wheatstone bridge to obtain temperature compensation over a wide range; sensor R_4 is sealed at 0% RH and responds only to temperature changes.

gears, to a single- or multiturn potentiometer. The potentiometer can be one arm of a Wheatstone bridge or used as a voltage divider. Often, a feedback linearization circuit is used with a bridge, such as shown in Figure 6.9, to produce a voltage output that varies linearly with potentiometer resistance (shaft angle). The potentiometers used in such shaft angle transduction must be linear and have high resolution. The resolution of a standard wire-wound, 3, 10- or 14-turn potentiometer is limited by the wiper connected to the shaft making contact with individual turns of the helically wound resistance wire. One way to avoid this problem is to use potentiometers that have a continuous, cermet or conductive plastic resistance materials.

Potentiometers used as angular position sensors typically are specified to have independent linearity over their total range of rotation. The Bourns 3400 series of 10-turn, wire-wound potentiometers typically have $\pm 0.15\%$ independent linearity and are

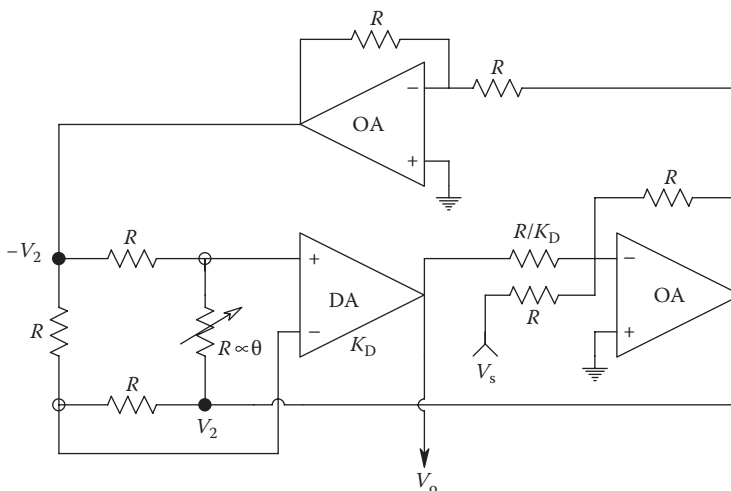


FIGURE 6.9

A feedback circuit giving Wheatstone bridge output linearization where the variable resistance varies linearly over a wide range as a function of the QUM, such as a shaft angle.

available with $\pm 0.05\%$ independent linearity. The Bourns 3400 series has a resistance tempco of $+0.007\%/^{\circ}\text{C}$. Conductive plastic potentiometers, such as the Beckman Instruments single-turn, 6550 series, have independent linearities of $\pm 0.5\%$ and tempcos of -150 to $+300$ ppm. Cermet potentiometers, such as the single-turn, Beckman Instruments 6300 series, have also have independent linearities of $\pm 0.5\%$ but have lower tempco ranges of ± 100 ppm.

In addition to precision rotational potentiometers for angle sensing, linear displacements can be measured with linear potentiometers, that is, pots whose resistance varies linearly with the linear position of the wiper. The Beckman Instruments series 400 rectilinear potentiometer series uses an infinite resolution, cermet resistance element. Wiper travel lengths range from $0.6''$ to $6.1''$ in various 400 series models. Linearity ranges from $\pm 1.35\%$ for the $0.6''$ model to $\pm 0.25\%$ in the $6.1''$ model.

One disadvantage of potentiometers as position sensors is that they require some extra torque or force to move their wipers. Another disadvantage is that because of the wiping action, they eventually wear out. Optical coding disks need no wipers and are limited in resolution by the number of output bits available. An LVDT can operate with negligible friction if a Teflon lining is used in its core. Still, potentiometers offer an inexpensive means of sensing angular or linear position in a number of applications.

6.2.1.6 Sensors Based on the Giant Magnetoresistive Effect

The giant magnetoresistive (GMR) effect is a physical phenomenon in which a magnetic field changes the resistance of a layered, thin-film sensor. GMR is a relatively new effect discovered in the late 1980s by Peter Gruenberg of the KFA Research Institute in Julich, Germany, and independently by Albert Fert at the University of Paris-Sud. Researchers at IBM extended the initial work on GMR devices; they developed sensitive GMR transducers based on sputtered thin-film technology. IBM's purpose for developing sensitive, miniature GMR devices has been to improve the data density stored on computer floppy and hard drive magnetic disks. IBM also calls their GMR sensor the *spin valve read head*.

GMR sensors have a resistance that is modulated by an imposed magnetic flux density; thus, a Wheatstone bridge can be used to sense the $\Delta R/R_M$, hence the ΔB . Unfortunately, GMRs are not particularly linear, which makes them acceptable for binary applications but presents problems for linear analog sensing unless feedback is used (Smith and Schneider 1999).

A basic GMR sensor is a multiple-layer, thin-film device. In its simplest form, it has a nonmagnetic, conductive spacing layer sandwiched between two ferromagnetic film layers. Usually, the magnetization in one ferromagnetic layer is fixed or *pinned*, along a set direction lying in the plane of the layer. In the absence of an external \mathbf{B} -field, the resistance of the conductive layer is maximum, R_m . The magnetization of the *free magnetic layer* is allowed to rotate in response to an externally applied magnetic field vector lying in the plane of the free layer. Lowest GMR sensor resistance occurs when the pinned and free layers are magnetically oriented in the same direction. There can be a 10%–15% decrease in R_m due to external applied \mathbf{B} , before the GMR effect saturates at high $|\mathbf{B}|$. If the \mathbf{B} vector reverses sign, we observe basically the same negative ΔR , except for the small hysteresis present. This is because electrons in the conductor layer with parallel spin directions move freely in both films. Higher resistance occurs when the magnetic orientations in the two films oppose each other, since movement of electrons in the conductor of either spin direction is hampered by one or the other magnetic films. Depending on the strength and orientation of the applied external \mathbf{B} field, the pinning direction, the materials used, and

their thicknesses, the $R_M = f(B_x, \theta, \varphi)$ curve can have a variety of shapes (Wilson 1996). (Note that ideally, for maximum sensitivity, $\mathbf{B} = \mathbf{B}_x$, i.e., \mathbf{B} is parallel with the x -axis; the GMR sensor lies in the x, z plane. Using polar coordinates, φ is the angle \mathbf{B} makes with the y -axis, which is \perp to the x, z plane, and θ is the angle the projection of \mathbf{B} on the x, z plane makes with the x -axis. So for maximum sensitivity, $\theta = 0^\circ$, and $\varphi = 90^\circ$.)

In many cases, the GMR $R_M = f(B_x)$ curve exhibits both hysteresis and some dead zone as shown in Figure 6.10. This GMR sensor has the *crossed easy axis* configuration in which the applied \mathbf{B} is perpendicular to the free layer easy axis and the pinning layer magnetization direction is parallel to \mathbf{B} . The response is linear around a Q-point, Q , and is relatively hysteresis free (Wilson 1996). Because of its magnetic sensitivity, the GMR must be magnetically shielded from the Earth’s field and stray magnetic fields from man-made power wiring and machinery. To use such a sensor as a linear device, it must be magnetically biased around the Q-point.

Much research is currently underway to improve the linearity and sensitivity of spin valves or GMR sensors. Attention is being paid to the use of multiple-layer thin films of various compositions, including two pinned layers; 6–12 layers are being studied by Fujitsu (Kanai et al. 2001).

GMR sensors are available commercially. Infineon Technologies AG offers a four-element GMR IC package (TLE5012) in which the four GMRs are connected in a bridge to measure the orientation of \mathbf{B} in X and Y . Their TLE5012 sensor is intended as a vector position sensor in that the bridge output is sinusoidal function of the input magnet’s \mathbf{B} -field angle. The magnetic-field-free value of each resistor is ca. 700Ω . Rhopoint Components of Oxted, Surrey, United Kingdom, offers a 12-sensor GMR device kit (AG910) for experimenters and R&D use.

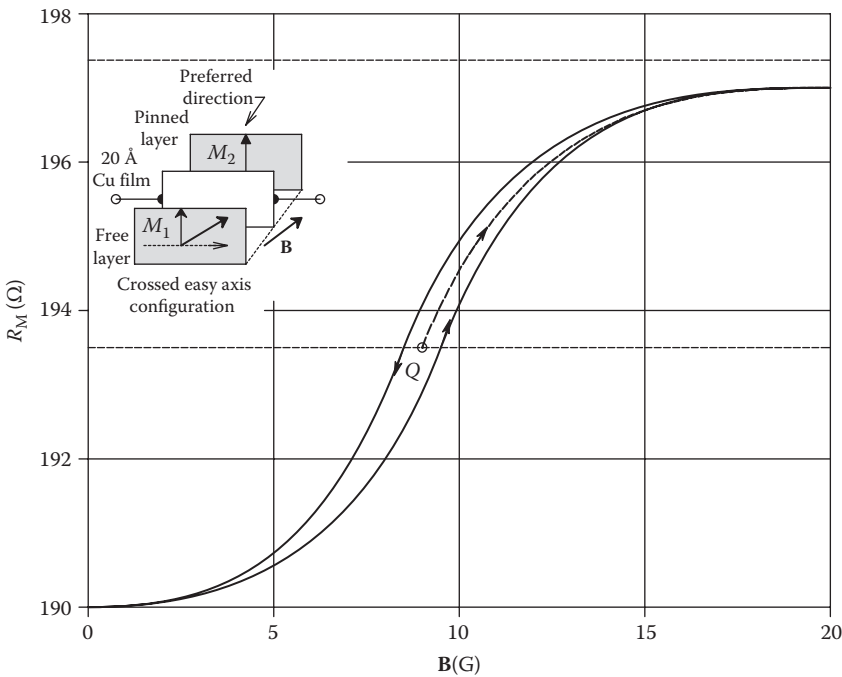


FIGURE 6.10 Resistance vs. applied magnetic field intensity in Gauss for a typical GMR magnetosensor.

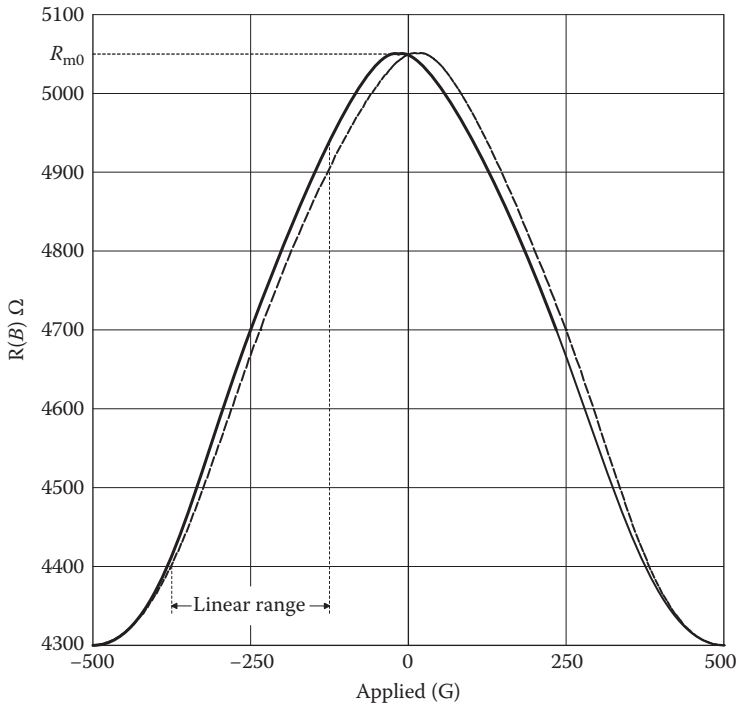


FIGURE 6.11 Magnetoresistance vs. Gauss characteristics of a commercial GMR sensor. Note that the curve is even in \mathbf{B} and has a slight hysteresis.

Nonvolatile Electronics Inc., Eden Prairie, MN (NVE), also markets GMR sensor assemblies based on antiferromagnetic coupling. The NVE GMR sensors have fairly linear $R(B)$ characteristics showing maximum resistance at 0 G (see Figure 6.11). The NVE sensor was determined graphically to have a maximum $\Delta R/\Delta B = 12.9\%$ and fractional sensitivity, S_B , of

$$S_B = \frac{\Delta R/R_{m0}}{\Delta B} = 3.575 \times 10^{-4} \text{ G}^{-1}. \tag{6.19}$$

NVE claims that if one of their GMR sensors uses a high-permeability external flux concentrator and 5 V excitation, 100 G will give a 350 mV voltage change. They compare that to a 5 mV output from a Hall sensor given the same input.

NVE also makes gradiometer GMR assemblies in which two matched GMR resistors are placed at each end of a rectangular bar, spaced at a distance of 0.5 or 1.0 mm. The four GMR sensors are connected in a Wheatstone bridge circuit. The gradiometer gives no output if the magnetic field is equal at each sensor. If a magnetic field density gradient exists over the separation length of the sensors, two at one end (S_1 and S_3) will be in a stronger field than the pair (S_2 and S_4) at the other end. Hence, the bridge is unbalanced and produces an output. Figure 6.12 illustrates the gradiometer chip and bridge circuit. If there is no gradient in \mathbf{B}_x , that is, $dB_x/dx = 0$, then assume all four of the GMR sensors are subject to equal \mathbf{B}_x , and all have resistance $R_0 - \delta R$, and the bridge is balanced. Now let \mathbf{B}_x at the right end of the chip be larger by some amount, ΔB . This causes the resistances of sensors S_1 and S_3

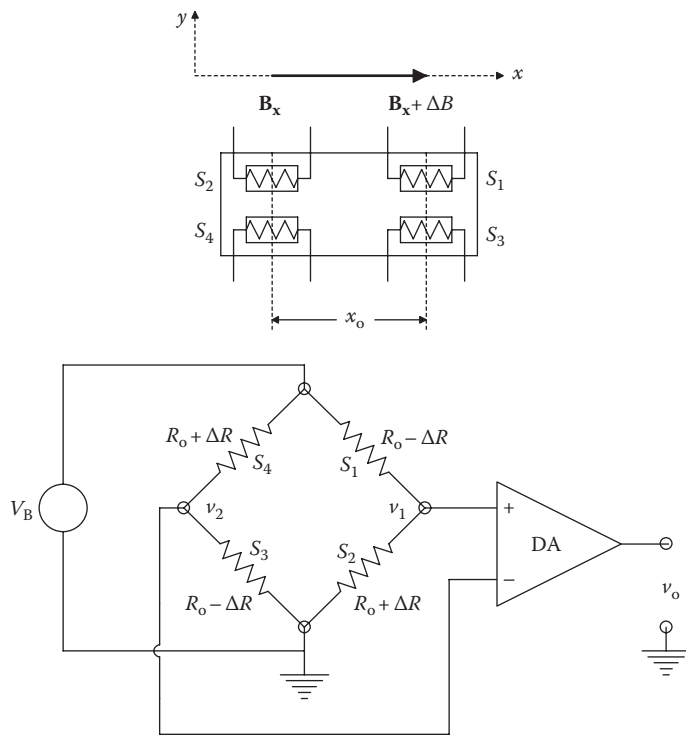


FIGURE 6.12
Circuit and diagram of a GMR gradiometer Wheatstone bridge.

to be smaller than those of S_2 and S_4 . In summary, with $\Delta B > 0$, $R_{S1} = R_0 - \delta R_+ = R_{S3}$, and $R_{S2} = R_0 - \delta R = R_{S4}$, where clearly, $\delta R_+ > \delta R$. Using the voltage divider relation, we can find the unbalanced bridge output:

$$v_o = v_1 - v_2 = V_B \left[\frac{R_0 - \delta R}{2R_0 - (\delta R + \delta R_+)} - \frac{R_0 - \delta R_+}{2R_0 - (\delta R + \delta R_+)} \right] \neq V_B \left[\frac{(\delta R_+ - \delta R)}{2R_0 - (\delta R + \delta R_+)} \right] \neq 0. \quad (6.20)$$

If $(\delta R + \delta R_+) \ll 2R_0$, then

$$v_o \cong \frac{V_B}{2R_0} (\delta R_+ - \delta R) = \frac{V_B}{2R_0} (\Delta \delta R) > 0. \quad (6.21)$$

Note that if the gradient in B_x is reversed, $v_o < 0$.

6.2.1.7 Anisotropic Magnetoresistance

The basic magnetoresistive effect was first described by Lord Kelvin (William Thompson) in 1856 in bulk ferromagnetic materials. Like many great discoveries in science, it took over 100 years for the development of thin-film, lithographic, semiconductor technology to realize practical AMR devices. The basic AMR device consists of strips of a thin film of permalloy (nickel-iron alloy) deposited on a silicon wafer and arranged in a *barber pole* geometry (Caruso et al. 1998). Figure 6.13 illustrates four AMR sensors arranged in a Wheatstone bridge.

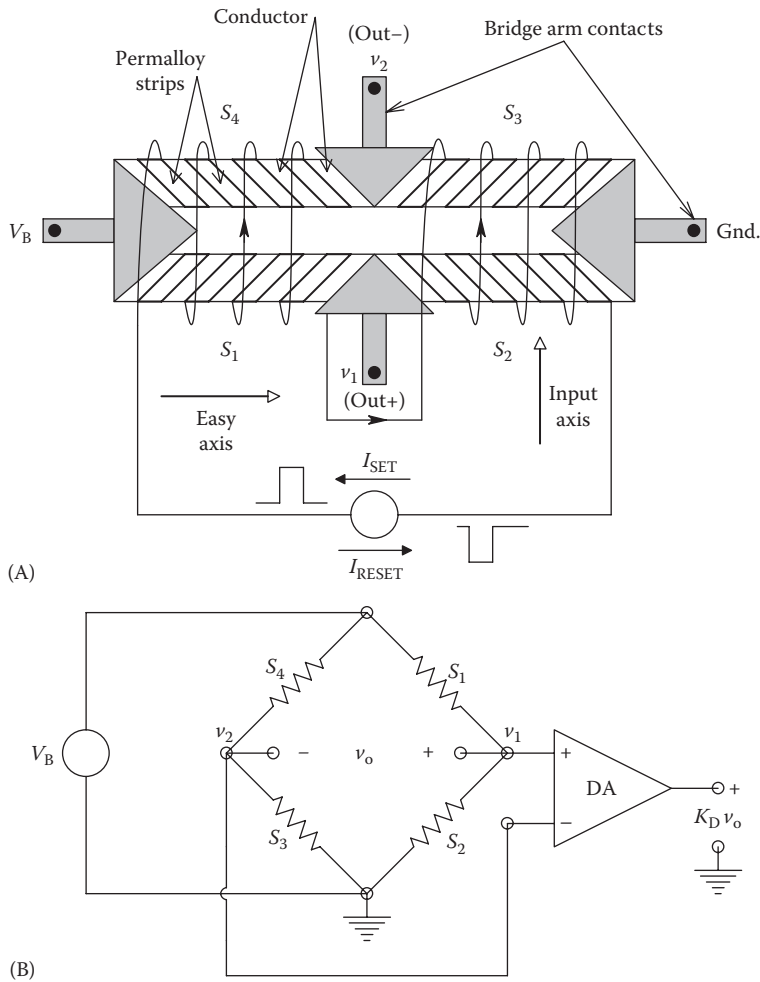


FIGURE 6.13 (A) Four anisotropic magnetoresistor devices arranged as a linearized magnetosensor. (B) Wheatstone bridge circuit for the four-element magnetosensor.

The phenomenon of anisotropic magnetoresistance can be observed in any ferromagnetic material. It is best described in terms of quantum mechanics; it is due to spin-orbit interactions in conduction band electrons in the metal. The *anisotropy* in AMR refers to the fact that the sensor's resistance is a function of the angle between the current density vector, \mathbf{J}_x in the sensor, and the internal magnetic flux density, \mathbf{B}_i . \mathbf{B}_i is the vector sum of the internal \mathbf{B}_e (the QUM) and the internal, fixed \mathbf{B}_m . That is, $\mathbf{B}_i = \mathbf{B}_e + \mathbf{B}_m$. Figure 6.14 illustrates the \mathbf{B} vector summation at a point P in the AMR film. Note that the larger \mathbf{B}_e is in the film, the lower the angle between the current density vector, \mathbf{J} , and the net internal flux density vector, \mathbf{B}_i . The AMR sensor's resistance *increases* nonlinearly with *decreasing* angle, θ_i ; it is maximum when $\theta_i = 0$. The graph in Figure 6.14 illustrates how the vector addition of \mathbf{B}_e and \mathbf{B}_m affect AMR sensor resistance when the initial angle between \mathbf{B}_m and \mathbf{J} is 45° .

The sensor's $R_{AMR} = f(|\mathbf{B}_i|)$ can be linearized by tinkering with the sensor geometry and biasing (Ripka et al. 1998). AMR device composition is generally proprietary, but thin films of permalloy are often used. AMR sensors made by Honeywell Sensor Products,

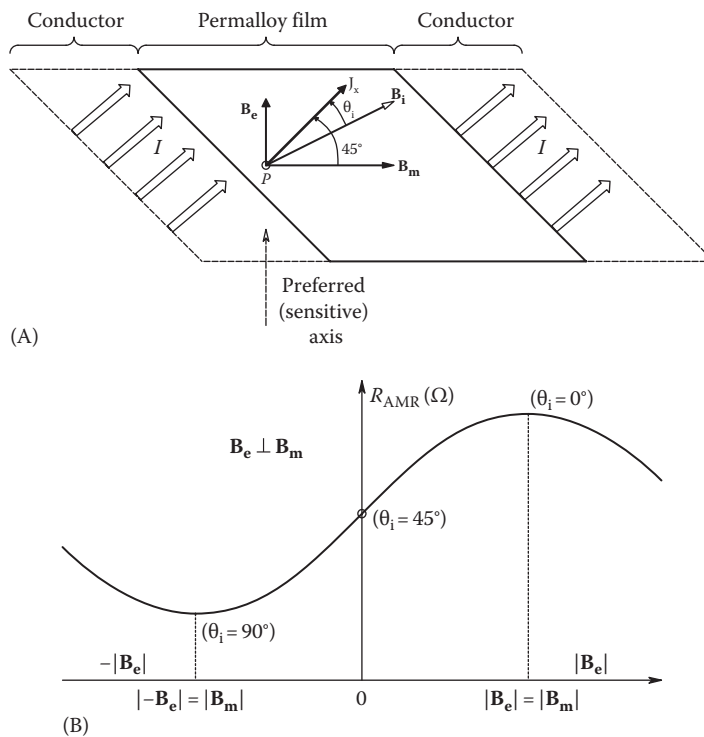


FIGURE 6.14 (A) \mathbf{B} vector summation at a point P in an AMR film. (B) R_{AMR} vs. applied $|B_e|$ for the AMR sensor.

NVE, and Philips have integrated feedback coils and reset (flipping) coils used to saturate the sensor material's B_m in a direction orthogonal to the input axis (preferred direction) to avoid distortion of the sensor characteristic caused by perpendicular fields (Ripka et al. 1998). Bridge output sensitivity is usually given in $V/V/\text{applied magnetic field}$. The first V is the bridge output voltage, the second V is the bridge excitation voltage, V_{br} , which is generally around 3–5 V. *Magnetic field* is given by some writers in *Oersteds* (cgs); an Oe is a unit of *magnetizing force*. SI magnetizing force is \mathbf{H} in amps/m. Still others specify magnetic field, \mathbf{B} , in Gauss (cgs). Note that 10^4 Gauss equal 1 Tesla = 1 Weber/meter² (SI units). Gauss and Teslas are units of *magnetic flux density*. In air, a uniform, 1 Oe magnetizing force creates a 1 G flux density, that is, $\mathbf{B} = \mu_0 \mathbf{H}$.

Useful input flux density range for AMR bridge sensors is from ca. 5 pT to 1 mT. By contrast, GMR sensors' useful ranges lies between 10 μT and 10^4 T. A SQUID sensor (see Chapter 8) works between 10 fT and 10 T (W/m^2) (Caruso et al. 1998). The horizontal component of the Earth's magnetic field in Cambridge, MA, is ca. $1.7 \times 10^{-5} \text{ W/m}^2$ or 0.17 G; the vertical component is ca. 0.55 G (Sears 1953). These values are well within the sensitivity range of AMR sensors.

Consider the example of designing a simple AMR sensor compass, designed for use on a level platform tangential to the Earth's surface. A handheld AMR compass with a bubble level used by hikers and a gimbaled AMR compass in a motor vehicle are representative applications. An AMR compass to be used on a platform that can experience roll, pitch, and yaw, such as a boat or aircraft, requires *three* AMR sensors with preferred directions along the X-, Y-, and Z-axes, as well as roll and pitch inputs from rate integrating gyros or clinometers (Caruso and Withanwasam 1999). The stable, horizontal platform compass

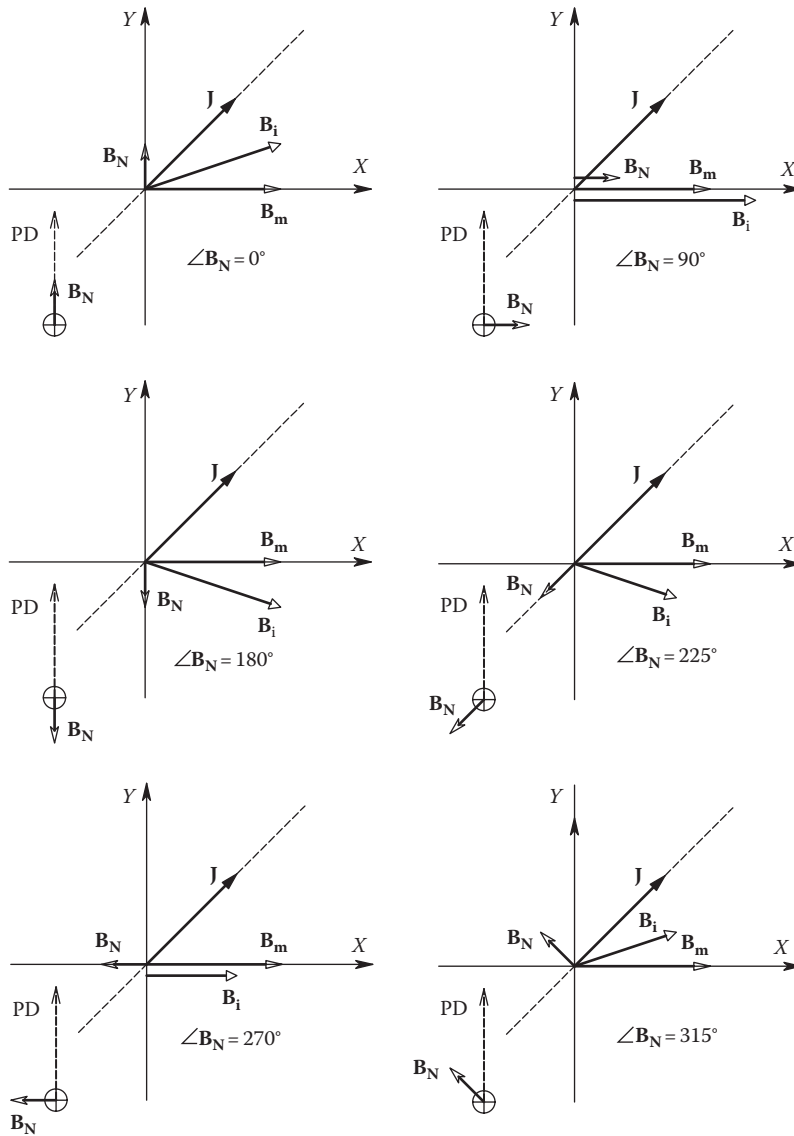


FIGURE 6.15

Six vector diagrams illustrating the conditions in an AMR sensor used to sense the Earth’s magnetic field, B_N . PD is the sensor’s preferred direction (direction of maximum response). J is the current density vector in the AMR film, B_m is the internal magnetizing flux density, and B_i is the net internal flux density vector.

uses two AMR sensors with orthogonal preferred axes along the X- and Y-axes defining the plane of the platform. Figure 6.15 illustrates the vector magnetic field conditions inside one sensor strip as a function of the direction of the external magnetic field vector, B_N , with respect to the sensor’s preferred input direction (PD). Many such sensor strips are incorporated in bridge sensor, S_1 . The second sensor bridge (S_2) is close to S_1 and has its PD rotated 90° clockwise from PD1. Because of the vector summation of the internal magnetizing flux density, B_m , and the horizontal component of the Earth’s magnetic field, B_N , the output of the S_1 bridge varies cosinusoidally with the compass direction; θ . θ in this

example is measured as the angle between PD1 and the north-pointing vector, \mathbf{B}_N . Thus, for example, if PD1 is pointing toward magnetic east, $\theta = +90^\circ$. We assume the S_1 sensor bridge output voltage, V_{o1} , follows the magnetically induced resistance change and can be modeled by the relation

$$V_{o1} = \left(V_{\min} + \frac{\Delta V}{2} \right) + \left(\frac{\Delta V}{2} \right) \cos(\theta). \quad (6.22)$$

It is be easily shown that the voltage output of the orthogonal S_2 bridge follows the approximate relation

$$V_{o2} = \left(V_{\min} + \frac{\Delta V}{2} \right) - \left(\frac{\Delta V}{2} \right) \sin(\theta). \quad (6.23)$$

Equations 6.22 and 6.23 can be solved simultaneously to find an expression for θ :

$$\theta = \tan^{-1} \left[\frac{(V_{\min} + \Delta V/2) - V_{o2}}{V_{o1} - (V_{\min} + \Delta V/2)} \right]. \quad (6.24)$$

A microcomputer can calculate θ from the variable inputs V_{o1} and V_{o2} and the constants V_{\min} and ΔV and display it digitally in degrees of heading, magnetic. Note that because $|\mathbf{B}_N|$ remains constant, and only its angle changes, V_{\min} and ΔV remain constant.

Many other interesting applications of AMR sensor bridges exist: for example, counting moving vehicles based on their perturbation of the Earth's magnetic field or the field of man-made magnets is possible using two- or three-axis, AMR sensor arrays. Vehicle counting can also be used to insure efficient operation of parking lots and garages. Vehicle direction of travel can also be determined. Another application of the AMR sensor bridge is galvanically isolated current measurements, from DC up to 10–15 MHz. The tangential component of the solenoidal B -field surrounding a current-carrying wire at some point P at a radius R from the center of the wire is given by

$$B_T = \frac{\mu_o I}{(2\pi R)} \text{ W/m}^2. \quad (6.25)$$

Here, we assume that the wire is of infinite length, and its radius $r \ll R$ (Sears 1953). If lower currents are to be measured, the field at the center axis at the end of an N -turn solenoid coil is

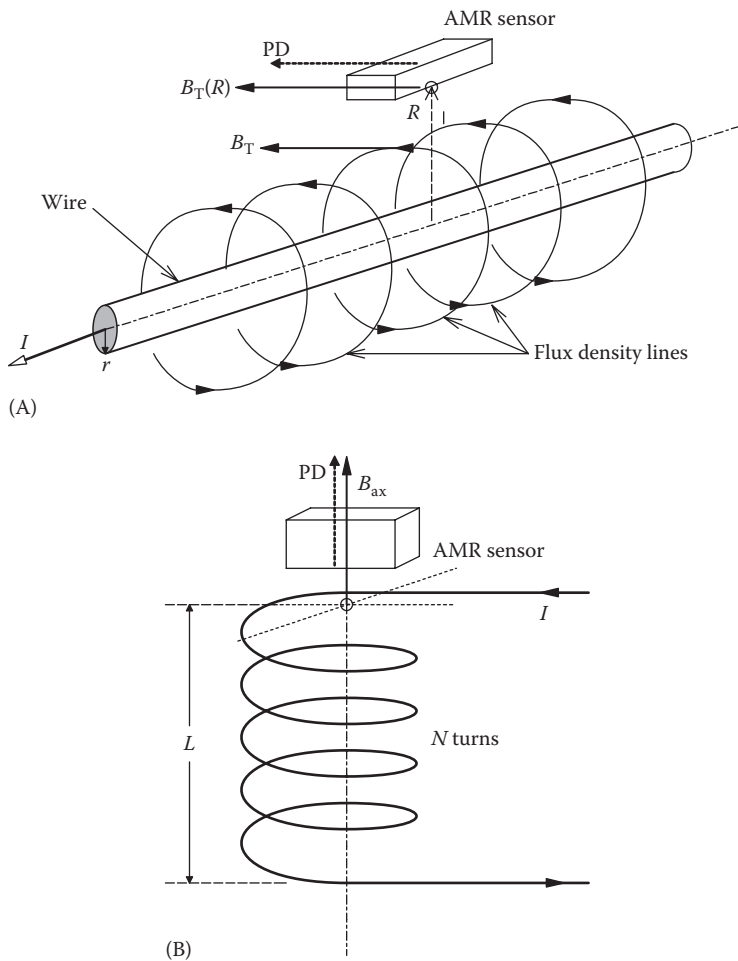
$$B_{\text{ax}} = \frac{\mu_o NI}{(2L)} \quad (6.26)$$

where

L is the solenoid's length

N is the number of turns in the coil (see Figure 6.16)

Curiously, the coil diameter drops out of the equation for B_{ax} . Figure 6.16 illustrates these two geometries for AMR current sensors.

**FIGURE 6.16**

Galvanically isolated current measurement using AMR magnetosensors. (A) The sensor responds to the solenoidal \mathbf{B} field surrounding a single, current-carrying conductor. The tangential B field is given by $B_T = \mu_0 I / (2\pi R)$. The current I can be DC or power line frequency AC. (B) A solenoidal coil can be used to increase sensitivity. It can be shown that the B field at the end of the solenoid is $B_{ax} = \mu_0 NI / (2L)$, where μ_0 is the permeability of empty space $= 4\pi \times 10^{-7}$ H/m, L is the solenoid's length in meters, N is the number of turns on the solenoid, R is the radial distance from the single conductor to the sensor, and $R \gg r$.

AMR sensors also find application in counting paper currency that has magnetic ink or a magnetic stripe embedded in the paper. A biasing magnet's field is disturbed by the bill as it moves by the magnet; this perturbation is sensed by the AMR sensor. This technique can easily detect counterfeit bills and be used to sort moving bills by denomination; it is far less expensive than optical character reading (OCR) techniques.

Still another application of AMR (or for that matter, GMR) sensors is a prototype IsoA design by the author. The heart of this circuit is shown in Figure 6.17. An AMR sensor is connected as a Wheatstone bridge. The bridge can be powered with either DC or AC; if AC is used, a PSR must be placed between DA₂ and the integrator, and also, the excitation frequency, ω_c , must be much higher than the highest frequency in $(v_1 - v_2)$. DC bridge excitation is shown. The voltage difference under isolated measurement is converted to a

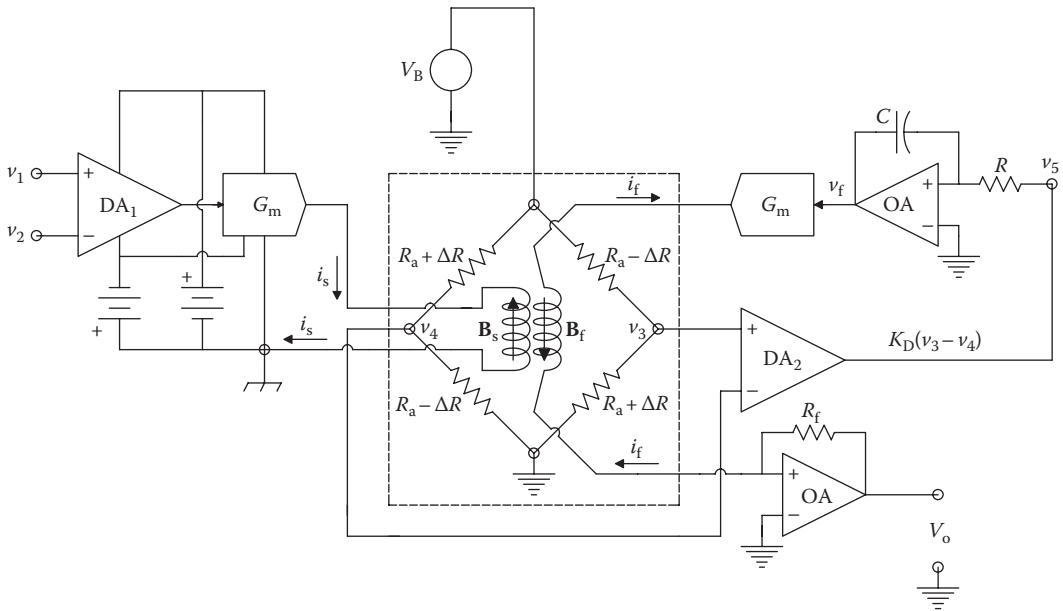


FIGURE 6.17
A prototype IsoA designed by the author using an AMR bridge in a servo-nulling mode.

signal current i_s by a VCCS with transconductance, G_m . The current i_s through the left coil creates a field \mathbf{B}_s that lies along the AMR sensor's PD. \mathbf{B}_s causes identical changes in the barber pole AMRs, ΔR , which in turn unbalance the bridge. The bridge output, $(v_3 - v_4)$, is amplified by DA_2 and serves as the input, v_5 , to the integrator. The integrator output, v_f , causes a feedback current, i_f , to flow, which forces $\mathbf{B}_f = -\mathbf{B}_s$. Because of the integrator, $v_5 = 0$ when $i_f = i_s$, $\mathbf{B}_f = -\mathbf{B}_s$, and the bridge is at null. The output voltage of the OA transresistor is

$$v_o = i_f R_f = G_m v_f R_f. \tag{6.27}$$

Thus, the isolated input voltage is

$$i_f = \frac{v_o}{R_f} = i_s = (v_1 - v_2) K_{D1} G_m \tag{6.28}$$

$$\downarrow$$

$$(v_1 - v_2) = \frac{v_o}{R_f K_{D1} G_m}. \tag{6.29}$$

It is also possible to show

$$(v_1 - v_2) = \frac{v_f}{K_{D1}}. \tag{6.30}$$

Galvanic isolation of the AMR bridge IsoA depends on the insulation between the chip containing the AMR bridge and the signal input coil. Both capacitance between the coils and the chip and the dielectric breakdown of this insulation are important for effective galvanic isolation.

Readers interested in the technology of magnetic field sensing and its applications should see the paper by Caruso et al. (1999a) and Honeywell Sensor Products applications note AN211 (2002).

6.2.2 Voltage-Generating Sensors

A wide variety of physical sensors generate an EMF in response to some input quantity. In most cases, the magnitude of the induced EMF is linearly proportional to the input quantity. Often, the induced voltage is very small, and special types of signal conditioning must be used.

6.2.2.1 Thermocouples and Thermopiles

Thermocouples are traditionally used to measure temperatures, often at extremes. Thermopiles are close arrays of thermocouples arranged in series to have great sensitivity and resolution; they are generally used to measure the energy of EM radiation at optical frequencies, such as from lasers; total absorption of the radiation causes a minute temperature rise, measured by the series thermocouples. Thermopiles also have a newer application as thermoelectric generators (TEGs), power sources for wireless sensors (WSs), and wireless data transmission (WDX) systems. Thermopiles can also be used in the NIR–FIR (700–30,000 nm) to sense blackbody heat radiation from a variety of sources, including human skin where subcutaneous human blood vessels (e.g., the human temporal arteries in the forehead) determine skin temperature. The FIR wavelength range for a noncontact, thermoelectric thermometer is set by the choice of *front window* material on the sensor. For example, see the transmission characteristics of various materials used in IR optical systems in Figure 7.117. IR/FIR thermopiles are made by companies such as Hamamatsu, Newport, Pasco, Solarlight, and Gammadata.

A basic thermocouple consists of a junction, often spot-welded, between two dissimilar metal wires. A basic thermocouple system must contain two couples (see Figure 6.18A). However, thermocouple systems may contain three or more couples. In a two-couple system, there is a net EMF in the loop causing current to flow if the two couples are at different temperatures. In a simple, two-couple system, one couple junction is maintained at a reference temperature, T_R , typically 0°C, formed by melting ice in water. Often, the thermocouple system (measuring and reference junctions), consisting of two different metals, must be joined to a third metal, such as copper, to connect to a millivoltmeter. The joining of the two thermocouple metals to copper produces one additional junction, as shown in Figure 6.18B. Fortunately, the addition of the third metal into the system will have no effect on the performance of the basic, two-metal, thermocouple system as long as the two junctions with the third metal are at the same (reference) temperature. This property is called the *law of intermediate metals* (Beckwith and Buck 1961).

Table 6.3 gives some common thermocouple materials and the temperature ranges over which they are effective. The net EMF in a thermocouple system is not a linear function of the temperature difference between the reference and measuring junction. In general, the net EMF can be expressed as a power series:

$$E_o = \frac{A(\Delta T) + B(\Delta T)^2}{2} + \frac{C(\Delta T)^3}{3} + \dots \quad (6.31)$$

where ΔT is the temperature difference above the reference junction temperature, usually made 0°C.

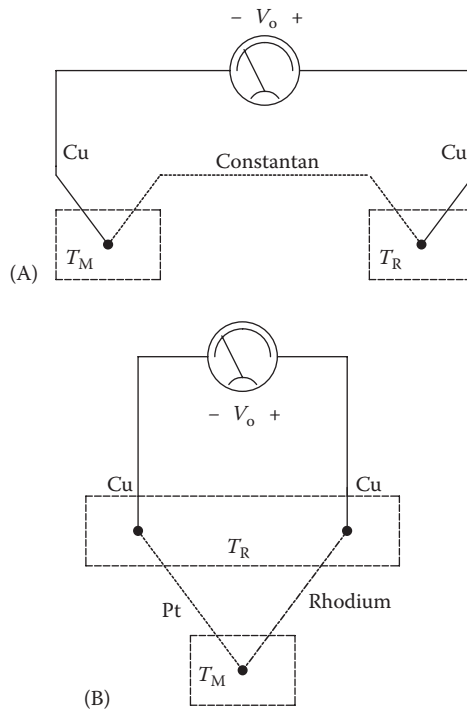


FIGURE 6.18

(A) A simple thermocouple circuit in which there are two couples (one for reference temperature; one for measurement) and two different metals. (B) A thermocouple circuit in which there are three metals. The two dissimilar junctions to copper at the reference temperature behave as the single reference junction in Figure 6.18A.

TABLE 6.3
Properties of Common Thermocouples

Material	S_T vs. Pt	Material	S_T vs. Pt
Bismuth	-72	Silver	6.5
Constantan	-35	Copper	6.5
Nickel	-15	Gold	6.5
Potassium	-9	Tungsten	7.5
Sodium	-2	Cadmium	7.5
Platinum	0	Iron	18.5
Mercury	0.6	Nichrome	25
Carbon	3	Antimony	47
Aluminum	3.5	Germanium	300
Lead	4	Silicon	440
Tantalum	4.5	Tellurium	500
Rhodium	6	Selenium	900

The thermoelectric sensitivity of a thermocouple pair is defined as $S_T = dE_o/dT = A + B(\Delta T) + C(\Delta T)^2$ and is usually expressed in microvolts/°C. Lion (1959) gave a table of thermoelectric sensitivities of various metals vs. platinum at temperatures near 0°C. These are presented in Table 6.4. Lion pointed out that sensitivities of thermocouples made from any pair of metals can be found by subtracting the tabular values of S_T for platinum. For example, S_T

TABLE 6.4

Thermoelectric Sensitivities of Thermocouples Made with Materials Listed vs. Platinum, in $\mu\text{V}/^\circ\text{C}$, with the Reference Junction at 0°C

Materials	Useful Temp. Range, $^\circ\text{C}$	Max Temp. for Short Periods	Sensitivity A in $\mu\text{V}/^\circ\text{C}$ at $^\circ\text{C}$
Copper/constantan	-300 to 350	600	15 at -200 60 at +350
Iron/constantan	-200 to +800	1000	45 at 0° 57 at +750
Pt/Pt90Rh10	0 to 1450	1700	0 at -138 5 at 0 12 at +1500
Iron/copnic	-200 to +860	1000	60 at 0
Chromel P/alumel	-200 to 1200	1350	40-55 between 250 $^\circ\text{C}$ and 1000 $^\circ\text{C}$
W95Re5/W26Re7	0 to 2316		

for copper vs. constantan is $(6.5 - [-35.5]) = 41.5 \mu\text{V}/^\circ\text{C}$. Of special note are the extremely high thermoelectric sensitivities of bulk semiconductor materials such as Ge and Si.

Thermocouple EMFs were traditionally measured using accurate potentiometers (calibrated with standard cells). Modern thermocouple measurement makes use of high input impedance, solid-state, electronic microvoltmeters. Electronically regulated, solid-state, temperature calibration instruments for thermocouple systems are also used, replacing melting ice and water in a thermos flask. Isotech North America, Colchester, VT, makes a *Model 938 Ice Point Reference* for thermocouple cold junctions. It has a temperature of $0^\circ\text{C} \pm 0.03^\circ\text{C}$. Mikron (2013) offers blackbody, IR radiation calibration sources covering from -40°C to 3000°C traceable to NIST standards.

Exergen Corp. of Watertown, MA, markets a line of self-powered, IR temperature sensors using thermocouples. Exergen IR sensors are used for noncontact measurements of object temperatures in industrial processes, ranging from 0°C up to 250°C . Object LIR blackbody radiation is sensed. Eight sensor categories with overlapping ranges cover this span. The range of IR wavelengths that can be sensed is determined by the sensor's *front window* material. Figure 7.117 illustrates the transmission spectra of IR-pass optical materials that can be used for front windows.

A *vacuum thermocouple* is used in conjunction with a sensitive DC microammeter to measure the true rms value of the current in its heater resistance element. A thermocouple junction is bonded to the heater wire and thus generates an EMF proportional to the temperature of the heater resistor. A reference junction is at room (ambient) temperature, so the net EMF of the thermocouple is proportional to ΔT , the temperature of the heater above ambient temperature. The temperature rise of the heater is given by

$$\Delta T = \overline{i^2} R_H \Theta_o \quad (6.32)$$

where

$\overline{i^2}$ is the mean squared heater current

R_H is the heater resistance (assumed constant)

Θ_o is the thermal resistance of the heater in vacuo; its units are degrees Celsius/watt

The DC microammeter current is then given by Ohm's law:

$$I_M = \frac{E_o}{(R_M + R_T)} = \frac{A\Delta T}{(R_M + RT)} = \frac{K_T \overline{i_H^2} R_H}{(R_M + R_T)} \quad (6.33)$$

where

R_M is the microammeter's resistance

R_T is the thermocouple's resistance

A is the thermocouple's EMF constant at room temperature

Typical parameters for the venerable Western Electric model 20D vacuum thermocouple are $R_H = 35$ ohms, $R_T = 12$ ohms, $K_T \cong 2.9$ V/W.

Note that vacuum thermocouples are used for true root-mean square ammeters or voltmeters by giving the associated DC microammeter a square-root scale. Meter deflection is proportional to the power dissipated in R_H , hence i_H^2 . Thermocouple true RMS meters work at DC and at frequencies from the tens of Hz up to the 10's of MHz. Because of their true MS deflection, they are widely used for noise voltage measurements.

6.2.2.2 Photovoltaic Cells

Photovoltaic cells are used to measure light intensity, as well as to generate DC electric power when used as solar cells. Photovoltaic cells include barrier-layer cells and *pn* junction photodiodes. We first examine the photovoltaic effect in *pn* junction diodes. When light of energy $h\nu$ is absorbed by an open-circuited *pn* junction device, the built-in electric field at the junction that results from the doping acts to concentrate the hole-electron pairs generated by photon interaction with valence band electrons in such a manner that the holes are swept toward the *p*-side and the electrons toward the *n*-side of the junction. The net effect is the production of an open-circuit EMF given by the relation (Navon 1975)

$$E_{oc} = v_T \ln \left\{ \left[\frac{qA(L_p + L_n)G(\lambda)}{I_{rs}} \right] + 1 \right\} \quad (6.34)$$

where

$v_T = nkT/q = 0.026$ V at 300 K

q is the magnitude of the electron charge in coulombs

A is the cross-sectional area of the junction

n is a constant ranging from one to two

L_n is the electron diffusion length

L_p is the hole diffusion length

$G(\lambda)$ is the hole-electron pair generation rate per unit volume, which is also a function of the light wavelength and intensity

This relation describes a logarithmic increase of the open-circuit photovoltage with light intensity, other factors being constant. Maximum E_{oc} in bright sunlight is typically 400–500 mV. Figure 6.19A illustrates the current–voltage curves for a photodiode or solar cell in the dark and when illuminated. Figure 6.19B shows typical V_{oc} and I_{sc} curves vs. illumination for a photovoltaic device.

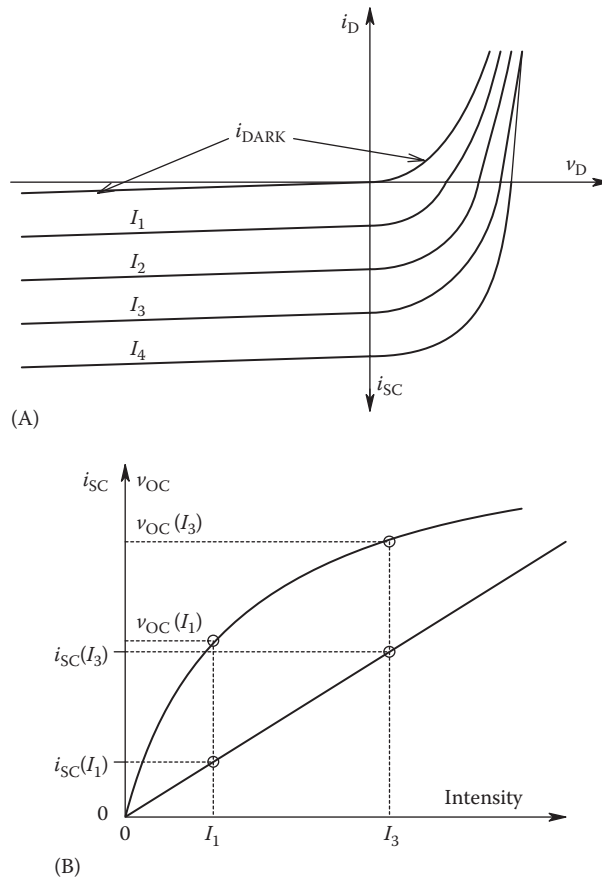


FIGURE 6.19 (A) Current–voltage curves for a photogalvanic (solar) cell or *pin* photodiode. This class of device is generally operated in the fourth quadrant ($v_D > 0, i_D < 0$). (B) Plots of the open-circuit voltage (V_{OC}) and short-circuit current (I_{SC}) of a photogalvanic device. Note that I_{SC} is linear with photon intensity and V_{OC} varies logarithmically.

When illuminated, it can be shown (Nanavati 1975) that the current curve is given by

$$i_D = I_{rs} \left[\exp\left(\frac{v_D}{v_T}\right) - 1 \right] - qA(L_p + L_n)G(\lambda). \tag{6.35}$$

(The parameters in Equation 6.35 are defined in the previous equation.)

Photodiodes operated as photovoltaic cells are maximally sensitive to photons with energies near their bandgap energy, ϕ_o . If $h\nu$ is less than ϕ_o , the photons will not generate hole–electron pairs. On the other hand, if the photons have energies far in excess of ϕ_o , they will be absorbed near the surface of the exposed semiconductor layer where the recombination rate is high, decreasing efficiency. Semiconductor photodiodes made from mixed compounds (PbSe, InAs, InSb) with low bandgap energies are used to create far IR detectors. A line of such detectors are offered by EG&G–Judson Infrared, Inc.

Germanium photodiodes have maximum sensitivity at a wavelength of 155 nm; however, their sensitivity ranges from visible wavelengths to 180 nm. An alternate means of

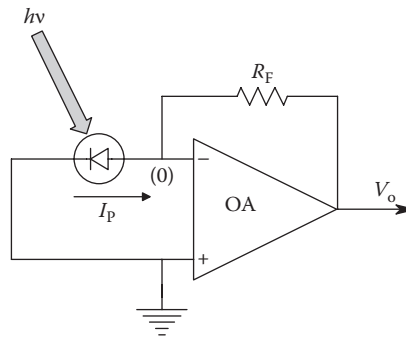


FIGURE 6.20

OA transresistor circuit in which a photodiode is operated in the short-circuit mode. Response speed is relatively slow because of the PD's large zero-bias junction capacitance.

using *pn* photodiodes is shown in Figure 6.20. An OA is used as a current to voltage converter. Because the OA's summing junction is at virtual ground, the photodiode is operated in the short-circuited mode, where $v_D = 0$ V. The OA output is thus given approximately by

$$V_o = R_F q A (L_p + L_n) G(\lambda) V. \quad (6.36)$$

Under zero-bias conditions, the diode's junction capacitance is quite large, and hence its speed of response to changes in the light flux is not as fast as it is for a reverse-biased photodiode operated as a photoresistor. EG&G–Judson Infrared's J16 series of Germanium photodetectors are useable over an optical input intensity range of 2×10^{-12} to 10^{-2} W. Their -3 dB frequencies range widely, from 0.07 to 250 MHz. The faster photodiodes have smaller active junction areas and shunt capacitances and higher shunt resistances.

A *solar cell* is operated in the fourth quadrant of its current–voltage curves (see Figure 6.19A). The open-circuit photovoltage causes a current to flow in a load, and thus the solar cell can deliver electrical power. A solar cell is made with a considerably expanded *pn* junction area, since it must do work. A typical solar cell of 2 cm^2 area might consist of a thin (0.5 mm), rectangular slab of *n*-type silicon, on top of which is diffused a thin, *p*-type (boron treated) surface layer, through which the light must pass. Such a 2 cm^2 cell can deliver about 10 mW in bright sunlight with about 12% conversion efficiency. Solar cells are designed to have maximum efficiency in delivering SS, DC power to a load, which might be charging batteries used for a portable instrumentation module. To obtain maximum efficiency, there is a trade-off between a number of design factors. For example, the thickness of the top (*p*) layer of the solar cell, d , measured from its surface to the center of the junction, must be less than L_n to allow electrons generated near the surface to diffuse to the junction before they recombine. Similarly, the thickness of the *n* region must be such that the electrons generated in the region can diffuse to the junction before they, too, recombine. This requirement implies a match between the hole diffusion length, L_p , the thickness of the *n* region, and the optical penetration depth in the surface *p* layer. It is also desirable in solar cell design to have a large contact potential in order to realize a large photovoltage. This means that heavy doping must be used. The need for heavy doping must be compromised with the requirement for long carrier lifetimes, which are reduced by heavy doping. Efficient solar cells require a low internal (ohmic) resistance. There is no problem in making metallic electrode contact with the bottom (dark) layer of the solar cell,

but low resistance also requires a large electrode contact area with the top layer. A large metallic area on the top semiconductor layer prevents light from reaching the cell, decreasing efficiency. The tops of solar cells are generally given an antireflection coating, similar to that used on camera lenses, to maximize light absorption.

6.2.2.3 Piezoelectric Transducers

There is a class of piezoelectric materials that can serve, among other applications, as mechanical input transducers, enabling the measurement of pressure, force, displacement, and other physical phenomena that can be related to them. Piezoelectric materials, when mechanically strained in a preferred manner, generate an open-circuit EMF. Actually, the mechanical strain causes a unidirectional separation of electric charges resident in the interior of the piezoelectric crystal structure. These displaced charges form an effective net charge on a capacitor formed by the bonding of metallic electrodes on the surface of the piezoelectric crystal. The transducer material itself, independent of piezoelectric activity, is an insulator having very low conductivity and a high dielectric constant. Thus, the effectively separated charges produce an open-circuit voltage on the capacitor equal to Q/C . In practice, the situation is not that simple. The piezoelectric transducer material is not a perfect dielectric, and the charges leak off through the exceedingly small volume conductance. Of course, any voltage measuring system attached to the transducer will have a finite input impedance and perhaps a DC bias current, further altering the stress-caused EMF. In addition, the transducer has mechanical mass, stiffness, and damping, giving it mechanical resonance properties that can further complicate the transducer's transfer function when the frequency of the mechanical input approaches the transducer's mechanical resonance frequency. In the following discussion, we have assumed that the piezoelectric transducers are operated at frequencies well below their mechanical resonant frequencies.

Piezoelectric materials include a number of natural crystals and man-made ceramic materials. Quartz, Rochelle salt, and ammonium dihydrogen phosphate (ADP) are examples of naturally occurring piezomaterials; barium titanate, lead zirconate-titanate, and lead metaniobate are synthetic piezoceramics. All piezoelectric materials have a *Curie temperature*, above which piezoelectric activity ceases to exist.

In order to understand how piezoelectric input transducers work, we will examine an example of a crystal responding to thickness compression. Figure 6.21 shows a rectangular

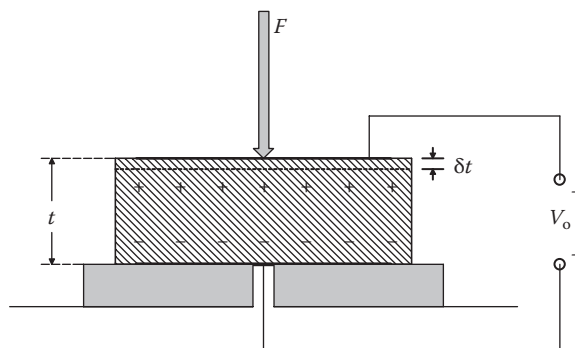


FIGURE 6.21

A piezoelectric transducer slab responding to thickness compression.

crystal block with metallic electrodes vapor deposited on its top and bottom surfaces. The transducer has thickness t and top area A . The capacitance of the transducer is approximately

$$C_x = \frac{\kappa \epsilon_0 A}{t} \text{ F} \quad (6.37)$$

where κ is the dielectric constant of the piezoelectric material.

A uniformly applied force F will cause the crystal to compress a minute amount, δt . The piezomaterial is assumed to have a Young's modulus, Y :

$$Y = \frac{\text{stress}}{\text{strain}} = \frac{(F/A)}{(\delta t/t)}. \quad (6.38)$$

The net charge displacement for the transducer is given by

$$Q = dF \text{ C} \quad (6.39)$$

where d is the *charge sensitivity* of the material in $(\text{C}/\text{m}^2)/(\text{N}/\text{m}^2)$.

Thus, we can write the open-circuit voltage across the transducer as

$$E_o = \frac{Q}{C} = \frac{dF}{\kappa \epsilon_0 A/t} = g \left(\frac{F}{A} \right) t \text{ V} \quad (6.40)$$

where

g is the *voltage sensitivity* of the transducer in $(\text{V}/\text{m})/(\text{N}/\text{m}^2)$

F/A is the pressure on the transducer's active surface

The axes along which a natural crystal of piezoelectric material is cut in order to make a transducer determine the type of mechanical input to which it will have maximum sensitivity. In addition to simple thickness compression, piezoelectric transducers also may be made to respond to length compression, thickness, and face shear, as well as twisting and bending. Synthetic piezomaterials may be cast in a variety of shapes not available with natural materials, such as thin-walled cylinders, half-cylinders, disks, hollow spheres, and hemispheres. Table 6.5 summarizes some of the salient properties of certain piezomaterials.

For mechanical input frequencies well below piezoelectric transducer mechanical resonance, the equivalent circuit shown in Figure 6.22 may be used to model the electrical behavior of the transducer. Although charge is displaced internally by strain, we must use a current source in the model. Since current is the rate of flow of charge, the equivalent current source is

$$i_{\text{eq}} = \frac{dQ}{dt} = d \left(\frac{dF}{dt} \right) = d\dot{F}. \quad (6.41)$$

Thus, from Ohm's law, the open-circuit voltage across the transducer is given by the transfer function

$$\frac{V_o}{F}(s) = \frac{sFdR_L}{s + sC_T/G_L}. \quad (6.42)$$

TABLE 6.5
Properties of Some Piezoelectric Materials

Material	Orientation	Charge Sensitivity, d	Voltage Sensitivity, g	Curie Temp.
Quartz	X cut; length along Y , length longitudinal	-2.25×10^{-12}	0.055	550°C
	X cut; thickness longitudinal	-2.04	-0.050	
	Y cut; thickness shear	4.4	0.108	
Rochelle salt	X cut 45°; length longitudinal	435	0.098	55°C
	Y cut 45°; length longitudinal	-78.4	-0.29	
Ammonium dihydrogen phosphate	Z cut 0°; face shear	48	0.354	125°C
	Z cut 45°; length longitudinal	24	0.177	
Barium titanate	Parallel to polarization	86–160	0.016	125°C
	Perpendicular to polarization	-56	0.005	
Lead zirconate titanate	Parallel to polarization	190–580	0.02–0.03	170°C–360°C
Lead metaniobate	Parallel to polarization	80	0.036	>400°C

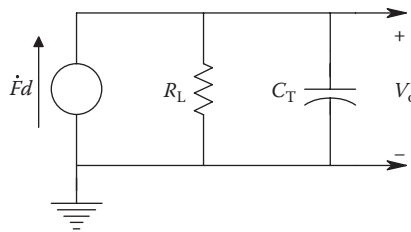


FIGURE 6.22

Electrical equivalent circuit of a piezoelectric transducer responding to applied force, $F(t)$. Mechanical frequencies in the applied force are assumed to be well below the transducer's natural mechanical resonant frequency. Note that the current source is proportional to dF/dt .

This transfer function is seen to be a simple HPF. Thus, because the charge displaced by the strain caused by force F leaks off the crystal, true DC response of a piezoelectric transducer is impossible. G_L is the total leakage conductance of the piezoelectric crystal plus the input conductance of the meter or oscilloscope used to measure V_o , plus the leakage conductance of the cable used to connect the transducer to the measuring system. C_T is the capacitance of the transducer plus the input capacitance of the measuring device. If the transducer is connected to an oscilloscope with a $10 \text{ M}\Omega$ input resistance, we can neglect the other leakage resistances because they are very large ($\approx 10^{13}$ ohms). C_T may be on the order of 1 nF for a piezoceramic transducer. Thus, the time constant of the transducer/oscilloscope system is ca. 10 ms , and the response to a step of force input will have decayed to zero in about 30 ms .

To improve and control LF response of a piezoelectric transducer system, we wish to connect the transducer to an amplifier with an ultrahigh input impedance. Such amplifiers are called *electrometer amplifiers*, and they generally have input resistances of 10^{13} – 10^{15} ohms. Another common means of conditioning the output of a piezoelectric transducer is to use *charge amplifier* configuration using an EOA (charge amplifiers are introduced in Section 2.6.1). Figure 6.23 shows one circuit for a charge amplifier. Here, we neglect any series resistance between the transducer and the OA's summing junction. If the OA

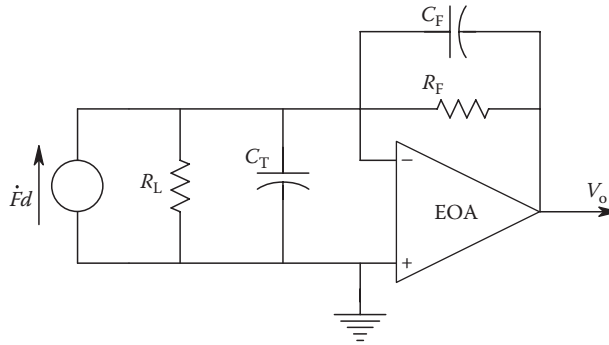


FIGURE 6.23

An EOA is used as a charge amplifier to condition the output of a piezoelectric force (or pressure) transducer.

is considered to be ideal, it is easy to see that the transducer current, $(dF/dt)d$, only flows through the amplifier’s feedback conductance. Hence, the transfer function of the transducer and amplifier can be written as

$$\frac{V_o}{F}(s) = \frac{-dR_F s}{s + C_F R_F} \text{ V/N.} \tag{6.43}$$

We still have a high-pass system with no DC response, but now the effect of C_T and G_T is negligible, and the LF pole is controlled with the known feedback elements, R_F and C_F . For example, let R_F be 10^{10} ohms and C_F be 5 nF. The time constant is then 50 s, and the break frequency is 3.18 mHz. The HF gain is then $-d/C_F$ volts/newton. For a quartz transducer, d is about 2.3×10^{-12} C/N; thus, the midband gain would be -4.6×10^{-4} V/N. The use of piezoceramics with larger d ’s obviously leads to greater sensitivities when used in the charge amplifier configuration. Because of their greater stability and higher Curie temperature, quartz pressure transducers are used in applications such as the measurement of internal combustion engine combustion chamber transient pressures.

Other applications of piezoelectric input transducers include measurement of minute, dynamic displacements (microphones, phonograph cartridges to translate the grooves in vinyl phonograph records into sound, measurement of minute pulsations on body surfaces, measurement of surface roughness, measurement of acceleration [with an attached mass], etc.). Piezoelectric transducers are also used as sonar receivers or, in ultrasonic applications such as medical imaging, are generally operated at their mechanical resonant frequencies for maximum sensitivity. At resonance, the electrical equivalent circuit of a transducer changes markedly.

Figure 6.24 illustrates the equivalent electric circuit model developed by Fox and Donnelly (1978) based on the Mason (1957) electrical model, for a piezoelectric input transducer loaded (radiating) on both faces, valid for frequencies near the mechanical resonance frequency, f_o , of the transducer. V is the applied driving voltage. η is the ideal transformer’s turns ratio in the Mason model. It can be shown that $\eta = Ae_{33}/t$, where A is the effective surface area of the transducer in m^2 , e_{33} is the piezoelectric stress constant in the material in C/m^2 , and t is the transducer thickness in m. (Here, we assume a transducer responding in the thickness mode.) In addition, the characteristic impedance of the piezoelectric material is given by $Z_x = 2\rho A t f_o$, where ρ is the density of the piezomaterial in kg/m^3 and f_o is the transducer’s Hz mechanical resonant frequency. The parameter $\alpha = \pi(\delta f/f_o)$, where $\delta f/f_o$ is the fractional deviation of the applied mechanical frequency from the mechanical resonance frequency.

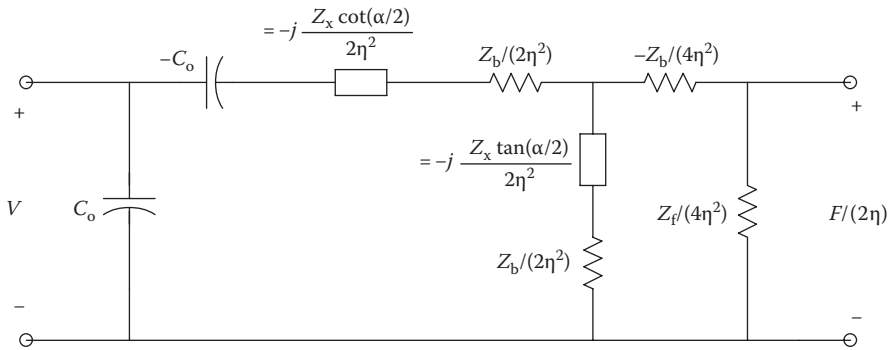


FIGURE 6.24

Electrical equivalent circuit of a piezoelectric transducer near its mechanical resonance frequency. See text for description of parameters.

When the transducer is driven at exactly its mechanical resonant frequency, f_o , its electrical equivalent circuit (Thevenin source impedance) reduces to a series RC circuit with capacitance C_o and a real part of $4\eta^2 / [(2\pi f_o C_o)2(Z_f + Z_b)]$ ohms. Here, C_o is the transducer's clamped capacitance and Z_f and Z_b are the (real) acoustic source impedances faced by the front and back faces of the transducer, respectively. As can be seen from the transducer model of Figure 6.24, operation of the transducer even slightly off the resonant frequency leads to a far more complex model for the Thevenin source impedance of the transducer's electrical output. In general, the real part of the Thevenin impedance of the transducer's output is numerically small, typically ranging from the tens of ohms to the hundreds of ohms. Contrast this with the extremely high resistances presented by piezoelectric transducers operated well below their resonant frequencies.

As a consequence of the low, real source impedance presented by a piezoelectric input transducer operated at its mechanical resonance frequency, a step-up transformer can be used to obtain optimum SNR when measuring threshold acoustic signals, as is done in sonar systems. Transformer optimization of output SNR is discussed in detail in Section 3.5.

6.2.2.4 Pyroelectric Sensors

Pyroelectric materials (PyMs) are crystalline or polymer substances that generate internal, electrical charge transfer in response to *internal heating* caused by the absorption of long-wave, IR radiation (LIR). The charge transfer can be sensed as a current or voltage change, depending on the kind of electronic signal conditioning associated with the PyM. In general, PyM materials are also piezoelectric; that is, they also respond to applied mechanical stress by internal charge transfer. PyM sensor materials include the polymers, *polyvinylidene fluoride* (PVDF) and *polyvinyl fluoride* (PVF), and the crystalline substances, *lithium tantalate* (LiTaO_3), *strontium* and *barium niobate*, *triglycine sulfate* (TGS), *Rochelle salt*, *KDP* (KH_2PO_4), *ADP* ($\text{NH}_4\text{H}_2\text{PO}_4$), *barium titanate* (BaTiO_3), *lead zirconate titanate* (LZT), etc.

PyM sensors are fabricated by taking a thin rectangle or disk of the PyM material and coating both sides with a very thin layer of vapor deposited metal such as gold, silver, or aluminum. Electrical contact is made with silver epoxied wires or pressure contacts. The side of the sensor that is to receive radiation is often given an extra, thin, rough, heat-absorbing coating such as platinum black. This coating maximizes the ratio of absorptivity to reflectivity for the PyM sensor. As IR radiation sensors, PyM sensors are used in the LIR wavelength range of 2–25 μm (5000 cm^{-1} to 400 cm^{-1} wavenumber).

Note that PyM sensors respond only to *changes* in temperature. Their input DC IR radiation must be chopped to be sensed. In the thermal SS, there is no net internal charge transfer and no voltage across their electrodes. Thus, if a constant input radiation power, P_i , is applied, the sensor's temperature rises to an equilibrium value, $T'_a > T_a$, where radiation and conduction heat losses equal the input power. In general, we can write a heat balance differential equation for the PyM sensor:

$$P_i(t) = C_T \frac{d(T'_a - T_a)}{dt} + \frac{(T'_a - T_a)}{\Theta} = C_T \frac{d\Delta T}{dt} + \frac{\Delta T}{\Theta} \tag{6.44}$$

where

- C_T is the PyM material's *heat capacity* in J/K
- Θ is its *thermal resistance* in °K/W
- $\Delta T = T'_a - T_a$

T_a is the starting, SS, ambient temperature of the PyM and T'_a is the temperature it rises (or falls) to as a result of absorbing (radiating) P_i over some time T_s .

Θ depends on the PyM material used, its configuration, and even how it is mounted. (Θ can be reduced by direct thermal conduction (heatsinking) and by air convection.) C_T is given by

$$C_T = cAh \tag{6.45}$$

where

- c is the PyM's *specific heat* in J/(cm³ K)
- A is the *absorbing surface area* in cm²
- h is the *PyM thickness* in cm

If $\Delta T/T$ is small, we can assume that P_i remains constant over T_s . The differential equation, (6.44), can be Laplace transformed and written as a transfer function:

$$\frac{\Delta T}{P_i}(s) = \frac{\Theta}{s\Theta C_T + 1} \tag{6.46}$$

Now, the short-circuit current from the irradiated PyM is given by

$$i_p(t) = K_p A \Delta \dot{T} \tag{6.47}$$

Laplace transforming

$$I_p(s) = K_p A s \Delta T(s) \tag{6.48}$$

When Equation 6.48 for $I_p(s)$ is substituted into Equation 6.46, we can finally write the transfer function

$$\frac{I_p}{P_i}(s) = \frac{sK_p A \Theta}{(s\Theta C_T + 1)} \tag{6.49}$$

K_p is the PyM's *pyroelectric coefficient* in C/(m² K). Table 6.6 gives the important physical constants of certain common PyMs (note units).

TABLE 6.6

Physical Properties of Certain PyMs

Pyroelectric Material	Pyroelectric Coefficient K_p in $\mu\text{Cb}/(\text{m}^2\text{K})$	Dielectric Constant $\kappa = \epsilon/\epsilon_0$	Thermal Resistance $\Theta^\circ\text{K}/\text{watt}$	Specific Heat in $\text{J}/(\text{cm}^3 \text{K})$
Triglycine sulfate (TGS)	350	3.5	2.5×10^{-3}	2.5
Lithium tantalate (LiTaO_3)	200	46	2.38×10^{-4}	3.19
Barium titanate (BaTiO_3)	400	500	3.33×10^{-4}	2.34
PVDF film	40	12	7.69×10^{-3}	2.4

Source: Data from Pállas-Areny and Webster (1991).

Assume the PyM sensor is at thermal equilibrium at temperature T_a . A radiation-blocking shutter is opened, permitting a *step* of radiation from a warm object at constant temperature T_b to reach the sensor.

The *short-circuit current* is given by Equation 6.22:

$$I_p(s) = \frac{P_{io}}{s} \frac{sK_p A \Theta}{(s\Theta C_T + 1)} = P_{io} \frac{K_p A / C_T}{(s + 1/(\Theta C_T))}. \quad (6.50)$$

In the time domain, this is simply an exponential decay waveform:

$$i_p(t) = \left(\frac{P_{io} K_p A}{C_T} \right) \exp \left[\frac{-t}{(\Theta C_T)} \right]. \quad (6.51)$$

ΘC_T is the sensor's *thermal time constant*, which is material and dimensionally dependent. P_{io} is assumed constant and is given by

$$P_{io} = ADH = K_s (T_b^4 - T_a^4) \quad (6.52)$$

$$K_s = \frac{A\sigma}{(1/e_a + 1/e_b + 1)}. \quad (6.53)$$

Now, examine the circuit in Figure 6.25. The PyM sensor is connected to an EOA, connected as a transimpedance amplifier. Note the equivalent circuit for the PyM can be configured as an *ideal current source*, $i_p(t)$, in parallel with the sensor's very low electrical leakage conductance, G_p , and its electrical, self-capacitance, C_p . $i_p(t)$ is given by Equation 6.51 for a step input of LIR power. Let us first neglect the OA's feedback capacitor, C_F . The output voltage of the OA is given by (note the direction of i_p)

$$v_o(t) = R_F i_p(t) = R_F \left(\frac{P_{io} K_p A}{C_T} \right) \exp \left[\frac{-t}{(\Theta C_T)} \right] \geq 0. \quad (6.54)$$

The peak $v_o(t)$ is $V_{opk} = R_F (P_{io} K_p A / C_T)$ volts; hence, we can calculate the temperature T_b of the warm object from V_{opk} and a knowledge of the system's constants:

$$T_b = \left[T_a^4 + \frac{V_{opk} C_T}{R_F K_p A K_s} \right]^{1/4} = \left[T_a^4 + V_{opk} K_{sys} \right]^{1/4}. \quad (6.55a)$$

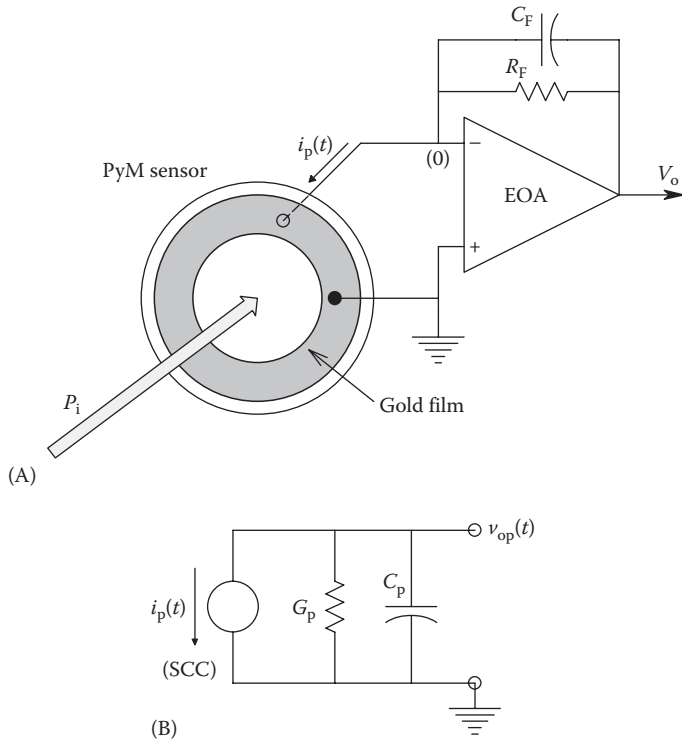


FIGURE 6.25

(A) An electrometer charge amplifier is used to condition the output of a pyroelectric IR sensor. (B) Equivalent circuit of the PyM sensor away from its mechanical resonance.

The problem with this approach is that while V_{opk} is measured fairly accurately, the system's constants and T_a are generally not that accurately known; hence, a known blackbody temperature source at T_{cal} can be used to find the lumped constant K_{sys} :

$$K_{sys} = \frac{(T_{cal}^4 - T_a^4)}{V_{ocal}}. \tag{6.55b}$$

Such a blackbody reference source can be built in to the transient radiation thermometer or be an external blackbody source, such as the Mikron™ Model M310 blackbody radiation calibration source.

Another approach to self-calibration is to keep the PyM at a known T_a , then expose it to T_{cal} and measure the peak V_{ocal} , then expose it to T_b and measure V_{opk} . The computer subtracts V_{opk} from V_{ocal} to form ΔV_o :

$$\Delta V_o = \left(\frac{1}{K_{sys}} \right) [T_{cal}^4 - T_b^4]. \tag{6.56}$$

Solving for T_b ,

$$T_b = [T_{cal}^4 - K_{sys}\Delta V_o]^{1/4}. \tag{6.57}$$

Thus, we see that the calculated T_b of the body relies on measurement of ΔV_o , K_{sys} , and the known T_{cal} .

To counteract HF measurement noise, a low-pass filter can be added to the current-to-voltage conversion OA in Figure 6.25 by placing capacitor C_F in the feedback path. With C_F in place, the system response to a *step input* of LIR irradiative power, P_i , is

$$V_o(s) = P_{i0} \frac{K_p A / (C_F C_T)}{[s + 1 / (C_F R_F)] [s + 1 / (C_T \Theta)]}. \quad (6.58)$$

The inverse Laplace transform of Equation 6.58 can be shown to be

$$v_o(t) = \frac{P_{i0} K_p A R_F \Theta}{(C_F R_F - C_T \Theta)} \left\{ \exp \left[\frac{-t}{(C_F R_F)} \right] - \exp \left[\frac{-t}{(C_T \Theta)} \right] \right\} V. \quad (6.59)$$

This is a positive waveform that rises with an initial time constant $C_T \Theta$ and falls more slowly with time constant, $R_F C_F$. Its peak is proportional to P_{i0} . The measurement noise comes from (1) unwanted mechanical vibration of the PyM sensor that is also piezoelectric, (2) electronic noise from the OA, and (3) thermal (Johnson) noise from the PyM's Norton conductance and from R_F .

Applications of PyM sensors include intruder sensors and burglar alarms, automatic patio light switches, and sensors used in various LIR spectrometers used in analytical chemistry. They are also used in the Braun Thermoscan™ IR fever thermometers that quickly and noninvasively measure the temperature of the eardrum (Northrop 2002). Commercial PyM sensors come in a variety of packages.

Some manufacturers (e.g., Molectron Detector, Inc., Portland, OR) package PyM sensors alone, with OAs or a JFET for the first-stage signal conditioning. Manufacturers such as GEC-Marconi (now BAE Systems Electronics Ltd.) list PyM sensor *responsivity* (nV/W(input)), *noise voltage* ($\mu\text{Vrms}/\sqrt{\text{Hz}}$), *noise-equivalent power* (NEP) (in $\text{nW}/\sqrt{\text{Hz}}$), *noise-equivalent irradiance* (NEI) (in $10^{-4} \text{ W}/(\sqrt{\text{Hz}} \text{ m}^2)$), and *specific detectivity* (D^*) ($10^6 \text{ m}\sqrt{\text{Hz}}/\text{W}$). For example, at 1.0 Hz, the GEC-Marconi/BAE PLT533 lithium tantalate PyM sensor has responsivity = 1800 V/W, noise = $1.1 \mu\text{V}/\sqrt{\text{Hz}}$, NEP = $0.6 \text{ nW}/\sqrt{\text{Hz}}$, NEI = $0.7 \times 10^{-4} \text{ W}/(\text{m}^2 \sqrt{\text{Hz}})$, and $D^* = 5 \times 10^6 \text{ m}\sqrt{\text{Hz}}/\text{W}$.

6.2.3 Sensors Whose Voltage Output Is Proportional to $d\Phi/dt$

It is well known from basic physics that an EMF will be induced in a coil of N turns surrounding a magnetic flux, Φ , when the magnetic flux changes in time. This effect may be simply stated mathematically as

$$E_o = N \frac{d\Phi}{dt} = L \frac{di}{dt} \text{ V}. \quad (6.60)$$

Many input transducers make use of this principle, including variable-reluctance phonograph pickups, dynamic microphones, accelerometers, and tachometers. The source of the magnetic flux may be either a permanent magnet or a DC- or AC-excited electromagnet (solenoid). The $d\Phi/dt$ may arise from a number of means, including modulation of the reluctance of the magnetic path by changing the size of an air gap in the flux path. Moving a coil perpendicular to the flux density also gives a $d\Phi/dt$, as does moving a permanent magnet in relation to a fixed coil.

6.2.3.1 Variable Reluctance Phonograph Pickup

The variable-reluctance phonograph pickup is a mechanoelectromagnetic sensor used to sense the mechanical variations in the width of sound-encoded grooves in a rotating vinyl audio disk. The *needle* in the spiral groove is forced from side to side as the disk rotates at a constant speed; this in turn differentially changes the magnetic reluctance in two magnetic flux paths, generating a small audio voltage proportional to needle deflection, hence the recorded sound. Vinyl audio disks of 33 1/3 rpm are still played by collectors, but have been largely displaced by magnetic tape media, digital optical media (compact audio disks), and purely digital media (RAM, flash drives).

Magnetic reluctance, \mathfrak{R} , is defined as the ratio of magnetomotive force to the magnetic flux, Φ , existing in a closed, magnetic *circuit*. It may be thought to be analogous to electrical resistance in an electrical circuit where the MMF is analogous to voltage and the flux Φ is analogous to current. A schematic of a simple, variable-reluctance phonograph pickup transducer is shown in Figure 6.26. The source of MMF is a permanent magnet (PM). Flux from the PM is split and passed through a right and left branch magnetic path so that the total flux in the magnet is the sum of left and right path fluxes. Thus,

$$\Phi = \Phi_R + \Phi_L \quad W. \tag{6.61}$$

The reluctances of the paths are modulated by the lengths of the air gaps caused to vary differentially by the small lateral displacement of the moveable arm in the air gap, δx . The arm is displaced by a lateral force, F , acting on the stylus, S , protruding downward from the arm. In playing a vinyl, nonstereo phonograph record, the stylus tip follows the lateral undulations in the helical groove of the record. Audio information is recorded on the record as lateral, side-to-side displacements of the groove. Because of its much higher inertia, the tone arm and cartridge do not respond to the small, audio-frequency lateral forces

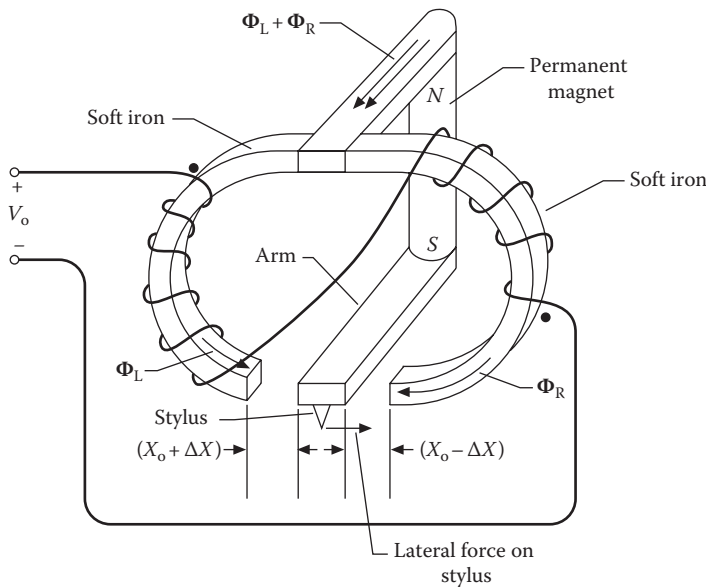


FIGURE 6.26 Diagram of a monophonic, variable-reluctance phonograph pickup.

the groove places on the stylus. Instead, the moveable (stylus) arm of the cartridge follows the lateral modulations of the groove. The reluctance of the right- and left-hand magnetic paths may be written as follows:

$$\mathfrak{R}_R = \mathfrak{R}_o - \frac{\delta x}{\mu_o A} \quad (6.62)$$

$$\mathfrak{R}_L = \mathfrak{R}_o + \frac{\delta x}{\mu_o A}. \quad (6.63)$$

The total magnetic flux passing through the magnet is then

$$\Phi = \frac{\text{MMF}}{(\mathfrak{R}_R + \mathfrak{R}_L)} = \frac{\text{MMF}}{2\mathfrak{R}_o} \text{ (constant)}. \quad (6.64)$$

An N -turn coil is wound with the dot polarity shown on the left arm of the magnetic circuit; this coil is connected in series with an N -turn coil on the right arm having the opposite dot polarity. The magnetic flux in the left arm is given by

$$\Phi_L = \frac{\text{MMF}/\mathfrak{R}_o}{(1 + \delta x/\mu_o A \mathfrak{R}_o)} \cong \frac{\text{MMF}}{\mathfrak{R}_o} (1 - \delta x/\mu_o A \mathfrak{R}_o). \quad (6.65)$$

Similarly, the flux in the right arm can be approximated by

$$\Phi_R \cong \frac{\text{MMF}}{\mathfrak{R}_o} (1 + \delta x/\mu_o A \mathfrak{R}_o). \quad (6.66)$$

The net EMF induced in the two series-connected coils is found by taking the time derivative of the flux in each arm, multiplying by the number of turns in each coil, and adding the EMFs algebraically:

$$E_o = \frac{2N(\text{MMF})}{\mathfrak{R}_o^2 \mu_o A} \left(\frac{d[\delta x]}{dt} \right) \ddagger \quad (6.67)$$

Thus, we see that the *output EMF is proportional to the stylus velocity*. The variable-reluctance transducer does not respond to constant (DC) stylus displacements, similar to a piezoelectric phonograph pickup.

6.2.3.2 Electrodynamic Accelerometer

Figure 6.27 illustrates a linear accelerometer made from a moving, cylindrical mass (which is also a permanent magnet), which moves inside a coil fixed to the accelerometer's case. The mass is constrained by a linear spring with stiffness k_s and is surrounded by oil, which gives viscous damping to the motion of the mass. The case of the accelerometer is attached to some structure of which we wish to measure the linear acceleration. The output voltage of the accelerometer is proportional to the relative velocity between the case (and structure) and the moving mass. In mathematical terms,

$$V_o = K_v(\dot{x}_c - \dot{x}_m). \quad (6.68)$$

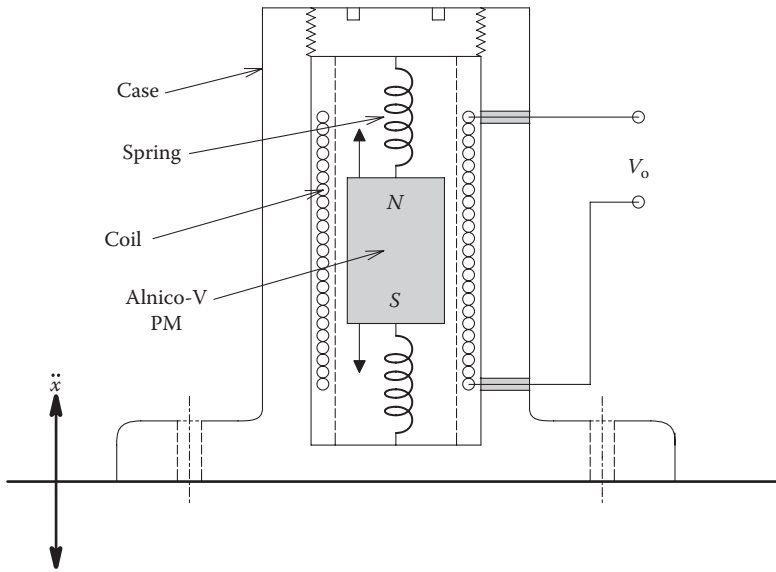


FIGURE 6.27
Cross section of an electrodynamic accelerometer.

The motion of the mass can be described by simple Newtonian mechanics:

$$M\ddot{x}_m + B(\dot{x}_m - \dot{x}_c) + k_s(x_m - x_c) = 0. \tag{6.69}$$

The acceleration, \ddot{x}_c , is considered to be the system input. If we separate terms of the force equation and take their Laplace transforms, we may write the transfer function:

$$\frac{X_m(s)}{X_c(s)} = \frac{sB + k_s}{s^2M + sB + k_s} = \frac{sB/k_s + 1}{s^2M/k_s + sB/k_s + 1}. \tag{6.70}$$

Now, the relation for $\dot{X}_m(s) = sX_m(s)$ is substituted into the equation for the output voltage. After some algebra, this yields

$$V_o(s) = \ddot{X}_c(s) \frac{sM/k_s}{s^2M/k_s + sB/k_s + 1} \tag{6.71}$$

Thus, we see that this transducer’s response at low mechanical input frequencies is proportional to *jerk*, that is, the rate of change of acceleration of the structure to which the case is attached. There is no output in response to constant acceleration, such as due to the Earth’s gravitational field.

6.2.3.3 Linear Velocity Sensors

A linear velocity sensor (LVS), as the name suggests, produces an output voltage that is directly proportional to the rate of linear displacement of an object attached to its core. Figure 6.28 illustrates a section through an LVS. Note that the two coils are connected in series-opposing configuration to improve linearity over the full displacement range.

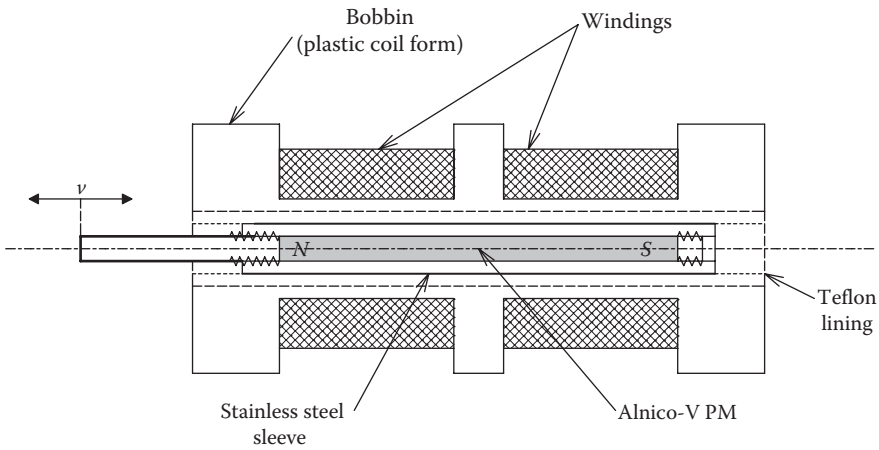


FIGURE 6.28
Cross section of an LVS.

Schaevitz Engineering (now Sherborne Sensors) offered LVSs covering ranges of motion from 12.5 to 500 mm. Sensitivities range from 4.8 to 26 mV/mm/s in various models using Alnico-V magnetic cores. Linearity is $\pm 1\%$ of output over the nominal linear range.

6.2.4 Sensors Whose Output EMF Depends on the Interaction of a Magnetic Field with Moving Charges

The fluid version of this class of generating sensor is generally used with a constant magnetic (\mathbf{B}) field to measure the average velocity of a fluid flowing in an insulating pipe or conduit, including blood vessels. The solid-state version of this class of sensor is the well-known Hall effect device, used to measure magnetic fields or to compute average power as a wattmeter.

6.2.4.1 Faraday Effect Flowmeters

Figure 6.29 illustrates the basic geometry of a Faraday flowmeter. Two electrodes on opposite sides of the circular cross section pipe make ohmic contact with the conductive fluid in the pipe moving with average velocity, \mathbf{v} . A uniform magnetic field of \mathbf{B} Tesla is perpendicular to the velocity \mathbf{v} vector and the \mathbf{L} vector between the centers of the electrodes. Faraday’s law of induction predicts that the EMF between the electrodes is given by

$$E_F = \int_0^d [(\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{L}] \quad \text{V} \tag{6.72}$$

where

- ($\mathbf{v} \times \mathbf{B}$) is the vector cross product vector (see Glossary)
- \cdot denotes the vector dot product operation that generates a real quantity (in brackets)

For orthogonal vector components, Equation 6.72 reduces to $E_F = BLv$. For effective measurement, the conductivity of the fluid should be greater than 10^{-5} per ohm/cm. If a circular

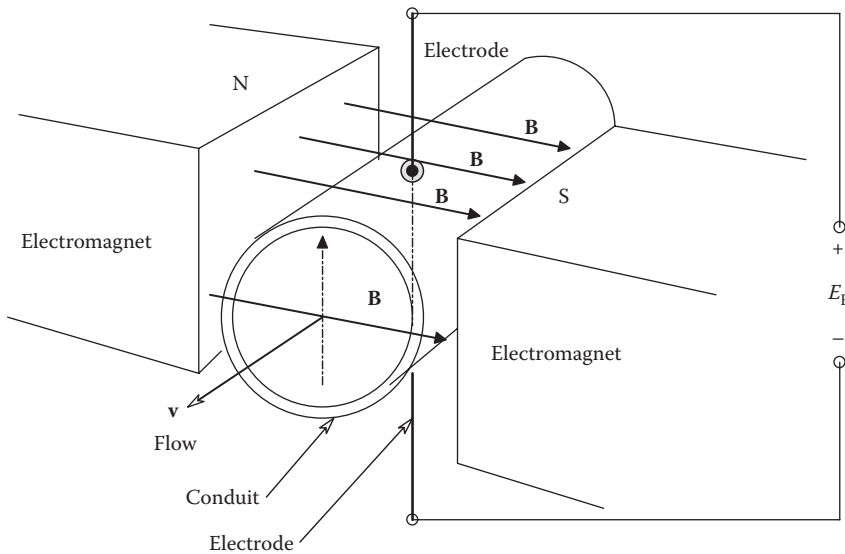


FIGURE 6.29 Diagram of a Faraday effect, liquid velocimeter. Note the flow vector is orthogonal to **B**.

pipe of diameter d is used for the conduit, and d is also the separation of the electrodes, then the average volume flow, Q , of the conducting fluid may be shown to be given by Lion (1959):

$$Q = \frac{E_F d \pi}{(4 \times 10^{-8} B)} \text{ mL/s.} \tag{6.73}$$

where

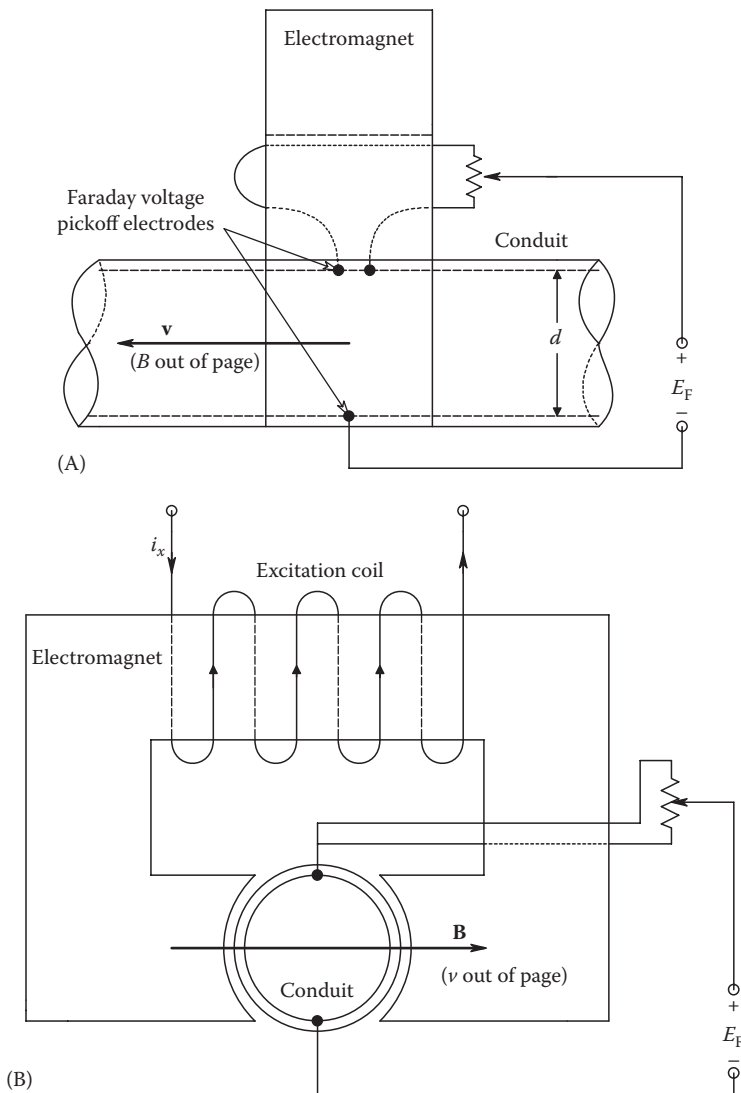
- d is measured in cm
- B in Gauss
- E_F in V

Often, the area of the conduit is not precisely known; hence, flow estimates using the Faraday method are not accurate, since E_F is proportional to the average velocity of the fluid.

In the Faraday flowmeter, the magnetic field is generally made to be an LF sinusoid in order that very small EFs can be better resolved. An AC E_o can generally be amplified above the amplifier's $1/f$ noise frequency band.

The Faraday flowmeter has been used to measure blood velocity and estimate blood flow. In one version, the electrodes make contact with the surface of the blood vessel, which is conductive itself. Due to geometrical errors in setting up the orthogonality between B , \bar{v} , and L in vivo, a net time-varying magnetic flux passes through the one-turn coil formed by the electrodes used to measure the Faraday EMF. Thus, the total output EMF seen across the electrodes will be the sum of the Faraday EMF and $d\Phi/dt = A(dB/dt)$ volts, which cuts the effective loop of area A formed by the E_F measurement circuit.

Webster (1978) described three methods of cancelling this unwanted $d\Phi/dt$ term. *One means* makes use of a cancellation circuit formed by making one of the electrodes a twin electrode, as shown schematically in Figure 6.30A. Each circuit will have a transformer

**FIGURE 6.30**

(A) Longitudinal section (side view) of a Faraday fluid velocimeter system using three sensing electrodes. The upper two electrodes permit cancellation of the unwanted $d\Phi$ (quadrature) term in the output voltage. (B) Axial (end-on) view of the three-electrode Faraday system.

error voltage; one of the form $E_{T1} = k_1 d\Phi/dt$, the other, $E_{T2} = -k_2 d\Phi/dt$. Thus, for some potentiometer wiper position, the two transformer EMFs will cancel, leaving the pure Faraday voltage. A *second means* of eliminating the effect of the transformer voltage is based on the fact that the transformer voltage is 90° out of phase (in quadrature) with the Faraday voltage (assuming sinusoidal excitation). Thus, if we use a phase-sensitive, sampling demodulator, we can sample the peak values of the Faraday voltage sine wave at the times the $d\Phi/dt$ sine wave is going through zero. The sampled Faraday voltage sine wave is held at sampling instants and low-pass filtered to obtain a DC signal proportional to \bar{v} . A *third means* of suppressing the unwanted transformer voltage makes use of a feedback system, shown in Figure 6.31. This system makes use of two PSRs and low-pass filters; one detects

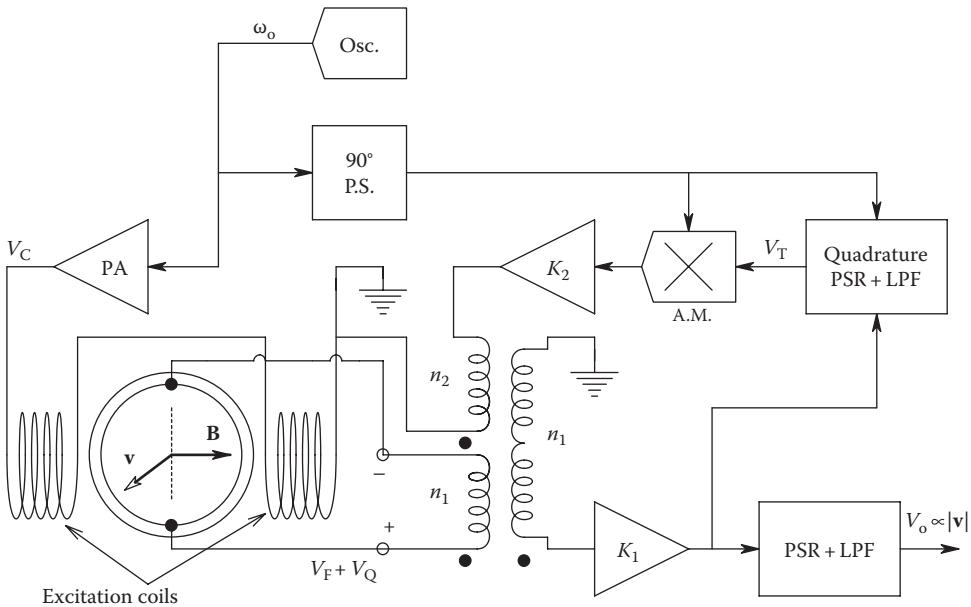


FIGURE 6.31 System to automatically null the unwanted quadrature output voltage from a Faraday liquid velocity sensor.

the desired Faraday voltage, the other, using a quadrature reference, detects the unwanted transformer voltage. A DC signal, V_T , proportional to the transformer voltage, acts through an analog multiplier to vary the level of a quadrature signal, V_{QF} , which is fed back through an auxiliary transformer winding, n_2 , to cancel the quadrature component, V_Q , which accompanies V_F . The quadrature detector loop acts to reduce the amount of transformer voltage contaminating the desired V_F signal, as shown in the following equation:

$$V_{QD} = \frac{V_Q}{[1 + (n_1/n_2)K_1K_{DQ}K_2(10)]} \tag{6.74}$$

If an integrator is put between the quadrature PSD/LPF and the analog multiplier, it is easy to show that there will be complete cancellation of V_{DQ} . Webster (1978) also described several other configurations of EM blood velocimeters using the Faraday streaming effect.

6.2.4.2 Hall Effect Sensors

Hall effect sensors, unlike Faraday effect devices, use no moving, conductive fluid. Instead, majority charge carriers (electrons or holes) have some average drift velocity as they traverse a bar of doped semiconductor. Figure 6.32 illustrates the geometry of a typical hall sensor. Note that as in the case of the Faraday transducer, the magnetic field vector \mathbf{B} , the drift velocity \mathbf{v} , and the voltage measuring axis are ideally orthogonal. Each hole or electron with velocity \mathbf{v} in the semiconductor will, in general, be subject to a Lorentz force given by the vector equation

$$\mathbf{F} = q(\mathbf{v} \times \mathbf{B}) = \{q|\mathbf{v}| |\mathbf{B}| \sin(\theta)\} \mathbf{u} \quad \text{N.} \tag{6.75}$$

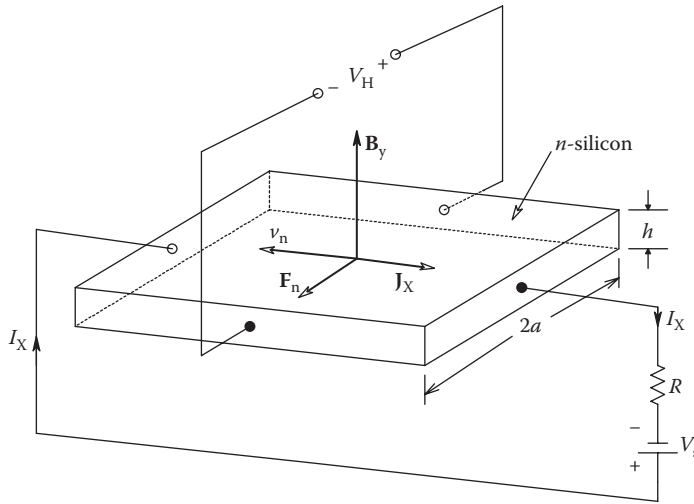


FIGURE 6.32

Schematic of a Hall sensor; an *n*-doped silicon slab is used. Note that the vectors J_x (current density), B_y (magnetic flux density sensed), and F_n (force on moving electrons) are orthogonal. v_{n-} is the electron velocity.

This force vector follows a right-hand screw rule in which its direction is found by rotating vector \mathbf{v} into \mathbf{B} and observing the unit vector direction, \mathbf{u} , that a screw with right-hand (normal) threads would advance, given that rotation. q is the charge of an electron (or hole). If electrons are the charge carriers, then we use a $-$ sign with q . As we see from Figure 6.32, the Lorentz force causes the moving electrons to crowd toward the front surface of the *n* semiconductor bar. An equilibrium density of electrons is reached when the average Lorentz force is balanced by the electrostatic force on the electron due to the electric field set up by their crowding. In other words, when

$$-qE_z = -qv_n B_y \quad \text{or} \quad E_z = v_n B_y \text{ V/m.} \tag{6.76}$$

Now, it is well known that in *n* semiconductor, the average drift velocity of electrons is given by

$$v_n = -\frac{J_x}{qn} \text{ m/s} \tag{6.77}$$

where

- J_x is the current density in the semicon bar in the $+x$ -direction
- n is the electron doping density

This relation for v_n can be substituted in Equation 6.77, giving

$$E_z = -\frac{J_x B_y}{qn} \text{ V/m.} \tag{6.78}$$

Now, if we consider the geometry of the semiconductor bar, $J_x = I/2ah$, and the Hall voltage developed across the two metallized faces of the bar is just $V_H = E_z 2a$ volts.

Substitution of these relations into Equation 6.78 yields the well-known expression for the Hall voltage, V_H :

$$V_H = -\frac{IB_y}{(qnh)} = \frac{R_H IB_y}{h} \quad (6.79)$$

where

h is the height of the semicon bar (in the y -direction)

R_H is the Hall coefficient; $R_H = -1/qn$ for n-doped semicon and $R_H = +1/qn$ for p semiconductor

Because the Lorentz force has the opposite sign for p semiconductor, all other vectors having the same direction, the Hall voltage for p material is $V_H = +IB_y/(qnh)$.

Hall sensors have many uses, most of which can be subdivided into either analog or switching applications. As analog sensors, Hall devices can be used to produce an output proportional to $|\mathbf{B}|\cos(\theta)$, where θ is the angle of the \mathbf{B} vector with the transducer's y -axis (axis of maximum sensitivity). Some commercial, IC, analog Hall sensors have built-in DC amplifiers to condition the V_H signal. Analog Hall sensors were listed by Sprague Electric Co. (1987) that have sensitivities of about 1.4 mV/G with HF bandwidths of 20–25 kHz (on \mathbf{B}). Other Hall sensors listed by F.W. Bell do not contain built-in amplifiers and claim sensitivities ranging from 10 to 55 mV/kG. Bell also offers a Model BH-850 high-sensitivity Hall sensor that uses a 9 in., high-permeability bar to concentrate the magnetic flux. This sensor claims 18 mV/G, which is adequate sensitivity to measure the Earth's magnetic field and thus has applications as a solid-state compass. Allegro Microsystems, Inc., makes a family of Hall effect switches and chopper-stabilized Hall effect switches. Optek Technologies, Inc., Carrollton, TX, makes *OH-series*, Hallogenic™ Hall effect switches (2006).

One obvious application of analog Hall sensors is the measurement of magnetic fields. They also can be used to measure other physical or electrical parameters that can be made to be proportional to magnetic fields (AC or DC). A less obvious application of an analog Hall sensor is as an audio-frequency wattmeter. The current through a reactive load is passed through a solenoid coil to produce a \mathbf{B}_y proportional to I . The voltage across the load is conditioned and used to generate a current through the Hall sensor proportional to V . The phase between the I and V sinusoids is θ . It is easy to show that the DC component of the Hall voltage is proportional to the average power in the load; that is, it is given by

$$\bar{V}_H = kP_{AVE} = kVI \cos(\theta). \quad (6.80)$$

where

k is a proportionality constant

V is the RMS voltage across the load

I is the RMS current through the load

$\cos(\theta)$ is the power factor

Switching Hall sensors have many applications in counting and proximity (of a permanent magnet) detection. Their output is generally a TTL logic signal. An IC DC amplifier, Schmitt trigger, and open-collector logic output are mounted in the same package with the Hall chip. Hall switches have uses in automobile electronic ignitions, brushless DC motors (rotor position sensing), tachometers, keyboards, thermostats, pressure and temperature alarms, burglar alarms, etc.

6.2.5 Sensors Based on Variable Magnetic Coupling

This class of sensors operates with AC excitation. Their outputs vary according to the degree of magnetic coupling between the excitation coil(s) and the output winding(s). Most sensors that operate on the principle of variable magnetic coupling between coils are mechanotransducers, that is, they can be used to measure or generate linear or rotational displacements. Because many other physical phenomena can be converted to small linear rotations or displacements, this class of transducer also is the basis for certain pressure, temperature, and acceleration sensors. Lion's Sec. 1–22 (1959) gives a good summary of this class of transducer, which includes the LVDT and the synchro transformer.

6.2.5.1 LVDT

The LVDT is one of the most widely used mechanical input sensors in modern instrumentation practice. LVDTs are used to measure linear mechanical displacement or position in control systems and in precision manufacturing gauging and can be used indirectly to measure force, pressure, acceleration, etc., or any quantity that can be made to cause a linear displacement. Figure 6.33A illustrates a section through an LVDT showing the central excitation coil; two, symmetrically located pickoff coils; and the hollow, cylindrical core tube through which slides the high-permeability magnetic core. The position of the core in the center tube relative to the excitation coil and the two pickoff coils changes the mutual inductance between the excitation coil and the pickoff coils, varying the induced voltages in the two pickoff coils. As one voltage drops, the other rises, etc.

A schematic representation of an LVDT is shown in Figure 6.33B. Note that the two, series pickoff coils have opposite dot polarity, so that when the moveable magnetic core is at magnetic center ($x = 0$), the induced EMFs in each coil sum to zero at the output. The AC excitation of an LVDT can range in frequency from about 50 Hz to 20 kHz. Peak output response for a given displacement, $\delta x > 0$, is maximum at some frequency that may range from 400 Hz to over 2 kHz. Some LVDTs contain built-in oscillators and PSRs so one has only to supply a DC power source (e.g., 24 V) and observe a DC output that has a voltage/core-position relationship similar to that shown in Figure 6.34. Several electronic IC manufacturers offer oscillator/PSR chips (Signetics' NE5520, Analog Devices' AD630) to facilitate the use of LVDTs, which do not contain this signal processing circuit internally.

LVDTs are not zero-force sensors. If their cores experience a lateral force, there will be increased friction with the inside of the core tube. In addition to friction forces, the solenoidal magnetic field from the excitation coil acts to pull the core toward the center ($x = 0$) position. The solenoid or axial force is proportional to the square of the excitation current. Typical peak axial force is less than 1 g, which is negligible in most position-sensing applications.

Specifications for LVDTs vary widely according to application. Those used for gauging applications (e.g., measuring the contour of a cam) have infinite resolution, linearity of $\pm 0.2\%$ of full range and repeatability as small as 4 micrometers. Gauge LVDTs generally have some sort of special tip to contact the work, such as a hardened, tungsten carbide ball, and also a helical spring to maintain a constant static contact force with the work. Other versions of LVDT gauges use air pressure to obtain a variable gauge contact force that can be set for delicate (soft) materials. The former Schaevitz Company (now Sherborne Sensors, Wyckoff, NJ) PCA-230 series precision, pneumatic, LVDT gauges had gauging ranges from $\pm 0.005''$ to $\pm 0.100''$. Other models had ranges up to $\pm 1.000''$, and some had built-in oscillator/demodulator ICs for DC operation.

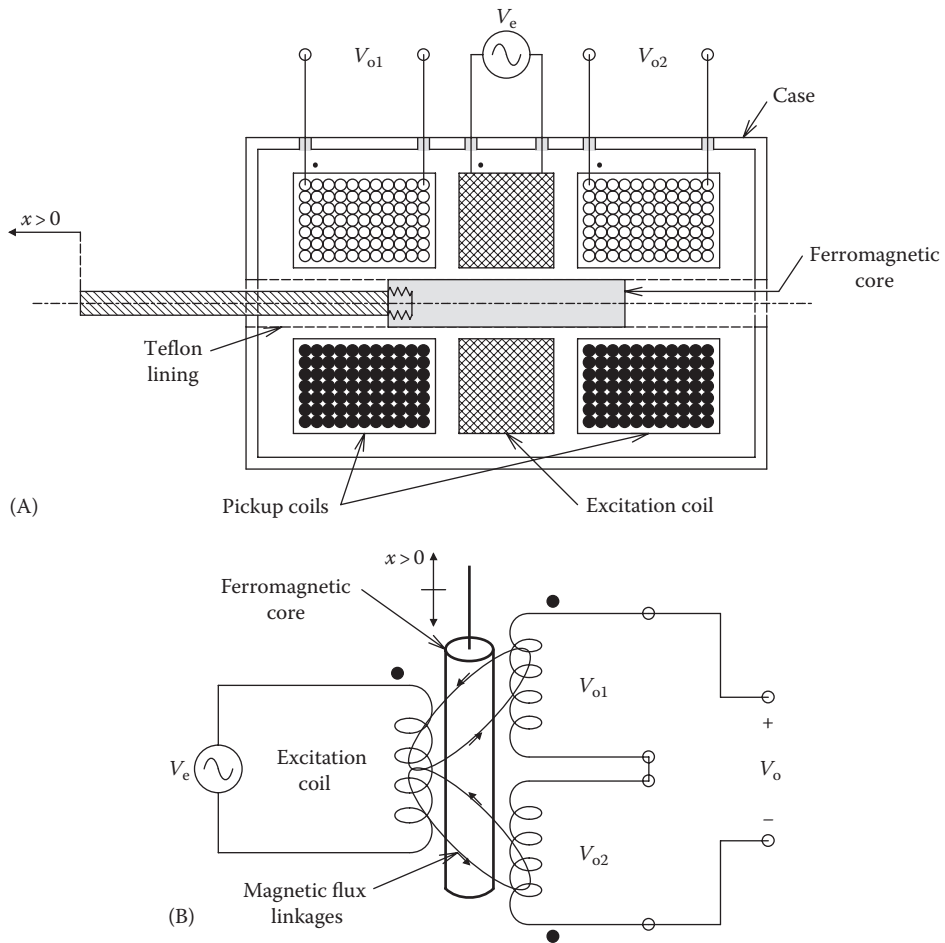


FIGURE 6.33 (A) Longitudinal cross section through a typical LVDT. (B) LVDT schematic. Note the dot notation on the output coils.

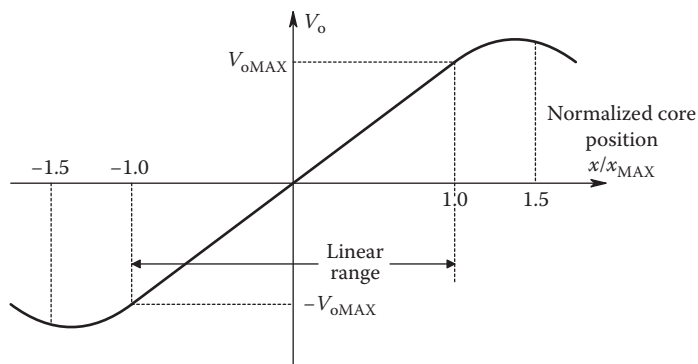


FIGURE 6.34 Typical output voltage vs. linear core displacement for an LVDT.

6.2.5.2 Synchros and Resolvers

We first consider *synchros*, also called *selsyns* or *autosyns*. These are angular position sensors that work on the principle of variable mutual inductance. They were perfected during World War II and find limited use in twenty-first century measurement system designs. They generally have multiple windings. There are three main classes of synchros: *transmitters*, *repeaters*, and *control transformers*. Figure 6.35A is a schematic diagram of a synchro. Rotor excitation is generally at 60 or 400 Hz. The three stator windings are arranged at 120° spacings, so the induced stator voltages may be written as follows:

$$v_1(t) = V_{pk} \sin(2\pi ft) \sin(\theta) \quad (6.81a)$$

$$v_2(t) = V_{pk} \sin(2\pi ft) \sin(180^\circ + \theta) \quad (6.81b)$$

$$v_3(t) = V_{pk} \sin(2\pi ft) \sin(240^\circ + \theta). \quad (6.81c)$$

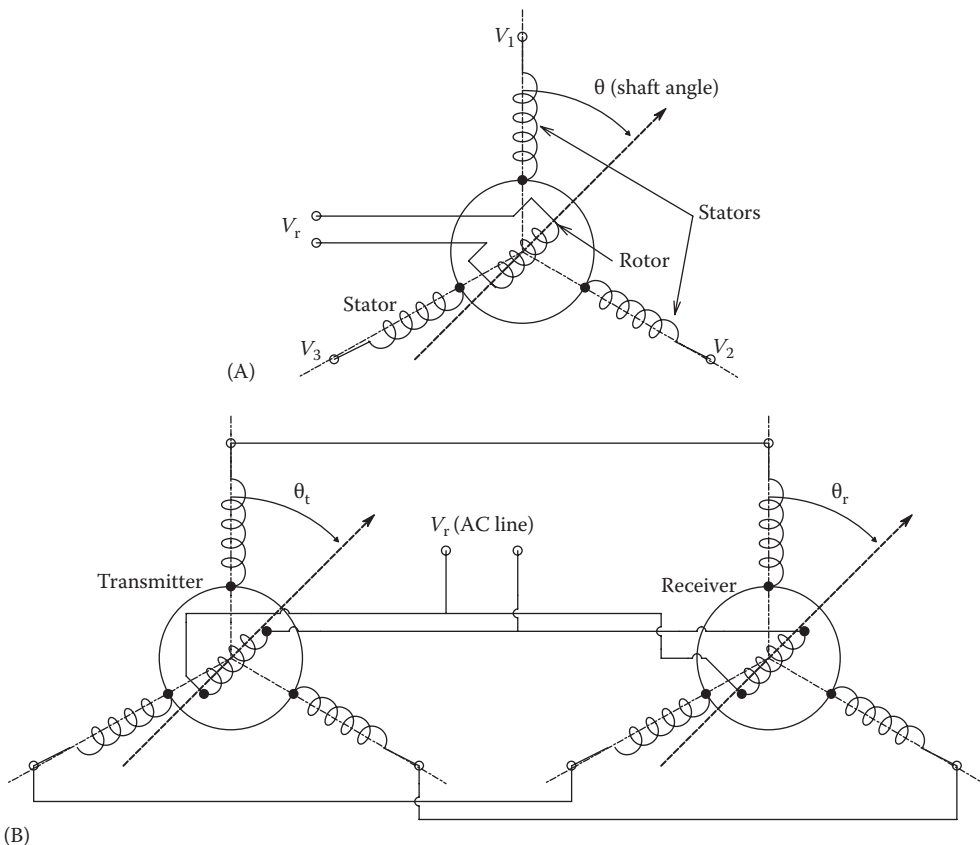


FIGURE 6.35

(A) Schematic diagram of a synchro transmitter. (B) Electrical connections between a synchro transmitter and repeater (receiver) enable the repeater rotor to follow the angle of the transmitter's rotor. No electronics are involved.

Synchros were first widely used in the 1940s in military systems as angular position sensors and components in servomechanisms used to train guns or gun turrets, sonar arrays, or radar antennas. Synchros come in all sizes, ranging from the massive US Navy 7G developed in World War II to modern, miniature synchros measuring 1" in diameter by 1.8" in length. Synchro accuracies range from $\pm 0.1^\circ$ to about $\pm 1.0^\circ$.

The simplest application of two synchros is signaling a shaft rotation angle to a remote location. In this application, two synchros are used, a transmitter and a repeater, shown in Figure 6.35B. The rotors of the transmitter and repeater are connected in parallel to the AC line, and corresponding stator lines are connected. The rotor of the transmitter is at some angle θ_t . The rotor of the repeater is initially at some angle θ_r and is free to rotate. The repeater rotor experiences a torque causing it to rotate so that $\theta_r = \theta_t$. The repeater rotor experiences zero torque when $\theta_r = \theta_t$. The repeater rotor generally has a light external inertial load, such as a pointer, and some viscous damping to prevent overshoots and oscillations when following sudden changes in the transmitter's θ_t .

Another major application of synchros is error signal generation in carrier-operated, angular position control systems (see Figure 6.36). A synchro transmitter and a synchro control transformer are used to generate a carrier frequency signal whose amplitude is proportional to $\sin(\theta_t - \theta_c)$, where θ_t is the rotor angle of the transmitter and θ_c is the rotor angle of the control transformer. If we consider θ_t to be the input, reference, or command signal and θ_c to be the output of a servomechanism designed to follow θ_t , then the output of the synchro transformer may be considered to be the servo system error for $\theta_e = (\theta_t - \theta_c) < 15^\circ$, where $\sin(x) \approx x$ in radians. The synchro transformer output can be shown to be a double-side-band, suppressed-carrier modulated signal. When demodulated by a PSR, the output will be a voltage proportional to $\sin(\theta_t - \theta_c)$. Obviously, the synchro transmitter and transformer can be located at some distance from one another, a distinct advantage.

Modern practice, however, is to replace the electromechanical synchro transformer with a commercially available, solid-state circuit system into which the input angle is read as a binary number from a coding disk, the output of which is displayed digitally, or in analog form, the same as would be seen from the rotor of a conventional synchro transformer. In spite of such solid-state conveniences, nonelectronic, electrical synchro systems are easy to work with; however, they lack the precision of modern, digital position-sensing systems. Modern angular position measurement systems generally make use of optical coding disks and directly code shaft angle into digital form for transmission to the control computer.

Resolvers are another rotational coordinate sensor. As shown schematically in Figure 6.37, resolvers differ from synchros in that they have two orthogonal stator windings and two orthogonal rotor windings. The stator windings are generally energized so that

$$v_{s1}(t) = V_{s1} \sin(\omega t) \quad (6.82a)$$

$$v_{s2}(t) = V_{s2} \sin(\omega t). \quad (6.82b)$$

Noting the dot convention shown in Figure 6.37, the induced output (rotor) voltages are

$$v_{r1}(t) = v_{s1}(t) \cos(\theta) - v_{s2}(t) \sin(\theta) \quad (6.83a)$$

$$v_{r2}(t) = -v_{s1}(t) \sin(\theta) + v_{s2}(t) \cos(\theta) \quad (6.83b)$$

where θ is the rotor angle with respect to the reference shown.

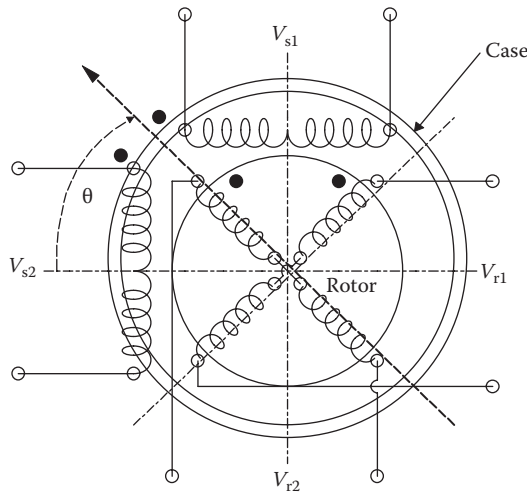


FIGURE 6.37
Schematic of a resolver.

Capacitive sensors have been designed to measure force by means of displacement of one or two capacitive electrodes. They also can be used to measure acceleration, thickness, depth of a dielectric liquid, and pressure. In its simplest form, the mechanical input quantity causes a change in the separation between two (or more) capacitor plates, as described in the following.

The capacitance between two, parallel, conducting plates, each of area A , separated by a dielectric of thickness d is given by (neglecting fringing effects)

$$C = \frac{\kappa \epsilon_0 A}{d} \tag{6.84}$$

where

- κ is the dielectric constant of the material separating the plates
- ϵ_0 is the permittivity of free space (8.85×10^{-12} F/m)

Small variations in C , ΔC , which may be measured by several means, may occur if the plate separation is caused to change by some physical quantity or if the dielectric constant, κ , changes due to pressure, temperature, mechanical strain, humidity, etc. The overall dielectric constant may also be modulated by sliding a substance with a dielectric constant greater than one between the capacitor plates so that some fraction, α , of A has a dielectric constant of $\kappa = \kappa_1$ and the remaining area between the plates, $(1 - \alpha)A$, has a dielectric constant $\kappa = 1$. The area of the capacitor plates may be effectively changed by sliding one plate over the other at constant d . An example of this latter means of effecting a ΔC is found in parallel-plate, rotary tuning capacitors used in radios.

Means of measurement of capacitance changes include various capacitance bridges or use of the capacitor in a tuned LC circuit determining the frequency of an oscillator; the oscillator's frequency is measured to obtain a signal proportional to $1/\sqrt{LC}$.

One of the most sensitive means of producing an output voltage proportional to ΔC is through the use of a de Sauty bridge, shown in Figure 5.7 and discussed in Section 5.5.1.4.

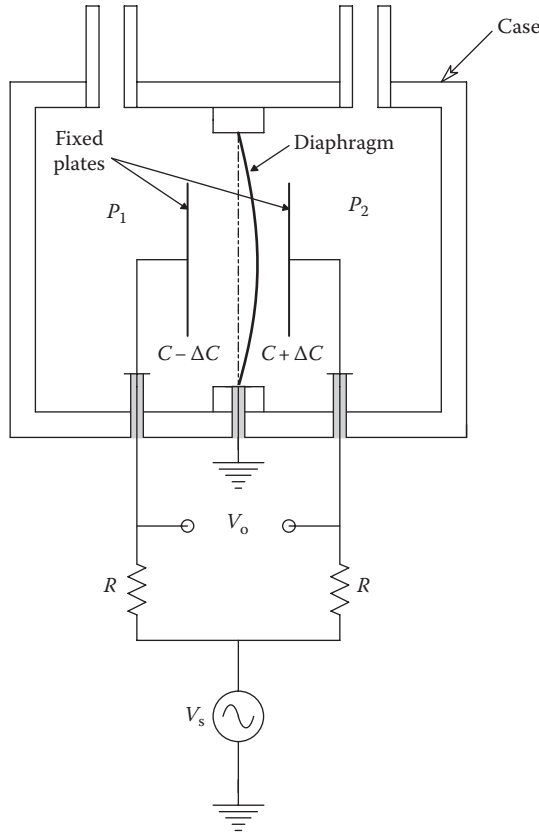


FIGURE 6.38

Section through a differential pressure transducer that uses the small change in capacitance caused by the deflection of the diaphragm to produce an AC output signal proportional to $(P_1 - P_2)$. A de Sauty bridge is used.

A capacitive pressure sensor in which a diaphragm deflects as the result of the pressure differential across it is shown in Figure 6.38. The diaphragm is grounded and thus forms a common plate with two capacitors such that $C_1 = C_o - \Delta C$, $C_2 = C_o + \Delta C$, and $\Delta C = k(P_1 - P_2)$. If a de Sauty bridge is operated at a frequency $f_o = 1/2\pi RC_o$ Hz, with peak excitation voltage V_s , it is easy to show that

$$\frac{V_o}{V_s}(s) = \frac{1/s(C_o - \Delta C)}{1/s(C_o - \Delta C) + R} - \frac{1/s(C_o + \Delta C)}{1/s(C_o + \Delta C) + R}. \tag{6.85}$$

After some algebra, we can write

$$\frac{V_o}{V_s}(s) = \frac{s2R\Delta C}{(1 + sRC_o)^2 - s^2R^2C^2(\Delta C/C_o)^2}. \tag{6.86}$$

Letting $s = j\omega = j/RC_o$, and noting that $(\Delta C/C_o)^2 \ll 1$, we obtain finally

$$\frac{V_o}{V_s}(j\omega) = \frac{\Delta C}{C_o} \tag{6.87}$$

or

$$v_o(t) = V_s \left(\frac{\Delta C}{C_o} \right) \sin \left(\frac{t}{RC_o} \right). \quad (6.88)$$

The bridge output, V_o , in this example is a double-sideband, suppressed-carrier signal whose amplitude is $V_o = V_s(\Delta C/C_o)$. Since it is possible to resolve a V_o as small as $1 \mu\text{V}$, when V_s is typically 5 V pk , and C_o is 100 pF , ΔC can be as small as $2 \times 10^{-4} \text{ pF}$!

Pressure measurements can also be made using a capacitor with fixed geometry but with a dielectric constant that changes with pressure. Certain nonpolar liquids, gasses, and piezoelectric materials can be used for this purpose. Lion (1959) reported that the dielectric constant of air at 19°C varies from 1.0006 at 1 atm to 1.0548 at 100 atm . Unfortunately, dielectric constants also change with temperature, so a null transducer can be used in a bridge circuit to obtain temperature compensation.

Dranetz, Howatt, and Crownover, cited by Lion, developed a titanate ceramic called Thermanon™, in which the dielectric constant decreased linearly over the temperature range of -40°C to $+160^\circ\text{C}$. One specimen was described, which had a dielectric constant of 490 at room temperature and a dielectric constant tempco of $-0.003/^\circ\text{C}$. However, the tempco increases dramatically as the Curie temperature is approached.

6.3 Fiber-Optic Sensors

Fiber-optic (FO) dielectric waveguides were developed primarily for broadband, long distance communications links. Their main advantages are low cost (silicon vs. copper), broad signal bandwidths, and immunity from interference caused by EM radiation, such as radio waves, EM pulses (EMPs) generated by lightning, or, perish the thought, nuclear explosions. They also may be used in harsh environments remote from their electro-optic signal conditioning systems.

Interestingly, FOCs are also useful in a variety of sensor applications (Wolfbeis 1991). They have been used in the measurement of electric current, magnetic fields, temperature, force, pressure, strain, acceleration, and pH. Several of their physical/optical properties have been used in realizing these applications. For example, (micro)bending of an optical fiber will result in the loss of light energy from the fiber core into the cladding at the bend, decreasing the output light intensity. This phenomenon is used in sensing force, pressure, strain, etc. (Hochberg 1986). pH may be sensed by measuring the emission of a fluorescent dye coating the end of the fiber; the emission characteristic of the dye changes with pH. pH has also been measured spectrophotometrically with FOCs; a pH-sensitive indicator dye such as phenol red is immobilized inside a cuprophan membrane at the end of a pair of FOCs. Light at two wavelengths is transmitted to the dye. Reflectance at one wavelength is pH dependent; reflectance at the other wavelength is independent of pH and is used as a reference. Reflected light at both wavelengths was collected by the output FOC and processed. System output was linear with pH over a limited range (Peterson and Goldstein 1982).

Another effect that is used in FO-based sensors is the modulation of the fiber's optical polarization properties by mechanical strain, temperature changes, or magnetic fields.

FOCs are also used in Sagnac-effect, FO, laser gyroscopes. Their use in FO gyros is expeditious, however, and it should be noted that the Sagnac effect will occur whether the light travels in a closed path in air directed by mirrors or in a multiturn FO coil. Sagnac-effect gyros are discussed in detail in Section 7.2.2.2.

6.3.1 Magneto-Optic Current Sensors

One of the major problems in measuring the electrical current in high-voltage, high-power transmission lines is that of measurement circuit isolation. One means of measuring heavy 60 Hz currents in a conductor is through the use of a current transformer. This transformer consists of a number of turns of wire wound on a high-permeability, toroidal magnetic core. The current-carrying conductor is passed through the *hole* in the toroid, and its magnetic field induces a voltage in the toroid's winding proportional to the current in the conductor. This induced voltage is measured with an AC millivoltmeter. This means of measuring heavy AC works fine at lower voltages on the conductor. However, when the voltage on the conductor reaches values too high to safely isolate the toroid's winding (and meter circuit) from the conductor, some other means must be used to insure safety.

The *Faraday magneto-optic effect* offers one means to measure high currents on conductors at extremely high voltages aboveground (50 kV and higher) with reasonable accuracy and excellent isolation. Magneto-optic current sensors (MOCSs) that make use of the Faraday magneto-optic effect can use either solid glass pathways or FO waveguides. When linearly polarized light (LPL) is passed through a transparent, diamagnetic material through which a magnetic field is also passed in the same direction, there will be, in general, a rotation of the polarization vector of the emergent ray. It may be shown that the polarization rotation angle is given by

$$\alpha = V \int \mathbf{H} \cdot d\mathbf{l}. \quad (6.89)$$

The rotation, α , is proportional to the line integral of the magnetic field intensity vector, \mathbf{H} , along the light propagation path, \mathbf{l} , which encloses the magnetic field. V is the *Verdet constant* for the material used.

Note that by the Ampere circuital law,

$$\int \mathbf{H} \cdot d\mathbf{l} = I, \quad (6.90)$$

where I is the current enclosed by one optical path loop.

Hence, if N loops of a FOC are wound on a form so that all turns are aligned and have the same area, we may express the Faraday magneto-optic polarization rotation as

$$\alpha = VNI. \quad (6.91)$$

Figure 6.39 illustrates one version of a Faraday current transducer consisting of N turns of single-mode optical fiber wound around a current-carrying conductor. The optical rotation of the LPL entering the fiber is given by Equation 6.90. There are several means of sensing small changes in the optical rotation of LPL, which are discussed in the following. The optical rotation, in theory, will remain unchanged regardless of the position of the current-carrying conductor in the FO coil. Also, by using a diamagnetic material such as glass fiber, the system has substantial temperature independence (Rogers 1973).

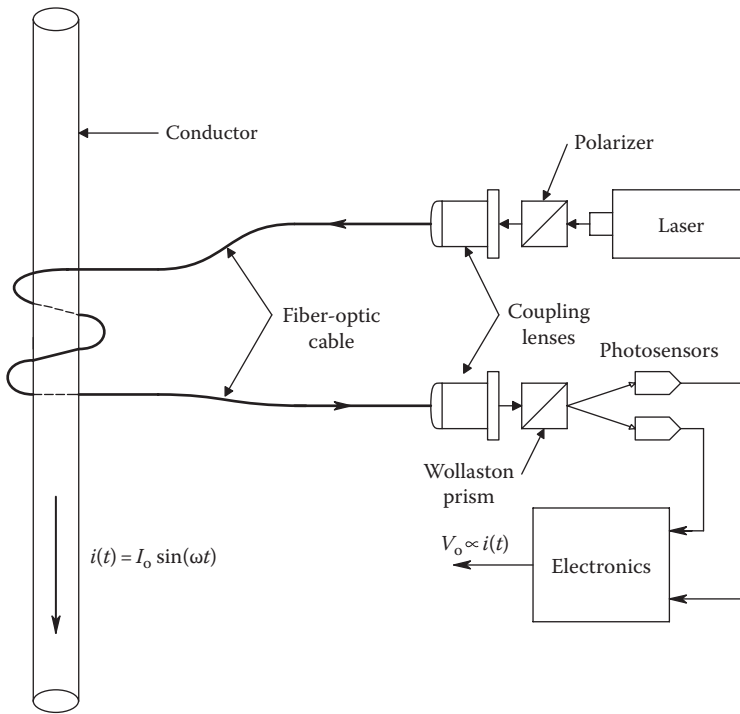


FIGURE 6.39

An FO, electrical current sensor using the Faraday magneto-optic effect. The WP separates the magneto-optically rotated E vector of the emergent LPL into two orthogonal components. The two components are detected by photodiodes and their signals are processed by a ratio detector.

Another version of the Faraday MOCS developed by Cease and Johnston (1990) is shown in Figure 6.40. Here, a single block of double extra dense flint (DEDF) glass is used. It acts both as the Faraday medium and as an electrical insulator for the bus bar. For this geometry, the optical rotation angle is given by

$$\alpha = VBL \tag{6.92}$$

where

V is the SI Verdet constant

B is the magnetic flux density in Teslas

L is the optical path length over which the Faraday effect occurs

Hecht (1987) gave a table of Verdet constants. From this table, we see that V for quartz at 20°C and $\lambda = 578 \text{ nm}$ is 1.66×10^{-2} minutes of arc/(gauss cm), V for light flint glass at 18°C and $\lambda = 578 \text{ nm}$ is 77.9×10^{-3} minutes of arc/(gauss cm), and V for pure water at 20°C and $\lambda = 578 \text{ nm}$ is 1.31×10^{-2} minutes of arc/(gauss cm).

The magnetic field magnitude, B , around a current-carrying conductor at a distance R , where $R \gg r$ and r is the conductor's radius, may be shown by the *Biot-Savart law* to be given by

$$B = \frac{\mu_0 I}{(2\pi R)} \text{ W/m}^2 \text{ (in SI units)}. \tag{6.93}$$

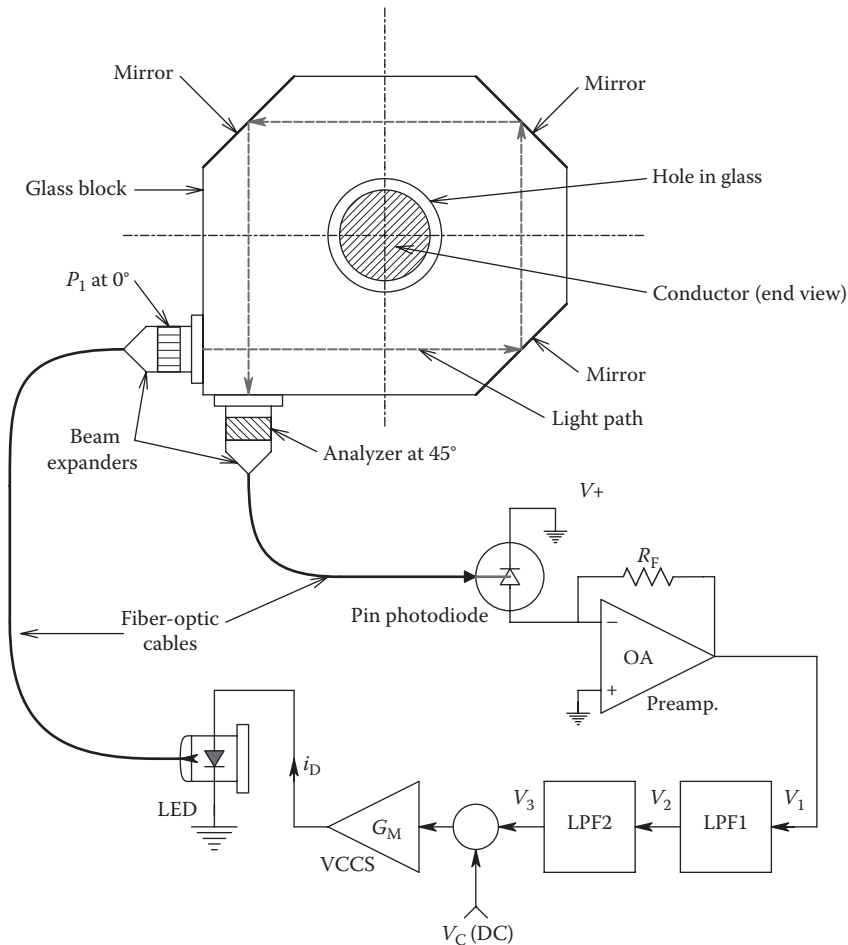


FIGURE 6.40
Diagram of another Faraday MOCS devised by Cease and Johnston (1990).

(B in gauss equals 10^4 w/m².) The vector \mathbf{B} is everywhere tangential to a circle of radius R centered on the conductor. Such a magnetic field is called solenoidal, and if the light path consists of N turns of an optical fiber wound on a circular form of radius R , concentric with the conductor, then the net Faraday rotation is given by

$$\alpha = V' \left[\frac{\mu_0 I}{(2\pi R)} \right] (2\pi RN)(10^6) = V' \mu_0 IN(10^6) \text{ minutes of arc.} \tag{6.94}$$

For example, a MOCS has 100 turns of a lead glass fiber with a Verdet constant $V' = 80 \times 10^{-3}$ minutes arc/(gauss cm), and a current of 1 A will produce a B-field, which gives a Faraday rotation of 10.055 min or 0.1676°/amp. Note that the constant, 10^6 , converts the Verdet constant, V' in min/(gauss cm), to the SI V in units min/(W/m² m); $\mu_0 = 1.257 \times 10^{-6}$ W/(amp m).

As you will see in the next section, it is entirely possible to measure small angles of optical rotation to better than $\pm 0.001^\circ$. This polarimeter resolution sets the ultimate resolution of a FO MOCS.

6.3.2 Means of Measuring the Optical Rotation of the Linearly Polarized Light Output of Certain Optical Sensors

In order to consider the measurement of the rotation angle of LPL, we must first review the mathematical description of light as a propagating transverse EM wave in a dielectric medium. As is well known, the propagation of light can be described by the Maxwell equations. In the simplest case, we will assume a plane wave propagating in the z -direction in a source-free, isotropic medium with permittivity, ϵ , and permeability, μ . From the Maxwell equations, we have

$$\nabla \times \mathbf{E} = -\frac{d\mathbf{B}}{dt} \quad (6.95a)$$

$$\nabla \times \mathbf{B} = (\mu\epsilon) \frac{d\mathbf{E}}{dt} \quad (6.95b)$$

$$\nabla \cdot \mathbf{E} = 0 \quad (6.95c)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (6.95d)$$

The second-order wave equation may be derived from the Maxwell vector equations, (6.95):

$$\nabla^2 \mathbf{E} - (\mu\epsilon) \frac{d^2 \mathbf{E}}{dt^2} = 0. \quad (6.96)$$

Assuming sinusoidal SS conditions, we can write

$$\nabla^2 \mathbf{E} + \mathbf{k}^2 \mathbf{E} = 0 \quad (6.97)$$

where k is the wavenumber defined as

$$k \equiv \frac{2\pi f}{c} = \frac{2\pi}{\lambda} = 2\pi f (\mu\epsilon)^{1/2} = \frac{\omega}{c}. \quad (6.98)$$

Note that in an anisotropic medium, k can be expressed as a vector

$$\mathbf{k} = \mathbf{a}_x k_x + \mathbf{a}_y k_y + \mathbf{a}_z k_z. \quad (6.99)$$

Here, we use \mathbf{a}_x as a unit vector pointing along the positive x -axis and k_x is the wavenumber in the x -direction, \mathbf{a}_y is a unit vector pointing along the positive y -axis, and k_y is the wavenumber in the y -direction, etc. The anisotropy may be considered to be due to different propagation velocities in the x -, y -, and z -directions (e.g., $c_x = 1/(\epsilon_x \mu_x)^{1/2}$).

We note that the \mathbf{B} - and \mathbf{E} -field vectors are mutually perpendicular to each other and to the direction of wave propagation, which we will take as the positive z -axis.

The *polarization* of an EM wave is defined in accordance with the IEEE standard (Balanis 1989). The polarization of a radiated EM wave is defined as "that property of a radiated electromagnetic wave describing the time-varying direction and relative magnitude of the electric field vector; specifically, the figure traced as a function of time by the extremity

of the vector at a fixed location in space, and the sense in which it is traced, as observed along the direction of propagation." There are three categories of EM wave polarization: *linear*, *circular*, and *elliptical*. If the \mathbf{E} vector at a point in space as a function of time is always directed along a line that is normal to the direction of wave propagation, the field is said to be *linearly polarized*. To illustrate this property mathematically, we assume that the \mathbf{E} vector has x and y components and may be written as follows:

$$\begin{aligned}\mathbf{E} &= \mathbf{a}_x E_x + \mathbf{a}_y E_y = \mathbf{Re} \left[\mathbf{a}_x E_x^+ \exp(j\{\omega t - kz\}) + \mathbf{a}_y E_y^+ \exp(j\{\omega t - kz\}) \right] \\ &= \mathbf{a}_x E_{xr} \cos(\omega t - kz + \phi_x) + \mathbf{a}_y E_{yr} \cos(\omega t - kz + \phi_y)\end{aligned}\quad (6.100)$$

where

E_x^+ and E_y^+ are complex
 E_{rx} and E_{yr} are real

Let us consider two cases of linear polarization: First, let $E_{yr} \equiv 0$ and $z = 0$ for convenience. This means that the locus of the instantaneous \mathbf{E} -field vector is a straight line that is directed along the x -axis. The \mathbf{E} field is given by

$$\mathbf{E} = \mathbf{a}_x E_{xr} \cos(\omega t + \phi_x). \quad (6.101)$$

In a more general case of linear polarization, we let $\phi_x = \phi_y = \phi$, $z = 0$, and we find that the \mathbf{E} vector is given by

$$\mathbf{E} = \left[\sqrt{(E_{xr})^2 + (E_{yr})^2} \right] \cos(\omega t + \phi). \quad (6.102)$$

This gives a straight line at an angle α with the x -axis for all time, as shown in Figure 6.41. The polarization angle α can be written as follows:

$$\alpha = \tan^{-1} \left(\frac{E_{yr}}{E_{xr}} \right). \quad (6.103)$$

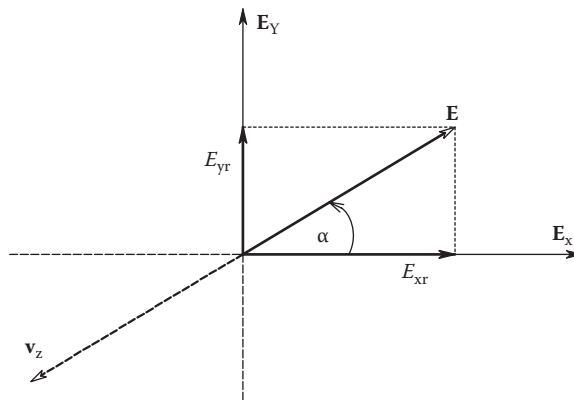


FIGURE 6.41
 \mathbf{E} vector of light linearly polarized along the α -direction.

Circular polarization is said to occur when the tip of the \mathbf{E} vector traces a circular locus in the XY plane when viewed by an observer standing at the origin looking in the positive z -direction. (The EM wave is assumed to propagate in the $+z$ -direction.) Circular polarization can be either clockwise (right handed) or counterclockwise (left handed), *as viewed looking from the rear of the EM wave in the $+z$ -direction*. It may be shown that circular polarization will occur when the \mathbf{E} field has two, orthogonal, linearly polarized components with equal magnitudes and the two orthogonal components must have a time-phase difference of odd multiples of 90° . As an example of circular polarization, let us assume in Equation 6.100 that $\phi_x = 0$, $\phi_y = -\pi/2$, $z = 0$, and $E_{xr} = E_{yr} = E_{CW}$. Thus,

$$\mathbf{E}_x = \mathbf{a}_x E_{CW} \cos(\omega t) \quad (6.104)$$

$$\mathbf{E}_y = \mathbf{a}_y E_{CW} \cos\left(\omega t - \frac{\pi}{2}\right) = \mathbf{a}_y E_{CW} \sin(\omega t). \quad (6.105)$$

The locus of the amplitude of the \mathbf{E} vector in the XY plane is just

$$|\mathbf{E}| = \sqrt{E_x^2 + E_y^2} = \sqrt{E_{CW}^2 [\cos^2(\omega t) + \sin^2(\omega t)]} = E_{CW}. \quad (6.106)$$

At any instant of time, \mathbf{E} is directed along a line making an angle α with the x -axis; so

$$\alpha = \tan^{-1}\left(\frac{E_{CW} \sin(\omega t)}{E_{CW} \cos(\omega t)}\right) = \tan^{-1}[\tan(\omega t)] = \omega t. \quad (6.107)$$

Thus, the angle α rotates clockwise with angular velocity of ω . The right-hand, circularly polarized \mathbf{E} vector can also be written (Balanis 1989) as follows:

$$\mathbf{E} = E_{CW} \mathbf{Re}\{(\mathbf{a}_x - j\mathbf{a}_y) \exp[j(\omega t - kz)]\}. \quad (6.108)$$

Using the development given earlier, it is easy to demonstrate that a counterclockwise (left hand) rotating circular polarization vector will occur when $\phi_x = 0$, $\phi_y = \pi/2$, and $E_{xr} = E_{yr} = E_{CCW}$. In this case, $E = E_{CCW}$ and $\alpha = -\omega t$. The \mathbf{E} vector may be written as follows:

$$\mathbf{E} = E_{CCW} \mathbf{Re}\{(\mathbf{a}_x + j\mathbf{a}_y) \exp[j(\omega t - kz)]\}. \quad (6.109)$$

It may be shown that LPL can be resolved into right- and left-hand circularly polarized components. This can be demonstrated physically by passing a beam of LPL through a *quarter-wave plate* optical device. In an isotropic medium, both right- and left-hand circular components of linear polarized light rotate at the same angular velocity. In an optically active, *anisotropic medium*, the velocity of light (in the z -direction) is different for left- and right-hand circularly polarized waves. This difference in the propagation velocity of the left- and right-hand circularly polarized waves is referred to as *circular birefringence*.

Coté et al. (1990) showed that in an optically active, circularly birefringent medium, the net rotation of LPL exiting the medium is

$$\alpha = \left(\frac{\pi z_0}{\lambda} \right) (n_R - n_L) \text{ rad} \quad (6.110)$$

where

z_0 is the distance the LPL propagates along the z -axis through the circularly birefringent medium

n_R and n_L are the indices of refraction in the medium for right- and left-hand circularly polarized light, respectively

(The index of refraction of a medium is defined as the ratio of the speed of light in vacuum to the speed of light in the medium; it is generally greater than unity.)

An EM wave is elliptically polarized if the tip of the E vector traces out an ellipse in the XY plane as the wave propagates in the $+z$ -direction. As in the case of circular polarization, ellipses traced clockwise are right handed, and left-handed ellipses are traversed counterclockwise. The treatment of elliptical polarization is beyond the scope of this text. The interested reader who wishes to pursue this topic should consult Section 4.4 in Balanis (1989).

Polarimeters are optoelectronic systems designed to measure the optical rotation produced by an active medium. As we have seen earlier, the medium can be a piece of glass, an optical fiber, or a solution containing optically active molecules (such as D -glucose).

The first and most basic type of electronic polarimeter design is shown in Figure 6.42. Monochromatic light from a source, S , is collimated and then linearly polarized at a reference angle of 0° by an *input polarizer*. The linearly polarized beam is then passed through the optically active medium, M , of length L . The E vector of the emergent beam is still linearly polarized but is rotated some angle α . The emergent beam is then passed through an *analyzer polarizer* whose pass axis is at 90° to the input polarizer. If $\alpha = 0$, then the detector, D , will sense zero light intensity. If $\alpha > 0$, the light intensity will increase according to the relation

$$I = \left(\frac{I_{MAX}}{2} \right) [1 - \cos(2\alpha)]. \quad (6.111)$$

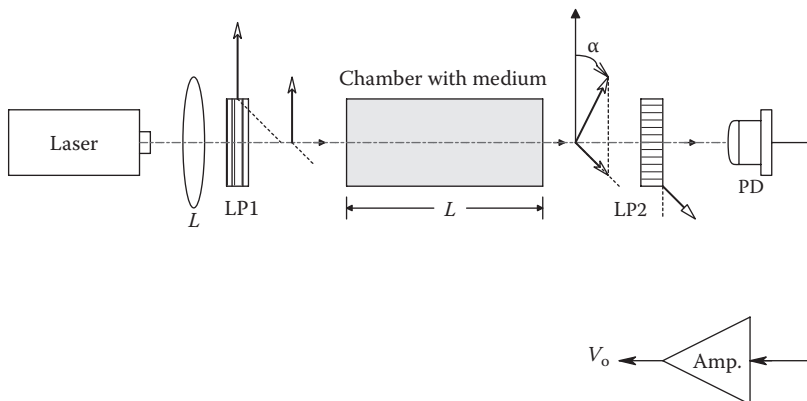


FIGURE 6.42

A basic, manually nulled, optical polarimeter. The sample chamber holding the optically active medium is typically $L = 10$ cm in length.

Thus, $\alpha = 90^\circ$ will result in maximum light transmission through the system to the detector. Obviously, the $I = f(\alpha)$ function is nonlinear. This nonlinearity is true even for $\alpha < 7^\circ$, where the intensity is given by $I \approx I_{\text{MAX}}(\alpha^2)$, α in radians. To eliminate system nonlinearity for small rotation angles, we can orient the polarization axis of the analyzer polarizer to 45° . Now, the light intensity at the detector will be maximum for $\alpha = 45^\circ$, and a constant level will be detected for $\alpha = 0$. The intensity at the detector for a 45° analyzer can be written as

$$I = \left(\frac{I_{\text{MAX}}}{2} \right) [1 + \sin(2\alpha)]. \quad (6.112)$$

Now, for $\alpha < 7^\circ$, I may be approximated by

$$I \approx \left(\frac{I_{\text{MAX}}}{2} \right) [1 + 2\alpha], \quad (6.113)$$

where α is in radians. Thus, the detector output will be linear in α , and the DC bias may be subtracted from the rotation-proportional voltage. Limitations to the resolution and precision of the basic polarimeter are set by the noise in I_{MAX} and the noise in the electronic photodetector.

A *second type of polarimeter*, described by Rogers (1973), was used in the design of a FO, Faraday magneto-optic, electric current sensor. The Faraday magneto-optic effect, described in Section 6.7.1, causes a rotation in the polarization angle of LPL traversing a single-mode FOC with turns around a current-carrying bus bar. The emergent beam is collimated and passed through a Wollaston prism, WP, as shown in Figure 6.43. Two beams of light emerge from the prism and are directed to photodetectors D_1 and D_2 . The WP is made from two sections of calcite, a birefringent mineral having different refractive indices in two different directions. As light travels through the first prism, both the ordinary and extraordinary rays travel colinearly with different refractive indices. At the junction of the two prisms, the beams are interchanged so that the ordinary ray propagates with a lower refractive index and is refracted away from the normal to the interface, while the extraordinary ray is acted on by a higher refractive index and is refracted toward the normal. The divergence angle between the two beams is increased upon exiting the prism; it depends on wavelength of the light used. (The divergence angle is about 22° for HeNe laser light at 633 nm.) If the input beam to the WP is linearly polarized at an angle α with the x -axis, the x - and y -components of the entering ray's E vector will be resolved into the emergent rays: The upper (ordinary) ray will have its E vector directed in the x -direction and be proportional to E_{xr} of the input ray, and the lower (extraordinary) ray will have its E vector directed in the y -direction and be proportional to E_{yr} of the input ray. The detector outputs thus will be proportional to E_{xr}^2 and E_{yr}^2 . In the Rogers detector, the output is formed by analog processing the detector outputs as described by the following equation:

$$V_o = K_D \frac{I_E - I_O}{I_E + I_O}. \quad (6.114)$$

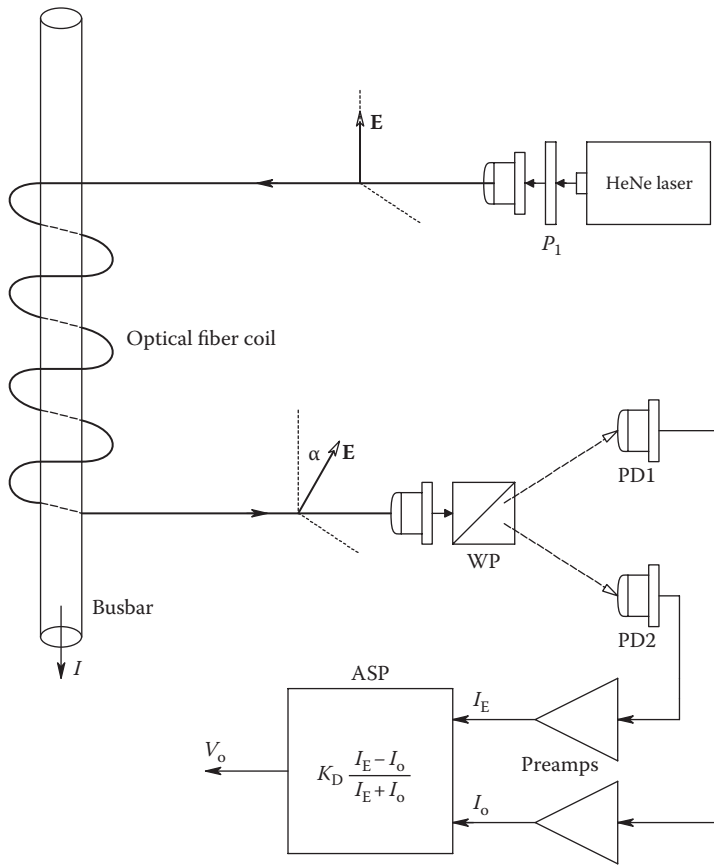


FIGURE 6.43
Rogers' (1973) polarimeter used in the MOCS system of Figure 6.39.

We note that if the analyzing WP is rotated 45° clockwise with respect to the x -axis of the polarized input ray to the FO coil, the intensity component of the extraordinary ray from the analyzer prism is given by

$$I_E = KE^2 \cos^2(45^\circ - \alpha). \tag{6.115}$$

And the intensity of the ordinary emergent ray is

$$I_O = KE^2 \sin^2(45^\circ - \alpha). \tag{6.116}$$

When these intensity expressions are substituted into Equation 6.114, it is easy to show by trigonometric identities that the detector output is given by

$$V_o = K_D \sin(2\alpha). \tag{6.117}$$

Equation 6.117 reduces to $V_o = K_D 2\alpha$, α being in radians, for $\alpha < 7^\circ$. Thus, the Rogers ratio detector is substantially independent of noise affecting E , but still is affected by detector

shot noise and noise in the ratio circuit. It should be noted that AC in the bus bar will create a power line frequency B field, which in turn will give a sinusoidally varying α . Thus, we will see for V_o ,

$$V_o(t) = K_D 2\alpha_o \sin(2\pi f t). \tag{6.118}$$

It is this signal that is proportional to $I(t)$ in the busbar.

A *third polarimetric means* of resolving very small optical rotation angles was developed by Coté et al. (1992). The Coté system, which is a true angle measuring system, is shown in Figure 6.44. A collimated beam of laser light is passed through a linear polarizer, then through a quarter-wave plate to produce *circularly polarized light*. The circularly polarized ray is then acted on by a rotating, linear polarizer (RP). The emergent beam is linearly polarized with the angle of its E vector rotating through 360° with the angular velocity of the rotating polarizer (RP), $2\pi f_r t/s$. The emergent beam is split into a reference beam and a beam that is passed through the optically active medium, M , in which we wish to measure the optical rotation, α_m . The reference beam and the measurement beam from the medium are then each passed through analyzer polarizers to two photodetectors, PDR and PDM. The voltage output of each photodetector is proportional to the magnitude squared of the E vectors emerging from the analyzer polarizers. In mathematical terms,

$$V_{oR} = \frac{K_D |\mathbf{E}_R|^2}{2} [1 + \cos(4\pi f_r t)] \tag{6.119a}$$

$$V_{oM} = \frac{K_D |\mathbf{E}_M|^2}{2} [1 + \cos(4\pi f_r t + 2\alpha_m)]. \tag{6.119b}$$

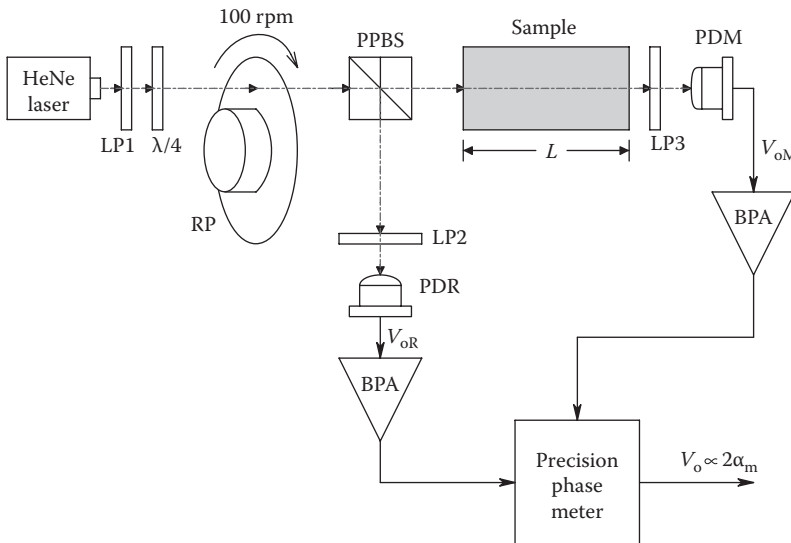


FIGURE 6.44

The millidegree polarimeter devised by Coté et al. (1990). *Note:* LP = linear polarizer, $\lambda/4$ = quarter-wave plate, RP = rotating linear polarizer (angular velocity ω_r , constant), PDR = reference photodetector, PDM = measurement photodetector, BPA = band-pass amplifier tuned to $2\omega_r$.

$V_{oM}(t)$ and $V_{oR}(t)$ are passed through band-pass filters tuned to $2f_r$ in order to eliminate their DC components and restrict the noise bandwidth. A precision phasemeter is used to directly measure the double optical rotation angle, $2\alpha_m$, from the cosine terms in Equations 6.119. The magnitudes of E_R and E_M and the noise associated with them have little effect on the measurement of $2\alpha_m$. System resolution is better than $\pm 0.0005^\circ$ (Coté et al. 1990). Note that the RP is effectively the *chopper* in the Coté et al. system and effectively sets system bandwidth by the Nyquist criterion. It is apparent that the Coté system will not work to measure α_m varying at power line frequencies unless the optical modulation frequency is made 600–1000 Hz. This system is best suited to measure small, static optical rotation angles such as found when measuring aqueous glucose concentrations.

A fourth type of polarimeter having an interesting closed-loop, self-nulling design was described by Gilham (1957). It had a noise-limited resolution of about $\pm 0.0006^\circ$. A schematic diagram of the Gillham design is shown in Figure 6.45. Full-scale sensitivity was 0.02°

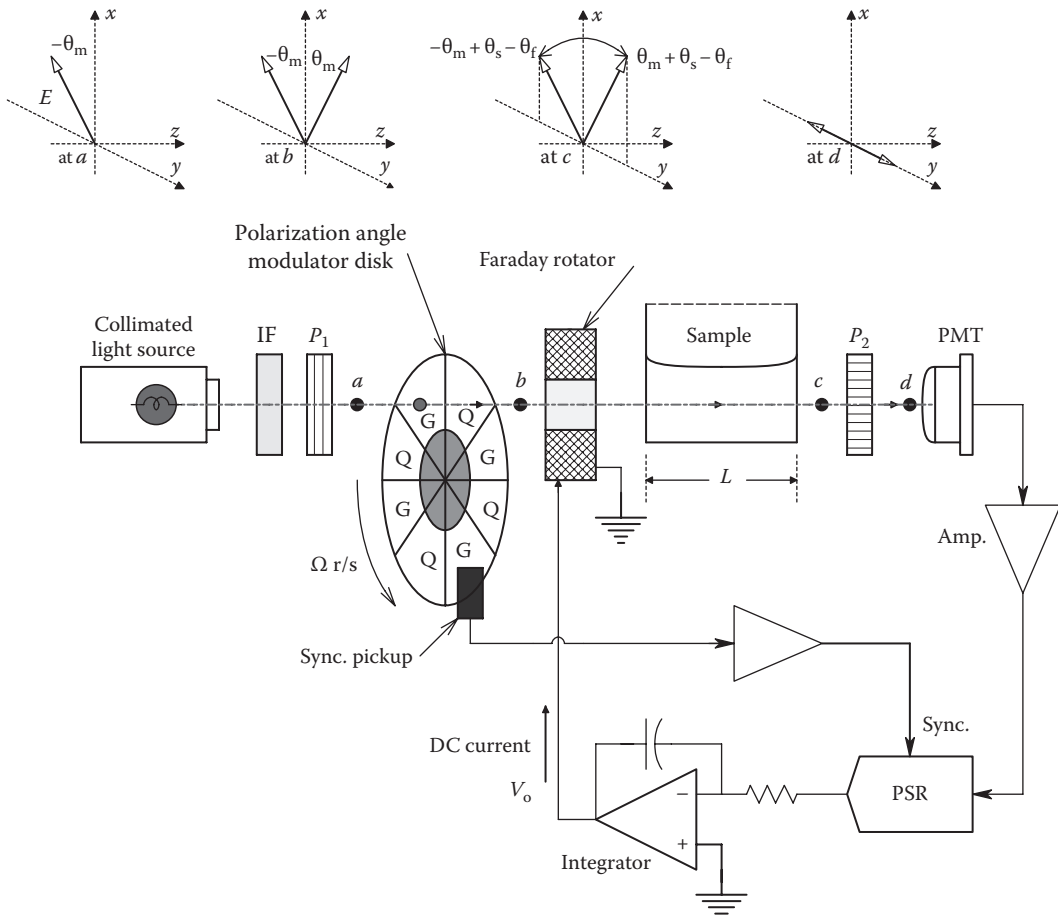


FIGURE 6.45 The self-nulling polarimeter of Gilham (1957). A constant optical rotation, α_o , is produced by the optically active sample in the cuvette of optical path length L . The chopper disk rotates at 1000 rpm. It has alternating segments of plain glass and linear polarizers that rotate the beam's E vector by 2θ degrees. The integrator forces a zero net DC optical rotation at (c), which gives zero DC light flux at (d). Note: P_1 and P_2 are orthogonal linear polarizers, PMT = photomultiplier tube photodetector, and IF = interference filter monochromator.

for the open-loop version of the Gillham system, which makes use of a double-sideband, suppressed-carrier detection scheme. Referring to Figure 6.45, we see that monochromatic light passing through polarizer LP1 is linearly polarized at an angle of $-\theta_m$ (clockwise) with respect to the x (vertical)-axis. A polarizing chopper was made by cementing four, equally spaced, quartz polarizers to a glass disk, which was rotated at 1000 rpm. When the beam emerging from LP1 passes through the disk, the polarization vector is either unchanged as it passes through the glass (or is rotated $+2\theta_m$ counterclockwise) when it passes through the quartz. Thus, the beam emerging from the chopper wheel is alternately polarized at $-\theta_m$ and $+\theta_m$ at 66.7 cycles/s with a square-wave waveform.

The polarization modulated beam next passes through a Faraday rotator (FR), whose DC coil current causes an additional rotation of the chopped beam so that the beam exiting the FR has the angles $(-\theta_m + \alpha_F)$ and $(+\theta_m + \alpha_F)$. The beam next interacts with the sample, S , whose optical rotation, α_S , we wish to measure. The ray emerging from the sample thus has alternate polarization vectors of $(-\theta_m + \alpha_S + \alpha_F)$ and $(\theta_m + \alpha_S + \alpha_F)$. The beam next passes through polarizer LP2 on its way to the photodetector. LP2's axis is put at -90° to the input x -axis, so that the photodetector's output is a 66.7 Hz square wave with peak-to-peak amplitude given by

$$V_{d(\text{PPK})} = \left(\frac{KI_o}{2} \right) \{ [1 - \cos(2(\theta_m + \alpha_S - \alpha_F))] - [1 - \cos(2(-\theta_m + \alpha_S - \alpha_F))] \} \quad (6.120)$$

which reduces to

$$V_{d(\text{PPK})} = \left(\frac{KI_o}{2} \right) \{ \cos(2(\theta_m + \alpha_F - \alpha_S)) - \cos(2(\theta_m - \alpha_F + \alpha_S)) \}. \quad (6.121)$$

By trigonometric identity, we simplify the Equation 6.121 to find

$$V_{d(\text{PPK})} = (KI_o) \sin(2\theta_m) \sin(2\alpha_S - 2\alpha_F). \quad (6.122)$$

Assuming that the magnitudes of the arguments of the sine functions are less than 7° , we may finally write, noting that the angle values are in radians,

$$V_{d(\text{PPK})} = 4KI_o \theta_m (\alpha_S - \alpha_F). \quad (6.123)$$

From Equation 6.123, we see that if the FR coil current is zero, the waveform observed at the detector is a true square-wave, double-sideband, suppressed-carrier modulated signal. Gillham used $\theta_m = 3^\circ$. If the FR is included in a feedback loop, as shown in the block diagram of Figure 6.46, it is easy to show that the SS output of the integrator, V_o , is exactly

$$V_o = \frac{\alpha_S}{K_F} \quad (6.124)$$

where K_F is the FR constant in radians/volt. This servo loop generates a Faraday rotation angle in the SS, which is equal and opposite to the sample angle. Because of its relatively low modulation frequency (66.7 pps), the Gillham system is not suitable to measure 50 or 60 Hz power line currents with Faraday MOCSS; it is intended to measure small ($<0.03^\circ$), static optical rotation angles.

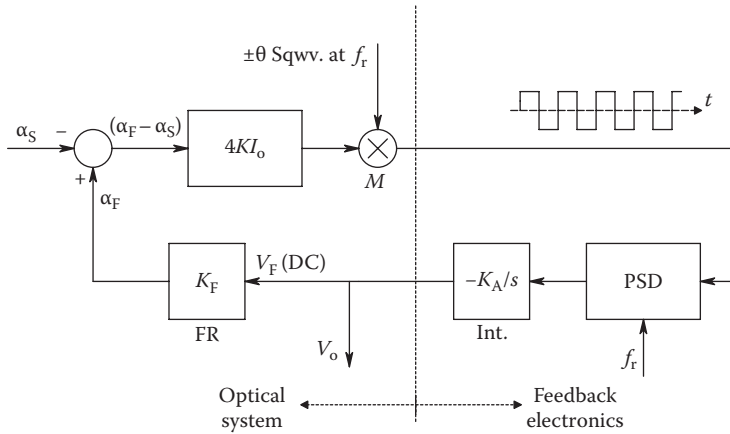


FIGURE 6.46

Systems block diagram of the feedback system used by Gilham (1957) to measure the optical rotation of the sample, α_s . The output of the phase-sensitive demodulator (PSD) is integrated and the integrator output V_o is used to drive the FR so that its rotation $\alpha_f \equiv -\alpha_s$ in the SS.

A fifth type of polarimeter that also uses feedback was described by Cease and Johnston (1990). Their system is shown schematically in Figure 6.40. Instead of using multiple turns of FOC wrapped solenoidally around the current-carrying busbar, they used a single block of quartz glass as a single-turn optical pathway through which the busbar passes. The AC in the busbar causes a proportional optical rotation, α , by the Faraday magneto-optic effect. (Positive α is taken here as counterclockwise.) An analyzer polarizer, LP2, is assumed to have its axis at $+45^\circ$ (counterclockwise) to the vertical (0°) polarization axis of the input polarizer, LP1. Thus, the detector output voltage, V_1 , may be expressed as

$$V_1 = \left(\frac{K_1 I_o}{2} \right) [1 + \sin(2\alpha)]. \tag{6.125}$$

For $\alpha < 7^\circ$, Equation 6.125 reduces to

$$V_1 \cong K_1 I_o [0.5 + \alpha] \tag{6.126}$$

where, as expected, α is in radians. As is seen in the system block diagram, Figure 6.47, the photodetector output, V_1 , is conditioned by two, simple low-pass filters. The output of the second filter, V_3 , is subtracted from a DC level, V_{c+} and this difference is amplified and used to control the average light intensity, I_o , of the LED light source for the system. Thus, I_o is given by

$$I_o = G_m K_{LED} (V_{c+} - V_3). \tag{6.127}$$

It is seen that this feedback system is nonlinear and that a simple pencil and paper analysis of the transfer function for V_2 vs. α is out of the question. Accordingly, we have simulated the system's behavior.

The original Cease and Johnston paper gave no constants for their system's components, so we arrived at the values used by reasonable estimates and trial and error. Figure 6.48 illustrates the results when the program given earlier is run with the constants given in the figure caption. Note that the photodetector output, V_1 , is a sinusoid of line frequency whose peak-to-peak amplitude in the SS is proportional to the current in the busbar. The DC level of V_2 ,

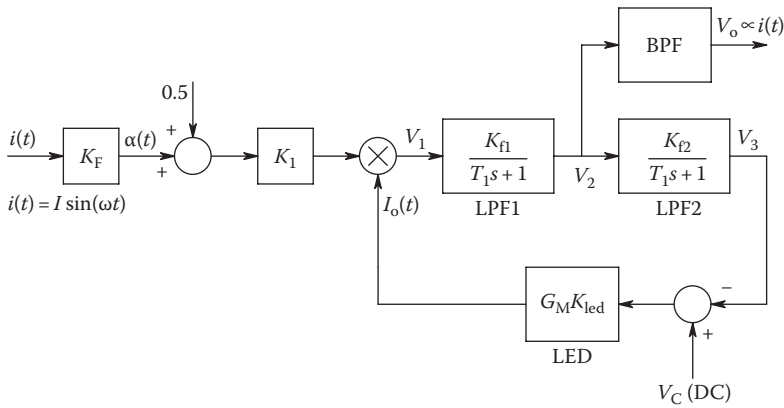


FIGURE 6.47 Block diagram of the polarimeter used in the MOCS described by Cease and Johnston (1990). See text for description.

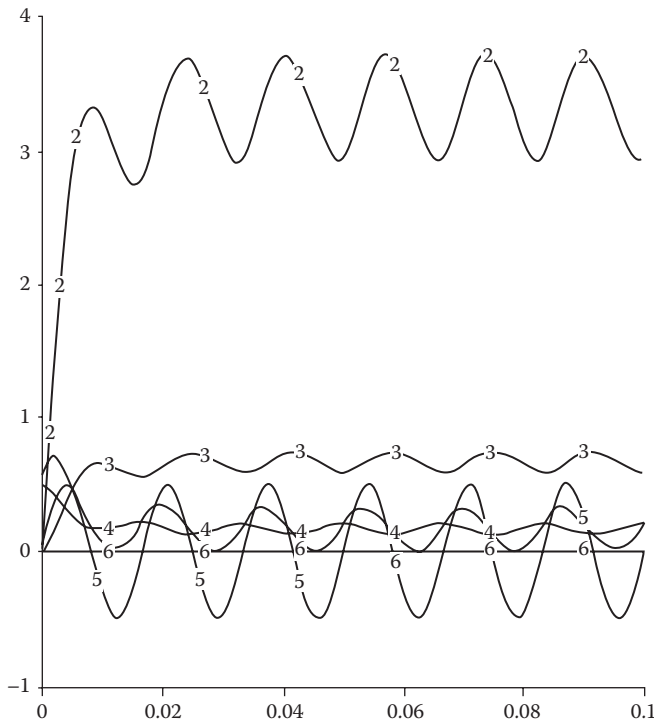


FIGURE 6.48 Results of a Simnon™ simulation of the block diagram system of Figure 6.47. The AC component of voltage V_2 (trace 2) is proportional to $I(t)$ in the busbar and would be isolated by a band-pass filter. Other traces are 1 = V_1 , 3 = V_3 , 4 = I_o , 5 = alpha, 6 = zero.

however, must be removed by passing V_2 through a band-pass filter (not shown in the original Cease and Johnston paper). It is the AC voltage output of the band-pass filter that is actually measured to give the value of $I(t)$ in the busbar. In our simulation of this system, we noted that the linear proportionality between the peak-to-peak V_2 and $I(t)$ is lost for $I_{pk} > 1200$ A, for the constants we used. An I_{pk} of 2000 A gives severely distorted peaks on the V_1 waveform.

In conclusion, we have seen that there are a number of ways that can be used to resolve optical rotation angles as small as 0.001° . Such small polarization angles may occur in instrumentation systems that make use of the Faraday magneto-optic effect, the electro-optic effect, the electrogyration effect (Rogers 1976), and in chemical polarimetry (cf. Section 13.2).

6.3.3 Fiber-Optic Mechanosensors

In this section, we will review the operation of various mechanosensors using optical fibers and light, (excluding FO, Sagnac-effect gyroscopes). FO systems have been devised to measure pressure, force, displacement, acceleration, etc., using a variety of mechanisms. Hochberg (1986) placed FO sensors in two classes: extrinsic FO sensors, in which the FOC(s) is/are passive communications link(s) to an optically active terminating element, and intrinsic FO sensors in which some optical property of the FOC itself is changed by the QUM.

Examples of extrinsic sensors are the intravascular pressure transducer developed by Hansen (1983) (see Figure 6.49A and B) and the feedback pressure sensor developed by Neuman (1988) for measuring intracranial pressures. The Neuman sensor, shown in

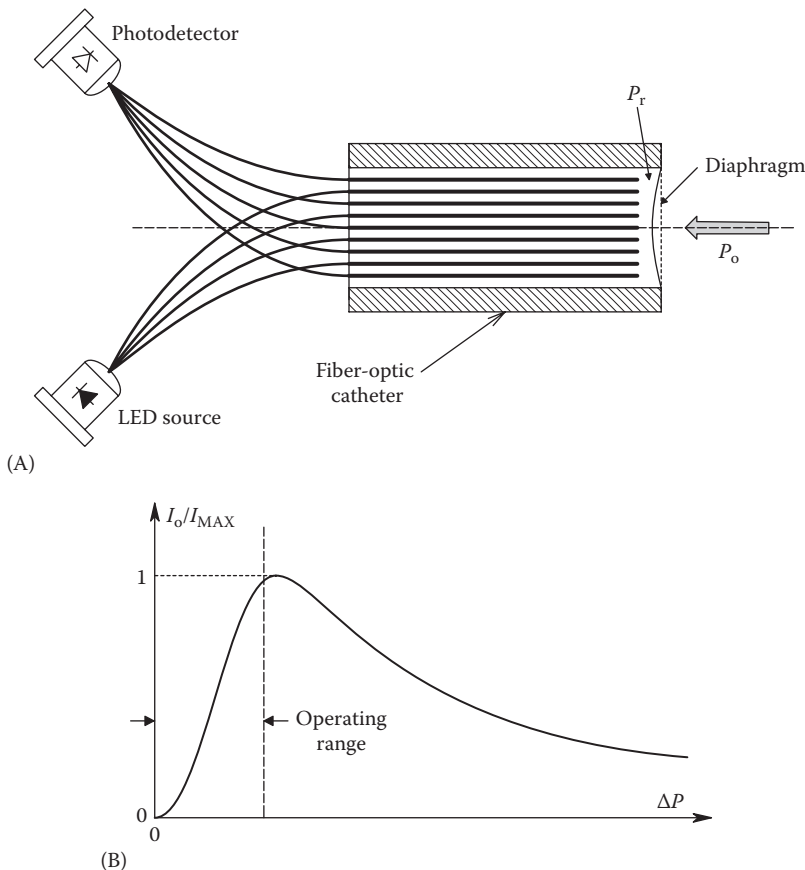


FIGURE 6.49

(A) Mixed fiber, FO pressure sensor. The pressure difference across the diaphragm causes a change in the measured light output. (B) Relative sensor output intensity vs. pressure difference for the Hansen (1983) FO pressure sensor.

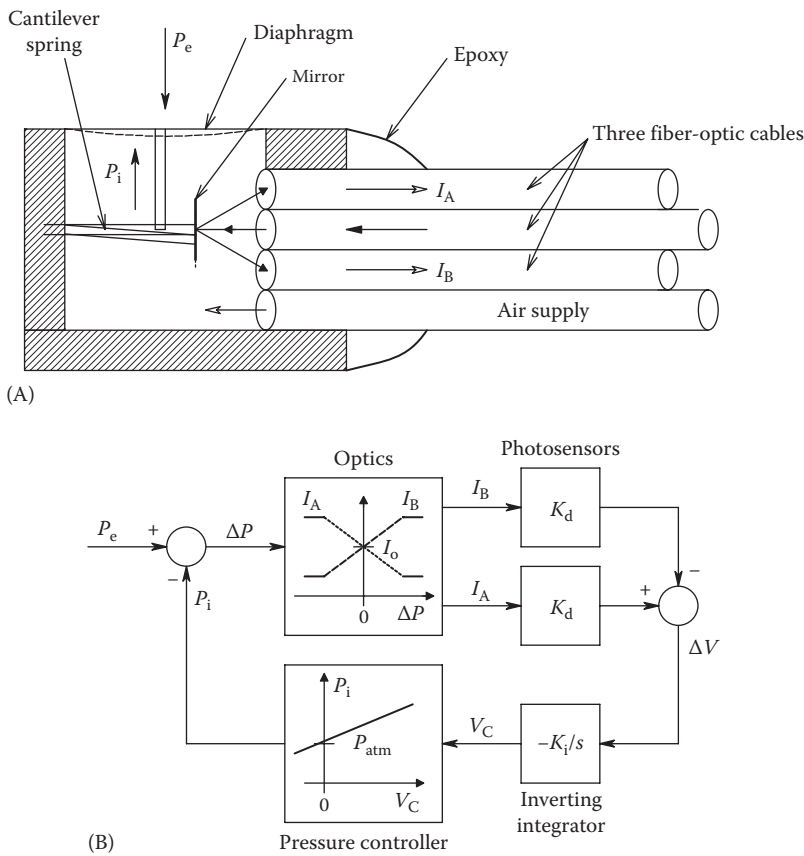


FIGURE 6.50

(A) The Neuman (1988) pneumatic feedback pressure sensor probe. When the internal pressure equals the external pressure being measured, there is no diaphragm deflection, and the light outputs are equal. (B) Block diagram of a simple controller for the Neuman pressure sensor. The system requires a voltage-to-pressure transducer to adjust the internal pressure of the probe.

Figure 6.50A, is a feedback or null system. The micro-diaphragm deflects proportional to the pressure difference across it. An increased external pressure, P_e , causes an inward deflection of the diaphragm, in turn causing less light to be reflected from the input optical fiber to fiber A, and more light to enter fiber B. Fibers A and B go to two photodiodes. The conditioned outputs of the photodiodes are subtracted to form an error signal that, if negative, causes the pressure inside the probe, P_i , to increase, forcing the diaphragm outward until the error signal is zero. At this point, the diaphragm is in its equilibrium position, and $P_i = P_e$. P_i is easily measured by a standard pneumatic pressure sensor and, in the SS, is equal to P_e . The block diagram for a simple controller for the Neuman probe is shown in Figure 6.50B. Other extrinsic FO sensors are shown in Figure 6.51A and B.

One intrinsic FO mechanosensor (FOM) makes use of the property of optical fibers in which microbending causes a net decrease in the transmitted light intensity because of the loss of light power into the cladding; the bending angle prevents internal reflection of light back to the core (see Figure 6.52). The microbending obviously requires small deflections of the deformer relative to the fiber and can be the result of the deflection of a diaphragm (pressure sensor) or elongation of a load cell (isometric force sensor). Another intrinsic

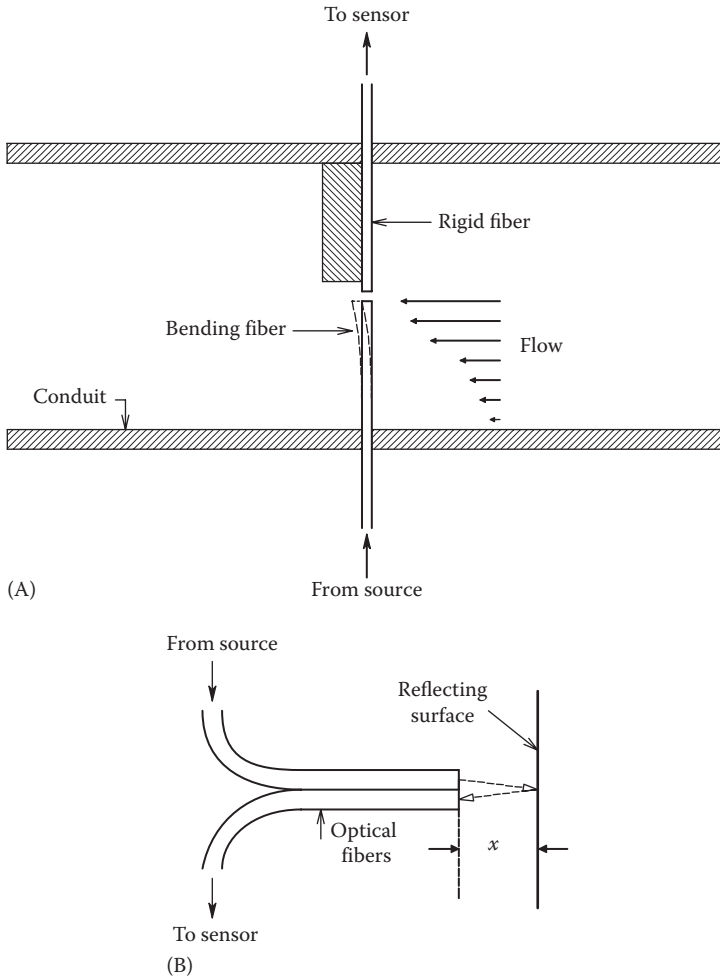


FIGURE 6.51

(A) A basic, single-fiber, extrinsic sensor in which some force (e.g., from flow) causes lateral displacement of one fiber, decreasing light coupling. This type of sensor is inherently nonlinear. (B) The optical coupling between two OFs varies with the distance from a plane reflecting surface in a nonlinear manner.

FOM design makes use of interferometry. Figure 6.53, after Hochberg (1986), illustrates an interferometric sensor system. The detector utilizes the light and dark interference patterns caused by combining the reference and measurement fiber light outputs. In this case, some net deformation of the measurement fiber causes a change in the fiber’s core’s refractive index, hence a phase change in the light output and an interference fringe shift that can be sensed by the detector. The spatial distribution of light in an interference pattern at its center generally follows a function of the form

$$I(x) = \left(\frac{I_o}{2} \right) \left[1 + \cos \left(\frac{4\pi x}{X} \right) \right] \tag{6.128}$$

where X is the spatial period of the fringes. The fringe shift occurs because of a small change (δx) in the spatial parameter, x . For maximum sensitivity, the fractional fringe shift should

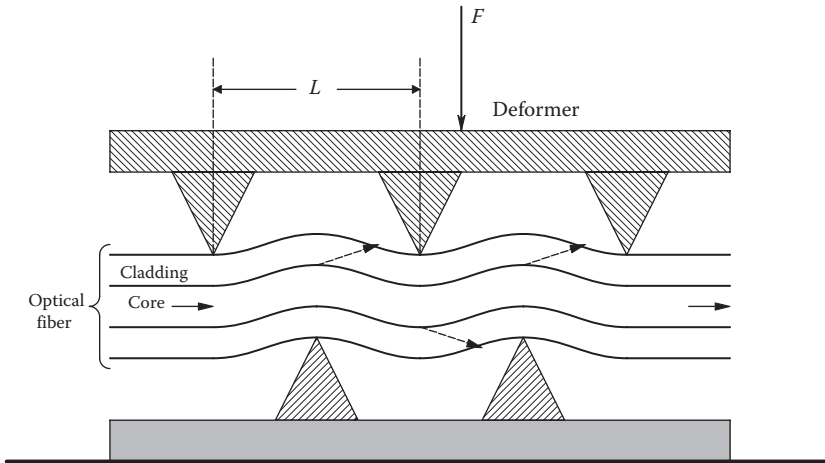


FIGURE 6.52

Intrinsic FOMs can be based on microbending a light-transmitting fiber. Microbending causes a decrease in the intensity of the exiting light because light energy is lost into the cladding. A pressure sensor can be made in which a membrane presses the microbending fixture into the OF.

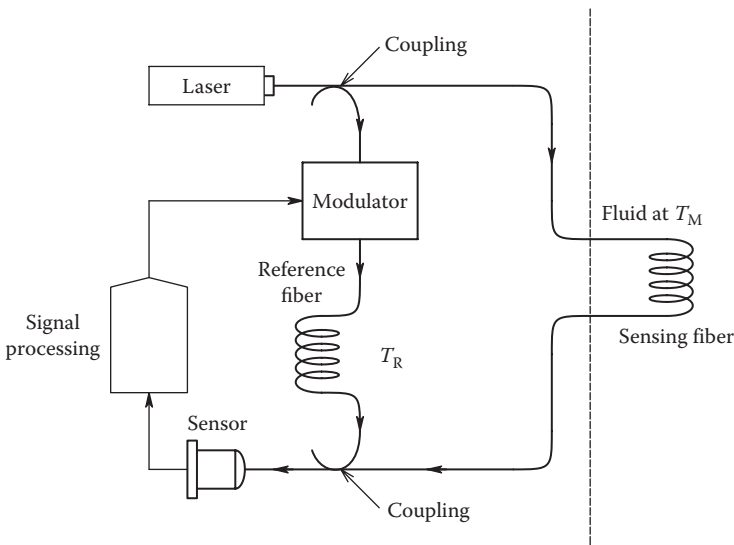
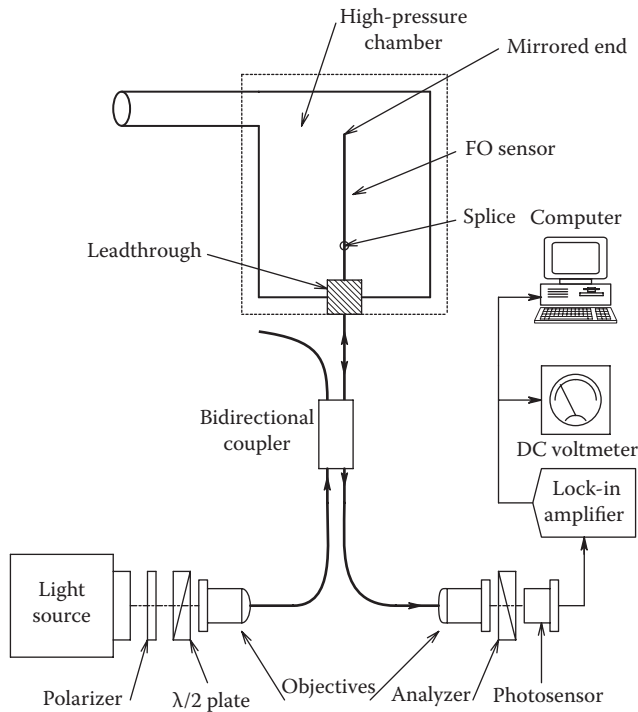


FIGURE 6.53

Schematic diagram of a two-fiber interferometer in which some physical quantity acting on the sensor fiber causes a phase change between the two light paths, hence a change in the detected intensity of the interference pattern generated.

be sensed at the points of the steepest slope of the cosinusoidal term in Equation 6.128. Thus, $I(\delta x) \cong (I_0/2)[1 + 4\pi\delta x/X]$.

An intrinsic, FO pressure sensor system using polarized light and interferometry was described by Bock et al. (1990). The Bock system is basically a one-fiber, Mach-Zehnder interferometer in which the sensing fiber carries two optical modes differing in polarization and whose phase delays change differentially in response to the applied external pressure. The pressure sensor of Bock et al. (1990) was shown to have almost perfect compensation

**FIGURE 6.54**

Schematic diagram of the interferometric, high-pressure sensor developed by Bock et al. (1990). It is a one-fiber, Mach-Zehnder interferometer that makes use of the pressure dependence of the sensing fiber's refractive index.

for temperature changes. Figure 6.54 shows their experimental setup, which used a HeNe laser as a source. The laser output was linearly polarized and launched parallel to one of the two principal axes of a highly birefringent (HB) fiber leading to the sensor. A bidirectional coupler couples light to the measurement fiber. The HB measurement fiber has a 45° axial splice (the reference section is rotated 45° with respect to the input segment). The reference section is of length L_1 . The sensing (terminal) section of HB fiber is axially spliced with a 90° rotation to the reference section. The terminal end of the sensing section of fiber has a gold mirror to reflect the light back down the fiber to the analyzer and detector. In another paper, Barwicz and Bock et al. (1990) reported further on their FO pressure sensing system. The photodetector output voltage was given by the rather complex formula

$$v(p, T) = \left(\frac{V_o}{2} \right) \left\{ 1 + \cos \left[\left(\frac{2\pi L}{\lambda} \right) \left(\frac{\lambda}{L_B} - C_p \beta \Delta T + |C|ap \right) \right] \right\} V \quad (6.129)$$

where

ΔT is the temperature difference from the reference temperature

p is the gauge pressure the FO sensor is exposed to

λ is the wavelength of light in the optical fibers

L_B is the beat length parameter of the fiber

L is the length of the fiber

a and β are material constants

C_p is the relative photoelastic constant

C is a pressure-dependent photoelastic constant given by

$$C = n^3 \left(1 + \frac{\sigma}{2E} \right) (p_{12} - p_{11}) \quad (6.130)$$

where

n is the refractive index of the fiber core

σ is the Poisson ratio

E is Young's modulus

p_{11} and p_{12} are optical strain coefficients

Barwicz and Bock (1990) showed that $v(p)$ obtained experimentally follows Equation 6.129 quite closely, the period of $v(p)$ varying inversely with sensing fiber length, L . For example, with $L = 26$ mm, the period in $v(p)$ is 26 MPa; that is, the cosinusoidal curve repeats itself every 26 MPa over a 100 MPa range. Note that 1 psi = 6895 Pa. By using three different lengths of FO sensor and computer processing of the sensor outputs, Barwicz and Bock were able to cover a 100 MPa range with 0.01 MPa resolution. This implies an 80 dB SNR for their system for high-pressure measurements.

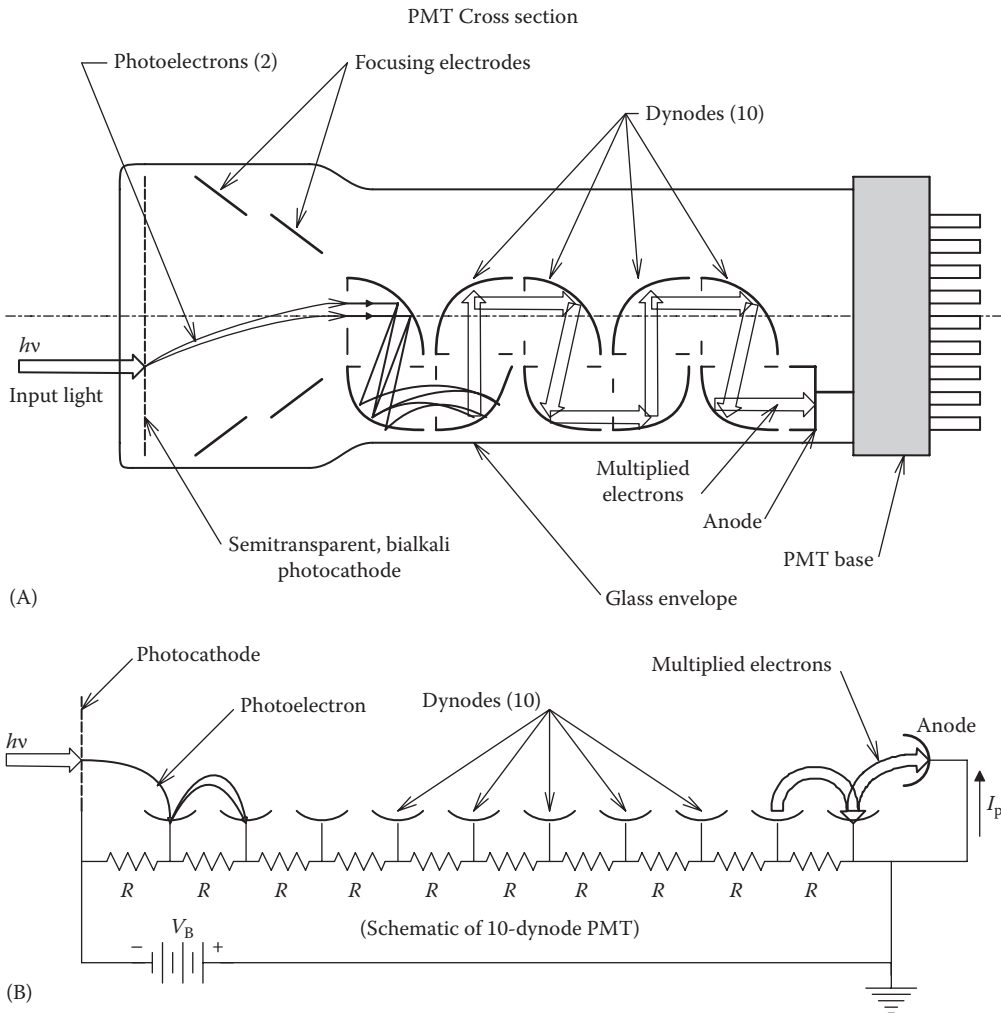
Additional examples of FOMs may be found in the texts by Culshaw (1984), Marcuse (1981), Davis (1986), and Cheo (1990). The design of FOMs is a rapidly growing field that has an active literature in instrumentation and in optoelectronics journals.

6.4 Photomultiplier Tubes and Related Photoelectron Multiplication Devices

6.4.1 Introduction

Photomultiplier tubes (PMTs) are the most sensitive of visible-range photon sensors. Under special operating conditions, they can respond in an almost 1:1 manner to single photons that strike their photocathodes. PMTs have been replaced in many applications by solid-state sensors such as avalanche photodiodes (APDs). However, they remain the photon sensor of choice for ultralow light intensity sensing such as in fluorescent molecular tagging applications, Raman spectrophotometry, single-bubble sonoluminescence (SBL) research, and spectrophotometry. PMTs are vacuum tubes. Early PMTs had glass vacuum envelopes about the size of a beer can; newer PMT designs are smaller. Figure 6.55A schematically illustrates a cross section through a cylindrical, end-on, PMT. Inside are *photocathode*, electron-focusing electrodes, and a series of 8–14 *dynode* electrodes where electron multiplication takes place. Ten dynodes is a more common number. The multiplied electrons are collected by the anode and form the PMT's photocurrent. In *side-on PMTs*, the geometry is changed so the photon (light) window is at right angles to the long axis of the envelope.

The photocathode is generally semitransparent to permit greater efficiency in photon capture. It is often made from bialkali, low-work function materials that determine its *spectral sensitivity* in mA of photoelectrons emitted per watt of light at a given wavelength. If a quantum of light with energy $h\nu \geq \phi$ strikes the photocathode, an electron may be emitted. ϕ is the cathode *work function* in joules. Statistically, the ratio of emitted photoelectrons to incident photons is the *quantum efficiency*, η . In general, $0 < \eta < 1$.

**FIGURE 6.55**

(A) Cross section of a 10-dynode, end-window PMT. (B) Schematic of the 10-dynode PMT. The dynodes are biased with a simple resistance voltage divider.

One photoelectron emitted from the cathode is accelerated in the PMT by the electric field between the cathode and the first dynode to the more positive first dynode electrode. It hits the *first dynode* with sufficient kinetic energy to release perhaps four electrons (on the average) that in turn are accelerated to the *second dynode*, which is at a fixed potential more positive than the first dynode. These electrons, too, gain enough kinetic energy to each dislodge say another four electrons. Now, a total of 16 electrons strike the third dynode, and the process repeats itself. Thus, it is easy to see that $4^{10} \cong 1.05 \times 10^6$ electrons strike the anode as the result of one photoelectron being emitted in a 10-dynode PMT. The dynode-to-dynode potential is maintained by a simple resistive ladder voltage divider biasing circuit as shown in the dynode schematic in Figure 6.55B. The delay time between the emission of the photoelectron from the photocathode until the current pulse is seen is typically ca. 20 ns in a 10-dynode PMT. A typical dynode-to-dynode potential is ca. 100 V.

Desirable characteristics in a PMT

1. High responsivity (A/W)
2. Low dark current
3. High output SNR
4. Wide spectral response
5. Low transport delay
6. High stability

Disadvantages of PMTs

1. Mechanically fragile (glass envelope).
2. Shapes and sizes are limited. Physically large.
3. Need stable high-voltage power supplies.
4. Expensive (\$100s).
5. Responsivity affected by magnetic fields. Needs magnetic shielding in critical applications.
6. Needs cooling to LN₂ temperatures for noise reduction in critical applications.

Competing with glass envelope PMTs are channel-plate photomultiplier (CPM) devices and single-channel PMTs, both of which are described in the succeeding text.

6.4.2 Operation of PMTs

Figure 6.56 illustrates the anode (output) spectral responsivity of five different Oriel PMTs. Each uses a different photocathode material. Note that some cathode materials permit PMT operation at UV wavelengths down to 185 nm (e.g., Hamamatsu R6925, a 9-dynode, side-on PMT that uses a bialkali cathode and a UV glass window) and others in the NIR to wavelengths as long as 1700 nm (e.g., Hamamatsu R5509-72, a 10-dynode PMT that uses an InP/InGaAs cathode and a borosilicate glass window). (See Hamamatsu 2013.)

The gain or amplification of PMTs can be expressed in several ways: (1) an anode *spectral responsivity*, $R_A(\lambda)$, in anode amps per watt of incident light power on the photocathode; (2) a pure current gain between cathode and anode, G_I ; and (3) *photocathode responsivity* $R_C(\lambda)$ in milliamps/watt. $R_A(\lambda)$ and G_I depend on the inter-dynode potentials, and the number of dynodes. For example, the Oriel model 77341 PMT operated at 1000 V has an inter-dynode potential of 100 V, a maximum $R_A(\lambda) = 2 \times 10^5$ A/W, $G_I = 5 \times 10^6$, and $R_C(\lambda) = 0.04$ A/W. In addition, this PMT has a maximum anode current, $I_A = 100$ μ A, from which we can conclude that the *maximum* input light power is 0.5 nW! The model 77341 PMT also responds to light wavelengths from 185 nm (UV) to 870 nm (NIR). It has a broad peak $R_C(\lambda)$ from ca. 300 to 550 nm. Its photocathode is 8 \times 24 mm, its 10%–90% rise time (anode current in response to a step of light) is 2.2 ns, and the transit time between a step of light and the anode current step is 22 ns. (Note that the electrons moving between dynodes travel at finite velocity, being accelerated by the 100 V inter-dynode potentials.) In addition to the anode photocurrent, all PMTs have an *anode* DC dark current, I_{DA} . I_{DA} has two principal components: amplified thermionic emission from the dark photocathode and an ohmic leakage component. Ohmic leakage can be the result of fingerprints and solder flux on

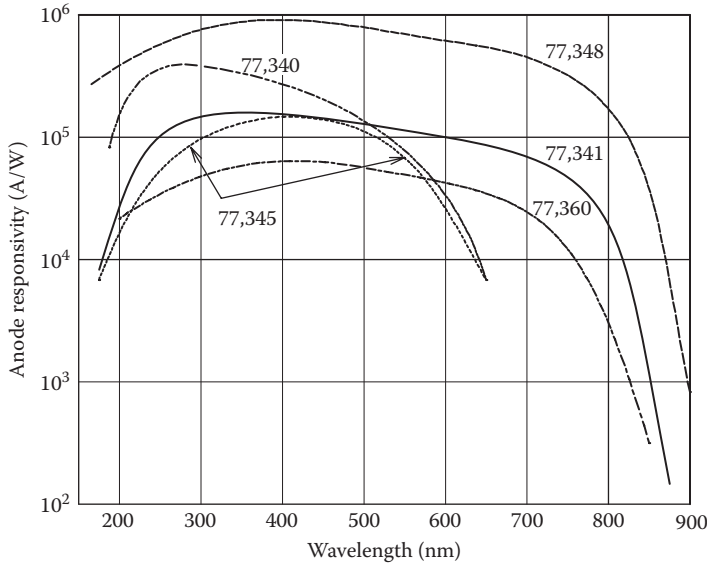


FIGURE 6.56
Spectral sensitivity of five different Oriel PMTs.

the PMT base and socket. The Oriel model 77341 PMT (sold by Newport) (Oriel 2013) has $I_{DA} = 2 \text{ nA}$. The thermionic emission from the cathode I_{TC} can be modeled by

$$I_{TC} = KT^2 \exp\left(\frac{-\phi}{kT}\right)A \tag{6.131}$$

where

- K is a constant
- T is the absolute (Kelvin) temperature
- ϕ is the cathode work function in joules
- k is Boltzmann’s constant ($1.380 \times 10^{-23} \text{ J/K}$)

All PMTs are noisy. Noise in the anode current comes from four sources: (1) white Johnson or thermal noise from the series resistor used in the anode circuit, (2) white noise from thermionic emission from the dynode surfaces (the resistors’ Johnson and dynodes’ shot noises are generally neglected in calculating PMT output SNR [SNR_0]), (3) white shot noise from the thermal emission of the (dark) photocathode, and (4) white shot noise from the photocathode’s actual photocurrent. Thus, the mean-squared anode (output) noise current can be written as follows:

$$\overline{i_{an}^2} = G_I^2 [2q(I_{CP} + I_{TC})]B \text{ mean squared amps} \tag{6.132}$$

where B is the Hz noise bandwidth (very high if we are counting pulses, low if DC light levels are concerned) and the *cathode* photocurrent is $I_{CP} = R_C(\lambda)P_L$. P_L is the light power in watts at a wavelength, λ , incident on the photocathode. I_{TC} is the cathode’s average thermionic emission current; it increases with the cathode’s Kelvin temperature.

Assume the output signal at the *anode* is the RMS photocurrent from the cathode multiplied by G_T . That is, $I_{AP} = I_{CP}G_T = G_T R_C(\lambda)P_L$. Thus, we can write the mean-squared output SNR as

$$SNR_o = \frac{G_T^2 R_C^2(\lambda)P_L^2}{2qBG_T^2[I_{TC} + R_C(\lambda)P_L]} = \frac{R_C^2(\lambda)P_L^2}{2qB[I_{TC} + R_C(\lambda)P_L]} \tag{6.133}$$

Note that SNR_o is independent of the PMT's current gain, and it *increases* with *increasing input light power* to the photocathode, P_L . A commonly used technique to increase SNR_o is to *decrease* I_{TC} by operating the PMT at the temperature of boiling liquid nitrogen (LN₂) (-196°C) or by cooling it thermoelectrically. Lion (1959) reported that there can be a 40 dB increase in a PMT's rms SNR_o between room temperature operation and operation in LN₂. Figure 6.57 shows that PMT SNR exceeds that of *pin* PDs and exceeds APDs for intensities $<1 \times 10^{-11}$ W.

Let us now calculate the light power at 555 nm wavelength to give a PMT $SNR_o = 1$. We assume the PMT is cooled with LN₂ so the cathode thermionic emission is ca. 1 electron/s, or $I_{TC} = 1.6 \times 10^{-19}$ A. We set the $SNR_o = 1$:

$$1 = \frac{R_C^2(\lambda)P_{L1}^2}{2qB[I_{TC} + R_C(\lambda)P_{L1}]} \tag{6.134}$$

Equation 6.134 yields a quadratic equation in P_{L1} :

$$-P_{L1}^2 + \left[\frac{2qB}{R_C(\lambda)} \right] P_{L1} + \frac{2qBI_{TC}}{R_C^2(\lambda)} = 0 \tag{6.135}$$

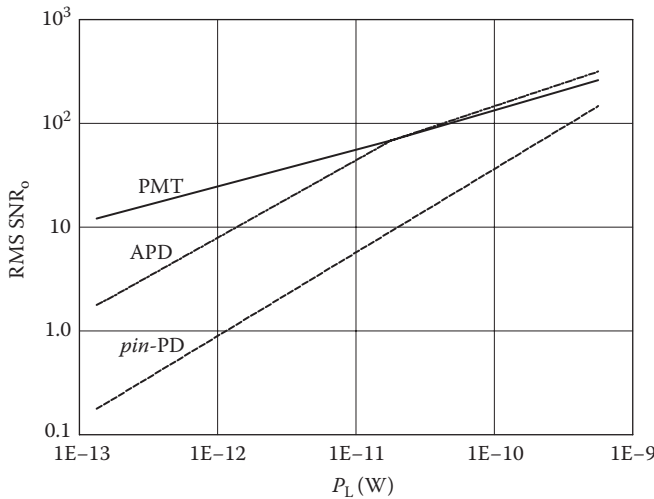


FIGURE 6.57 Representative comparison of output SNRs for three major categories of photosensors under varying input light power conditions. *Note:* PMT = photomultiplier tube, APD = avalanche photodiode, *pin*-PD = *pin* photodiode.

Solving this quadratic equation for its real root, we find a good numerical approximation:

$$P_{L1} \cong \frac{2qB}{R_C(\lambda)}. \quad (6.136)$$

Substituting the numerical values from the Oriel 77431 PMT cited earlier, $R_C(\lambda) = 0.04 \text{ A/W}$ and $B = 10^3 \text{ Hz}$. Thus, $P_{L1} \cong 8 \times 10^{-15} \text{ W}$, mighty small!

An LN₂-chilled, 12-dynode PMT properly operated can be used to statistically detect single photons. Under extreme low-light power conditions, a single photon may or may not generate a single photoelectron. If it does, the electron is multiplied and gives an output current pulse. Under these conditions, the dark current is due to single thermionic electrons from the cathode, which occur at about 1 s^{-1} . One means of resolving photoelectrons over thermionic events is differential counting. A count of output current pulses over a fixed time is made with the PMT covered to get the mean emission rate for cathode thermionic electrons, then the constant photon source is presented to the PMT and the count redone. The extra counts are presumed to be due to photons. A pulse-height window is used to exclude smaller anode output pulses coming from thermionic emission from the dynodes. Cathode thermionic electrons and photoelectrons are assumed to follow Poisson distributions. *Ortec AN51, Rev. 1 (2000)*, describes the statistical theory required to estimate the number of single photons/time. Maximum sensitivity of an LN₂-chilled, magnetically shielded 14-dynode PMT is on the order of 10^{-16} W at $\lambda = 555 \text{ nm}$, using a lock-in amplifier signal recovery system such as the Oriel Merlin™ digital lock-in radiometry system, sold by Newport.

For fast pulse operation of a PMT, the dynodes #10, 11, and 12 are bypassed to ground with capacitors on the order of $0.02 \text{ }\mu\text{F}$. The size of the capacitor bypassing the last (12th) dynode should be at least 100 times the output charge per pulse, sic:

$$C_{12} \geq \frac{100(I_{\text{Appk}}T_p)}{V_{12}}, \quad (6.137)$$

where

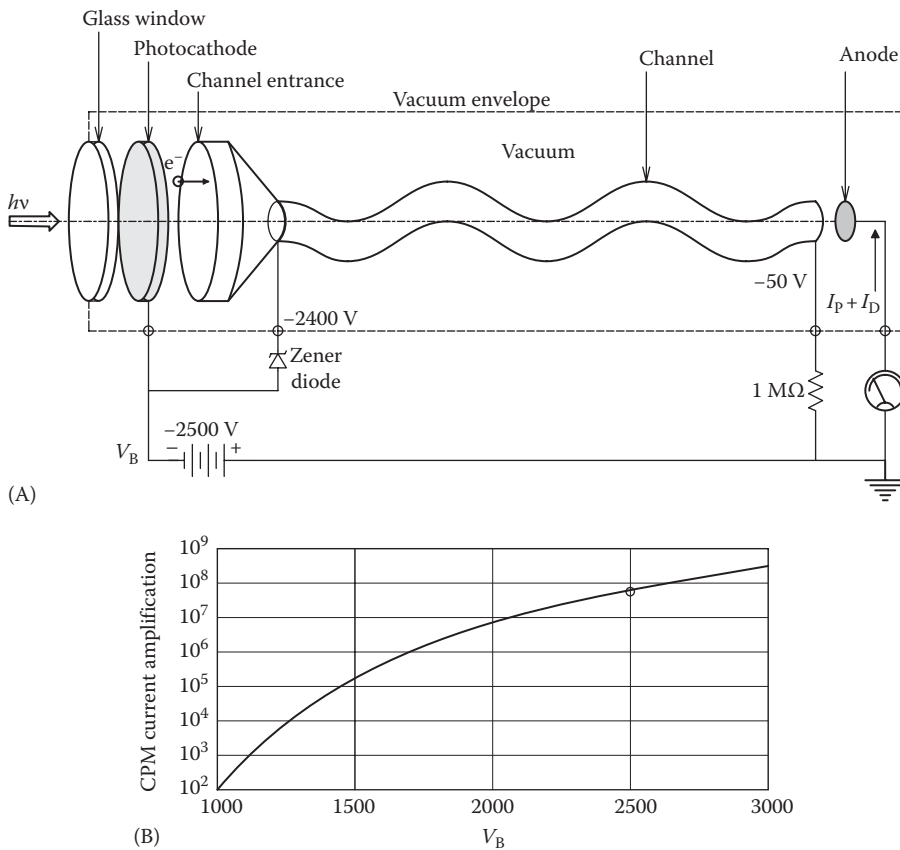
I_{Appk} is the peak anode output current

T_p is the output pulse width

V_{12} is the potential to ground at the 12th dynode (see the Hamamatsu (2013) online catalog, Sect. 11, *Voltage Divider Consideration*)

6.4.3 Single-Channel Photomultiplier

The Perkin-Elmer (P-E) CPM tubes (C900-C1300-C1900 series) have an incredibly simple and ingenious design (see P-E Optoelectronics 2003). The envelope package is cylindrical, 75 mm long, and 10.5 mm in diameter. Four electrical leads come out the bottom. A semi-transparent cathode is deposited on the inner surface of the entrance window. Figure 6.58A illustrates a cross section of the P-E CPM tube. (The photocathode is shown separated from the glass window for artistic simplicity.) In a major departure from PMT architecture using dynodes for electron multiplication, P-E found that by accelerating photoelectrons from the anode into the cork-screw-shaped, hollow, resistive channel, they would collide with the inner walls of the channel, releasing secondary electrons that would also be accelerated and release still more electrons. Current amplifications is given as 5×10^7 with

**FIGURE 6.58**

(A) Cross section of the P-E channel PMT (PE-CPM tube). (B) Plot of current gain (anode current/primary photocurrent vs. operating voltage, V_B).

a supply voltage of $V_B = 2400\text{ V}$. Supply (bias) current flowing up the resistive channel is about $50\text{ }\mu\text{A}$ with $V_B = 2400\text{ V}$. Thus, the total channel resistance is about $47\text{ M}\Omega$. Anode sensitivity is $R_A(\lambda) = 3 \times 10^6\text{ A/W}$ at $\lambda = 410\text{ nm}$ for the P-E model C492 CPM. This CPM has a quartz window and a bialkali photocathode; it has a spectral response that lies between 165 and 650 nm . Also, $I_D = 80\text{ nA}$, the anode current rise time is 3 ns , and the pulse width (FWHM) is 6 ns for this CPM with a 2400 V supply. One P-E CPM (model C911) works into the vacuum UV, down to 115 nm ; it has a magnesium fluoride window. P-E claims that because the channel electron multiplication process is essentially silent, the CPM has a noise level one or two orders of magnitude lower than dynode PMTs, resulting in a significantly higher dynamic range. The CPM tube's current gain is shown in Figure 6.58B.

6.4.4 Microchannel-Plate Photomultipliers

The microchannel plate PMT (MCP) is another dynodeless PMT design. The heart of an MCP is a lead glass disk through which have been inserted from 10^4 to 10^7 small holes. The holes are typically $8\text{--}25\text{ }\mu\text{m}$ in diameter on $10\text{--}40\text{ }\mu\text{m}$ centers and $40\text{--}100$ times longer than their diameters. Each hole acts as a miniature CPM device. A potential of ca. 4000 V is placed across the ends of the holes to accelerate the multiplied photoelectrons traversing

the holes (Pankratz 1998). It has been found that the electron multiplier effect of the holes does not depend on the length l of the holes or on their diameters, d , but rather on the ratio, $\alpha = l/d$. Hence, size reduction is possible permitting many small holes to be placed in parallel in an MCP disk. The glass disk itself is on the order of 2.5 mm thick. The axis of each hole is perpendicular to the disk's surfaces or can be inclined ca. 8° from perpendicular. To establish the E-field across the tubes, a thin layer of a metal such as nichrome or inconel is vapor deposited on the flat surfaces of the disk. The resistance between electrodes is on the order of $10^9 \Omega$. MCPs not only respond to photoelectrons emitted from a photocathode but can also directly detect x-ray and γ -ray photons. The electron gain of each MCP hole is from 10^4 to 10^7 (Wiza 1979).

As an example of an MCP, examine the Burle (now Photonis USA, Pennsylvania, Inc.) model 85104 MCP device. This MCP uses a semitransparent GaAs photocathode that responds over $350 \leq \lambda \leq 860$ nm. It uses a two-stage microchannel plate structure, similar to that shown schematically in cross section in Figure 6.59. The 85104 MCP device is physically small; the case is a cylinder 52.5 mm long and 44.5 mm in diameter. The window is 18 mm in diameter. The maximum $V_B = -3600$ V, and the maximum average anode

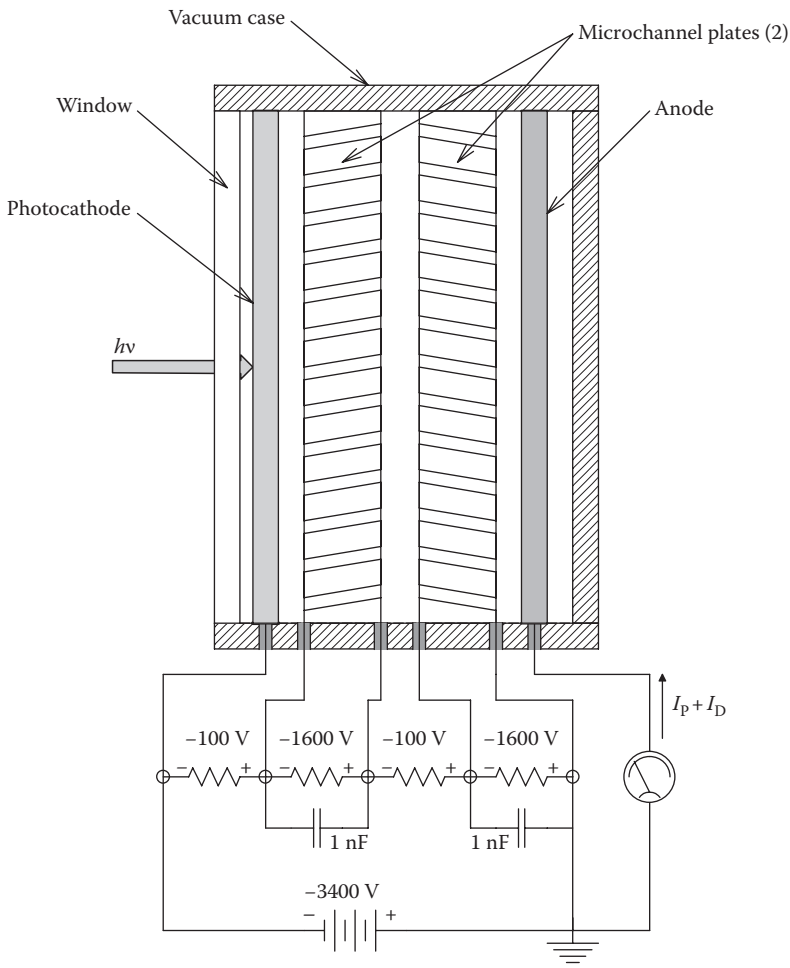


FIGURE 6.59

Cross section (not to scale) of a microchannel plate PM. *Note:* biasing resistors.

current is 300 nA. The peak cathode spectral sensitivity is $R_C(\lambda) = 120 \text{ mA/W}$ at 860 nm; the net electron gain, $G_1 = 2 \times 10^6$. Thus, $R_A(\lambda) = 2.5 \times 10^5 \text{ A/W}$ at 860 nm. The dark current, $I_D = 10 \text{ nA}$, max. Because of the small length of the holes, typical electron transit time is ca. 150 ps, and the pulse spread is ca. 550 ps.

MCP applications include all the applications of vacuum; dynode PMTs, including single-photon detection when cooled with LN_2 ; and image intensifiers. MCPs are also used in the design of ultrahigh-speed CRT oscilloscopes. The MCP device is placed in proximity focus with the phosphor screen of a CRT. This allows use of low beam currents and low acceleration potentials necessary for high writing speeds; the latter are necessary in a high-bandwidth (~several GHz) real-time oscilloscope. MCP devices are also used for the direct detection of charged particles and energetic photons (x- and γ -rays). In this application, they compete with the Hamamatsu position-sensitive, dynode PMTs used with scintillation crystal plates. MCP devices can also be gated by placing a gate grid between the photocathode and the outer channel plate. Holding the grid negative repels photoelectrons and effectively turns the MCP device off.

6.4.5 Summary

In this section, we have described the salient characteristics of the three classes of photomultiplier sensors: vacuum PMTs with dynodes, the single CPM, and MCPs. All three are capable of counting single photons, equivalent to incident light levels on the order of 10^{-17} W . Resolution was seen to be enhanced by reducing device shot noise. Thermionic emission white noise from the cathode and Johnson noise from resistors are reduced by cooling the sensor with LN_2 or thermoelectrically.

Because of their extreme sensitivity, all types of PMTs can be damaged by excessive input light power. Manufacturers generally specify a maximum anode current for an active device. Exposure to high light intensity can also compromise a photomultiplier's sensitivity even when it is not energized; however, the tube will recover if given *rest* in the dark (Lion 1959).

The Hamamatsu *PMT Catalog* (2013) lists and illustrates 16 applications of PMTs. These include the following: (1) spectrophotometry, (2) gamma cameras (used in nuclear medicine), (3) positron emission tomography (PET), (4) in vitro radioassay, (5) liquid scintillation counting systems, (6) biotechnology (cell sorters/counters, fluorometers, DNA sequencers), (7) high-energy physics experiments, (8) oil well logging, (9) mass spectrometry/solid surface analysis, (10) environmental measurements, (11) applications of laser measurements (ladar, side-looking ladar, Doppler laser velocimetry), (12) plasma applications, (13) color scanners, (14) industrial measurements applications, (15) space research applications, and (16) low-light measurements (Raman spectrophotometry). To these, I can add single- and multiple-bubble sonoluminescence and surface plasmon resonance (SPR).

6.5 Ionizing Radiation Sensors

6.5.1 Introduction

Ionizing radiation can be either high-energy, EM radiation that includes photons, gamma radiation, and x-rays or can be a directed beam of high-energy, subatomic particles such as beta particles (electrons), alpha particles (helium nuclei, less the two orbital electrons), or certain ions, moving at high velocity. Ionizing radiations are given that name because

as they pass through various media that absorb their energy, additional ions, photons, or free radicals are created. In a biological context, such ions or free radicals can, under certain circumstances, lead to genetic damage and mutations. On the other hand, the production of ions, electrons, or photons under controlled circumstances can lead to a means of counting the incident ionizing particles and determining their energy. We describe in the succeeding text several designs for radiation sensors, including the familiar gas-filled Geiger–Muller tube (GMT), solid-state crystal detectors, and scintillation counters.

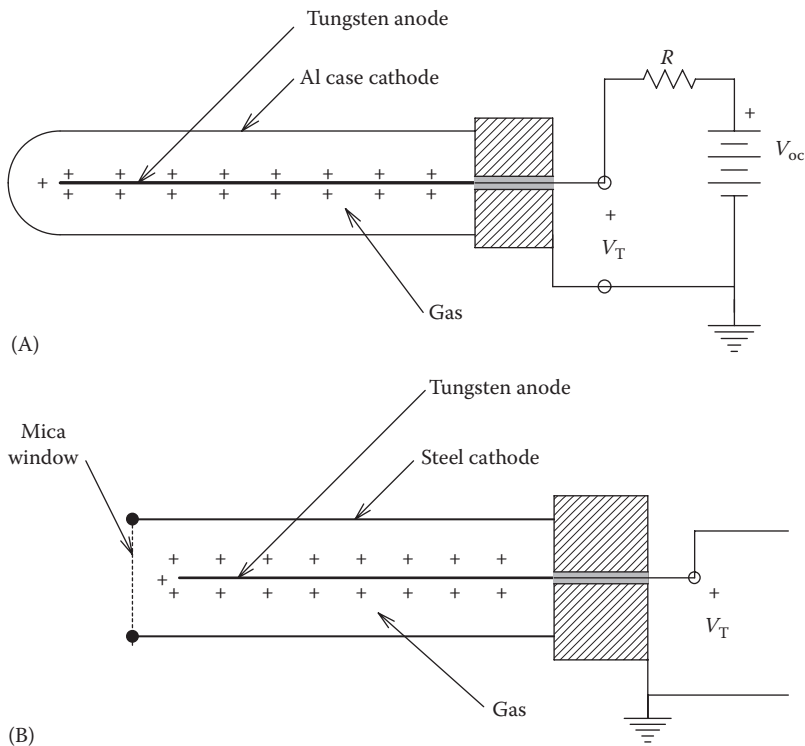
There are several units of radiation dose measurement. The most basic is the number of radioactive emission events per minute (counts per minute, cpm). The *curie* (Ci) is defined as that quantity of any radioisotope undergoing 3.70×10^{10} events/s. Also, 1 becquerel = 27 picocuries (pCi). The *Roentgen* (R) measures the energy produced by gamma radiation (photons) in 1 cm^3 of air. The *rad*, or radiation absorbed dose, recognizes that different materials that receive the same exposure may absorb different amounts of the radiation's energy. A rad measures the amount of radiation energy transferred to some mass of material, typically humans. One Roentgen of gamma radiation exposure results in about 1 rad of absorbed dose in humans. The *rem* stands for Roentgen equivalent man. This unit relates the dose of any radiation to the biological effects of that dose. The dose in rem = $Q \times$ dose in rad, where Q is a *quality factor* dependent on the animal species and type of radiation. For man, given γ radiation or beta particles, 1 rad \rightarrow 1 rem, so $Q = 1$. In the SI system of measurements, 1 Sievert (Sv) \equiv 100 rem, and 1 Gray (Gy) \equiv 100 rad (CDC 2011, Quayle 2013).

6.5.2 Geiger–Muller Tubes

A basic GMT with coaxial symmetry is shown in Figure 6.60A. The outer wall of the tube is generally a thin metal such as aluminum. The outer wall is the negative electrode or cathode of the GMT. The anode is a thin wire of tungsten or platinum run axially at the center of the tube and is, of course, insulated electrically from the cathode. The other major type of GMT is the end-window design, shown in Figure 6.60B. The end window is a thin sheet of mica or mylar and serves to admit particles that have low penetration such as α and β particles. (Although α particles have very high energies, in the range of 4–10 MeV, their penetration in air is relatively short; a 3 MeV α particle travels about 1.7 cm in air, and α particles can be stopped by a sheet of paper or aluminum foil! β particles (electrons) also arise from the decay of certain natural and man-made radioisotopes. An electron with 3 MeV energy will travel about 13 m in air before being stopped and may penetrate as much as 1 mm of aluminum.)

GMTs are generally filled with an inert gas such as helium, argon, or nitrogen at pressures ranging between a few mm Hg and atmospheric. Modern GMTs are self-quenching; that is, they also contain a polyatomic gas such as methane or alcohol vapor that prevents energetic photons emitted from the activated inert gas molecules from reaching the cathode and stimulating the release of secondary electrons by photoelectric emission. The polyatomic gas is either dissociated or decomposed by the high-energy photons in the process of absorbing their energy. Because the polyatomic gas is decomposed, the self-quenching GMT gradually loses its quenching ability with its total number of counts. The effective life of a GMT may be as high as 10^{10} counts. Methane as a quenching gas gives a shorter count life (10^8) than does alcohol vapor.

The DC potential difference across a GMT determines its properties as a radiation counter. Lion (1959) described five distinct ranges of applied voltage to the GMT. (The voltage ranges given in this text are for example; every type of GMT will have numerical values peculiar to its own design.)

**FIGURE 6.60**

(A) Conventional, coaxial, Geiger–Mueller radiation counting tube (GMT). The thin aluminum shell passes most ionizing radiation particles except alpha particles and low-energy electrons (β -particles). (B) A GMT with a thin mica end window to enable counting of alpha particles.

The first region (100–400 V) is called the ionization chamber region. In this region, an ionizing particle entering the GMT causes the generation of one or more electron-ion pairs. These charged particles drift toward the electrodes in the E-field inside the tube. If the field is high enough so that they reach their destinations before an appreciable number of them recombine, and not so strong that additional ionization by collision occurs, the number of charges moving to and arriving at the electrodes is equal to the number initially produced by the collision(s) of the ionizing particle with the inert gas molecule(s). Each surge of ions from an ionizing particle collision thus causes a very small current pulse, the time integral of which is proportional to the number of ions, n_j , produced by the j th collision.

The second region (400–700 V) is called the *proportionality region*. Because of the higher electric field strength, ions and electrons produced in the GMT by an energetic particle are accelerated to higher kinetic energies and thus cause the generation of new ions and electrons by inelastic collisions. This process lacks the energy to become self-sustaining, but a current pulse produced has an area proportional to kn_j . The multiplication factor, k , depends on many factors including the type and pressure of inert gas used. k can range from 10 to 10^4 . Because the current pulses caused by the collisions of ionizing particles with inert gas in the GMT are larger, proportional operation of the GMT allows the measurement of n_j and thus the energy of the ionizing particles.

The third region (700–1000 V) is the region of limited proportionality. In this region, the sizes of the pulses are no longer proportional to the initial number of ions formed by the collision and hence not proportional to ionizing particle energy. This region has little practical use.

The fourth region (1000–1300 V) is the well-known Geiger counter region. Here, a collision of an ionizing particle with an inert gas molecule in the GMT causes output pulses that are all of the same size, independent of n_i and the particle's energy. The mechanisms by which this mode of operation occurs have been described by Lion (1959): The initial + ion and electron formed in the collision are accelerated in the strong E-field and gain sufficient kinetic energy to cause further ionizations through collisions with inert gas molecules. The additional ions and electrons so produced form an *avalanche* of charge. Photons are emitted from the excited gas atoms in the avalanche, and the discharge forms along the center wire anode where the field strength is highest.

According to Lion (1959):

The light [photon] emission proceeds in every direction, but only in the vicinity of the central wire is the field strength high enough to facilitate the formation of new avalanches. The discharge spreads along the wire, therefore, until the entire anode is surrounded by a narrow cylinder of ions, and the discharge becomes self-sustaining.

The termination of the discharge comes about in the following manner. The positive ions surrounding the central anode move toward the cathode with a velocity which is considerably smaller (because of their mass) than that of the electrons. These ions cause a positive space charge surrounding the positive wire; as they move toward the cathode, the positive space-charge cylinder ("the virtual anode") increases in diameter and, therefore, the field strength in the counter decreases.

The recovery time in GMTs following a *count* ranges from 30 to 300 μs (Lion 1959); thus, the maximum GMT counting rate can be from 3,000 to 30,000 counts/s. The action of the quenching gas is described earlier.

The fifth region (over 1300 V) is called the *unstable region*. Here, the field strengths are so large that the quenching gas is ineffective, and one ionizing particle collision with an inert gas molecule can lead to multiple pulses or, in extreme cases, a sustained plasma discharge that may destroy the GMT.

It should be stressed that the current pulses from a GMT operated in the Geiger or plateau region are about 1000 times the size of the pulses observed when the GMT is operated in the proportionality region. The circuit of a simple, portable Geiger counter is shown in Figure 6.61. Several output modes are commonly used: an audio output through headphones or a loudspeaker, an analog meter output, and a digital counter output reading in counts per minute, or other time unit. A GMT counter system does not provide information as to what the ionizing particles or radiation is. This information must be inferred from an a priori knowledge of the source of radiation.

6.5.3 Solid-State, Crystal Radiation Sensors

Crystal radiation sensors have used crystals of materials such as lightly chlorine-doped cadmium telluride (a *p* semiconductor), thallium chloride–thallium bromide, silver chloride, silver bromide, and diamond on which metallic electrodes have been deposited on opposite surfaces. A DC potential is applied to the electrodes, generating an internal electric field in the crystal of several kV/cm. If an ionizing particle enters the crystal and collides with its atoms, a burst of free electrons is released that then drift through the crystal

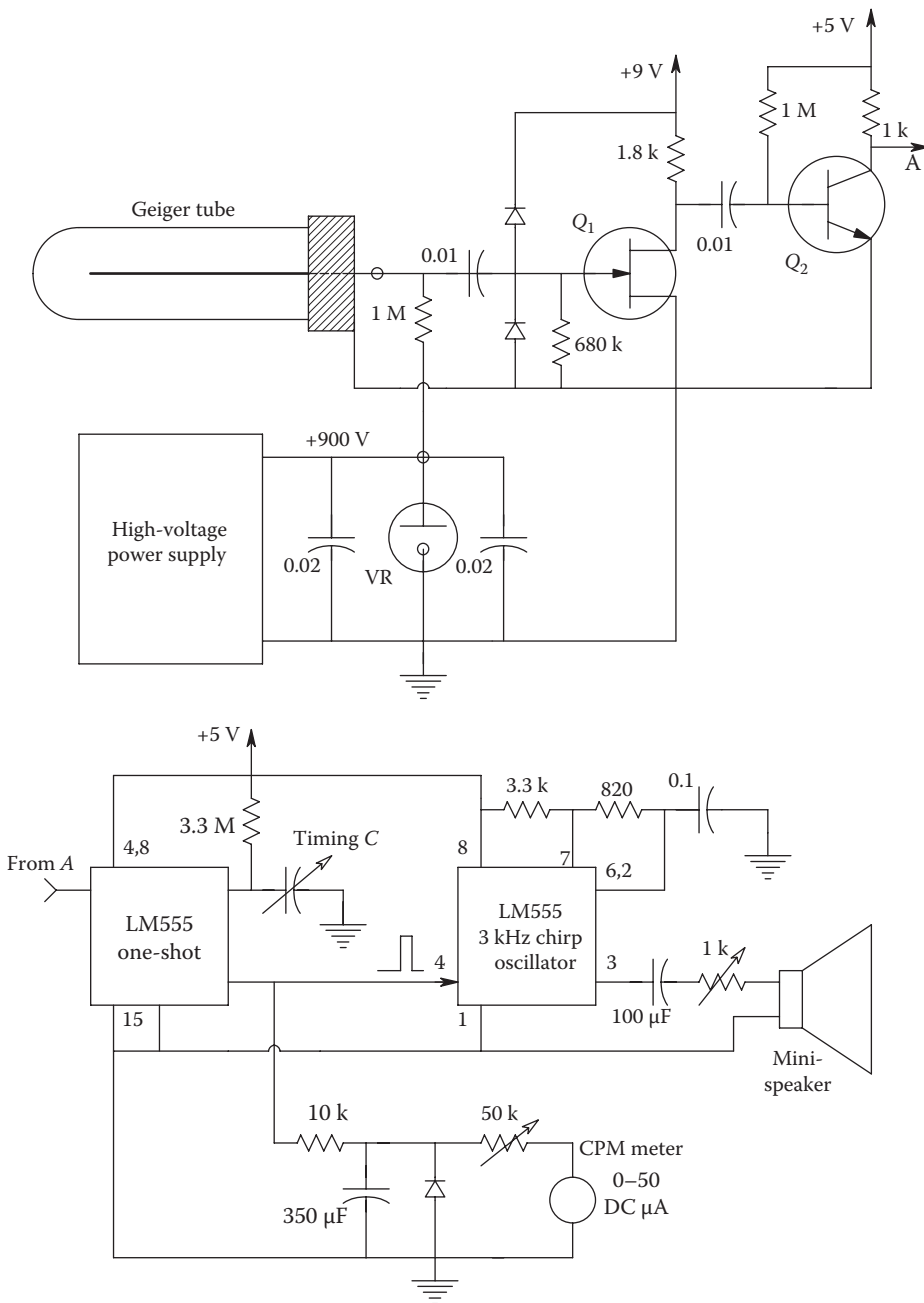


FIGURE 6.61

Circuit of a simple, analog radiation (Geiger) counter designed by the author. Voltage pulses from the GMT are conditioned by transistors Q_1 and Q_2 and used to trigger an LM555L-1, which acts as a one-shot multivibrator, producing output pulses of standard width and height. The output pulses from the one shot are averaged by a low-pass filter preceding a microammeter whose current is calibrated in PPM (pulses per minute). The one shot's output pulses also trigger a second LM555L-1 chip that generates an audio *chirp*. The chirps are sent to headphones or a small loudspeaker so the operator can be aware of abrupt changes in radiation level above background.

to the positive electrode where their arrival causes a brief current pulse. Only 1.0 eV is required in a Cl-doped CdTe crystal to put an electron into the conduction band.

Let us assume the total electron charge released by the ionizing event is n_0q coulombs (n_0 is the number of electrons released, q is the charge on one electron), and these electrons are created near the cathode and drift en masse over the thickness, d , of the crystal to the positive electrode. In their travel to the anode, some of the electrons recombine with positive charges (holes) or are trapped by impurities in the crystal such that the charge reaching the anode is given by

$$Q_A = n_0q \exp\left(\frac{-d}{\lambda}\right). \tag{6.138}$$

Thus, n_0q coulombs start to move at some velocity, v , toward the anode, decreasing in number as they travel. Because the applied electric field in the crystal is high (about 5 kV/cm) and d is small, the travel time of the electrons, $T_d = d/v$, is small compared with the RC time constant of the external circuit in Figure 6.62. Note that the collecting (right-hand) electrode is a solid disk, but the anode (left-hand) electrode is annular, providing a center hole through which the photons strike the crystal. The magnitude of the charge bundle in the crystal may be found as a function of time:

$$Q(t) = n_0q \exp\left(\frac{-tv}{\lambda}\right) \text{ for } 0 \leq t \leq T_d \tag{6.139}$$

$$Q(t) = 0 \text{ for } t > T_d.$$

We note that the peak voltage on the capacitor formed by the crystal and its electrodes is given by $\bar{V}_c = \bar{Q}/C$. \bar{Q} is the average charge flowing over the interval, $[0, T_d]$.

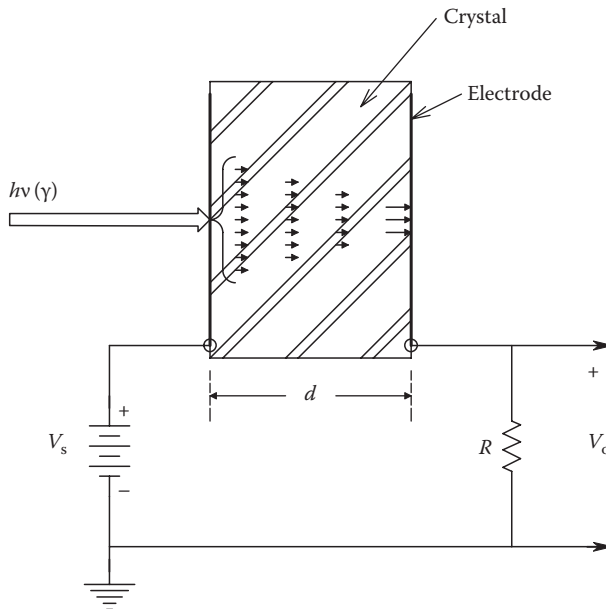


FIGURE 6.62
Schematic of a solid-state, crystal radiation sensor.

Integrating Equation 6.139 from 0 to T_d and dividing by CT_d yields the peak value of the voltage induced by the moving electrons:

$$V_c = \left(\frac{n_o q \lambda}{Cd} \right) \left[1 - \exp\left(\frac{-d}{\lambda} \right) \right]. \tag{6.140}$$

V_c decays exponentially as $v_c(t) = V_c e^{-t/RC}$. In the equations given earlier, the constant λ is known as the *Schubweg* of the crystal; this is the space constant for electron decay in the crystal, as seen from Equation 6.138. The value of the *Schubweg* depends on the crystal material and is proportional to the applied field strength. Lion (1959) stated that the crystal will behave as a good proportional counter when the *Schubweg* lies between a few tenths of a millimeter and 1 cm. He stated: "If the incident particle penetrates deeply into the crystal and causes the release of electrons uniformly throughout the crystal volume." The peak voltage is then given by

$$V_c = \left(\frac{n_o q \lambda}{Cd} \right) \left[1 - \left(\frac{d}{\lambda} \right) (1 - e^{-d/\lambda}) \right]. \tag{6.141}$$

As in the case of GMTs, there is a critical range of applied DC voltage where the pulse size is proportional to the energy of the ionizing particle. The particle energy for single ion-pair production in a crystal is of the order of 5–10 eV. Lion states that a 1 MeV beta particle causes the release of about 1.2×10^5 electrons in a AgCl crystal. The lowest detectable energy for an incident particle is in the order of 1 keV; this produces from 100 to 200 electrons. With a 5 kV/cm internal DC field, an AgCl crystal can resolve about 10^6 cps, and a diamond can count about 10^7 events/s. According to Lion, space charges may accumulate in the crystal due to the accumulation of immobile holes in the crystal lattice or from electrons trapped near the anode. Such space charges may seriously limit the counting life of a crystal, but their effect may be overcome by irradiating the crystal with IR light or periodically reversing the polarity of the DC applied to the crystal.

An interesting, simple, analog radiation sensing circuit using a *pn* silicon junction diode was described by Prazak and Scott (1975). This system was used to sense gamma radiation and high-energy x-rays. It was not used to resolve single radiation photon events, such as done by a Geiger or scintillation counter. Figure 6.63 illustrates the simple circuit.

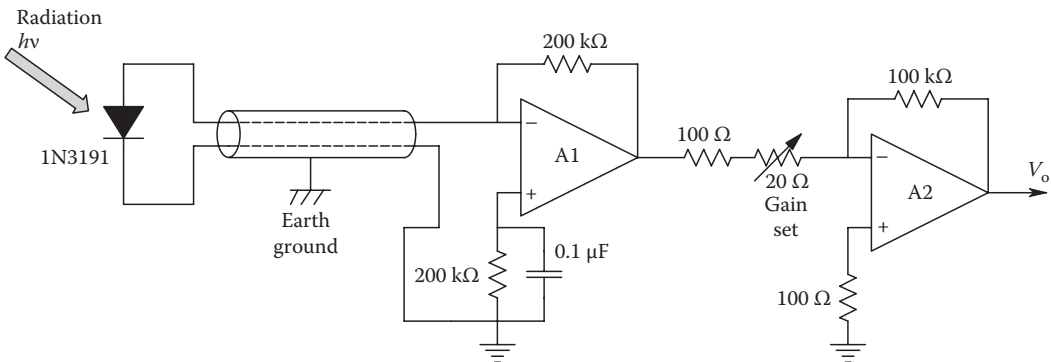


FIGURE 6.63

In this simple circuit, a *pn* diode operated at $v_D = 0$ is used as a high-energy radiation sensor. Absorbed radiation causes an increase in the reverse diode current, i_D , which, in turn, causes V_o to go positive. This radiation sensor was used to detect high levels of gamma radiation and high-energy x-rays. See text for description.

A 1N3191 silicon diode operated under zero-bias conditions was used as the high-energy photon sensor. The first-stage amplifier A1 (BB 3521L) is a low-DC drift, FET-input OA, and A2 (BB 3292) is a low-drift, chopper-stabilized OA with a maximum offset voltage drift of $\pm 0.3 \mu\text{V}/^\circ\text{C}$. The DC output voltage of the circuit was measured to be $V_o = 0.01D$, where D is the dose rate in rad/min. Thus, $D = 1000 \text{ R/m}$ gives $V_o = 10.0 \text{ V}$. Total system error was claimed to be $<1\%$ over $10 \leq D \leq 1000 \text{ R/m}$. The 20Ω variable resistor is used to calibrate the system. In the author's opinion, small capacitors in parallel with the OA feedback resistors would act to reduce HF output noise.

6.5.4 Scintillation Counters

A scintillation counter has three main components: a scintillation crystal, a PMT used to count flashes (scintillations) from the crystal, and an electronic pulse forming and pulse-height discriminating circuit. The scintillation crystal is generally sodium iodide, about one-half inch thick. The energy from ionizing radiation that collides with the crystal lattice is converted in part to a flash of light (secondary photons), the intensity of which is proportional to the energy of the absorbed radiation. The intensity of the flashes or scintillations are very weak and must be converted to electrical pulses by a PMT. The PMTs used in scintillation counters have flat, transparent photocathodes. Photons from the scintillation crystal strike the PMT's photocathode and cause the release of a number of electrons proportional to the intensity of the flash, the intensity of which is proportional to the energy of the ionizing particle. The few photoelectrons released at the photocathode are amplified by the dynode electrodes of the PMT to form a large output current pulse. The current amplification of a PMT (anode current/photocathode current) depends on the supply voltage used, the tube's geometry, etc., and can range from 10^2 to 5×10^6 (Lion 1959). Overall photosensitivity of PMTs can range from 0.01 to 2×10^4 amps/lumen. The PMT current amplification increases proportional to the square root of the supply voltage. With liquid nitrogen cooling to reduce noise, PMTs can resolve scintillations as small as 2×10^{-16} lumen. The response time of a PMT is in general less than 10 ns; hence, such tubes should be capable of counting as fast as 10^7 cps. The output pulses from PMT circuits are often reshaped to make them sharper. One such sharpening circuit, taken from Webster (1978), uses a shorted delay line to cancel the long tail on the primary output pulse, making the pulse narrow (see Figure 6.64).

Scintillation systems are widely used in nuclear medicine in conjunction with gamma-emitting isotopes to image tissues with cancerous growth. The patient swallows or is injected with a small amount of an isotope that is selectively taken up metabolically by the organ in question. The presence of a tumor generally means that more of the radioisotope will be taken up than by normal tissue. A scintillation system with a lead collimator is used to scan over the tissue in question and to detect areas of abnormal (high) radioactivity (Northrop 2002). As an example of such a system is the use of the isotope I^{131} , which has a half-life of 8.1 days. I^{131} is selectively taken up by cells in the thyroid gland in the process of the formation of the hormone thyroxin. Planar arrays of scintillation crystals and their associated PMTs are called gamma cameras and are used in nuclear medicine to visualize parts of the body that have selectively absorbed a radioisotope. This type of gamma camera provides relatively coarse resolution because of the physical size of the scintillation crystals and PMTs. In the Anger camera, a single, large scintillation crystal is used with an array of 19 PMTs. The location of a single radioactive event in the crystal is estimated by a formula involving the intensity of the flash produced at each PMT and the location of each PMT. Another configuration of scintillation camera used in nuclear medicine is the pinhole collimator. In this mode of operation, pixel resolution is traded off for sensitivity (Macovski 1983).

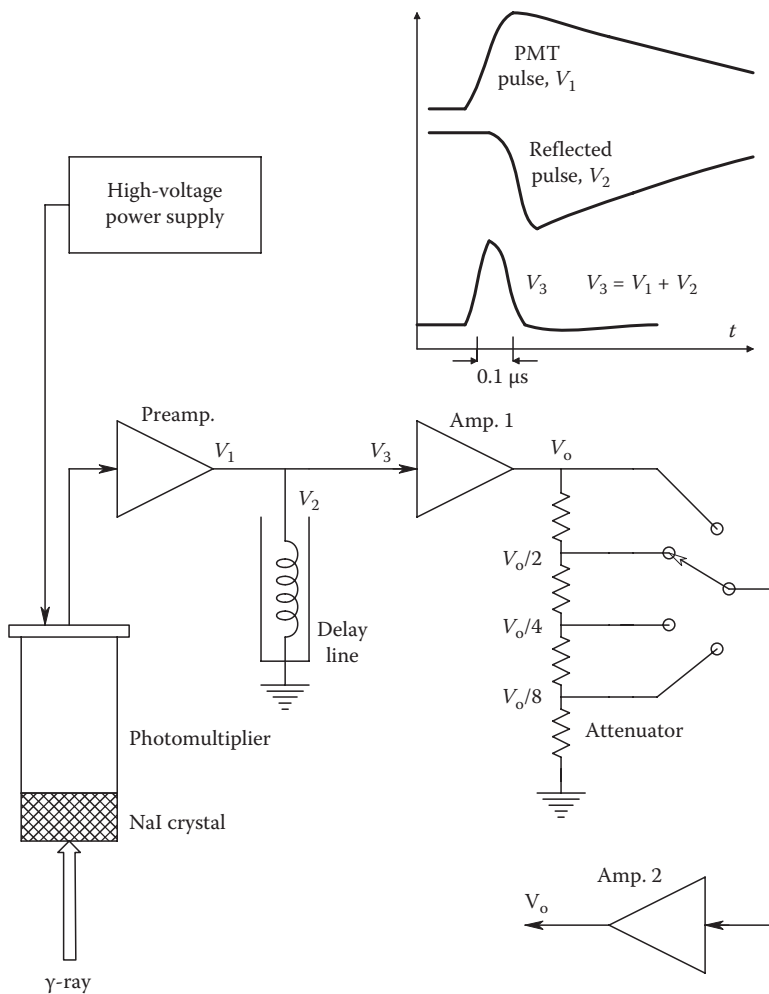


FIGURE 6.64

Schematic of a sodium iodide crystal-based scintillation radiation detector used in nuclear medicine. The energy of a light flash (scintillation) from the NaI crystal, which results from the absorption of a gamma particle, is proportional to that particle's energy. Light flash energy is converted to voltage pulses by a fast PMT and associated pulse sharpening circuitry. Electrical output pulses are scaled by height and counted to isolate the activity a specific radioisotope.

6.6 Electrochemical Sensors

6.6.1 Introduction

This class of sensor is generally used in a *wet* environment, that is, one in which the substance to be measured is dissolved in water or another solvent. The outputs of such sensors are generally currents or voltages that are functions of the concentration of the substance to be measured. Electrochemical sensors may be subdivided into three categories: those that generate EMFs as the result of the sum of two half-cell potentials, (i.e., they are

electrochemical cells), those that operate polarographically, and those that are basically fuel (redox) cells. All three categories involve electrochemical reactions. Chapter 11 treats solid-state chemical microsensors.

6.6.2 pH- and Ion-Specific Electrodes

The traditional pH measurement system is composed of two half-cells forming a *battery* or cell whose EMF is a function of pH. pH is defined as the logarithm to the base 10 of the reciprocal of the hydrogen ion concentration. In terms of an equation,

$$\text{pH} \equiv \log_{10} \left(\frac{1}{[\text{H}^+]} \right). \quad (6.142)$$

A traditional pH cell is shown in Figure 6.65. The pH half-cell electrode itself is made from a special glass membrane across which the EMF is a function of hydrogen ion concentration. The pH electrode is filled with 0.1 N HCl, and electrical contact is made internally through a silver chloride-coated silver wire (Ag/AgCl). The resistance of the glass pH electrode is quite high, around 50 megohms or more; this limits the current that may be drawn from the pH cell so that an electrometer amplifier interface must be used to measure the cell's EMF without IR drop errors. The reference half-cell electrode is traditionally a saturated calomel design, shown on the right in Figure 6.65. Undissolved potassium chloride

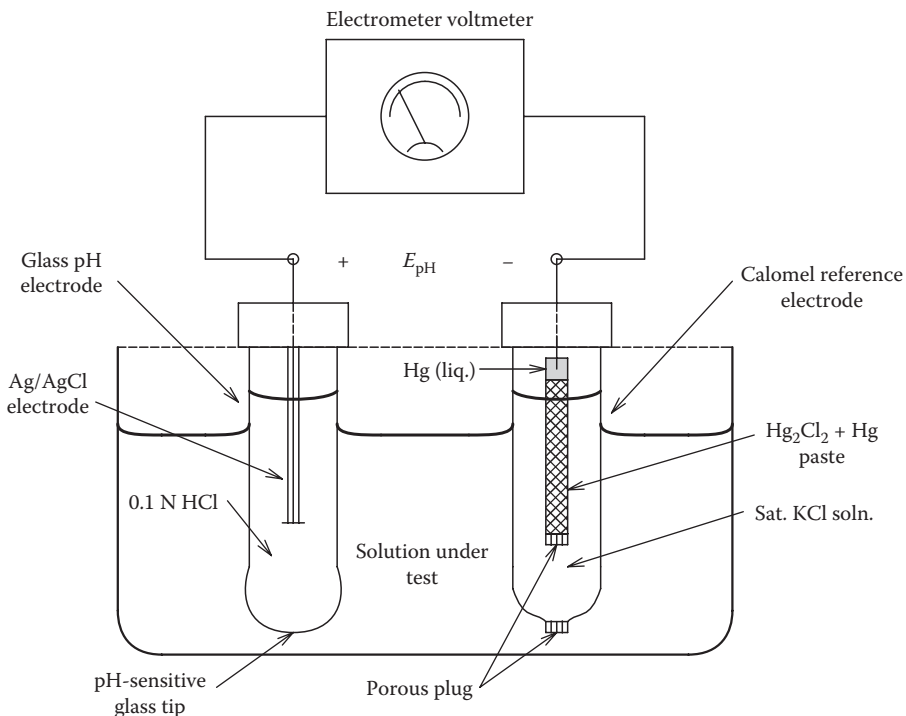


FIGURE 6.65

A basic, laboratory pH-measuring system. The pH cell's EMF is generated by the sum of the half-cell potential of a pH-sensitive glass electrode and that of a calomel reference electrode whose half-cell potential is independent of pH. No temperature compensation is shown.

crystals saturate the internal solution of the reference half-cell. The total pH cell's EMF is given by summing the half-cell EMFs:

$$E_{\text{GL}} = E_{\text{GL}}^0 - \left(\frac{RT}{\mathfrak{F}} \right) \ln(a_{\text{H}^+}) = E_{\text{GL}}^0 + \left(\frac{2.3026RT}{\mathfrak{F}} \right) (\text{pH}) \quad (6.143\text{a})$$

$$E_{\text{CAL}} = E_{\text{CAL}}^0 - 7.6 \times 10^{-4} (t - 25^\circ\text{C}) \quad (6.143\text{b})$$

$$E_{\text{NET}} = E_{\text{GL}} - E_{\text{CAL}} = (E_{\text{GL}}^0 - 0.2415) + 7.6 \times 10^{-4} (t - 25) + \left(\frac{2.3026RT}{\mathfrak{F}} \right) (\text{pH}) \text{ V} \quad (6.143\text{c})$$

where a_{H^+} is the electrochemical activity of the hydrogen ions in solution; activity may be considered to be equal to concentration at low concentrations. E_{GL} is the glass half-cell electrode's EMF, E_{CAL} is the calomel electrode's half-cell EMF, R is the MKS gas constant (8.3143 J/K), T is the temperature in Kelvin, t is the Celsius temperature, \mathfrak{F} is the Faraday number (96,500 absolute coulombs), E_{CAL}^0 is the standard potential for the calomel half-cell (0.2415 V), and E_{GL}^0 is the standard potential for the glass electrode.

E_{GL}^0 varies between glass electrodes and is compensated for when the individual pH measuring system is calibrated. Note that acidic solutions have pHs ranging from 0 to 7.0, and basic solutions' pHs range from 7.0 to 14. Also note that the pH potential, E_{NET} , is highly temperature dependent, and therefore all pH meters use a temperature probe of some sort also immersed in the solution under measurement. The signal from the temperature probe is used to null the $(t - 25^\circ)$ term in the pH cell EMF, E_{NET} , and to correct for changes in the $(2.3026RT/\mathfrak{F})$ term if t of the measurement solution differs from the temperature of the standard pH buffer solutions used to calibrate the pH meter. It should be noted that there are many physical variations on the standard glass pH electrode and the calomel reference electrode. Designs exist, which have been optimized for the measurement of blood pH, the pH of high temperature solutions, pH measurements in the stomach, etc. In addition to the common glass pH electrodes, under limited conditions, other half-cells can be used whose half-cell potentials are proportional to pH. Tungsten metal, quinhydrone, and palladium oxide electrodes are responsive to pH (Maron and Prutton 1958, Grubb and King 1980).

Specific ion electrodes are used in conjunction with a reference electrode to measure the logarithm of the activity (concentration) of a specific anion or cation in solution (Orion 2013). Specific ion electrodes are temperamental in that most have ranges of pH for optimum operation, and they also may require the complete absence of certain ion(s) that compete in the electrode's half-cell chemical reaction and therefore cause errors. We mention a few, representative, specific ion electrodes and their properties:

The Orion model 94-17 chloride electrode has a log-linear half-cell potential that increases by 59.2 mV/decade change in concentration over a 10^{-4} to 1 M range.

This electrode tolerates a pH range of 0–14, as well as nitrate, sulfate, phosphate, fluoride, and bicarbonate ions. It will not function in strongly reducing solutions nor in the presence of sulfide, iodide, or cyanide ions. The chloride level must be at least 300 times the bromide level, 100 times the thiosulfate level, and 8 times the ammonia level for proper operation.

The Orion model 94-06 cyanide ion electrode has a log-linear half-cell potential over a 10^{-6} to 10^{-2} M range. The cyanide electrode does not respond to cations except silver, and most common anions, including nitrate, fluoride, and carbonate, do not interfere with normal operation. The electrode will not work with sulfide ions in the solution. The chloride

concentration may not be more than 10^6 times the cyanide concentration, the bromide level may not be more than 5000 times the cyanide level, and the iodide level may not be more than 10 times the cyanide before error will occur in measuring the cyanide level.

The Orion model 92-19 potassium ion electrode produces a log-linear, 59.6 mV/decade concentration change of K^+ ions over a 10^{-6} to 1 M concentration range. It operates from pH 1 to 12. Selectivity constants, given as the ratio of the electrode's response to the interfering ion to the response to K^+ , are given by Orion as $Cs^+(1.0)$, $NH_4^+(0.03)$, $H^+(0.01)$, $Ag^+(0.001)$, $Na^+(2 \times 10^{-4})$, and $Li^+(10^{-4})$. Thus, we see that the potassium electrode is relatively insensitive to other metal ions but has a 1:1 ambiguity with cesium ions and is affected at the 1% level by pH.

The concentration of cadmium ions may be measured with the Orion model 94-48 electrode. The EMF of this half-cell is log-linear from 10^{-7} to 10^{-1} M $[Cd^{++}]$. The pH range is 1–14, and silver, mercury, and copper must not be present in the solution. The level of free lead or ferric ions must not exceed the cadmium ion concentration for absence of error.

The Severinghaus pCO_2 electrode described by Wheeler (in Webster 1978) makes use of the property that the pH of a buffer solution is proportional to the $\log[pCO_2]$ over the range of 10–90 mm Hg partial pressure. A Teflon membrane separates the solution in which CO_2 gas is dissolved (blood) and the buffer solution. The membrane is permeable to CO_2 gas but blocks the passage of ions such as H^+ and HCO_3^- . The pH of the Severinghaus cell may be shown to be equal to

$$pH = \log[HCO_3^-] - \log(k) - \log(a) - \log[pCO_2] \quad (6.144)$$

where

k is the equilibrium constant for CO_2 gas in water going to bicarbonate ion and hydrogen ion

a is the constant relating the CO_2 gas pressure above blood to the concentration of CO_2 dissolved in blood

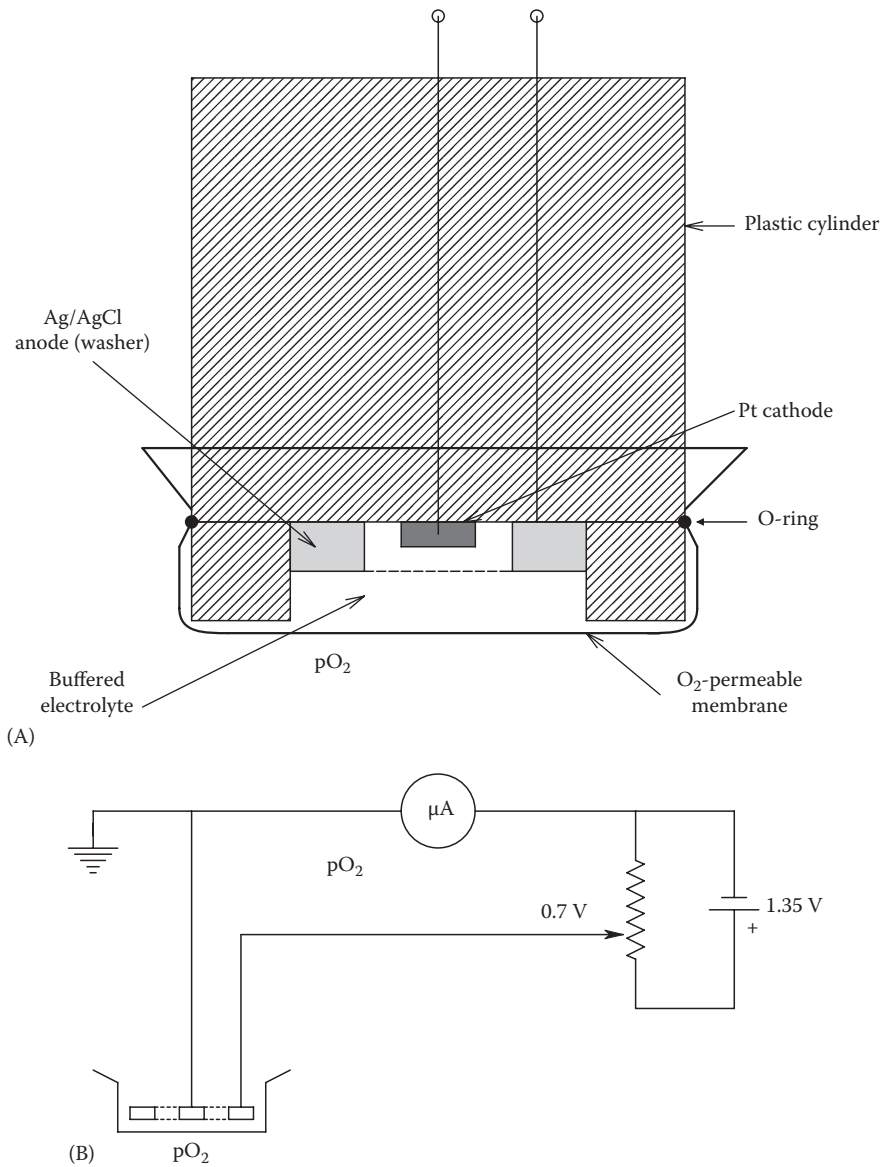
$[HCO_3^-]$ is the concentration of bicarbonate ions in the buffer solution around the pH electrodes

The Severinghaus pCO_2 electrode requires calibration with two, known partial pressures of CO_2 , and, like any chemical cell, is affected by temperature.

6.6.3 Polarographic Electrodes

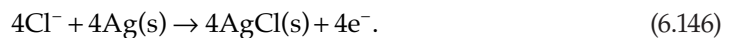
Polarographic electrodes differ from pH- and ion-specific cells in that a polarographic cell is run at a constant potential to force two electrochemical reactions to take place. The current that flows in the cell is determined by and is proportional to the limiting concentration of one reactant. The Clark cell for the measurement of the partial pressure of oxygen dissolved in water, blood, or of oxygen in the air is shown schematically in Figure 6.66A. The Clark cell has four components: a platinum or gold cathode, an Ag/AgCl anode, an electrolyte solution, and a plastic membrane (polypropylene, Teflon, or mylar), which is permeable to O_2 gas and little else. The electrolyte is generally a buffered, saturated KCl solution. At the noble metal cathode, a reduction reaction takes place, which is described as



**FIGURE 6.66**

(A) Cross section through a Clark oxygen-sensing, polarographic electrode. (B) Schematic of the basic circuit used with a Clark O_2 probe. No temperature compensation is shown.

At the Ag/AgCl anode, the *oxidation reaction*, (6.146), takes place:



It is found experimentally that when the anode is held at 0.7 V positive with respect to the cathode, a current flows, which is described by the relation

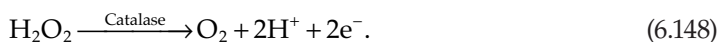
$$I = I_o + \beta[pO_2]. \quad (6.147)$$

β is typically about 10 nA/mm Hg pO₂. Both I_0 and β are increasing functions of temperature. The response time of the Clark electrode to step changes in pO₂ is limited by the ease of O₂ diffusion through the membrane to the cathode surface. Typically, in the 10s of seconds, the response time of a fast Clark O₂ electrode has been cited by Geddes and Baker (1968) as 0.3 ms. According to Wheeler (in Webster 1978), the settling time for a Clark cell is proportional to the pO₂ being measured. The basic circuit for operating a Clark probe is shown in Figure 6.66B.

The Clark cell has two major sources of error in measuring pO₂. One is the effect of temperature. The Clark cell must be calibrated and operated at a controlled, fixed temperature that varies no more than $\pm 0.1^\circ\text{C}$, or, knowing its tempco, we can measure its temperature and calculate compensation terms for Equation 6.102.

6.6.4 Fuel Cell Electrodes

A fuel cell is a battery (cell) in which the chemical reactions producing the EMF and output current occur at one or both half-cell electrodes and consume two or more reactants that are broken down to reaction products. If one of the reactants is present in excess concentration, and the other is present at a reduced concentration, then the cell's EMF may be proportional to the logarithm of the activity of the limiting reactant. A specific example of a fuel cell system is the catalyzed oxidation of glucose to form gluconic acid, hydrogen peroxide, and heat. A glucose fuel cell described by Wingard et al. (1982), used a mixture of the bioenzymes *glucose oxidase* and *catalase* immobilized on the surface of a platinum screen electrode. *Glucose oxidase* catalyzes the oxidation of glucose to gluconic acid and H₂O₂. *Catalase* catalyzes the breakdown of H₂O₂ according to the oxidation reaction



The authors state that the source of the cell's EMF is the hydrogen peroxide reaction, although firm evidence for this contention was not given. They speculate that the reaction at the platinum anode is the oxidation of H₂O₂, and the cathodic half-cell reaction is the reduction of AgCl to Ag. Wingard et al. showed that the open-circuit EMF of their fuel cell was proportional to the logarithm of the glucose concentration from 5 to 500 mg/dL. Another type of glucose fuel cell is described in Chapter 8 of Wise (1989).

It should be pointed out that there are many other electrochemical cells that have been devised to measure glucose. Some use polarographic methods in which the current through the cell is a function of the glucose concentration, others measure the decrement of pO₂ across a membrane caused by the oxidation with a Clark O₂ cell, and others measure the temperature rise in a chamber resulting from the oxidation. The interested reader should see the review of glucose sensors by Peura and Mendelson (1984).

6.7 Mechano-Optical Sensors

6.7.1 Introduction

In this section, we discuss certain sensor mechanisms, excluding FOs, in which mechanical parameters such as displacement and velocity cause changes in the intensity of transmitted or reflected light, either directly or as the result of induced changes in the linear

polarization angle of the output light or from optical interference. In the simplest case, shaft angle is sensed with an optical encoder disk. Shaft velocity can be found by differentiating the position signal.

Another means of measuring angular velocity is through the use of the *Sagnac effect*, discussed in the following. The displacement of interference fringes is proportional to the angular velocity. The Sagnac effect forms the basis for modern optical gyroscopes, which are discussed in the following chapter. While their primary response mode is to sense angular velocity, gyroscopes can be combined with electromechanical feedback systems in order to sense angular position; such systems are covered in Chapter 7. The linear velocity of a fluid medium can be measured optically by laser Doppler velocimetry (LDV), the basics of which are described in the following.

6.7.2 Optical Coding Disks

Optical coding disks are used to convert analog shaft rotation angle to a digital word proportional to the angle. Optical coding disks may be subdivided into two categories: incremental position encoders and absolute position encoders. In both types of sensors, light is either transmitted to or blocked from a linear phototransistor array, depending on the angle of the shaft to which the photoetched coding disk is attached. Thus, coding disk output is digital and may be input directly into a computer controlling the shaft angle or recording its value. Figure 6.67A shows a three-bit, absolute position encoding disk etched for a straight binary output code and Figure 6.67B illustrates a gray code etched disk. Practical resolution of an absolute position disk is 14 bits. The MSB (2^0) track is the innermost track; higher-order bits are encoded radially outward, with the LSB (2^{N-1}) lying on the outside edge of the disk. The arc length of the smallest (outer) sectors can be shown to be given by

$$l \cong R \tan\left(\frac{360}{2^{N+1}}\right). \tag{6.149}$$

Thus, for $N = 14$, and $R = 4''$, the radius to the outer track, the resolvable sector length l is $0.000767''$ or $19.5 \mu\text{m}$. This also implies that one LSB of shaft revolution is $\theta = 360/2^{N+1} = 1.099 \times 10^{-4}$ degrees. To obtain finer resolution, one only has to connect a second coding disk to the primary shaft through a step-up gear ratio of $2K$, where K is 1, 2, 3, etc., and use the outer K tracks of the second disk as the K LSBs.

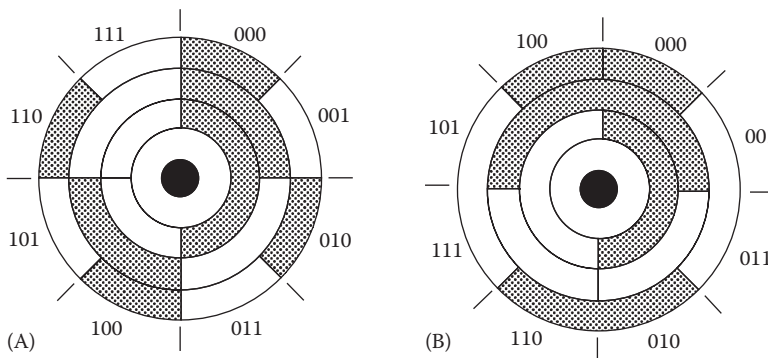


FIGURE 6.67
 (A) A three-bit, straight binary, absolute position shaft coding disk. (B) A three-bit, gray code disk.

Incremental coding disks are far simpler in their construction. On their outer edges, they have photoetched two, simple, light/dark square wave patterns, one of which is 90° out of phase with the other. Depending on the application, from 100 to 1000 or 1024 cycles may be used. Once a reference angle is established, total shaft angle can be obtained by counting the output pulses with TTL up/down counters, such as the 74LS192 or 74LS193. Comparison of the phase of the two square wave outputs of the incremental coding disk gives rotation direction and sets the up/down control on the counters.

It should be pointed out that angular velocity information can be obtained from the digital signals from optical coding disks by simple digital differentiation, in the case of absolute position encoders, and by simply counting the pulses per unit time for the output of an incremental disk, or doing digital differentiation on the position counts stored in the up/down counters.

Optical mask techniques similar to those used in angle coding disks can be used to realize direct linear analog displacement to digital coding sensors. Linear, charge-coupled photodetector arrays can also be used with an opaque mask and lens optics to measure linear displacements with a precision on the order of $1/1024$. The output of such a CCD system is basically analog, however, and the array element number at which the light/dark transition occurs gives the desired displacement.

Still another method of optical detection of linear displacement, described in Barney (1985, Sec. 7.4.1), makes use of the moving interference fringes caused by the superposition of two, linear Moiré patterns. An optical detection scheme is used that uses the relative motion of the fringe pattern to obtain direction and displacement information.

6.7.3 Sagnac Effect Sensing of Angular Velocity

The *Sagnac effect* is the principle underlying the modern FO gyroscope (Juang and Radharamanan 2009, Napolitano 2010). In this section, we will describe the basic Sagnac effect. FO gyros will be discussed in detail in the next chapter. We stress that the Sagnac effect occurs whether the light traverses a closed path in air or in a glass optical fiber loop.

The Sagnac effect was first observed in 1911 by Harress. Harress constructed a polygonal ring interferometer using glass prisms, diagrammed in Figure 6.68. His purpose was to measure the dispersion properties of glasses. When the ring was rotated, a shift in the interference fringe pattern was noted. Harress assumed that this shift was due to Fresnel–Fizeau drag. This drag effect is observed when light propagates through a linearly moving optical medium.

Sagnac performed his experiments in 1914. A mercury green line source was collimated and directed at a beam splitter (a half-silvered mirror at 45°). The two beams were directed in opposite but colinear paths in air around an optical table, as diagrammed in plan view in Figure 6.69. The half-silvered mirror served to direct the beams to a screen where the incident beams produced interference fringes. Sagnac's loop area was 866 cm^2 , and he rotated his interferometer at 2 revolutions/s (r/s). Sagnac observed a fringe shift of 0.07 fringes as the entire optical table containing mercury lamp, lenses, mirrors, and screen was rotated about an axis perpendicular to the table. Sagnac was able to show that the fringe shift was *not* due to Fresnel–Fizeau drag, and he derived a relation for the fractional fringe shift given by Equation 6.166. In 1926, Pogany repeated Sagnac's experiment with improved apparatus in which $\lambda_0 = 546 \text{ nm}$, $A = 1178 \text{ cm}^2$, and $\dot{\phi} = 157.4 \text{ r/s}$. He measured $\Delta z = 0.906$; the theoretical value was 0.900. Pogany repeated his experiment in 1928 and obtained results within 1% of theoretical values.

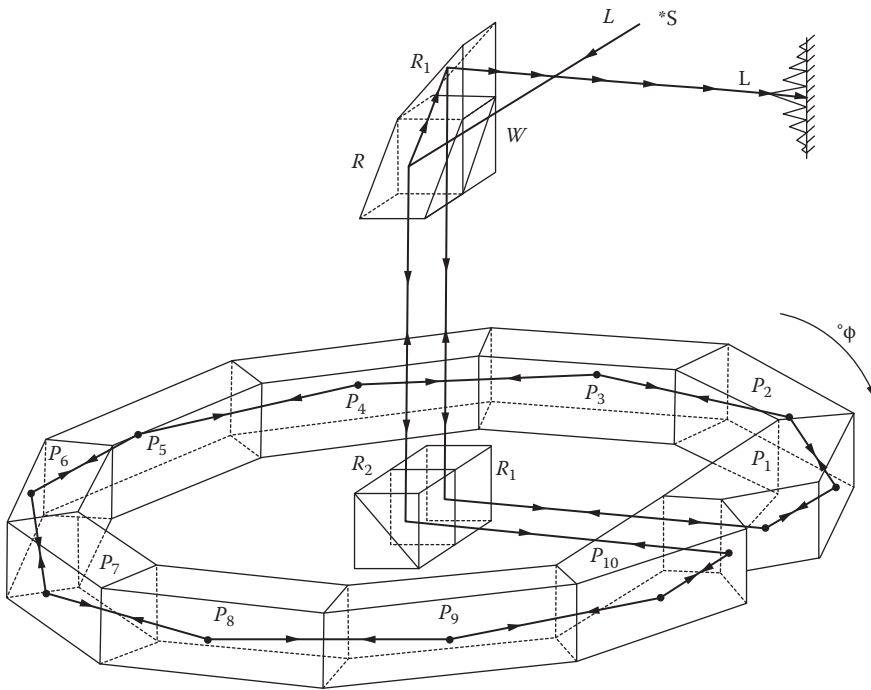


FIGURE 6.68 Harress' ring interferometer (a Sagnac system). Two, counter-rotating light beams were used.

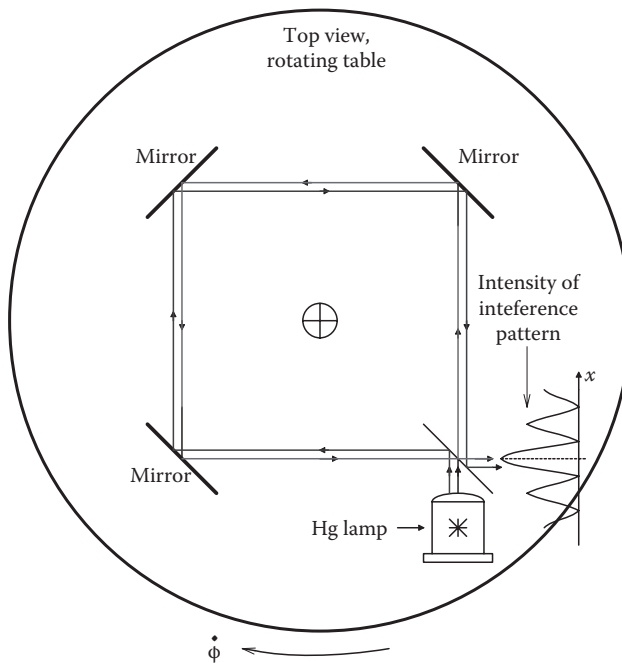


FIGURE 6.69 Top view of Sagnac's ring interferometer in which fringe shifts are proportional to the angular velocity ϕ of the apparatus around the center of the table.

From the early work on the Sagnac effect, four principles emerged: (1) Equation 6.166 applies for a one loop system; (2) The fringe shift does not depend on the shape of the closed light path, *only on its area*; (3) The fringe shift does not depend on the axis of rotation; it does depend on the path tilt, θ , with respect to the plane of the optical path, however; and (4) The fringe shift does not depend on the propagation medium. The Sagnac equation for fringe shift is

$$\Delta z = \frac{[4\dot{\phi}A \cos(\theta)]}{(\lambda_0 c)} \tag{6.150}$$

where

Δz is the fractional fringe shift

λ_0 is the wavelength of light

A is the area enclosed by the light path

θ is the angle the rotation axis makes with the perpendicular to the plane of the light path

c is the velocity of light around the path

$\dot{\phi}$ is the angular velocity of path rotation around an axis at an angle θ to a perpendicular to its plane, in rad/s

It is easy to derive Equation 6.166 if we consider a *circular light path* rather than a rectangular one. Such a circular path, shown in Figure 6.70 could be the limiting result of using

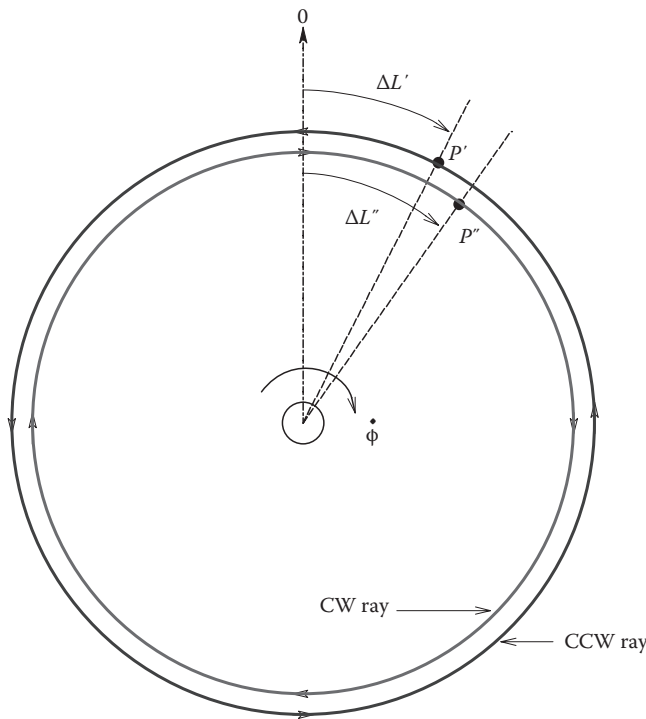


FIGURE 6.70

Rays on a single optical fiber ring, relevant to the derivation of the Sagnac gyro equation.

a very large number of prisms or mirrors, as shown in Figure 6.68 (Harress' polygonal approximation), or of using a single circular turn of a glass optical fiber.

Assume that at $t = 0$, the input/output point of the ring is at 0° . First, we assume that the ring is stationary, that is, $\dot{\phi} = 0$. In this case, it takes $\tau_o = 2\pi R/c$ s for a light wave to circle the ring, either clockwise or counterclockwise.

Now, let the ring rotate clockwise at a constant angular velocity, $\dot{\phi}$. In the time, τ_{cw} , a fixed point on the ring rotates an arc length $\Delta L''$ to point P'' . This may be expressed as

$$\tau_{cw} = \frac{\Delta L''}{R\dot{\phi}}. \quad (6.151)$$

In the same time, the clockwise lightwave travels

$$\tau_{cw} = \frac{(2\pi R + \Delta L'')}{c}. \quad (6.152)$$

Obviously, $\tau_{cw} > \tau_o$. A wave propagating counterclockwise takes less time to reach point P' : Sic:

$$\tau_{ccw} = \frac{(2\pi R - \Delta L')}{c} = \frac{\Delta L'}{R\dot{\phi}}. \quad (6.153)$$

The fractional fringe shift, Δz , is given by

$$\Delta z = \frac{c\Delta\tau}{\lambda_o} \quad (6.154)$$

where $\Delta\tau$ is given by

$$\Delta\tau = \{\tau_{cw} - \tau_{ccw}\} = \left\{ \frac{(2\pi R + \Delta L'')}{c} - \frac{(2\pi R - \Delta L')}{c} \right\} = \frac{(\Delta L'' + \Delta L')}{c} \quad (6.155)$$

Now it is easy to show that

$$\Delta L'' = \frac{2\pi R^2 \dot{\phi}}{(c + R\dot{\phi})} = \frac{(2\pi R^2 \dot{\phi}/c)}{(1 + R\dot{\phi}/c)} \approx \frac{2\pi R^2 \dot{\phi}}{c}. \quad (6.156)$$

Similarly,

$$\Delta L' = \frac{2\pi R^2 \dot{\phi}}{(c - R\dot{\phi})} = \frac{(2\pi R^2 \dot{\phi}/c)}{(1 - R\dot{\phi}/c)} \approx \frac{2\pi R^2 \dot{\phi}}{c}. \quad (6.157)$$

So the fractional fringe shift can be written finally as

$$\Delta z = \frac{(\Delta L'' + \Delta L')}{\lambda_o} = \frac{4\pi R^2 \dot{\phi}}{(c\lambda_o)} = \frac{4A\dot{\phi}}{(c\lambda_o)} \quad (6.158)$$

which is the same as Equation 6.166, with $\cos(\theta) = 1$.

As an example, the basic sensitivity of a one-turn Sagnac system can be calculated from Equation 6.166. θ is assumed to be 0° :

$$\frac{\Delta z}{\dot{\phi}} = \frac{4A}{(c\lambda_o)} = \frac{4 \times 1 \text{ m}^2}{(3 \times 10^8 \text{ m/s} \times 4 \times 10^{-7} \text{ m})} = 0.033 \text{ fringe/r/s.} \quad (6.159)$$

Clearly, sensitivity will be raised if multiple optical paths are used, a situation addressed in the discussion of FO gyroscopes in the next chapter.

6.7.4 Laser Doppler Velocimetry

LDV provides a *no touch* means of measuring the linear velocity of fluids (and particles). As the name suggests, LDV makes use of the Doppler effect at optical frequencies. In this section, we will review the basics of LDV and discuss certain key applications.

First, we describe the Doppler effect. We are all familiar with the Doppler effect on sound. A car moving toward us blows its horn. As it passes, there is a perceptible downward shift in the pitch of the horn. Johann Christian Doppler gave a paper, "On the colored light of double stars and some other heavenly bodies," in 1842 before the Royal Bohemian Society of Learning. Doppler was Professor of elementary mathematics and practical geometry at the Prague State Technical Academy. He apparently got little recognition for his work and died of consumption in 1854 at the age of 45. A contemporary of Doppler, Buys Ballot, in 1844 contested Doppler's theory as an explanation for the color shift of binary stars. Ballot actually did an experiment using sound waves where a trumpet player played a constant note while riding on a flatcar of a train moving at constant velocity. A musician with perfect pitch, standing at trackside, perceived the trumpet tone to be a half-tone sharp as the train approached and a half-tone flat as it receded. In spite of this evidence, Ballot continued to object to Doppler's theory. Ballot's publications apparently served to discredit Doppler for a number of years.

To derive the Doppler effect for either EM or sound waves, we assume a moving, reflecting target and a stationary source/observer, as shown in Figure 6.71. Assume that the sinusoidal waves leaving the stationary transmitter (TRX) propagate at velocity, c , over a distance, d , to the target, T . The target is moving at velocity, V , toward the source at an angle θ . Velocity V can thus be resolved into a component parallel to the line connecting TRX and the reflecting target and a component perpendicular to the TRX-T line. These components are $V \cos(\theta)$ and $V \sin(\theta)$, respectively. The reflected wave from T propagates back to the stationary receiving transducer, RCX, along path d . The receiving sensor output waveform can be written as

$$V_o = B \sin[\omega_o t + \psi]. \quad (6.160)$$

The transmitted radian frequency is ω_o , and ψ represents the phase lag between the transmitted signal and the received signal. In general, the phase lag, ψ , is given by

$$\psi = \frac{2\pi 2d}{\lambda} = \frac{2(2\pi)d}{(c/f_o)} = \frac{\omega_o 2d}{c} \text{ rad.} \quad (6.161)$$

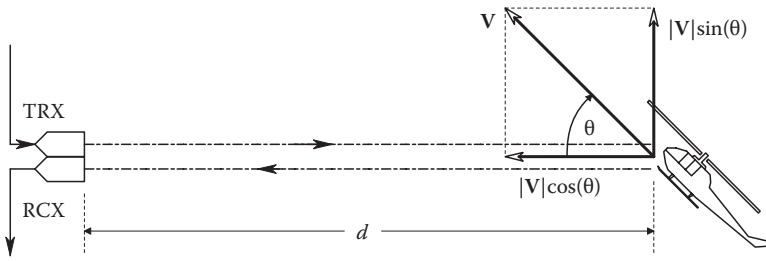


FIGURE 6.71 A simple Doppler system with stationary, collinear transmitter (TRX) and receiver (RCX) directed at a moving object with vector velocity, \mathbf{V} .

However, the distance $2d$ is changing because of the target velocity component along the line from the transducers to the target. Thus, the frequency of the received signal, ω_r , is the time derivative of its phase:

$$\omega_r = \frac{d[\omega_0 t + \omega_0 2d/c]}{dt} = \omega_0 \left(1 + \left[\frac{2}{c} \right] \dot{d} \right) = \omega_0 \left(1 + \left[\frac{2V}{c} \right] \cos(\theta) \right). \quad (6.162)$$

The frequency, ω_D , is the Doppler frequency shift that contains the velocity information:

$$\omega_D = \left(\frac{\omega_0 2V}{c} \right) \cos(\theta) \text{ r/s.} \quad (6.163)$$

Note that $\omega_r > \omega_0$, because in this example, the target is approaching the source/sensor. In LDV systems, the wave return path is seldom collinear with the path from the source to the reflecting object. A basic LDV system geometry is shown in Figure 6.72. The source, S , emits light with wavelength, λ_s , that propagates toward reflecting particle, P . Unit vector, \mathbf{i} , is directed from S to P . P moves with velocity, \mathbf{V} . It is assumed that $(|\mathbf{V}|/c)^2 \ll 1$, so that relativistic effects are negligible. If $\mathbf{V} = 0$, the number of wave fronts striking the particle per unit time is $f_s = c/\lambda_s$. (f_s for a HeNe laser is $3 \times 10^8 / 632.8 \times 10^{-9} = 4.741 \times 10^{14}$ Hz.) The number of wave fronts hitting a moving particle, P , is

$$f_p = \frac{(c - \mathbf{V} \cdot \mathbf{i})}{\lambda_s} \text{ Hz.} \quad (6.164)$$

Note that the vector dot product gives the particle velocity component parallel to \mathbf{i} . The wavelength λ_p apparent to the particle is

$$\lambda_p = \frac{c}{f_p} = \frac{(\lambda_s c)}{(c - \mathbf{V} \cdot \mathbf{i})} = \frac{(\lambda_s c)}{[c - V \cos(\theta)]}. \quad (6.165)$$

Now, the unit vector from the instantaneous particle position, P , to the stationary receiver, R , is \mathbf{r} . An observer at R sees a scattered wavelength, λ_r , from P

$$\lambda_r = \frac{(c - \mathbf{V} \cdot \mathbf{r})}{f_p} = \frac{[c - V \cos(\theta)]}{f_p}. \quad (6.166)$$

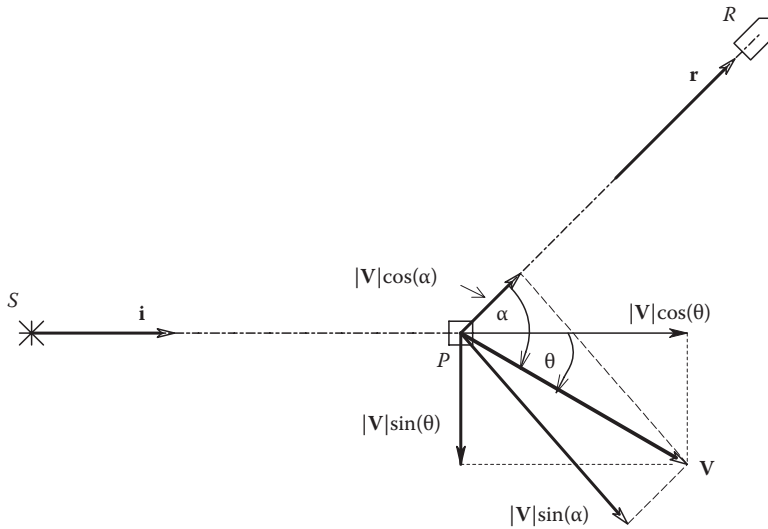


FIGURE 6.72

2D geometry of a Doppler system used in LDV. Note that the transmitter laser (S) and the receiver (R) are at different angles with respect to the object's velocity vector, \mathbf{V} .

Here again, $\mathbf{V} \cdot \mathbf{r}$ represents the velocity component of the particle in the \mathbf{r} direction. Hence, the frequency of the received, scattered radiation is

$$f_r = \frac{c}{\lambda_r} \text{ Hz.} \quad (6.167)$$

Substituting from Equation 6.166 above, we find

$$f_r = \frac{c(c - \mathbf{V} \cdot \mathbf{i})}{[\lambda_s(c + \mathbf{V} \cdot \mathbf{r})]}. \quad (6.168)$$

Now the observed Doppler frequency shift is

$$f_D = f_r - f_s = \left(\frac{c}{\lambda_s} \right) \left[\frac{c - \mathbf{V} \cdot \mathbf{i}}{c - \mathbf{V} \cdot \mathbf{r}} - 1 \right]. \quad (6.169)$$

For $V \ll c$, Equation 6.185 reduces to

$$f_D \cong \frac{\mathbf{V} \cdot (\mathbf{r} - \mathbf{i})}{\lambda_s} = \frac{V[\cos(\theta) - \cos(\alpha)]}{\lambda_s} \text{ Hz} \quad (6.170)$$

where

- θ is the angle between \mathbf{V} and \mathbf{i}
- α is the angle between \mathbf{V} and \mathbf{r}

Note that if the reflected radiation is directed back along \mathbf{i} to R coincident with S , the angle $\alpha = 180^\circ + \theta$. Because $\cos(180^\circ + \theta) = -\cos(\theta)$, f_D given earlier reduces to the value given by Equation 6.179.

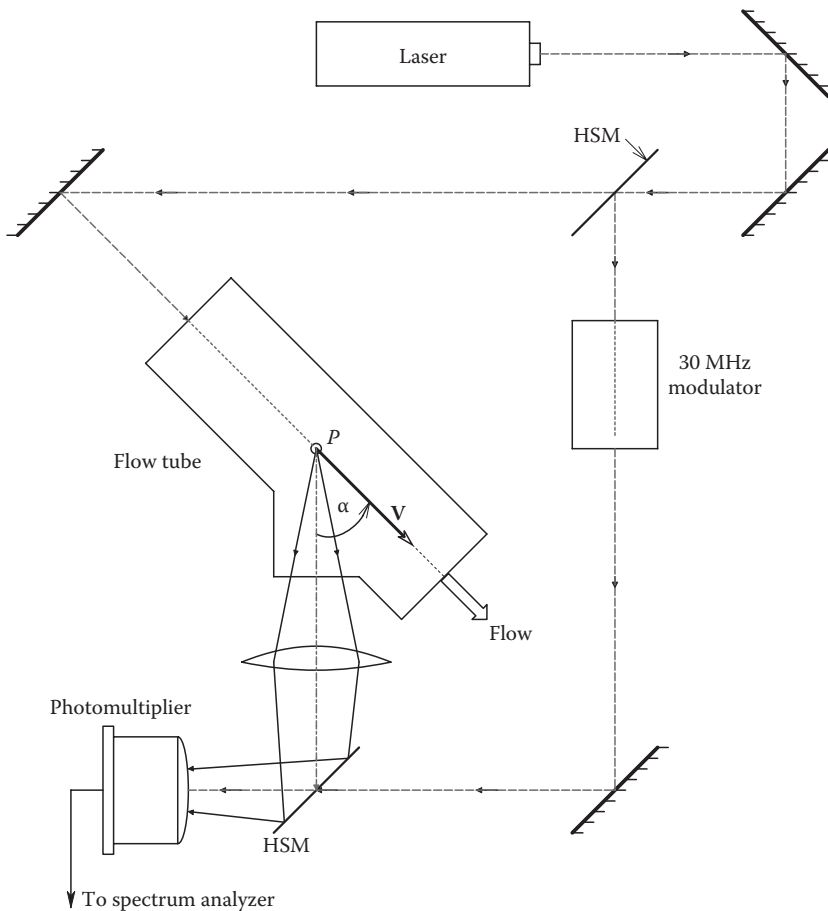


FIGURE 6.73

Diagram of the LDV system used by Yeh and Cummins (1993) to first measure fluid velocity by Doppler shift. In their system, $\alpha = 30^\circ$.

Figure 6.73 shows the basic system first used for LDV by Yeh and Cummins (1964). A HeNe laser was used with $\lambda_s = 632.8$ nm. $0.5 \mu\text{m}$ diameter polystyrene spheres were used in water at a density of 1:30,000 by volume to obtain scattering. Sphere velocity is assumed to be water velocity. The laser reference beam was frequency shifted by 30 MHz using a Bragg cell acousto-optic modulator. The measurement beam was directed parallel to the fluid velocity vector, \mathbf{V} ; hence, $\theta = 0^\circ$. The scattered, Doppler-shifted beam was taken off at 30° relative to the angle of \mathbf{V} and passed through a half-silvered mirror and directed to the photocathode of a PMT, colinear with the frequency-shifted reference beam. Optical mixing occurs at the surface of the PMT photocathode. Optical mixing is a process analogous to square-law (heterodyne) detection in a radio, where two sine waves of different amplitudes and frequencies (f_s and f_r) are added together, then squared. This process produces sinusoidal terms at the PMT output with frequencies $2f_s$, $2f_r$, $(f_s + f_r)$, $(f_r - f_s)$, and 0 (DC). Of interest in LDV is the difference frequency term. In the Yeh and Cummins system, $f'_s = f_s + 30$ MHz, so the difference term equals 30 MHz for $\mathbf{V} = 0$. The Doppler frequency shift, f_D , is added to the 30 MHz center frequency in the Yeh and Cummins system, which was used to measure the parabolic laminar flow profile in water. They could measure

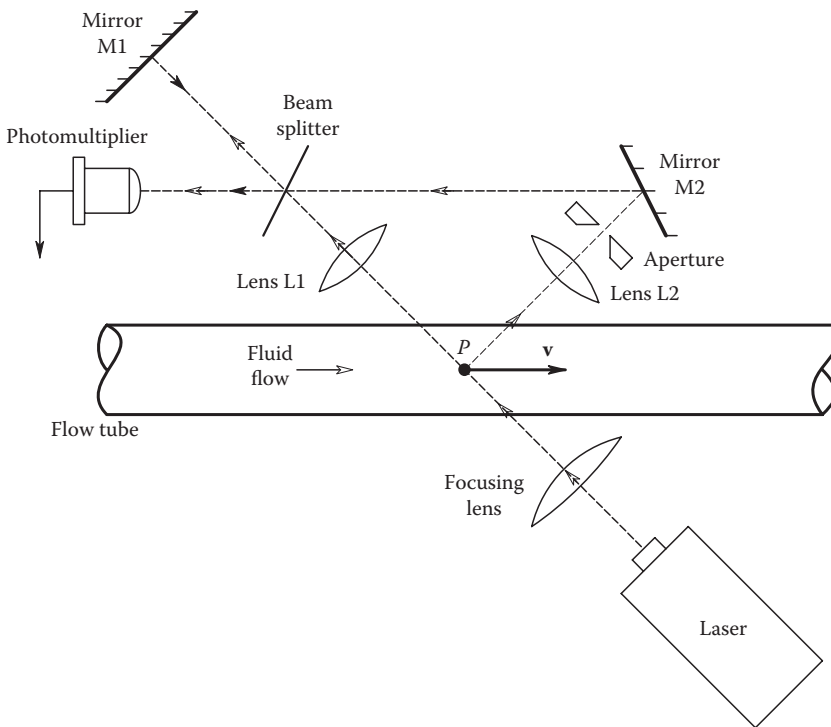


FIGURE 6.74

The LDV system of Foreman et al. (1993a,b). This is a so-called single-beam anemometer.

velocities as low as 0.007 cm/s, corresponding to a Doppler frequency shift of 17.5 Hz. The maximum velocity they measured was 0.05 cm/s, corresponding to a Doppler shift of 125 Hz. The minimum detectable f_D was 10 Hz, or one part in 10^{14} .

Many architectures for LDV systems have been described. See, for example, texts by Watrasiewicz and Rudd (1976) and Thompson and Stevenson (1979). Figure 6.74 illustrates a *single-beam laser anemometer* described by Foreman et al. (1993a,b). The laser is focused on a microregion of the liquid whose velocity is to be measured. Part of the light is scattered and forms the signal beam; part is transmitted to form the reference beam. Through the use of lenses, slits, mirrors, and a half-silvered mirror, the signal and reference beams are combined at the surface of the PMT's photocathode. This system was effectively used to measure both laminar flow in a liquid and the flow of air. No particles needed to be added to the liquid, but smoke particles were required to obtain adequate signal strength in the air measurements. Air velocities of up to 1000 fps were measured, corresponding to $f_D = 33$ MHz. Foreman et al. found that the signal energy due to scattering was several hundred times stronger for small scattering angles in the forward direction, rather for light scattered back along the input path. It should be noted that the Foreman system design that picks off the signal beam at a shallow forward angle suffers from the disadvantage that the laser and the PMT are on opposite sides of the fluid flow path.

The dual-beam LDV design, shown in Figure 6.75A, was first reported by Goldstein and Hagen (1993). Note that the laser output (reference) beam is perpendicular to \mathbf{V} and passes

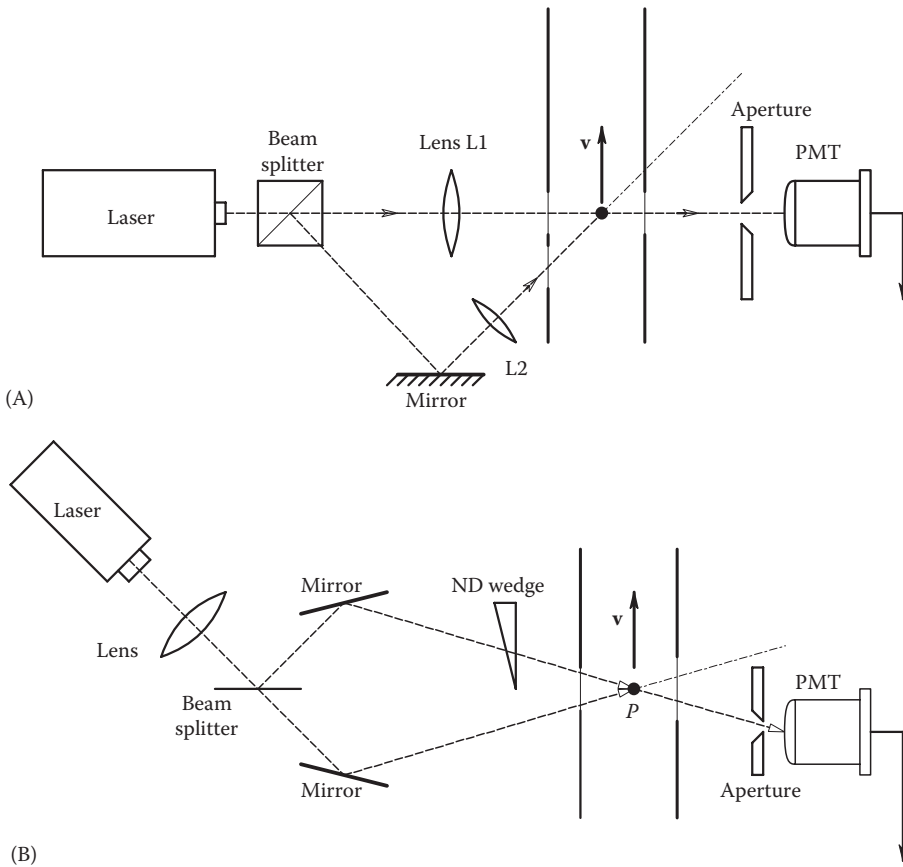


FIGURE 6.75 (A) A dual-beam LDV system described by Goldstein and Hagen (1967). (B) Another dual-beam LDV system described by Goldstein and Kreid (1967).

directly to the PMT through aperture A . The signal beam irradiates the scattering particle at angle θ . Some of the scattered signal light also passes through aperture A and mixes with the reference beam at the PMT photocathode. Observe that the dual-beam LDV is similar to the single-beam system except that the laser and the PMT are interchanged. Goldstein and Kreid (1993) used this system to measure the air flow profile at the inlet of a square duct (Figure 6.75B).

There are many other considerations in the design of LDV systems that we will only touch upon; space considerations do not permit us to treat them in detail. First, we have noted that the detected LDV signal originates from the interaction of particles suspended in the moving fluid with the laser beams. Particle size, density, and concentration in the fluid medium are important considerations in LDV system design. Particles are generally assumed to be randomly distributed in the sampling volume. Under conditions of low Reynolds number flow (laminar flow, no turbulence), one can generally assume that the particle velocity is the same as the fluid velocity. However, in oscillating or turbulent flows, particles generally do not follow the fluid without some lag, creating measurement errors. As an example of the error caused by particle lag, consider the force on a smoke

particle in moving air. From the Stokes drag formula (Watriasiewicz and Rudd 1976), we can write the drag force on the particle in one direction as

$$F_d = 3\eta d(v_p - v_f) \quad (6.171)$$

where

- η is the dynamic viscosity of the gas
- d is the diameter of the smoke particle
- v_p is its absolute velocity
- v_f is the velocity of the fluid

The Newtonian equation of motion for a particle of density ρ is thus

$$F = ma = \left(\frac{\pi d^3}{6} \right) \rho \dot{v}_p = -3\eta d(v_p - v_f). \quad (6.172)$$

Equation 6.188 is an ordinary, first-order, linear differential equation in v_p , and if there is a step change in fluid velocity of size v_{f0} , then the ODE's solution is

$$v_p(t) = v_{f0} \left[1 - \exp\left(\frac{-t}{\tau}\right) \right]. \quad (6.173)$$

The particle time constant is

$$\tau = \frac{(\pi d^2 \rho)}{(18\eta)} \text{ s}. \quad (6.174)$$

This analysis implies that the smoke particles lag behind gas velocity changes with first-order dynamics. (The situation is actually far more complicated.) A 1 μm diameter smoke particle with a density of 2.5 will have a τ of 20 μs and will follow velocity fluctuations well at frequencies up to about 8 kHz. Some other particles used in LDV are silicone oil aerosols in air (density difference 900), magnesium oxide in flame (density difference 1.8×10^4), PVC spheres in water (density difference 1.54), and polystyrene spheres in water (density difference 1). A 0.8 μm diameter particle with a relative density of 10^3 will follow 10 kHz velocity fluctuations with 1% error.

The calculation of the minimum laser power to achieve a given SNR in an LDV system is ordinarily quite complex, and we refer the interested reader to the many texts on LDV. Watriasiewicz and Rudd (1976) derived expressions for the SNR at the photosensor as a function of incident power. For example, a PMT photosensor will give a 40 dB SNR for 10 μW power incident on the detector. The same 10 μW incident power gives a 50 dB SNR for PIN and APD sensors. At incident light powers below 10^{-7} W, the PMT generally has a better SNR than the photodiodes.

LDV systems are used in a wide variety of applications ranging from meteorology (studies of wind shear, clear air turbulence, tornadoes, etc.), studies of air flow in automotive streamlining, respiratory physiology, wind tunnel studies in aircraft design, investigation of laminar flows with recirculation and turbulent flow in ducts, studies of combustion dynamics in furnaces and gas turbines, and investigations of blood flows.

6.8 Chapter Summary

In this chapter, we have examined various basic physical mechanisms underlying the operation of sensors and sensor systems used in making mechanical, chemical, physical, and electrical measurements. In some cases, we have also described certain sensor applications in order to give a better understanding of the sensor mechanisms.

Many sensors are seen to work through the QUM causing changes in resistance or capacitance. Other sensors involve the generation of voltages through the separation or release of charges (electrons, holes, ions); these include piezoelectric, pyroelectric, Faraday, and Hall effect sensors. Still other sensors make use of certain physical wave effects, for example, the Doppler and the Sagnac effects.

Electrochemical sensors are seen to involve ion-specific chemical half-cells (batteries), polarographic electrochemical reactions, such as used in the Clark oxygen cell, and fuel cell (redox) reactions. The outputs of electrochemical sensors are voltages or currents proportional to the concentration of the QUM.

Radiation sensors were shown to be basically event-indicating devices, where the radiation event generates ionized atoms, molecules, or photons that are then detected by a variety of means, optical (by scintillation) or electrical (transient currents resulting from ionization in crystals or gasses). PMTs and CPMs were shown in Section 6.8 to convert EM radiation (NIR through x-rays) to electron currents and are able to count single photons under special conditions.

Our survey of sensor mechanisms in this chapter has focused on the more important principles whereby physical quantities are converted to electrical signals. Our listing is by no means exhaustive. Chapter 7 deals with the various means by which we can measure specific physical quantities, using various types of transducers, devices, and systems.

Problems

- 6.1 Two, wire strain gauges are bonded to the outside of a SCUBA air tank in order to measure the circumferential strain when it is filled. With atmospheric pressure in the tank (zero strain), the circuit of Figure P6.1 is balanced so $V_o = 0$ V with all resistors $R = R_{27} = 300 \Omega$:
- Find an expression for $V_o = f(R, V_B, \Delta R)$ for the circuit. Assume the OA is ideal.
 - When the tank is pressurized to 2500 psi, it is strained; also, its temperature rises to 67°C due to adiabatic air compression. Calculate the tank's strain, ϵ , at 2500 psi filling pressure. Assume: $\Delta R = R_{27}[eGF + \alpha(T - 27)]$, $V_o = -35.4$ mV. The excitation voltage is $V_B = 5$ V; the strain gauges have $GF = 2.9$, and $\alpha \text{ temp-cos} = 1 \times 10^{-4}$.
 - The air in the tank is now rapidly exhausted to atmospheric pressure. Due to air expansion, the tank temperature falls below 27°C . The output voltage is now $V_o = +10$ mV. Find the tank's temperature.

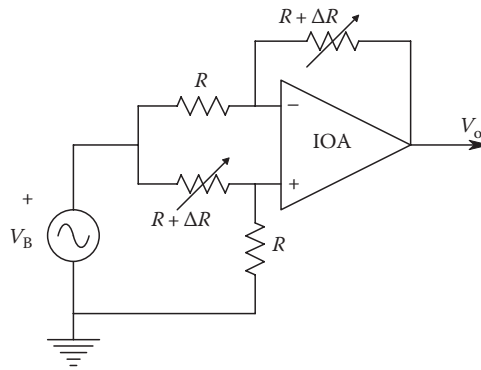


FIGURE P6.1

- 6.2 A thin, intrinsic silicon slab is used as a PC to sense 633 nm laser light intensity. The slab measures $0.5 \times 0.05 \times 0.001$ cm. The light hits the 0.5×0.5 cm surface. Electrodes are attached to the two 0.5×0.001 cm surfaces. One electrode goes to a -10 Vdc supply and the other to the summing junction of an ideal OA, shown in Figure P6.2:
- Calculate the dark current through the Si PC. The resistivity at room temperature for Si is $\rho = 2.30 \times 10^3$ ohm cm. Find $V_{o(\text{dark})}$.
 - What absorbed power (radiant flux), P_i , at 633 nm light in watts, on the PC will give a photocurrent equal to the dark current? Assume: $\mu_n = 1500$ cm²/(V·s), $\mu_p = 450$ cm²/(V·s), $\tau_p = 10^{-4}$, $\lambda = 6.33 \times 10^{-5}$ cm, $q = 1.60 \times 10^{-19}$ Cb, $\eta = 0.8$ (efficiency),

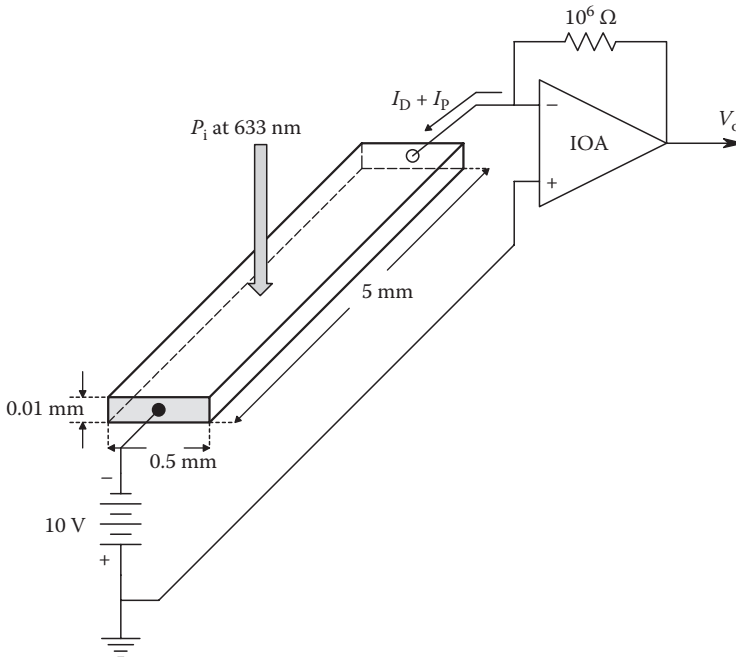


FIGURE P6.2

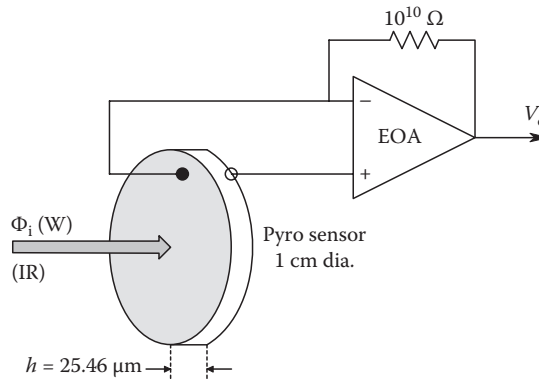


FIGURE P6.3

$c = 3 \times 10^{10} \text{ cm/s}$, $h = 6.624 \times 10^{-34} \text{ J} \cdot \text{s}$ (Planck's constant). What radiant intensity in W/cm^2 does this absorbed power correspond to? The PC's photoconductance is given by

$$G_p = \frac{q\eta\tau_p(\mu_p + \mu_n) [P_i\lambda]}{L^2 [hc]} \text{ S.}$$

C. Show how you can modify the circuit so that V_o will be zero in the dark.

- 6.3 A pyroelectric IR sensor is connected in the short-circuit current mode to an EOA as shown in Figure P6.3. The sensor is irradiated with a rectangular pulse of Φ_o watts IR power for T seconds. The Laplace transform of the input pulse is

$$\Phi_i(s) = \frac{\Phi_o(1 - e^{-sT})}{s}.$$

The sensor is 1 cm in diameter, $h = 25.46 \mu\text{m}$ (thickness), $K_p = 60 \mu\text{Cb}/(\text{m}^2 \text{K})$ (pyroelectric constant), $c = 2.34 \times 10^6 \text{ J}/(\text{m}^3 \text{K})$ (specific heat), and $\Theta = 200 \text{ K/W}$ (thermal resistance):

- A. Find the sensor's thermal time constant (note units).
 - B. Find an expression for $v_o(t)$. Plot and dimension $v_o(t)$. Let $\Phi_o = 100 \text{ nW}$, $T = 3 \text{ s}$, $R_F = 10^{10} \Omega$.
- 6.4 An LZT piezoelectric disk is used to make a *crystal* microphone. The LZT disk has an effective area of 1 cm^2 , a compression piezostrain coefficient of $d_{33} = 320 \text{ pCb/N}$, $\kappa = 1500$ (dielectric constant), and $h = 3 \text{ mm}$ (thickness). The total leakage conductance of the sensor = 10^{-11} S :
- A. Calculate the capacitance of the sensor, C_x . Assume a simple parallel-plate capacitor model.
 - B. Draw the equivalent circuit of the microphone. Calculate the peak open-circuit voltage across the microphone alone when a 1 kHz sinusoidal sound wave with peak pressure of 150 dyn/cm^2 impinges on the crystal.
 - C. Now, the microphone is connected to the EOA circuit as shown in Figure P6.4. Ten feet of coax is used that adds 300 pF in parallel to the sensor's C_x . Calculate the peak output voltage of the amplifier for the same sound input. Let $R_F = 10^{10} \Omega$, $C_F = 10^{-10} \text{ F}$.

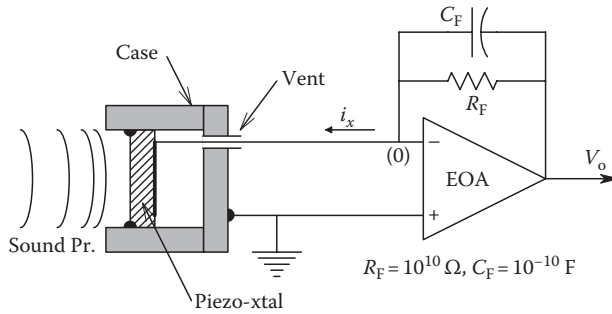


FIGURE P6.4

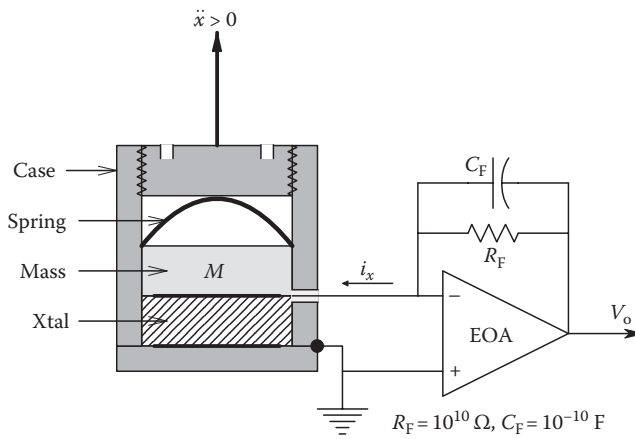


FIGURE P6.5

- 6.5 A piezoelectric accelerometer of the compression type has its output current conditioned by an EOA charge amplifier as shown in Figure P6.5. The proof mass in the accelerometer is 10 g; the piezo-crystal's d coefficient is 140 pCb N. The crystal's capacitance $C_x = 500$ pF; its leakage conductance $G_x = 10^{-12}$ S:
- The accelerometer is given a vertical step of acceleration of 1 g. Find, sketch, and dimension $v_o(t)$.
 - The piezoelectric accelerometer is attached to the tip of a helicopter rotor blade undergoing a static vibration test. At one part of the test, the blade tip vibrates in the SS at 10 Hz in the vertical plane with a peak sinusoidal amplitude of 10 cm. Calculate the peak amplitude of $v_o(t)$ for this condition.
- 6.6 A piezoelectric crystal accelerometer is used to measure the horizontal deceleration of a vehicle in order to deploy its airbags (Figure P6.6). The airbags are to deploy if the deceleration exceeds $5g = 5 \times 9.8 \text{ m/s}^2$. A charge amplifier conditions the accelerometer current output. A comparator senses $v_o(t)$ over a limit, V_R , and triggers airbag inflation:
- Draw the equivalent circuit of the accelerometer and amplifier.
 - Assume the vehicle is traveling at a velocity of 5.0 m/s (11.19 mph) when it hits the crash test barrier.

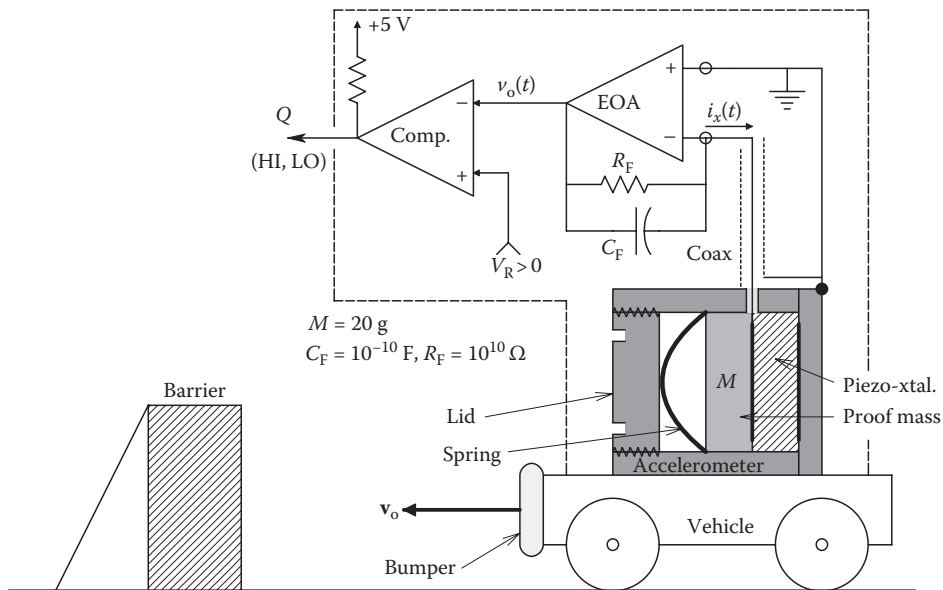


FIGURE P6.6

Its velocity decreases linearly to zero in 100 ms. Graph and dimension the velocity $v(t)$, dv/dt , and d^2v/dt^2 .

- C. Calculate, plot, and dimension the amplifier output $v_o(t)$ for the crash. What must the value of the DC comparator reference voltage, V_R , be so that the comparator output goes LO and triggers the airbags for this crash. Assume that the crystal's polarity is such that the initial output, $v_o(0^+)$, is positive. The crystal's $d = 150 \text{ pCb/N}$. [Note the Laplace x form pair: $1/(s + a) \leftrightarrow e^{-at}$.]
- 6.7 A simple linear accelerometer is made by suspending a mass M in a case from a cantilever spring arm. Two strain gauges are cemented to the cantilever arm. The strain on the cantilever at the gauges is proportional to its relative displacement: $\epsilon = K_k(x_1 - x_2)$. x_1 is the case displacement; x_2 is the mass displacement inside the case. The case is filled with oil to give the accelerometer viscous damping, D . A Wheatstone bridge circuit (see Figure 6.7B) gives an output, $V_o = (V_1 - V_2)$, proportional to the strain, ϵ . For downward deflection of the cantilever, the top strain gauge is under tension and the bottom gauge is under compression. Thus, for the top gauge, $R_{\text{top}} = R_o(1 + \epsilon GF)$, and for the bottom gauge, $R_{\text{bot}} = (1 - \epsilon GF)$:
- A. Write the Newtonian force balance ordinary differential equation for the accelerometer. Laplace transform it to find $X_2/X_1(s)$.
- B. Derive an expression for the strain, $\epsilon(s)$, as a function of $\ddot{X}_1(s)$.
- C. Give an expression for $V_o(s)$ as a function of $\ddot{X}_1(s)$.
- 6.8 In the circuit of Figure P6.8, a silicon PIN photodiode is used to sense the output of an optical system powered by a 532 nm (green) laser diode (LAD). The dark leakage current of the diode is 5 nA at 27°C; the OA is ideal:
- A. Find the R_1 value required to make $V_o = 0$ in the dark.
- B. What optical power incident on the photodiode is required to produce $V_o = 1.0 \text{ V}$? Assume $\eta = 0.5$.

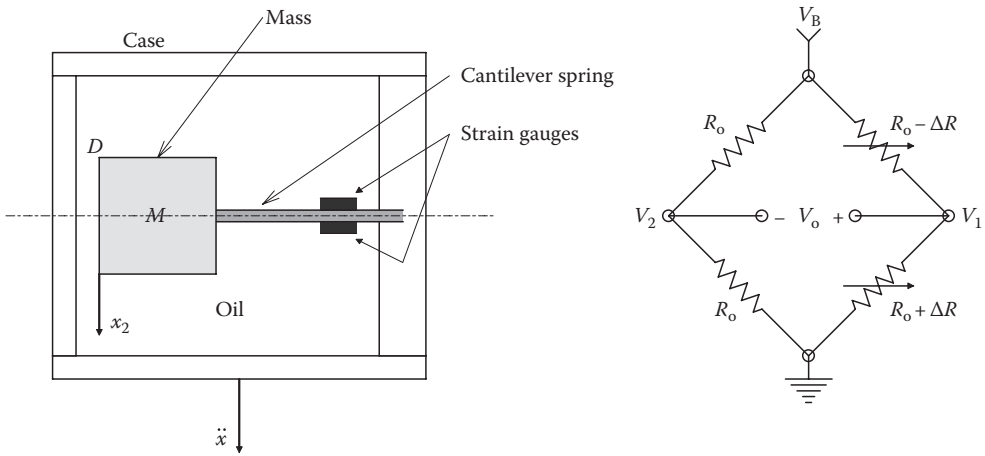


FIGURE P6.7

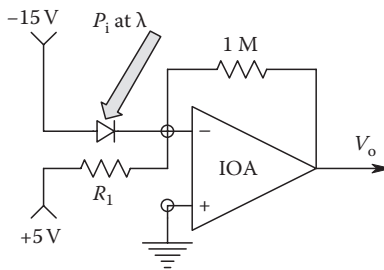


FIGURE P6.8

6.9 In the circuit of Figure P6.9, a silicon photodiode is operated in the zero-current mode ($i_D = 0$). Let $I_{rs} = 1 \text{ nA}$, $V_T = 0.026 \text{ V}$, $I_p = 0.25 P_i$ (photocurrent at 512 nm wavelength):

- A. Sketch the general form of the $i_D = f(v_D, P_i)$ curves for the photodiode.
- B. Derive an expression for $v_D = g(P_i)$.
- C. Sketch and dimension a plot of V_o vs. P_i .

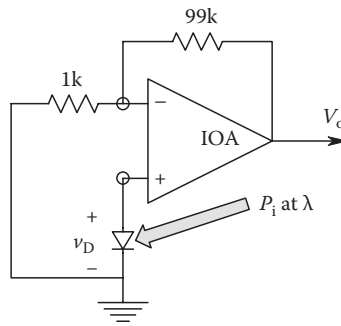


FIGURE P6.9

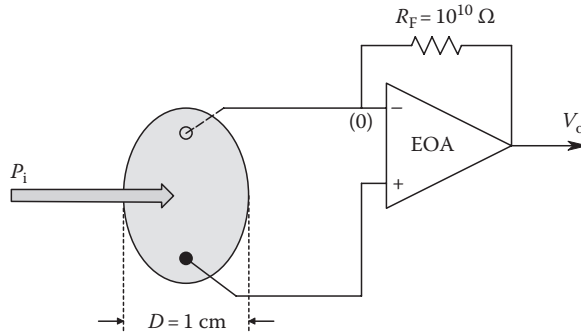


FIGURE P6.10

6.10 In the circuit of Figure P6.10, a disk pyroelectric sensor is used to sense changes in LIR radiation.

The exposed area is 1 cm in diameter, its thickness is $h = 25.46 \mu\text{m}$, its pyroelectric coefficient $K_p = 25 \mu\text{Cb}/(\text{m}^2 \text{K})$, its specific heat is $c = 2.5 \times 10^6 \text{ J}/(\text{m}^3 \text{K})$, its thermal resistance is $\Theta = 20 \text{ K}/\text{W}$, and 50% of the incident IR radiation is absorbed by the sensor and 50% is reflected away:

- A. Calculate the time constant of the sensor.
- B. Assume a step input of LIR power to the sensor. What input power is required to produce a peak output voltage of 100 mV?
- C. Sketch and dimension $v_o(t)$, given the step input of (B).

6.11 In Figure P6.11, a Si PIN photodiode is used to sense a 10 kHz, sinusoidally intensity-modulated, 633 nm light beam. To improve output SNR, the OA is followed by a band-pass filter centered at 10 kHz and having a 1 kHz noise bandwidth. In the dark, the RMS output noise at V_o is $v_{\text{on}} = 0.5 \times 10^{-6} \text{ RMS V}$. Use $h = 6.624 \times 10^{-24} \text{ J}\cdot\text{s}$, $q = 1.60 \times 10^{-19} \text{ Cb}$, $c = 3 \times 10^8 \text{ m/s}$, $\eta = 0.55$. The input light power is described by

$$P_i(t) = 0.5P_{i\text{pk}}[1 + \sin(2\pi 10^4 t)] \text{ W.}$$

The photocurrent is given by

$$I_P(t) = P_i(t) \left[\frac{\eta q}{(h\nu)} \right].$$

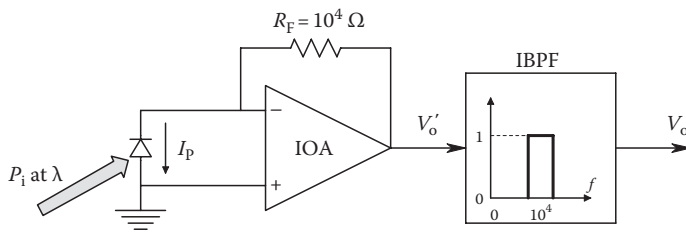


FIGURE P6.11

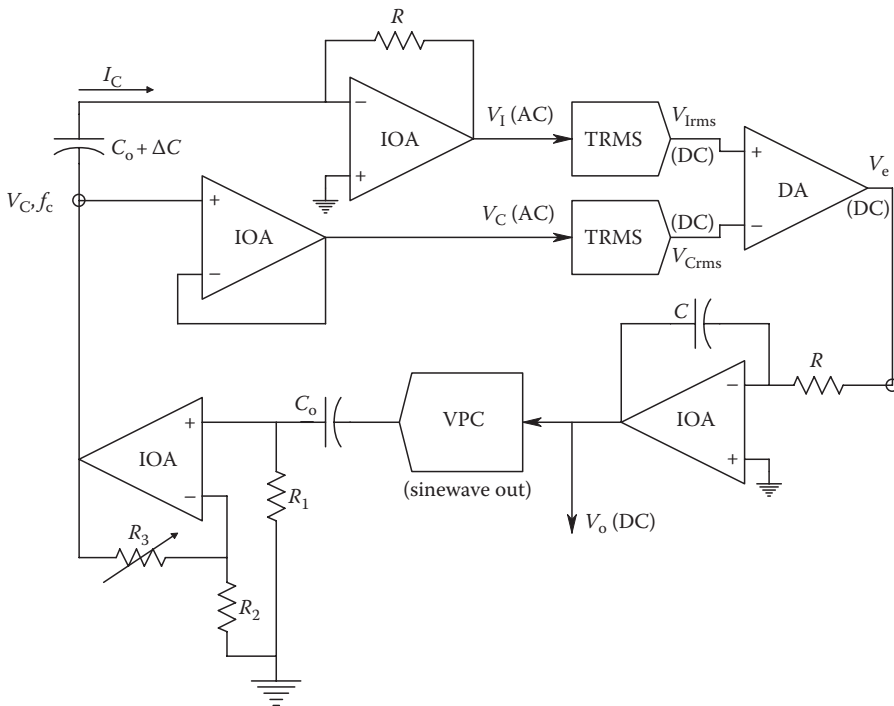


FIGURE P6.12

- A. Give an expression for and sketch and dimension $V_o'(t)$.
- B. Give an expression for and sketch and dimension $V_o(t)$.
- C. Find the P_{ipk} value so that the RMS $V_o = 0.5 \times 10^{-6}$ V.
- 6.12** The feedback system illustrated in Figure P6.12 keeps the root-mean-square value of the AC current through the variable capacitor constant by changing the frequency at constant rms V_c to compensate for changes in C . $C = C_o + \Delta C$. Instead of a VCO, a voltage-to-period converter (VPC) is used in which the output sine wave's period, T , is proportional to V_o . That is, $T = K_p V_o + b$ and $f_c = 1/T$ Hz. In the SS, the integrator forces $V_e \rightarrow 0$, so the rms $V_c =$ the RMS V_1 :
- A. Derive an expression for V_o at null.
- B. What must b equal so that $V_o \propto \Delta C/C_o$?
- 6.13** Figure P6.13 illustrates a circuit of a manually nulled, ΔC sensing circuit. The AC radian frequency is made $\omega = 1/(RC_o)$. The ideal OA makes an all-pass, phase-shifting circuit. Its input is $V_1 \angle 0^\circ$; its output is $V_o = V_1 \angle \theta$. R_2 is manipulated to match the phase of V_o to V_2 . R_3 adjusts the amplitude of V_o to match the amplitude of V_2 :
- A. Derive an expression for $V_o/V_1(j\omega)$.
- B. Derive an expression for the magnitude of V_e as a function of $\Delta C/C_o$ and other circuit parameters.
- Assume $V_e = 0$ for $\Delta C/C_o = 0$ and $|\Delta C/C_o| \ll 1$ in the denominator of $|V_e|$.
- 6.14** Figure P6.14 shows a differential, *commutated capacitor bridge* (cf. Section 5.5.1.6) used to convert the $\Delta C/C_o$ of a differential capacitive sensor to a DC output voltage. In the circuit, MOS switches periodically connect $C_o + \Delta C$ and $C_o - \Delta C$ to alternately discharge

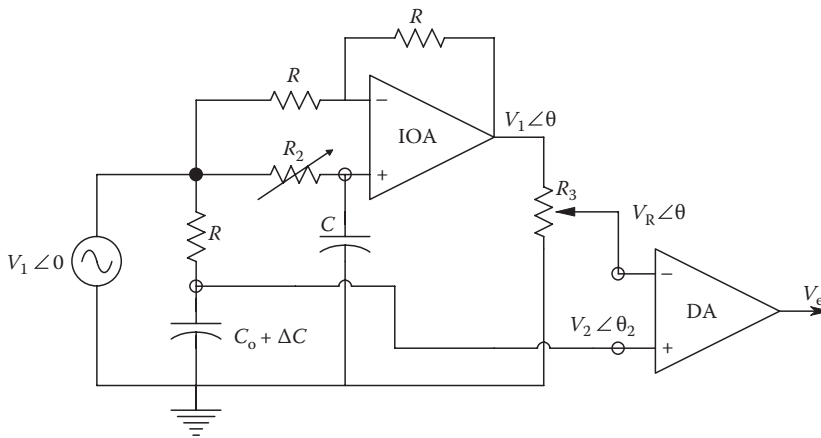


FIGURE P6.13

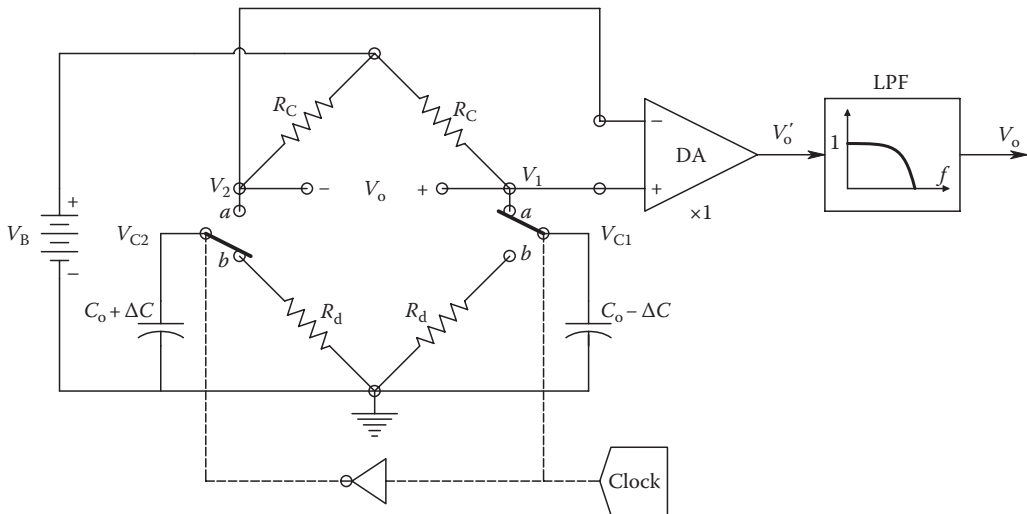


FIGURE P6.14

through the small resistors, R_d , or charge through the larger resistors, R_c . When $C_o + \Delta C$ is connected to R_d , $C_o - \Delta C$ is connected to R_c and vice versa. The bridge output voltage, $v_o(t)$, is the difference between $v_1(t)$ and $v_2(t)$. Assume each capacitor discharges completely through R_d in negligible time. Also, when connected to R_c , each capacitor charges slowly toward V_B . The switching clock has period T . Note that $R_d C_o \ll T/2 \ll R_c C_o$. Although the charging capacitor voltage is an exponential, because $T/2 \ll R_c C_o$, it may be treated as a linear (triangular) waveform over $0 \leq t \leq T/2$. Thus, you can use the approximation $e^{-x} \cong 1 - x$:

- A. Sketch and dimension the SS: $v_{c1}(t)$, $v_{c2}(t)$, $v_1(t)$, $v_2(t)$, and $v_o(t)$ over several clock periods.
- B. $v_o(t)$ is passed through a unity-gain low-pass filter to give an output proportional to its average (DC) value. Find an expression for the average value of $v_o(t)$ in terms of V_B , bridge parameters, and $\Delta C/C_o$.

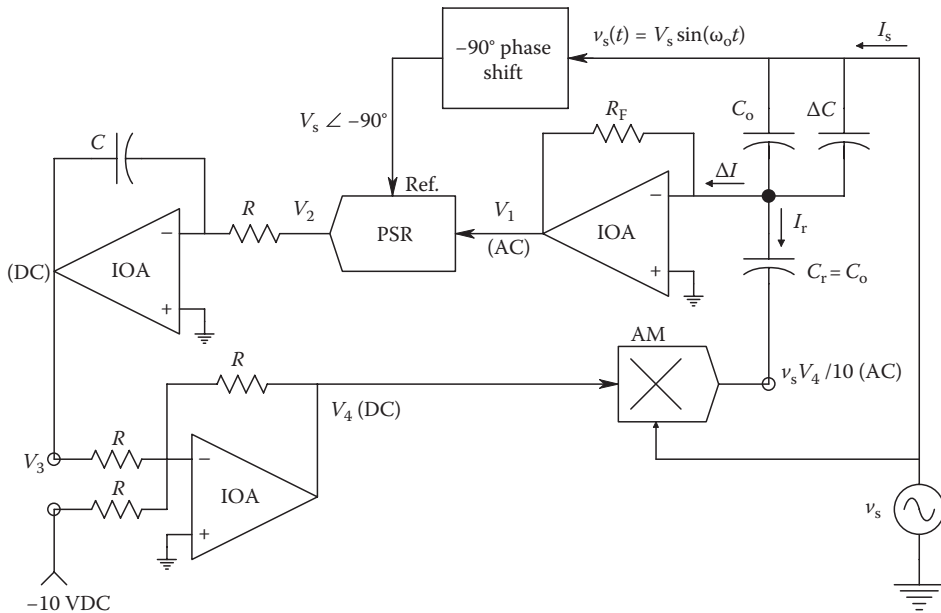


FIGURE P6.15

6.15 Figure P6.15 describes a *self-nulling system* used to measure small changes in a capacitance caused by the presence of foreign objects near the test capacitor’s plates. With no foreign object present, $\Delta C \rightarrow 0$, V_3 is set to zero, and the reference capacitor is set to C_0 so that ΔI into IOA1’s summing junction is zero; thus, V_1 and V_2 are zero. The output of the summer (IOA3) is $V_4 = +10$ Vdc, so the multiplier output is $-v_s(t)$. In the figure, PSR = phase-sensitive rectifier and AM = analog multiplier.

When a foreign object is present, it *increases* the sensing capacitor’s capacitance by ΔC . Now, $|I_s| > |I_r|$, and ΔI , V_1 (AC), and V_2 (DC) all increase. The integrator output V_3 (DC) goes negative by a small amount; this causes V_4 (DC) to *increase*, making v_m (AC) more negative, increasing $|I_r|$, until $\Delta I \rightarrow 0$ again:

- A. Write a general algebraic expression for $\Delta I(j\omega)$ (system not at null).
 - B. Write a general expression for $v_1(t)$.
 - C. Sketch and dimension $v_2(t)$. Why is the -90° phase shift necessary for the PSR sync?
 - D. At null, $\Delta I \rightarrow 0$. Derive an expression for V_3 (DC) at null in terms of ΔC , C_0 , and other system parameters. (V_3 is the system output.)
- 6.16 In Figure P6.16, a capacitive sensor is used to make a low-pass filter. The radian frequency of the sinusoidal source is made $\omega = 1/(RC_0)$; $v_s(t) = V_s \sin(\omega t)$:
- A. Give an expression for $V_2/V_s(j\omega)$ in time-constant form.
 - B. Find an expression for the transfer function’s magnitude sensitivity, $S_{\Delta C}^{|v_2|}$, defined as $d|V_2|/d \Delta C$.
 - C. Find an expression for the transfer function’s phase sensitivity, $S_{\Delta C}^\phi$, defined as $d\phi/d\Delta C$.

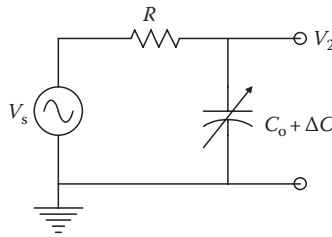


FIGURE P6.16

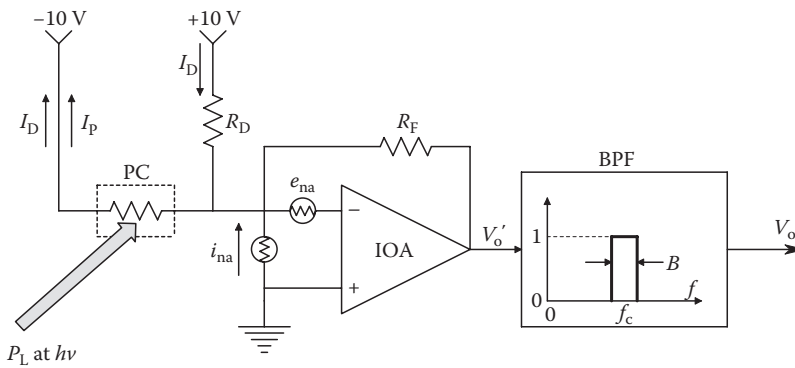


FIGURE P6.17

- 6.17 An ideal but noisy OA is used to condition the signal from a silicon PC light sensor. Resistor R_D is used to compensate for the PC's dark current, I_D . In the circuit, $e_{na} = 3 \text{ nV}/\sqrt{\text{Hz}}$, $i_{na} = 0.4 \text{ pA}/\sqrt{\text{Hz}}$, $R_F = 10^6 \text{ } \Omega$, $R_D = 2.3 \times 10^5 \text{ } \Omega$, and $V = 10 \text{ Vdc}$. Because the light is chopped, the OA is followed by a (ideal) noiseless BPF with center frequency of $f_c = 400 \text{ Hz}$, pass gain = 1, and bandwidth $B = 100 \text{ Hz}$. See Figure P6.17:
- Derive an algebraic expression for and calculate the numerical value of the RMS noise output voltage, v_{on} , in the dark. Assume all resistors make thermal white noise and $T = 27^\circ\text{C}$.
 - Give an expression for V_o' as a function of the photocurrent, I_p . What is the PC's DC dark current?
 - The photocurrent may be shown to be given by $I_p = P_L \eta V \tau_p (\mu_p + \mu_n) / (l^2 h \nu)$ amps, where $l = 0.01 \text{ cm}$, $h \nu = 1.2 \text{ eV}$, $\tau_p = 10^{-4} \text{ s}$, $(\mu_p + \mu_n) = 1800 \text{ cm}^2 / (\text{V} \cdot \text{s})$, $\eta = 0.8$. What photon power, P_L , in watts must be absorbed by the PC in order that $V_o' = v_{on}$ from part (A)? Find the ratio of the I_p at this P_L to I_D .
- 6.18 A square pyroelectric sensor is used to sense square-wave-chopper LIR light. The chopping has a 50% (ON/OFF) duty cycle. The circuit is shown in Figure P6.18. The sensor is 3 mm on a side and $10 \text{ } \mu\text{m}$ thick. The specific heat is $c = 2.5 \times 10^6 \text{ J/m}^3 \text{ K}$. The pyroelectric constant is $K_p = 25 \text{ } \mu\text{Cb/m}^2 \text{ K}$, and $\Theta = 20^\circ/\text{W}$:
- Calculate the sensor's thermal time constant, τ_{th} .
 - Find the P_i value required to give a 1 V peak V_o initially.
 - Assume the input square wave of peak height P_i found in (B), and period 2 ms is applied to the sensor for a long enough time for the system to be in the SS. Sketch and dimension $v_o(t)$ in the SS, showing the maxima and minima.

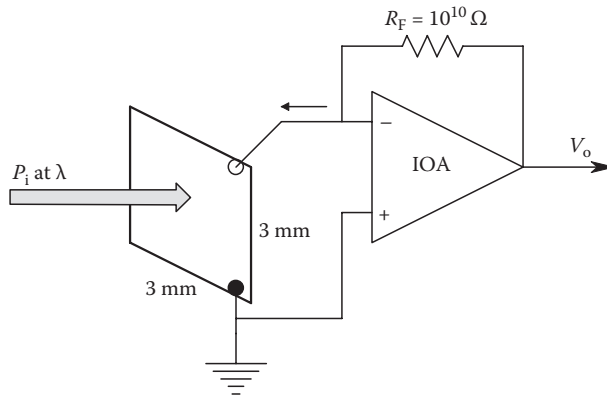


FIGURE P6.18

- 6.19 A 1 lb weight is suspended from a 0.005" diameter wire. The unstrained resistance of the wire, R_o , is 270 Ω , and its unstrained length is $L_o = 36"$. The wire's gauge factor, GF, is 3.2:
- Calculate the psi tensile stress on the loaded wire.
 - Assume Young's modulus is 3.2×10^7 psi. Find the strain on the wire and ΔR .
 - Assume a Wheatstone bridge incorporates the wire as one arm. The other arm resistances = R_o . The bridge is excited with 5 Vdc. Find V_o due to the 1 lb load.
- 6.20 The *Lion circuit* is used to condition the output of differential capacitor sensors. It is shown in Figure P6.20. If we assume that the diodes are ideal, the waveform at the v_o' node can be represented by

$$v_o'(t) = \left(\frac{V_s}{2}\right) \left[1 - \exp\left(\frac{-t}{t_1}\right) \right], t \text{ for } V_s = +12 \text{ V}$$

$$v_o' = \left(\frac{-V_s}{2}\right) \left[1 - \exp\left(\frac{-t}{t_2}\right) \right], t \text{ for } V_s = -12 \text{ V}$$

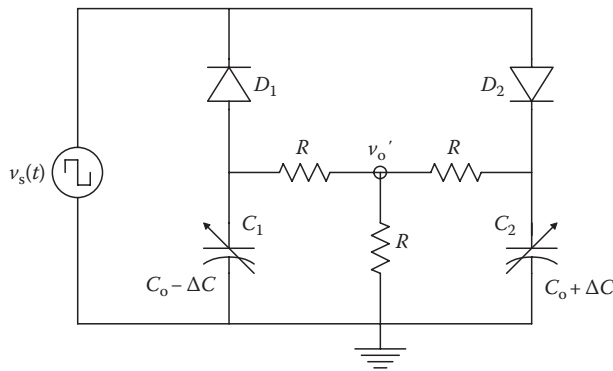


FIGURE P6.20

where $C_1 = C_o - \Delta C$, $C_2 = C_o + \Delta C$, $\tau_1 = C_1 R(3/2)$, and $\tau_2 = C_2 R(3/2)$. Also, we assume that τ_1 and $\tau_2 \gg T/2$ of the input square wave, $v_s(t)$:

- A. Derive an algebraic expression for the average $v'_o(t)$ in the SS. Let $R = 10^6 \Omega$, $C_o = 30 \text{ pF}$, $T = 10^{-5} \text{ s}$, $V_s = 12 \text{ Vpk}$. Evaluate the system's sensitivity in DC volts/ $(\Delta C/C_o)$.
 - B. Use SPICE or MicroCap™ to simulate the Lion bridge output, $v'_o(t)$. Use a $\pm 12 \text{ V}$, 100 kHz square wave input, 1N4148 diodes; let $\Delta C/C_o = 10 \text{ pF}/30 \text{ pF}$. Place a 10 fF capacitor across the v'_o node to ground (this prevents register overflow and doesn't affect the results of the simulation). Examine five cycles of the output.
- 6.21** The RMS magnitude of the 60 Hz magnetic field vector in a small volume of 3D space is to be measured. Three, identical, orthogonal, Hall sensors are used, as shown in Figure P6.21A. The AC output voltage of each sensor is proportional to the perpendicular component of the \mathbf{B} field through that sensor. For example, $v_{Hy}(t) = K_H B_y(t) = K_H |\mathbf{B}| \sin(377t) \sin(\beta)$ (the DC current through each Hall sensor is not shown):
- A. Give expressions for $v_{Hx}(t)$ and $v_{Hz}(t)$.
 - B. Consider the resolver circuit of Figure P6.21B. Derive an expression for $v_2(t)$, $v_3(t)$, and V_4 .
- 6.22** An analog multiplier and a low-pass filter are used to demodulate the output of a LVDT length sensor, as shown in Figure P6.22. The excitation voltage is $v_s(t) = V_s \sin(6283t)$. The output voltage is $v_o(t) = K_s x(t) \sin(6283t - \theta)$. The low-pass filter has unity DC gain and removes all signals of carrier frequency and above from $v_m(t)$:
- A. Derive an expression for $v_x(t)$.
 - B. Design a passive or active circuit that can be put in the v_r line to make the reference phase agree with the phase shift in $v_o(t)$. The peak value of $v'_r(t)$ should remain the same.

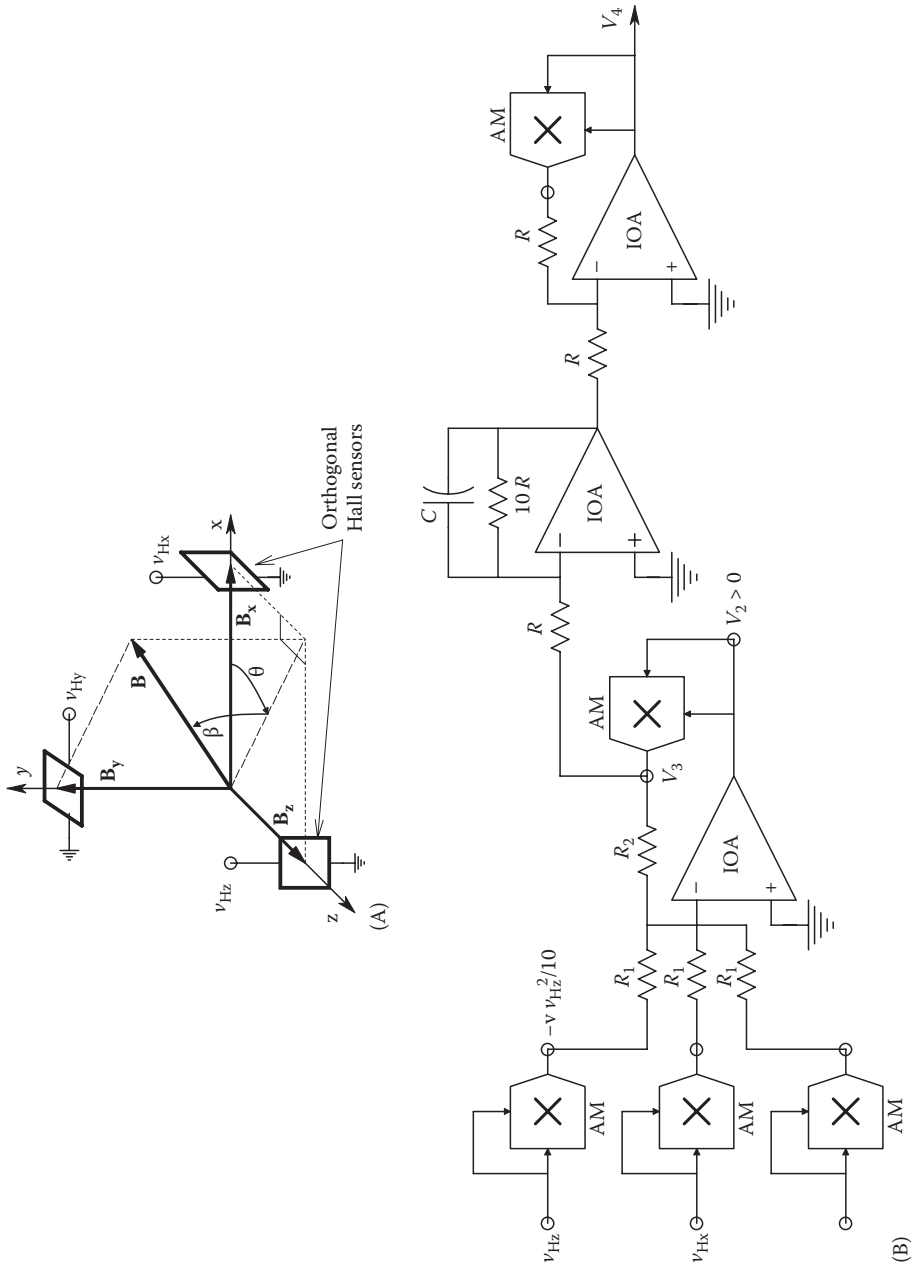


FIGURE P6.21

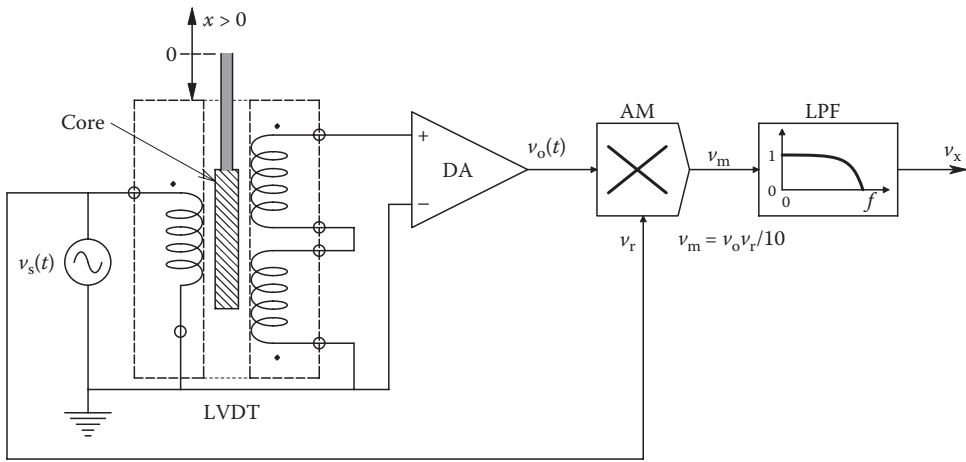


FIGURE P6.22

7

Applications of Sensors to Physical Measurements

7.1 Introduction

In this chapter, we examine various means available to measure certain physical quantities. In Chapter 6, we reviewed various transduction mechanisms underlying measurement system function, such as changes in resistance. (Many physical quantities cause resistive changes, such as temperature, strain [caused by force or pressure], light intensity, and humidity.) Our purpose in this chapter is to examine the most commonly employed sensor types used to measure a given physical quantity, such as temperature. The quantities that we shall consider here are *acceleration, velocity and position (angular and rectilinear), temperature, pressure, force, torque, gyroscopic measurement of angular velocity and position, and the GPS system*. We also give an introduction to spectrophotometric measurements of substances.

7.2 Measurement of Angular Acceleration, Velocity, and Displacement

There are many applications where it is important to know information about the attitude of a vehicle in inertial space. In the case of aircraft, instrumentation of roll, pitch, and yaw is critical for the design of robust, closed-loop, fly-by-wire autopilots and flight stabilization systems. Similar knowledge is important for the control of high-speed hydrofoils, hovercraft, submarines, and unmanned autonomous vehicles (UAVs). Drone surveillance aircraft, guided missiles, and *smart* bombs also require feedback on their roll, pitch, and yaw to perform their tasks effectively. In addition to vehicle stabilization and guidance, the measurement of angular acceleration, velocity, and position is also important in the design of constrained mechanical systems that face variable loads and inputs, for example, automobile and truck suspension systems.

In the following sections, we describe some of the common sensor systems used to measure angular acceleration, angular velocity, and angular position.

7.2.1 Angular Acceleration Measurement

In this section, we describe several means by which angular acceleration can be measured. One is tempted to obtain angular acceleration by simply differentiating the output of a rate gyroscope. While theoretically correct, this procedure will generally result in poor resolution and a poor signal to noise ratio because the process of differentiation enhances the noise present in the angular velocity signal. It is generally better to measure angular acceleration directly, as described below.

7.2.1.1 Angular Acceleration Measurement with a Constrained Mechanical Gyro

The inner gimbals of a mechanical gyro (see Section 7.2.2 for the development of the linear ODEs describing the behavior of mechanical gyroscopes) are restrained by a stiff, piezo-electric twister crystal as shown in Figure 7.1A. We may assume $\dot{\theta}$ and $\ddot{\theta}$ are zero, and the torsional moment for the inner gimbals is given by the gyro equation, (7.26A), as

$$M_{\theta} = -H_r \dot{\phi}. \tag{7.1}$$

The twister piezoelectric transducer has an equivalent circuit given by Figure 7.1B at mechanical frequencies well below its mechanical resonance frequency. A net charge is displaced by the mechanical torque, $q = M_{\theta} d$ (d is a constant). Current is simply the rate of flow of charge, so the equivalent current source in the model for the transducer is $i_x = -H_r \dot{\phi} d$, and the output voltage of the transducer is described by the transfer function:

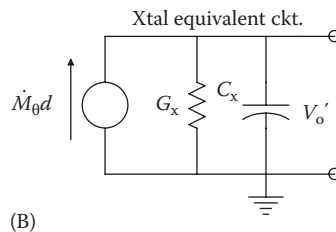
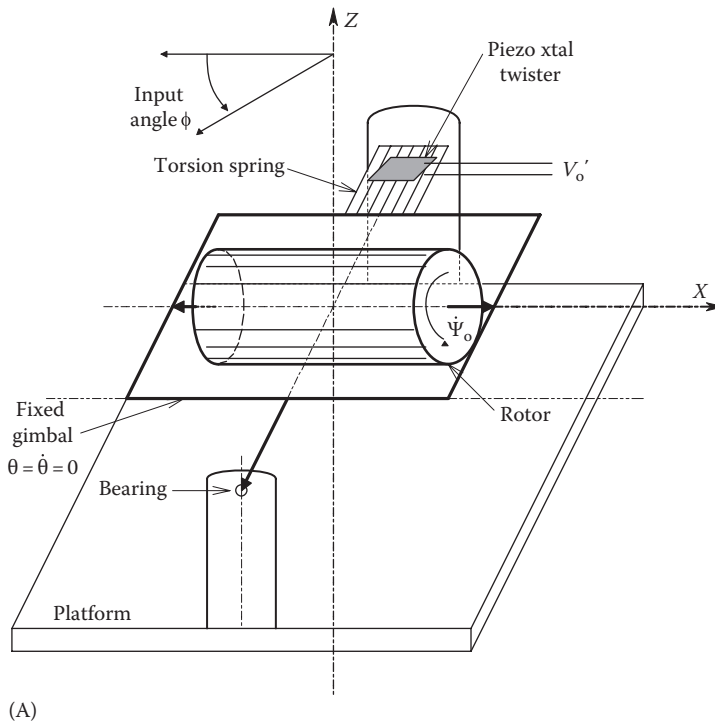


FIGURE 7.1
 (A). Simple mechanical gyroscope with a gimbal restrained by a torsion spring. Gimbal torque is proportional to the input angular velocity, $\dot{\phi}$. A piezoelectric *twister* transducer senses the gimbal torque, M_{θ} . (B). Equivalent circuit of the piezoelectric transducer at mechanical frequencies well below the transducer’s mechanical resonance frequency.

$$V_o(s) = \frac{H_r R_x d \ddot{\Phi}(s)}{(s C_x R_x + 1)} \tag{7.2}$$

Thus, the SS output of the transducer is proportional to the angular acceleration of the gyro around the ϕ -axis. If a charge amplifier is used (see Section 7.2.6), the time constant in the output transfer function becomes $C_F R_F$, the feedback C and R of the charge amplifier. A far simpler means of measuring $\dot{\phi}(t)$ is shown in Section 7.2.12.

7.2.1.2 Simple Inertia Wheel–Spring–Dashpot Angular Accelerometer

Figure 7.2 illustrates a simple inertia wheel angular accelerometer in which the output voltage is proportional to the *relative position* of the rotor wheel in the case, that is,

$$V_o = K_o (\phi_R - \phi_C) \tag{7.3}$$

where

ϕ_R is the instantaneous angle of the rotor with a fixed, external reference

ϕ_C is the instantaneous angle of the angular accelerometer case with a fixed, external reference

K_o is a sensor gain constant

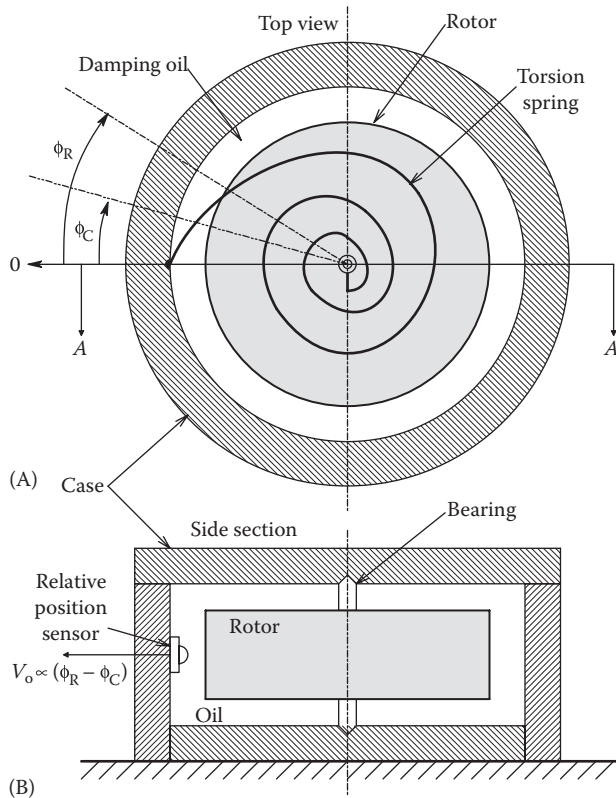


FIGURE 7.2

Two views of an inertia wheel angular accelerometer. (A) Top view. Wheel position in the case can be sensed photooptically or by a potentiometer. The inertia wheel is restrained by a helical spring. (B) Side-sectional view.

The Newtonian torque equation is

$$J\ddot{\phi}_R + B(\dot{\phi}_R - \dot{\phi}_C) + K_s(\phi_R - \phi_C) = 0 \quad (7.4)$$

where

B is the viscous friction between the wheel and the case

K_s is the torsion spring constant

If we Laplace transform the ODE of Equation 7.4 and solve for ϕ_R , and then substitute into the expression for V_o , we obtain

$$\frac{V_o}{\ddot{\phi}_C}(s) = \frac{-K_o J / K_s}{s^2 J / K_s + s B / K_s + 1}. \quad (7.5)$$

Thus, the SS output of this simple angular accelerometer is $V_o = -\ddot{\phi}_C (K_o J / K_s)$. Such an accelerometer should be well damped so that the rotor wheel will not oscillate. The accelerometer's damping factor is easily seen to be $\zeta = B / (2\sqrt{JK_s})$ and should be made to be about 0.5–0.7 by adjusting B , to obtain a rapid transient response without excessive overshoots. The natural frequency of this accelerometer is $\omega_n = \sqrt{K_s / J}$ r/s. Thus, extending the accelerometer's bandwidth decreases its sensitivity, a good example of a gain–bandwidth trade-off. To see how it is possible to maintain high sensitivity and bandwidth, we will examine an active accelerometer design.

7.2.1.3 Servoangular Accelerometer

In this design, shown in Figure 7.3, a torque motor is used as an active spring. The torque motor has a flat torque–speed characteristic over most of its speed range, the torque being proportional to the motor's input current, I_M . The accelerometer output voltage is made proportional to I_M . The torque balance ODE for the servoaccelerometer is

$$T_M = K_T I_M = J\ddot{\phi}_R + B(\dot{\phi}_R - \dot{\phi}_C). \quad (7.6)$$

Also, we have

$$V_2 = K_P (\phi_C - \phi_R) \quad (7.7)$$

$$I_M = G_M V_2 \quad (7.8)$$

$$V_o = K_O I_M. \quad (7.9)$$

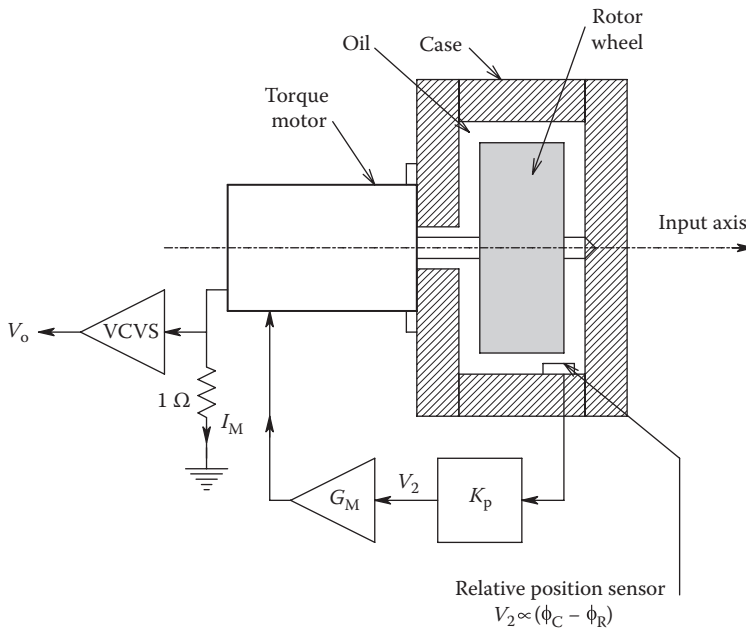


FIGURE 7.3

Vertical section through a servoangular accelerometer. The torque motor current, I_M , may be shown to be proportional to the angular acceleration of the case around the input axis. G_M = transconductance amplifier; K_p and K_o = gains. See text for analysis.

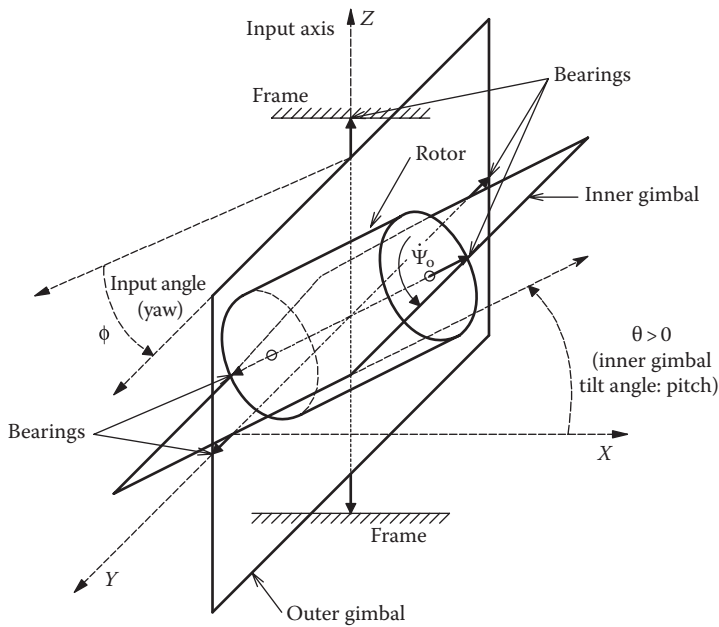
Combining the earlier equations, it can be shown that the angular accelerometer has the low-pass transfer function:

$$\frac{V_o}{\ddot{\Phi}_C}(s) = \frac{K_o/K_T}{s^2 J / (K_T K_p G_M) + B / (K_T K_p G_M) + 1} \tag{7.10}$$

Now, the natural frequency governing the high-frequency response is set by the electrical parameters K_T , K_p , and G_M . In this design, K_o can be made large to compensate for a high K_T , thus preserving sensitivity and bandwidth. The system's natural frequency is $\omega_n = \sqrt{K_T K_p G_M / J}$ r/s. In the SS, $V_o = \ddot{\Phi}_C K_o / K_T$.

7.2.2 Angular Velocity Measurement with Mechanical Gyroscopes

The angular velocity of vehicles, such as associated with the roll, pitch, or yaw rate of aircraft, missiles, boats, submarines, or spacecraft (or even birds and insects in flight), needs to be measured in order to control the vehicle's attitude in inertial space and, in the case of certain modern hypersonic aircraft, stabilize their flight attitude under all flying conditions. The principal means of sensing angular velocity is by a *rate gyro*. A rate gyro can use the traditional electromechanical spinning rotor, or be optoelectronic, using the Sagnac effect, introduced in Section 6.7.3, or use a variety of novel mechanisms including an air jet, vibrating tuning fork (TF), or a vibrating mass or inertia wheel. In this section, we will first discuss conventional mechanical gyroscopes and then describe the design and sensitivity limitations of fiber-optic, Sagnac-effect gyros.

**FIGURE 7.4**

Schematic diagram of a two-gimbal mechanical gyroscope. The input angle, ϕ , is a forced, CCW rotation of the outer gimbal around the z-axis (a yaw input).

The innards of a conventional, 2-gimbal, mechanical gyroscope are illustrated in Figure 7.4. The rotor is spun electrically at a high, constant angular velocity, typically as high as 24,000 rpm for 400 Hz excitation. The rotor is held in an inner frame, or gimbal, which is in turn suspended inside an outer gimbal. In early (e.g., 1950s) gyroscopes used in autopilots, the angle between the inner and outer gimbals, θ , was crudely sensed by a wiper on a wire-wound resistor potentiometer. Resolution therefore was limited by the diameter of the resistance wire vs. the area of the wiper. In the analysis that follows, infinite resolution in sensing the gyro tilt angle, θ , is assumed.

Referring to Figure 7.4, we will define ϕ as the input angle. If the rotor is spinning in the direction shown with angular velocity, ψ , a positive (CCW) increase in ϕ will result in the gyroscopic effect of having the inner gimbals and rotor tip up to the right, giving a $\theta > 0$ (CCW rotation). The detailed analysis of the dynamics of a mechanical gyro results in three, coupled, highly nonlinear differential equations. Analysis proceeds smoothly if we define three mutually orthogonal coordinates for the gyro system; these are in fact the angular momentum axes of the rotor and inner gimbals. \mathbf{H}_1 is the angular momentum vector component perpendicular to the rotor spin axis and the plane of the inner gimbals. \mathbf{H}_2 is the angular momentum vector of the inner gimbals and rotor as they pivot on the outer gimbals. \mathbf{H}_2 is in the plane of the inner gimbals and perpendicular to the rotor spin axis and \mathbf{H}_1 . \mathbf{H}_3 is the angular momentum vector of the rotor, along the rotor axis and perpendicular to the other two momentum vectors. The direction of the angular momentum vectors is given by the right-hand rule; the fingers curl in the direction of rotation and the thumb points in the vector direction.

The best way to describe the dynamics of a mechanical gyroscope is by the Euler-Lagrange equations (Aseltine 1958, Cannon 1967). These equations can be derived from Newton's laws and provide a general means of describing the motion of complex

mechanical systems having rotating or fixed coordinate systems. The Lagrange equations are of the form

$$\frac{d}{dt} \left[\frac{\partial U}{\partial \dot{x}_j} \right] - \frac{\partial U}{\partial x_j} = M_j, \quad j = 1, 2, 3, \dots, N \quad (7.11)$$

where

U is the total KE of the system

x_j is the j th general coordinate

M_j is the j th generalized reaction force, moment, or torque

The differential work done in causing a small change in the coordinates can be written as

$$\delta W = \sum_j M_j \delta x_j \quad j = 1, 2, 3, \dots, N. \quad (7.12)$$

In the case of the gyroscope shown in Figure 7.4, x_1 will be taken as the rotor angle, ψ ; x_2 is defined as the inner gimbals' angle, θ , and x_3 is the input angle, ϕ . The *total KE* of the gyro is the sum of the KEs associated with each orthogonal coordinate and can be written as

$$U = \left(\frac{1}{2} \right) J_r [\dot{\psi} + \dot{\phi} \sin(\theta)]^2 + \left(\frac{1}{2} \right) J_1 \dot{\theta}^2 + \left(\frac{1}{2} \right) J_1 [\dot{\phi} \cos(\theta)]^2 \quad (7.13)$$

where

J_r is the moment of inertia of the rotor around its spin axis

J_1 is the moment of inertia of the rotor and inner gimbals around the axes normal to the spin axis

From Equation 7.11, we can write the first nonlinear ODE for the system. Consider $j = 1$, so $x_1 = \psi$. Thus, it is easy to show that

$$\frac{\partial U}{\partial \psi} = 0 \quad (7.14)$$

$$\frac{\partial U}{\partial \dot{\psi}} = J_r \dot{\psi} + J_r \dot{\phi} \sin(\theta) \quad (7.15)$$

$$\frac{d}{dt} \left[\frac{\partial U}{\partial \dot{\psi}} \right] = J_r \ddot{\psi} + J_r \ddot{\phi} \sin(\theta) + J_r \dot{\phi} \dot{\theta} \cos(\theta). \quad (7.16)$$

Hence, the Lagrange equation for the gyro rotor spin axis friction moment is

$$M_\psi = J_r [\ddot{\psi} + \ddot{\phi} \sin(\theta) + \dot{\phi} \dot{\theta} \cos(\theta)]. \quad (7.17)$$

The rotor friction moment, M_ψ , is usually taken as zero. The rotor is run at constant (synchronous) velocity, so $\ddot{\psi} = 0$. Now, the second Lagrange equation is written for $x_2 = \theta$, the inner gimbals' angle. In this case, we have

$$\frac{\partial U}{\partial \theta} = J_r \dot{\psi} \dot{\phi} \cos(\theta) + J_r \dot{\phi}^2 \sin(\theta) \cos(\theta) - J_1 \dot{\phi}^2 \cos(\theta) \sin(\theta) \quad (7.18)$$

$$\frac{\partial U}{\partial \dot{\theta}} = J_1 \dot{\theta} \quad (7.19)$$

$$\frac{d}{dt} \left[\frac{\partial U}{\partial \dot{\theta}} \right] = J_1 \ddot{\theta}. \quad (7.20)$$

Hence, the Lagrange equation for the θ coordinate is

$$M_\theta = J_1 \ddot{\theta} - J_r \dot{\psi} \dot{\phi} \cos(\theta) - J_r \dot{\phi}^2 \cos(\theta) + J_1 \dot{\phi}^2 \cos(\theta) \sin(\theta). \quad (7.21)$$

The third Lagrange equation is written for $x_3 = \phi$, the input angle:

$$\frac{\partial U}{\partial \phi} = 0 \quad (7.22)$$

$$\frac{\partial U}{\partial \dot{\phi}} = J_r \dot{\psi} \sin(\theta) + J_r \dot{\phi} \sin^2(\theta) + J_1 \dot{\phi} \cos^2(\theta) \quad (7.23)$$

$$\begin{aligned} \frac{d}{dt} \left[\frac{\partial U}{\partial \dot{\phi}} \right] &= J_r \ddot{\psi} \sin(\theta) + J_r \dot{\psi} \dot{\theta} \cos(\theta) + J_r \ddot{\phi} \sin(\theta) + 2J_r \dot{\phi} \dot{\theta} \sin(\theta) \cos(\theta) \\ &\quad - 2J_1 \dot{\phi} \dot{\theta} \cos(\theta) \sin(\theta) + J_1 \ddot{\phi} \cos^2(\theta). \end{aligned} \quad (7.24)$$

The Lagrange equation for ϕ is just

$$M_\phi = \frac{d}{dt} \left[\frac{\partial U}{\partial \dot{\phi}} \right]. \quad (7.25)$$

To linearize the gyro ODEs, Equations 7.21 and 7.25, we note again that $\ddot{\psi} \equiv 0$, and we assume that system operation is such that $|\theta| < 15^\circ$ so $\sin(\theta) \cong \theta$ in radians. Also, $\cos(\theta) \cong 1$, and terms containing products of the numerically small quantities $\dot{\phi}$, $\ddot{\phi}$, $\dot{\theta}$, and $\ddot{\theta}$ are negligible. Under these approximations, we finally obtain the two, simultaneous, *linear* ODEs that approximately describe the dynamic behavior of the gyro:

$$M_\theta = J_1 \ddot{\theta} - H_r \dot{\phi} \quad (7.26A)$$

$$M_\phi = H_r \dot{\theta} + J_1 \ddot{\phi} \quad (7.26B)$$

where H_r is the angular momentum of the rapidly spinning rotor, $J_r \dot{\psi}$.

Under most conditions, the outer gimbals' angle, ϕ , is considered to be a rotation of the vehicle about a principal axis, such as its roll axis. The response of the roll gyro having its outer gimbals' axis pointing in the direction of the roll axis is considered to be the inner gimbals' angle, θ , or, in some cases, the torque, M_θ . The linear gyro differential equations earlier can be Laplace transformed and written as

$$M_\theta(s) = J_1 s^2 \Theta(s) + H_r s \Phi(s) \tag{7.27A}$$

$$M_\phi(s) = H_r s \Theta(s) + J_1 s^2 \Phi(s). \tag{7.27B}$$

We shall use these linear equations to describe the dynamics of the rate gyro and, later, the behavior of the RIG.

7.2.2.1 Mechanical Rate Gyro

In the one-gimbal mechanical rate gyro, the gimbal's motion is restrained by a spring with torque constant, K_s , and a viscous damper (dashpot) with rate constant B , as shown in Figure 7.5. In this gyro, we have dispensed with the outer gimbal and assume the gyro base rotates an angle ϕ , as shown, around the input (z) axis. The rotor gimbal's moment is thus given by

$$M_\theta = -\Theta(s)[Bs + K_s]. \tag{7.28}$$

The gyro equations can now be written as

$$0 = \Theta(s)[s^2 J_1 + sB + K_s] - H_r s \Phi(s) \tag{7.29A}$$

$$M_\phi = \Theta(s) s H_r + J_1 s^2 \Phi(s). \tag{7.29B}$$

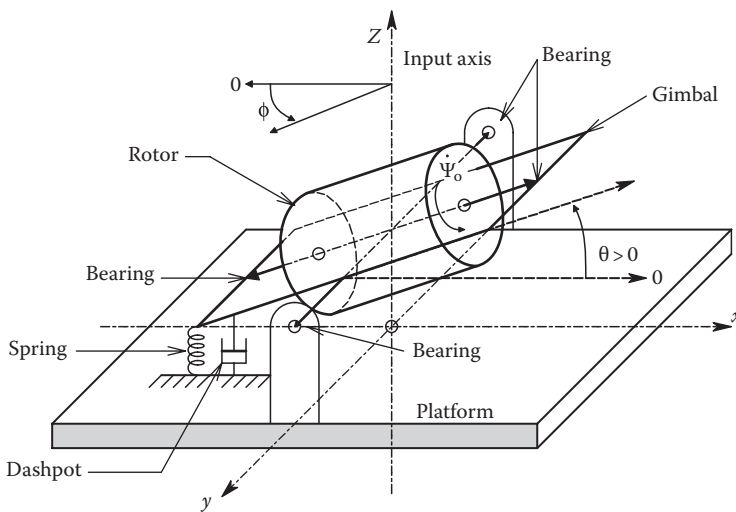


FIGURE 7.5 Schematic diagram of a single-gimbal, mechanical, rate gyro. In the SS, the gimbal (pitch) angle $\theta = \dot{\phi} H_r / K_s$.

Using Cramer's rule, we can solve Equations 7.29A and 7.29B for the relationship between θ and $\dot{\phi}$:

$$\frac{\Theta}{\dot{\Phi}}(s) = \frac{\begin{vmatrix} 0 & -sH_r \\ M_\phi & s^2J_1 \end{vmatrix}}{\begin{vmatrix} s^2J_1 + sB + K_s & 0 \\ sH_r & M_\phi \end{vmatrix}} = \frac{M_\phi s H_r}{M_\phi(s^2J_1 + sB + K_s)} = \frac{sH_r/K_s}{s^2J_1/K_s + sB/K_s + 1}. \quad (7.30)$$

This transfer function is easily seen to be a relation between the gimbal's tilt angle, θ , and the angular velocity of the gyro around the input axis, $\dot{\phi}$:

$$\frac{\Theta}{\dot{\Phi}}(s) = \frac{H_r/K_s}{s^2J_1/K_s + sB/K_s + 1}. \quad (7.31)$$

For a constant velocity input,

$$\dot{\Phi}(s) = \frac{\dot{\phi}_o}{s} \quad (7.32)$$

and the output angle is given by

$$\Theta(s) = \frac{\dot{\phi}_o(H_r/K_s)}{s(s^2J_1/K_s + sB/K_s + 1)}. \quad (7.33)$$

In the time domain, the response $\theta(t)$ is shown in Figure 7.6. It is a classic second-order, low-pass system step response, with SS value, $\theta_{SS} = \dot{\phi}_o H_r/K_s$. Thus, in response to a constant angular velocity input, $\dot{\phi}_o$, the rate gyro rotor tips an SS angle, θ_{SS} . θ can be measured by a low-friction potentiometer wiper or by a frictionless optical coding disk.

Another approach to rate gyro design is shown in Figure 7.7. Here, the rotor gimbal is constrained not to rotate by a very stiff torsion spring. The spring has strain gauges

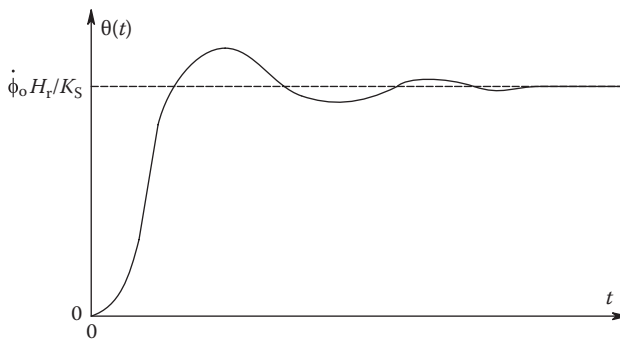


FIGURE 7.6

Second-order response of the rate gyro of Figure 7.5 to a step input of angular velocity, $\dot{\phi}(t) = \dot{\phi}_o U(t)$.

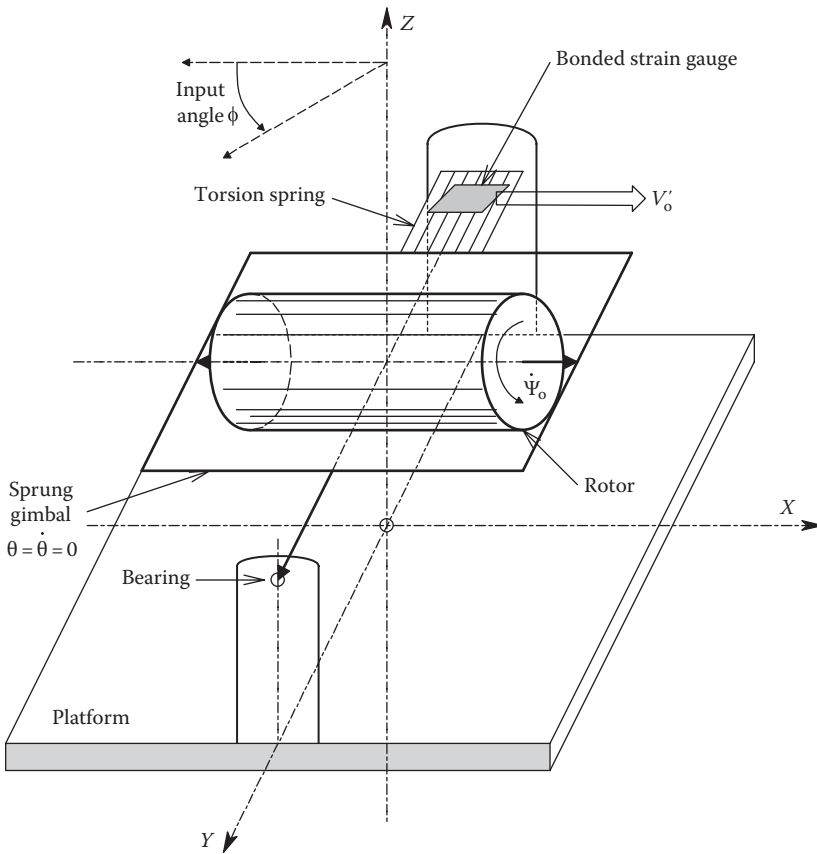


FIGURE 7.7 Constrained gimbal rate gyro. The gimbal torque is measured by a strain gauge and is proportional to $\dot{\phi}$. Compare this system to the gyro angular accelerometer shown in Figure 7.1.

cemented to it and arranged so that torque exerted on the spring produces an output voltage. Referring to the linearized gyro equations, Equations 7.27A and 7.27B, and for practical purposes, we assume that $\ddot{\theta} = \dot{\theta} \equiv 0$. Thus, Equation 7.27A reduces to

$$M_{\theta}(s) = -H_r s \Phi(s). \tag{7.34}$$

Hence, the output of the strain gauge bridge attached to the torsion spring can be written as

$$v_o(t) = K_B H_r \dot{\phi}. \tag{7.35}$$

Thus, we see that the rotor gimbal does not need to move appreciably to sense $\dot{\phi}$. Note that the $\dot{\phi}$ term is multiplied by a $\cos(\beta)$ term in the relations presented earlier when the gyro is rotated around some axis at an angle β with its input or z-axis.

7.2.2.2 Sagnac-Effect Fiber-Optic Gyroscopes

The Sagnac gyro is actually a mechanooptoelectronic gyro. In Section 6.9.2, we derived the relation for interference fringe shifts as a function of rotation rate for a Sagnac interferometer. The fractional fringe shift, Δz (generally measured electro-optically), was shown to be related to the input rotation rate by Equation 6.115. When we include the cosine term for off-axis rotation, the relation for Δz becomes

$$\Delta z = \dot{\phi} \cos(\beta) \left[\frac{4A}{c\lambda} \right]. \quad (7.36)$$

Equation 7.36 describes the fractional fringe shift for a single optical circuit. Sensitivity of the Sagnac gyroscope can be increased by increasing the number of optical circuits; a convenient way of doing this is by using multiple turns of an OF waveguide. A very simple FO gyro is shown in Figure 7.8. Since N turns of FOC are used, the fringe shift equation is easily seen to become

$$\Delta z = \dot{\phi} \cos(\beta) \left[\frac{4NA}{c\lambda} \right] = \dot{\phi} \cos(\beta) \left[\frac{2(N2\pi R)r}{c\lambda} \right] = \dot{\phi} \cos(\beta) \left[\frac{2L_F R}{c\lambda} \right] \quad (7.37)$$

where

R is the radius of the FO coils

N is the number of turns

c is the speed of light on the OF

λ is the wavelength of the light

L_F is the total linear length of the FOC used

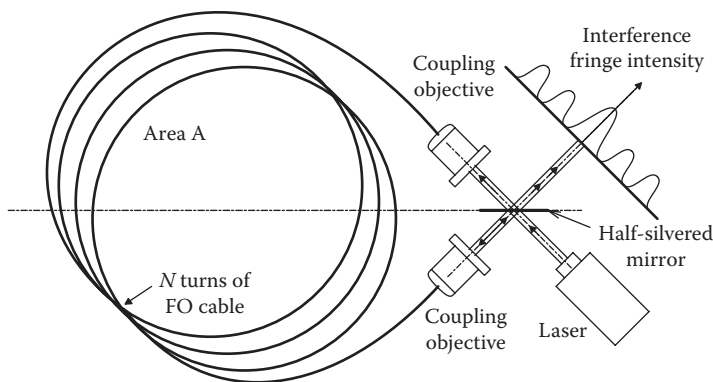


FIGURE 7.8
Simple Sagnac effect, DC, FO gyroscope.

Practical considerations concerning SNR limit the maximum number of turns used or L_F . Light intensity from the laser or LAD source is attenuated by the OF as a function of path length. This attenuation is given by the well-known relation

$$P_o = P_i 10^{-\alpha L_F} \quad (7.38)$$

where

α is the attenuation constant that is wavelength and fiber dependent

P_i is the input power in watts

P_o is the output power

L_F is the length of the fiber

The noise associated with the photodetector used to sense fringe shifts may be expressed in terms of an RMS uncertainty in the fringe shift, Δz_{RMS} . This noise is given by

$$\Delta z_{\text{RMS}} = \left(\frac{1}{2\pi} \right) \sqrt{\frac{h\nu}{\eta P_o}} \sqrt{B} = \left(\frac{1}{2\pi} \right) \sqrt{\frac{h\nu}{\eta P_i 10^{-\alpha L_F}}} \sqrt{B}. \quad (7.39)$$

Hence, the fringe shift RMS SNR can be written as

$$\text{SNR} = \frac{\Delta z}{\Delta z_{\text{RMS}}} = \frac{\phi(2L_F R / c\lambda)}{(1/2\pi) \sqrt{(h\nu B / \eta) P_i 10^{-\alpha L_F}}} = \phi \frac{4\pi R \sqrt{P_i}}{c\lambda \sqrt{h\nu B / \eta}} (L_F 10^{-\alpha L_F / 2}) \quad (7.40)$$

where

h is Planck's constant

ν is the frequency of the light in Hz

B is the Hz noise bandwidth

η is the photodetector's quantum efficiency (typically 1/2)

Regardless of ϕ , the SNR has a maximum with respect to L_F . This optimum L_F can be found by differentiating the right-hand term in parentheses in Equation 7.40 with respect to L_F and then setting the derivative equal to zero.

This procedure yields

$$L_{F(\text{OPT})} = \frac{2}{\alpha \ln(10)} = \left(\frac{2}{\alpha} \right) (0.4343). \quad (7.41)$$

The optimum number of turns, N_{OPT} , can be found from

$$N_{\text{OPT}} = \text{INT} \left[\frac{L_{F(\text{OPT})}}{2\pi R} \right] = \text{INT} \left[\frac{0.13824}{(\alpha R)} \right]. \quad (7.42)$$

Other embodiments of the Sagnac FO gyro have been developed to improve resolution by allowing more precise measurement of the fringe shift and carrier operation to avoid 1/f noise associated with DC and low-frequency measurements. By introducing AC modulation of the light phase traversing the FO coil, the intensity of the interference fringes is also AC modulated, and fringe shift detection can be made more precise and noise free because photosensor outputs can be amplified out of the range of amplifier 1/f noise. The use of single-mode, polarization-preserving OFs with attenuations (α) of less than 5 dB/km and integrated coherent optical systems has permitted the design of Sagnac FO gyros with noise-limited

sensitivities of $ca. 2.8 \times 10^{-6}$ deg/s using a 1 Hz noise bandwidth centered at the modulating frequency. By way of contrast, the Earth's rotation rate is 4.167×10^{-3} deg/s. Hence, an FO gyro with its input axis pointing north can easily measure the Earth's axial spin velocity. Sagnac fiber-optic gyros are the most sensitive angular rate sensors that have been developed to date.

7.2.2.3 Vibrating Mass Rate Gyro

The first design we will consider is the single-mass, *vibrating mass rate gyro* (VMRG). A schematic (top and side view) of the VMRG is shown in Figure 7.9. This particular design is interesting because of the following:

1. It illustrates the generality of the Lagrange method for analyzing complex electro-mechanical systems.
2. By extending the single, torsional, vibrating mass gyro to a paired structure in which the masses vibrate 180° out of phase, we can examine a crude model of the TF rate gyro.
3. The single VMRG forms a model of an ancient, biological, rotational rate sensor system that evolved millions of years ago and acts to stabilize the flight of dipteran insects (flies and mosquitoes).

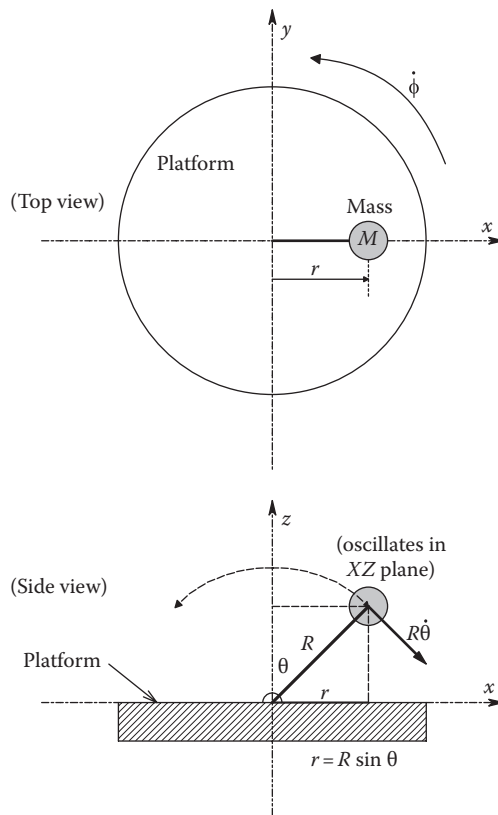


FIGURE 7.9
Top and side views of a single VMRG.

Prehistoric insects had four wings like present-day dragonflies. In the process of evolution, short-bodied dipteran flies had their hind wings morph into two (bilateral) organs known as *halteres*. A haltere has an inverted tenpin-shaped structure that is attached and pivoted at its small end from the insect's body below and behind each wing. It is basically a mass at the end of a rod that vibrates in one plane at the wing-beat frequency when the insect is flying. Flies have many mechanoreceptor neurons in the elastic cuticle at the base of each haltere that evidently sense strains in the cuticle resulting from rotation-rate-induced gyro moments (Bullock and Horridge 1965). The complex manner in which this sensory neural information is processed and used to control the flight of flies and mosquitoes remains to be described. However, we do know that if both halteres are removed, a fly no longer is capable of controlled, coordinated flight; it crashes.

To derive a mathematical description for the torques at the base of the vibrating mass, we use the well-known Lagrange method (Cannon 1967), which is based on Newtonian mechanics. For a complex mechanical system, we can write N equations of the form

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_k} \right) - \frac{\partial L}{\partial q_k} = M_k \tag{7.43}$$

where

- q_k is the k th, independent coordinate. The q_k can be rotational or translational
- M_k is the k th moment (torque) if rotational dynamics are considered or the k th force if the dynamics involve linear translation
- T is the total KE of the system
- U is the total potential energy of the system
- $L \equiv T - U$. L is called the system's Lagrangian function (Cannon 1967)

Consider Figure 7.9 in which the linear flight velocity vector of a fly is along the x -axis, and for simplicity, let the mass (haltere) vibrate only in the XZ plane. In this example, the independent coordinates are θ , the angle of the vibrating mass in the XZ plane with the z -axis, and ϕ , the input rotation around the z -axis (yaw left is shown). The system's potential energy is $U = (1/2)K_\theta\theta^2$ joules. K_θ is the spring constant of the torsion spring that restores the resting mass to align with the z -axis. A torque motor (muscles) (not shown) keeps the mass vibrating sinusoidally in the XZ plane so that $\theta(t) = \theta_m \sin(\omega_m t)$. There are two components to the system's KE:

$$T = \left(\frac{1}{2} \right) \dot{J}_o \theta^2 + \left(\frac{1}{2} \right) M (r \dot{\phi})^2 \tag{7.44}$$

where

- \dot{J}_o is the moment of inertia of the mass on the end of a rod of length R (haltere). It is well known that $\dot{J}_o = MR^2$
- r is the projection of R in the XY plane: $r = R \sin(\theta)$
- $(r \dot{\phi})$ is the tangential velocity of the mass M in the XY plane due to the input rotation rate, $\dot{\phi}$

Thus, the system's Lagrangian is

$$L = \left(\frac{1}{2} \right) \dot{J}_o \dot{\theta}^2 + \left(\frac{1}{2} \right) M (r \dot{\phi})^2 - \left(\frac{1}{2} \right) K_\theta \theta^2 = \left(\frac{1}{2} \right) \dot{J}_o \dot{\theta}^2 + \left(\frac{1}{2} \right) MR^2 \left(\frac{1}{2} \right) [1 - \cos(2\theta)] \dot{\phi}^2 - \left(\frac{1}{2} \right) K_\theta \theta^2. \tag{7.45}$$

And we have

$$\frac{\partial L}{\partial \theta} = -\left(\frac{1}{2}\right)MR^2\left(\frac{1}{2}\right)[-2\sin(2\theta)]\dot{\phi}^2 - K\theta = \left(\frac{1}{2}\right)MR^2\sin(2\theta)\dot{\phi}^2 - K\theta \quad (7.46)$$

$$\frac{\partial L}{\partial \dot{\theta}} = J_o\dot{\theta} \quad (7.47)$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) = J_o\ddot{\theta}. \quad (7.48)$$

Hence, $M_\theta = J_o\ddot{\theta} + K\theta - (1/2)MR^2\sin(2\theta)\dot{\phi}^2$ is the torque required to be produced by the muscles to maintain the haltere in simple harmonic motion.

Our interest is in the (instantaneous) torque around the pivot in the XY plane, M_ϕ , due to the input pitch angular velocity, $\dot{\phi}$. We can write as follows:

$$\frac{\partial L}{\partial \phi} = 0 \quad (7.49)$$

$$\frac{\partial L}{\partial \dot{\phi}} = Mr^2\dot{\phi} = MR^2\sin^2(\theta)\dot{\phi} = MR^2\dot{\phi}\left(\frac{1}{2}\right)[1 - \cos(2\theta)] \quad (7.50)$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\phi}}\right) = MR^2\ddot{\phi}\left(\frac{1}{2}\right) - MR^2\dot{\phi}\left(\frac{1}{2}\right)\cos(2\theta) + MR^2\dot{\phi}\dot{\theta}\sin(2\theta). \quad (7.51)$$

Thus,

$$M_\phi = \ddot{\phi}\left(\frac{1}{2}\right)MR^2 - \dot{\phi}\left(\frac{1}{2}\right)MR^2\cos(2\theta) + \left(\frac{1}{2}\right)MR^2\dot{\phi}\dot{\theta}\sin(2\theta). \quad (7.52)$$

Because $|2\theta_m| \gg 10^\circ$ in the haltere system, we cannot use the approximations: $\sin(2\theta) \cong 2\theta$, and $\sin^2(\theta) \cong \theta^2$. The M_ϕ moment can be written as

$$M_\phi = \ddot{\phi}\left(\frac{1}{2}\right)MR^2 - \dot{\phi}\left(\frac{1}{2}\right)MR^2\cos[2\theta_m\sin(\omega_m t)] + \dot{\phi}\left(\frac{1}{2}\right)MR^2\theta_m\omega_m\cos(\omega_m t)\sin[2\theta_m\sin(\omega_m t)]. \quad (7.53)$$

Thus, the M_ϕ moment lies in the XY plane and contains a DC term proportional to $\ddot{\phi}$ and also double frequency terms in $\dot{\phi}$ and $\ddot{\phi}$. The desired angular velocity term is larger by an amount ω_m that can be about 900 r/s in mosquitoes.

Note that if a haltere were mounted so that it vibrated in the animal's XY plane, then it would appear to be responsive to angular acceleration and angular velocity due to *roll and pitch*.

Gyro with two, 180° out-of-phase, vibrating masses: Figure 7.10 illustrates a side and top view of a haltere-type VMRG with two, symmetrically vibrating masses. It can be easily shown that the second harmonic torque around the two combined pivots adds in-phase for the two

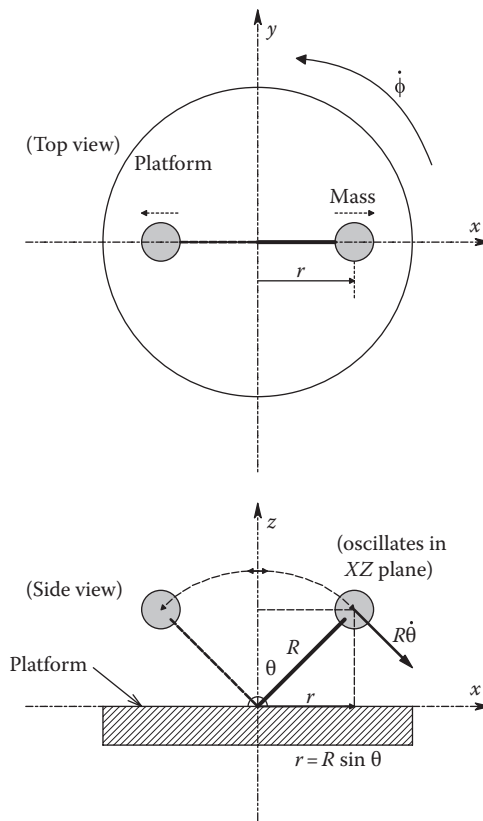


FIGURE 7.10

Top and side views of a two-mass VMRG. The masses vibrate at the same frequency, but 180° out of phase.

masses, and we get twice the second harmonic M_ϕ generated. A second advantage of this geometry is that the out-of-phase vibrations cancel the reaction forces in the x -direction. By analogy, a TF whose arms are vibrating 180° out of phase will experience a second harmonic torque around its handle when rotated around the handle axis at angular velocity $\dot{\phi}$. Such a TF design is the basis for the Systron Donner QRS14 Inertial GyroChip™ angular rate sensor.

The upper TF is made to resonate at its natural frequency; the arms vibrate in the ZX plane. Input angular rotation at $\dot{\phi}^\circ/\text{s}$ around the z -axis causes induced moments at the TF base that excite the bottom TF at its resonant frequency, but with different phases. A phase-sensitive detector (PSD) (see Section 2.8) compares the piezoelectric output of the bottom TF with a reference signal from the driven, upper TF. The DC output voltage of the LPF following the PSD is thus proportional to $\dot{\phi}$. The QRS14 has four standard ranges: ± 50 , ± 100 , ± 200 , and $\pm 500^\circ/\text{s}$. Threshold resolution for the $\pm 100^\circ/\text{s}$ is $\leq 0.004^\circ/\text{s}$, and the output noise is $\leq 0.02^\circ/\text{s}/\sqrt{\text{Hz}}$ (-103 series). The TF gyro's bandwidth can be ≤ 50 Hz.

7.2.2.4 Humphrey Air Jet Rate Gyro

This innovative angular rate sensor did not use a vibrating or rotating mass. However, its operation can be heuristically described in terms of simple Newtonian mechanics and the Coriolis effect. Its operation makes use of the fact that once a mass is in linear motion, it

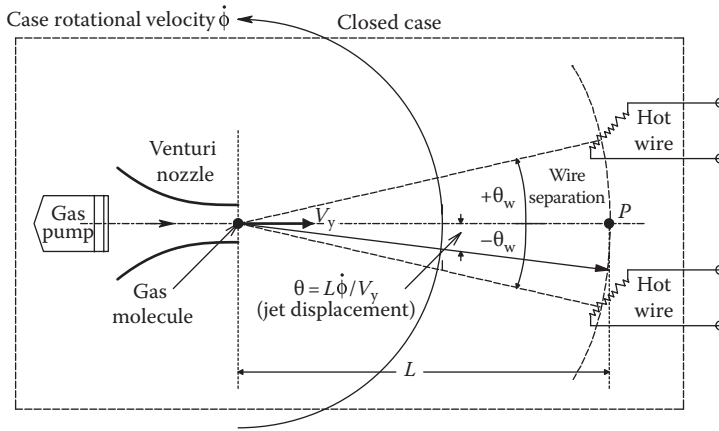


FIGURE 7.11 Top schematic view of a Humphrey air jet (Coriolis) rate gyro. (Not to scale.)

will continue to move in a straight line at constant velocity unless acted on by an external force. The geometry of the air jet gyro is shown schematically in Figure 7.11. For simplicity in the analysis, we will consider a single gas molecule mass and neglect its interactions with other gas molecules in the stream of gas emitted from the venturi and with the sides of the chamber. The gas is forced through the venturi by a novel piezoelectric gas pump.

Assume a molecule leaves the nozzle at $t = 0$ with velocity V_y in the y -direction. If the input rotation rate, $\dot{\phi}$, is zero, the molecule will arrive at a point P midway between the hot wire sensors after $T = L/V_y$ seconds, assuming the velocity remains constant. L is the distance from the nozzle mouth to point P . Now, if the input rotation is in the direction shown, the gyro's hot wires, used to sense the air jet's mean velocity, will have rotated CCW by an angle, $\theta = \dot{\phi}T = \dot{\phi}L/V_y$ radians. The distance along the circumference of the circle of radius L is $\delta = L\theta = L^2\dot{\phi}/V_{y0}$ meters. Thus, the gas molecule will strike more toward the lower hot wire instead of at P . The displacement of the molecule perpendicular to v_{y0} is a result of the Coriolis effect.

The distribution of gas molecule velocities in the jet of gas in the gyro is assumed to be modeled by a Gaussian distribution in angle at radius L . That is, at distance L , assuming zero angular velocity:

$$v_y(L, \theta, \dot{\phi} = 0) = V_y \exp\left[\frac{-\theta^2}{(2\sigma^2)}\right]. \tag{7.54}$$

Thus, at P ($\theta = 0$), $v_y = V_y$. Assume the hot wire velocity probes are separated by an angle of $\pm \theta_w$ around point P . Thus, the lower wire at $\theta = -\theta_w$ will see a gas velocity of

$$v_y(-\theta_w, \dot{\phi} = 0) = V_{y0} \exp\left[\frac{-(-\theta_w)^2}{(2\sigma^2)}\right]. \tag{7.55}$$

The upper wire also sees a velocity of

$$v_y(\theta_w, \dot{\phi} = 0) = V_{y0} \exp\left[\frac{-(+\theta_w)^2}{(2\sigma^2)}\right] \tag{7.56}$$

Now, if $\dot{\phi} > 0$, the Coriolis effect will deflect the center of the jet by an angle $\theta_\phi = \dot{\phi}/V_{y0}$. Now, the gas velocity at the *lower hot wire* is

$$v_y(-\theta_w, \dot{\phi}) = V_{y0} \exp\left[\frac{-(-\theta_w + L\dot{\phi}/V_y)^2}{(2\sigma^2)}\right] \text{ (lower wire)}. \quad (7.57)$$

The gas velocity at the *upper hot wire* is lower than at the lower wire; therefore, it is hotter and has a higher resistance:

$$v_y(+\theta_w, \dot{\phi}) = V_{y0} \exp\left[\frac{-(+\theta_w + L\dot{\phi}/V_y)^2}{(2\sigma^2)}\right] \text{ (upper wire)}. \quad (7.58)$$

Let us now indulge in some approximations to linearize the system. We will use the approximation, $e^{\pm\gamma} \cong 1 \pm \gamma$, for $|\gamma| \ll 1$. Assume $\theta_w^2 \gg 2\theta_w L\dot{\phi}/V_y \gg L^2\dot{\phi}^2/V_y^2$. Also, assume $L^2\dot{\phi}^2/V_y^2 \ll 1$. Thus, the gas velocity at the lower wire is

$$\begin{aligned} v_y(-\theta_w, \dot{\phi}) &\cong V_y \exp[-(\theta_w^2 - 2\theta_w L\dot{\phi}/V_y + L^2\dot{\phi}^2/V_y^2)/(2\sigma^2)] \cong V_y \exp[-\theta_w^2/(2\sigma^2) + \theta_w L\dot{\phi}/(V_y\sigma^2)] \\ &= V_y \exp[-\theta_w^2/(2\sigma^2)] \exp[\theta_w L\dot{\phi}/(V_y\sigma^2)] \cong V_y \exp[-\theta_w^2/(2\sigma^2)] [1 + \theta_w L\dot{\phi}/(V_y\sigma^2)]. \end{aligned} \quad (7.59)$$

Using the same development, we find that the gas velocity at the lower probe is

$$v_y(+\theta_w, \dot{\phi}) \cong V_y \exp\left[\frac{-\theta_w^2}{(2\sigma^2)}\right] \left[\frac{1 - \theta_w L\dot{\phi}}{(V_y\sigma^2)} \right]. \quad (7.60)$$

We now turn our attention to the circuit that provides the differential output from the HWAs in the Humphrey angular rate sensor. As you will see in Section 7.3.2.1, HWAs are widely used to measure gas velocity in a variety of applications, including subsonic vehicle aerodynamics and in heating and ventilating system design. The theory behind the HWA is simple: When a wire is heated, its *resistance increases*. Also, when moving gas is directed at the wire, heat is removed, its temperature drops, and its *resistance decreases*. These phenomena can be described mathematically. The resistance *increase* can be modeled by

$$R(\Delta T) = R_0 [1 + \alpha \Delta T]. \quad (7.61)$$

Here, R_0 is the wire's resistance at reference temperature, T_0 , ΔT is the temperature difference of the wire above the reference temperature caused by electrical (I^2R) heating, and α is the wire's alpha tempco (α is material- and temperature dependent). The SS ΔT can be modeled by

$$\Delta T = P\Theta \quad (7.62)$$

where

P is the average electrical power dissipated in the wire ($P = I_0^2 R$) in watts

Θ is the thermal resistance of the wire in $^\circ\text{C}/\text{watt}$

Θ depends on the wire's geometry (shape, length, diameter, etc.) and the heat conductivity and heat capacity of the gas surrounding the hot wire and also if the gas surrounding the wire is moving. In moving gas, it has been found that

$$\Theta \cong \Theta_0 - \beta\sqrt{v}. \tag{7.63}$$

Here, Θ_0 is the thermal resistance in still gas, β is a nonnegative constant, and v is the gas velocity at the wire. Using Relations (7.61)–(7.63), we can solve for $R(v)$:

$$R(v) = \frac{R_0}{1 + \alpha I_0^2 R_0 [\Theta_0 - \beta\sqrt{v}]} \text{ Ohms.} \tag{7.64}$$

Now, refer to the schematic shown in Figure 7.12. Each of the hot wires is heated by a DC current source, I_0 . The voltage across each hot wire is simply $I_0 R(v)$. Thus, the voltage across the top wire can be written as

$$V_+ = \frac{I_0 R_0}{(1 - \alpha I_0^2 R_0 \Theta) + \alpha I_0^2 R_0 \beta \sqrt{v_y (+\theta_w, \dot{\phi})}} \cong \frac{I_0 R_0 / (1 - \alpha I_0^2 R_0 \Theta)}{1 + \frac{\alpha I_0^2 R_0 \beta \sqrt{v_y (+\theta_w, \dot{\phi})}}{(1 - \alpha I_0^2 R_0 \Theta)}}$$

↓

$$V_+ \cong \frac{I_0 R_0}{(1 - \alpha I_0^2 R_0 \Theta)} \left\{ 1 - \frac{\alpha I_0^2 R_0 \beta \sqrt{v_y (+\theta_w, \dot{\phi})}}{(1 - \alpha I_0^2 R_0 \Theta)} \right\}. \tag{7.65}$$

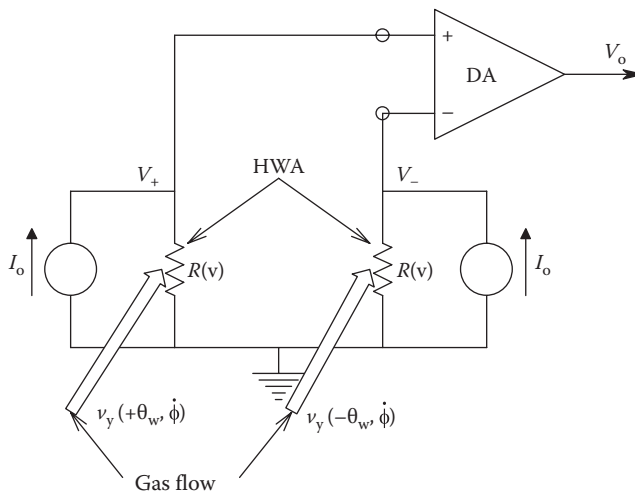


FIGURE 7.12
Simplified schematic of the HWAs used in the air jet rate gyro.

Using a similar calculation for the lower wire, we find

$$V_- \cong \frac{I_o R_o}{(1 \mp \alpha I_o^2 R_o \Theta)} \left\{ 1 - \frac{\alpha I_o^2 R_o \beta \sqrt{v_y(-\theta_w, \dot{\phi})}}{(1 \mp \alpha I_o^2 R_o \Theta)} \right\}. \tag{7.66}$$

In the calculations earlier, we have assumed that

$$\frac{\alpha I_o^2 R_o \beta \sqrt{v_y(\pm\theta_w, \dot{\phi})}}{(1 \mp \alpha I_o^2 R_o \Theta)} \ll 1, \tag{7.67}$$

and have used the approximation $1/(1 + \varepsilon) \cong 1 - \varepsilon$. Next, we use the binomial theorem to write $\sqrt{1 \pm \varepsilon} \cong 1 \pm \varepsilon/2$. This latter relation is applied to Equations 7.65 and 7.66:

$$\sqrt{v_y(\pm\theta_w, \dot{\phi})} \cong \sqrt{V_y \exp[-\theta_w^2/(2\sigma^2)]} [1 \pm \theta_w L \dot{\phi}/(V_y 2\sigma^2)]. \tag{7.68}$$

When Equation 7.68 is substituted into the approximations for V_+ and V_- , V_o can be found from

$$V_o = K_A (V_+ \mp V_-) \cong \dot{\phi} \frac{K_A I_o^3 R_o^2 \alpha \beta (\theta_w L / V_y \sigma^2) \sqrt{V_y \exp[-\theta_w^2 / 2\sigma^2]}}{(1 - \alpha I_o^2 R_o \Theta_o)^2}. \tag{7.69}$$

Thus, it appears that within the bounds of the inequality assumptions made earlier, the SS differential output of the pair of differential HWAs is proportional to the input rotation rate, $\dot{\phi}$. It should be noted that there are more sophisticated means of processing the outputs of HWAs. One such means used for large signal changes is to remove the \sqrt{v} proportionality by squaring. Another means uses feedback to automatically increase the I_o through the HWA to compensate for an increase in \sqrt{v} , thus forcing the HWA resistance to remain constant at $R(v = 0)$.

The Humphrey angular rate sensor had a built-in transport lag in its transfer function. This is simply the average time it takes a gas molecule to travel from the venturi nozzle to point P . In one Humphrey rate sensor, $L = 0.4''$ and $V_y = 80$ in./s. Thus, $\delta T = 5$ ms, the transport lag. If a step input of angular velocity is given at $t = 0$, it takes 5 ms before the sensor responds. What will the SS output voltage be if the same angular rate sensor is given a 200 Hz sinusoidal input? That is, $\dot{\phi}(t) = \dot{\phi}_o \sin(2\pi 200t)$. It appears that the Humphrey rate sensor had a response bandwidth of ca. 0–25 Hz, which was adequate for application in most vehicles (drone aircraft, boats, submarines, etc.).

7.2.3 Angular Velocity Measurement with Tachometers

A *tachometer* is a sensor/system that converts the angular velocity of a *rotating shaft* to an electrical output, generally a voltage. Tachometers are typically used to measure the angular velocity, generally in RPM, of rotating machines such as electric motors, generators, turbines, pumps, and internal combustion engines.

A commonly found form of tachometer is the DC, permanent magnet generator, in which the output voltage linearly follows the shaft velocity. Shaft velocity is read by a

calibrated, DC voltmeter. Response time of the DC generator tachometer is basically set by the mechanical natural frequency of the analog voltmeter. In the case of a digital voltmeter, the response time is limited by the low-pass filtering at the input ADC.

Other tachometers avoid the expense of the DC generator and use either optical or magnetic means to generate a train of pulses whose repetition rate is proportional to the shaft speed. In the tachometer illustrated in Figure 7.13, a light beam is bounced off one or more reflecting strips cemented to the rotating shaft. Every time the shaft comes around, the light beam is reflected onto a photosensor, producing a train of pulses that is shaped by a comparator and one-shot multivibrator. The average value of this conditioned pulse train is seen to be proportional to the shaft's RPMs. An op amp LPF averages the pulse train and is also used to correct for DC offset and set the output calibration, as shown in Figure 7.13.

An alternate means of generating a pulse train whose rate is proportional to shaft speed is to cement permanent magnets to the rotating shaft and place a digital Hall sensor in a fixed location near the path of the magnet. Every time the magnet comes close to the Hall sensor, its output changes state, generating the pulse train. The disadvantage of the pulse averaging tachometer is that the LPF time constant needs to be long to suppress output voltage ripple at low RPM. The long time constant means a slow response to changes in RPM.

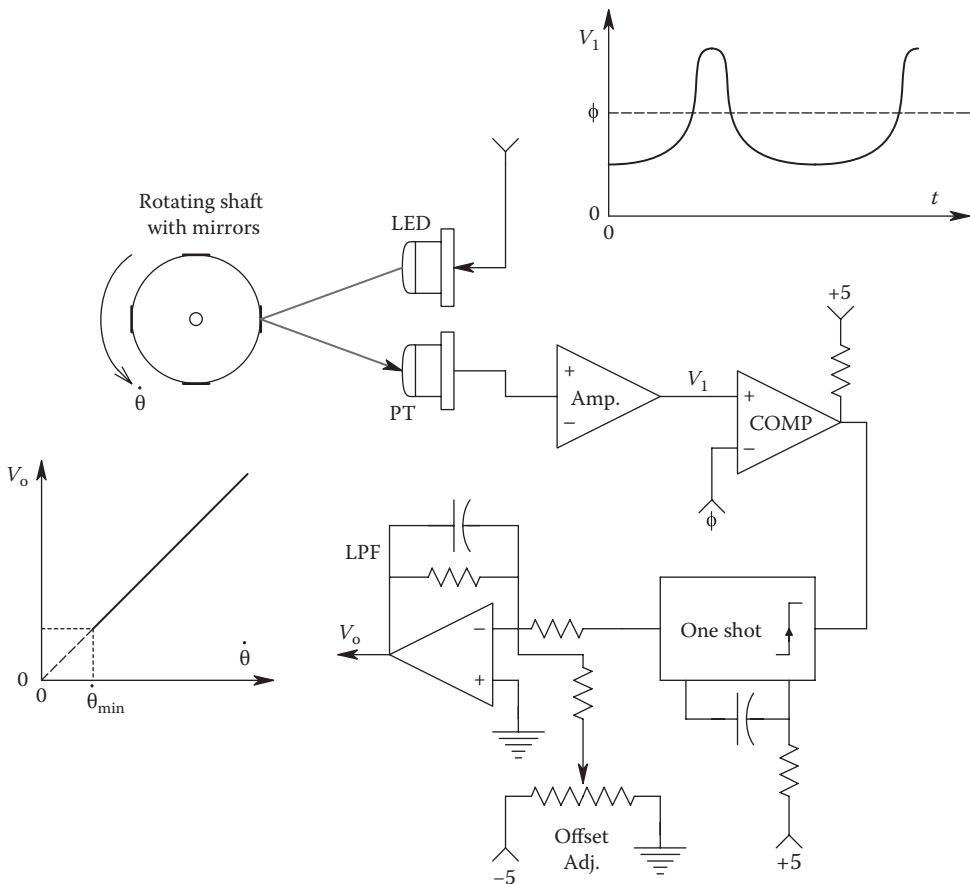


FIGURE 7.13

Schematic diagram of a simple photoelectric tachometer system. Standard duty-cycle pulses are LPF averaged to yield a DC voltage $V_o \propto \dot{\theta}$.

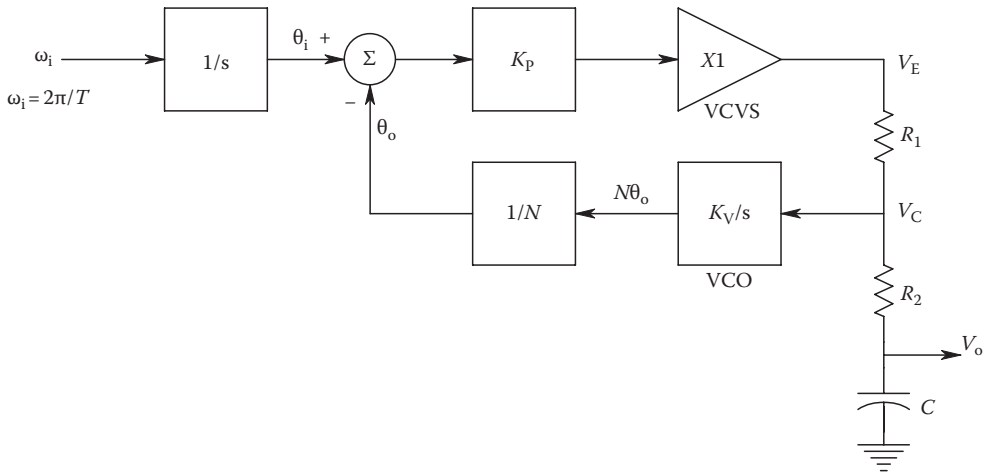


FIGURE 7.14 Block diagram of a 4046 CMOS PLL used as a tachometer. The (virtual) input integrator converts the input event rate (frequency, ω_i) to the phase input, θ_i , of the PD.

One way to get around the problem of a long filter time constant being required for ripple-free output at low RPMs is to use a PLL to convert the rate of the pulse train to a DC output proportional to RPM. The block diagram of a PLL tachometer is shown in Figure 7.14. A 4046 CMOS PLL is used to convert input frequency (ω_i) to a proportional DC output voltage, V_o . In the block diagram, the first integrator is required to convert the input signal’s frequency to phase, the quantity the PLL operates on. In the SS, $\theta_i = \theta_o$, and it is easy to show that the DC output voltage is given by

$$V_o = \frac{2\pi f_i N}{K_V} \tag{7.70}$$

where

- N is a digital frequency divider ratio ($N = 1, 2, 4, \dots 64, \dots$)
- K_V is the PLL VCO’s constant in (r/s)/V

Reduction of the block diagram yields the transfer function

$$\frac{V_o}{f_i}(s) = \frac{2\pi N / K_V}{s^2 C(R_1 + R_2)N / (K_P K_V) + s(1 + CR_2 K_P K_V / N)N / K_P K_V + 1} \tag{7.71}$$

where K_P is the phase detector (PD) gain. The PLL tachometer is seen to be a second-order system with a natural frequency of $\omega_n = \sqrt{K_P K_V / (C(R_1 + R_2)N)}$ tr/s. In this design, the VCO runs at N times the input frequency at lock. This allows the measurement of very low frequencies. All VCOs have a minimum frequency at which they will oscillate, as well as a maximum frequency. The minimum and maximum VCO frequencies, divided by N , set the input range of this PLL tachometer.

The final means for the measurement of angular velocity that we will describe can be used with turbine wheels as well as rotating shafts. A system known as an *instantaneous pulse frequency demodulator* (IPFD) operates on a pulse sequence obtained by the rotation

of magnets attached to the turbine wheel or shaft passing fixed Hall switch or, alternately, by the rotation of illuminated, alternating white and black markings past a fixed photo-sensor. The IPFD measures the period or time interval between each pulse pair in the sequence and then calculates the reciprocal of that interval. The reciprocal of the i th interval is defined as the i th element of instantaneous frequency of the pulse sequence, r_i , and is proportional to the angular velocity of the shaft or wheel. If, for example, each revolution produces 36 pulses, then 2400 RPM would produce 86,400 ppm or 1440 pps. Since the IPFD is responsive to the reciprocal of the length of each period, any change in the speed of rotation from 2400 RPM will be sensed within about 0.7 ms. This is a considerably faster response time than is afforded by pulse averaging or PLL tachometers. The IPFD is thus a useful system to study subtle, short-term changes in angular velocity. A block diagram for an analog IPFD tachometer developed by the author is shown in Figure 7.15A. The processing time, that is, the time required to charge the capacitor to the maximum voltage, sets the maximum possible instantaneous frequency (shortest input period). The maximum input

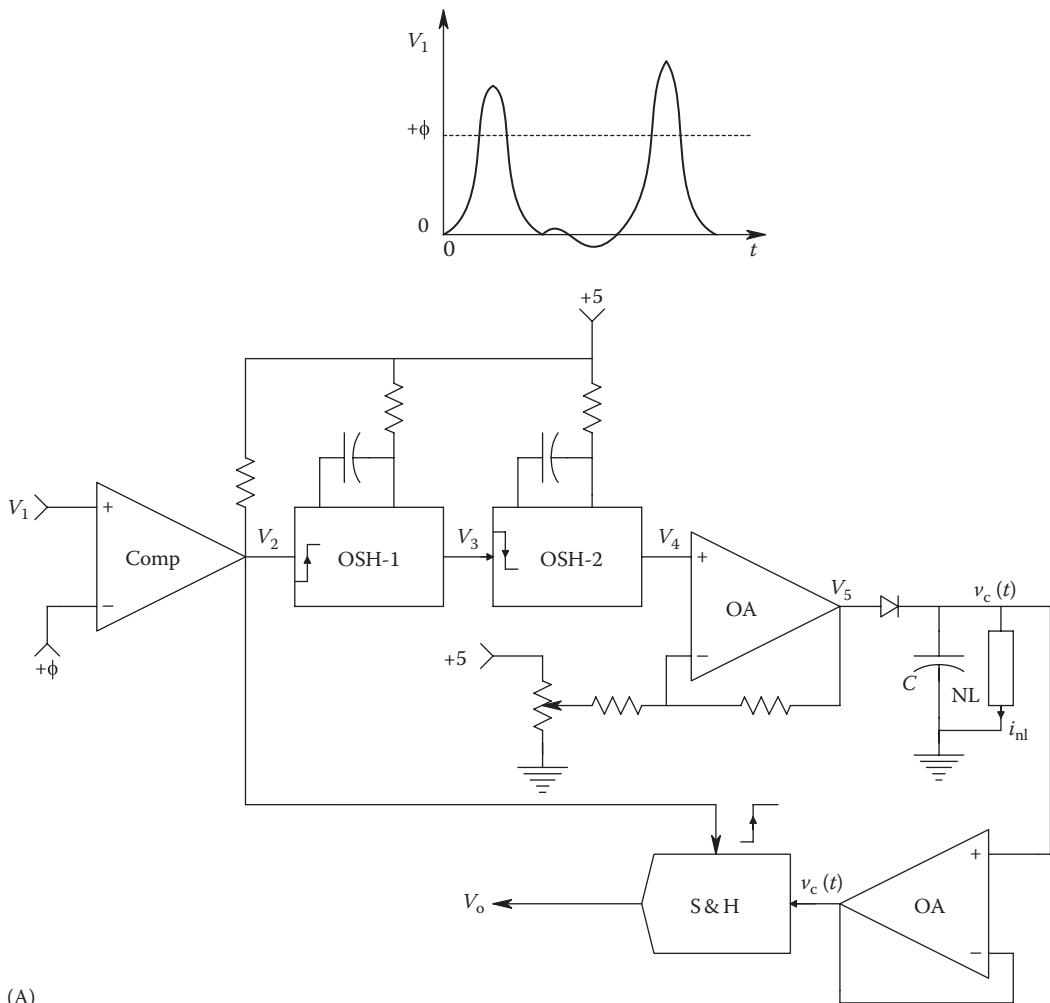


FIGURE 7.15

(A) Schematic of an analog IPFD designed by the author.

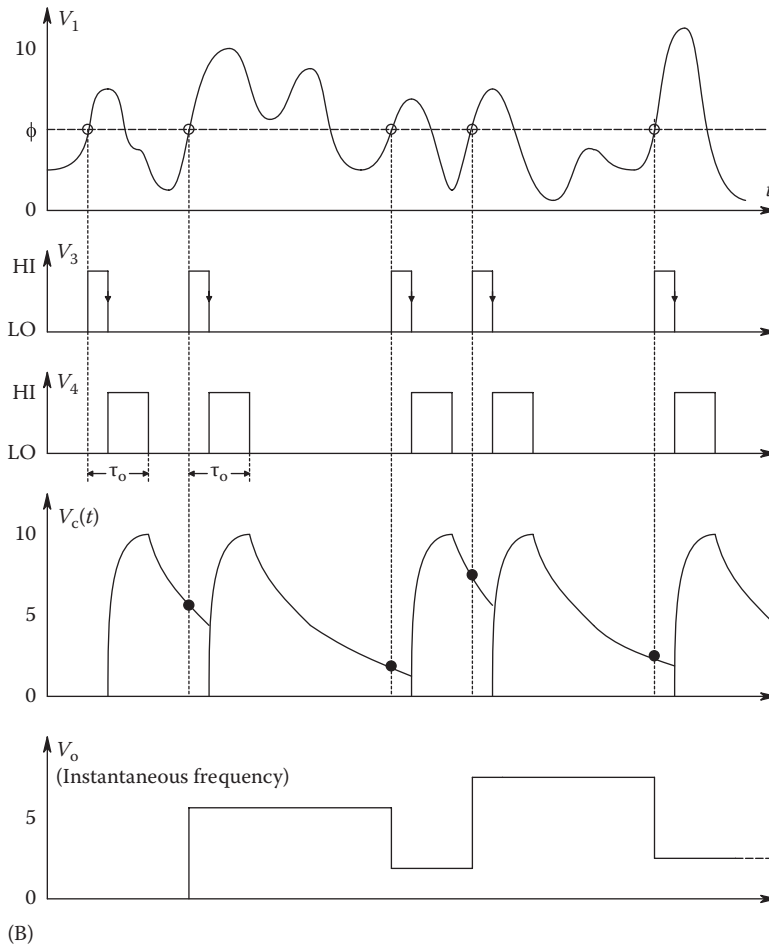


FIGURE 7.15 (continued)

(B) Waveforms observed in the analog IPFD. Note that the output voltage appears stepwise; the height of the step is proportional to the reciprocal of the interval between the wop preceding pulses.

period (lowest input frequency) is determined by the lowest capacitor voltage at which the decay waveform follows the curve given by

$$v_c(t) = \frac{C/\beta}{t + \tau_o} \tag{7.72}$$

V_o is the peak voltage the capacitor is charged to at time $t = 0$, when the voltage begins decaying hyperbolically. Time τ_o is the processing time between the $(i - 1)$ th pulse and the beginning of the hyperbolic decay, $v_c(t)$, which is sampled at time t_i when the i th pulse occurs. $\beta \equiv C/(\tau_o V_o)$. The hyperbolic decay of voltage from the capacitor occurs because the capacitor is discharged through a nonlinear resistor with the volt-ampere relation:

$$i_{nl} = \beta v_c^2 \tag{7.73}$$

Hence, the node equation for the capacitor discharge is

$$\dot{v}_c + \left(\frac{\beta}{C}\right)v_c^2 = 0, \quad t \geq 0. \quad (7.74)$$

This is a form of *Bernoulli's equation*, which has a solution for the initial condition of $v_c = V_o$ at $t = 0$ given pulse, generating a stepwise voltage waveform, $q(t)$. The height of each step is given by Equation 7.87. The value of the hyperbola is sampled and held at the occurrence time of the $(i + 1)$ th instantaneous frequency of the preceding pulse pair interval and is proportional to RPM. In mathematical terms, the IPFD output can be written as

$$q(t) = k \sum_{i=1}^{\infty} r_i [U(t - t_i) + \mathbf{t}U(t - t_{i+1})]. \quad (7.75)$$

Here, $U(t - a)$ is the unit step function, zero for $t < a$ and 1 for $t \geq a$. r_i is the i th element of instantaneous frequency, defined over $t_i \leq t < t_{i+1}$ as

$$r_i \equiv \frac{1}{(t_i - t_{i-1})} \text{ s}^{-1}. \quad (7.76)$$

Figure 7.15B illustrates key waveforms in the operation of the analog IPFD.

Webster (1978) described a digital IPFD that was used as a cardiometer (to measure heart rate). The Webster system had an analog output that is also given by Equation 7.90. Illustrated in Figure 7.16, the reciprocal relation between interpulse interval and r_i is ingeniously generated by the use of a 12-bit DAC whose output resistance increases linearly with the size of its binary input, N_i . N_i is proportional to $(t_i - t_{i-1})$. A timing diagram for the Webster IPFD is shown in Figure 7.17. The op amp's output is thus $V_o = V_R R_F / R_{DAC} = K r_i$ over the interval (t_i, t_{i+1}) .

In summary, we have seen that tachometers can have a wide variety of embodiments, ranging from simple DC generators whose output voltage is proportional to shaft angular velocity to systems that generate pulse trains whose frequencies (and instantaneous frequencies) are proportional to angular velocity. The pulse trains can be processed by averaging PLLs or IPFDs to obtain DC signals proportional to angular velocity.

7.2.4 Angular Position Measurement with Gyroscopes

Modern inertial navigation systems and attitude control systems such as those used by space vehicles, helicopters, aircraft, submarines, etc., need estimates of *angular position* for their operation. There are several simple modifications of mechanical and Sagnac FO gyroscopes that will allow them to sense angular position, rather than just angular velocity. These include the RIG and various servo platform gyroscope configurations.

The RIG is the same as the rate gyro shown in Figure 7.5, except that the spring is removed. The linearized gyro equations now become

$$M_\theta = -B\dot{\theta} = J_1 \ddot{\theta} - H_r \dot{\phi} \quad (7.77A)$$

$$M_\phi = H_r \dot{\theta} + J_1 \ddot{\phi}. \quad (7.77B)$$

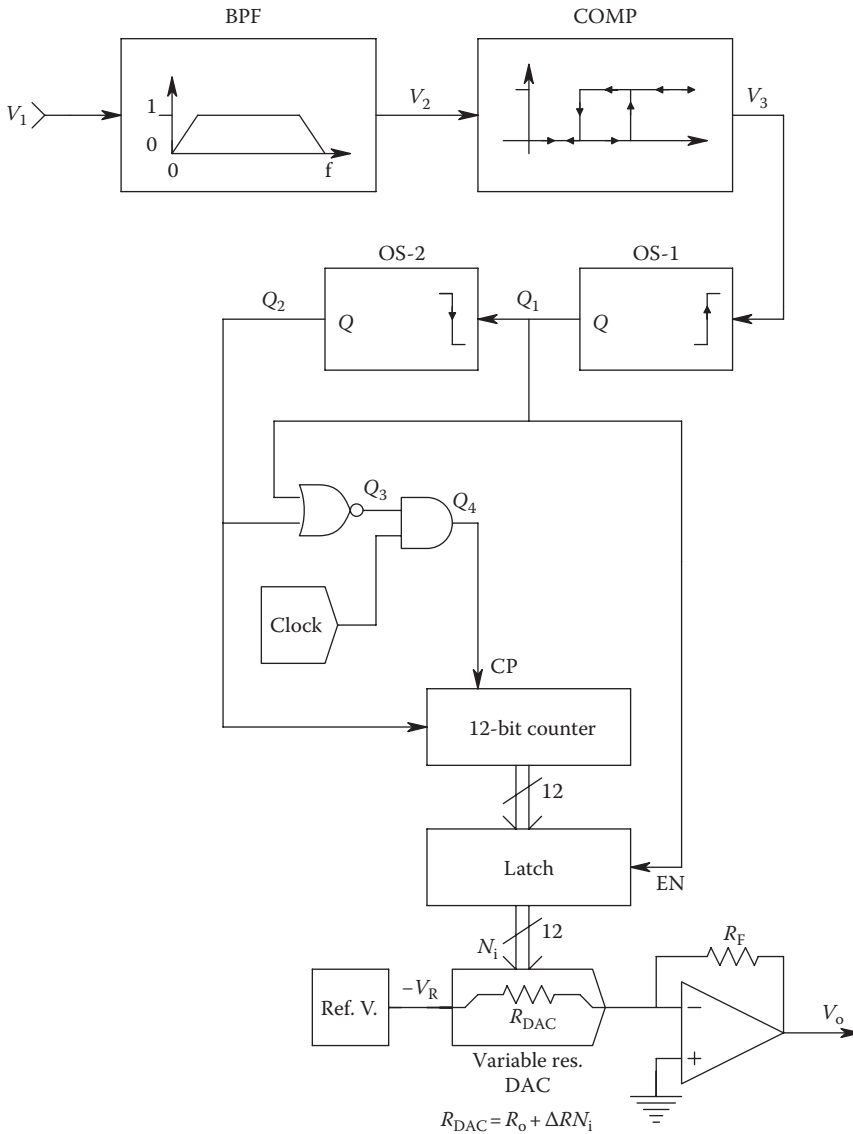


FIGURE 7.16
Digital IPFD (with analog output) described by Webster (1978).

The solution of Equations 7.92A and 7.92B yields

$$\frac{\Theta}{\Phi}(s) = \frac{H_r/J_1}{s + B/J_1} \tag{7.78}$$

Thus, in the SS, from the Laplace final value theorem, a constant rotation ϕ_o around the input axis will produce a constant inner gimbal deflection of $\theta_{ss} = \phi_o H_r/B$ radians. The RIG is seen to be approximately a first-order, low-pass sensor system with a time constant of J_1/B seconds.

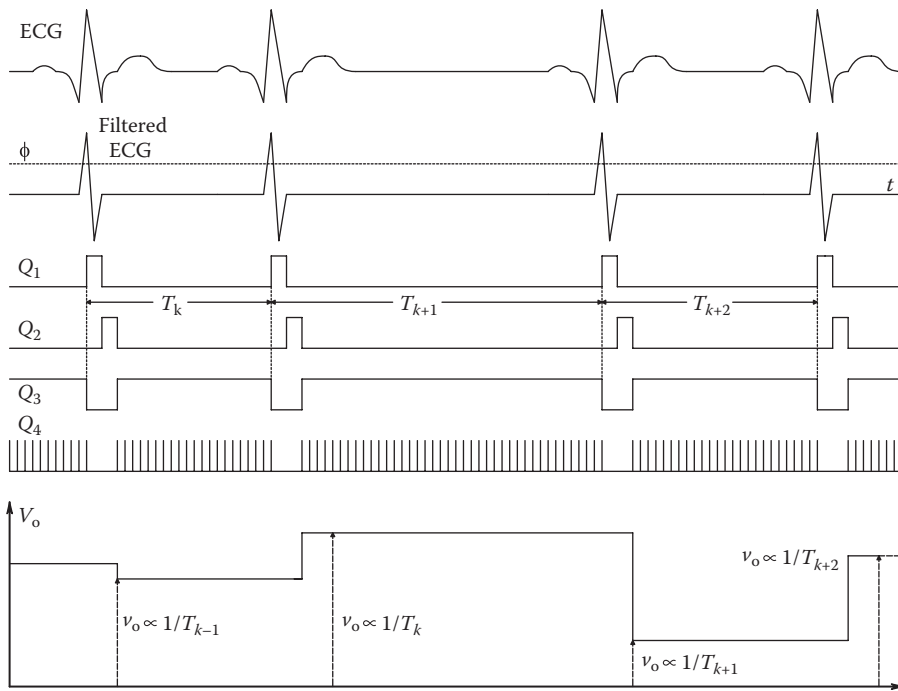


FIGURE 7.17 Timing diagram for the Webster IPFD of Figure 7.16. This IPFD is used as an instantaneous frequency cardiometer.

Figure 7.18 illustrates the use of a feedback platform to obtain an output signal proportional to the SS input rotation, ϕ_{io} . In this case, we use a servomotor-driven platform to counterrotate the RIG as the assembly rotates with the vehicle. The inner gimbals' angle, θ , is sensed by a potentiometer or optical sensor. The servo-gyro equations are

$$\phi_e = \phi_i \mp \phi_m \tag{7.79A}$$

$$M_{\theta} = 0 = B\dot{\theta} + J_1\ddot{\theta} - H_t\dot{\phi}_e \tag{7.79B}$$

$$M\phi = H_t\dot{\theta} + J_i\ddot{\phi}_e. \tag{7.79C}$$

The platform position is driven by the θ signal through a proportional plus derivative compensated servomotor.

The platform position may be related to θ by the relation

$$\frac{\Phi_m}{\Theta}(s) = \frac{K_M K_A (\tau_c s + 1)}{s(\tau_M s + 1)}. \tag{7.80}$$

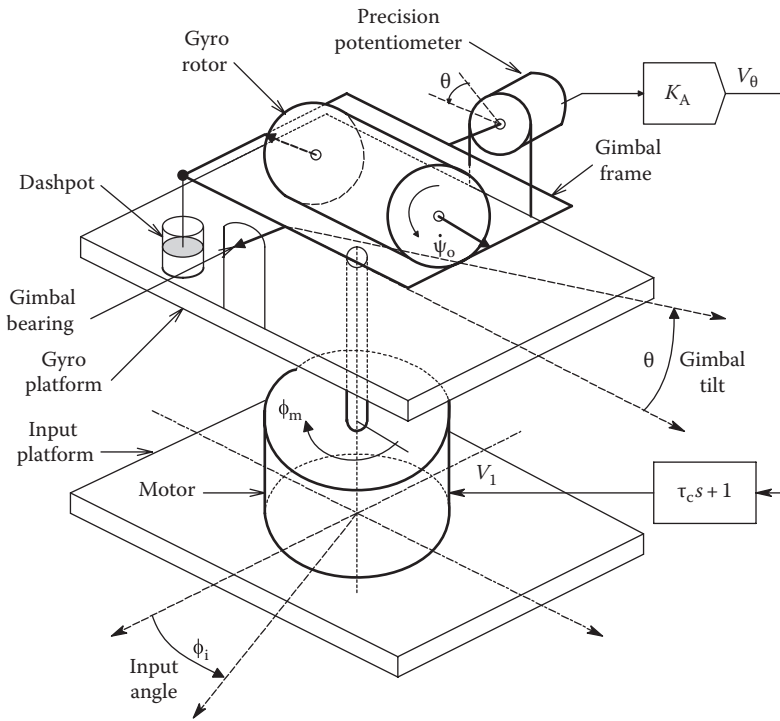


FIGURE 7.18 Diagram of a servoplatform, RIG. In the SS, $\phi_m = \phi_i$. Proportional plus derivative feedback is required for stability.

The block diagram describing the servo-RIG is shown in Figure 7.19. System output is proportional to the platform angle, ϕ_m . From the block diagram, we can write the servo-RIG's transfer function:

$$\frac{V_o}{\Phi_i}(s) = \frac{K_o(\tau_c s + 1)}{\frac{3}{2} J_1 \tau_M / (H_r K_A K_M) + s^2 (B \tau_M + J_1) / (H_r K_A K_M) + s (B + \tau_c H_r K_A K_M) / (H_r K_A K_M) + 1} \quad (7.81)$$

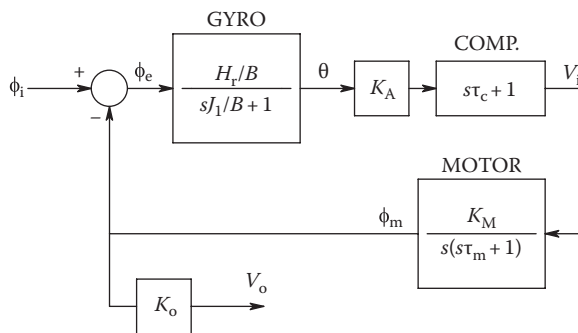


FIGURE 7.19 Systems block diagram describing the dynamics of the servo-RIG of Figure 7.18.

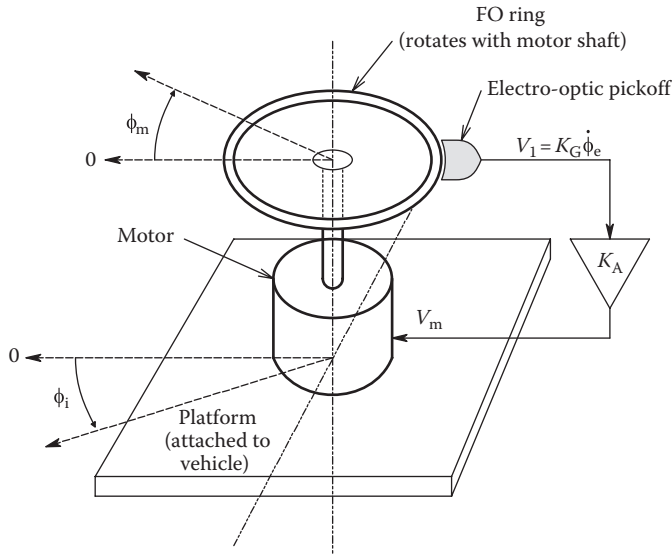


FIGURE 7.20
Diagram of a servoplatform FO (Sagnac) gyro.

In the SS, $V_o = K_o \Phi_i$. The advantage of the servo-RIG configuration is that sensitivity is set by the parameter, K_o , and the dynamic range is not limited by inner gimbal tilt angle, θ . In fact, it is easy to show that in the SS, $\theta/\Phi_i \rightarrow 0$. The natural frequency of the servo-RIG is seen to depend on the gains K_A and K_M .

A Sagnac FO gyro can also be placed on a servo platform to realize a RIG. Figure 7.20 illustrates such a system. As in the previous example, the platform counterrotates against the input angle so we have

$$\phi_e = \phi_i \mp \phi_m. \tag{7.82}$$

The Sagnac FO gyro SS output can be summarized by

$$V_1 = K_G \dot{\phi}_e. \tag{7.83}$$

The Sagnac gyro output is further conditioned by an amplifier, K_A , whose output acts on the servomotor system. The servo-gyro block diagram is shown in Figure 7.21. Again, the system output is proportional to the motor rotation, which can be measured with a potentiometer or a coding disk:

$$\frac{V_o(s)}{\Phi_i} = \frac{K_o}{s\tau_m / (K_G K_A K_M) + 1}. \tag{7.84}$$

In the SS, we find that $V_o = K_o \phi_{io}$, assuming $K_G K_M K_A \gg 1$, and the fringe shift is nulled to zero.

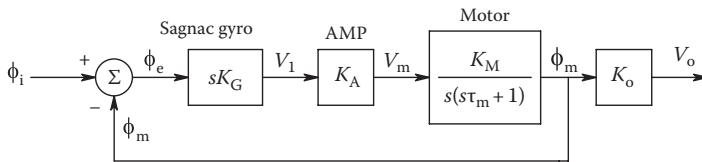


FIGURE 7.21
Systems block diagram describing the dynamics of the servo-FO gyro of Figure 7.20.

Placing a gyroscope on a servo-controlled platform is seen to be an effective means of increasing the gyro's dynamic range of response and its natural frequency. In designing such a system, it is clear that there is a trade-off between the considerable extra cost of the servo platform and the increase in performance obtained.

7.2.5 Angular Position Measurement with Clinometers

Gyroscopes, discussed in the previous two sections, were seen to operate either in a gravitational field or in a microgravity environment, such as in an orbiting satellite vehicle (SV). Clinometers, on the other hand, require gravity to work. There are two classes of clinometer: pendulum and liquid filled (the liquid is either a dielectric or an electrolyte, depending on the design). Clinometers are used on ships to sense roll and pitch angle and on heavy earth-moving machinery and cranes to sense tilt. Obviously, excessive tilt or roll can lead to capsizing and loss of life and/or the equipment. A clinometer is considerably simpler in design than is a mechanical gyroscope and consequently is a far less expensive instrument. A schematic design of a *pendulum clinometer* is shown in Figure 7.22. The roll axis is perpendicular to the plane of the page. A pendulum of total mass, M , and moment of inertia, J , is suspended in the center of a cylindrical case. The case is filled with light oil for viscous damping, and the position of the clinometer pendulum relative to the case is measured by a potentiometer wiper with negligible friction. A torsion spring acts to restore the pendulum to the center position of the case. Summing the torques around the pivot point, we can write

$$0 = -J\ddot{\theta}_p + B(\dot{\theta}_c - \dot{\theta}_p) + K_s(\theta_c + \theta_p) - MgR \sin(\theta_p). \tag{7.85}$$

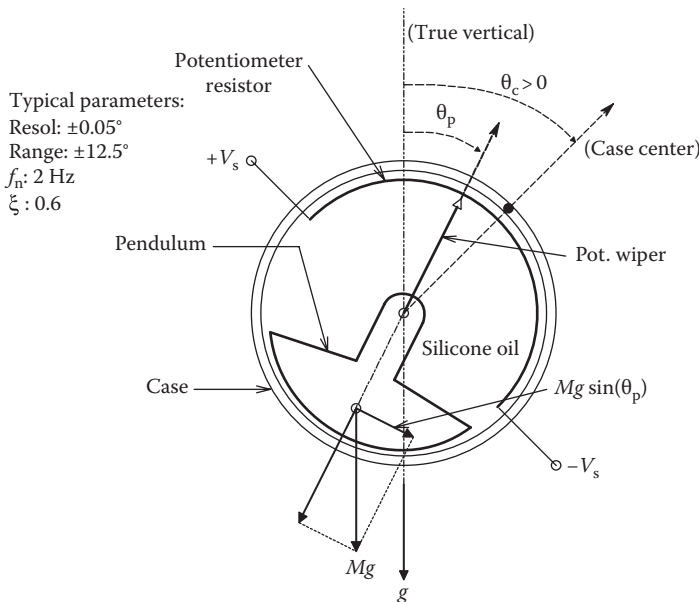


FIGURE 7.22 Vertical section through a passive, pendulum clinometer. The axis around which the clinometer tilts is perpendicular to the plane of the page.

This ODE can be linearized by assuming $|\theta_p| < 15^\circ$, so $\sin(\theta_p) \approx \theta_p$ (in radians). Laplace transforming and separating Θ terms, we can write

$$\frac{\Theta_p}{\Theta_c}(s) = \frac{(sB + K_s)}{s^2J + sB + (K_s + MgR)}. \tag{7.86}$$

Now, the clinometer output is assumed to be given by

$$V_o = K_o(\theta_c - \theta_p). \tag{7.87}$$

Substituting Equation 7.101 in the relation for V_o , we obtain finally

$$\frac{V_o}{\Theta_c}(s) = \frac{K_o(s^2J/MgR + 1)MgR}{[s^2J/(K_s + MgR) + sB/(K_s + MgR) + 1](K_s + MgR)}. \tag{7.88}$$

From Equation 7.103, we see that the I/O transfer function of the pendulum clinometer is that of a *mechanical notch filter*. The transfer function has a pair of conjugate zeros on the $j\omega$ axis in the s -plane at $\omega_o = \sqrt{MgR/J}$ r/s and a pair of complex-conjugate poles with $\omega_n = \sqrt{(K_s + MgR)/J}$ r/s. A typical SS frequency response of the pendulum clinometer to $\theta_c(t) = \theta_o \sin(\omega t)$ is shown in Figure 7.23. The response is flat from DC to just below the notch frequency, ω_o . In order that the pendulum follow rapid changes in θ_c , $\omega_o = \sqrt{MgR/J}$ should be made as large as possible. In the limiting case, the pendulum can be considered to be a mass point, M , at a distance R from the pivot. It is well known that such a pendulum has a moment of inertia given by $J = MR^2$. Thus, the notch frequency of the pendulum is independent of its mass and is given by

$$\omega_o = \sqrt{\frac{g}{R}} \text{ r/s}. \tag{7.89}$$

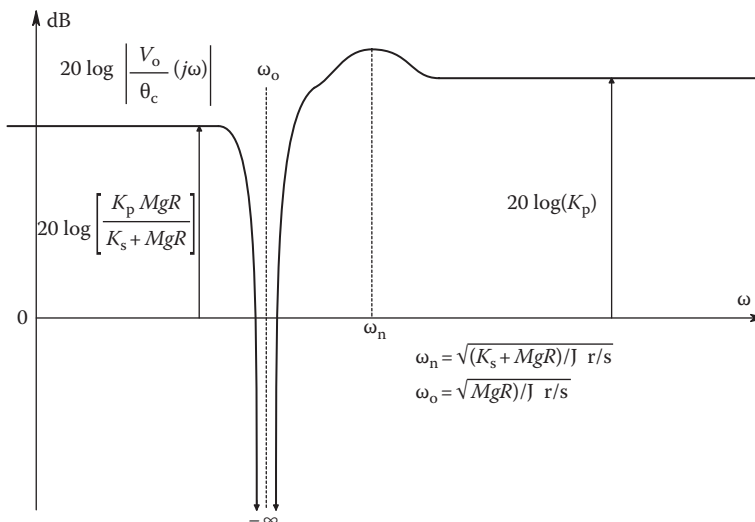


FIGURE 7.23 SS frequency response of the pendulum clinometer, subject to a sinusoidal tilting of its case. Note that it is a mechanical notch filter.

Hence, a 10 cm pendulum in the earth's gravitational field will have an f_0 of about 1.6 Hz; a 2 cm pendulum has an f_0 of 3.5 Hz. The measurement of most vehicle roll angles is done under almost static conditions, so the dynamic performance of the clinometer angle sensor is generally not an issue. Even the roll rate of large ships is slow enough so that most pendulum clinometers can follow the roll angle accurately.

A second class of clinometer has no moving mechanical parts. Figure 7.24A illustrates the simplified sectional drawing of a dielectric liquid-filled, differential, capacitive-type clinometer. This design principle is used by Measurement Specialties™ in their AccuStar® line of electronic clinometers. A sealed, thin, cylindrical chamber is half filled with a dielectric liquid with a high dielectric constant. There are no moving parts. The top half is filled with a dry inert gas. A differential capacitor is formed by depositing a film of noble metal on glass or plastic plates: one electrode is a disk and the other two are sector shaped

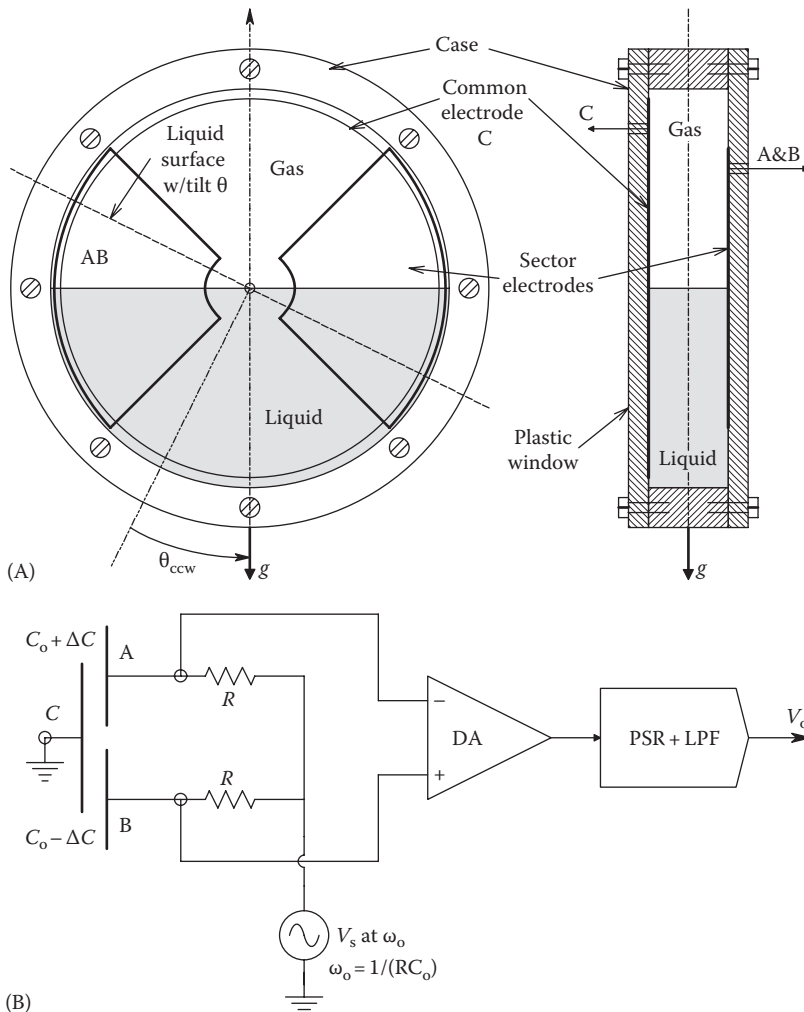


FIGURE 7.24 (A) Front and side view of a dielectric liquid-filled clinometer. The sector electrodes and metal backplate make a differential capacitor system. The differential capacitance is a linear function of the case tilt angle, θ . (B) Schematic of a De Sauty bridge used to sense $V_o \propto \Delta C \propto \theta$. Maximum sensitivity occurs when $\omega_0 = 1/(RC_0)$.

so that when the dielectric liquid between the plates tilts, an equal area of the sector plates covered by the liquid is added or subtracted to the area covered at zero tilt. Since the area of a parallel plate capacitor is proportional to the area of the plates, it is easily seen that

$$\Delta C = \frac{\kappa \epsilon_0 \Delta A}{d} \quad (7.90)$$

where

κ is the filling liquid's dielectric constant

ϵ_0 is the permittivity of empty space, $8.85 \times 10^{-12} \text{ C/N-m}^2$.

From the electrode geometry, $\Delta A = k\theta$, so it is clear that $\Delta C = K_c\theta$.

The differential change in capacitance caused by tilting the clinometer case can be sensed by using a simple De Sauty bridge as shown in the schematic in Figure 7.24B. From the schematic, we can write

$$f_{E_{\max}} = 3.15 / (0.2RC) \text{ pps.} \quad (7.91)$$

Now, set $\omega \equiv 1/RC_0$ r/s, so

$$V_o = K_A V_s \left\{ \frac{2j\Delta C/C_0}{[(j+1) + j\Delta C/C_0][(1+j) - j\Delta C/C_0]} \right\} \cong K_A V_s \frac{\Delta C}{C_0}. \quad (7.92)$$

Here, we assumed $(\Delta C/C_0)^2 \ll 2$. Thus, the bridge output voltage is proportional to the case tilt, θ . The AccuStar® clinometer has a total range of $\pm 60^\circ$ and a linear range of $\pm 45^\circ$. Its resolution is 0.001° , its time constant is 0.3 s, and its -3 dB bandwidth is 0.5 Hz. AccuStar offers analog voltage output, 16-bit serial digital output, and a digital pulse-width output model.

An interesting variation on the dielectric liquid AccuStar® clinometers is the electrolyte liquid clinometers manufactured by HL Planartechnik GmbH of Dortmund, Germany, now a subsidiary of Measurement Specialties. Their D- and DPG-Series inclinometer sensors use a conductive liquid instead of a dielectric liquid and operate as two, differential arms of an AC-powered Wheatstone bridge. The electrodes must be a noble metal such as gold or platinum to avoid corrosion. Their DPG clinometers offer ranges of $\pm 5^\circ$, $\pm 10^\circ$, $\pm 15^\circ$, and $\pm 30^\circ$ with a resolution of 0.001° . Output options include RS-232 digital format at 2400 or 9600 baud, analog voltage or current, a pulse-width modulation (PWM) signal, or switch output signals.

7.2.6 Angular Position Measurement of Shafts

We have already discussed the use of optical coding disks in Section 6.9.1 and the use of precision potentiometers in Section 6.2.5 in the measurement of shaft angles. Another magnetic sensor used in angular position sensing is the Inductosyn® (Finden and Horlock 1957), manufactured by Farrand Controls, Ruhle Inductosyn International Division (Farrand 1996, Ruhle 2009). Inductosyns have been used in the inertial navigation systems, the periscopes, and the fire control systems of Polaris, Poseidon, and Trident submarines and in controlling rotary jig borers to achieve positioning accuracies of ± 3 arcsec

and ± 0.0002 in. They have also been used as position-sensing transducers in the *remote manipulator system arm* installed in the cargo hold of the US space shuttle. They also have applications in missile fire control systems such as the British Rapier missile launcher.

Physically, the rotational Inductosyn consists of two, opposing, flat, toroidal cores, each with a flat printed circuit coil with repeated, parallel hairpin turns attached to one flat surface of the flat toroidal cores. The length of one complete cycle of the hairpin coil is called the pitch, P . Inductive coupling between the two flat coils is used to measure the angular displacement between the two coils. An AC current, generally between 2.5 kHz and 100 kHz in frequency, is used as the excitation to one coil. The voltage induced in the pickup coil is given by

$$v_i(t) = kV_s \cos\left(\frac{2\pi\theta}{P}\right) \sin(2\pi f_e t) \quad (7.93)$$

where

k is the coupling coefficient

V_s is the input voltage

θ is the angular displacement between the Inductosyn disks

P is the angular period of one hairpin turn

f_e is the frequency of the excitation voltage

If a second pickup winding is located adjacent to the first, but given an angular displacement of $P/4$ with respect to the first pickup coil, the voltage induced in the second coil is given by

$$v'_i(t) = kV_s \sin(2\pi\theta/P) \sin(2\pi f_e t). \quad (7.94)$$

Only the amplitudes of v_i and v'_i change with θ/P ; their phase and frequency remain the same. If the space quadrature windings of an Inductosyn are excited by constant amplitude carriers in time quadrature, the resulting output signal has a constant amplitude, but its phase undergoes a 360° shift for an angular rotation of $\theta = P$.

Inductosyns (Farrand 1996, Koujili 2008) can be divided into *absolute rotary sensors* and *incremental rotary sensors*. Inductosyns that have only one pattern of coils with a large number of electrical cycles have a small absolute measurement span ($\theta \ll 360^\circ$) and are suitable for sensing incremental position data. Excitation frequencies can range from 200 Hz to 200 kHz, but the typical excitation is 2V peak-to-peak at 10 kHz.

Inductosyns that are used for absolute measurements where θ can range up to 360° have two coil patterns: in one approach, a coarse one-cycle pattern is combined with a fine multicycle pattern. The outputs of both patterns are digitized and combined in synchronization or correlation logic to provide an output that is both accurate and absolute. In the other approach, two multicycle patterns are used, which differ by one cycle. The outputs of the two, $N/(N-1)$, patterns are digitized and subtracted. The result of this subtraction is digital coarse data that are then combined with one of the digitized fine outputs to provide an output that is again accurate and absolute.

The available nominal outside diameters of Inductosyn *rings* range from 2 to 42 in. Limiting accuracies for Inductosyn diameters of 3", 4", 7", and 12" are ± 2 , ± 1.5 , ± 1.0 , and ± 0.5 s of arc, respectively. Initial alignment and concentricity of the Inductosyn rings is essential in order to realize the high angular precision cited earlier. Care also must be

taken to avoid electrostatic coupling and ground loops to avoid artifact errors in the operation of these sensors. Inductosyns are used with custom electronics packages to give a DC voltage output proportional to angular position between the coils. Because there is no physical contact required between the Inductosyn stator and rotor, there are no brushes or wipers to wear out, similar to the case of optical coding disks.

Farrand also makes a family of linear Inductosyns to sense absolute position displacement. In some low-bandwidth applications, these sensors are competitive with LVDTs (see Section 6.2.5.1).

7.3 Measurement of Linear Acceleration, Velocity, Displacement, and Position

In this section, we examine certain means to measure the linear acceleration of structures and vehicles with sensors known as *accelerometers*, as well as certain techniques for measuring linear velocity and position. The output of a linear accelerometer can be integrated to obtain a velocity estimate; a second integration yields a position or deflection estimate. There are also other means for finding the velocity of particles, fluids, structures, and vehicles, including the use of the Doppler effect with light, microwaves, or sound. Knowing the precise position of a vehicle is really a navigation problem, and if done from within a closed vehicle such as a submerged submarine or an aircraft flying under IFR, it is an inertial navigation problem. On the other hand, the relative position of small objects and structures can be measured with a wide variety of position sensors, including ones that do not physically touch the object. This latter class of no-touch position sensors includes the use of ultrasound, microwaves, and light in various schemes that utilize the waves' propagation velocity, interferometry, or phase-lock techniques.

7.3.1 Accelerometers

7.3.1.1 Basic Newtonian Accelerometer

Figure 7.25 illustrates a basic, passive, moving-mass, linear accelerometer. A mass, M , is restrained to slide inside a cylindrical case. There is viscous damping, B , between the case and the mass, and a spring capable of compression or stretch is used to return the mass

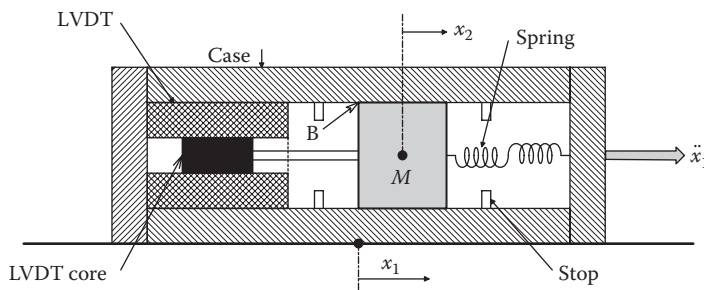


FIGURE 7.25

Passive, linear accelerometer. Mass position ($x_1 - x_2$) is sensed by the LVDT and is proportional to the case acceleration in the SS. See text for analysis.

to its equilibrium position in the center of the case when there is no acceleration. The relative position between the moving mass and the case is measured by an LVDT. The LVDT's body is attached to the case, and the LVDT's moving core is attached to the movable mass. The case itself is firmly attached to the object whose linear acceleration is to be measured. Typical application of an accelerometer includes the measurement of the acceleration of a vibration table used to test the mechanical stability of components or to measure the up/down acceleration of vehicle body while driving over a bumpy road or to sense vehicle collisions and trigger airbag deployment. For a mass and spring accelerometer moving in the x -direction, we may write, using Newton's law,

$$0 = M\ddot{x}_2 + B(\dot{x}_2 - \dot{x}_1) + K_s(x_2 - x_1). \quad (7.95)$$

Here, x_1 is the position of the case (and the structure whose acceleration we wish to measure), and x_2 is the position of the mass inside the case with respect to an external inertial reference point ($x = 0$). This ODE may be Laplace transformed and put in the form of a transfer function:

$$\frac{X_2}{X_1}(s) = \frac{sB/K_s + 1}{s^2M/K_s + sB/K_s + 1}. \quad (7.96)$$

Now, we assume that the output of the LVDT is demodulated and filtered to give a voltage proportional to the relative position of the mass in the accelerometer case:

$$V_o = K_o(X_1 - X_2). \quad (7.97)$$

Substitution of Equation 7.111 into Equation 7.112 gives us the accelerometer's transfer function:

$$\frac{V_o}{\ddot{X}_1}(s) = \frac{K_o(M/K_s)}{s^2M/K_s + sB/K_s + 1}. \quad (7.98)$$

Thus, for a constant acceleration, $V_o = \ddot{X}_1 K_o(M/K_s)$. Linearity of this simple mechanical accelerometer depends on spring linearity over the working range of displacement, as well as LVDT linearity. The natural frequency of this sensor is $f_n = (1/2\pi)\sqrt{K_s/M}$ Hz, and the sensor's damping constant is given by $\zeta = B/(2\sqrt{MK_s})$. If the magnitude of \ddot{x}_1 becomes too large, the mass will hit the stops in the case and the output voltage will saturate. On the other hand, if \ddot{x}_1 is very small, small amounts of static (coulomb) friction between the mass and case will prevent the mass from moving relative to the case, and there will be an output dead zone. The friction problem can be solved by lubrication, by precision machining of the mass/case assembly, and by dithering the mass with an AC solenoidal magnetic field. The accelerometer frequency response and dynamic range can be extended by the use of the servoaccelerometer design, described in the following.

7.3.1.2 Servoaccelerometer Design

A servoaccelerometer is shown in cross section in Figure 7.26. We see that the same components are present as for the simple design described earlier, except the spring has been

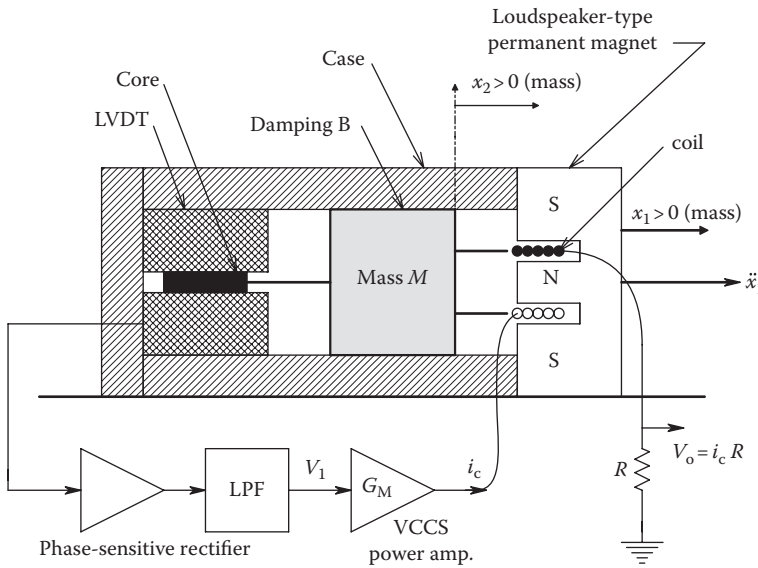


FIGURE 7.26 Section through a servolinear accelerometer. The servodesign allows the accelerometer’s natural frequency to be raised while preserving sensitivity. See text for analysis.

replaced with a force coil, not unlike the voice coil of a standard loudspeaker. The force coil moves in the air gap of a strong, coaxial permanent magnet, also like that used for loudspeakers. The Newtonian force balance equation can be written as

$$F_C = K_F I_C = M \ddot{x}_2 + B(\dot{x}_2 + \dot{x}_1). \tag{7.99}$$

Here, we assume the force produced by the coil and magnet is proportional to the current through the coil. In addition, the LVDT system output is $V_1 = K_D(x_1 - x_2)$, and the current is produced by a VCCS with transconductance, G_M . We assume further that the servoaccelerometer output voltage is proportional to I_C . The combination of these relations after Laplace transforming yields the quadratic transfer function:

$$\frac{V_o}{\ddot{X}_1}(s) = \frac{K_o(M/K_F)}{s^2 M / (K_F G_M K_D) + s B / (K_F G_M K_D) + 1}. \tag{7.100}$$

Now, the system’s natural frequency is $f'_n = (1/2\pi)\sqrt{K_F G_M K_D / M}$ Hz, and its damping factor is $\zeta' = B / (2\sqrt{M K_F G_M K_D})$. The feedback design of this active sensor allows the natural frequency to be increased using $G_M K_D$, without lowering sensitivity, as happens with the passive accelerometer. Of course, the complexity of the servoaccelerometer design makes it more expensive than the simple moving-mass design.

7.3.1.3 Piezoelectric Accelerometers

One way to eliminate the problem of coulomb friction that accompanies any moving-mass design is to not have the mass move relative to the case. Figure 7.27 illustrates a single-mass, compression-type, linear accelerometer in which the mass is in contact with the

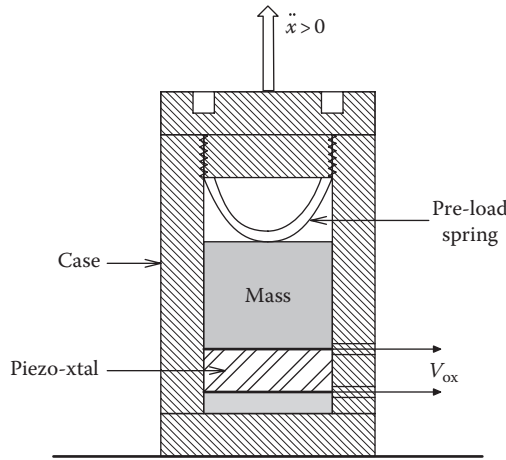


FIGURE 7.27
Simple piezoelectric crystal accelerometer design shown in cross section.

surface of a piezoelectric crystal such as quartz, barium titanate, and LZT. The mass is often preloaded with respect to the case with a very stiff spring. Acceleration of the case in the $\pm x$ -direction causes a corresponding acceleration of the mass, M . Hence, an additional force, $F_x = \pm M\ddot{x}$, is exerted on the piezoelectric material, which is compression sensitive. Thus, an equivalent current, $i_x = \pm \dot{F}d$, is generated in parallel with the below-resonance, R_x - C_x parallel equivalent circuit for the piezomaterial. It is easy to show that

$$V_o = \frac{\ddot{X}(s)MdR_x}{sR_xC_x + 1} \tag{7.101}$$

Thus, for the acceleration, we may write

$$\frac{V_o}{\ddot{X}}(s) = \frac{\ddagger MdR_x}{sR_xC_x \ddagger 1} \tag{7.102}$$

Equation 7.117 is a simple high-pass transfer function. It tells us that the piezoelectric accelerometer will not measure low-frequency or constant accelerations (such as due to gravity), but will work well for time-variable accelerations containing frequencies above $f_b = 1/(2\pi R_x C_x)$ Hz. Thus, piezoaccelerometers are widely used for vibration testing, where the frequencies of the applied displacement are above f_b . If a charge amplifier is used, the break frequency is now set by the feedback R_F and C_F of the electrometer op amp. The transfer function becomes

$$\frac{V_o}{\ddot{X}}(s) = \frac{-sMdR_F}{sR_FC_F \ddagger 1} \tag{7.103}$$

The low break frequency, $1/(2\pi R_FC_F)$, can generally be made lower than that for the piezo-element alone, given earlier. The HF gain is $-Md/C_F$ volts/meter/second².

Simple compression-type piezoaccelerometers suffer from two drawbacks: *one*, a high temperature sensitivity (temperature equivalent acceleration output caused by rapidly

changing transducer temperatures) and, *two*, sensitivity to base bending (under the piezo-sensor) and off-axis accelerations (linear and angular) that cause the mass to exert force asymmetrically on the piezomaterial, producing erroneous outputs. The DeltaTron® piezoaccelerometer design developed by Brüel and Kjaer Instruments, Inc., minimizes these problems. B&K DeltaTron® accelerometers use LZT piezomaterial. The accelerometers have a built-in MOSFET charge amplifier to buffer the crystal output. Typical DeltaTron® sensor frequency range is 0.3 Hz to 18 kHz, and crystal self-resonant frequency is ca. 48 kHz. The maximum acceleration range is $\pm 7500 \text{ m/s}^2$ at 100°C . Axial sensitivity is 1 mV/ms^2 at 159 Hz and 100 m/s^2 and 25°C for Types 4394, 4395, 4397, and 4398. Equivalent input residual noise (from 1 Hz to 22 kHz) is $< 0.025 \text{ m/s}^2$.

Because there is no relative motion between the mass and the case, there is no coulomb friction to cause a dead zone, and the dynamic range of piezoelectric accelerometers can be an enormous 160 dB. Because of noise and output saturation, amplifiers do not have this dynamic range, and so signal-conditioning gains would have to be switched to accommodate an \ddot{X} that varies 160 dB. The mechanical resonance frequency of piezoaccelerometers is orders of magnitude higher than that of moving mass and spring accelerometers. This difference exists because the stiffness of the piezomaterial is orders of magnitude larger than that of the spring.

Because of their HF bandwidth, piezoaccelerometers are well suited for vibration and impact testing applications in the kilohertz range.

7.3.2 Linear Velocity Measurement

Measurement of linear velocity is subdivided into two classes: *velocity measurements on solid or mechanical objects* and *velocity measurements of fluids*, both from a stationary reference position. There appears to be no way to inertially measure vehicle velocity internally except by integrating accelerometer output. In some cases, the velocity observer is inside the moving object (e.g., an automobile, boat, or airplane); in other cases, the observer is without, but in all cases, there must be an external reference point.

An obvious means of estimating the linear velocity of a mechanical (solid) object is to integrate the output of a linear accelerometer attached to it. This method works well when the SNR at the accelerometer output is high and the integrator does not drift. If the accelerometer output is integrated digitally, at least a 16-bit ADC is needed, and the integration routine and sampling interval must be carefully specified.

Generally, other means of velocity measurement depend on what velocity is being measured. We have already discussed the use of LDV to measure the velocity of fluids containing small, reflecting particles in Section 6.9.3. LDV can be used with any moving, reflecting surface to measure velocity. LDV has recently been adapted to work with an IR LAD to measure vehicle velocity, competing with microwave Doppler radar speed detection. Not only is the IR laser beam highly focused, but it is invisible; the IR radar speed *gun* has to be aimed with a telescopic sight. Within the limitations of propagation range, ultrasound can also be used to measure vehicle velocity in a Doppler system. A major problem with all Doppler systems for measuring velocity of moving, reflecting objects is the necessity to know θ , the angle between the velocity vector and the line connecting the transmitting and receiving transducers with the moving object. This problem has been largely overcome by the crossbeam techniques developed by George and Lumley (1973) for LDV and by Fox (1978, 1987) for ultrasound velocimetry.

7.3.2.1 Means of Measurement of Fluid Velocity

We have already described systems to measure fluid velocity by LDV and by the Faraday effect. We now consider several other means:

Hot wire and hot film anemometers: The hot wire and hot film anemometers make use of the fact that the resistance of a metallic wire or film changes with its temperature. The wire's temperature, on the other hand, depends on the electric power dissipated in the wire (Joule's law) and the rate that the heat from the power is removed by the surrounding medium. The rate that the heat is dissipated is a nonlinear, increasing function of the velocity of the fluid surrounding the wire or film. Several representative hot wire and hot film probes made by TSI, Inc. (2013), Shoreview, MN, are illustrated in Figure 7.28.

Hot wire sensors are used in fluid mechanics research for gas velocity and turbulence measurements. They are generally very fine wires (e.g., 13 μm diameter and 1.3 mm length) stretched between two, gold-plated needle supports. Depending on the gas temperature and velocity range, materials used for the wires can be platinum-coated tungsten that has a tempco of 0.0042 $\Omega/^\circ\text{C}$, pure platinum (tempco of 0.003 $\Omega/^\circ\text{C}$) useful up to 300 $^\circ\text{C}$, and platinum-iridium alloy (tempco of 0.0009 $\Omega/^\circ\text{C}$) useful up to 750 $^\circ\text{C}$. Less fragile are metal film sensors. A high-purity platinum film is bonded to a fused quartz substrate and then covered with alumina for gas velocity measurements or with quartz for measuring the velocity of conducting liquids. Metal film sensors can be configured as wires (by bonding the platinum to quartz filaments as small as 0.001" in diameter) or as wedges (the film is bonded to the end of a quartz *screwdriver*) or as a flat strip on a flat quartz disk or as a thin ring around the circumference of the tip of a quartz cone. Film sensors can be used to about 300 $^\circ\text{C}$ on high-temperature probes.

All hot wire and hot film probes have minimum and maximum fluid velocities over which they work effectively. The minimum velocity detectable is set by the amount of free convection from the heated sensor. Velocities down to 0.15 m/s (0.5 ft/s) can be measured in air and down to 0.003 m/s (0.01 ft/s) in water. The maximum velocities that can be reliably measured are primarily limited by mechanical considerations, that is, the strength of the fine wires. Special hot film sensors have been used in gasses with supersonic velocities and in liquids with velocities up to 15 m/s.

Although most fluid velocity measurements are made with single probes or pairs of probes, true, 3D vector flow now can be measured with a special three-wire probe developed by TSI, Inc. This probe is used with a special computer that solves *Jorgenson's equations* for the three-wire probe to give the exact 3-D velocity vector.

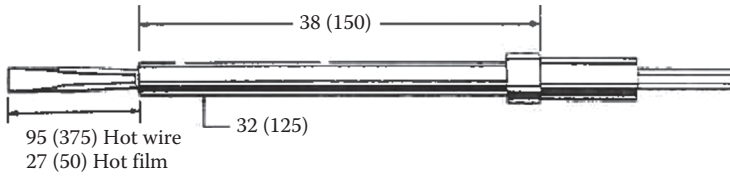
A heuristic, static analysis of a single, hot wire probe is based on several simplifying assumptions. First, we assume that the probe's resistance, R_w , is linearly related to its temperature. Mathematically, this may be expressed as

$$R_w = R_o(1 + \alpha \Delta T) \quad (7.104)$$

where

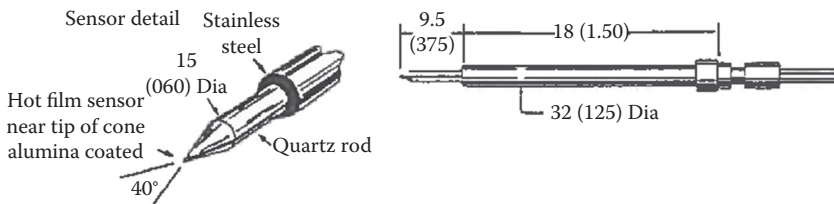
R_o is the probe's resistance at some reference temperature, T_o
 α is the first-order, positive tempco of the metal wire

High temperature straight probe—Model 1220



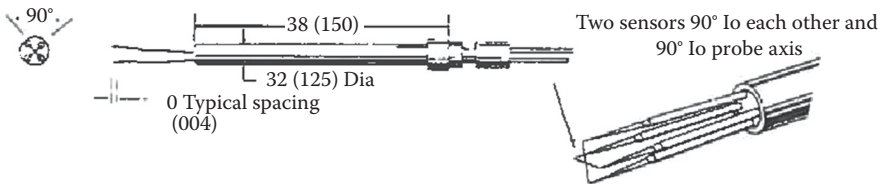
Designed for ambient temperature to 300°C. Suitable for one-component velocity measurements.

Straight conical probe—Model 1230



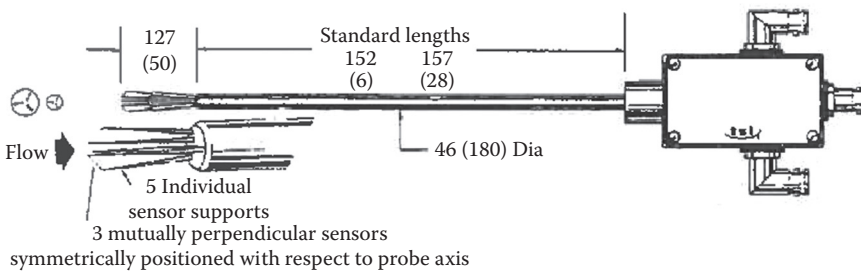
Rugged sensor design resists contamination and breakage. Conical probe shape is used widely in liquid flow measurements.

Standard "X" probe—Cross flow—Model 1240



Features two sensors 90° to each other and 90° to the probe axis. For measurements of two turbulence components. Correlations and flow vectors in 2D flows. Ideal for measurements where space is limited.

3D probe—Model 1294



Features three mutually perpendicular sensors symmetrically positioned with respect to probe axis. Recommended for high-level turbulence (intensities 10%). Model 1146 (1147 for-60 sensor) shield include (see pg. 28)

FIGURE 7.28 Four, hot wire and hot film fluid velocity probes. (Courtesy of TSI Inc., Shoreview, MN.)

Second, we assume that the temperature rise of the probe, ΔT , is given by the *heat sink equation*

$$\Delta T = P \Omega \tag{7.105}$$

where

P is the average, SS electrical power dissipated in the probe

Ω is the *thermal resistance* of the probe, in degrees Celsius per watt

Ω is determined by the geometry of the hot probe; the heat-conducting properties of the surrounding medium, whether the medium is moving or not; and the velocity of the medium relative to the probe's geometry. For the case where the fluid velocity is perpendicular to the axis of the wire, the thermal resistance is given approximately by

$$\Omega = \Omega_o - \beta \sqrt{|\bar{v}|} \tag{7.106}$$

where

Ω_o is the thermal resistance in still fluid

β is a geometry- and fluid-dependent constant

$|\bar{v}|$ is the average magnitude of the fluid velocity, assumed perpendicular to the axis of the wire

In still fluid, ΔT_o can be found from the heat sink equation

$$\Delta T_o = P_o \Omega_o = I_o^2 R_o (1 + \alpha \Delta T_o) \Omega_o. \tag{7.107}$$

Solving for ΔT_o , we obtain the following:

$$\Delta T_o = I_o^2 R_o \Omega_o / (1 - \alpha I_o^2 \Omega_o) \tag{7.108}$$

where I_o is the RMS current through the probe resistor.

This relation for ΔT_o can be substituted into the equation for resistance as a function of temperature, and we find

$$R_{w_o} = R_o / (1 - \alpha I_o^2 \Omega_o). \tag{7.109}$$

R_{w_o} is the new, elevated, resistance above the ambient resistance of the hot wire probe in still fluid due to electrical power dissipation. If the fluid is moving, we may write

$$R_w(I, v) = R_o / [1 - \alpha I^2 (\Omega_o \mp \beta \sqrt{|v|})]. \tag{7.110}$$

Figure 7.29 illustrates a simple, feedback bridge circuit in which the hot wire probe is maintained at nearly constant resistance, hence operating temperature. In the circuit, we assume that $R_1 \gg R_w(I, v)$, so that the current through the probe is given by $I = V_B / R_1$. V_o is a DC source used to set R_{w_o} . With the switch SW1 open, $V_B = V_S$ and $V_1 = V_{1o}$. Thus,

$$V_{1o} = V_S R_{w_o} / R_1 = V_S R_o / [R_1 (1 - \alpha V_o^2 \Omega_o / R_1^2)]. \tag{7.111}$$

V_2 is adjusted to have this same value, that is, $V_{2o} = V_{1o}$, so $V_3 = 0$. Exact analysis of the closed-loop system is tedious and involves the algebraic solution of a cubic equation. It is

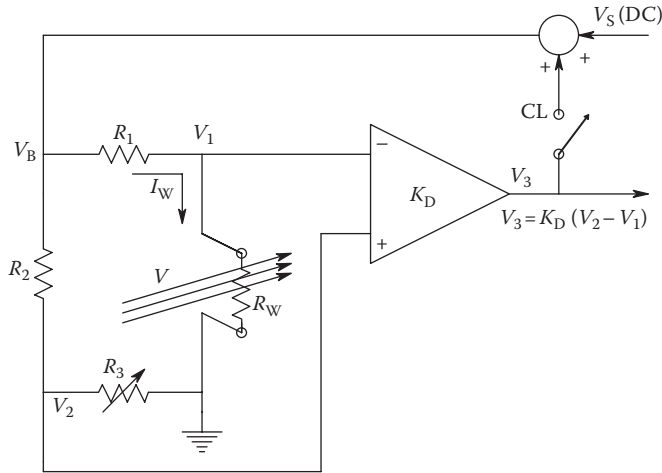


FIGURE 7.29
Schematic of a simple feedback bridge circuit useful with hot wire probes.

useful to assume that the high gain of the amplifier, K_D , forces the closed-loop $R_W(l, v)$ to equal the R_{W_0} in still fluid. Thus,

$$\frac{R_o}{1 - \alpha V_S^2 \Omega_o / R_1^2} = \frac{R_o}{1 - \alpha V_B^2 (\Omega_o - \beta \sqrt{|v|}) / R_1^2}. \tag{7.112}$$

From Equation 7.128, we find that

$$V_B^2 = V_S^2 / (1 - \beta \sqrt{|v|} / \Omega_o). \tag{7.113}$$

If we assume that $1 - \beta \sqrt{|v|} / \Omega_o \ll 1$, then it is easy to show that

$$V_{B+} \cong V_S (1 + \beta \sqrt{|v|} / 2\Omega_o) \tag{7.114}$$

and finally,

$$V_3 = V_S \beta \sqrt{|v|} / 2\Omega_o. \tag{7.115}$$

Thus, the electrical output of this system for conditioning the HWA response is proportional to the square root of the fluid velocity. Further linearization can be obtained by passing V_3 through a square-law circuit. When fluid flows at an angle to the hot wire probe axis, a reduced response is obtained, the exact form of which depends on the probe geometry, including the supports.

7.3.2.2 Other Means of Measuring Fluid Velocity and Flow

Fluid flow through a uniform pipe may be subdivided into two major categories: laminar flow, where the velocity vectors have a parabolic distribution, shown in Figure 7.30A, and turbulent flow, shown in B, which is no longer characterized by a parabolic velocity profile and which no longer contains not only parallel streamlines but also eddies and

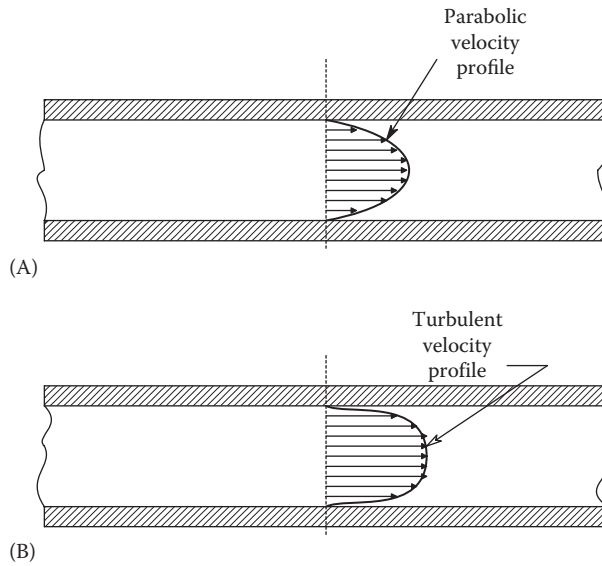


FIGURE 7.30

(A) Axial section through a pipe carrying fluid under laminar flow conditions. Note the parabolic distribution of axial fluid flow with near zero velocity at the pipe walls. (B) This pipe is carrying fluid under turbulent flow conditions. Note that the axial velocity distribution is no longer parabolic with the pipe’s radius.

whorls. The transition between laminar and turbulent flow occurs over a narrow range of (average) velocities. It has been found that the transition between laminar and turbulent flow occurs when a diagnostic variable called the *Reynolds number* exceeds a threshold value. The Reynolds number is given by

$$N_R \doteq \bar{V}\rho D/\eta \tag{7.116}$$

where

- \bar{V} is the average velocity (LT^{-1}) in the pipe
- ρ is the mass density (ML^{-3}) of the fluid
- η is the absolute viscosity of the fluid ($ML^{-1}T^{-1}$)
- D is the internal diameter of the pipe (L)

Combining the dimensions provided earlier, N_R is seen to be dimensionless. If English units are used (lb, ft, s), a velocity above that which will produce $2000 < N_R < 2300$ will result in turbulent flow. Needless to say, the mathematics describing turbulent flow is beyond the scope of this text, so we will concentrate on flow and velocity measurements under laminar conditions ($N_R < 2000$). Laminar flow velocity in a pipe with circular cross section may be shown to be described by

$$V(r) = \frac{(p_1 - p_2)}{4\eta L} (R^2 - r^2) \quad 0 \leq r \leq R. \tag{7.117}$$

Here

- $(p_1 - p_2)$ is the pressure difference across a length L of pipe
- R is the pipe’s radius
- r is the radial distance from the center of the pipe

The total volume flow across the cross-sectional area of the pipe may be found by integrating the parabolic velocity profile over the pipe's cross-sectional area:

$$Q = \int v(r) dA = \int_0^R V_m (1 - r^2/R^2) 2\pi r dr = V_m A / 2 \text{ m}^3/\text{s} \quad (7.118)$$

where

V_m is the peak velocity, equal to $(p_1 - p_2)R^2/(4\eta L)$

A is the pipe's cross-sectional area = πR^2

Thus, under laminar flow conditions, if we can measure the peak velocity (at the center of the pipe), we can calculate the net volume flow rate, Q .

One common means of measuring the volume flow of incompressible fluids (generally, liquids) is by *obstruction meters* (Beckwith and Buck 1961). These include venturi tubes, flow nozzles, and orifice nozzles. Figure 7.31A–C, reprinted with permission from Beckwith and Buck (1961), illustrates three of these obstruction meters. What is actually measured is the pressure differential across the nozzle. Beckwith and Buck show that the volume flow through a venturi tube can be expressed as

$$Q_v = CMA_2 \sqrt{2g(P_1 - P_2)/\alpha} \text{ m}^3/\text{s} \quad (7.119)$$

and the volume flow through nozzles and orifices is given by

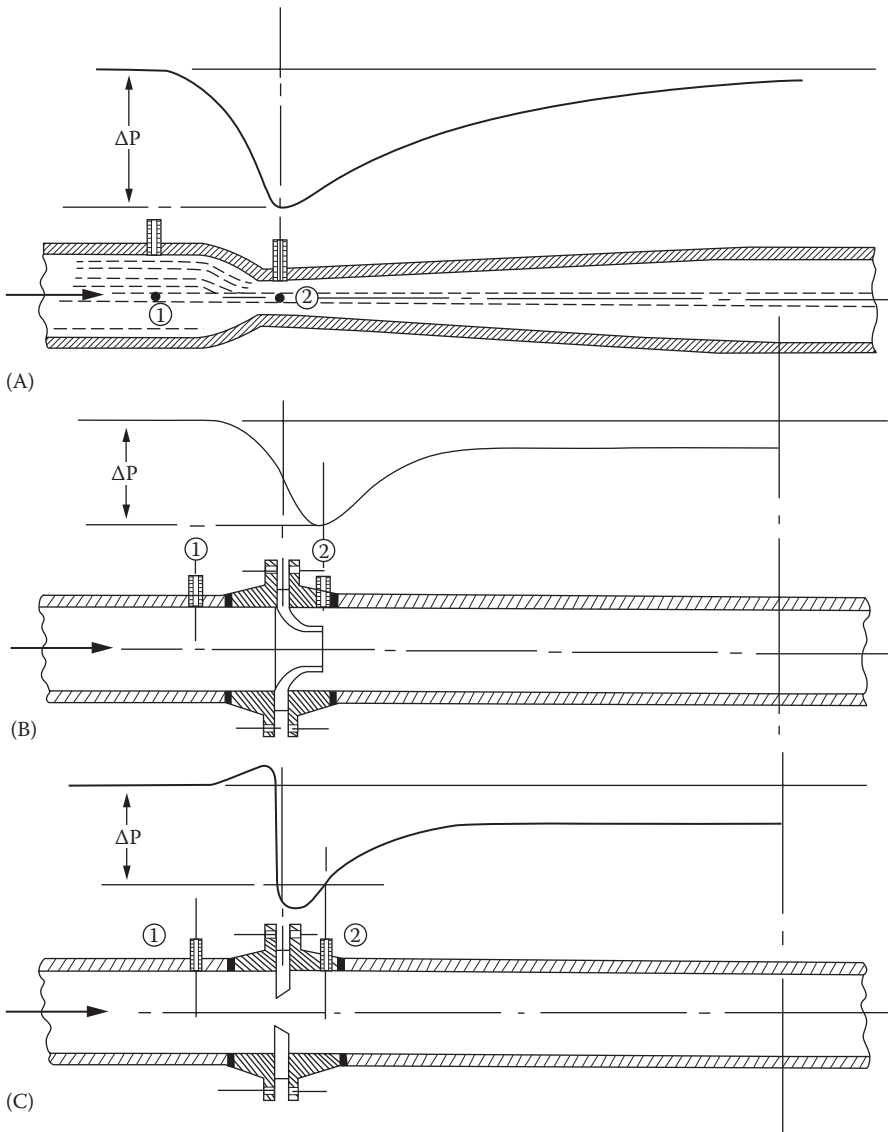
$$Q_n = KA_2 \sqrt{2g(P_1 - P_2)/\alpha} \text{ m}^3/\text{s} \quad (7.120)$$

where C and K are constants that depend on, among other factors, the Reynolds number and the geometry of the obstruction. M is given by

$$M = 1/\sqrt{(1 - (A_2/A_1))^2} \quad (7.121)$$

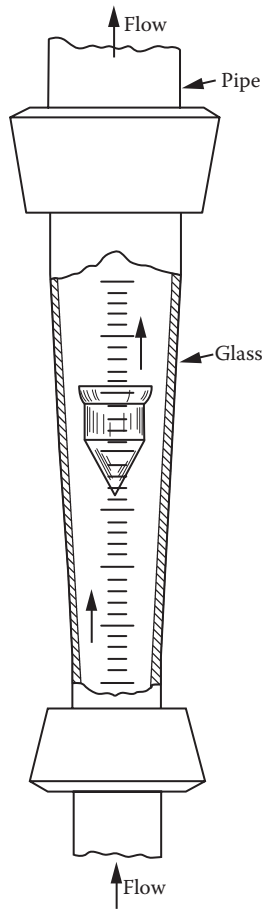
$K = CM$, g is the acceleration of gravity, α is the specific weight of the liquid, A_1 is the upstream pipe cross-sectional area, and A_2 is the minimum area of the obstruction. Since the pressure difference, $P_1 - P_2$, is measured, it is clear that the actual volume flow is proportional to its square root. Because of the turbulence losses associated with obstruction flowmeters, there is always an SS downstream pressure loss, in some cases up to 30% for an orifice flowmeter. There is significantly less SS pressure loss with a venturi system because of its streamlining and reduced turbulence.

Obstruction meters can also be used with compressible fluids (gasses). The situation here is far more complex, because of the adiabatic compression and expansion of the gas as it passes through the obstruction. In this case, the volume flow is no longer a simple square-root function of the pressure difference; power law terms involving the ratio of pressures across the obstruction are involved (cf. Beckwith and Buck 1961, Sects. 12–15).

**FIGURE 7.31**

(A) Section through a venturi nozzle. The volume flow through the nozzle is proportional to the square root of the pressure difference ΔP for incompressible liquids. (B) and (C) Orifice flowmeters. The pressure drop across the orifice is proportional to flow. (From Beckwith, T.G. and Buck, N.L., *Mechanical Measurements*, Addison-Wesley, Reading, MA. With permission.)

The fixed geometry obstruction meters suffer from the common problem that the pressure drop is proportional to the square of the flow rate. Thus, a very wide range differential pressure meter is required. If a large range of flow is to be measured, then accuracy may be low at the low end of the range. The *rotameter*, illustrated in Figure 7.32, is a flow-measuring device of extreme simplicity. It consists of a vertical, glass or clear plastic tube whose inner diameter increases toward the top. Inside the tube is a *float*, which is pushed upward by the upward-flowing fluid. The equilibrium position of the float in the tapered

**FIGURE 7.32**

Section through a typical rotameter flowmeter. (From Beckwith, T.G. and Buck, N.L., *Mechanical Measurements*, Addison-Wesley, Reading, MA. With permission.)

tube is determined by the balance between the downward gravity force and the upward pressure, viscous drag force, and buoyant force. Beckwith and Buck (1961) gave a complex relationship between the volume flow of a noncompressible fluid and the height of the float in the tapered tube:

$$Q = (\pi/4)[(D + by)^2 - d^2] \frac{\sqrt{2gv_f(\rho_f \mp \rho_w)}}{\sqrt{A_f r_w}} C \quad (7.122)$$

where

D is the effective diameter of the tube at the height of the float, y

v_f is the volume of the float

b is the change in tube diameter (taper) with y

d is the maximum diameter of the float

A_f is the effective cross-sectional area of the float

ρ_f is the float's density

ρ_w is the liquid's density

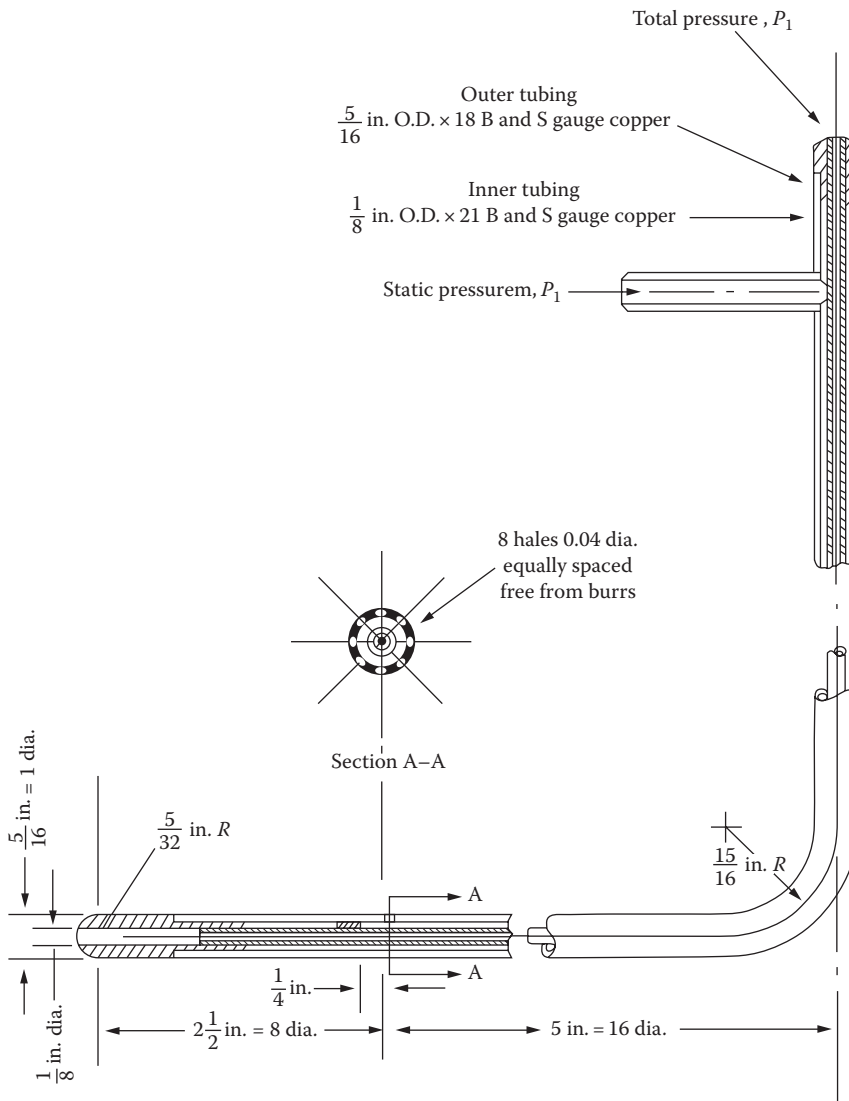
C is the discharge coefficient

Rotameters are used for gasses as well as incompressible liquids. A major disadvantage in their use is that they cannot be used with opaque fluids or fluids with large amounts of particles in suspension. Rotameters must also be read out manually. Advantages of rotameters are their nearly linear flow scale and their pressure loss is constant over the entire flow range. Another type of flow or velocity meter makes use of a free-spinning turbine wheel in the pipe carrying the fluid. The flow or velocity is a function of the speed of the turbine, which can be measured by a tachometer generator that must be coaxial with the turbine in the pipe. Alternately, if small magnets are placed in the tips of the turbine's vanes, a Hall effect switch sensor can be used outside the pipe to measure the rotating vanes passing under it. The number of pulses/second is proportional to the turbine's angular velocity, which is in turn an increasing function of volume flow. Because no analytical expression exists for turbine velocity as a function of flow or velocity, each different turbine meter design must be calibrated for a given application.

Pressure probes, which include *Pitot tubes*, are also useful in measuring fluid velocity in ducts or in free space. The Pitot static probe is used to estimate the relative airspeed of aircraft through the air. Here again, we see that the velocity is approximately proportional to the square root of the pressure difference for air. A differential PT may be used to effect readout. A typical Pitot tube geometry is illustrated in Figure 7.33. Pitot tubes lose calibration when the fluid flow is no longer coaxial to the probe axis. A yaw angle of as small as 8° can produce significant errors in Pitot tube calibration. The *Kiel ram tube*, shown in Figure 7.34, operates similarly to the Pitot tube except the reference pressure is not taken from the probe. The Kiel probe is evidently far more insensitive to off-axis air velocity, maintaining its calibration up to yaw angles of 40° (Beckwith and Buck 1961).

In medicine and physiology, it is important to measure blood flow and velocity in various blood vessels and in the heart. One means of estimating blood flow is by the Faraday effect, already discussed in Section 6.3.5.1. Faraday flow probes have been designed that clamp on to the blood vessel in question. They do not actually contact the blood, but do require surgery to expose the vessel in question. Their use is obviously invasive and not without risk. Webster (1978) described the designs of several Faraday blood flowmeters. It should be stressed that the Faraday system is responsive to the average velocity of the conducting fluid (blood) and that flow is a derived quantity based on a knowledge of the internal diameter of the blood vessel.

The use of *Doppler ultrasound* transmitted through the skin to the blood vessel of concern offers a risk-free, noninvasive means to estimate blood velocity and flow (see Webster 1978, Sect. 8.4). All ultrasonic blood velocity/flow systems make use of the fact that blood contains erythrocytes or red cells suspended in it. Erythrocytes are typically found in densities of about 5×10^6 per mm^3 and are biconcave disks normally about $8 \mu\text{m}$ in diameter, $2 \mu\text{m}$ thick at their edges, and $1 \mu\text{m}$ thick at their centers. The velocities of individual erythrocytes follow the laminar flow streamlines of velocity in the larger blood vessels (arteries and veins). Ultrasound incident on the moving blood in a vessel is therefore reflected off myriads of tiny scatterers whose velocities range from zero to maximum and whose cross-sectional areas vary randomly in time and space (they tumble as they move). Hence, the return signal is composed of two major components: The largest component is at the carrier frequency and is reflected from all nonmoving tissues in the beam path. The information-carrying component is much smaller in energy and is made up from the superposition of many amplitudes and frequencies reflected from the ensonified, moving red cells. The exact nature of the Doppler return signal depends on

**FIGURE 7.33**

Pitot airspeed tube. (From Beckwith, T.G. and Buck, N.L., *Mechanical Measurements*, Addison-Wesley, Reading, MA. With permission.)

how tightly the ultrasound beam can be focused in the interior of the vessel to discriminate individual regions of velocity.

The most basic form of Doppler ultrasound system is the continuous-wave (CW) configuration, where the return signal is heterodyned with a sine wave at the transmitted frequency, then low-pass filtered and amplified. The frequency of this output signal is proportional to the average velocity of the ensonified blood. The output signal is listened to by the operator and may be used to detect fetal heartbeats or pulsating aneurysms in artery walls, in addition to blood velocity in vessels such as the common carotid arteries. In the latter case, obvious asymmetries in carotid blood velocity may indicate an asymmetrical obstruction, such as by atherosclerotic plaque in one carotid sinus. CW Doppler

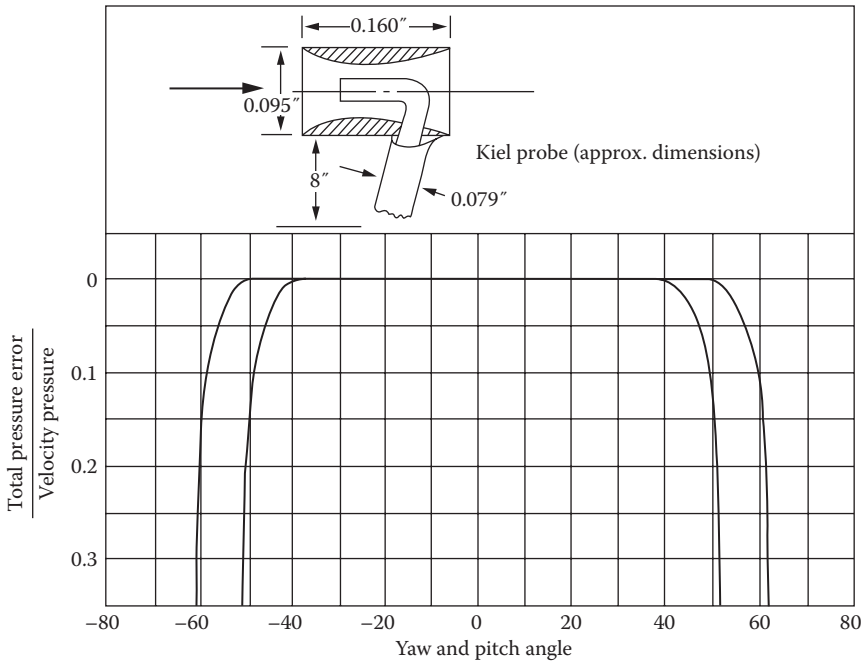


FIGURE 7.34

Kiel-type air velocity probe. Note its relative insensitivity to off-axis air velocity. (From Beckwith, T.G. and Buck, N.L., *Mechanical Measurements*, Addison-Wesley, Reading, MA. With permission.)

ultrasound systems are generally more qualitative in their medical applications than are the more sophisticated pulsed Doppler systems. Pulsed Doppler systems can be range-gated to enable a blood velocity profile to be constructed for large vessels such as the descending aorta.

To be used quantitatively, all conventional Doppler blood velocity systems require a precise knowledge of θ , the angle between the blood velocity vector and a line connecting the transmitting/receiving transducers with the small volume of blood whose velocity is being sensed. Recall that the frequency of the Doppler-shifted return signal is given by the relation

$$f_r = f_t \left(1 + \frac{2v \cos(\theta)}{c} \right) \tag{7.123}$$

where

- c is the average velocity of sound in blood
- v is the velocity of the small blood volume ensounded
- f_t is the transmitted (carrier) frequency

The Doppler shift, here taken as positive because the velocity of the reflector has a component toward the transducers, is detected electronically and is given by

$$f_d = \left[\frac{f_t 2v \cos(\theta)}{c} \right]. \tag{7.124}$$

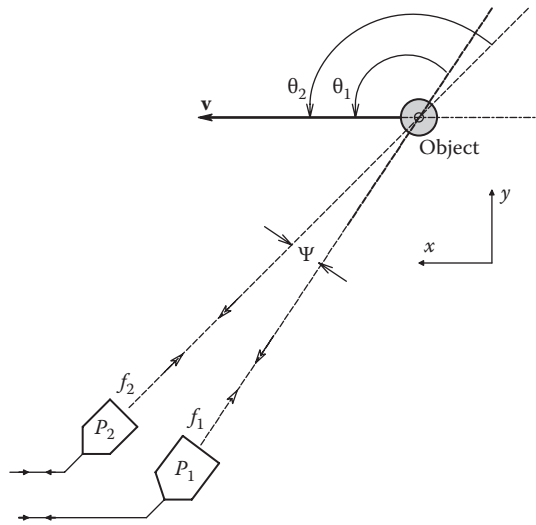


FIGURE 7.35 Geometry relevant to Fox’s angle-independent Doppler system. TR1 ultrasound transducer 1 transmits at frequency f_1 . Transducer TR2 transmits at f_2 .

Obviously, v is proportional to $f_r/\cos(\theta)$, and an error in θ will give an error in v . In 1978, Fox presented a closed-form, angle-independent solution to the 2-D Doppler situation that utilizes the outputs of two, independent transmit–receive probes (see Figure 7.35). Fox’s solution yields the velocity magnitude, $|\mathbf{v}| = \sqrt{v_x^2 + v_y^2}$, and the angle θ_1 between the velocity vector and a line from number 1 probe. Note that the probes lie in the XY plane with the velocity vector. The probes are separated by an angle ψ , and their beams converge on the moving, reflecting object at P . The Doppler frequency returned to each probe is given by

$$f_{d1} = 2f_1 |\mathbf{v}| \cos(\theta_1)/c \tag{7.125a}$$

$$f_{d2} = 2f_2 |\mathbf{v}| \cos(\theta_2)/c. \tag{7.125b}$$

Now, the angle θ_2 can be written in terms of θ_1 as

$$\theta_2 = \theta_1 + \psi. \tag{7.126}$$

After some arcane trigonometry and algebra, it can be shown that

$$|\mathbf{v}| = \sqrt{v_x^2 + v_y^2} = \left(\frac{cf_{d1}}{2f_1} \right) \sqrt{\frac{1 \mp 2R \cos(\psi) \mp R^2}{\sin^2(\psi)}} \tag{7.127}$$

and

$$\theta_1 = \tan^{-1} \left[\frac{\cos(\psi) - R}{\sin(\psi)} \right] \tag{7.128}$$

where R is the Doppler shift ratio corrected for carrier frequency, given by

$$R \equiv \frac{(f_{d2}/f_2)}{(f_{d1}/f_1)}. \quad (7.129)$$

In order to calculate $|\mathbf{v}|$ and θ_1 , one must measure f_{d1} and f_{d2} , knowing f_1 , f_2 , and ψ . The accuracy of the method is limited by the accuracy that one can determine f_{d1} and f_{d2} . Because of the parabolic (laminar) flow profile in blood vessels, the finite size of the ensonified volume (typically 3 mm³ for a 2.25 MHz carrier, Fox [1978]), and the random scattering nature of moving red blood cells, there typically is a bell-shaped distribution of f_{ds} , rather than a single, sharp peak. The mode of the distribution is generally taken as the desired f_d . The Fox two-probe method of determining the velocity vector in two dimensions is best implemented with a computer system that algorithmically processes the Fourier-transformed Doppler return signals to determine their modes to estimate f_{d1} and f_{d2} and then calculates $|\mathbf{v}|$ and θ_1 using the relations presented earlier and the known parameters, Ψ , f_1 , and f_2 .

Fox and Gardiner (1988) extended the 2-D, closed-form solution for \mathbf{v} to three dimensions. Their equations are too long to include here, but they have the same general form as the far simpler 2-D case described earlier. Their experimental results showed that the calculated $|\mathbf{v}|$ remained within 5.6% of the theoretical value for Doppler angles up to 50°. Also, their angle estimate agreed with the theoretical values with a correlation coefficient, $r = 0.99937$.

The 2- and 3-D Doppler flow velocimetry technique developed by Fox and colleagues is of course not restricted to the ultrasonic measurement of blood velocity. The Fox technique can be extended to the other Doppler modalities (lasers and microwaves), when a 2- or 3-D estimate of object velocity is required.

The final system that we will discuss, which is used for the measurement of gas velocity and flow and which is used in modern pulmonary diagnostic instruments, is the *pneumotachometer* or pneumotach. Basically, a pneumotach is a pneumatic analog of a resistor. A resistor obeys Ohm's law, that is, $I = V/R$. In the case of laminar gas flow through the pneumotach,

$$Q = \frac{(P_1 - P_2)}{R_p} \quad (7.130)$$

where

Q is the volume flow of gas, $(P_1 - P_2)$ is the pressure drop across the pneumotach
 R_p is the equivalent pneumatic resistance of the pneumotach

The pneumotach itself can have one of several forms. The body or duct of a pneumotach is typically 1.5" in diameter, and it uses either a fine-mesh screen or hundreds of parallel capillary tubes, about 1" to 2" long to obtain the desired acoustic resistance. Unfortunately, a capillary tube does not exhibit a pure, frequency-independent

resistance, but appears reactive at high frequencies. The acoustic impedance of a single capillary tube is given by (Olson 1943)

$$\mathbf{Z}_A(j\omega) = \frac{8L\eta}{\pi R^4} + j \frac{4L\omega\rho}{3\pi R^2} \text{ cgs acoustic ohms} \quad (7.131)$$

where

η is the viscosity of air at 20°C (1.86×10^{-4} poise)

R is the radius of the tube in cm

L is its length in cm

ρ is the density of air (1.205×10^{-3} g/cm³)

ω is the radian frequency of pressure variation across the pneumotach elements

If N identical capillary tubes are packed in parallel, the resultant acoustic impedance can be shown to be

$$\mathbf{Z}_{A(\text{NET})} = \frac{8L\eta}{N\pi R^4} + j \frac{4L\omega\rho}{3N\pi R^2} \text{ cgs acoustic ohms.} \quad (7.132)$$

The acoustic impedance of a capillary tube pneumotach appears resistive for frequencies from DC up to about 1/10th of the frequency where $R_p = X_p$, or

$$f_{\max} \approx \frac{0.3\eta}{(\pi\rho R^2)} \text{ Hz} \quad (7.133)$$

f_{\max} is about 50 Hz for $R = 0.01715$ cm and $L = 1$ cm. This means that as long as the frequency components in the pressure applied across the pneumotach impedance are below 50 Hz, the pneumotach will appear acoustically resistive, and the pressure drop will be proportional to volume flow. A broader bandwidth pneumotach can be made from thin, parallel, *rectangular slits*. The acoustic impedance of one such slit is given by Olson (1943):

$$\mathbf{Z}_A(j\omega) = \frac{12\eta L}{t^3 d} + j \frac{\omega 6\rho L}{5td} \text{ cgs acoustic ohms} \quad (7.134)$$

where

η is the viscosity of air

ρ is the density of air

L is the length of the slit in the direction of flow

t is the height of the slit perpendicular to the direction of flow

d is the width of the slit perpendicular to the flow direction

The critical frequency for this type of pneumotach is again taken as 1/10th the frequency where $X_p = R_p$, or

$$f_{\min} \approx \frac{\eta}{(2\pi t^2 \rho)} \text{ Hz.} \quad (7.135)$$

For slits with $t = 0.01$ cm, the acoustic impedance of the pneumotach would appear real up to about 240 Hz.

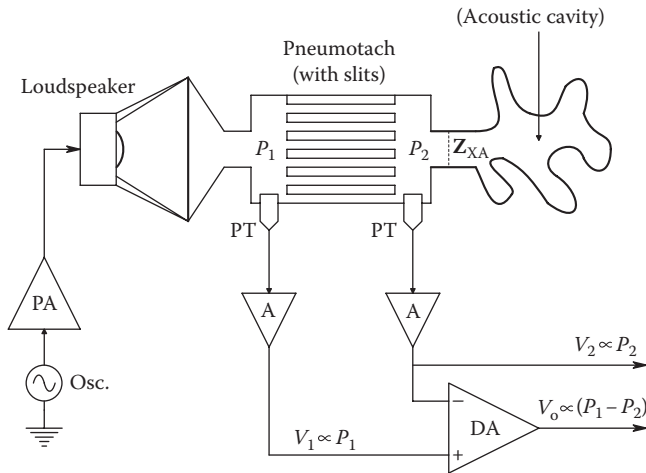


FIGURE 7.36 Diagram of a pneumotach system used to measure acoustic impedance Z_{XA} . Note: Osc = audio oscillator. PA = power amplifier. A = voltage amplifiers. DA = differential amplifier. PT = pressure transducer.

We see that the pneumotach can be used to measure volume flow under laminar conditions by sensing the pressure drop across it with a differential pressure sensor. Average air velocity can easily be derived by multiplying the volume flow by the pipe’s area. The pneumotach can also be used to measure the acoustic impedance of a series acoustic load at frequencies where the pneumotach appears resistive itself. Figure 7.36 illustrates such a system. A pressure source (a loudspeaker) is coupled to the pneumotach. Two independent pressure sensors are used to measure the pressure difference across the pneumotach ($P_1 - P_2$) and the pressure across the unknown acoustic impedance (P_2). The unknown acoustic impedance is then just

$$Z_{XA}(j\omega) = \frac{P_2}{(P_1 - P_2)/R_p} (j\omega) \text{ acoustic ohms.} \tag{7.136}$$

Such means were used by Pimmel et al. (1977) to investigate the acoustic impedance of the human respiratory system over a frequency range of 1–16 Hz and by Peslin et al. (1975) over a frequency range of 3–70 Hz.

7.3.3 Measurement of Linear Position

7.3.3.1 Introduction

The technology of linear position measurement is a very broad area, ranging from position measurements of vehicles from within (navigation, inertial navigation, and a global positioning system [GPS]) to position measurements of mechanical objects, living animals, humans, vehicles, and radiation sources and estimation of vehicle position from without (as in the location and tracking of aircraft around an airport or the tracking of submarines). It is not within the scope of this text to consider vehicle location problems by radar or sonar nor will we consider the specialized topics of inertial navigation and Long Range Navigation (LORAN). We will consider satellite navigation using GPS, however. First, we will focus on the measurement of mechanical position.

Most mechanical systems operate under coordinate constraints, that is, they may only move in a path determined by their associated connections. For example, the position of the

printing head of a dot-matrix printer at any instant is located somewhere along the length of its guide bar; this is a bounded 1-D position problem that has been solved cleverly by not measuring the position of the print head as the printer operates, but by locating the zero or reference position at printer power-up by the use of a simple photoelectric position indicator at the left-hand side of the guide bar. Once the zero position is set, a stepping motor is used to position the print head along the guide bar as required. Because the stepping motor rotates in discrete angular increments with every applied pulse sequence, all the printer CPU has to do is use the net count of stepping motor command pulses stored in an up/down counter to *know* the location of the print head. The discrete nature of printing and symbols allows this simple indexing system to work effectively. However, there are many cases where there must be continuous feedback on position to optimize position control system performance.

In Section 6.4.1, we described the operation of the LVDT, a sensor widely used to give precise information in bounded linear displacement. LVDTs come in all sizes and cover the range of displacements from millimeters to ten centimeters. An LVDT is an AC, carrier-operated sensor with an AC, double-sideband, suppressed-carrier output that requires demodulation to a voltage proportional to core displacement by a PSR, followed by an LPF. In some LVDT models, the carrier oscillator and PSR/LPF are built into the LVDT's case. In all cases, the response speed of the LVDT is limited by the carrier frequency and the time constants of the LPF. Also used to sense bounded linear displacements is the linear Inductosyn® (Farrand 1996).

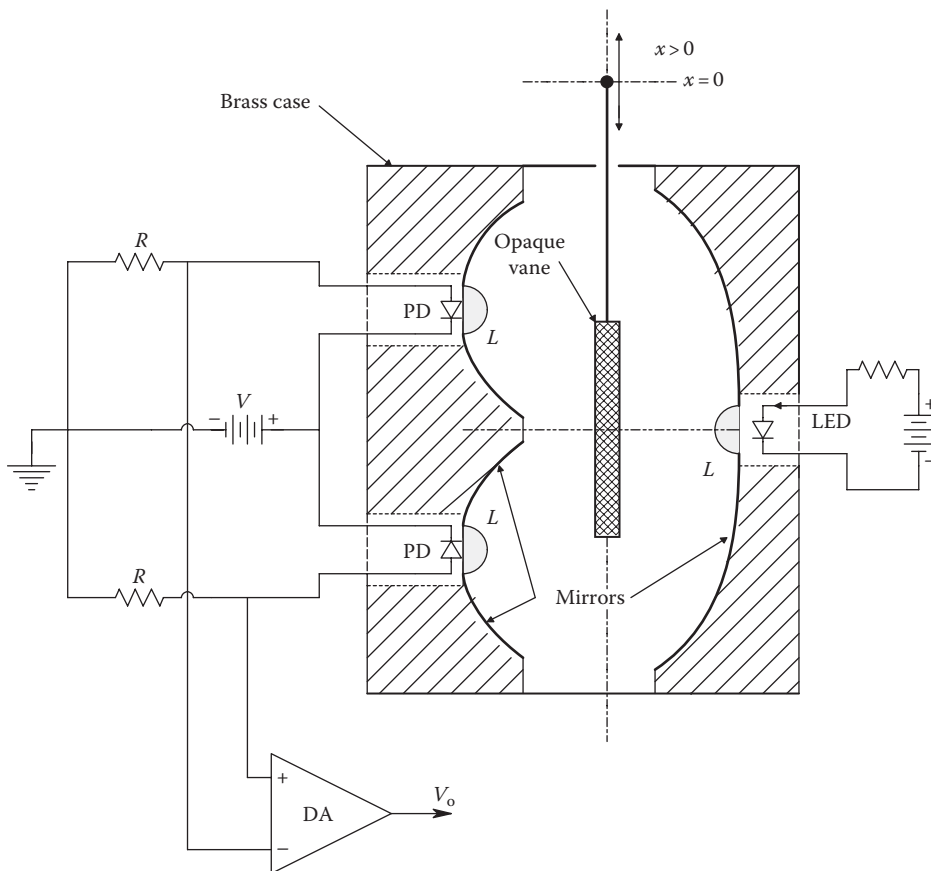
One-dimensional mechanical position can also be sensed photoelectrically. One scheme we have used makes use of a red LED and two photodiodes or phototransistors connected differentially as shown in Figure 7.37. An opaque vane attached to an axially moving shaft blocks the amount of light impinging on the photodiodes. Near the center of its range, the voltage output is linear with shaft position. This system was used by the author to provide position feedback to a servomechanism used to provide controlled stretches to small muscles in physiological studies.

The United Detector Technology, Co. (UDT), described a series of position-sensing systems using four photodetector elements that can determine the 2-D position of a collimated spot of light (UDT 2012). The UDT position sensor systems are operated differentially to resolve the light spot's coordinates, X , Y , on the sensor head. The four-quadrant photosensors are arranged so that B is the response of the 1st quadrant sensor to the fraction of the light spot on it, A is the second quadrant sensor's response, D is the 3rd quadrant sensor's response, and C is the 4th quadrant sensor's response. The center of the spot in X , Y coordinates can be shown to be found from the analog computations (UDT 2012):

$$X = \frac{(A+D)-(B+C)}{A+B+C+D}, \quad Y = \frac{(A+B)-(C+D)}{A+B+C+D}.$$

The spectral response of the UDT sensors peaks at about 850 nm. The sensor can also be used in the reverse-biased photodiode mode. Rise times range from 0.5 to 7 μ s, depending on the model. The UDT linear position sensors have maximum displacement ranges of 0.21" or 1.18". A 1% output nonlinearity in the central 75% of the length is claimed for the displacement of a 1 mm diameter light spot. A listing of UDT optical position sensor systems can be found at Gamma-Sci (2012a,b).

In some cases, linear position can be converted to rotational position by use of a cable passed around a pulley to a weight or spring. The weight or spring keeps the cable tight, and the pulley drives one of many possible angular position-measuring sensors. Probably the simplest angular position-measuring sensor for this application is a 10-turn potentiometer.

**FIGURE 7.37**

Section through a mechanooptical system used to measure small, linear displacements in the mm range. Photodiodes are used to sense differential light flux caused by changes in vane position. *Note:* PD = photodiode. L = lens. LED = light-emitting diode.

The potentiometer can be used as a voltage divider or be one arm of a Wheatstone bridge. This type of system has been used to measure the static level of liquid in a tank or well.

Linear position of distant objects can also be measured by means of reflected laser light, ultrasound waves, and microwaves. In the simplest implementation, as developed for the Polaroid™ camera range finder, a short ultrasound pulse is transmitted toward the object whose position is to be sensed, and the time delay for the echo to return is used as a measure of the distance. The same principle applies to laser ranging, used in modern surveying and geophysics, and with radar, whose uses should be obvious. All of the aforementioned pulse–echo systems rely on a precise knowledge of the propagation speed of the energy in the propagating medium. Unfortunately, the speed of sound is affected by air temperature, pressure, and moisture content. The speed of sound in water is also affected by temperature, pressure, density, and salinity. The speed of light in various media is only slightly affected by factors that affect the speed of sound. Also critical in any pulse–echo system design is the need for a precise, stable clock to measure the time intervals between pulse transmission and echo reception. Such clocks are often derived from thermostatted quartz crystal oscillators using frequency multiplication techniques.

Often, we do not wish to measure the absolute position of an object, but rather a small change in its position. In these cases, we may make use of *interferometric techniques*; coherent light is commonly used. Fox and Puffer (1978) described the use of a holographic interferometric system to measure the 3-D growth of plants in response to applied stimuli such as short flashes of light. Resolution of displacement was on the order of 1/10 of a fringe, or $0.16\ \mu\text{m}$, using a HeNe laser. Growth movement velocities under constant illumination of $0.4\text{--}0.05\ \mu\text{m}/\text{min}$ were observed using the cactus, *Stapelia*.

Drake and Leiner (1984) described the use of a single-mode, FO, Fizeau interferometer to measure the displacement of the tympanic membrane of the cricket in response to a sinusoidal sound stimulus of 90 dB SPL over a range of 1–20 kHz. Their system could resolve peak displacements ranging from 0.1 to 300 Å. The 0.1 Å limit was set by system noise and the 300 Å limit by linearity considerations in the interferometer. A HeNe laser was used. Drake and Leiner's system is illustrated schematically in Figure 7.38. Note that the distal end of the single-mode OF is brought to within about a mm from the reflecting tympanic membrane, this distance not being exceptionally critical except if too large; the power of the return signal picked up by the OF is too small for interferometer operation. The specimen is mounted on a thickness-mode piezoelectric crystal that acts as displacement vernier to adjust the exact working distance between the end of the OF and membrane for maximum interferometer sensitivity.

Northrop (1980) and Northrop and Decker (1977) described a 40 kHz ultrasonic system used as an incremental motion detector. The application of Northrop's system was a prototype, no-touch, infant apnea monitor and convulsion alarm for at risk newborn babies. The system's transmit and receive transducers were installed in the top of a conventional incubator system, and the CW ultrasound beam was directed down to ensonify the surface of the baby's skin, a good reflecting surface. The return signal was amplified and processed electronically as shown in the block diagram of Figure 7.39. The apnea monitor system output

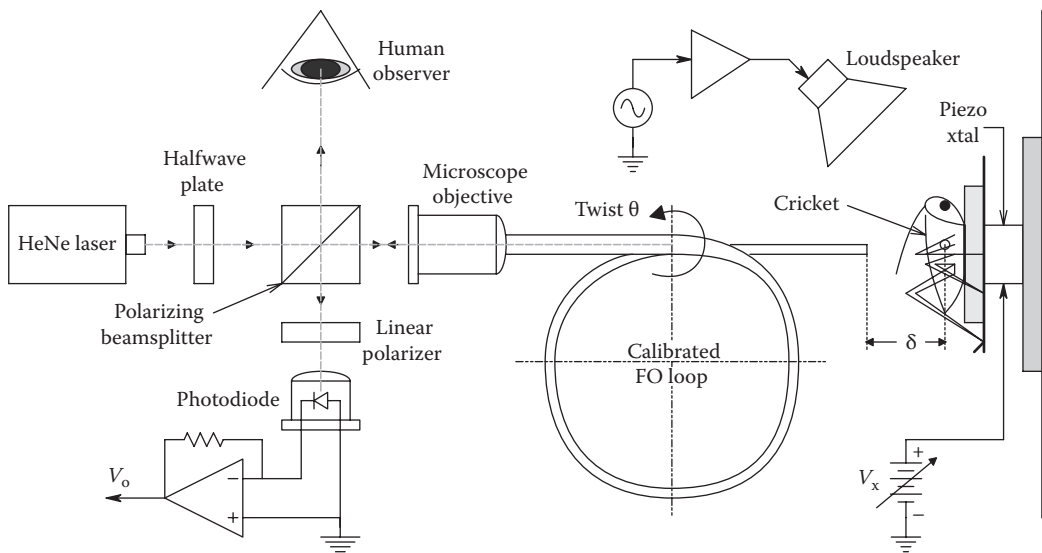


FIGURE 7.38

Diagram of the single OF Fizeau interferometer used by Drake and Leiner (1984) to measure 0.01–30 nm displacements in a sound-excited, insect tympanic membrane. The piezocrystal was used to null the interferometer and introduce a static test displacement.

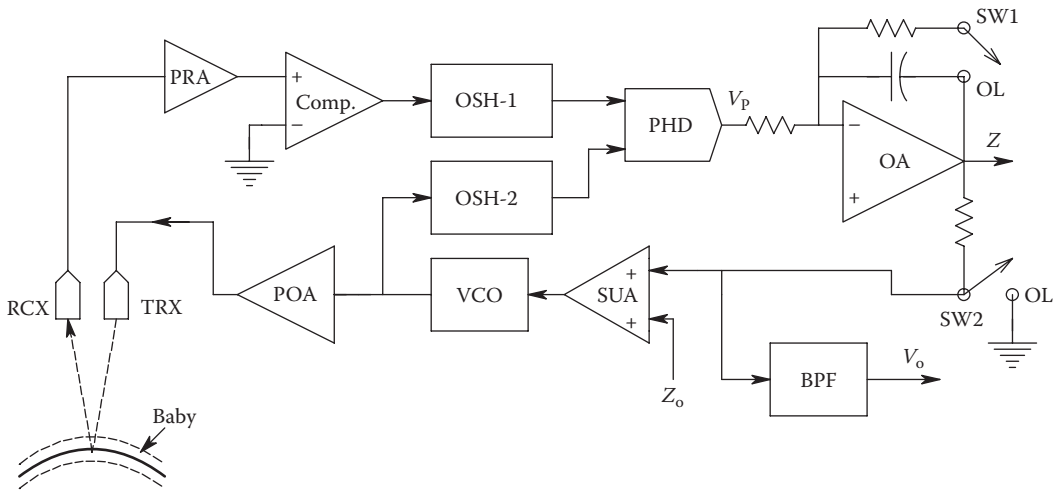


FIGURE 7.39 Block diagram of the 40 kHz ultrasonic, infant apnea monitor system of Northrop (1980). *Note:* TRX = transmitting ultrasound transducer. RCX = receiving ultrasound transducer. PRA = preamplifier. POA = power amplifier. Comp. = analog comparator. OSH = one-shot multivibrator. PHD = phase detector. OA = op-amp integrator. BPF = band-pass filter. SUA = summing amplifier. VCO = voltage-controlled oscillator. OL = open-loop. Z = output voltage. Z_o = DC reference voltage.

is responsive to the small changes in the ultrasound path length. The frequency of the closed-loop system's VCO is changed by feedback so that at any instant, there are a constant number of wavelengths in the air in the round-trip airpath (TRX to BABY to RCX). That is, the controlled variable in this feedback system is the total phase lag between TRX and RCX. It may be shown that the analog output of the apnea monitor is proportional to incremental changes in the total airpath length. Such changes in airpath length would normally be from chest and abdominal movements caused by normal breathing of the sleeping infant. Changes in the normal amplitude and rate of these displacements are what are used to trigger the apnea alarm to summon medical help. The analog output is given by

$$V_o = K_o \frac{2\pi(N + 1/2)c}{2X_o^2 K_v} \Delta x \tag{7.137}$$

where

- K_o is a constant
- K_v is the VCO constant in (r/s)/V
- c is the speed of sound in air
- X_o is the average path length from the transducers to the baby
- N is the integer number of ultrasound wavelengths in $2X$, the total airpath length
- Δx is the (incremental) displacement of the baby's chest and abdomen

N is given by

$$N = \text{INT} \left[\frac{2XK_v}{2\pi c} \right] V_c. \tag{7.138}$$

Here, V_C is a constant (DC) voltage. The ultrasonic apnea monitor was calibrated using an 18" woofer loudspeaker whose DC displacement sensitivity in $\mu\text{m}/\text{amp}$ was known. Limiting resolution of the Northrop system was found to be about $\pm 12 \mu\text{m}$, due to system noise and the wavelength at 40 kHz (0.86 cm). Output sensitivity for $X = 21 \text{ cm}$ was 0.92 V/mm. Once lock is established, the distance X may vary slowly over a wide range, limited only by the working range of the VCO and its ability to keep the SS relation satisfied:

$$f_o = \frac{(N+1/2)c}{2X}, \quad (7.139)$$

f_o is the VCO output frequency. Large, rapid changes in X may cause the system to lose lock. This event can be detected electronically (Northrop and Decker 1977) and used to relock the system to a new N value. Note that the phase-lock design principle of the Northrop system can be extended to microwaves. Such a microwave system might be useful in detecting motion of living persons buried in opaque but otherwise microwave transparent media, such as snow or rubble from collapsed buildings.

The apnea monitor system suffered from the disadvantage of having a range-dependent gain (Equation 7.153). This was not considered important in a system with fixed geometry designed to detect the cessation of movement, not its exact amplitude. You will see in Section 7.3.5 that if the VCO in the feedback loop is replaced with a VPC, both target velocity and range become proportional to voltages independent of X_o .

7.3.3.2 GPS Background

It is general knowledge that the Earth and the other eight planets are in orbits around the sun. The moon is in an orbit around the Earth, as are many man-made satellites used for surveillance, weather, communications, climate research, astronomy, and, of course, navigation. The latter class includes the US GPS SVs, the Russian GLONASS SVs, and certain SVs in equatorial geosynchronous orbits. The orbits of all orbiting celestial bodies, including SVs, obey *Kepler's three laws*.

Johannes Kepler was truly a great mind of his era. Kepler lived between 1571 and 1630. He published his *first law* in 1605. In 1609, he published his *second law* in his text, *Astronomia Nova*. He published his *third law* in 1619 in his monograph, *Harmonices Mundi*. It was his 3rd law that led Isaac Newton to his law of gravitation, not an apple. Kepler certainly was the founder of the science of celestial mechanics; in fact, he coined the word *satellite*. Kepler's three laws can be stated simply for satellites: (1) *The orbits of satellites are ellipses.* (A circle is a special case of an ellipse where the two foci are at the center.) (2) *If two bodies revolve around each other under the influence of mutual gravitational attraction (i.e., the Earth and an SV), a line joining their mass centers sweeps out equal areas in the orbital plane in equal intervals of time.* (3) *The square of the orbital period is proportional to the cube of the orbit's semi-major axis* (NASA 2013). See Figure 7.40 for an illustration of the geometry of an elliptical satellite orbit. Note that the elliptical orbit lies in a plane. The long axis of an ellipse is called its major axis. One-half its major axis is by definition the *semimajor axis*, \mathbf{a} . (In the limiting case where an ellipse becomes a circle, the semimajor axis is the circle's radius.) One-half of an ellipse's minor axis is its *semiminor axis*, \mathbf{b} . The *eccentricity of an ellipse* is defined as $\epsilon \equiv \sqrt{(\mathbf{a}^2 - \mathbf{b}^2)} / \mathbf{a}$. It can be shown that the sum of distances from any point P on the ellipse to the two foci is equal to $2\mathbf{a}$. Note that ellipse focus F' is not centered on a mass, while F is. The GPS SVs are deliberately put in what is as close to a circular orbit as can be obtained. Because of mass

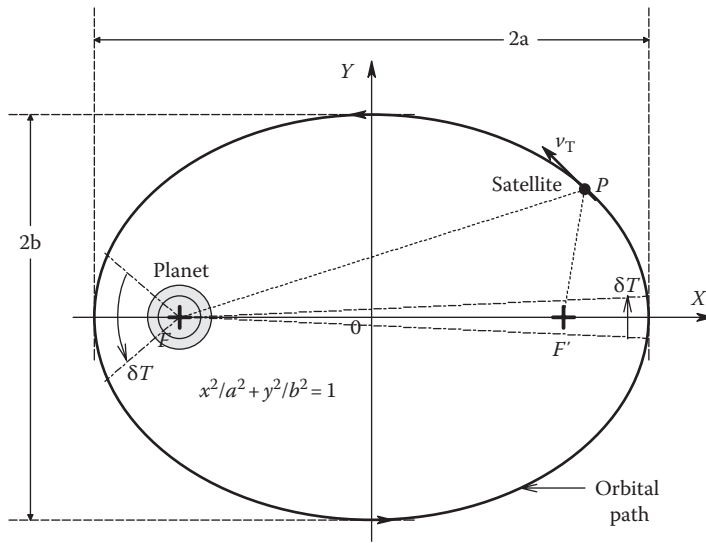


FIGURE 7.40
Elliptical orbit of an SV around a planet.

inhomogeneities inside the Earth and its very slight degree of oblateness at the poles, the actual GPS SV orbits do have a slight eccentricity that is measured and corrected for in the operation of the GPS system.

As we all know, the Earth is not a perfect sphere; for example, it is slightly fatter at the equator, having a mean equatorial radius of 6378 km, a polar mean radius of 6357 km, and a mean radius of 6371 km. Navigation on the Earth's surface has been traditionally done in terms of latitude and longitude, which requires accurate maps having latitude and longitude imprinted thereon. Latitude and longitude are polar coordinates (φ and λ , respectively) in a spherical trigonometric system for locating a point P on the Earth's spherical surface. Specifically, *latitude* (φ) is a north/south measure of angular position on a series of circular planes passing through the Earth, perpendicular to its *polar axis* (the axis the Earth rotates on). The equator is the 0° latitude locus farthest from the poles. *Longitude* (λ) is an east/west measure of angular position in relation to the prime (Greenwich) meridian (0° longitude).

7.3.3.3 Use of the Global Positioning System to Locate Objects in Three Dimensions

The US GPS has been in existence since the early 1980s (Nelson 1999). It was put in place by the US DoD as a response to cold war threats. Twenty-four SVs were put into overlapping, medium Earth orbits at 26,562 km above the Earth's center. There are six orbital planes, each with four equally spaced SVs. Three of the 24 SVs are nonactive *spares*. The orbital period of each SV is 11 h, 56 min. Each SV weighs ca. 787 kg. All GPS SVs carry four atomic clocks: two cesium vapor maser clocks and two rubidium vapor maser clocks. Synchronized precision time keeping between all GPS SVs is essential to precisely triangulate a GPS receiver on the Earth's surface. The cesium beam clocks run with an error of $\pm 2 \times 10^{-12}$, amounting to an error of several ns/day. (See Sect. 14.5 in Northrop [2012] for more details on the GPS SVs and how the UHF L1 and L2 signals are coded.)

Figure 7.41A illustrates a GPS receiver at point P in Cartesian coordinates (x, y, z) on an ellipsoid. Spheroidal polar coordinates $(\varphi, \lambda, R_N + h)$ are transformed to latitude, longitude, and height in the final GPS display, shown in Figure 7.41B. h is the receiver's height above the (ideal) ellipsoid surface. The semimajor axis a of the ellipsoid Earth model is 6378.137 km. The semiminor axis of the ellipsoid is $b = 6356.752$ km. The Earth is a spheroid, flattened at the poles relative to the mean equatorial radius, a . This 0.3353% distortion of the Earth's volume (hence mass) from an ideal sphere causes slight, unique distortions in each of the SVs' ideally circular orbits that must be compensated for in calculating the receiver's position, given the four SVs' time and ID signals. The actual algebra of coordinate transformations and position determination is beyond the scope of this section.

The purpose of the US GPS system is to locate a GPS receiver at a position $P(R_N + h, \varphi, \lambda)$ in spherical coordinates or at $P(x, y, z)$ in rectangular coordinates on or above the Earth's

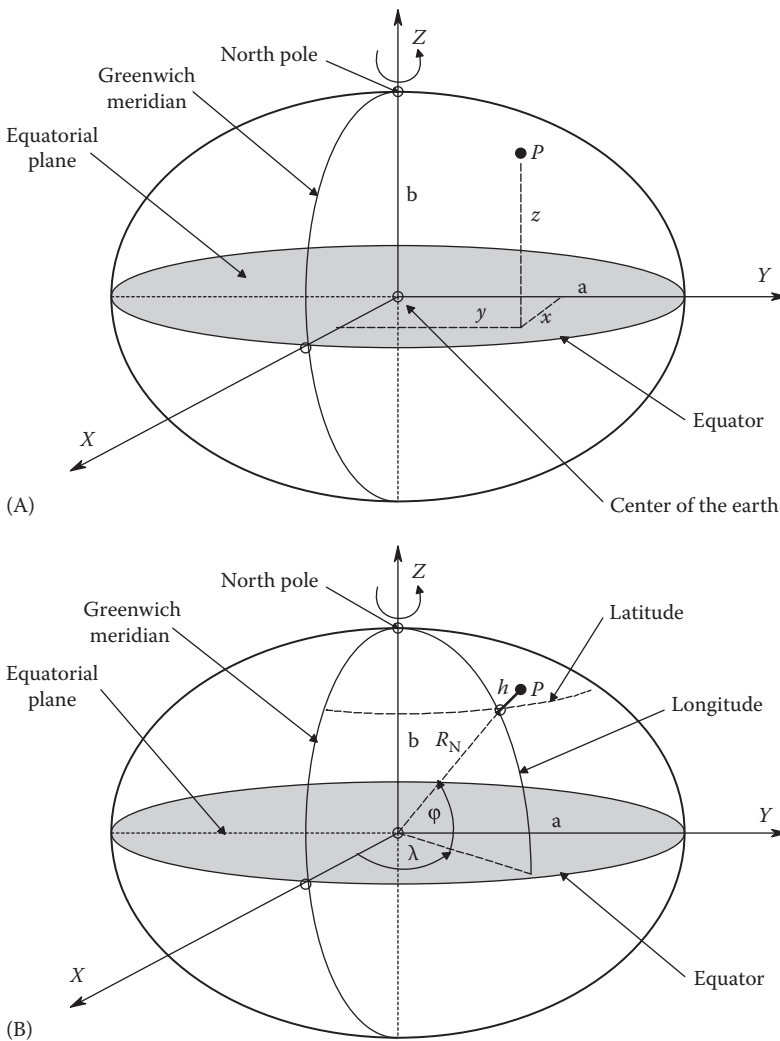


FIGURE 7.41 (A) Rectangular coordinates used to describe the location of GPS receiver, P . (B) Spherical coordinates used to describe the location of P . h is the height of P above the Earth's surface.

surface (see Figure 7.40). GPS location has obvious application to the navigation of moving vehicles including ships, wheeled and tracked vehicles, aircraft, and missiles. GPS location is also used by hikers, explorers, surveyors, mapmakers, oceanographers, and farmers to lay out fields and guide harvesting machinery. Most recently, its applications have been extended to tracking delivery trucks, rental vehicles, and cell phones and keeping track of children at play and convicts on parole. In these applications, a precision GPS receiver radios the receiver's coordinates to a base station where it is shown on a detailed map. Shipboard GPS readouts are generally in latitude–longitude; modern units can also give course vectors to entered waypoints (e.g., buoys), receiver departure from the desired course, and position on a moving map. GPS units used by hikers can plot to scale the path the hiker has taken. GPS receiver systems have evolved from the ponderous 73 kg *Macrometer-1000*, developed at MIT in 1982 (Langley 2000), to the modern, handheld, wrist mounted cell phone units in use today. Receiver antennas for the L1 and L2 GPS satellite signals at 1.57542 GHz and 1.22760 GHz, respectively, now are small as a 2.5 cm². Note that larger antennas give stronger signals. (See Sect. 14.3.2.1 in Northrop [2012] for a description of broadband fractal antennas.)

In summary, the four components of the basic GPS system include (1) the SVs, (2) the signals they send, (3) the ground control system for satellite signal correction, and (4) the GPS receiver design.

7.3.3.4 GPS System Development

Early (1980s–1990s) civilian GPS receivers had position (circle radius) accuracies of ca. 15 m. Differential GPS (DGPS) was developed by the US Coast Guard in March 1999 using fixed shore stations to augment accuracy. DGPS gives ship position accuracies of ca. 0.75 m. DGPS covers all US coastal waters, the Great Lakes, Puerto Rico and the US Virgin Islands, and portions of waters around Alaska and Hawaii (US Coast Guard 2013). A simple DGPS receiver normally calculates its position by measuring the time it takes the L1 signal from the SV to reach its position. From an a priori knowledge of where the SV is in its orbit at the time of transmission (ephemeris data), how long it takes to send the signal, and knowing the speed of the signal, it can compute what is called the *pseudorange* (PR) to the SV. PRs from four SVs are necessary to calculate the receiver's position, $\mathbf{P}(R_N + h, \varphi, \lambda)$. (Arcane spherical trigonometry [not covered here] is used to find \mathbf{P} .) Each PR is corrected by data from a DGPS base station. The correction data are sent at an amazingly slow 100 or 200 baud, which generates a delay of 2–5 s before the corrected receiver position is available for display (DePriest 2006). PR error can be attributed to the following causes, as shown in Table 7.1.

TABLE 7.1

Sources and Magnitudes of GPS PR Error

Error	Value
Ionospheric (refraction) delays	4.0 m
Clock	2.1 m
Ephemeris (orbital position)	2.1 m
Multipath (reflections of UHF signal)	1.0 m
Tropospheric delays	0.7 m
Receiver	0.5 m
Cumulative	10.4 m

The Achilles' heel of the DGPS system is that the farther the receiver is from the base station, the poorer the accuracy of the correction, hence the poorer the position estimate by the DGPS receiver. For New England waters, USCG DGPS base stations are located at Sandy Hook, NJ; Horiches, LI, NY; Acushnet, MA; Portsmouth Harbor, NH; Brunswick, ME; and Penobscot, ME.

The GPS *wide area augmentation system* (WAAS) is also called WAGPS or space-based augmentation system (SBAS). The European Community (EC) developed their own version of the US WAGPS called EGNOS, and the Japanese have their version, called MTSAT. Similar to the USCG-sponsored DGPS that uses base stations largely on the coast, the US FAA-sponsored development of the WAAS system uses ca. 25, precisely surveyed ground reference stations positioned across the United States and several master stations. The WAAS system uses its own three GEO SVs that are in high, *geosynchronous equatorial orbits*, plus the 24 GPS SVs. The geosynchronous WAAS SVs useful to the United States are called AOR-W, AOR-E, and POR. An equatorial geosynchronous orbit has a radius of 35,768 km from the Earth's center; the GEO SV's tangential velocity is 3,075 m/s. SVs in geosynchronous equatorial orbits are stationary above a fixed point on the Earth's equator; that is, their orbital angular velocity equals the Earth's rotational angular velocity (4.167×10^{-3} deg/s). AOR-W is at lat. 0° (equator) and long. 54°W . AOR-E is at lat. 0° , 15.5°W . POR is at lat. 0° , long. 178°E , almost mid-Pacific.

WAAS works in the following manner: The WAAS ground stations (and EGNOS ground stations) harvest GPS position information across the country and compare it to their accurate ground positions and each generates an error vector with ionospheric delay, clock errors, and ephemeris error components. Each station transmits this information to a master ground station that analyzes and compiles a master error matrix, which is a function of latitude–longitude coordinates on a rectangular grid over the Earth's surface. The grid is subdivided into eight, north–south strips approximately 45° of longitude wide. Some 1607 points are in the grid. The computed, interpolated/extrapolated GPS error estimates at the grid points are then uplinked to the geosynchronous satellites. An updated clock error matrix is uplinked every minute; ionosphere and ephemeris data are uplinked every two minutes. A WAAS receiver first determines its position from the 4 or 5 GPS SVs it *sees*, and then from the downloaded correction data from the GEO SV, it determines where it is (approximately) on the grid, determines the closest four points, and interpolates its position correction from the error matrix. The WAAS system can work with three, two, or one correction point if others are not available (DePriest 2006). Typical 95% horizontal accuracy of the WAAS system was 1.198 m (X and Y) in Minneapolis, MN, and the 95% vertical accuracy was 1.937 m (Z) (FAA 2005).

Recall that WAAS was originally developed for aircraft navigation. Because of its wider area availability, several companies have developed WAAS receivers that use not only the corrections downloaded from a GEO satellite but also phase information between the L1 and L2 carriers and sophisticated postprocessing algorithms. For example, the OmniSTAR 8000 HP™ WAGPS system has an advertised horizontal accuracy of 10 cm 95% and a height accuracy of 20 cm 95%. The 95% means that there is a probability of 0.95 that a position reading will lie within a 10 cm radius from the true location for (X , Y) and within 20 cm of the true height (Z). OmniSTAR lists applications of this precise WAGPS system, including mooring; crane monitoring; seismic surveying; oil drilling; robot vehicle guidance; inshore dredging (e.g., for power cable trenches); trenching for pipelines, power cables, and FOCs; farming; and rough surveying.

To function, a standard GPS receiver requires simultaneous signals from *four* (overhead) orbiting SVs to determine its position in longitude, latitude, altitude (and time). From these

data, the receiver can also compute a course vector (speed and direction). Position calculation requires that the atomic clocks onboard the GPS SVs, and the receiver, be synchronized in order that the propagation time of the RF signal from an SV to the receiver is precisely known. (An error in propagation time, $\Delta\tau$, of only $1\ \mu\text{s}$ can give a position error of 300 m!) The determination of the four unknowns in receiver position P (longitude $[\varphi]$, latitude $[\lambda]$, height $[R_N + h]$, and time error $[\Delta\tau]$) requires the simultaneous solution of four nonlinear equations for which the four SVs provide input data. Quoting from the u-blox 2009a GPS tutorial:

The details displayed and calculations made by a GNSS [GPS] receiver primarily involve the WGS-84 (World Geodetic System 1984) reference system. The WGS-84 coordinate system is geocentrically positioned with respect to the center of the Earth. Such a system is called ECEF (Earth Centered, Earth Fixed). The WGS-84 coordinate system is a three dimensional, right-handed, Cartesian coordinate system with its original coordinate point at the center of mass (=geocentric) of an ellipsoid, which approximates the total mass of the Earth.

The most precise GPS receivers now in use are the real-time kinematic (RTK) systems. RTK systems are capable of making survey-grade position measurements in real time and provide immediate accuracy to within 1–4 cm. RTK systems require that a *fixed base station* (reference GPS receiver) be within 6–10 miles of the roving (mobile) receiver. They also require ca. 30 min to initialize before operation! The roving receiver initially must acquire the signals from 5 GPS SVs. It then must maintain contact with four SVs, as well as the fixed base station (RTK GPS 2010). RTK systems are expensive, costing from \$15,000 to \$50,000. Perhaps in the future, we may see autonomous crop cultivation robots, guided by RTK GPS.

7.3.3.5 GPS Signals

Data broadcast from all GPS SVs has the following signal structure: Two pseudorandom noise (PRN) codes modulate the transmitted signal. Each SV has its own, unique PRN code. One is the *coarse/acquisition* (C/A) code; the other is the *precision* (P) code. The C/A code is used by the receiver for satellite acquisition and identification and for coarse position determination by lower-accuracy receivers. The P -code is for higher-precision position determination; it can be encrypted to create a secure signal called the Y -code that is used by the military and other authorized users. The C/A code is a gold code of register size 10, which has a binary sequence length of 1023 and a clocking rate of 1.023 MHz; thus, it repeats itself every 1 ms. The P -code is of length 2.3547×1023 with a clock rate 10 times the C/A code or 10.23 MHz. The L2 carrier carries only the P -code. The second P -code signal permits a dual-frequency measurement of the ionospheric group delay (Nelson 1999). Broadcast from the GPS SVs, the GPS L1 carrier, is at $10.2300\ \text{MHz} \times 154 = 1.57542\ \text{GHz}$; the L2 carrier center frequency is at $10.2300\ \text{MHz} \times 120 = 1.22760\ \text{GHz}$. The new L5 (*safety of life*) channel center frequency is at $10.2300\ \text{MHz} \times 115 = 1.17645\ \text{GHz}$. The L5 civilian-use signal is being implemented in new GPS SVs launched during and after 2010. The L1, L2, and L5 carrier frequencies are derived from a master clock in the SV by PLL techniques.

Modulation of the GPS L1 and L2 signals is complicated, as is their demodulation and position calculation by the receiver. Figure 7.42 illustrates the basic modulation process used on the L1 and L2 GPS SV carriers: NM is the *navigational message* data that also modulate the L1 and L2 carriers. The NM code describes an SV's orbit, clock corrections, and

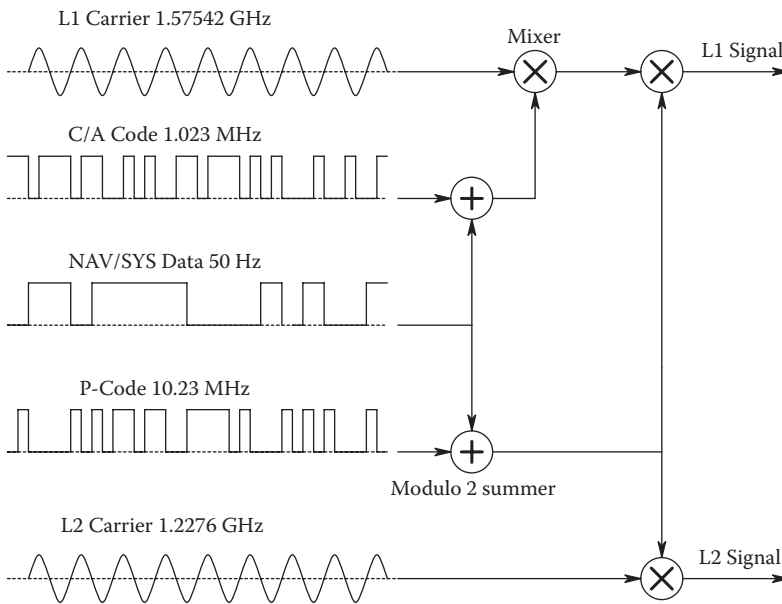


FIGURE 7.42 Basic GPS satellite transmitter modulation process, showing the structure of the L1 and L2 signals.

other arcane system parameters. The C/A code is a gold code of register size 10, which has a binary sequence length of 1023 and a clocking rate of 1.0230 MHz; thus, it repeats itself every 1 ms. The C/A code is used by the receiver for SV acquisition and ID. The P-code is used for high-precision position determination (Northrop 2012).

The *control segment* for the GPS system consists of a master control facility located at Schriever AFB in Colorado Springs, CO, and observer stations located in Hawaii, Ascension Island in the S. Atlantic, Diego Garcia Island in the Indian Ocean, and Kwajalein Island in the Pacific. These monitor stations measure GPS SV signals and develop precise orbital models (ephemeris data) for each SV that include clock corrections. This information is sent to the master control station in Colorado that uploads the necessary ephemeris data and clock corrections to each SV. The SVs then send subsets of the orbital ephemeris data to GPS receivers in the slow navigation message. This continuous updating and correcting is a necessary process to maintain the accuracy of the GPS system (Northrop 2012).

7.3.3.6 GPS Hardware

A GPS antenna must have a compact, broadband design; such a design will usually use a fractal or other space-filling geometry (see Northrop 2012, Section 14.3.2.1). The GPS L-signals from the antenna are conditioned by a low-noise preamplifier and sent to an integrated GPS receiver chip for processing. Siakavara (2007) presented the design for a compact, fractal microstrip antenna composed of the union of four inverse Koch fractal patches. This broadband antenna operates in three frequency ranges (GPS L bands, Digital Cellular System-DCS1800, and 2.7 GHz). In the GPS bands, the antenna generates a circularly polarized field with the maximum strength at the broadside direction. Another planar fractal antenna design for the GPS L1 (1.57542 GHz), GSM (1.850 GHz),

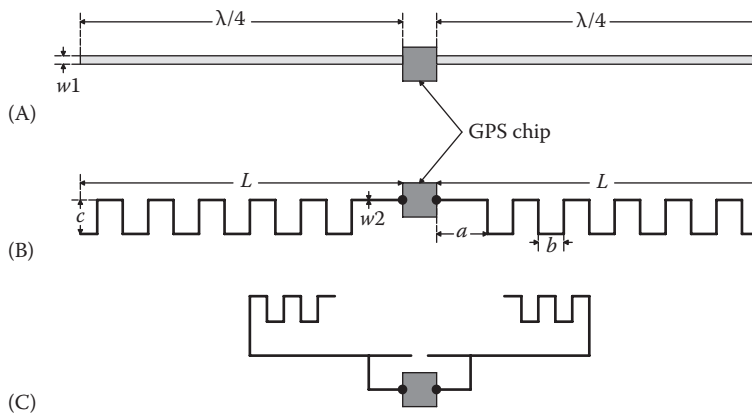


FIGURE 7.43
 (A) Basic quarter-wave dipole antenna. (B) A Greek key dipole antenna. (C) Variation on the Greek key dipole.

and Wi-Fi (2.440 GHz) bands was described by Azaro et al. (2007). Their antenna, before the Koch triangle nibbles were taken, measured 4.5×4.5 cm. (An excellent tutorial paper summarizing and evaluating various GPS antenna designs can be found at u-blox 2009b.)

There are several other antenna designs used with GPS receivers other than planar fractal geometries. Perhaps the simplest GPS antenna is the two-element ($2 \times \lambda/4$) dipole at the L1 frequency; each $\lambda/4$ arm of the dipole is 4.761 cm. This dipole size can be shrunk by making the dipole in 2-D, using a flat meander pattern, such as the Greek key crank, on the arms (cf. Figure 7.43).

If design space permits, greater omnidirectional sensitivity can be found using the quadrifilar helix antenna (QHA) design described in u-blox 2009b. For the GPS L1 frequency, if the QHA antenna is surrounded with air, its outline will be ca. 6 cm high \times 4.5 cm in diameter; if gold wire is deposited on a high-dielectric constant ceramic, the size can be shrunk to 1.8 cm high \times 1.0 cm diameter without loss of performance. Because of its omnidirectional sensitivity, the QHA can pick up F1 signal reflections from terrain and vehicles such as aircraft, often causing degraded navigational performance. Many other passive GPS antenna designs exist (e.g., chip antennas, loop antennas, planar inverted F antennas, and patch antennas), too many to describe here.

The so-called active GPS antennas are used when the antenna must be placed over a few cm from the receiver, such as on the roof of a metal vehicle. Basically, an LNA is located directly at the antenna terminals and is used to condition the received GPS L-signals (gain typically ca. 15 dB) and drive a transmission line/coaxial cable of up to ca. 5 m at its characteristic impedance (e.g., 52Ω) to prevent standing waves. The LNA typically may have a noise factor $F \cong 2$ dB (Northrop 2012). A SAW BPF can be used to enhance the SNR at the receiver chip.

In GPS tags used for tracking animals, the antenna is often a space-saving, planar fractal design, on or near the printed circuit board (PCB) carrying the GPS chipset.

The electronic circuitry bulk required for GPS receiver functions has been steadily shrinking, driven by the need to embed GPS in mobile telecommunications and computer systems and also by the need to actively track vehicles, goods, people, and animals from a distance. Several manufacturers offer compact GPS chips for size-critical applications.

A laptop computer can be turned into a GPS system with the GlobalSat WorldCom Corp. GSAT™ USB Dongle GPS receiver, model ND-100S (GSAT 2013). This device is the size of

a large flash drive ($6.55 \times 2.3 \times 1.1$ cm) and contains an SiRFstarIII® chip; it runs with all popular OSs (Windows 98/2000/XP/Vista/7, Mac OS8/9/10, and Linux RedHat 7.8/8.0/9.0).

The Swiss corporation, u-blox AG, makes a complete line of GPS and Galileo receivers, chipsets, and antennas (u-blox 2013). Their UBX-G6000 series GPS chips contain an RF front end with an LNA and have the following characteristics/specifications: The receiver has 50 channels and receives GPS L1 C/A code, also adapted to GALILEO L1 open service (with upgrade, Europe). Maximum update rate is up to 5 Hz ROM, 2 Hz flash memory. Accuracy (all SV signals ≥ -130 dBm)—position, 2.5 m CEP, SBAS 2.0 m CEP. Acquisition times with TCXO—cold starts, 28 s; warm starts, 28 s; hot starts, 1 s. RF sensitivity (with crystal osc.)—acquisition, 160 dBm; tracking, 160 dBm; cold start, 147 dBm. Operational limits—velocity 500 m/s (972 knots), altitude 50,000 m. Serial interfaces—1 UART (UBX-G6010), 2 UARTS (UBX-G6000), 1 USBv.2.0 12 Mbits/s, 1 DDC (I²C compliant), 1 SPI. Power consumption (important for animal tagging)—59 mW continuous, 30 mW at 1.8 V power save mode. Supply voltages, single voltage supply—1.8 or 2.5–3.6 V; dual voltage supply—1.4/1.8 or 1.4/2.5–3.6 or 1.8/2.5–3.6 V. (*u-blox* makes several other single-chip and two-chip GPS systems [e.g., UBX-G5010 and G5000/G0010], as well as a variety of GPS antennas.)

One of the smallest GPS chipsets is made by *Rakon*, a New Zealand company (www.rakon.com/ [accessed April 16, 2013]). Originally designed to go into cell phones, one Rakon chip measures only 2.5×2.0 mm. Its small size makes the Rakon chip attractive for animal GPS tag designs. Clearly, an effective antenna for the Rakon chip would have to be larger than the chip, a case of the *tail wagging the dog*.

A Telonics™ ST-18 GPS tag measures 12×5 cm and weighs ca. 200 g in air. The Telonics ST-14 GPS tag is larger, 16×10 cm, and weighs ca. 750 g (larger batteries). Other companies offer GPS tags: the Sirtrack KiwiSat 101™ is 18×6 cm (cf. www.lotek.com/ [accessed April 17, 2013]). Wildlife Computers, Inc., offers several models of ARGOS-compatible tags. GPS tags are not inexpensive—for example, the Wildlife Computers, Inc. SPOT5 tag costs \$1,350 or more, depending on custom features (see www.wildlifecomputers.com/ [accessed April 17, 2013]).

A survey of GPS chip receivers may be seen at <http://gpstekreviews.com> (accessed April 17, 2013).

Figure 7.44 illustrates the block diagram of a simple, civilian-use, GPS receiver based on the design using the Zarlink (now Microsemi October 13, 2011) chipset used in an Orion GPS (www.oriongps.com/ [accessed April 17, 2013]) receiver. Note that it uses a superheterodyne architecture; the output of a 10 MHz, temperature-controlled quartz crystal local oscillator is processed by an on-chip PLL frequency multiplier and mixed to give intermediate frequencies of 175.42, 35.42, and, finally, 1.405 MHz. The 2-bit, quantized output of the triple-conversion receiver/detector is sent to a DSP package consisting of a parallel, 12-channel signal correlator used for SV identification by its unique PRN signature. The correlator's operation and its output are managed by a 32-bit, RISC microprocessor operating at 20 MHz. On-chip memory includes 512 kB 20 ns SRAM and 256 kB EPROM holding Orion GPS firmware. The firmware contains the *GPS Almanac*, which has coarse position (ephemeris) data for all possible GPS SVs. (All GPS receivers must have almanacs.)

An inexpensive, handheld GPS receiver generally only works with the L1 signal and processes one GPS SV's signal at a time, extracting the data it needs to calculate its position on an alternating basis. More expensive receiver designs have five, independent DSP channels for parallel computation of position. Very high-end GPS receivers may have a dedicated channel for each SV in the constellation.

All GPS systems are essentially noisy. That is, if a series of position determinations are made using a fixed receiver, each reading is likely to be some distance from the receiver's

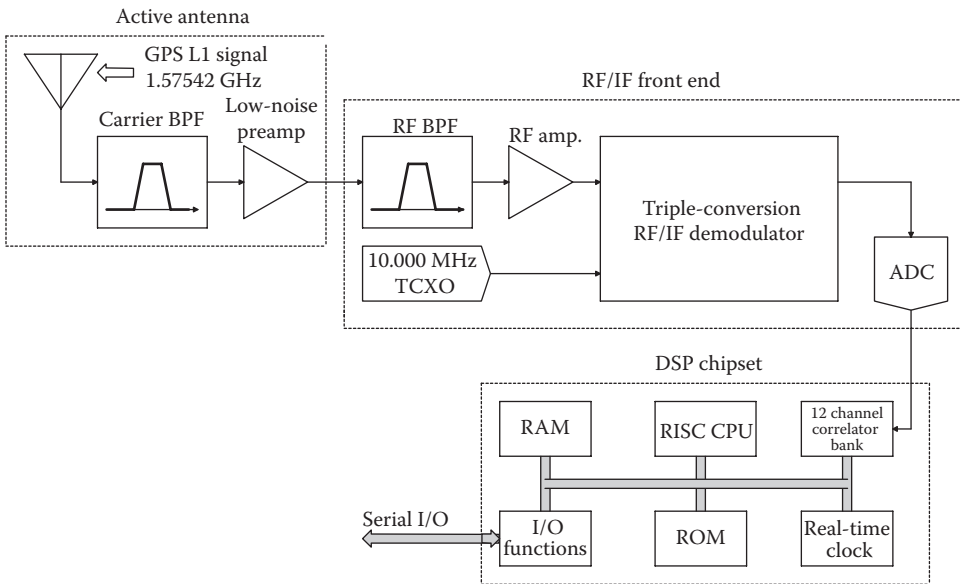


FIGURE 7.44 Block diagram describing the architecture of a simple GPS receiver.

true position. GPS error, ϵ , is defined as the vector distance from the true receiver position to the indicated position. There are several ways of statistically describing sets of errors. One can start by making a histogram of error magnitudes in the horizontal plane for a large number of measurements for the same receiver location. It is found that most such histograms can be fit by the continuous Weibull distribution with shape factor $\beta = 2$ or a continuous Rayleigh distribution. By integration of the Weibull distribution, it can be shown that

$$\text{Probability } (\epsilon \leq r) = 1 - \exp\left[\frac{-r^2}{\text{MSE}^2}\right] \tag{7.140}$$

where MSE is the *mean-squared error* calculated by

$$\text{MSE} = \left(\frac{1}{N}\right) \sum_{j=1}^N \epsilon_j^2. \tag{7.141}$$

The root-mean squared error (RMSE) is simply the square root of the MSE. From Equation 7.156, we can illustrate the three principal error statistics used by GPS manufacturers and evaluators. CEP or circular error probable assumes that $\text{Pr}(\epsilon \leq r) = 0.5$. $r = 0.83$ for this to happen. This means that for large N , about half the positions determined will lie inside a circle of radius $r = 0.83 \times \text{RMSE}$. If $r = 1 \text{ RMSE}$, 63% of a large number of measurements will lie inside a circle of this radius. Finally, the Rayleigh probability model tells us that if $r = 1.73 \times \text{RMSE}$, the probability is that 95% of the measurement points will lie inside a circle with this radius drawn around the known location of the receiver. One problem in evaluating GPS receiver performance is that manufacturers often just give a statement such as *error = 1 m* without specifying how the *error* is evaluated.

7.3.3.7 GPS Tag Applications in Biology

One important biological application of the GPS system is in ecological and behavioral studies where GPS-tagged wild animals, including certain fish, cetaceans, seals, and sea turtles, are tracked. The GPS-equipped tags must frequently communicate the animal's position data, plus recorded behavioral measurements, via a radio link to a data-compiling station (on the Earth's surface, or to a satellite, thence to an Earth station).

One application of GPS radio tags has been the tracking of great white sharks off the Californian and Mexican coasts. This work has been done by Michael Domeier, sponsored by the National Geographic Society (NOAA 2010). After catching a shark, the fish is lifted from the water and unhooked and then measured, blood and tissue samples taken, and a GPS radio tag is literally bolted to its dorsal fin before release. Every effort is taken to minimize stress on the shark. The tag periodically samples parameters such as shark depth and water temperature, and when the shark surfaces with its dorsal fin and tag antenna in the air, the tag senses the antenna is in air and it broadcasts its GPS coordinates plus time, date, and sample data to an ARGOS SV. These GPS radio tags are reported to have a functional lifetime of ca. 6 years, giving behavioral biologists and ecologists a long-term record of a shark's feeding, migration, and mating behavior.

Long-life GPS tags have also been used to track ocean sunfish (*Mola mola*) in the Western Atlantic (Sims et al. 2009). These GPS tags transmit to the ARGOS satellite system when on the surface. The sunfish GPS tags were attached behind the fish's dorsal fin to its body with a 1.5 m long permanent monofilament tether. The tag body was roughly conical, 15 cm long by 80 mm in diameter; the ARGOS satellite antenna was 17.1 cm in length. Sunfish like to bask at the ocean surface, giving the tag good opportunities to broadcast data, but they also have been observed to dive as deep as 472 m (Sims et al. 2009).

GPS tags have also been used to study the behavior of elephant seals in the ocean. Glued to the seal's neck behind the head, these tags periodically sample and record animal depth and data on swimming dynamics supplied by x -, y -, z -axis accelerometer ICs. Acceleration data and the animal's depth record and position are uploaded to a satellite when the animal surfaces to breathe (Tremblay et al. 2009, Costa et al. 2010). Sea turtles have also been tagged with GPS data loggers epoxied to their shells (Seaturtle 2010).

In Alaska and in the high Rocky Mountains (Glacier National Park), the elusive wolverine (*Gulo gulo*) has been tagged with GPS radio collars in order to study its habits (Harrington 2010). In the Glacier National Park study (Copeland et al. 2004), all captured animals were surgically implanted with intraperitoneal VHF radio transmitter tags (Telonics Models 200 for kits and 400 for adults) (Telonics, Inc., Mesa, AZ. www.telonics.com/ [accessed April 17, 2013]). In addition, some GPS-/ARGOS-equipped collars were also used. In the Alaska study (Lewis et al. 2008), store-on-board GPS collar tags were used (GPS-3300S, by Lotek Wireless, Inc., Newmarket, Ontario, Canada). The collars had a programmable, remote-release mechanism, set to come off the animal 24 weeks after tagging. The tagged collar was retrieved and data downloaded with GPS locations and time/date, permitting the wolverine's extensive meanderings to be detailed.

GPS tags have also been used to study elephant migrations and the behavior of elephants in Mali in Africa (Kiger 2010).

7.3.3.8 GPS Summary

We have seen that the evolution of accurate electronic location of a point on the Earth's surface, the ability to navigate on sea and in the air, has evolved from man's ability to launch SVs into orbits around the earth. This evolution has also depended on the development

of ICs and microcomputers, precision maser clock oscillators, and effective microwave (L-band) radio equipment. Single point measurement position accuracy has increased from early civilian GPS receivers using L1 signals from GPS SVs with selective availability (SA) enabled, which gave a whopping 100 m horizontal error and 156 m vertical error. Government concern for Cold War threats and terrorists using GPS to guide munitions kept SA active until May 2000, when it was finally turned off. By the late 1990s, several schemes had been developed to enhance basic GPS accuracy, with SA on, rendering SA obsolete. Basic civilian GPS accuracy with SA off is ca. 15 m. (The DGPS system developed by the coast guard for general ship navigation initially gave 3–5 m accuracy and now is on the order of 0.75 m using both L1 and L2 satellite signals.)

WAGPS receivers with postprocessing can realize ca. 1 cm accuracy. (Note that signal averaging can improve resolution.) We can expect to see mm accuracy in the future, accurate enough for AA or A2 land surveying. Will the transit and theodolite eventually become obsolete for land surveying? Note that GPS accuracy is not without dollar cost. The most accurate GPS receivers must receive and use correction signals, as well as use parallel processing on the received L1, L2, and L5 signals. This electronic and DSP complexity is expensive. Current prices (2013) for top-of-the-line global navigation satellite systems (GNSSs) lie in the range of 13–15 k\$ (USD). The high-precision, surveying grade GNSSs make effective use of the GPS L1, L2, L2C, and L5 signals. Handheld units for hikers, boaters, golfers, and archaeologists cost about \$100–\$300 USD. Marine WAGPS systems for large vessels having map displays cost around 1–5 k\$.

To further explore GPS technology, see the excellent tutorial on GPS satellite navigation available online from u-blox (2009a).

7.3.4 Use of Optical Interferometry to Measure Δx

7.3.4.1 Introduction

There are many applications where it is necessary to accurately measure very small linear mechanical displacements and distances between objects on the order of nanometers. Such applications include the measurement of deflection of diaphragms and cantilevers in force, acceleration, and pressure sensors and thickness measurements in the microfabrication of LSI circuit chips and nanomachines. One means of measuring small distances without contact is to use optical interferometry. An interferometer compares the phase between a reference light beam and a measurement beam by using the phenomena of constructive and destructive interference that occurs when two, coherent, lightwave beams are summed on a surface or a photodetector. Many kinds of interferometers have been developed for various applications in metrology. These include the *Fabry–Pérot (F-P)*, *Fizeau*, *Mach–Zehnder*, *Michelson*, *Nomarski*, and *Sagnac* designs. A number of other interferometer designs exist; see for example, Chapter 9 in Hecht (1987) and Chapter 4 in Sirohi and Kothiyal (1991). See Figures 7.45 through 7.48 for simplified diagrams of some of these interferometers.

There have been several areas of noninvasive medical diagnosis in which information is gained by the no-touch, optical interferometric measurement of extremely small (e.g., on the order of nanometers) mechanical displacements of external surfaces. One such displacement is the extremely small vibration of the eardrum in response to incident sound pressure (Drake and Leiner 1984). Such measurement can be useful in the study of the mechanics of hearing, including the tympanal reflex in response to sudden, loud sounds. Another physical displacement is the movement of a tooth in situ in response to a lateral force. Measurement of the mechanical compliance of teeth in gums can be useful in orthodontic

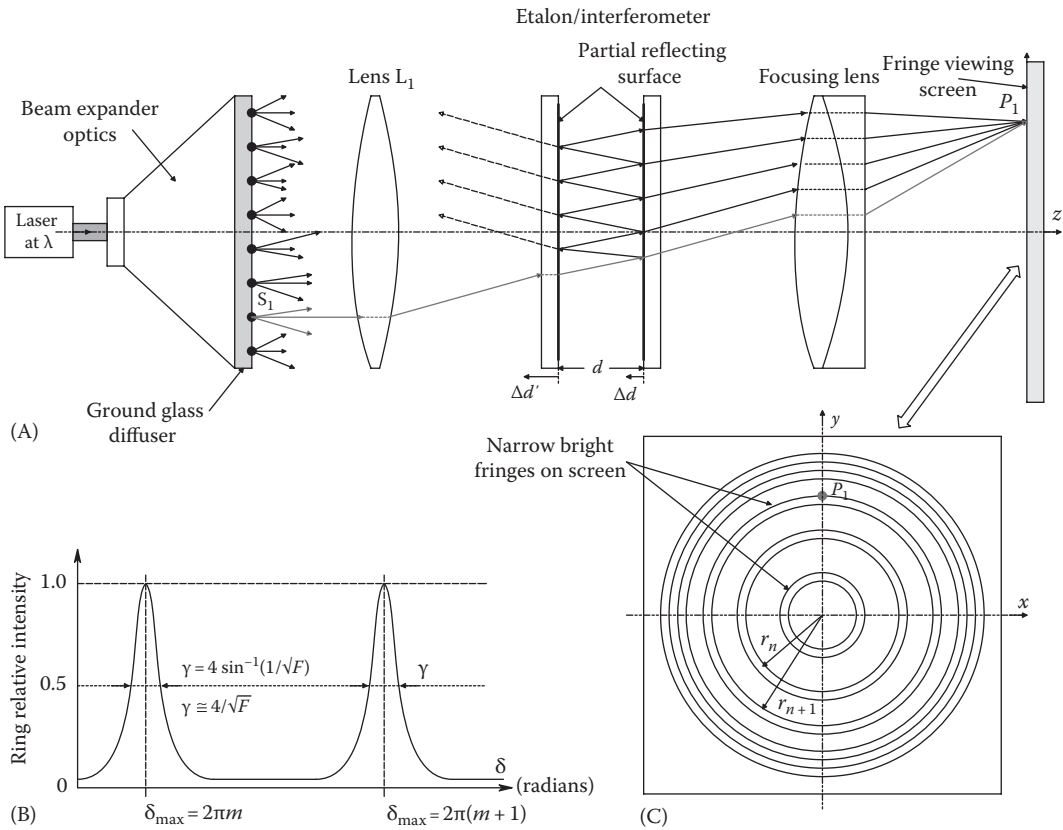


FIGURE 7.45 (A) FPI. (B) Relative ring intensity in the interferogram. Note sharpness of peaks. (C) View facing the interferogram on the screen. F is the finesse. See text for description.

research and in the study of dental health. Still another application of optical interferometry lies in the measurement of very small displacements of the skin caused by blood flow in capillary beds. Microplethysmography of the skin surface can have application in studies of the control of peripheral circulation and in assessing microcirculation following vascular surgery. Quantitative measurement (amplitude and frequencies) of skin displacement caused by fine tremors in underlying muscles can have application in studies of neurological disorders, such as Parkinson’s disease and traumatic brain injury.

There are many types of optical interferometers that can be used to make the measurements described earlier; most use a single coherent light source (i.e., a laser) and split the coherent beam into two optical paths, R and M , then combine the beams in such a manner that there is a linear summation of the output beam E vectors (E_M and E_R) at a common output plane or photosensor. At the output plane, there is alternate *constructive* and *destructive interference*. There are bright rings or lines where there is constructive addition of E vectors; dark rings or lines appear where there is destructive (subtractive) interference. In order to measure a physical quantity such as a small surface displacement, the displacement must cause a phase change between E_M and E_R by changing the M -beam’s lightwave travel time. The wave travel time (or phase) can be altered by changing the length of the M path or passing the M ray through a medium having an index of refraction, $n = c/v$, different from the R path.

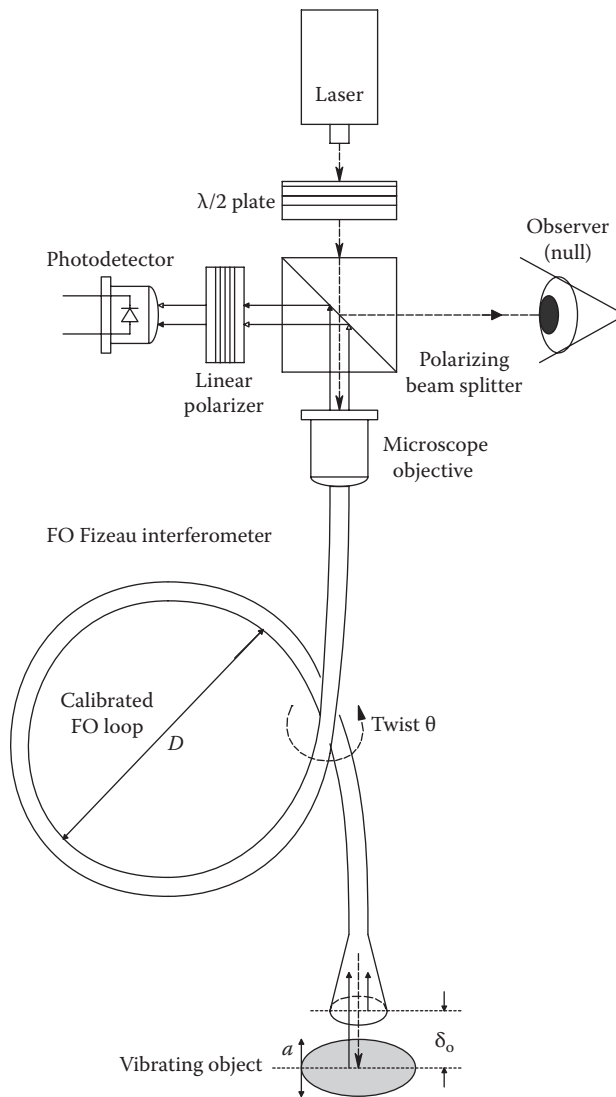


FIGURE 7.46
Diagram of a Fizeau interferometer used to measure small $a(t)$.

As you will see, a convenient way of changing path length is to stick a mirror on the surface under measurement and to use the phase change in the reflected light in the M path when the surface moves. Other physical quantities can be measured by interferometry when they cause a phase change between the M and R beams. For example, when FOs are used for the R and M paths, differential temperature changes and mechanical strain or bending of the M fiber can produce phase changes. For an elementary discussion of the use of interferometry in length or displacement measurement, see Chapter 4 and others in the text by Sirohi and Kothiyal (1991). Detailed mathematical analyses of several major types of interferometer can be found in Chapter 9 in Hecht's *Optics* (1987).

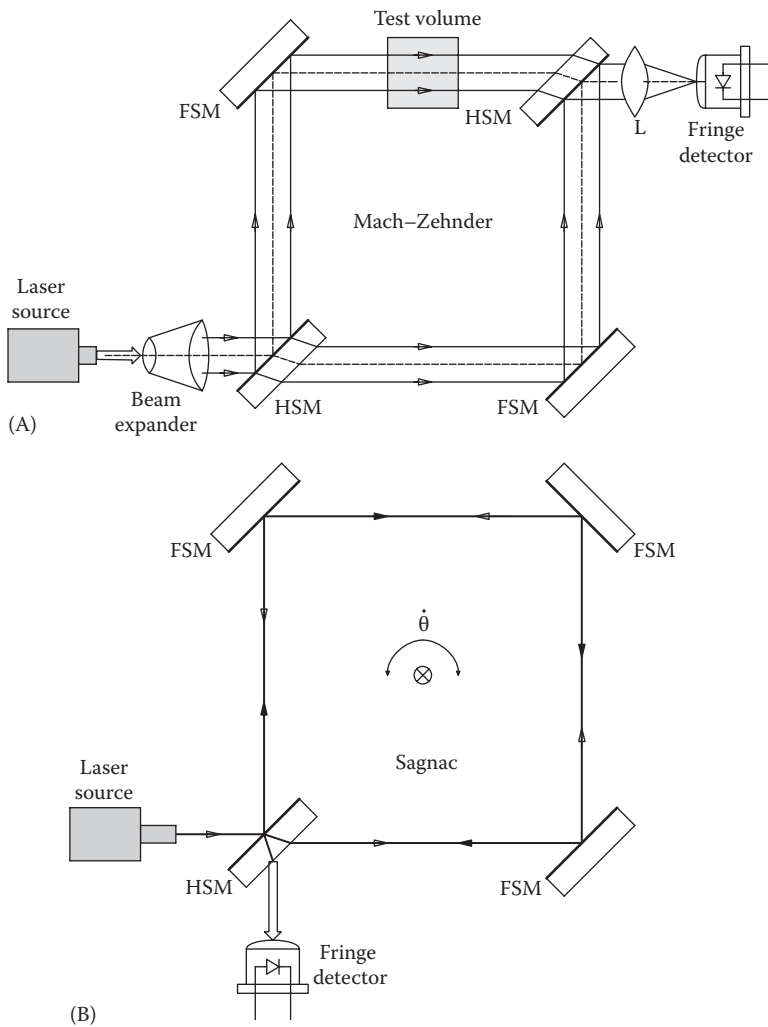


FIGURE 7.47
 (A) Mach-Zehnder interferometer. (B) Sagnac interferometer.

Direct measurement of an object's displacement with an interferometer is limited to motions of less than $\pm\lambda/4$. Greater motion will result in a periodic output of the interferometer's photodetector, and some means must be used to count interference *fringes* during the motion of the object to keep track of the total object displacement.

Before describing some specific types of interferometers, let us examine a general, heuristic treatment of optical interference. For interference to occur, the EM lightwave must be temporally and spatially *coherent*. That is, the frequency and phase of the E vectors of the interfering waves must remain constant in time and space for stable interference to occur. A stable, monochromatic source such as a laser is generally used for interferometry. The optical intensity changes associated with interference occur as the result of the superposition of the E vectors at a point on the detector. The intensity at the detector is proportional to the net E vector squared. We will assume that the R and M rays are linearly polarized

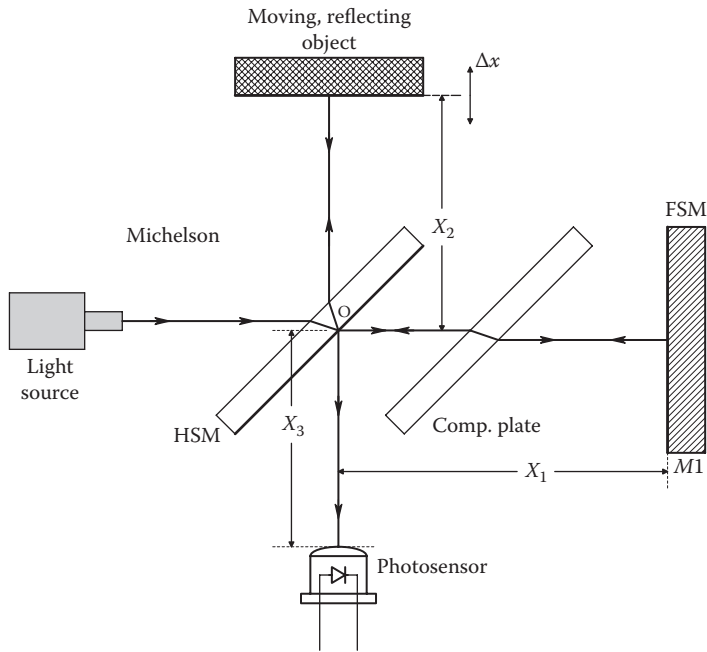


FIGURE 7.48
 Traditional Michelson interferometer. HSM = half-silvered mirror. FSM = front surface mirror.

in the same direction, for example, along the x -axis. The E vector amplitudes vary in time and space. At a point on the output plane of the generalized interferometer, the x -directed E vector amplitudes can be written as

$$e_r = e_r(t) = E_{R0} \cos[(2\pi\nu)t] \tag{7.142}$$

$$e_m = e_m(t) = E_{M0} \cos[(2\pi\nu)t + \phi_m + k\Delta x], \quad k \equiv \frac{2\pi}{\lambda} = \frac{2\pi\nu}{c} \text{ r/m} \tag{7.143}$$

where

- $\nu = c/\lambda$ is the EM lightwave Hz frequency
- $\nu = 4.74 \times 10^{14}$ Hz for a 633 nm HeNe laser

Now, the net intensity at the detector is proportional to the E vector sum squared:

$$I_d \propto (e_r + e_m)^2 \text{ W/m}^2. \tag{7.144}$$

e_m and e_r are the time-varying E vectors of EM plane waves at a point P at the detector surface. The fixed phase of e_r is taken as zero; the phase of the measurement wave, e_m , has a fixed phase lag component, ϕ_m , due to the extra propagation delay, and a variable phase component, $k\Delta x$, due to a small, relative displacement of the reflecting surface, Δx , that

we are trying to measure. Note that the frequency of both waves is $\nu = c/\lambda$ Hz. Substituting Equations 7.158 and 7.159 into Equation 7.160, we find

$$I_d \propto \left(\frac{E_{R0}^2}{2} \right) \{1 + \cos[2\pi(2\nu)t]\} + \left(\frac{E_{M0}^2}{2} \right) \{1 + \cos[2\pi(2\nu)t + 2\phi_m + 2k\Delta x]\} + E_{R0}E_{M0} \{ \cos[2\pi(2\nu)t + \phi_m + k\Delta x] + \cos[\phi_m + k\Delta x] \} \text{ W/cm}^2. \quad (7.145)$$

Of course, the photodetector does not respond to terms at lightwave frequencies ν and 2ν Hz. Thus, the photodetector output is given by the time average of $(e_r + e_m)^2$ times a constant:

$$V_o = K_d \bar{I}_d = K_d \left\{ \left(\frac{E_{R0}^2}{2} \right) + \left(\frac{E_{M0}^2}{2} \right) + E_{R0}E_{M0} \cos[\phi_m + k\Delta x] \right\} \text{ V}. \quad (7.146)$$

We now use the trig. identity, $\cos(A + B) \equiv [\cos(A)\cos(B) - \sin(A)\sin(B)]$, where $A = \phi_m$ and $B = k\Delta x$ radians. If Δx is slowly varying (i.e., if it is a DC quantity), we can force $\phi_m = q\pi/2$ ($q = 1, 5, 9, \dots$) so that $\cos(\phi_m) = 0$ and $\sin(\phi_m) = +1$. Thus, the detector output is given by

$$V_o = K_d \left\{ \left(\frac{E_{R0}^2}{2} \right) + \left(\frac{E_{M0}^2}{2} \right) + E_{R0}E_{M0} \sin[k\Delta x] \right\} \text{ V}. \quad (7.147)$$

From Equation 7.163, we see that the photodetector output contains DC terms and a sinusoidal term that has sinusoidal peaks that add to give V_o maxima for $q\lambda/4$, where $q = \dots, -15, -11, -7, -3, 1, 5, 9, 13, \dots$, and minima for V_o at $q = \dots, -13, -9, -5, -1, 3, 7, 11, 15, \dots$. If the net displacement of the object exceeds $\pm\lambda/2$, it is necessary to count maxima and minima in V_o to find the total displacement of the object and to determine whether the displacement is toward or away from the interferometer. By placing a second (quadrature) photodetector next to the first and delaying the phase of the output light ray to it by $\pi/2$ radians, we can write the second PD output voltage as

$$V'_o = K_d \left\{ \left(\frac{E_{R0}^2}{2} \right) + \left(\frac{E_{M0}^2}{2} \right) + E_{R0}E_{M0} \cos[k\Delta x] \right\} \text{ V}. \quad (7.148)$$

The quadrature detector output, V'_o , has its maxima for $p\lambda/4$, $|p| = 0, 4, 8, 12, \dots$ V'_o 's minima are at $|p| = 3, 7, 11, \dots$. Electronic threshold logic can detect the maxima and minima of V_o and V'_o and determine from their phase relation in time whether the object is approaching or receding from the interferometer.

If the object displacement, Δx , is moving sinusoidally at an audio frequency so that $\Delta x = a_o \sin(\omega_m t)$, the DC terms are filtered out by high-pass filtering, and again, ϕ_m is set equal to $q\pi/2$, $q = 1, 5, 9, 13, \dots$, there results

$$v_o(t) = K_d E_{R0} E_{M0} \sin[ka_o \sin(\omega_m t)] \text{ V}. \quad (7.149)$$

To keep the system in its linear output range, the argument of the $\sin[*]$ term must not exceed $\pm\lambda/4$. That is, a_o must be $< \lambda/4$. If $a_o > \lambda/4$, periodicity in the output vs. a_o is observed, and unless one counts successive output maxima and minima, there will be ambiguity in the true value of a_o . Equation 7.165 can be expanded by the Bessel–Jacobi identity (Stark et al. 1988):

$$v_o(t) = K_d E_{R_o} E_{M_o} \sum_{n=-\infty}^{\infty} J_n(\xi) \sin(n\omega_m t) \quad \text{V} \tag{7.150}$$

where, clearly, $\xi \equiv ka_o$ and $J_n(\xi)$ is a *Bessel function of the first kind*. It can be shown that $J_n(\xi)$ can be approximated by the series (Stark et al. 1988):

$$J_n(\xi) = \sum_{k=0}^{\infty} \frac{(\xi/2)^{(n+2k)(-1)^k}}{(n+k)!k!} \tag{7.151}$$

If $\xi \ll 1$, Equation 7.167 can be approximated by

$$J_n(\xi) \cong \frac{(\xi/2)^n}{n!} \quad n > 0, \text{ integer} \tag{7.152}$$

and in general (Dwight 807.4 1969),

$$J_{-n}(\xi) \equiv (-1)^n J_n(\xi). \tag{7.153}$$

Thus, the first few terms of Equation 7.167 can be written, assuming $\xi \ll 1$, as

$$v_o(t) \cong \sum_{n=0}^{\infty} K_d E_{R_o} E_{M_o} \left\{ \xi \sin(\omega_m t) + 0 + \left(\frac{\xi^3}{24}\right) \sin(3\omega_m t) + 0 + \left(\frac{\xi^5}{1920}\right) \sin(5\omega_m t) + 0 + \dots \right\} \tag{7.154}$$

Because even high-order harmonic terms are zero, and $\xi \ll 1$, the fundamental frequency term dominates the series of Equation 7.167. a_o , the QUM, can be found by phase-sensitive rectification of $v_o(t)$ using a ω_m -frequency reference signal, followed by low-pass filtering (i.e., by using an LIA). If the LPF has gain K_F , then its DC output, \bar{V}_o , will be

$$\bar{V}_o = K_F K_d E_{R_o} E_{M_o} k a_o \text{ DC V.} \tag{7.155}$$

We next examine two examples of optical interferometers that have been used in biomedical applications. The first is the fiber-optic, *Fizeau interferometer* developed by Drake and Leiner (1984).

7.3.4.2 Measurement of Tympanic Membrane Displacement by Fiber-Optic Fizeau Interferometer

This prototype instrument, shown in Figure 7.38, was used to sense the minute displacement of the tympanic membrane of a common cricket subject to external audible sound. (Also see Figure 7.46.) Drake and Leiner used a 15 mW HeNe (633 nm) laser as their source. The half-wave plate was rotated to minimize the observed intensity, indicating that all of the LPL from the laser was going through the polarizing beam splitter into the 10×0.255 NA microscope objective used to direct all of the incident LPL into the proximal end of the polarization-preserving, glass optical fiber (PPOF). The PPOF was of two-step design; it had a $5 \mu\text{m}$ core diameter and a $125 \mu\text{m}$ cladding diameter. (Note that the expanded distal end of the OF is drawn for clarity to illustrate the fate of rays at the fiber's end; it is not really there.) By experimentally adjusting the PPOF loop diameter, D , and twisting it some small angle, θ , around the axis from the objective to the object, the PPOF loop could be made to act as a *quarter-wave plate*. Thus, the light returning to the proximal end of the PPOF was rotated 90° with respect to the entering light. Drake and Leiner stated:

All of the interfering light is then reflected toward the detector by the polarizing beam-splitter, while light reflected from the back side of the beam-splitter, from the elements of the microscope objective, and from the proximal end of the fiber, all pass straight back through the polarizing beam splitter and does not contribute unwanted radiation at the detector.

Because the Fizeau interferometer gives an output signal proportional to *relative object displacement*, Drake and Leiner mounted their object (a cricket tympanum) on a piezoelectric crystal that could be displaced toward or away from the PPOF's distal end by $>\pm\lambda/2$ by applying a 10 Hz, triangular, voltage waveform and observing the magnitude of the AC signal output at the acoustic stimulus frequency, f_m . The AC output was maximum when the distance, a , was some multiple of $\lambda/4$. This particular DC voltage was applied to the crystal for subsequent tests of tympanic membrane displacement response, Δx . Calibration of the Fizeau interferometer was accomplished by applying a 1 kHz triangular wave around the DC bias voltage to the piezocrystal. The triangle wave amplitude was adjusted until the output peaks just began to fold over as seen on an oscilloscope display. The target was then known to be undergoing a $\pm\lambda/4$ displacement. Thus, an output calibration factor was determined in V/nm displacement ($\lambda/4 = 158.2 \text{ nm}$).

Drake and Leiner displayed their output signal on an oscilloscope and on a spectrum analyzer. The typical ratio of fundamental frequency to second harmonic amplitude was 0.16. They found that their instrument could resolve $0.01 \text{ nm} < \Delta x_0 < 30 \text{ nm}$ linearly. This is an amazing sensitivity considering that an LIA was not used. They commented that the working distance, a , from the distal end of the fiber to the moving object can be increased to over a few mm by adding a Selfocâ lens to the tip of the PPOF cable.

Drake and Leiner mentioned that future applications of their Fizeau interferometer can include in vivo studies of human eardrum displacement, measurement of the basilar membrane deflection in the cochlea (presumably done in vitro), measurement of nerve axon displacement during an action potential, and dental strain measurements.

Kaiser et al. (2001) reported the RMS resolution of their improved, phase-shifting, Fizeau interferometer to be ca. $\lambda/10^3$ or 0.6 nm. Their Fizeau interferometer altered the lightwave frequency of their LAD source to obtain a fine null.

7.3.4.3 Measurement of Skin Vibration by Optical Interferometry

The *Michelson interferometer* has been used to measure skin surface microvibrations (Hong and Fox 1993, Hong 1994). A traditional Michelson interferometer is illustrated in Figure 7.48 (Hecht 1987). Note that it uses a compensating plate in optical path X_1 so that each beam travels the same distance through glass. The compensating plate is identical to the HSM beam splitter, except that it does not have silvering. The use of the compensation plate corrects for refractive index dispersion with λ and lets the Michelson interferometer be used with broadband (semicoherent) light sources. If a monochromatic, laser source is used, the compensation plate is not necessary. A Michelson interferometer without the compensation plate is often called a *Twyman–Green interferometer* (Sirohi and Kothiyal 1991). As we showed earlier with the Fizeau interferometer, the optical intensity at the photodetector is proportional to the square of the sum of the two E vectors impinging on the detector. Let us take point **O** on the HSM as the phase origin. Neglecting the HSM thickness, we can write for the Twyman–Green/Michelson interferometer the following:

$$e_r(t) = E_{R0} \sin[2\pi\nu t + k(2X_1 + X_3)] \quad (7.156)$$

$$e_m(t) = \eta E_{R0} \sin[2\pi\nu t + k(2X_2 + X_3)] \quad (7.157)$$

where

the lightwave frequency $\nu = c/\lambda$ Hz (4.74×10^{14} Hz for a HeNe laser)

$k \equiv 2\pi/\lambda$, the distances are defined in Figure 7.48

$\eta < 1$ represents the fraction of incident light reflected back to the detector

We know that the intensity at the detector is proportional to $[e_r(t) + e_m(t)]^2$:

$$I_d \propto \left(\frac{E_{R0}^2}{2} \right) \{1 - \cos[2\pi(2\nu)t + k(2X_1 + X_3)]\} + \eta^2 \left(\frac{E_{R0}^2}{2} \right) \{1 - \cos[2\pi(2\nu)t + k(2X_2 + X_3)]\} \\ + 2(E_{R0}\eta E_{R0}) \sin[2\pi\nu t + k(2X_1 + X_3)] \sin[2\pi\nu t + k(2X_2 + X_3)]. \quad (7.158)$$

Because the detector does not respond to the light frequency terms, Equation 7.174 reduces to

$$I_d \propto \left(\frac{E_{R0}^2}{2} \right) (1 + \eta^2) + \eta E_{R0}^2 \cos[2k(X_1 - X_2)]. \quad (7.159)$$

Using the $\cos(A - B) \equiv [\cos(A)\cos(B) + \sin(A)\sin(B)]$ trig identity, we can write

$$I_d \propto \left(\frac{E_{R0}^2}{2} \right) (1 + \eta^2) + \eta E_{R0}^2 \{ \cos[2kX_1] \cos[2kX_2] + \sin[2kX_1] \sin[2kX_2] \}. \quad (7.160)$$

Now we adjust X_1 so that $\cos[2kX_1] \rightarrow 0$ and $\sin[2kX_1] = 1$. We can now consider a small sinusoidal skin vibration at frequency f_s and peak amplitude, $\Delta x_o < \lambda/4$:

$$X_2(t) = X_{2o} + \Delta x_o \sin(2\pi f_s t). \quad (7.161)$$

Thus,

$$I_d \propto \left(\frac{E_{R0}^2}{2} \right) (1 + \eta^2) + \eta E_{R0}^2 \sin[2kX_{20} + 2k\Delta x_o \sin(2\pi f_s t)]. \quad (7.162)$$

Using the $\sin(A + B) \equiv [\sin(A)\cos(B) + \cos(A)\sin(B)]$ trig identity, we can finally write

$$I_d \propto \left(\frac{E_{R0}^2}{2} \right) + \eta^2 \left(\frac{E_{R0}^2}{2} \right) + \eta E_{R0}^2 \{ \sin[2kX_{20}] \cos[(2k\Delta x_o) \sin(2\pi f_s t)] + \cos(2kX_{20}) \sin[(2k\Delta x_o) \sin(2\pi f_s t)] \}. \quad (7.163)$$

Again, the $\cos[\xi \sin(2\pi f_s t)]$ and $\sin[\xi \sin(2\pi f_s t)]$ terms can be represented as Bessel functions (Stark et al. 1988), thus,

$$\cos[\xi \sin(2\pi f_s t)] = \sum_{n=-\infty}^{\infty} J_n(\xi) \cos(n2\pi f_s t) \quad (7.164)$$

$$\sin[\xi \sin(2\pi f_s t)] = \sum_{n=-\infty}^{\infty} J_n(\xi) \sin(n2\pi f_s t) \quad (7.165)$$

where $\xi \equiv (2k\Delta x_o)$. Thus, the electrical output of the detector contains DC terms, a fundamental frequency term (at f_s), and harmonics (at nf_s , $n \geq 2$). The DC energy can be eliminated by an HPF, and the fundamental frequency terms can be selected by a BPF or by the use of an LIA. Thus, the output of the Michelson interferometer is nonlinear, even with a reflecting object's small motion.

If the object motion is fast enough, there will also be a significant *Doppler frequency shift* associated with the lightwave frequency of $e_m(t)$. This shift can be expressed as a change in frequency of the e_m **E** vector:

$$e_m(t) = \eta E_{R0} \sin[2\pi\nu_m t + k(2X_2 + X_3)] \quad (7.166)$$

and

$$e_r(t) = E_{R0} \sin[2\pi\nu t + k(2X_1 + X_3)] \quad (7.167)$$

where $\nu_m \equiv \nu + \Delta\nu$. The *Doppler shift* is $\Delta\nu \equiv \nu(2\dot{X}_2/c)$ Hz (Northrop 2002). Now, ignoring light frequency terms, the intensity is of the form

$$I_d \propto \left(\frac{E_{R0}^2}{2} \right) (1 + \eta^2) + 2\eta E_{R0}^2 \cos(A)\cos(B) \quad (7.168)$$

where

$$A \equiv [2\pi\nu t + k(2X_1 + X_3)]$$

$$B \equiv [2\pi\nu(1 + 2\dot{X}_2/c)t + k(2X_2 + X_3)]$$

Using the identity, $\cos(A) \cos(B) \equiv (1/2) [\cos(A - B) + \cos(A + B)]$, we have

$$I_d \propto \left(\frac{E_{R0}^2}{2} \right) \left(1 + \eta^2 \right) + \eta E_{R0}^2 \left\{ \cos[(2k\dot{X}_2)t + (2k)(X_2 - X_1)] \right. \\ \left. + \cos[2\pi(2\nu)t + (2k\dot{X}_2)t + (2k)(X_1 + X_2 + X_3)] \right\}. \quad (7.169)$$

The second term in the brackets of Equation 7.185 is dropped because it is twice the light-wave frequency. The first term is at the Doppler shift frequency, $\Delta\nu = \nu(2\dot{X}_2/c)$ Hz, which is generally in the audio range and resolvable by the photosensor. We can now use the trig. identity, $\cos(C + D) \equiv [\cos(C)\cos(D) - \sin(C)\sin(D)]$, to find

$$I_d \propto \left(\frac{E_{R0}^2}{2} \right) \left(1 + \eta^2 \right) + \eta E_{R0}^2 \left\{ \cos[(2\pi\nu)(2\dot{X}_2/c)t] \cos[(2k)(X_2 - X_1)] \right. \\ \left. - \sin[(2\pi\nu)(2\dot{X}_2/c)t] \sin[(2k)(X_2 - X_1)] \right\}. \quad (7.170)$$

If the skin motion is sinusoidal so that

$$X_2(t) = X_{20} + a_o \sin(\omega_s t) \quad (7.171)$$

then

$$\dot{X}_2 = \omega_m a_o \cos(\omega_s t). \quad (7.172)$$

We set $X_1 = X_{20}$, and after some algebra, we get the messy result:

$$I_d \propto \left(\frac{E_{R0}^2}{2} \right) \left(1 + \eta^2 \right) + \eta E_{R0}^2 \left\{ \cos[(4\pi\delta_2/\lambda)(\omega_s \cos(\omega_s t))t] [\cos(2k a_o^\xi) \sin(\omega_s t)] \right. \\ \left. - \sin[(4\pi\delta_2/\lambda)(\omega_s \cos(\omega_s t))t] \sin[(2k a_o^\xi) \sin(\omega_s t)] \right\} \quad (7.173)$$

The $\cos[\xi \sin(\omega_s t)]$ and $\sin[\xi \sin(\omega_s t)]$ terms in Equation 7.189 give rise to the motion-frequency Bessel function series of Equations 7.180 and 7.181. It is not known what sort of temporal modulation of the Bessel series the Doppler $\cos[\xi \omega_s t \cos(\omega_s t)]$ and $\sin[\xi \omega_s t \cos(\omega_s t)]$ terms will provide. For example, let us calculate the maximum Doppler shift for a peak sinusoidal skin deflection of 120 nm at 1 Hz, using a HeNe laser with $\lambda = 633$ nm. From Equation 7.188, $\Delta\nu_{\max} = (k\omega_s a_o)$ Hz. So

$$\Delta\nu_{\max} = [2 \times 2\pi(1) \times 1.2 \times 10^{-7} / (633 \times 10^{-9})] = 2.38 \text{ Hz}. \quad (7.174)$$

Thus, the peak Doppler shift was about three times the 1 Hz displacement frequency.

Hong and Fox (1993) reported on an *optical stethoscope* that used a 670 nm diode laser as a coherent light source with a Michelson (actually, a Twyman–Green) interferometer to sense skin vibrations. Their interferometer was housed in a metal box about $5'' \times 3.5'' \times 3.5''$; its AC-coupled photodiode output was observed in real time on a digital oscilloscope with FFT spectrum analysis capability. A more refined version of this instrument was described by Hong (1994), in which he used a HeNe laser source and an FOC to impinge the light on the skin. In most of the phantoms and human subjects studied by Hong, object displacements were periodic at the heart rate, but displacements were in *hundreds of μm* rather than hundreds of nm! Thus, the Twyman–Green interferometer was well out of its $\pm\lambda/4$ linear range, and the apparatus evidently served as a *mixer* to extract the lightwave Doppler shift frequency in the measurement beam. The Doppler frequency shifts extracted followed the shape of the *magnitude* of the time derivative of the pressure waveform in the phantom studies, as evidenced by the output of a frequency-to-voltage converter. (Apparently, the *skin* surface displacement of the phantom was proportional to the vessel pressure.) In human studies, Hong cited the peak displacement of the skin over the common carotid artery to be $510 \mu\text{m}$ over 86 ms. This is a velocity of $5.93 \times 10^{-3} \text{ m/s}$, and the peak Doppler lightwave frequency is $\Delta\nu = 2 \times 5.93 \times 10^{-3} / 633 \times 10^{-9} = 18.74 \text{ kHz}$. Obviously, this output frequency will vary over the cardiac cycle. Because there is no way to sense the sign of $\Delta\nu$, the VFC V_o was determined by

$$V_o \propto \Delta n \propto |L_2| \propto |\dot{p}|. \quad (7.175)$$

Thus, in summary, Hong's interferometer was apparently used as a Doppler mixer, not a true, nm distance-measuring system.

7.3.4.4 Two-Frequency, Heterodyne Interferometry

Figure 7.49 illustrates one form of a two-frequency interferometer. A special laser produces two optical frequencies a few MHz apart (a fractional separation of $10^7/10^{14} = 10^{-7}$). The EM waves at the two frequencies, ν_1 and ν_2 , are linearly polarized at a right angle (orthogonally polarized). Both waves are partially reflected at beam splitter BS1 and sent to polarizer P1 that has its axis at 45° to both waves. The output of P1 is LPL with its E vector directed along P1's pass axis. Thus, LPL that is the sum of the beams at frequencies ν_1 and ν_2 strikes the detector P1's photocathode. P1's output current, as we have shown earlier, is proportional to the average of the square of the sum of the E vectors of the laser's two output EM waves. The detector P1 cannot respond to terms at the lightwave frequencies, ν_1 , ν_2 , $2\nu_1$, $2\nu_2$, $(\nu_2 + \nu_1)$, but only responds to DC intensity terms and radio-frequency AC term at frequency $(\nu_2 - \nu_1)$, which is in the MHz range. Hence, the lower counter, when gated, counts pulses at the fixed frequency $(\nu_2 - \nu_1)$. The horizontal, two-component beam from BS1 passes through a polarizing beam splitter, PBS2. The LPL beam of frequency ν_1 is directed to the movable retroreflector prism, RR1, where it is returned to PBS2 with a Doppler-shifted frequency of

$$\nu_{1r} = \nu_1 + \Delta\nu_d = \nu_1 \left(\frac{1 + 2\delta\dot{x}}{c} \right) \text{Hz}. \quad (7.176)$$

The orthogonally polarized beam of fixed frequency, ν_2 , is sent to the fixed retroreflector, RR2, where it is returned to PBS2. Both of these orthogonally polarized beams are directed to a full reflecting mirror, M , and thence to polarizer P2 that also is at 45° with respect to

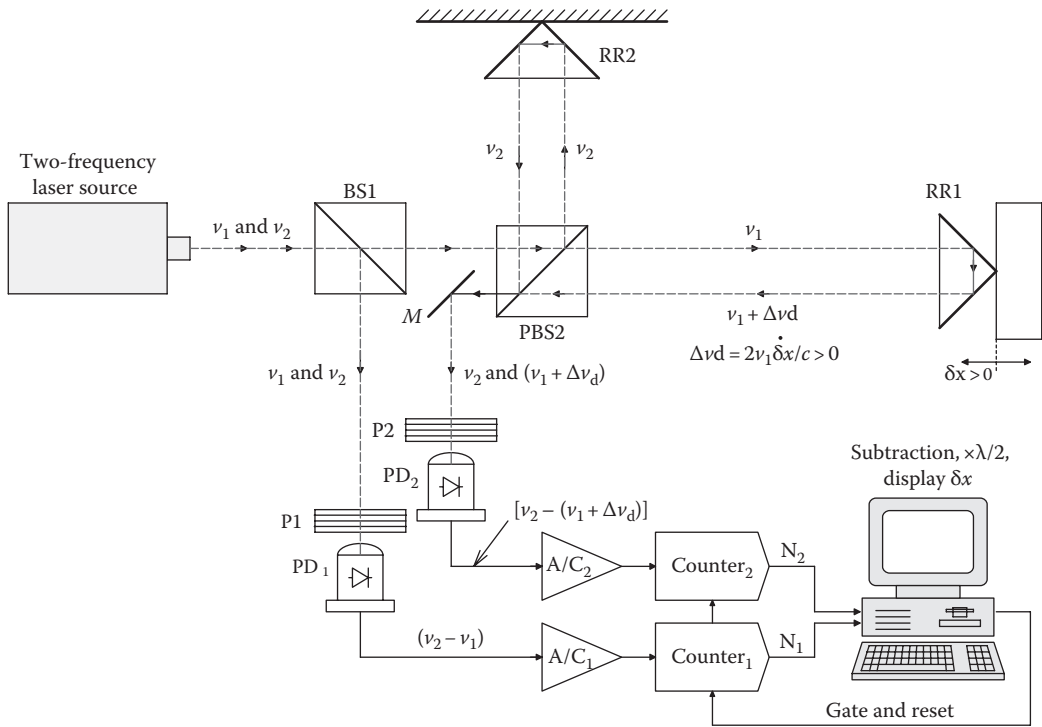


FIGURE 7.49
Heterodyne interferometer uses two light frequencies, ν_1 and ν_2 .

the E vectors of the two impinging beams. The output light from P2 is LPL with its E vector along P2's pass axis. Photodetector PD2 also responds to the average of the square of the sum of the two E vector components striking its photocathode, that is, DC light intensity components and a low-frequency, AC term at frequency $[\nu_2 - (\nu_1 + \Delta\nu_d)]$. Only the AC terms at the PD outputs are of interest; the DC is blocked.

The amplifier/comparators (A/C₁, A/C₂) produce TTL output waves of frequencies $(\nu_2 - \nu_1)$ and $[(\nu_2 - \nu_1) - \Delta\nu_d]$, respectively. Assume the counters are reset and gated to count for ΔT seconds. The count $N_1 = \Delta T(\nu_2 - \nu_1)$ cycles. Similarly, $N_2 = \Delta T[(\nu_2 - \nu_1) - \Delta\nu_d]$. Assuming the velocity of RR1 is constant over ΔT , the computer subtracts N_2 from N_1 and finds $\Delta N = \Delta T\Delta\nu_d = \Delta T(2\nu_1\delta\dot{x}/c)$ cycles. Thus, the velocity, $\delta\dot{x}$, can be written as

$$\delta\dot{x} = \frac{\Delta Nc}{\Delta T2\nu_1} = \frac{\Delta N}{\Delta T} \left(\frac{\lambda_1}{2} \right). \tag{7.177}$$

If ΔN is calculated periodically with period $T_c > \Delta T$, we can generate a series of velocity samples, $\delta\dot{x}(nT_c)$, which can be integrated to find X_o , the total RR1 displacement. Thus, with the use of DSP, this interferometer can measure both target velocity *and* the distance traveled by a cooperative target.

Essential to the operation of this interferometer is the generation of two coherent rays of frequencies ν_1 and ν_2 such that $\Delta\nu = (\nu_2 - \nu_1)$ is on the order of 10s of kHz to 10s of MHz.

Phase-locked, coherent frequency shifting of the LPL rays can be done in several ways (Sirohi and Kothiyal 1991). Quoting these authors:

One technique makes use of rotating polarization components, namely quarter- and half-wave plates and [a] polarizer. For example, if circularly polarized light is transmitted through a rotating half-wave plate, the orthogonal components of the transmitted light show a frequency difference equal to twice the rotation rate of the plate. The principal difference in this method is that for reasonable rotation rates it is not possible to obtain a frequency shift $[\Delta\nu]$ much larger than a few kHz.

A moving diffraction grating can also be used as a frequency shifter, since it shifts the N th diffracted order by an amount Nvf , where v is the velocity component of the grating perpendicular to the grating lines [m/sec] and f is the spatial frequency of the grating [lines/meter]. In other words, the frequency shift is equal to N times the number of grating lines that pass a given point per second. For a [moving] grating, the frequency shift is independent of the wavelength. To obtain a large frequency shift in the range of several kHz, a circular grating is useful.

An acoustooptic Bragg cell shifts the frequency of diffracted light in the same way as a moving diffraction grating. In this case the traveling acoustic wave serves as the moving grating. The frequency shift of the first diffracted order is equal to the frequency used to drive the Bragg cell independent of the wavelength of [the input] light. A Bragg cell gives a frequency shift in the range of several MHz.

Some lasers can be designed to give two frequency outputs, for example, the Zeeman split HeNe laser.

Indeed, the frequency shift in a *Bragg cell* can be written as

$$\Delta\nu = \frac{2v_s}{\lambda_s} = f_s \text{ Hz.} \quad (7.178)$$

where

v_s is the velocity of sound in the Bragg medium

λ_s is the wavelength of the sound waves in the Bragg medium

Electrooptic modulation and diffraction of coherent light by Bragg cells are described in great detail in Ch. 12 of Yarov (1991).

7.3.4.5 Fabry–Pérot Interferometer

Figure 7.45 illustrates one form of the Fabry–Pérot interferometer (FPI). This optical measurement system has several unique features. It uses a diffuse, distributed-parameter light source. (Such a source can be made by expanding a laser beam with a telescope to about a 2 cm diameter disk directed on the back of a ground glass sheet.) The ground glass then acts as a secondary source of very many rays at many angles. In Figure 7.45, we trace one ray from point S_1 on the source, through a convex lens L_1 , thence to the *etalon*. The light ray reflects back and forth many times on the inside of the half-silvered, parallel plates of the etalon. A small fraction of the internally reflected rays are refracted and pass out of the etalon's plates. The emergent rays from the right surface of the etalon are collimated to a focus at point P_1 . Using a complex optical superposition, it is possible to derive an expression for the intensity of the concentric circular interference rings of the FPI on its viewing screen (cf. Hecht 1987, Sect. 9.6.1, Vaughan and Vaughan 1989, Betzler 2002). However, we will not repeat this derivation here.

Two things are unique about the F–P diffraction rings:

1. Their radial separation *decreases* with their radii.
2. The rings are very sharp, unlike the diffraction rings of a Michelson interferometer, in which the intensity varies roughly sinusoidally with ring radius.

The etalon diffraction cavity supports many internal reflections from each ray directed to it from lens L_1 . The parallel internal faces of the etalon's glass plates are vapor coated with a thin, partially reflecting film of metal that produce intense internal reflections. When a light ray *inside the etalon* encounters the metal film, a large fraction of its intensity, R , is reflected internally; a very small fraction, A , is absorbed; and a small fraction, T , is transmitted outside the etalon cavity. Hence, $1 \equiv R + T$, $R \gg T > A$.

Betzler (2002) defined a *quality factor* parameter, Q_R , for the FPI:

$$Q_R \equiv \frac{4R}{(1-R)^2} \dagger \quad (7.179)$$

Betzler showed, using superposition of reflected waves, that the ratio of output (transmitted) to input (incident) intensity of the etalon is given by

$$\frac{I_t}{I_i} = \frac{1}{1 + Q_R \sin^2(\delta/2)} = \frac{1}{1 + \frac{4R}{(1-R)^2} \sin^2(\delta/2)} \quad (7.180)$$

where

Q_R is defined in relation (7.195)

R is the incident intensity fraction reflected internally

δ is the phase difference in radians between each succeeding reflection in the etalon

The *finesse*, F , of an FPI is defined as the ratio of the peak-to-peak distance ($\Delta\lambda$) (to the next adjacent peak) of a pair of diffraction rings to the peak's half-intensity width ($\delta\lambda$) in terms phase in radians. To derive F , we examine the points where the intensity transfer function is equal to 0.5. From relation (7.196),

$$\frac{1}{1 + Q_R \sin^2(\delta/2)} = 0.5, \dagger \text{ so } Q_R \sin^2(\delta/2) = 1, \text{ and for } |d - 2N\pi| \ll 1, \text{ we have } \delta = \pm 2/\sqrt{Q_R}.$$

The phase difference between two adjacent diffraction peaks is 2π radians; thus, the finesse F is shown to be (Betzler 2002, Eq. 19)

$$F = \frac{\pi}{2} \sqrt{Q_R} = \frac{\pi\sqrt{R}}{1-R}. \quad (7.181)$$

The larger the internal reflectance fraction R , the larger the finesse F , the sharper the output peaks, and the higher the resolving power of the FPI. The finesse typically can range from about 20–200 in a basic, single-pass, FPIs. F can be measured from the diffraction pattern

as the wavelength separation between output peaks, $\Delta\lambda$, divided by the wavelength width ($\delta\lambda$) at half peak intensity of a peak. That is, $F = \Delta\lambda/\delta\lambda$.

Hecht (1987) showed that each narrow, bright, F-P interference ring has half its peak intensity for a phase difference of γ radians. Hecht defined the *fineness* by $F \equiv 2\pi/\gamma$. The half width of a peak can be shown to be $\gamma \cong 4/\sqrt{F}$ r, where $F = 4R/(1 - R)^2$ (the same as Q_R defined by Betzler). Thus, $F = \pi\sqrt{R}/(1 - R)$, as in Equation 7.197.

The usual mode of operation of an FPI is to observe fringe movements as the etalon plate separation is varied by Δd by moving, say, the left-hand plate. By detecting fringe movements optically, it is possible to drive a piezomotion generator connected to the right-hand plate to renull the fringes. Since the piezodisplacement to renull the fringe pattern is $\Delta d' = V_c K_p = \Delta d$, we can measure V_c at null to find Δd .

FPIs have been used to measure a diversity of measurands, including atmospheric CO_2 , magnetic fields, small (nm) displacements, large (mm) displacements, high-pressure transients, and underwater sound. They also can be applied as dichroic filters and, in telecommunications networks, as filters in wavelength division multiplexing, optical wavemeters, and laser resonators and in astronomy, an F-P etalon is used to select a single atomic transition for imaging. They can also serve effectively as photon delay lines.

7.3.4.6 Discussion

Interferometers are generally suited for precise direct measurement of surface displacements when the range of deflection is $<\lambda/4$ meters. When the surface moves at an HF, the interferometer output contains many harmonics due to the $\cos[\xi\sin(\omega_s t)]$ terms and light-wave Doppler shift terms. Most interferometers are better suited to measuring small, slow displacements, such as tooth drift in orthodontic research or plant growth. Heterodyne interferometers using two lightwave frequencies are suitable for measuring object velocities and displacements.

When a single-wavelength interferometer is used to measure a distance that is a large multiple of $\lambda/4$, an additional quadrature detector is used. The sign of Δx is found by examining the phase of the fringes in the direct and quadrature detector outputs as the object moves in time. The total number of fringes gives the distance to the nearest λ , and interpolation of the residual distance can be done on $\Delta V_o = (V_o - V'_o)$ from Equations 7.163 and 7.164.

FPIs offer the advantage generating sharp diffraction rings; hence, it is easier to detect object motion-caused fringe shift. FPI precision can range from 1:1000 to over 1:10⁵, depending on the design and signal processing used. Accuracy is on the order of $\lambda/2000$. Sources of noise in interferometry include mechanical vibrations, changes in air refractive index due to temperature, barometric pressure and humidity changes, electronic noise in photo-sensors, and drift in source laser amplitude, frequency, and polarization.

7.3.5 Phase-Lock Velocity and Range Sensing with Ultrasound or EM Radiation

In this section, we will examine a class of closed-loop measurement system that gives simultaneous analog output voltages proportional to target range and velocity along the range vector. Recall that conventional radar works by measuring the time it takes a transmitted pulse to return to the receiver after being reflected from the target. Doppler radar measures the frequency shift of the reflected pulse and computes the target velocity from this information. Sonar works in essentially the same manner. The *laser velocity and range finder* (LAVERA) phase-lock method adjusts the transmitted frequency under closed-loop control in order to keep the phase lag constant between received and transmitted waves.

Sixteen years ago, the author and students developed a means of measuring target range and velocity without explicit consideration of the Doppler effect (Nelson 1999). This system was developed for use with an amplitude-modulated laser beam. Many conventional laser-based velocimeters are based on a time-of-flight range finder principle in which laser pulses are reflected from a reflective (cooperative) target back to the source position and the range is found from the simple relation

$$L = \frac{c\Delta t}{2} \text{ m.} \quad (7.182)$$

where

c is the speed of light

Δt is the time from when a pulse is emitted to when it is received

L is the distance to the target

If the target is moving toward or away from the laser/photosensor assembly, the relative velocity can be found by approximating $V = \dot{L}$ by numerically differentiating the sequence of return times for a sequence of velocimeter output pulses; $Dt = \sum \Delta t_k$. Thus, $v_k = (\Delta t_k - \Delta t_{k-1})/T$. (Other, more complicated numerical differentiation algorithms more resistant to noise can also be used; see Section 10.3.) In 1992, Laser Atlanta LLC first marketed a vehicular velocimeter that used this principle (cf. www.SpeedLaser.com/ [accessed April 17, 2013]). Their SpeedLaser® Type R uses 238 pps; it has a target speed range of 10–200 mph \pm 1 mph, over a 0–30,000 ft target range.

The LAVERA system we will describe in this section uses quite a different principle to provide simultaneous velocity and range output. The LAVERA system was reduced to practice by graduate student Todd Nelson (1999) for his MS research in Biomedical Engineering at the University of Connecticut. The motivation for developing the system was to make a walking aid for the blind that could be incorporated into an electro-optical cane.

The system uses sinusoidally amplitude-modulated, CW light from an NIR LAD. The beam from the LAD is reflected from the target object and travels back to the receiver photosensor where its intensity is converted to an AC signal whose phase lags that of the transmitted light. The phase lag is easily seen to be

$$\Delta\phi = (2\pi)\left(\frac{\Delta t}{T}\right) = (2\pi)\left(\frac{2L/c}{T}\right)r \quad (7.183)$$

where

Δt is the time it takes a photon to make a round trip to the target

T is the period of the modulation, and c is the speed of light

The system we developed is a closed-loop system that tries to keep $\Delta\phi$ constant regardless of the target range, L . This means that as L decreases, T must decrease, and conversely, as L increases, T must increase to maintain a constant phase difference. Figure 7.50 illustrates the functional architecture of the LAVERA system. Note that the DC bias voltage, V_B , is required to set a mean intensity of the diode laser beam (light intensity cannot go negative) that is sinusoidally amplitude modulated. The transmitted intensity is thus

$$I_t = I_o + I_m \sin\left(\frac{2\pi t}{T}\right), \dagger I_o > I_m. \quad (7.184)$$

R-C HPFs remove the I_o component of the transmitted and received light signals before they are input to the comparators.

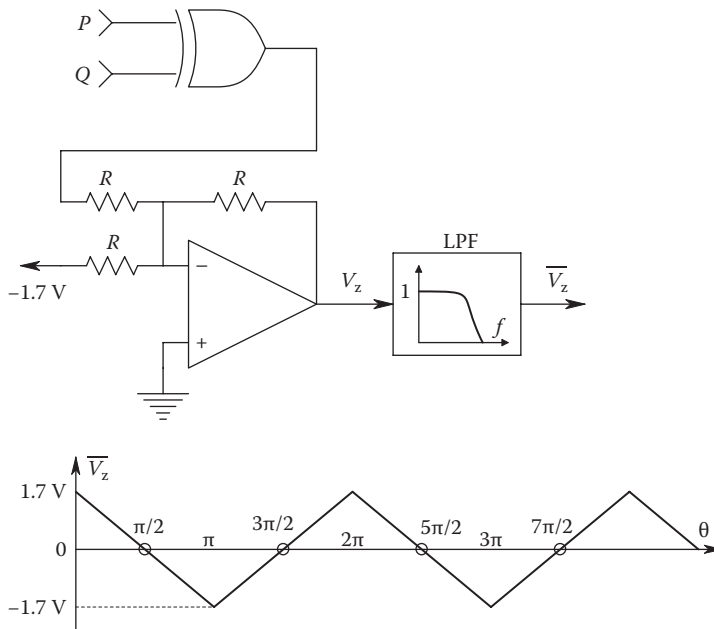


FIGURE 7.51 Exclusive OR gate used as a quadrature PD. The DC voltage added by the OA gives V_z zero average value when the EOR gate output has a 50% duty cycle. Note that the LPF output, \bar{V}_z , has zero value when the phase difference, θ , between TTL input square waves P and Q is equal to odd multiples of $\pi/2$.

In the SS with fixed L , $V_v = 0$ in a type 1 feedback loop. For $V_v = 0$, $\phi_{ss} = \pi/2$. But we have already established that the phase lag for a returning, modulated light beam is

$$\phi = (2\pi) \left(\frac{2L}{c} \right) \left(\frac{1}{T} \right) r. \tag{7.188}$$

If Equation 7.205 for ϕ is solved for T , and T is substituted into the VPC equation, we have

$$T_{ss} = (2\pi) \left(\frac{2L}{c} \right) \left(\frac{1}{\phi_{ss}} \right) = K_{VPC} V_{Pss} + b \tag{7.189}$$

↓

$$\frac{8L}{c} = K_{VPC} [V_{Lss} - V_r] + b. \tag{7.190}$$

If I let $V_r \equiv b/K_{VPC}$, the final expression for V_{Lss} is obtained:

$$V_{Lss} = \frac{8L}{(cK_{VPC})}. \tag{7.191}$$

Thus, the SS V_L is indeed proportional to L . Going back to the VPC equation, we find that for a stationary target, the SS VPC output frequency is given by

$$f_{ss} = \frac{1}{T_{ss}} = \frac{c}{8L} \text{ Hz.} \tag{7.192}$$

At $L = 1 \text{ m}$, $f_{ss} = 37.5 \text{ MHz}$; when $L = 30 \text{ m}$, $f_{ss} = 1.25 \text{ MHz}$.

Now, suppose the target is moving away from the transmitter/receiver at a constant velocity, $V = \dot{L}$, V_v is no longer zero, but can be found from the integrator relation

$$V_v = \frac{-\dot{V}_L}{K_i} = \frac{-8\dot{L}}{K_i K_{VPC} c}. \tag{7.193}$$

Notice the sign of V_v is negative for a receding target; also, as the target recedes, the VPC output frequency drops. Clearly, practical considerations set a maximum and a minimum for range L . The limit for near L is the upper modulation frequency possible in the CW system; the limit for far L is the diminishing intensity of the return signal and noise.

To appreciate the closed-loop dynamics of the LAVERA system, it is possible to make a systems block diagram as shown in Figure 7.52. Note that this system is nonlinear in its closed-loop dynamics. It can be shown (Nelson 1999) that the closed-loop system's undamped natural frequency, ω_n , and damping factor are range dependent, even though the SS range and velocity sensitivities are independent of L . The undamped natural frequency for the LAVERA system of Figure 7.50 can be shown to be

$$\omega_n = \sqrt{\frac{4(3.4)cK_i K_{VPC}}{64\tau_i L_o}} \text{ r/s.} \tag{7.194}$$

The closed-loop LAVERA system's damping factor can be shown to be

$$\xi = 1/2 \left[\frac{64L_o}{4(3.4)cK_i K_{VPC}} + \tau_i \right] \sqrt{\frac{4(4.3)cK_i K_{VPC}}{64\tau_i L_o}}. \tag{7.195}$$

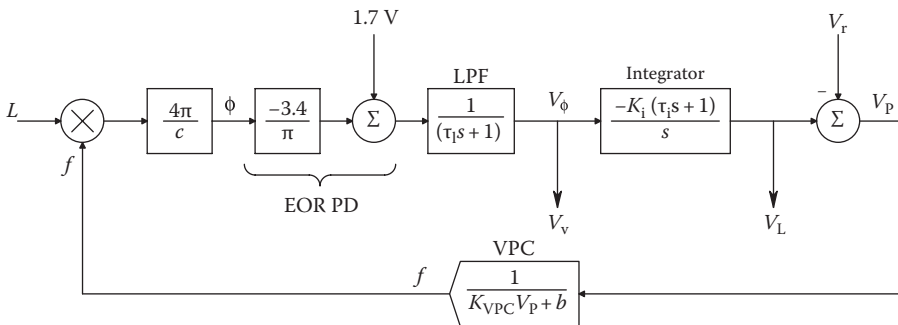


FIGURE 7.52 Systems block diagram of the LAVERA system, relevant to tracking a moving target. V_v is proportional to target velocity and V_L is proportional to target range, L . See text for analysis.

The LAVERA system architecture was also extended to ultrasound and microwave distance measurement systems (Nelson 1999). At the heart of all LAVERA systems is the use of a VPC instead of a voltage-to-frequency converter (VFC). The VPC is a VCO in which the *period* of the output waveform is directly proportional to the input (control) voltage, that is,

$$T = \frac{1}{f} = b + K_p V_c \text{ s.} \tag{7.196}$$

Thus, the output *frequency* of a VPC is given by

$$\omega_o = \frac{2\pi}{[b + K_p V_c]} \text{ r/s.} \tag{7.197}$$

For small changes in V_c we can write

$$\Delta\omega = \left(\frac{\partial\omega_o}{\partial V_c} \right) \Delta V_c = \frac{-2\pi}{(b + K_p V_c)^2} \Delta V_c. \tag{7.198}$$

Note that there is no distance dependence in the velocity and range sensitivity expressions; they remain constant over the system's operating range.

Examples of an analog and a digital VPC are shown in Figures 7.53 and 7.54, respectively. Note that the analog VPC uses a standard VFC but takes the reciprocal of the input analog control voltage. Note that $f = K_F V_2$ for the VCO. In the nonlinear op amp circuit, $V_R = V_2 V_3 / 10$. Also, from Figure 7.53, $V_3 = K_V (K_P V_P + b)$, so finally,

$$f = \frac{K_F 10 V_R / K_V}{K_P V_P + b} \text{ Hz.} \tag{7.199}$$

As in the case of VFCs, there are practical limits to the range of the output frequency in VPCs.

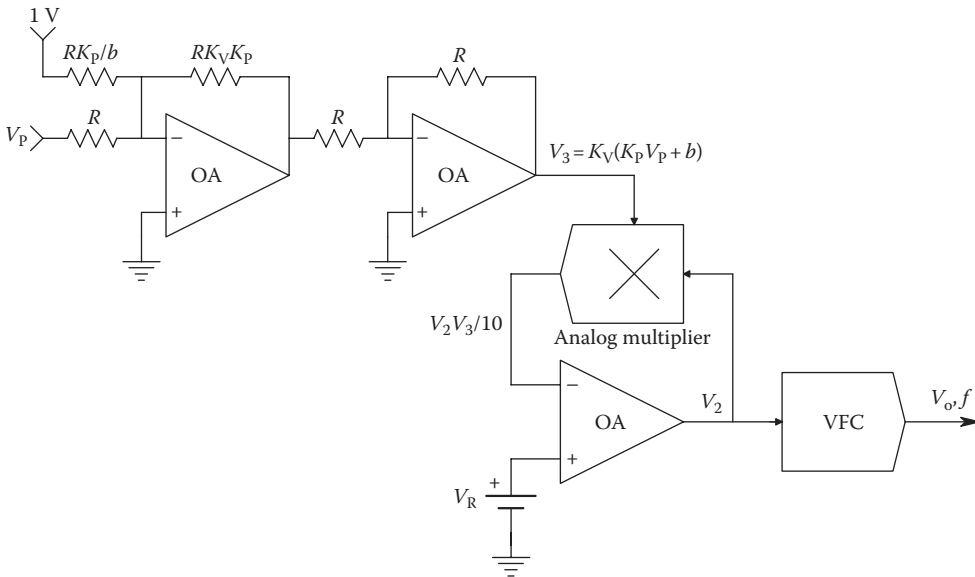


FIGURE 7.53 Analog, VPC oscillator. The analog VFC can have a sinusoidal or digital (TTL) output. See text for analysis.

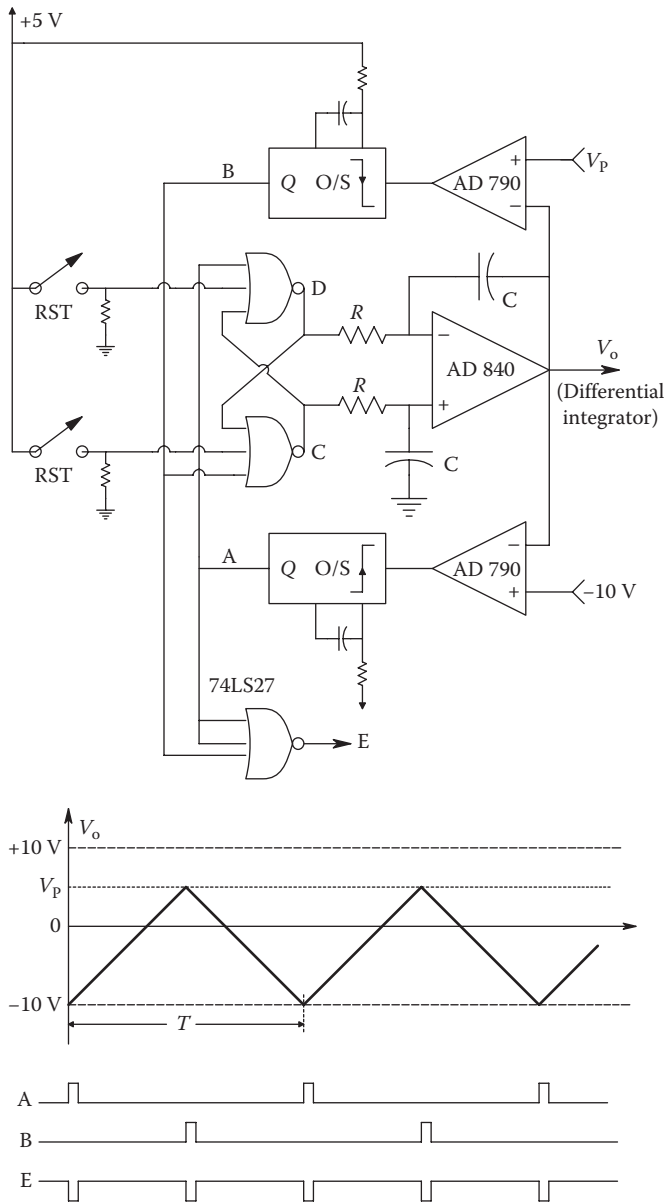


FIGURE 7.54 Hybrid (digital + analog) VPC circuit. Note: AD790 = high-speed analog comparator. AD840 = op amp. O/S = one-shot multivibrator. 74LS27 = 3-input NOR gate. The waveforms illustrate analog and digital voltages in the system when $V_p = +5\text{ V}$.

The digital VPC of Figure 7.54 is really a hybrid circuit in which the outputs of the R-S flip-flop (FF) are integrated by the op amp. Assume C is HI and D LO: V_o ramps positive until it reaches the input signal, V_p . The upper AD790 comparator then goes LO, and the upper one-shot generates a narrow positive output pulse that causes the NOR gate R-S FF to change states. Now, C is LO and D HI; the integrator output now

ramps negative until it reaches -10 V, the threshold for the lower comparator to go HI. When the lower comparator goes HI, it triggers the lower one-shot to give a narrow positive output pulse that again causes the FF to change state, and the process repeats itself. One can see intuitively that the more positive V_P is, the longer will be the E output pulse period.

Let us now derive the transfer function of this DVPC, $T = f(V_P)$. First, we need the transfer function of the differential integrator. Using superposition, and assuming an ideal op amp,

$$V_o(s) = V_D(s) \left[\frac{-1}{sRC} \right] + V_C(s) \left(\frac{1}{[(s+1/RC)RC]} \right) \left(1 + \frac{1/sC}{R} \right) = [V_C(s) - V_D(s)] \left[\frac{1}{sRC} \right]. \quad (7.200)$$

Note that the R-S FF's outputs $V_D(s)$ and $V_C(s)$ can be treated as steps of voltage occurring at $t = 0$ and thus can be written as $V_C(s) = 3.4/s$ and $V_D(s) = 0.25/s$ or $V_C(s) = 0.25/s$ and $V_D(s) = 3.4/s$. Thus, with appropriate attention to initial conditions, and assuming $V_P(t)$ remains constant over the VPC period, it is easy to show that the integrator sawtooth output has slope of $m = \pm RC/3.15$. So starting at $V_o(t) = -10$ V, it is found that it takes one-half sawtooth period for $V_o(t)$ to reach V_P . $T/2$ is easily found to be

$$\frac{T}{2} = \frac{(V_P + 10)RC}{3.15} \text{ s.} \quad (7.201)$$

Note that it also takes the same $T/2$ for $V_o(t)$ to ramp down to -10 V again.

Because the bottom 3-input NOR gate gives an output pulse every time the R-S FF changes state, the pulse output of the VPC, E , has period $(V_P + 10)RC/3.15$ s. In terms of frequency, the *lowest* output frequency of E will be

$$f_{E\min} = \frac{3.15}{20RC} \text{ pps.} \quad (7.202)$$

If the minimum V_P allowed is -9.8 V, the *maximum* digital VPC output frequency will be

$$f_{E\max} = \frac{3.15}{0.2RC} \text{ pps.} \quad (7.203)$$

For example, let $C = 1$ nF, $R = 1$ k, then $f_{E\max} = 15.75$ Mpps, and $f_{E\min} = 157.5$ kpps occurs when $V_P = +10$ V.

The LAVERA system is unique in that it provides simultaneous range and velocity output voltages for a moving target. If the second integrator is added in the loop, the closed-loop system becomes type II, giving zero SS error in V_V when the target is moving at constant velocity (not accelerating). The velocity output is independent of range. This independence disappears if a VFC is used in the system instead of a VPC. The general LAVERA systems architecture can be extended to pulsed or CW ultrasound, microwaves,

and pulsed lightwaves. Sinusoidally amplitude-modulated, CW lightwaves were used in the system described earlier because lower-bandwidth amplifiers can be used than required for pulsed systems.

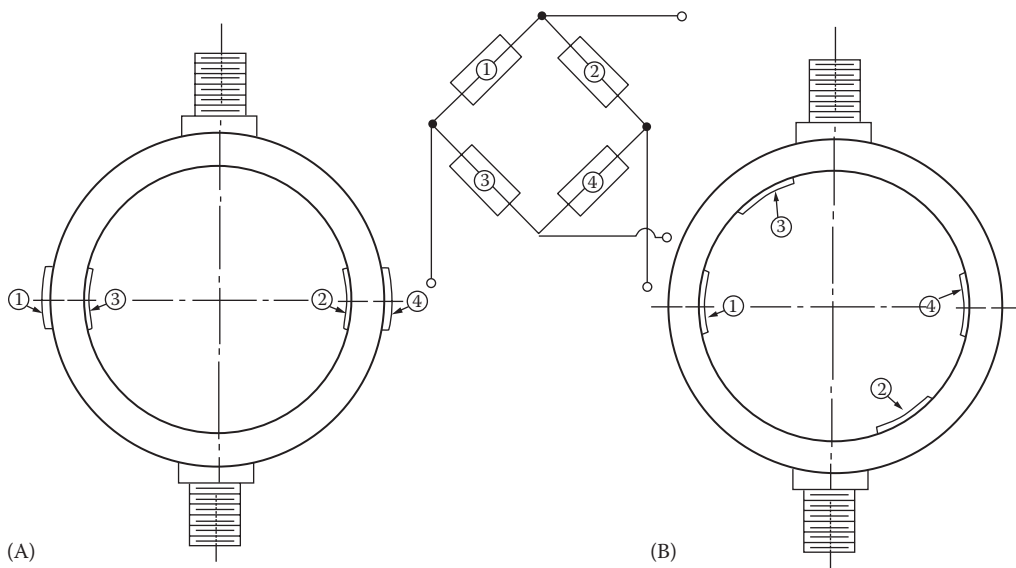
7.4 Measurement of Force and Torque

Measurement of force is important in many branches of science, engineering, and biomedicine. For example, the magnetic attractive or repulsive force between current-carrying conductors is used in the definition of ampere as the unit of electric current. The measurement of forces in a uniform gravitational field is how we weigh things. The thrust of a jet engine is measured as a force. In biomechanics, force plates are used to study the forces involved in various gaits (walking, trotting, running, sprinting). Physiologists and physical therapists are concerned with the contractile force produced by muscles under various physiological and pharmacological conditions. Aerospace engineers and naval architects are concerned with the fluid dynamic forces on airframes, wings, control surfaces, hulls, etc. A locomotive or tractor-trailer exerts force on a load to accelerate it to speed and then keep it rolling. Traction force times velocity is the power expended in moving the load (horsepower). The range of forces that are commonly measured is from micrograms to hundreds of thousands of kilograms, $\times 1$ g. Obviously, no one sensor can cover this dynamic range, so there are many types and sizes of force sensors (load cells) available to the instrumentation engineer, depending on the application.

Torque measurements are generally made on rotating shafts connecting a power source (electric motor, diesel engine, etc.) to a device doing work (pump, rolling mill, conveyor belt, etc.) Torque measurements are used to calculate rotary output power (torque \times angular velocity). In general, direct torque measurements are more difficult to make than are force measurements because the shaft is rotating, and slip rings must be used to couple the output signal to the stationary, outside world. There are means, as will be seen in Section 7.2.5, of measuring torque optically or magnetically that do not require direct contact with the rotating shaft.

7.4.1 Load Cells for Force Measurement

Paradoxically, most force measurements involve the small displacement of a stressed, linearly elastic member. Displacement or *strain* is proportional to the force load-produced *stress*. The transduction of the small displacement may be done optically by interferometry or electrically by a sensitive LVDT, by a bonded or unbonded wire strain gauge bridge, by a semiconductor strain gauge bridge, by a circuit that senses small changes in capacitance, or by a piezoelectric crystal (crystal force transducers only respond to time-varying forces). Any force applied to a piston acting on a fluid will produce a rise in pressure, which, if there are no leaks, will remain constant and can be measured with a PT. The most basic type of load cell is one that uses four, bonded strain gauges in a Wheatstone bridge configuration, illustrated in Figure 7.55. The four-arm bridge is inherently temperature compensated, and the circular load cell configuration works for compressive as well as tension loads.

**FIGURE 7.55**

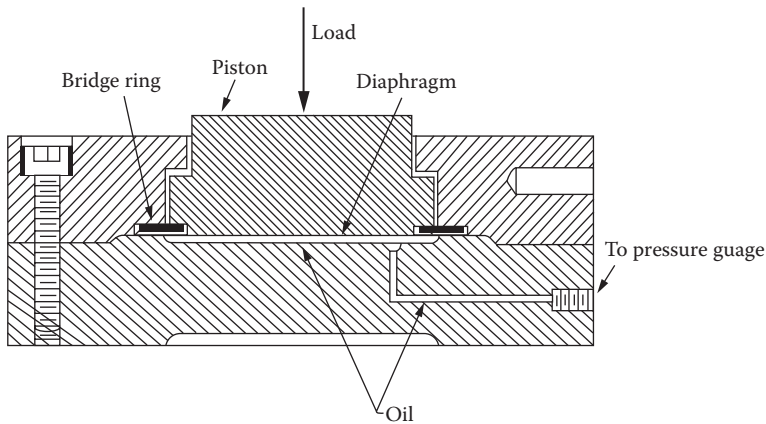
Two, ring-type load cells used to measure tensile or compressive forces. (A) The strain gauges respond only to the bending strains; the axial strain components cancel out in the bridge. (B) Wheatstone bridge has greater sensitivity because gauges 1 and 4 respond to both axial and bending strain components. (From Beckwith, T.G. and Buck, N.L., *Mechanical Measurements*, Addison-Wesley, Reading, MA. With permission.)

Interface, Inc., of Scottsdale, AZ (cf. www.interfaceforce.com/ accessed April 17, 2013), offers a large line of bonded strain gauge bridge load cells with ranges from 5 to 200,000 lb full scale in tension or compression. The Interface load cells are quite precise and linear. Nonlinearity is $\pm 0.03\%$ of rated output, hysteresis is 0.02% of rated output, and nonrepeatability is $\pm 0.01\%$ of rated output.

Schaevitz Engineering, now Sherborne Sensors, markets LVDT-based load cells with ranges from 10 g full scale to as much as 500,000 lb full scale. Frequency response of an LVDT-based load cell is set by the carrier frequency used with the LVDT, which can be as high as 10 kHz. The Schaevitz load cells have linearities of 0.2% of full range, resolutions of 0.1% of full range, and repeatability of 0.1% of full range. Deflection at full-scale force may be 0.005 in. if the Schaevitz 004-XS-B subminiature LVDT is used in the cell.

Still another type of load cell is the capacitive cell marketed by Kavlico Corp., Chatsworth, CA, a brand of CST sensors. In this cell, the applied force causes the deflection (5×10^{-4} in. at full-scale force) of a ceramic diaphragm, one surface of which is metallized, which in turn causes a proportional capacitance change that is sensed electronically. Kavlico load cells have full-scale ranges of 100 g–60 lb. System accuracy is $\pm 5\%$ of full scale, and repeatability and hysteresis are typically $\pm 0.36\%$.

Grass Instruments, Quincy MA, now Grass Technologies, a brand of Natus Medical, Inc. (cf. Natus 2013), made a physiological force sensor that is basically a cantilever to which are bonded four strain gauges that made a temperature-compensated Wheatstone bridge. The Grass force sensors had sensitivities that could be decreased by the addition of insertable spring sets. The Grass FT-03C had a working range of 50 g with no insertable springs, but could be given a full-scale range of 2 kg with the stiffest spring set. Deflection at full-rated load is 1 mm. The mechanical natural resonant frequency was 85 Hz for the 50 g range and

**FIGURE 7.56**

Section through a hydraulic load cell. This system responds only to compressive forces. (From Beckwith, T.G. and Buck, N.L., *Mechanical Measurements*, Addison-Wesley, Reading, MA. With permission.)

500 Hz for the 2 kg range. The strain gauge bridge output was 1.5 mV/mm/V applied to the bridge. Obviously, differential amplification was needed.

Accurate, linear load cells for extreme compressive forces can be made using hydraulic technology. Figure 7.56 illustrates such a hydraulic cell. Beckwith and Buck (1961) stated that capacities as high as 5×10^6 lb with accuracies of $\pm 0.1\%$ of full scale have been attained with this design. Pneumatic load cells offer the advantage over hydraulic fluid designs over a lower force range in terms of temperature insensitivity. Air is compressible and does not freeze or boil over a very wide temperature range. Pneumatic load cells are generally made using a compliant diaphragm, rather than a piston. In addition, they use a feedback design in which the internal air pressure is adjusted by a needle valve attached to the moving diaphragm such that if the external, downward load force exceeds the upward pneumatic force on the diaphragm, the diaphragm drops downward and decreases the outflow of air by decreasing the pneumatic bleed conductance. This causes the internal air pressure to rise, raising the diaphragm and increasing the outflow conductance. An equilibrium is reached in which the internal pressure times the effective diaphragm area equals the applied load. Also, the volume inflow must equal the volume outflow at equilibrium. Refer to Figure 7.57. Assume the needle valve has a pneumatic conductance that is a function of diaphragm displacement given by

$$G_o(x) = G_{oMAX} \quad x < 0 \quad (7.204a)$$

$$G_o(x) = G_{oMAX} \left(\frac{1-x}{x_{MAX}} \right) \quad 0 \leq x \leq x_{MAX} \quad (7.204b)$$

$$G_o(x) = 0 \quad x \geq x_{MAX}. \quad (7.204c)$$

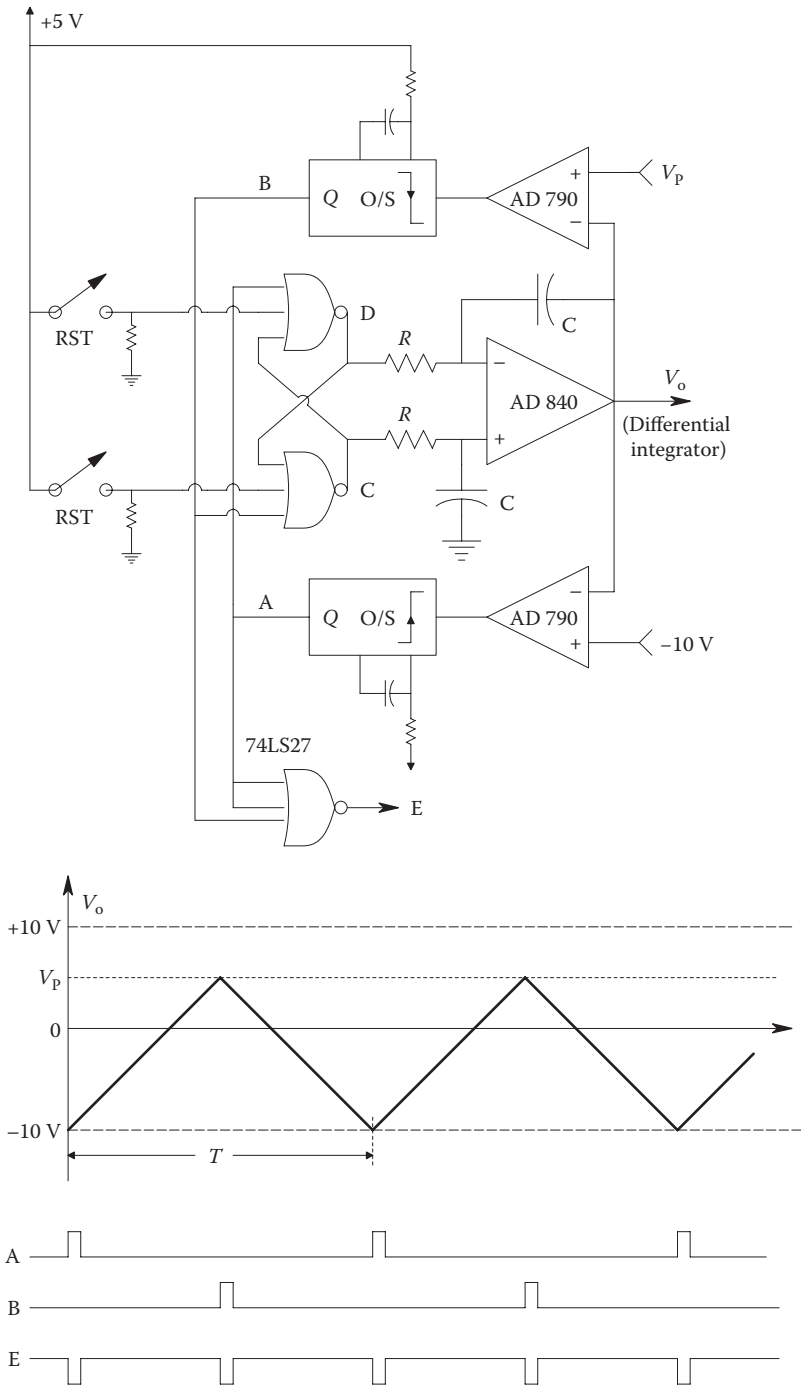


FIGURE 7.57 Section through a pneumatic load cell. (From Beckwith, T.G. and Buck, N.L., *Mechanical Measurements*, Addison-Wesley, Reading, MA. With permission.)

Here, negative x is upward displacement of the membrane from $x = 0$. It is not difficult to show that in the SS,

$$P_i = \frac{F}{A} = \frac{P_o G_i}{(G_i + G_o)} \quad (7.205)$$

where

G_i is the fixed, pneumatic input conductance

A is the effective area of the diaphragm

F is the external load force

P_o is the pressure in the air reservoir supply

From Equation 7.222, we see that the practical bounds on the measurement of F are set by $F_{MAX} = P_o A$ at the upper end and $F_{MIN} = P_o A G_i / (G_i - G_{oMAX})$ at the lower end. One problem with simple pneumatic load cells is that they may become unstable under certain conditions. If the force on the pneumatic load cell is a mass, the air in the chamber behaves like a nonlinear spring, and oscillations and instability in the operation of this sensor can occur unless pneumatic damping is used in its design. Beckwith and Buck (1961) reported that 40-ton capacity pneumatic load cells are commercially available.

Closed-loop, EM force cells are used for weighing light objects in the range of 0.1 mg–100 g. In this design, an upward force is generated electromagnetically to equal the external, downward force, as shown in Figure 7.58. Equilibrium position can be sensed either electro-optically or with a high-resolution LVDT. The EM force is generated by a coil and permanent magnet assembly similar to that used in a loudspeaker, except that there is no spring required. The coil current is made proportional to the integral of the downward displacement. A system block diagram describing this system's dynamics is shown in Figure 7.58B. There is zero SS displacement error, and it can be shown that

$$\frac{V_c}{Mg} = \frac{1}{G_M K_F} \text{ V/N.} \quad (7.206)$$

Other obvious and far less technical means for measuring forces less than several hundred kilograms include spring scales that make use of Hooke's law and a calibrated scale; displacement of the moving end of the spring is proportional to the applied force. Steelyards, laboratory beam balances, platform balances, etc., all make use of calibrated weights and lever arms pivoted on knife-edge bearings. A human operator adjusts the position of one or more known weights on the long arm of the balance until the arm is level and stationary, signifying that the unknown force times its lever arm equals the sum of the known forces times their lever arms.

7.4.2 Torque Measurements

Torque measurements on rotating shafts are made with more difficulty than are those on stationary shafts. The reason for this difficulty is that the wires from sensor elements, such as a strain gauge bridge cemented to the shaft, must be brought out to the nonrotating world through slip rings. Noise introduced by the brushes making contact with the rotating slip rings can reduce the resolution of a rotating, torque-measuring system below than that of an identical system used on a stationary shaft without slip rings. Figure 7.59 illustrates the configuration of bonded strain gauges on a shaft in order to

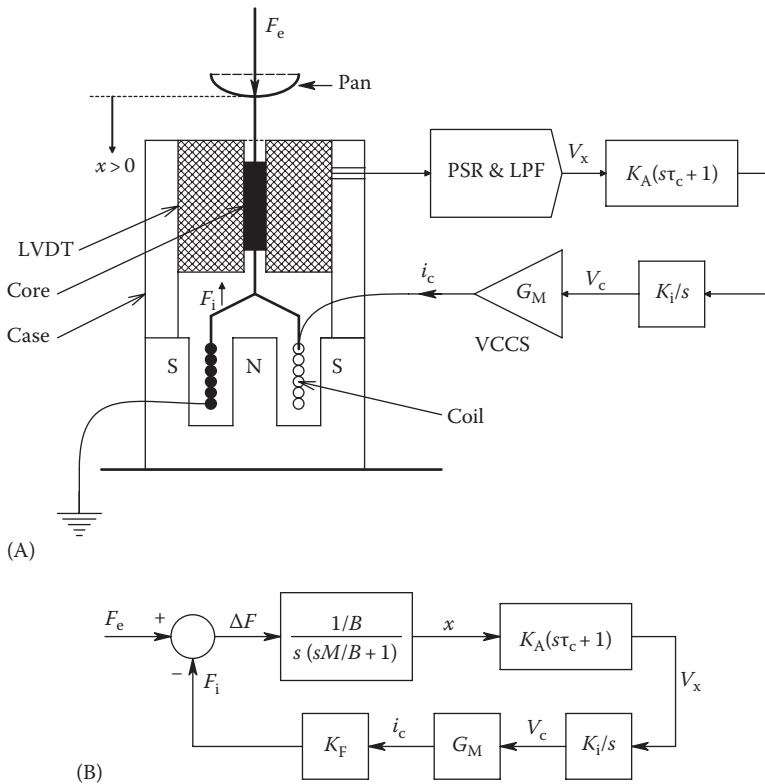


FIGURE 7.58

(A) Section through an electrodynamic, feedback load cell. Position is sensed by an LVDT. The LVDT output is conditioned by a PSR and an LPF, then by a proportional plus derivative compensator and integrator. The integrator's output, V_c , is then conditioned by a transconductance amplifier (VCCS) whose output, i_c , is the input to the force coil. (B) Systems block diagram of the electrodynamic, feedback load cell.

measure torsional strain caused by a torque couple applied to the ends of the shaft. Such a configuration may be shown to be insensitive to temperature changes, axial load, and bending of the shaft.

Fortunately, there are other means of measuring torque on a rotating shaft that do not require electrical contact with the shaft. In one approach, uniformly spaced, alternating black and white markings are painted around the circumference of each end of the torque-carrying shaft under measurement. A photosensor consisting of an IR LED and a phototransistor is placed above both rotating, black/white bands. The output of each photosensor is a square wave with frequency proportional to the shaft's angular velocity. However, the phase between the square waves is determined by the torque-produced displacement angle between the black/white bands. This angle is given by

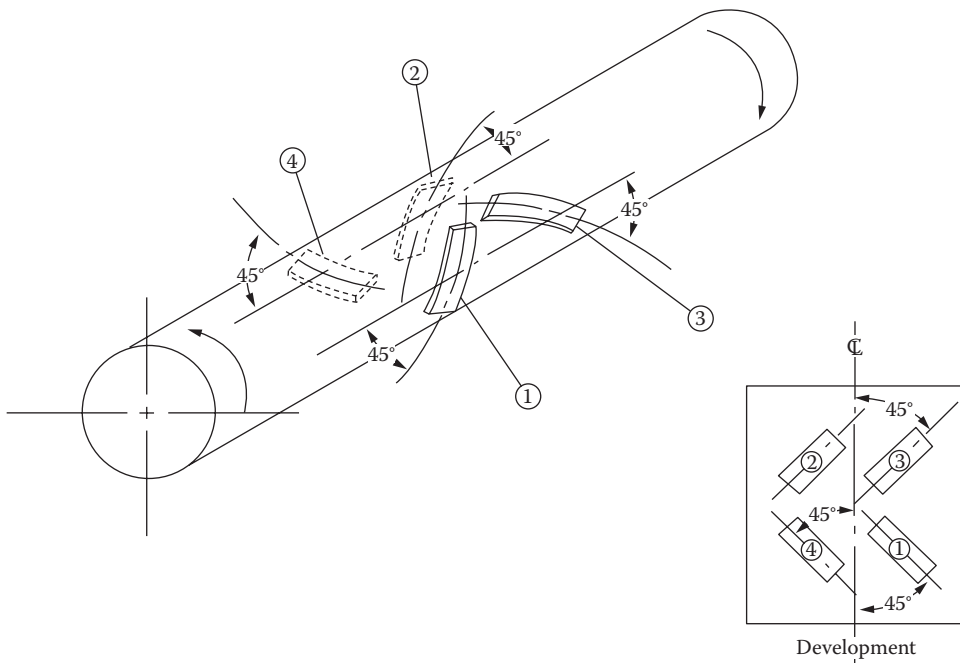
$$\phi = KTd \tag{7.207}$$

where

d is the distance between the bands

T is the torque on the shaft

K is a constant determined by the material and dimensions of the shaft

**FIGURE 7.59**

Four bonded strain gauges cemented to a rotating shaft under torque. This configuration is insensitive to temperature, axial load, or bending. Slip rings must be used to make electrical connections to the gauges (not shown). Slip rings are noisy, so a wireless telemetry transmitter on the shaft can be used. (From Beckwith, T.G. and Buck, N.L., *Mechanical Measurements*, Addison-Wesley, Reading, MA. With permission.)

The electrical phase angle between the two square wave trains, θ , can be measured and is related to ϕ by

$$\theta = N\phi \quad (7.208)$$

where N is the integer number of black/white cycles inscribed around the circumference of the torque shaft at each end. The frequency of each square wave is just

$$f = N \times \frac{RPM}{60}. \quad (7.209)$$

RPM is the shaft's revolutions per minute. One way of measuring the phase angle, θ , is to use a PD IC, such as the Motorola MC4044. The MC4044 has a working range of $\pm 360^\circ$ and generally can be used to resolve angles as small as $\pm 0.1^\circ$. Obviously, to measure low values of torque, a PD is needed that can resolve hundredths of a degree or better and work over a wide range of frequencies.

A *frequency-independent millidegree resolution PD design* was devised by Du (1993), based on a design by Nemat (1990). The Du system is illustrated schematically in Figure 7.60. The periodic reference signal (A) is converted to a TTL waveform by a fast comparator, the output of which is the input to a PLL in which the VCO output is made to follow the input frequency and multiply it by a factor of 360,000. Thus, every input cycle produces 360,000 PLL

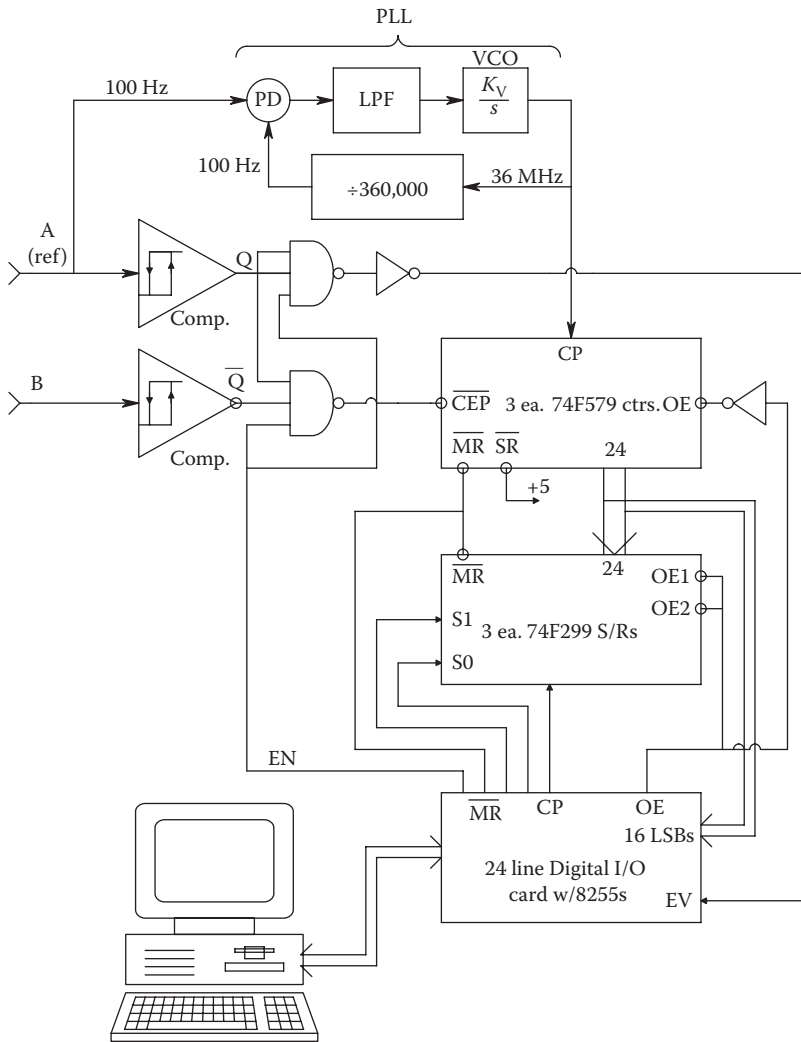


FIGURE 7.60
Schematic of the millidegree phase meter designed by Du (1993).

clock cycles, and the input phase is effectively divided into millidegrees, that is, there is one PLL clock pulse for every 0.001° of input, regardless of the exact input frequency. The reference and complimentary, phase-shifted square wave inputs act through a NAND gate (G1) to enable a 24-bit, high-frequency binary counter that counts the high-frequency output of the PLL for N cycles of the input wave. For purposes of discussion, assume that the input signals are 100 Hz; thus, the PLL's high-frequency output will be 36 MHz. The number of high-frequency pulses counted depends, of course, on the phase shift between the 100 Hz input square waves, as well as N .

If the phase shift is zero, no pulses are counted. On the other hand, if, for example, the phase shift is 3.6° , then 3600 PLL clock cycles will be counted for every one of the N gated input cycles. After N input waveform cycles, the cumulative count in the binary counter will be 9.216×10^5 , if $N = 256$ and $\theta = 3.6^\circ$. In order to obtain an average count for N input cycles, the cumulative count in the counter is loaded in parallel to a 24-bit

shift register (3, 74F299 ICs). Division by N is easily accomplished by shifting the register contents to the right by k clock cycles from the controlling computer. We make $N = 2^k$ in this system, with $k = 0, 1, 2, \dots, 8$. The 16 LSBs of the shifted binary number in the register is then loaded into the computer and displayed as a decimal angle in the format, 88.888°. For our 3.6° phase shift example, this would be 3600 or 3.600°. The PLL allows precise measurement of the phase difference between the input signals A and B even though their frequency may not be fixed. In fact, it will vary as the shaft speed varies.

A limitation of the Du system emerges when the phase angle, θ , is large. Now, the counter must count upward of 360,000 pulses per cycle. The cumulative count for $N = 256$ would be about 92.2 million before division by 256. Thus, if intended to work with large phase angles, the Northrop and Du system needs counters capable of holding raw counts of over 2^{26} . A practical upper bound on the input signal frequency is set by the maximum operating frequency of the PLL and the maximum counting rate of the counter. A Signetics NE564 PLL will operate up to 50 MHz. The 74F579 decade counters work up to 115 MHz.

An alternate to the use of photoelectronic measurement of the shaft's torque angle, ϕ , is the use of a variable reluctance pickup, such as shown in Figure 7.61. Here, a voltage pulse is induced in the coil every time a steel gear tooth passes under the permanent magnet. Two identical steel gears, concentric with the torque shaft, are used, as in the case of the photoelectronic pickup described earlier. One gear is at the motor end of the shaft, the other at the load end. The angular difference, ϕ , between the two gears depends on the shaft length L , the torsion spring constant of the shaft K_v , and of course the load torque, τ_L ; thus, $\phi = LK_v\tau_L$. Because the output pulses are biphasic, fast analog comparators, such as LT1016s, are used to convert them to TTL pulse trains for input to the PD.

Other schemes of no-touch torque measurement in rotating shafts have been described that make use of the magnetoelastic (Villari) effect (Lion 1959). Here, the magnetic permeability of ferromagnetic materials is seen to change in a nonlinear manner when the

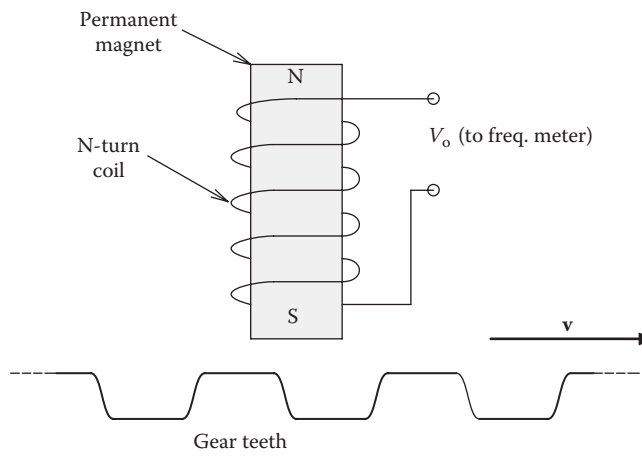


FIGURE 7.61

Diagram of a variable reluctance tachometer pickup. As each steel gear tooth passes under the magnet and coil, there is a transient increase in the flux linkages through the coil because of lower magnetic reluctance. The biphasic voltage pulses across the coil, V_o , occur every time a gear tooth passes the pickup, and the height of the pulses increases with gear speed, as well.

material is subjected to mechanical stress, such as torsion and tension. Lion reports that the Villari effect (see *Glossary*) has been used to measure forces in load cells. As the force is increased, there is a nonlinear decrease in the inductance of a coil wound around the ferromagnetic core under compression. Villari effect-based sensors may be adaptable to the measurement of torque (cf. *NASA Tech Briefs* 15(3): 1991, p. 50). Besides their obvious nonlinearity, the Villari effect has a strong temperature dependence that would need compensation in a field-grade instrument. The inductance change resulting from mechanical strain of the core must be detected with some form of inductance bridge or by sensing the change in the output frequency of an oscillator in which the inductance is a frequency-determining component. (See the tutorial paper on magnetostrictive linear position sensors that use the Villari effect by MTS [2013].)

7.5 Pressure Measurements

7.5.1 Introduction

Many interesting transducer systems have been developed to measure fluid pressure, as gauge pressure, absolute or differential pressure. Applications of pressure sensors are broad, including but not limited to aircraft altimeters; submarine depth meters; pressure meters for gas tanks; sensors for aircraft rate of climb and airspeed indicators; medical/physiological pressure sensors for blood, cerebrospinal fluid, intraocular pressure, intrathoracic pressure, kidney dialysis pressure, etc.; hydraulic system pressure; oil pressure in bearings for turbines, generators, diesel engines, etc.; liquid level indicators for storage tanks; and purge liquid pressure in oil well reconditioning. Pressures involved in controlled explosions, such as combustion chamber pressures in internal combustion engines or the gas pressure in the chamber or barrel of firearms, can be measured with special pressure sensors having high-frequency response and high-temperature and high-pressure capabilities. Also of interest are pressure sensors used in vacuum systems, that is, sensors responsive to very low pressures.

In the following sections, we discuss the important, commercially available pressure sensors and the mechanisms by which they work. It will be seen that nearly all pressure sensors use some type of membrane or diaphragm that deflects linearly with pressure, relative to the zero gauge pressure condition. As you will see, measurement of this deflection has been done in a number of ingenious ways.

There are a number of units used to describe pressure. In the United States, *pounds per square inch* (psi) is commonly used, also the manometer pressures, *inches of water* and *inches of mercury*. The SI pressure unit is the *pascal* (N/m^2); other metric pressures include the *bar* (*millibar*), the kgf/cm^2 , and the manometer pressures in *mmHg*, *cmH₂O*, and, of course, the *atmosphere*. BP is commonly read in mmHg. Note that one *atmosphere* = 14.696 psi = 1.0133×10^{-1} Pa = 1.0133×10^3 mbar; also, 10^6 dyn/cm² = 1 bar. Modern tire pressure gauges now commonly have scales in psi *and* kilopascals (in case one drives outside of the continental United States).

7.5.2 High-Pressure Sensors

In the category of high-pressure sensors, we place all transducers acting on pressures around and above atmospheric pressure, including those that measure gauge pressure

(pressures above atmospheric), absolute pressure, and pressure differences. At the low end of the measurement range, we find pressure sensors specialized for biomedical applications and applications in automobile fuel and emission control systems. One design approach for low-pressure sensors used by Schaevitz Engineering (now Sherborne Sensors) is to use a compliant bellows attached to the core of an LVDT. Bellows deflection is proportional to the pressure difference across its walls, and models are available that use vacuum, vented gage, or differential pressure references. Such designs can be made quite sensitive (0–2 in. H₂O in the Schaevitz Model P-3000 series). Because of the high compliance (low stiffness) of the bellows and the mass of the LVDT core, such sensitive pressure sensors have limited frequency response and are generally suited for measuring static pressures or pressures that vary slowly.

Another approach to pressure measurement was seen in the design of pressure sensors made by Conal Precision Instruments, Inc., Franklin Lakes, NJ. Conal pressure sensors use the pressure of a bellows against a quartz crystal used in a nominal 40 kHz oscillator to reduce the oscillator's frequency. Unfortunately, the relation between oscillator period and input pressure is nonlinear, and a calibration equation must be used to convert period to a function proportional to pressure. Conal claimed resolution of 0.0001% of FS and hysteresis of 0.01% of FS. Full-scale pressure ranges from 20 to 1000 psiA are available.

Many of the pressure sensors described below use the deflection of a thin diaphragm or the strain produced in this diaphragm by pressure as the primary step in converting pressure to an electrical output. When pressure is applied to a flat diaphragm on the side opposite the gauges (see Figure 7.62), the central gauge or gauges experience tension, while the gauge cemented to the edge of the diaphragm sees compression. In terms of the deflection of the center of the diaphragm, deflection is linear with applied pressure for deflections up to about 30% of the diaphragm thickness (Beckwith and Buck 1961). The central deflection of a diaphragm is given approximately by

$$\delta y = \frac{3\Delta P R^4(1-\mu^2)}{16 Y t^3} \quad (7.210)$$

where

ΔP is the pressure difference across the diaphragm

R is the diaphragm's radius

t is its thickness

Y is its Young's modulus

μ is its Poisson's ratio

The radial stress in the diaphragm at its edge is

$$\sigma_r = \left(\frac{3}{4}\right) \left(\frac{R}{t}\right)^2 \Delta P \text{ psi.} \quad (7.211)$$

The tangential stress at the center of the diaphragm is given by

$$\sigma_t = \left(\frac{3}{8}\right) \left(\frac{R}{t}\right)^2 (1+\mu) \Delta P \text{ psi.} \quad (7.212)$$

The parameters are the same as for Equation 7.227 (Beckwith and Buck 1961, Ch. 12).

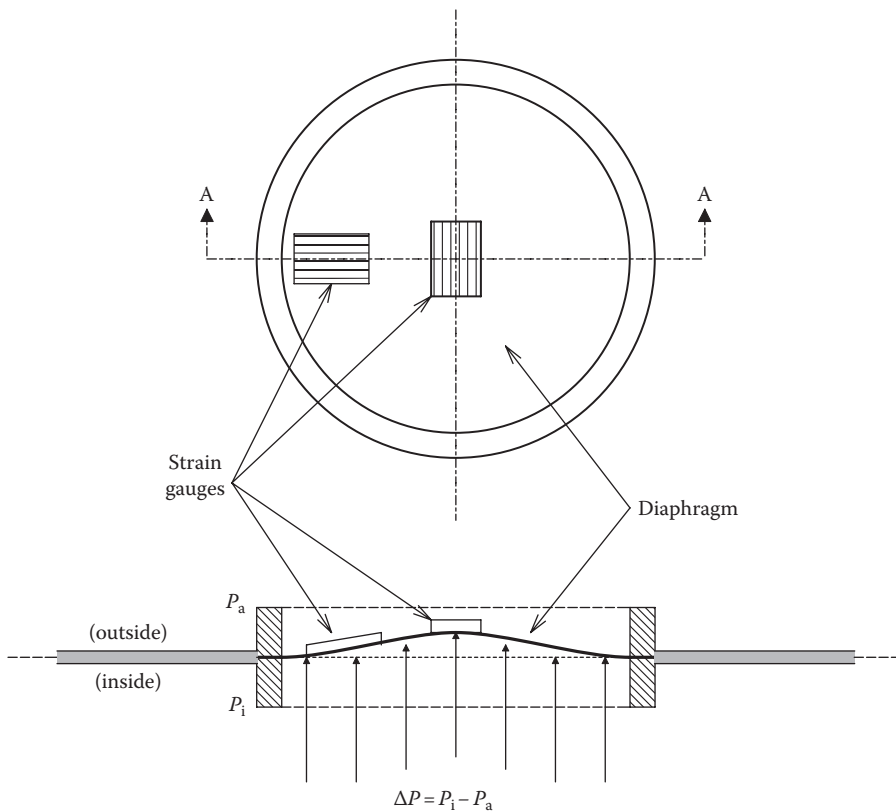
**FIGURE 7.62**

Diagram of a circular diaphragm pressure sensor in which the sensing elements are a pair of strain gauges bonded to the thin steel diaphragm as shown. The strain gauges are put in one arm of a Wheatstone bridge, giving temperature compensation. (From Beckwith, T.G. and Buck, N.L., *Mechanical Measurements*, Addison-Wesley, Reading, MA. With permission.)

Stow Laboratories, Inc., Hudson, MA, made miniature, Pitran™ pressure sensors. In the Stow design, illustrated in Figure 7.63, pressure acting on a diaphragm caused a pointed, insulated stylus to bear mechanically on the *layer cake* of an *npn* silicon transistor. This transferred pressure produces a large, reversible, effective decrease in the transistor's β , lowering I_C and thus raising V_{CE} . Stow claimed that Pitran sensors had 0.5% linearity and hysteresis and a 65 dB dynamic range. Maximum membrane deflection was 2 $\mu\text{in.}$, and the Pitran's mechanical resonant frequency was about 150 kHz. Pitran models had FS pressure ranges of 0.1–20 psid. An early technical paper on the design and testing of the Pitran pressure sensor was written by Quinn and Posipanko (1970).

A family of pressure sensors based on a differential, variable reluctance transducer principle is sold by Validyne Engineering Corp., Northridge, CA. A cross-sectional view of the Validyne pressure sensor is shown in Figure 7.64. The Validyne sensor design uses a diaphragm of magnetically permeable stainless steel clamped between two identical blocks. Embedded in each block are identical *E* core and coil assemblies. With zero deflection of the diaphragm, both coil and core assemblies see an identical air gap, and therefore, both coils have identical inductances. When a pressure difference exists across the diaphragm, its deflection causes an increase in the air gap or reluctance of one coil assembly and a

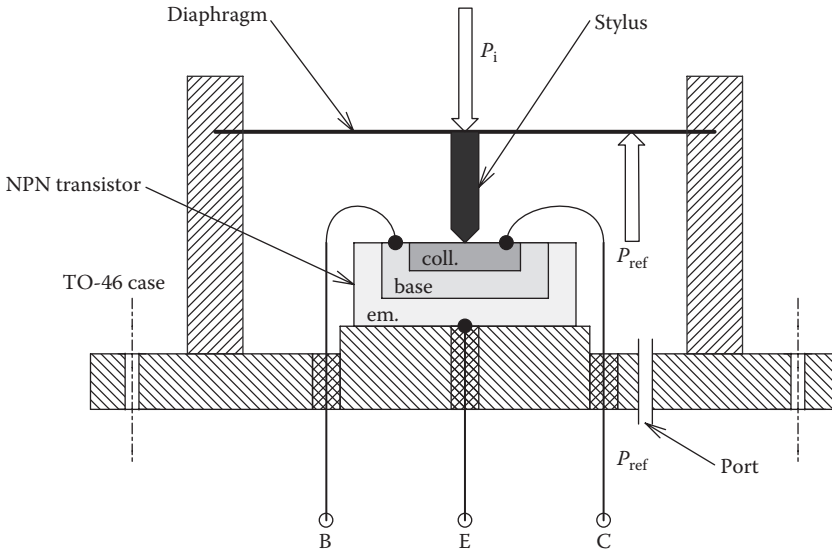


FIGURE 7.63 Section through a Pitran™ pressure sensor. (Reprinted with permission from Stow Laboratories, Inc., Hudson, MA.)

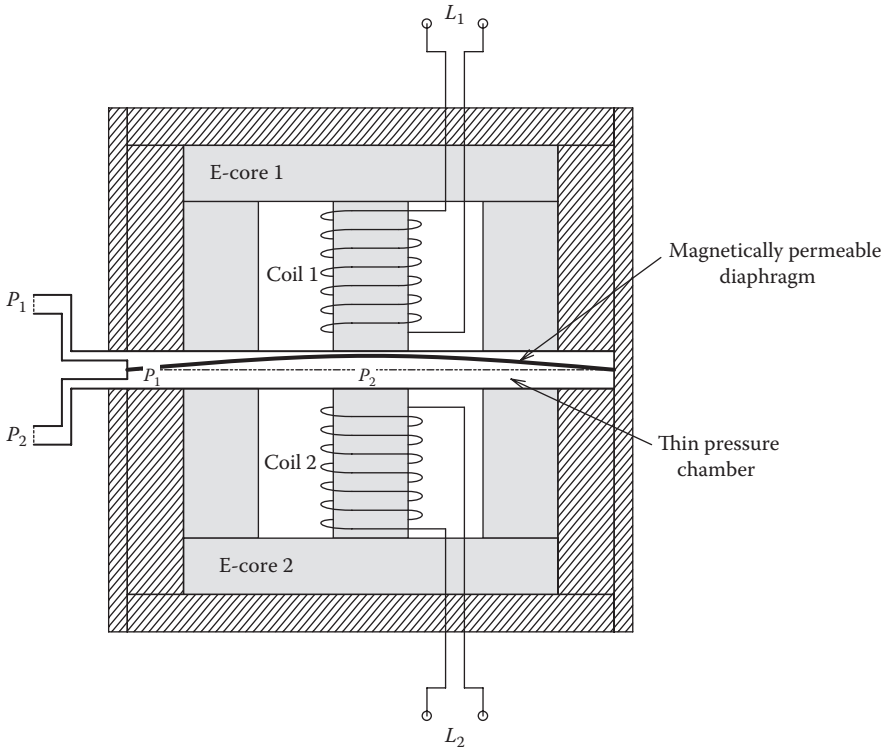
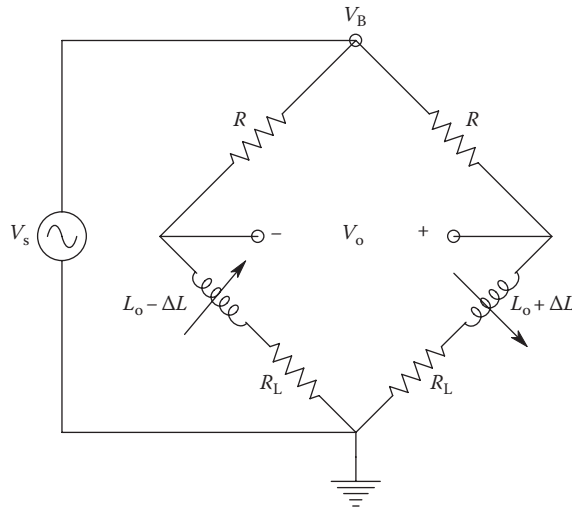


FIGURE 7.64 Section through a differential pressure sensor made by Validyne Engineering Corp. Sensitivity can be altered in the field by changing diaphragms. Diaphragm deflection caused by ΔP alters the magnetic flux linking the coils, causing the top coil's inductance to increase by some small ΔL , while the lower coil's inductance decreases by ΔL .

**FIGURE 7.65**

Simple AC inductance bridge that can be used to give an output proportional to ΔL . The frequency of the bridge excitation, V_B , is made $\omega = (R_L + R)/L_o$ r/s.

decrease in the reluctance of the flux path of the second coil. Coil inductance, L , is inversely proportional to the reluctance of the magnetic flux path, so we see that a small bowing of the diaphragm due to ΔP will cause a corresponding $+\Delta L$ in one coil and a $-\Delta L$ in the other. This inductance change is symmetrical and can be sensed with a simple AC bridge, such as shown in Figure 7.65. It is left as an exercise to show that when the bridge sinusoidal excitation frequency is $\omega_o = (R + R_L)/L_o$ r/s, the bridge output voltage is given by

$$\frac{|V_o|}{V_B} \cong \frac{\Delta L}{L_o} \left[\frac{R}{R + R_L} \right]. \quad (7.213)$$

The bridge output is a DSBSC signal and can be demodulated by a PSR and an LPF. The output of the PSR/LPF is seen to be linearly proportional to ΔP across the transducer diaphragm. For greater range, stiffer (thicker) diaphragms are used. Diaphragms are available to give full-scale ranges from 0.0125 to 12,500 psi. Excitation voltage is typically 5 Vrms, 3–5 kHz. Linearity and hysteresis are typically $\pm 0.5\%$.

A pressure sensor design that is more commonly encountered than the types described earlier uses four strain gauges bonded to the protected side of a deflecting diaphragm. In such a full Wheatstone bridge design, two of the bridge arms are configured to increase resistance with diaphragm strain and two to decrease resistance. The close proximity of the strain gauges affords nearly complete temperature compensation. In other designs, only two gauges are active, increasing resistance with strain. The other two are inactive and in close proximity to the active gauges, so a temperature-compensated output can be realized. There are a number of miniature pressure sensor designs that make use of IC fabrication techniques. In the design described by IC Sensors, Inc., Milpitas, CA, four piezoresistive strain gauges are diffused into the surface of a single silicon crystal diaphragm to form a fully active Wheatstone bridge. The single-crystal diaphragm has negligible hysteresis. (Measurement Specialties™, Inc., acquired IC Sensors, Inc., from Perkin-Elmer, Inc., in 2000.) Measurement Specialties™ offers an EPIH series of miniature pressure sensors

using a fully active, four-arm Wheatstone bridge of piezoresistive material bonded to a silicon diaphragm. In the EPIH-17 sensor, the body diameter is 1.27 mm, and the body length is 11.4 mm; the reference pressure port extends 6.4 mm behind the sensor. Full-scale pressure ranges from 5 to 300 psi, and the corresponding mechanical resonance frequencies range from 500 kHz to 1.7 MHz. Useful mechanical frequency response is given as 20% of the resonance frequency. Ultraminiature sensors such as the EPIH series can be mounted in the tips of catheters and used invasively to measure pressures in the chambers of the heart, blood vessels, etc., in physiological studies, in wind tunnel tests, etc.

Another approach to pressure sensor design has been devised by Motorola Semiconductors, Phoenix, AZ. Motorola's MPX IC series of pressure sensors do not use a conventional, four-active-arm Wheatstone bridge. Instead, a single, *p*-type, diffused silicon strain resistor is deposited on an etched, single-crystal silicon diaphragm, as shown in Figure 7.66. A second resistor is deposited to form an X-shaped, 4-terminal resistor with two current taps (1 and 3) and two output voltage taps (2 and 4). When current is passed through terminals 1 and 3, and pressure is applied at right angles to the current flow, a transverse electric field is generated, producing an EMF between pins 2 and 4. Motorola suggests that this effect can be considered to be an electromechanical analog of the Hall effect, with pressure replacing the magnetic field. In the Motorola MPX3100 sensor series, a complete signal-conditioning system with four op amps and laser-trimmed resistors has been built on the margin of the silicon wafer holding the diaphragm and the transverse voltage strain gauge (see Figure 7.67A). The output of the MPX3100 is quite linear, $\pm 0.2\%$ FS. Temperature compensation of full-scale span is $\pm 2\%$ FS over a 0°C – 85°C range. In the design of the MPX2000 series pressure sensors, five laser-trimmed resistors and

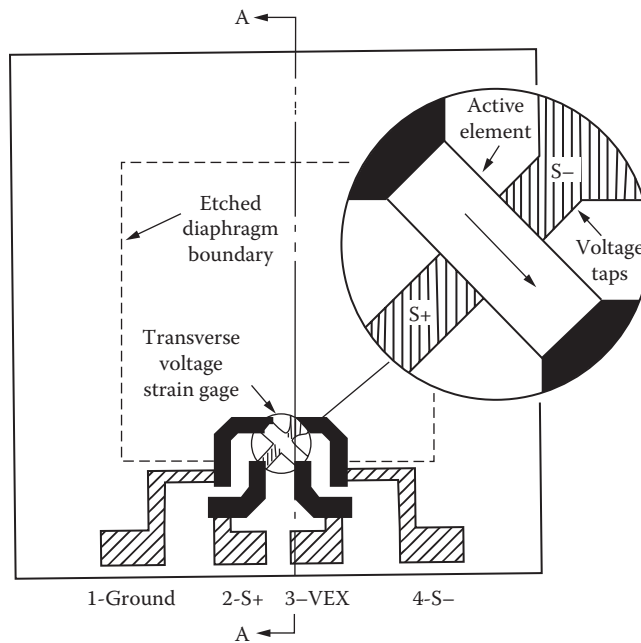
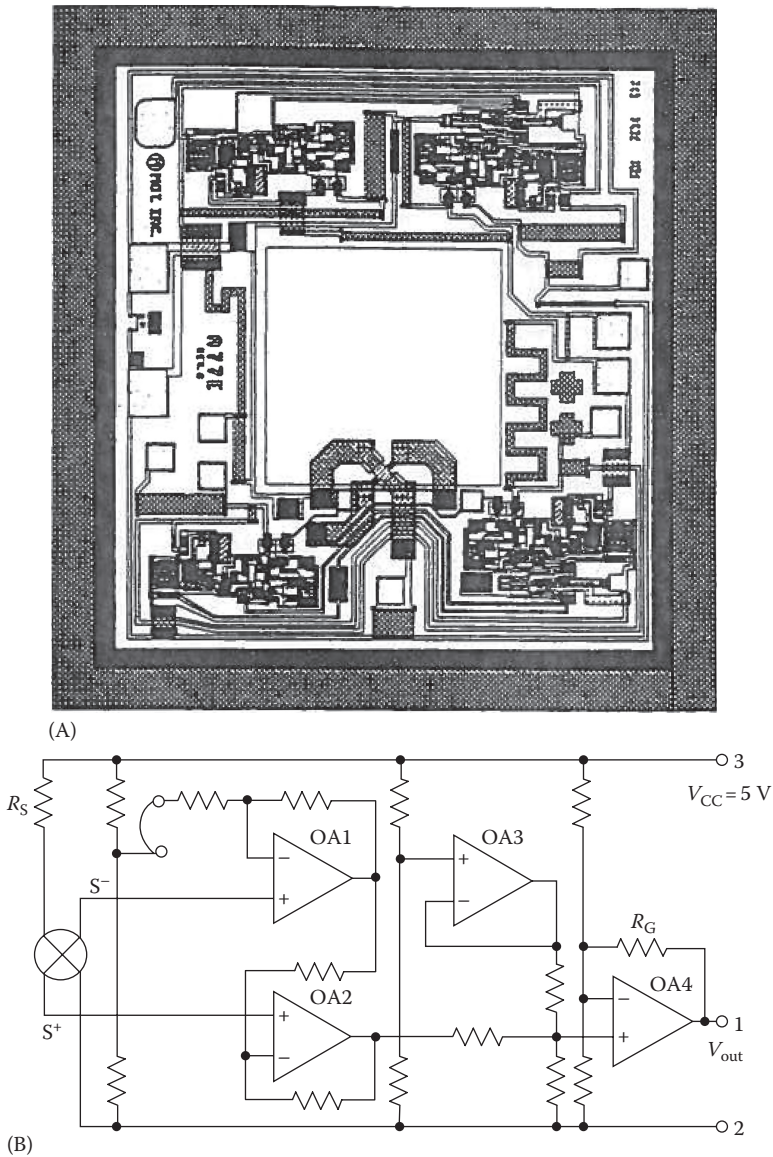


FIGURE 7.66

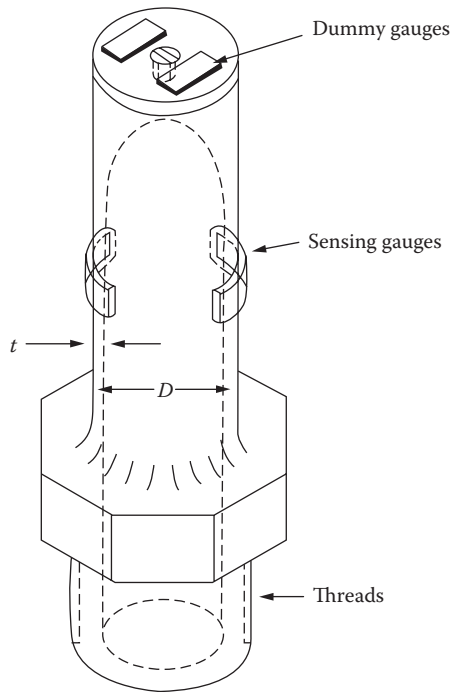
Diagram of the etched microcircuit diaphragm of a Motorola MPX series, MEM pressure sensor. The patented X-shaped structure is a four-terminal resistor with two voltage and two current taps. Motorola claims that the sensor is the electromechanical analog of a Hall effect device. (Courtesy of Motorola Semiconductors, Schaumburg, IL.)

**FIGURE 7.67**

(A) Top view of the IC of a Motorola MPX3100 pressure sensor. (B) Simplified schematic of the temperature-compensated MPX3100 pressure sensor. (Courtesy of Motorola Semiconductors, Schaumburg, IL.)

two thermistors are deposited on the margin of the silicon chip around the pressure diaphragm. These resistors and thermistors are used to achieve a $\pm 1\%$ FS temperature effect over a 0°C – 85°C span. A similar temperature range on one of the non-thermally compensated sensor models, such as the MPX10, would be a whopping $\pm 16.2\%$ FS. The simplified circuit schematic for a MPX3100 pressure sensor is shown in Figure 7.67B.

Measurement of *extremely high pressures* in the range of 10,000 to over 250,000 psi (6.895×10^4 kPa to $>1.724 \times 10^6$ kPa) requires special transducer designs. The simplest is a cylindrical pressure cell, illustrated in Figure 7.68. Two active bonded strain gauges respond to

**FIGURE 7.68**

Cylindrical high-pressure sensor cell that uses four bonded strain gauges in a Wheatstone bridge; the two gauges on the end of the cell are inactive and are used for temperature compensation. (From Beckwith, T.G. and Buck, N.L., *Mechanical Measurements*, Addison-Wesley, Reading, MA. With permission.)

the outer surface's hoop stress caused by pressure difference across the cylinder's wall. Two reference strain gauges are located on the top of the cylinder. It may be shown that the outer surface's hoop strain is given by

$$\epsilon_H = \frac{P_i d^2 (2 - \mu)}{Y(D^2 - d^2)} \text{ in./in.} \quad (7.214)$$

where

Y is Young's modulus

D is the cylinder's outer diameter

d is its inner diameter

μ is Poisson's ratio

P_i is the internal pressure (Beckwith and Buck 1961)

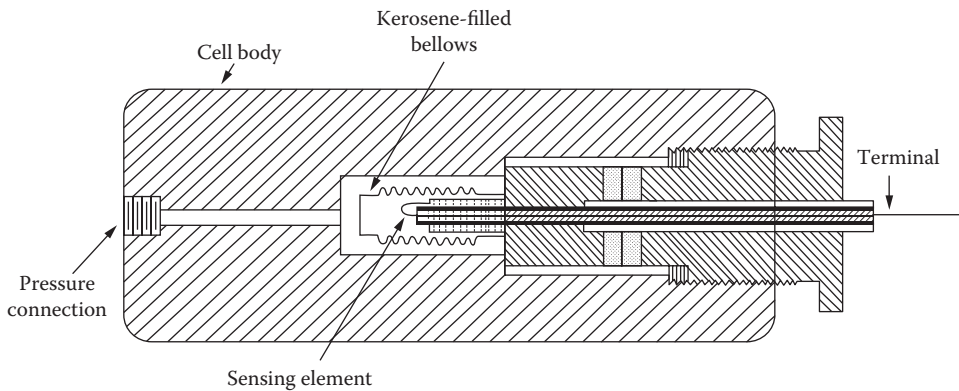
A different sensor mechanism that is used to measure extremely high pressures makes use of the fact that the electrical resistance of a conductor changes if it is compressed by an applied, external pressure. Such sensors are called bulk compression gauges (see Figure 7.69). The resistance of the compressed conductor may be approximated by

$$R(P) = R_o(1 + bP) \quad (7.215)$$

where

R_o is the resistance of the conductor at 1 atm and at standard temperature

b is the pressure coefficient of resistance

**FIGURE 7.69**

Section through a very high-pressure, bulk modulus compression pressure sensor. The sensing resistor is made one arm of a Wheatstone bridge. (From Beckwith, T.G. and Buck, N.L., *Mechanical Measurements*, Addison-Wesley, Reading, MA. With permission.)

Two metals commonly used for bulk compression gauges are an alloy of gold and 2.1% chromium and manganin. The gold alloy is preferred because of a much lower tempco than manganin ($\Delta R/R = 0.01\%$ for a 21.1°C – 82.2°C change in temperature for the gold alloy vs. 0.2% for manganin). The $(\Delta R/R)/P$ for the gold alloy is 0.673×10^{-7} ohm/ohm/psi and is 1.692×10^{-7} ohm/ohm/psi for manganin. Although the gold alloy has a lower pressure sensitivity, its lower tempco makes its use preferred in bulk compression gauges (Beckwith and Buck 1961, Ch. 12). A bulk compression pressure sensor that works up to 1.5×10^6 psi was described by Hall (1958). Bulk compression pressure sensors must be temperature compensated. Technical difficulties are encountered at very high pressures in bringing the insulated resistance element's leads out of the case without leaks or plastic flow problems.

The final type of pressure sensor that we will describe in this chapter is the quartz crystal, piezoelectric gauge. As we have seen in Section 6.3.3, quartz transducers are only suitable for measuring transient and time-varying pressure changes that contain frequencies above their low-frequency (high-pass) pole. Quartz transducers do not respond to static (constant) pressures. Because quartz pressure sensors work well at elevated temperatures and respond to transient changes in pressure with frequency responses that typically extend to hundreds of kilohertz, they are often used to measure the transient pressure changes associated with combustion in internal combustion engines and the transient pressure associated with internal ballistics of firearms (peak chamber pressures in military and hunting rifles are typically 50–65,000 psi). Some quartz dynamic pressure sensors are water cooled to prevent the quartz crystal from reaching its Curie temperature and failure in certain high-temperature applications such as measurement of cylinder pressure in a running internal combustion engine. The output of a quartz piezoelectric pressure sensor is generally conditioned by a charge amplifier (see Section 2.6.1). Kistler Instrument Corp., Amherst, NY, makes a wide selection of quartz pressure sensors. For example, the Kistler model 6205 sensor is specifically designed for ballistics research and ammunition testing. It has a working range of 0–75,000 psi, a charge sensitivity of -0.09 pC/psi, a resonant (upper) frequency of 300 kHz, a rise time of $1.5 \mu\text{s}$, nonlinearity and hysteresis of $\leq 1\%$ FS, and an operating temperature range of -50°C to 200°C . The Kistler model 6213 piezoelectric pressure sensor has a maximum range of 145,000 psi, and their

special Z-series of sensors operate over an extended temperature range of -196°C to 400°C . PCB Piezotronics, Inc., Buffalo, NY, also makes a full line of quartz pressure sensors, accelerometers, and load cells.

7.5.3 Low-Pressure Sensors

Here, we define low-pressure sensors as those sensors designed to measure pressures significantly lower than atmospheric. Their primary application is in vacuum system measurements. Two means of measuring low pressures do not have electrical outputs: the *McLeod gauge*, shown in Figure 7.70, is a mercury-filled glass system that is operated manually and read out visually. Its useful range is typically 5–0.005 mmHg pressure. The McLeod gauge makes use of Boyle's law: $P_1 = P_2 (V_2/V_1)$. P_1 is the pressure we desire to measure; P_2 is the increased pressure in the McLeod gauge obtained by compressing the gas from volume V_1 to a smaller volume, V_2 . The gas whose pressure is under measurement should not

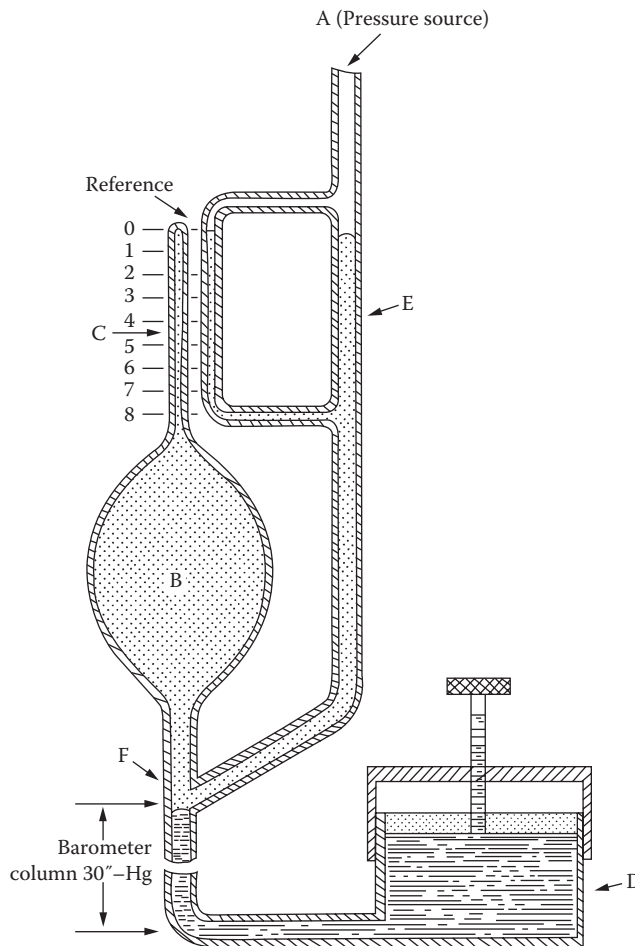


FIGURE 7.70
Section through a McLeod vacuum gauge. Mercury is the fluid used.

contain water vapor, or the compression process will cause condensation. Beckwith and Buck (1961) described the operation of a McLeod gauge:

Measurement is made as follows. The unknown pressure is connected to the gauge at point A, and the mercury level is adjusted to fill the volume represented by the darker shading. Under these conditions the unknown pressure fills the bulb B and capillary C. Mercury is then forced out of the reservoir D, up into the bulb and reference column E. When the mercury level reaches the cut-off point F, a known volume of gas [V_1] is trapped in the bulb and capillary.

The mercury level is then further raised until it reaches the zero reference point in E. Under these conditions the volume remaining in the capillary is read directly from the scale, and the difference in heights of the two columns is the measure of the trapped pressure [P_2]. The initial pressure may then be calculated by use of Boyle's law.

Another directly indicating pressure gauge was the venerable GE *molecular vacuum gauge*. The GE gauge was calibrated over a 0.002–20 mmHg range. A synchronous motor drove a fan in a cylindrical chamber that is at the pressure under measurement. The gas in the chamber was given an angular momentum by the fan, and its swirling molecules struck the blades of an adjacent, restrained fan connected to a pointer and a helical spring. The higher the gas pressure in the chamber, the higher the density of molecules striking the pointer's fan, and the more torque produced on it, hence the greater the deflection of the pointer. The GE gauge's scale was nonlinear, with the highest resolution at the low end of the scale (0–0.1 mmHg covered the first 90° of meter deflection, while 0.1–20 mmHg covered the remaining 150° of scale).

It is of course important to have low-pressure sensors that can be read out electrically. The first and simplest electrical vacuum gauge is the *Pirani gauge*. The Pirani gauge consists of a filament of platinum, tungsten, or nickel in a glass bulb connected to the vacuum system. Various gasses have different heat conductivities, so the Pirani gauge must be calibrated for the gas or gasses being used. A simple operating mode of the Pirani gauge is to place it in one arm of a Wheatstone bridge circuit, as shown in Figure 7.71. An identical gauge in vacuo

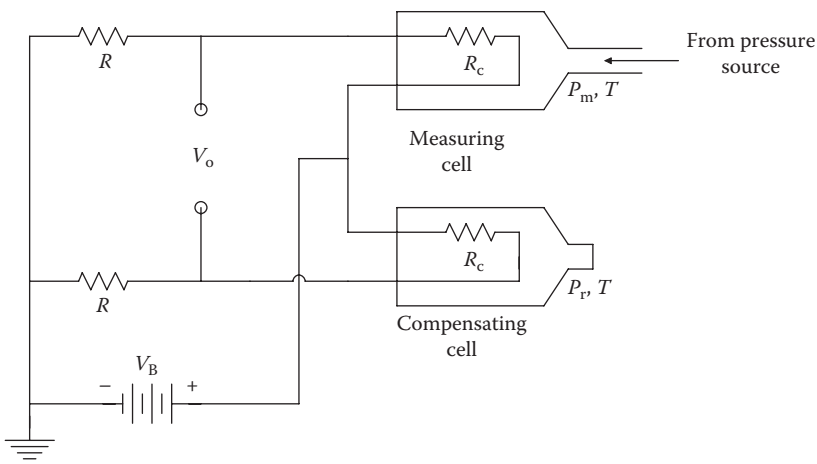


FIGURE 7.71
Schematic diagram of a Pirani low-pressure sensor.

is used as a temperature reference. Recall from Sections 7.2.2.4 and 7.3.2.1 on HWAs that in a vacuum, there will be a rise in the Pirani gauge's filament temperature given by

$$\Delta T = \frac{I_o^2 R_o \Omega_o}{(1 - \alpha I_o^2 \Omega_o)} \quad (7.216)$$

where

I_o is the RMS current through the Pirani filament

R_o is the resistance of the filament at ambient (reference) temperature

α is the tempco of the wire

ΔT is the equilibrium increase in temperature of the wire above ambient temperature

Ω_o is the thermal resistance of the filament in vacuo, in degrees Celsius per watt

We also showed in Section 7.2.2.1 that the resistance of the hot filament in vacuo is given by

$$R(\Delta T, 0) = \frac{R_o}{(1 - \alpha I_o^2 \Omega_o)}. \quad (7.217)$$

Now, as the gas pressure is increased from zero, filament heat is no longer only dissipated by radiation and conduction through the leads but also by heating gas molecules and giving them increased KE. Thus, as the pressure increases, Ω decreases from its vacuum value. At low pressures (below 10 mmHg, Lion [1959]), the thermal resistance may be given by

$$\Omega(P) = \Omega_o(1 - \sigma P) \quad (7.218)$$

where σ is dependent on system geometry, filament material, and the heat conductivity of the gas.

If Equation 7.235 is substituted into Equation 7.234, then we see that the Pirani gauge filament resistance may be approximated by

$$R(P) = R(\Delta T, 0) \left[1 - \frac{\alpha I_o^2 \Omega_o \Delta P}{1 - \alpha I_o^2 \Omega_o} \right]. \quad (7.219)$$

The linear working range of most Pirani gauges is from 10^{-5} to 1 mmHg, although modifications have allowed measurements of pressures as low as 5×10^{-9} mmHg. Of interest is the closed-loop, feedback operation of the Pirani gauge. In this mode of operation, the Wheatstone bridge is balanced at some reference pressure, P_r . As P increases from P_r , the thermal resistance decreases, and the Pirani filament cools and its temperature and resistance decrease. Decreasing resistance causes the bridge output to increase. This increase in V_o is applied to increase the voltage across the bridge, thereby increasing the power dissipation in the Pirani filament and thereby increasing its temperature and resistance, restoring bridge balance. The other resistors in the bridge are assumed to have very low thermal resistances and tempcos.

Another low-pressure (vacuum) sensor makes use of the heater/VTC device introduced in Section 6.3.1 as a means of measuring true rms current. In the vacuum version of this

sensor, the thermal resistance of the heater decreases with increasing gas pressure. This is because at increased gas pressure, there are more gas molecules present to absorb heat energy from the heater and convey it to the walls of the glass envelope where it is lost by radiation and convection. Hence, the heater temperature drops (at constant power input) as the gas pressure increases, and thus, the EMF of the associated thermocouple also decreases with the pressure increase. Thermocouple vacuum sensors work in the range from 10^{-3} to 10 mmHg. Below pressures of 10^{-3} mmHg, heat loss through the heater wire supports exceeds that from the gas molecules, and the thermocouple pressure sensor loses linearity and sensitivity. VTC pressure sensors are ideally suited for feedback operation; one such closed-loop system is shown schematically in Figure 7.72. In this system, feedback is used to nearly match the DC EMF from the pressure-responsive thermocouple to the DC EMF from the reference (vacuum) thermocouple. For the reference thermocouple, we have a heater temperature rise given by

$$\Delta T_r = I^2 R_W \Omega_o. \tag{7.220}$$

The thermocouple EMF is proportional to ΔT_r :

$$V_r = K_T (I^2 R_H \Omega_o). \tag{7.221}$$

For the pressure-responsive thermocouple, we have

$$V_p = K_T [I_o^2 R_H \Omega_o (1 - \sigma P)]. \tag{7.222}$$

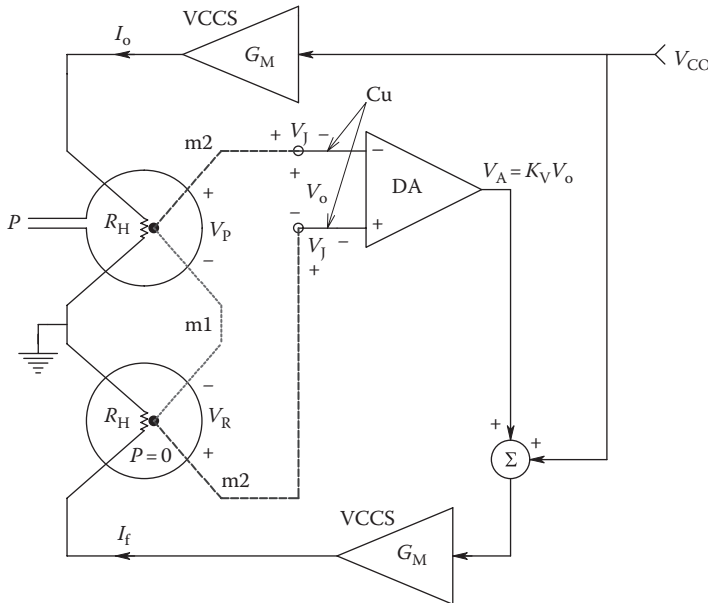


FIGURE 7.72 Schematic diagram of a feedback VTC low-pressure sensor. The two thermocouples must have matched characteristics. The amplifiers, G_M , are VCCSs; DA is a differential voltage amplifier.

Here, we assume that at very low pressures, the thermal resistance of the heater decreases linearly with increasing pressure, P . The current in the pressure-responsive thermocouple heater is constant:

$$I_o = G_M V_{co} \quad (7.223)$$

where

G_M is the transconductance of the two VCCSs

V_{co} is a DC voltage

Thus, the current in the reference thermocouple's heater is

$$I = G_M(V_{co} - K_V V_o) \quad (7.224)$$

where

K_V is the DA gain

V_o is the difference between the reference thermocouple's EMF and the pressure-responsive thermocouple's EMF

That is,

$$V_o = V_r - V_P. \quad (7.225)$$

If the aforementioned relations are substituted into Equation 7.242, we obtain a quadratic equation in V_o :

$$0 = V_o^2 - V_o \frac{1 + 2V_{co}\mu}{\mu K_V} + \frac{P\sigma V_{co}^2}{K_V^2} \quad (7.226)$$

where

$$\mu = K_V K_T R_W \Omega_o G_M^2. \quad (7.227)$$

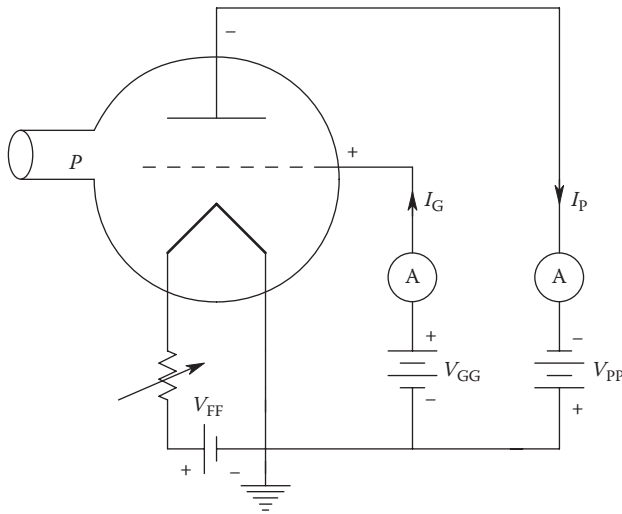
If we assume that $2V_{co}K_V K_T R_W \Omega_o G_M^2 \gg 1$, then one root of the quadratic equation can be shown to be

$$V_o = \frac{P\sigma V_{co}}{2}. \quad (7.228)$$

The DA's output is

$$V_A = \frac{P\sigma K_V V_{co}}{2}. \quad (7.229)$$

Thus, the proposed design for the feedback thermocouple, low-pressure sensor system is directly proportional to the pressure. Note that the calibration depends on the value of the constant, σ , which will depend on the composition of the gas. Hydrogen would have a larger σ than would air, for example.

**FIGURE 7.73**

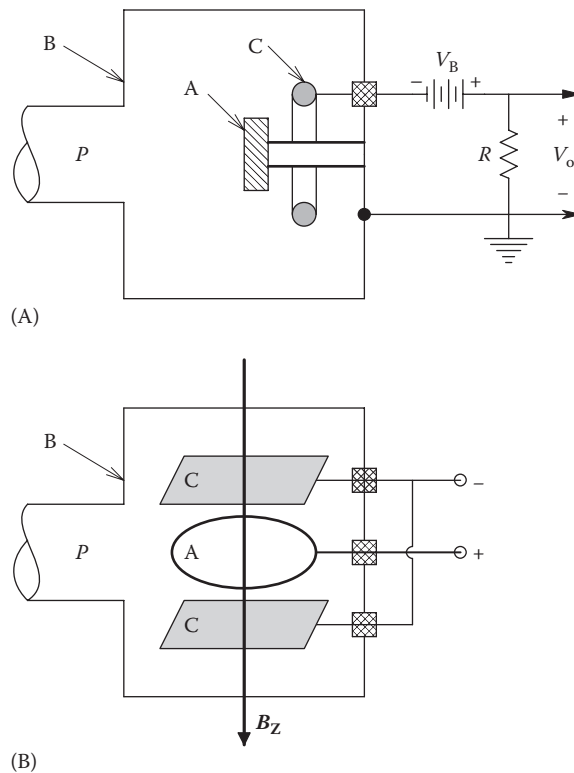
Vacuum triode ionization gauge used to measure very low pressures ($10^{-8} \leq P \leq 10^{-3}$ mm Hg).

The *ionization vacuum gauge* is routinely used to measure very low pressures, in the range of 10^{-8} – 10^{-3} mmHg. Special designs have been used down to 10^{-10} mmHg (Lion 1959, Sect. 1-57). The ionization gauge is constructed similar to a triode vacuum tube, as shown in Figure 7.73. A glass envelope is kept at the pressure under measurement. In the envelope is a heated cathode that emits electrons. Surrounding the cathode is a grid, and outside the grid is a plate. Unlike a vacuum tube triode, the ionization detector's grid is maintained at a high (100–250 V) positive potential with respect to the cathode. The plate is kept at a negative potential (–2 to –50 V). Electrons from the hot cathode are accelerated by the field from the grid; they collide with gas molecules and ionize them. The + ions in the space between the grid and the plate are collected by the negative plate; those formed in the space between the cathode and the grid are accelerated toward the cathode. The free electrons and negative ions are collected by the positive grid. The rate of ion production is proportional to the density of gas molecules in the tube (or to the gas pressure) and to the number of electrons available to ionize the gas. Lion (1959) showed that the gas pressure is proportional to the ratio of the positive ion current at the plate to the grid current. Thus,

$$P = K \left(\frac{i_p^+}{i_g} \right) \quad (7.230)$$

where K is about 10^{-2} mmHg(mA/ μ A). K depends on the gas composition, the geometry of the tube, and the voltages used. The emission of photoelectrons from the plate, caused by light and soft x-rays from the grid, sets the lower bound of ion gauge sensitivity. One mode of operation of the ion gauge is to keep the grid current constant (1–20 mA) and to measure the plate current. A log ratio I_c (cf. Section 2.5.7) can also be used to measure P over the wide range of this sensor.

Instead of creating an ion current by collisions with energetic electrons, a radioactive, α particle emitter can be used to make a radioactive ionization gauge, illustrated in

**FIGURE 7.74**

(A) Radioactive ionization vacuum sensor. Note: A = α particle-emitting radioisotope. C = positive ion collector electrode (cathode). B = metal housing (anode). (B) Diagram of a Philips–Penning vacuum sensor in which C = cathodes, A = anode. (From Lion, K.S., *Instrumentation in Scientific Research: Electrical Input Transducers*, McGraw-Hill Book Co., Inc, New York, 1959.)

Figure 7.74A. The number of + ions formed is proportional to the gas pressure as long as the range of the α particles is longer than the dimensions of the chamber. The useful range of pressures is 10^3 – 10^{-3} mmHg. Below 10^{-3} mmHg, the mean free path of the α particles increases beyond the dimensions of the chamber so that the probability of an ionizing collision with a gas molecule is reduced and the ion current drops off sharply. One advantage of the radioactive ionization gauge is that it is not damaged if it is turned on at atmospheric pressure, and it requires no degassing.

The final low-pressure sensor we will describe in this section is the *Philips–Penning gauge*, illustrated in Figure 7.74B. In this gauge, collisions between gas molecules and moving electrons create an ion plasma, which contributes to the gauge current. Electrons are emitted as a result of the high electric field and bombardment of the electrode surfaces by ions. The working range of the Philips–Penning gauge is 4×10^{-7} to 10^{-3} mmHg (Lion 1959). The electrodes of this gauge are run at potentials of 1–3 kV, and a transverse, DC magnetic field 300–8000 Oe is used to decrease the mean free path of electrons and thus extend the low-pressure range. One version of the Philips–Penning gauge, the inverted magnetron, was reported to work at pressures as low as 10^{-12} mmHg (Lion 1959).

7.6 Introduction to Substance Detection and Measurement Using Photons

7.6.1 Introduction

In this section, we will describe various methods of detecting the presence of and measuring the concentration or mass of certain chemicals (analytes) using various analytical techniques. The first technique we will examine is based on the substance's molecular responses to applied EM radiation, including wavelength-specific energy absorbance (or transmission) and fluorescence (stimulated photon emission). The EM wavelengths can range from the FIR to the far ultraviolet (UVC).

Figure 7.75 illustrates the EM spectrum. The (photon) wavelength range used for spectrometry spans $430\ \mu\text{m}$ – $430\ \text{pm}$ going from FIR to UVC. When EM waves in this wavelength range interact with matter of finite thickness, some fraction R of the incident energy is reflected, a fraction A is absorbed as the radiation passes through, and a fraction T is transmitted. It should be obvious that $R + A + T = 1$. In using light to measure how much of a substance is present, we generally ignore R and examine A or T as a function of wavelength, that is, $A(\lambda)$. The *transmittance* of a substance through which *photon* EMR passes is defined by

$$T(\lambda) \equiv \frac{P_{\text{out}}}{P_{\text{in}}} < 1 \quad (7.231)$$

where

P_{in} is the radiant input power to the substance

P_{out} is the emerging radiant power

The transmittance is a function of the wavelength of the incident light in space. The *absorbance*, A , also known as *OD* is defined as

$$A \equiv OD \equiv -\log_{10} [T(\lambda)] = \log_{10} \left[\frac{1}{T(\lambda)} \right]. \quad (7.232)$$

The optical absorbance as a function of wavelength or wave number is a unique, physical-chemical *fingerprint* of the analyte. The particular absorbance peaks are caused by the interaction of incident photons having energy $h(c/\lambda)$ of certain molecular bonds. Molecular subgroups such as $-\text{H}$, $-\text{OH}$, $-\text{CH}_3$, and $-\text{NH}_3$ are attached to other larger molecular units such as benzene rings, purine rings, sugars, and fatty acids. All chemical bonds are elastic and allow the joined molecules to move in complex vibrational modes when they absorb low-energy photons having their mechanical natural frequencies. Stretching, twisting (rotating), waving, rocking, etc., are some of the complex elastic motions possible. When the photons striking a particular molecular bond have the correct energy to excite a resonant mode of the bond, they are selectively absorbed, causing enhanced bond resonance. Thus, the EMR exiting the sample has absorbance peaks at the wavelengths (energies) of the photons that excite molecular resonances. Table 7.2 illustrates the types of photon/molecular interactions as a function of wavelength and wave number (Northrop 2002).

Beer's law is often used to model absorbance; see Figure 7.76 for an illustration. First, note that a small fraction of the impingent radiant power, $R = R_{\text{inC}} + R_{\text{inS}} + R_{\text{outS}} + R_{\text{outC}} + R_{\text{PMT}}$ is

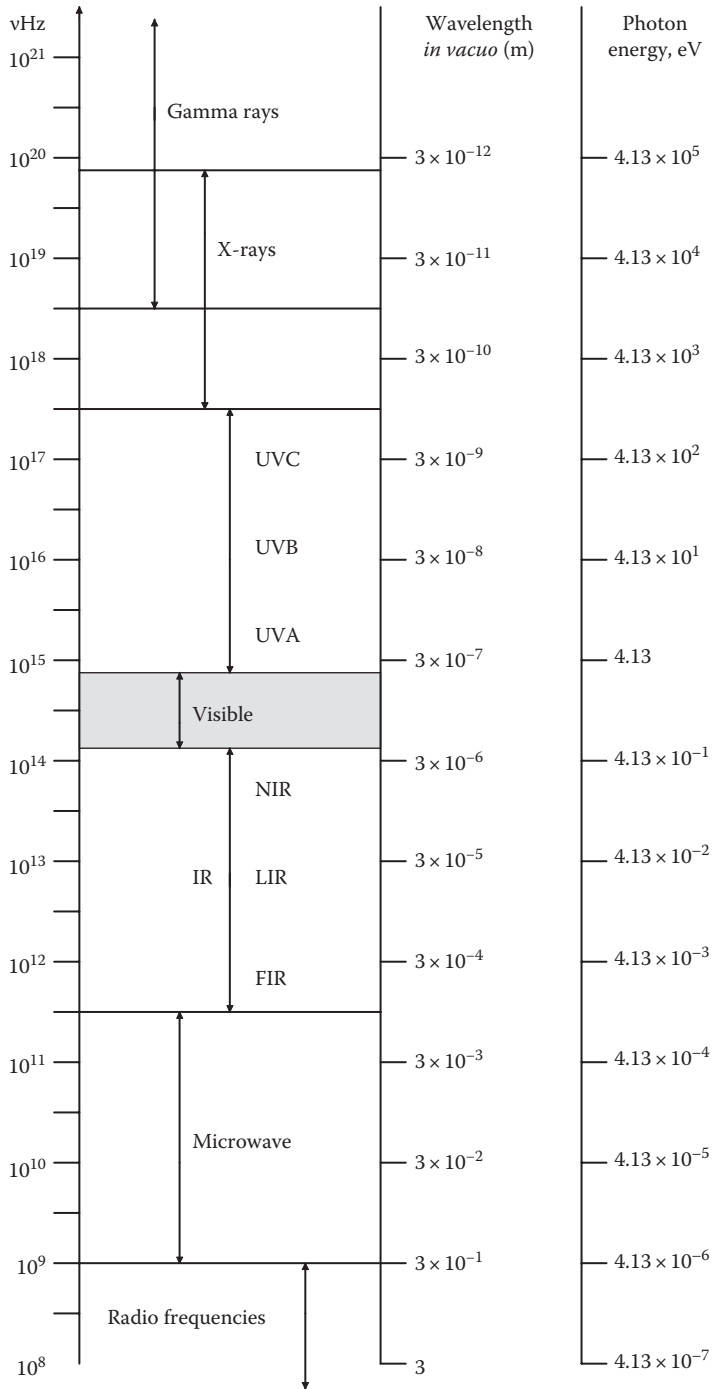


FIGURE 7.75
EM spectrum.

TABLE 7.2

Wavelength λ , Wave Number ξ , Energy Levels, and Type of Molecular Energy Transition When Photons Are Absorbed by Molecules

λ Range	ξ Range, cm^{-1}	Type of Radiation	Energy Transition	$h\nu/q$ Range, eV
100–10 cm	0.01–0.1	VHF, μ wave radio	Spin orientation	1.24E–5 to 1.24E–4
10–1 cm	0.1–1	μ wave	Molecular rotations	1.24E–4 to 1.24E–3
1–0.1 cm	1–10	μ wave, FIR	Molecular rotations	1.24E–3 to 1.24E–2
100–10 μm	10–100	FIR, MIR	Molecular vibrations	1.24E–2 to 1.24E–1
10–1 μm	100–10 ³	MIR, NIR	Molecular vibrations	1.24E–1 to 1.24
10 ³ –100 nm	10 ³ –10 ⁴	Visible, UV	Valence e-transitions	1.24 to 1.24E1
100–10 nm	10 ⁴ –10 ⁵	UVC (vacuum UV)	Valence e-transitions	1.24E1 to 1.24E2
10–1 nm	10 ⁵ –10 ⁶	UVC, x-rays	Inner shell e-transitions	1.24E2 to 1.24E3
1–0.1 nm	10 ⁶ –10 ⁷	X-rays	Inner shell e-transitions	1.24E3 to 1.24E4
<0.1 nm	>10 ⁷	γ -rays	Nuclear transitions	>1.24E4 eV

reflected at the air/cuvette interfaces, the cuvette/analyte interfaces, and the air/PMT interface where there is a difference in refractive index. The actual spectral radiant power absorption is assumed to take place in the medium of interest (the analyte) inside the cuvette. Cuvettes can be made of glass, quartz, plastic, or exotic materials such as germanium metal or IRTRAN™ depending on the range of wavelengths used. Beer's law states that the rate of decrease of radiant power with distance as light passes through a homogeneous medium is proportional to the concentration C of the absorbing substance (for solutions) times the power. In mathematical terms,

$$-\frac{dP}{dx} = \alpha'PC \quad (7.233)$$

↓

$$-\int_{P_{in}}^{P(L)} \frac{dP}{P} = \int_{x=0}^L \alpha' C dx \quad (7.234)$$

↓

$$\log_{10} \left[\frac{P_{in}}{P(L)} \right] = \alpha CL, \quad \alpha = \frac{\alpha'}{2.303}. \quad (7.235)$$

Thus, it is clear that $OD \equiv \alpha CL$ and $\%T \equiv 10^{[-\alpha(\lambda)CL+2]}$. Note that α , the *extinction coefficient*, also known as the *absorptivity*, of the solution, is a function of wavelength with many peaks and valleys that can be related to photon energy being absorbed by aspects of the analyte's molecular structure. The absorptivity is the analyte's *spectrometric fingerprint*.

Note that in Figure 7.76, the monochromator was represented as a prism. Modern spectrometers generally use one or two *diffraction gratings* and slits to obtain a narrow wavelength beam of light (high full power/half width or Q) for analysis. For example, the Czerny–Turner single-grating monochromator uses a plano grating and two concave focusing mirrors (cf. Figure 7.77). A detailed mathematical treatment of grating monochromator

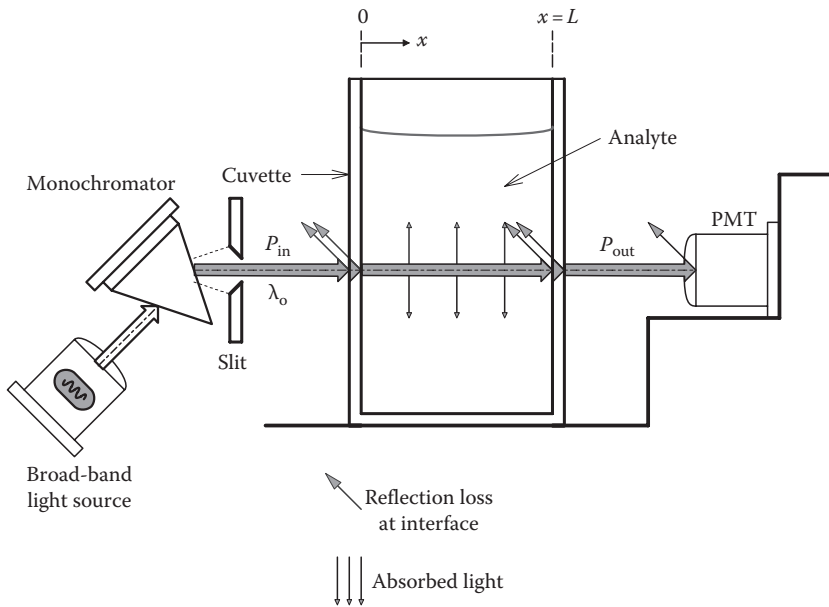


FIGURE 7.76
Simple single-beam optical system illustrating optical absorption, reflection, and Beer's law.

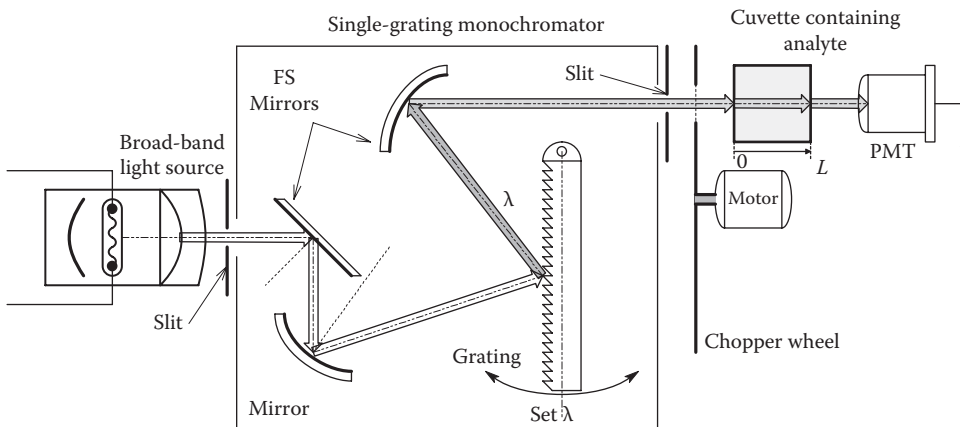


FIGURE 7.77
Single-beam spectrophotometer.

design, dispersion, errors, and aberrations can be found in the Newport *Diffraction Grating Handbook* (2012).

If continuous scanning of a range of the spectrum is not necessary, then a group of LADs emitting at a set of peak wavelengths, $\{\lambda_{k_i}\}$, can be used to quantify known peak absorption wavelengths of the analyte(s).

7.6.2 Dispersive Spectrophotometry

Dispersive spectrophotometry (DS) methods constitute a large number of UV, visible light, and IR spectroscopy systems. DS refers to the fact that the light or EM radiation passed

through the sample (analyte) must pass through a monochromator either before or after the analyte to describe the analyte's absorption/transmission characteristic spectrum. The simplest monochromator is a prism followed by a slit to select a narrow band of wavelengths centered around some λ_o . A beam of *white* light is directed into the prism; a fan-shaped beam of light, dispersed by wavelength, exits the prism. The desired λ_o is selected by rotating the prism. The slit defines the desired wavelength band. Modern monochromators use one or two *diffraction gratings* that consist of a glass plate into which are etched very fine, closely spaced, parallel lines. Refraction, reflection, and interference occur when a beam of white light impinges on a grating, creating an output light beam with a peak λ_o dependent on the incident and exit angles. A grating output generally has a higher Q (peak power wavelength λ_o /half power wavelength width) than a prism, and by using two gratings in series, one can obtain an even more spectrally pure output beam, albeit at the expense of intensity.

Figure 7.77 illustrates a single-beam spectrophotometer. White light from a tungsten or halogen source is directed to a grating. The exit slit further increases the Q of the exit beam, which is passed through a chopper wheel that periodically interrupts the beam. The chopping rate is typically in the hundreds of pps. The chopped beam then passes through a solution of the analyte in a special chamber known as a cuvette. The light path through the analyte in the cuvette is L cm, where L can be 0.1, 1, 2, 10 cm, etc., depending on the absorptivity of the solution. The exiting light power from the cuvette is measured by a calibrated PMT. The PMT DC dark current is rejected by the PSR, which responds only to the AC (chopped) signal from the PMT. The intensity of the light source must be stabilized by using a regulated DC power supply. The test beam's λ_o is selected by varying the grating angle. Unfortunately, as λ_o is varied, so does the intensity of the input beam to the cuvette. To compensate for this effect, a run is made with a cuvette full of distilled water or a reference solution and the $P_{\text{outRef}}(\lambda)$ recorded and used to compensate for the run with the analyte. A phase-sensitive demodulator is used to detect $P_{\text{outAn}}(\lambda)$, and a computer calculates $A(\xi)$, $A(\lambda)$, $T(\xi)$, or $T(\lambda)$.

Figure 7.78 illustrates one version of a *dual-beam spectrophotometer*. This design is more complex optically because it simultaneously measures PMT DC dark current, reference beam transmission, and analyte transmission. A special, 3-state chopper wheel is used that has holes and mirrors in it such that the PMT alternately sees dark, the analyte beam and the reference beam. Sample and hold circuits allow appropriate corrections and comparisons to be made from the 3-state-in-time PMT waveform.

The spectrogram for an analyte is usually plotted as $\%T$ vs. *wave number*, ξ , in cm^{-1} . $\xi \equiv 10^4/\lambda(\mu\text{m}) \text{ cm}^{-1}$. ξ is proportional to the EM wave's frequency, $\nu(\text{Hz}) = c(\text{m/s})/\lambda(\text{m})$. By using a wave number abscissa, $\%T$ peaks crowded together at longer wavelengths are separated and given in more detail.

Figure 7.79 illustrates the use of an *attenuated total reflection (ATR)* plate in spectrophotometry where we wish to measure the wavelength-dependent absorption of backscattered light from thin films of analyte, including superficial tissue (dermis, capillaries, etc.). This approach can be used to measure analytes in liquid blood samples and may find application in measuring capillary blood constituents such as glucose, cholesterol, alcohol, and heroin (diacetylmorphine). The input light enters the ATR prism and is directed through its bottom surface into the absorbing sample. The light that is backscattered from the sample reenters the ATR prism and is totally reflected from its top surface, being directed again into the tissue where it is again backscattered. The repeated reentry and backscattering of the beam from the tissue increase the sensitivity of the spectroscopic process by effectively increasing L in Beer's law. The beam exiting the ATR plate thus has had a multitude of opportunities to interact with analyte molecules in intimate contact with the lower surface and contains the spectral absorption signature of the analyte. Increases in sensitivity of a factor of 20 can sometimes be had.

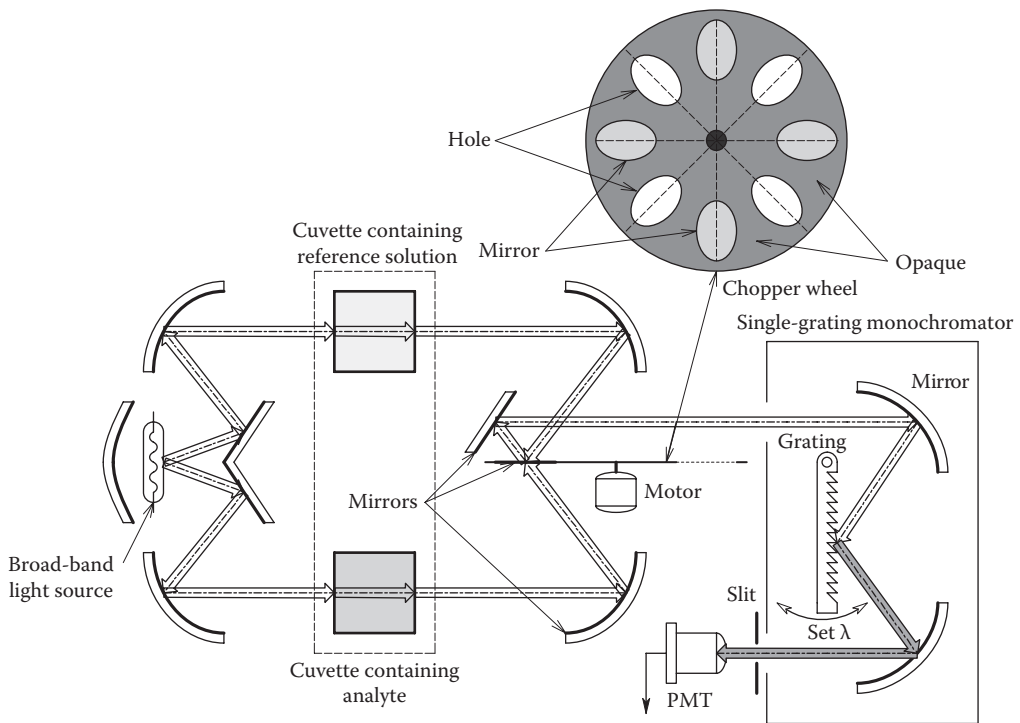


FIGURE 7.78

Dual-beam spectrophotometer. A grating monochromator is used. The mirror/glass/opaque chopper wheel allows the alternate comparison between a reference cuvette's absorption with the analyte's absorption.

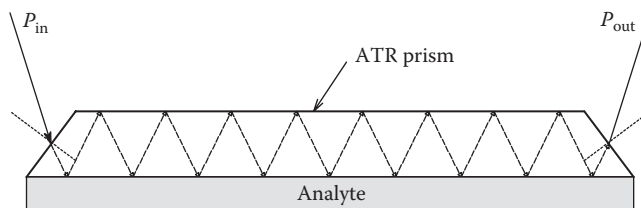


FIGURE 7.79

Side view of an ATR plate or prism. The ATR plate allows multiple beam interactions with the surface of the analyte.

Raman spectroscopy is a special type of dispersive spectroscopy. It is based on the *Raman effect*, named for the Indian physicist, C.V. Raman, who was awarded the 1930 Nobel prize for his discovery. In his seminal paper, published in the journal *Nature* in 1928, Raman observed that when light interacted with matter, some small fraction of its energy was *scattered*, and the frequency of the scattered light was shifted in a complex manner from that of the incident beam (*Stokes* and *anti-Stokes scattering*). Most of the scattered light energy is returned at the input wavelength; this light undergoes elastic, *Rayleigh scattering*.

Since its discovery, the Raman effect has been developed into a sensitive, analytical chemical tool with results that rival conventional spectroscopic techniques (UV, VIS, and IR) (Kaiser Optical Systems 1998, Berger et al. 1999, Princeton Instruments 2013). The advent of

monochromatic laser sources has expedited the rise of Raman spectroscopy as an analytical tool. In the past few years, Raman spectroscopy has been applied to various biomedical applications, such as measuring various important analytes in blood and serum (Frank 1998, Koo et al. 1999, Lambert et al. 2002, Shafer-Peltier et al. 2002). Also, attention has recently been given to developing noninvasive, biomedical Raman systems (Hanlon et al. 2000).

There are two kinds of Raman spectroscopy systems currently in use: *conventional Raman* and *stimulated Raman spectroscopy (SRS)*. Figure 7.80 illustrates a typical, conventional, in vitro Raman system. A high-power diode laser emitting in the NIR, typically at 830 or 850 nm, is used as the source. Its beam is chopped or otherwise amplitude modulated to enable synchronous detection of the scattered light. NIR source light is generally used to minimize any natural, background fluorescence from a complex, biological sample. Note that the shorter

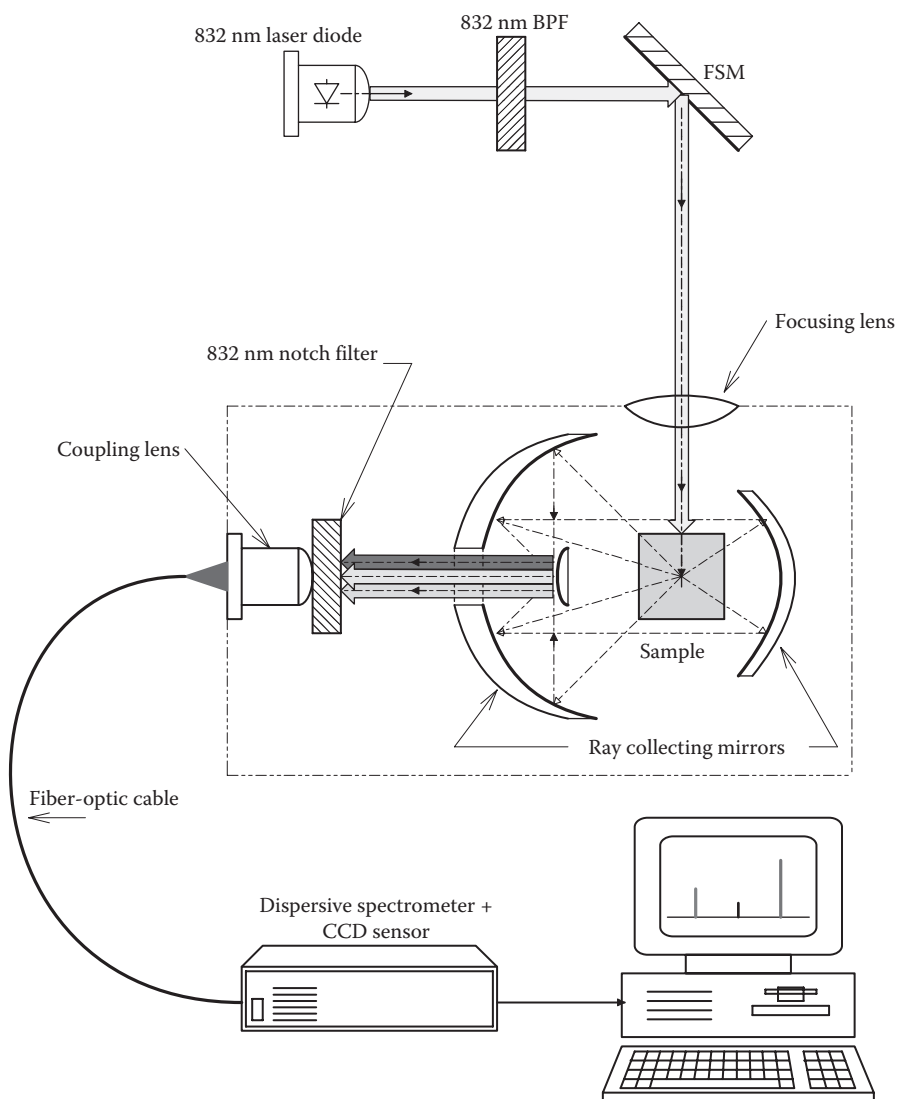


FIGURE 7.80
Diagram of a conventional Raman spectroscope.

visible wavelengths and UV light excite fluorescence. Most of the input light is elastically scattered from the sample (Rayleigh light) with no frequency shift by analyte molecules. Inelastic (Raman) scattering in all directions occurs at about 10^{-6} – 10^{-8} times the intensity of the Rayleigh-scattered light. Energy from the monochromatic input beam is absorbed in the process of exciting vibrations of the various interatomic bonds of the analyte. The Raman scattered (output) light thus has a complex spectrum depending on the chemical structure of the analyte(s). In general, the Raman red shifts in the output spectrum have frequencies given by $(\nu_{\text{in}} - \nu_{\text{Bk}}) = \Delta\nu_{\text{Sk}}$ Hz. ν_{in} is the input (excitation) lightwave frequency; ν_{Bk} is the *natural frequency* of the k th class of atomic bond. $\Delta\nu_{\text{Sk}}$ is also called the *Stokes shift* of the k th bond class. The energy absorbed in the photon excitation of a particular, interatomic bond resonance is proportional to the $\Delta\nu_{\text{Sk}}$ of its spectral peak. That is, $E_{\text{abs}} = h \Delta\nu_{\text{Sk}}$ J. A weaker emission is seen to occur for the *anti-Stokes*, or blue shift, Raman emission:

$$(\nu_{\text{in}} + \nu_{\text{Bk}}) = \Delta\nu_{\text{aSk}} \text{ Hz.}$$

To effectively gather more of the weak, Raman-scattered light, often a parabolic mirror is used near the sample. An *optical notch filter* is also used to attenuate the Rayleigh-scattered light of frequency, ν_{in} . A Hi-Q, holographic, optical notch filter attenuates by ca. 10^{-6} at the center of its stop band and passes ca. 90% of the input light at other wavelengths. The weak Raman light is passed through a dispersive monochromator that either projects the desired Raman spectrum on a linear, CCD sensor array where the k th pixel intensity is proportional to $\Delta\nu_{\text{Sk}}$, or light at a given $\Delta\nu_{\text{Sk}}$ is directed onto a single photomultiplier sensor or some photodiode. For in vitro Raman spectroscopy, relatively large input intensities are used, typically in the hundreds of milliwatts. Such high input power cannot be used for in vivo Raman measurements because of potential heat damage to tissues, including the retina.

Raman spectra are generally displayed in units of *wave number*, $\xi \text{ cm}^{-1}$, which is proportional to the Hz frequency of the light. Note that 1 cm^{-1} equals 3×10^{10} Hz or 30 GHz. For example, the frequency of 830 nm input light is $\nu_{\text{in}} = 3.61 \times 10^{14}$ Hz, and its wave number is $\xi_{\text{in}} = 1.203 \times 10^4$.

Figure 7.81 illustrates both Raman and Fourier transform infrared (FTIR) absorbance spectra from methanol. Notice that in the two spectrograms, some congruent peaks are sharper and stronger than in the other spectrogram.

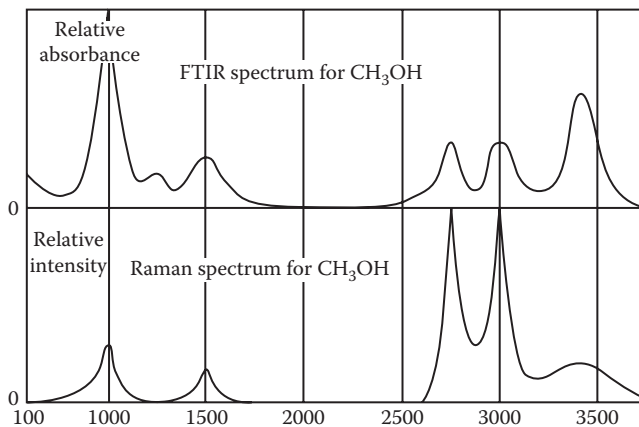


FIGURE 7.81

Comparison of the absorbance spectra of methanol by FTIR and by Raman spectroscopy.

Stimulated Raman spectrophotometry (SRS) is a technique that avoids the nondirectional scattering of the frequency-shifted light emitted in conventional, single-source Raman spectroscopy. In SRS, two lasers are used to excite the analyte: a *pump laser* and a tunable *probe laser*. The intersection of their beams defines an analyte volume. When the frequency (or wave number, $\Delta\xi$) difference between the two lasers equals the $\Delta\nu_{Sk}$ of a Raman-active mode, there is an *increase* in the irradiance of the transmitted probe beam and a corresponding *decrease* of the irradiance of the transmitted pump beam. In one SRS system designed to sense D-glucose in biological samples, the $\Delta\nu$ between the pump and probe lasers is made to be the $\Delta\nu_{Sk}$ of a major glucose Raman peak, that is, $\xi = 518 \text{ cm}^{-1}$ (Tarr and Steffes 1993, 1998). (Recall that $\nu\lambda = c$, and $\xi \text{ cm}^{-1} = 3.333 \times 10^{-11} \nu$, ν in Hz.) The actual frequencies of the pump and probe lasers are not as important as the $\Delta\nu$ between them. The pump frequency should be in the NIR, if possible, to avoid exciting intrinsic fluorescence of the specimen. The advantage of the SRS technique is that no expensive monochromator is used. As shown in Figure 7.82, the probe laser's output is both wavelength tunable

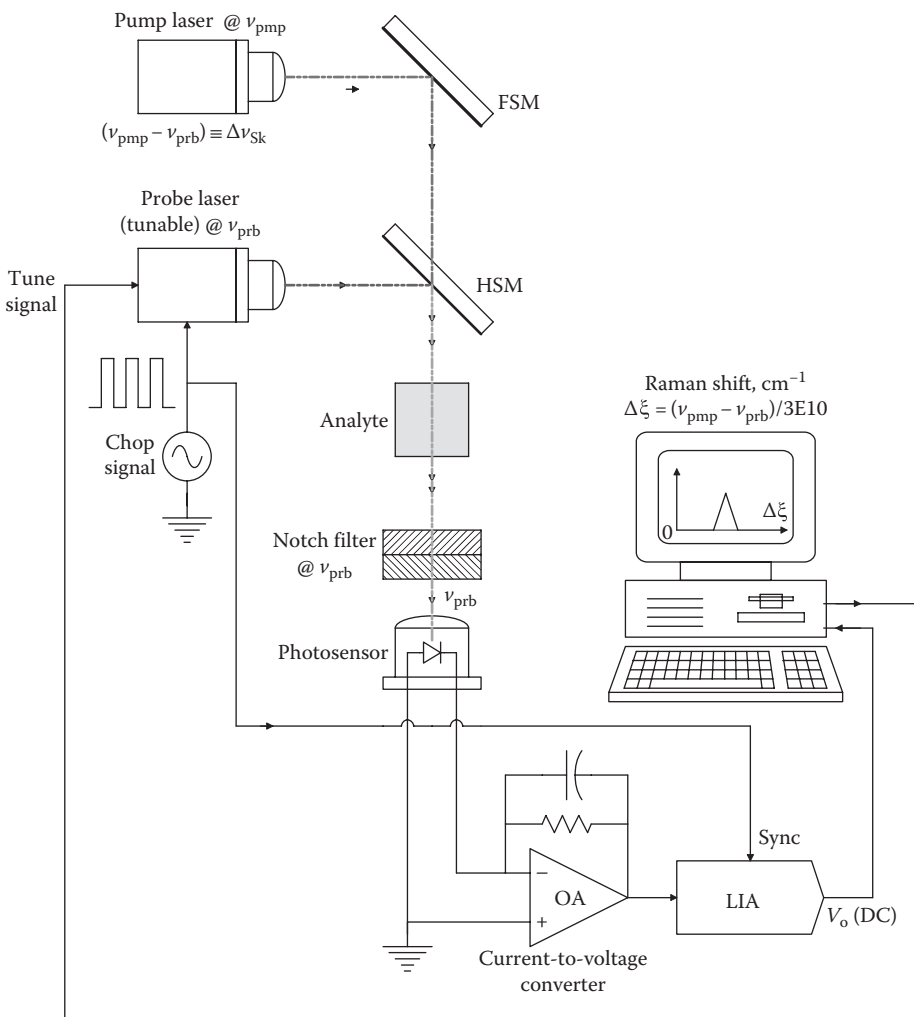


FIGURE 7.82
Diagram of an SRS system.

and chopped, either mechanically or electronically by switching the LAD current on and off. The beams from the probe and the pump lasers are combined, using either FO mixing or mirrors and an HSM, and then directed into the sample. Transmitted light from the pump laser is greatly attenuated by an optical notch filter, letting through the modulated probe laser beam that is directed to a suitable photosensor. The sensor's output contains a modulated signal component consisting of the probe beam increased as a function of the Raman–Stokes energy absorbed by the analyte, as well as modulated components from sample background fluorescence, plus Raman emissions from the water solvent and from glass (cuvette, lenses, and FOs). These latter components are artifacts with little $\Delta\nu$ dependence and are easily subtracted from the desired signal following synchronous (lock-in) demodulation of the sensor output. Note that in SRS, the *probe laser* beam's frequency is not shifted; rather, its amplitude-modulated intensity is changed.

Quoting Tarr and Steffes (1998):

The stimulated Raman effect requires no phase matching between the two laser beams, only that they be spatially overlapping in the Raman-active medium. When the frequency difference of the two lasers equals the vibrational frequency of the Raman-active mode [ν_{sk}], a nonlinear optical interaction occurs between the optical field and the medium. The strength of the nonlinear optical interaction depends upon the third-order nonlinear susceptibility, $\chi^{(3)}$, of the medium. The result of this interaction is a coherent increase in the [transmitted] irradiance of the Stoke's (probe) beam and a corresponding decrease in the irradiance of the [transmitted] pump beam.

Tarr and Steffes derived a theoretical model for the gain of the Stokes beam.

We have seen that good *in vitro* results have been obtained with RS and SRS in measuring certain analytes in water and blood plasma or serum (Northrop 2002). Mildly invasive medical procedures such as cystoscopy, bronchoscopy, and colonoscopy can provide a means to use Raman spectroscopy to examine lesions (candidate tumors) on the surface of the epithelium and mucosa in these locations.

For noninvasive, SRS diagnostic applications, the pump or excitation laser's energy must be introduced transdermally, and the very weak, backscattered, Raman-shifted light must be collected and analyzed dispersively. SRS could be used transdermally by shining the pump and probe lasers through a thin vascular tissue such as a finger web or an earlobe. A serious requirement in all forms of *in vivo* Raman analysis is to be able to do it at laser power levels that will not damage the biological tissues being irradiated (by heating) (Hanlon et al. 2000).

7.6.3 Nondispersive Spectrophotometry

Nondispersive spectrophotometry (NDS) is a chemical analytical method that avoids using an expensive monochromator and a PMT to quantify a *specific analyte*. The analyte can be liquid, solid, or gas. Instead of plotting $\%T(\lambda)$ or $A(\lambda)$ for a sample, and using the resulting peaks and valleys to quantify the analyte, an NDS instrument selects a narrow range of wavelength, $\Delta\lambda$, in which the analyte has a *unique* peak and valley in its $\%T(\lambda)$ curve. The $\Delta\lambda$ band is generated by passing broadband light through a BPF made from one or more *interference filters*. Note that there are several possible configurations for an NDS system (Northrop 2002); we will consider one of them here. Figure 7.83 illustrates a basic, OL, two-beam, NDS system that uses manual nulling. Light in the band, $\Delta\lambda$, is first chopped, then passed through a half-silvered, beam splitter mirror. The direct beam passes through the sample cuvette, thence to a photosensor, PS_1 . The reference beam passes through a

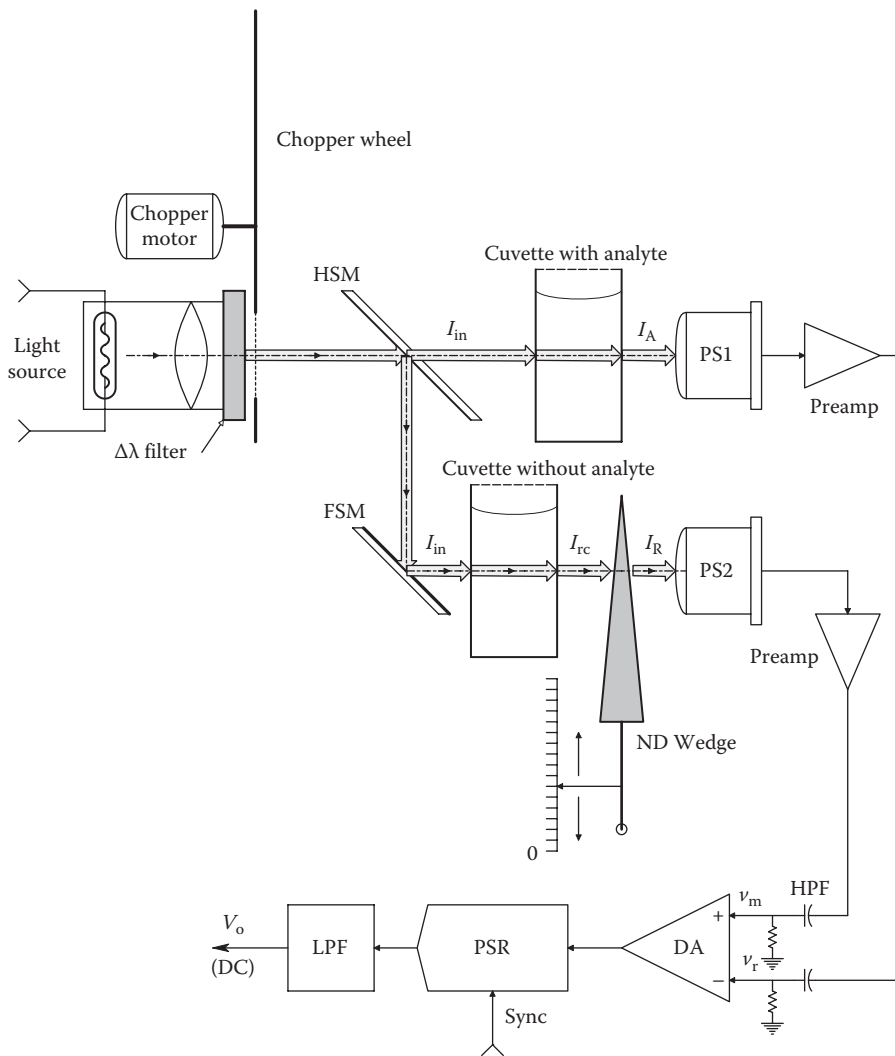


FIGURE 7.83

Diagram of a manually nulled, two-beam, nondispersive spectroscopy system. The $\Delta\lambda$ filter is a band-pass interference filter that provides photon energy in a $\lambda_1 - \lambda_2$ band.

blank cuvette filled with solvent, thence through an adjustable, calibrated, neutral density wedge (NDW), thence to photosensor PS_2 . The outputs of the photosensors are high-pass filtered to remove DC, then amplified by a difference amplifier (DA). The DA output is a square wave at chopper frequency whose amplitude and phase (0° or 180°) are determined by the relative intensities at PS_1 and PS_2 . In system calibration, with no analyte present in the sample cuvette, the NDW is manually set to zero (no extra attenuation in the reference beam). The DC output voltage, V_o , is nulled by adjusting the gain of either preamp so $v_{m0} = v_{r0}$. With the analyte present, more light in the passband, $\Delta\lambda$, in the upper beam is absorbed and the total intensity at PS_1 , I_A , is reduced. Now, the NDW must be advanced to renull V_o . The neutral density (ND) required to renull the system can be shown to be proportional to the analyte concentration.

A further refinement of the NDS system is to make it self-nulling (see Figure 7.84). In this design, the DC error voltage, V_o , is integrated to make a type 1 control system with zero SS error, and the integrator output, V'_o , is conditioned to drive a linear positioning system that moves the NDW. Now, it can be shown that the difference in the voltage V'_o with no analyte and V'_o with analyte is proportional to the (linear) change in ND, hence in the

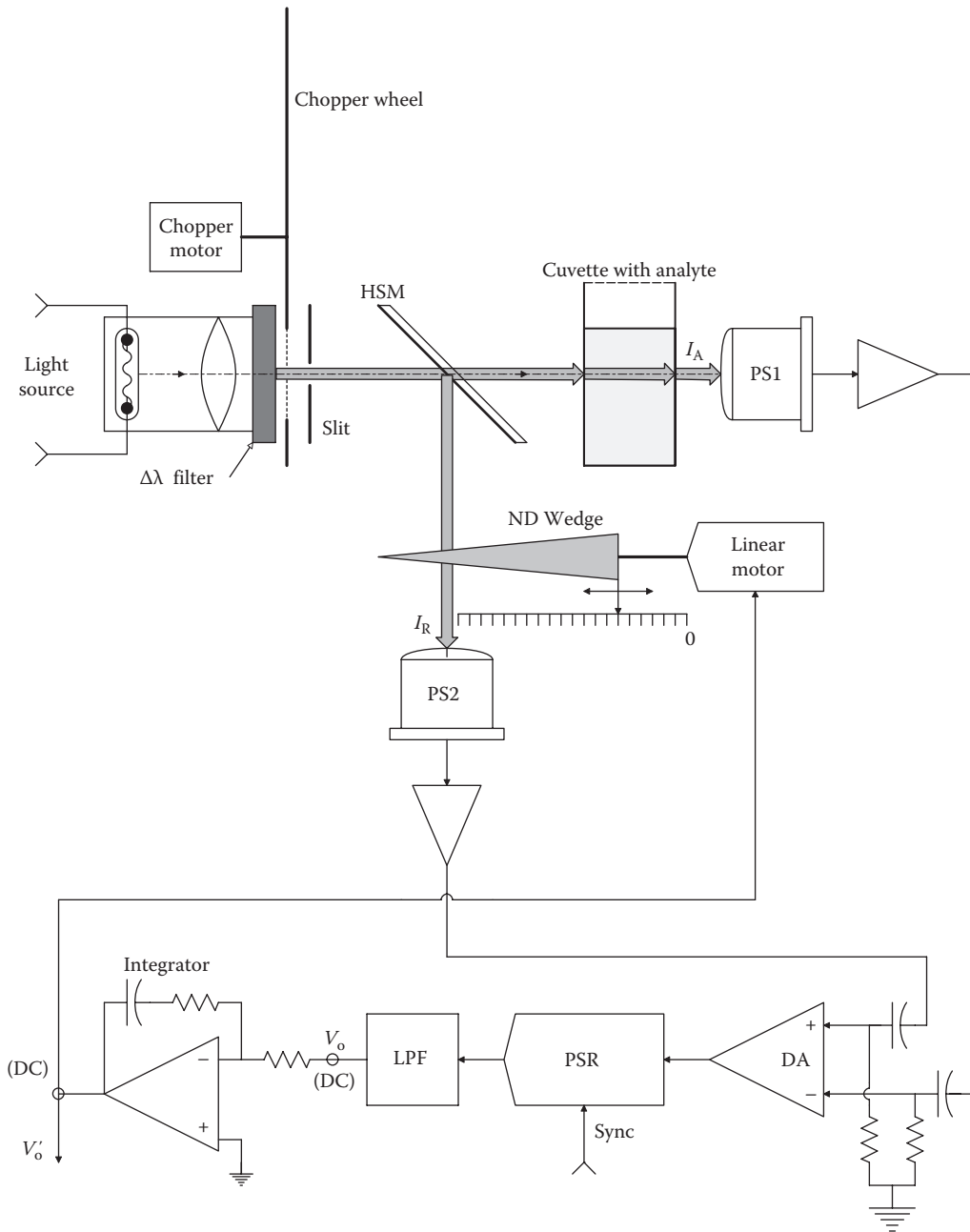


FIGURE 7.84
Servonulled NDS system.

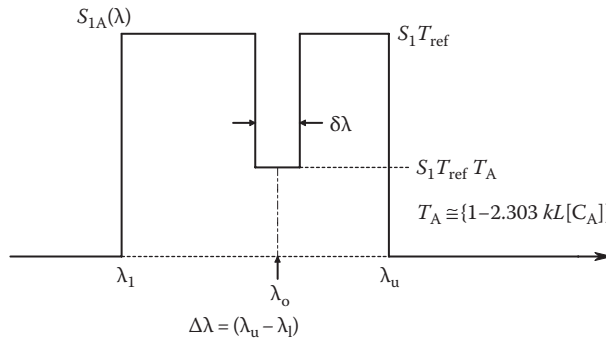


FIGURE 7.85

Idealized light intensity spectrum, $I_A(\lambda)$, from Figure 7.84. The notch represents photon energy absorbed by the analyte.

analyte concentration. An idealized transmittance spectrum is shown in Figure 7.85. Note the absorbance notch caused by the analyte.

We now describe mathematically how an OL NDS system works (see Figure 7.83). Assume that the output of the optical BPF is an ideal rectangular spectrum of width, $\Delta\lambda$. After being split into two beams by the HSM, both source spectrums have width $\Delta\lambda$ and equal magnitudes, $S_1(\lambda)W/nm$. The reflected reference spectrum is attenuated by the reference cuvette's reflections and the solvent's absorption. The reference cuvette has a transmission $T_{ref} = (I_{rc}/I_{in})$. $I_{in} = (S_1 \Delta\lambda) W$. I_{rc} next passes through the NDW where it is further attenuated by its transmittance, T_{ND} . In general, $T_{ND} \equiv (I_{out}/I_{in}) = (I_R/I_{rc}) = 10^{-\beta x}$, where x is the displacement of the wedge from 0 on the linear x scale, and β is the NDW's attenuation constant. (Note that $ND \equiv \beta x$; also that an ND of 1 means the input light is attenuated by a factor of 0.1, or $I_{out} = I_{in}/10$.) Thus, the intensity at the reference photo-sensor is

$$I_R = (S_1 \Delta\lambda) T_{ref} T_{ND} W. \tag{7.236}$$

The spectrum that passes through the HSM also passes through the walls of the sample cuvette twice (in and out) and also through the test solution with analyte. Thus, the measurement spectrum is also attenuated by a factor $T_{ref} < 1$. We assume that in a narrow band, $\delta\lambda$, around λ_o , the analyte has a strong absorption. Hence, the spectrum exiting the cuvette, $S_{1A}(\lambda)$, has a reduced total intensity as shown in Figure 7.83. Using Beer's law, the intensity of $S_{1A}(\lambda)$ is simply

$$I_A = (S_1 \Delta\lambda) T_{ref} - \delta\lambda S_1 T_{ref} (1 - T_A). \tag{7.237}$$

The transmittance minimum of the analyte, T_A , is given by Beer's law:

$$T_A = 10^{-kL[A]} \cong 1 - 2.303kL[A]. \tag{7.238}$$

Here, we have assumed that $kL[A] \ll 1$. $[A]$ is the concentration of the analyte, and the value of k is both wavelength and concentration dependent. Thus, after some algebra, we can write the intensity of the light exiting the cuvette as

$$I_A = (S_1 \Delta \lambda) T_{\text{ref}} \left\{ 1 - 2.303 \left(\frac{\delta \lambda}{\Delta \lambda} \right) kL[A] \right\} W. \quad (7.239)$$

To null the system, the wedge is advanced until, again, $I_R = I_A$ and $V_o = 0$. Thus, at null we have

$$(S_1 \Delta \lambda) T_{\text{ref}} T_{\text{ND}} = (S_1 \Delta \lambda) T_{\text{ref}} \left\{ 1 - 2.303 \left(\frac{\delta \lambda}{\Delta \lambda} \right) kL[A] \right\}. \quad (7.240)$$

So

$$T_{\text{ND}} = \left\{ 1 - 2.303 \left(\frac{\delta \lambda}{\Delta \lambda} \right) kL[A] \right\}. \quad (7.241)$$

Now, we note that the analyte concentration is small, so we can use $T_{\text{ND}} = 10^{-\beta x} \cong 1 - 2.303 \beta x$. Thus,

$$\begin{aligned} (1 - 2.3 \beta x) &= \left\{ 1 - 2.303 \left(\frac{\delta \lambda}{\Delta \lambda} \right) kL[A] \right\}. \\ \downarrow & \\ x &\cong \left(\frac{\delta \lambda}{\Delta \lambda} \right) \frac{j kL[A]}{\beta} \end{aligned} \quad (7.242)$$

Equation 7.259 tells us that the displacement of the NDW by x to renull the NDS when an analyte is present is simply proportional to the analyte's concentration, $[A]$, assuming Beer's law holds and the NDW's ND is given by $10^{-\beta x}$, where $\beta x \ll 1$. (The 2.3 factor comes from approximating $10^{-\beta x}$ by the use of $e^{-\epsilon} \cong 1 - \epsilon$. Note that $\log_{10}(a) = \ln(a)/2.303$.)

To demonstrate the effectiveness of the NDS approach in quantifying an analyte, Fellows (1997) designed, built, and tested an NDS to measure the % oxyhemoglobin in vitro. Percutaneous measurement of the % O₂ saturation of RBC hemoglobin (Hb) is an established NI measurement technique. The *pulse oximeter* (described in detail in Northrop [2002]) makes use of the differential absorption of red and near infrared (NIR) light by Hb vs. HbO. Pulse oximeters use a red and an NIR light-emitting diode (LED) (not LADs) as a 2λ source to make the measurement. Fellows' NDS system was developed as a proof-of-concept design, rather than a competing instrument for the very simple pulse oximeter. Fellows' system architecture closely followed the design shown in Figure 7.84, except instead of a linear, NDW, she used a rotating, fine-mesh screen as a variable, ND attenuator. Unlike the linear wedge, the screen's transmittance is a function of its angle with respect to the beam being attenuated and can be shown to be given by

$$T_{\text{sc}}(\theta) = \left(\frac{1-2r}{d} \right) \left[1 - \left(\frac{2r}{d} \right) \sec(\theta) \right]. \quad (7.243)$$

Each wire in the screen has radius, r , and d is the center-to-center spacing of the wires; a square mesh is assumed. The ND screen was rotated by a servogalvanometer whose angle, θ , was proportional to its DC input voltage.

Fellows used a filter that passed light in a 650–750 nm band. One-centimeter cuvettes were used for the test solutions, the standard and the compensation. Sigma® freeze-dried human hemoglobin was made up at a concentration of 12 g/L in water buffered to pH 7.4. To make either HbO or Hb, either O₂ gas or CO₂ gas was bubbled through the sample and reference cuvettes, respectively.

Fellows' servonulling NDS system was able to measure % HbO, but was difficult to calibrate because of the very nonlinear ND relation. A linear wedge or ND disk would have simplified the operation of her system.

A logical application of the NDS system would be to try to noninvasively and percutaneously measure (tissue) blood glucose in the FIR range. Glucose has three absorption peaks (%*T* nulls) in the FIR: one at 10.97 μm, one at 11.98 μm, and one at 12.95 μm. If the LIR BPF passed from 10.5 to 11.5 μm, the peak at 10.97 μm could be used. Figure 7.86 illustrates a prototype, FIR, NDS system proposed by the author that might use a finger web or earlobe (always with constant *L*).

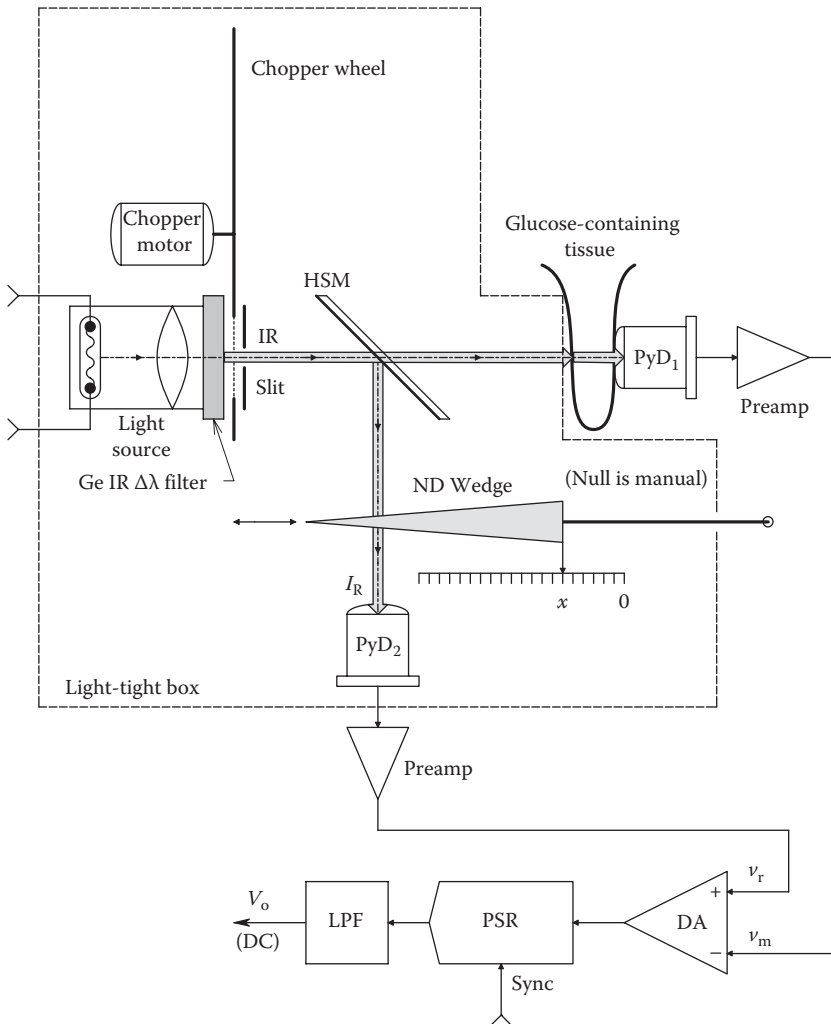


FIGURE 7.86 Prototype NDS system that might be able to measure blood glucose by IR absorption. An earlobe or finger web might be used.

The FIR beam is defined by an IR BPF and chopped, and the null is detected using a PSR. A light-tight box excludes stray IR from the pyroelectric photosensors (e.g., PVDF). FIR-compatible HSM and NDW optical elements must be used. Such a system would be easy to build, but its accurate use would depend on individual calibration with a standard blood glucose test and the absence of other tissue and blood substances that might overwhelm the glucose absorption peaks.

7.6.3.1 Use of Nondispersive Spectrophotometry to Quantify CO_2 Gas in Air

The inspection of an IR absorption spectrum of pure CO_2 gas shows it has three major IR absorptivity peaks: two large, sharp peaks occur at $4.2 \mu\text{m}$ and at ca. $2.6 \mu\text{m}$ wavelength and a smaller, broad peak is at ca. $2 \mu\text{m}$. The absorptivity peak at $2.6 \mu\text{m}$ is coincident with a large, broad, absorptivity peak of water vapor. However, water vapor does not interfere significantly with the $4.2 \mu\text{m}$ CO_2 peak. Thus, an NDS sensor system using IR radiation optically band-pass filtered between 3.5 and $4.4 \mu\text{m}$ can potentially quantify CO_2 concentration in dried air.

The Vaisala CARBOCAP[®] NDS IR sensor is specifically designed to measure $[\text{CO}_2]$ in air and flue gasses. It consists of a chamber that has an IR source, a mirror to effectively increase IR beam length, a tunable FPI filter, and an IR detector. The FPI filter is electrically tuned to make $[\text{CO}_2]$ and reference measurements. Note that atmospheric CO_2 measurements are important in mushroom farming, greenhouse agriculture, and fermentation applications, as well as human environmental health issues.

Vaisala makes the GM70 handheld CO_2 meter for spot checking applications; it uses their patented CARBOCAP[®] sensor and has a number of ranges spanning 0%–20% (GMP221 probe) and 0–10,000 ppm (GMP222 probe). Accuracy for the GMP221 and GMP222 probes is $\pm(1.5\%$ of range $+2\%$ of reading).

7.6.3.2 Nondispersive Photoacoustic Spectrophotometry

A photoacoustic NDS spectrometer is shown schematically in Figure 7.87. Such instruments have been used to quantify gas-phase concentrations of CO , CO_2 , N_2O , halothane, and other gasses in air (Webster 1992).

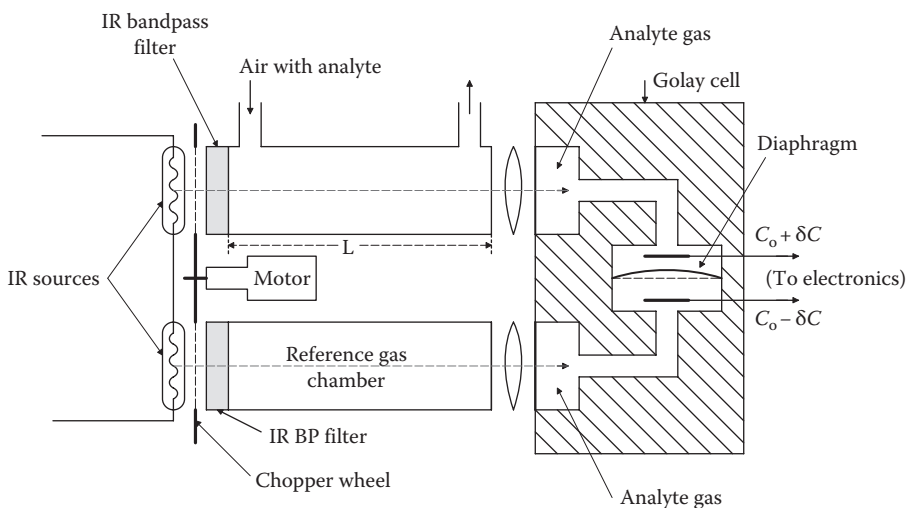


FIGURE 7.87

Cross-sectional diagram of a nondispersive, photoacoustic spectrometer designed to quantify analyte gasses.

To quantify $[\text{CO}_2]$ in an air sample using NDPAS, broadband IR is band-pass filtered around the CO_2 absorption peak at $4.2 \mu\text{m}$. The reference cell contains air with no CO_2 , so the photon energy transmitted through it is larger than from the sample with CO_2 because the CO_2 absorbs heavily in the $4.2 \mu\text{m}$ band. Both exiting beams are focused through IR-pass windows into the Golay (aka Luft) cell (Moseley et al. 1991, Webster 1992). The pure CO_2 in the Golay cell efficiently absorbs the pulsed (due to the chopper) photon energy from the sample and reference cells and undergoes a transient thermal expansion. The CO_2 in the bottom detector chamber in general absorbs more thermal energy than that in the upper chamber; thus, the transient acoustic pressure wave produced by this heating pushes the diaphragm upward, causing a momentary increase in the capacitance between the grounded diaphragm and the upper electrode and simultaneously a decrease in capacitance between the diaphragm and the lower electrode. These transient changes in capacitance can be detected electronically with an AC De Sauty bridge (cf. Figure 5.7) and associated circuitry. Note that the optical chopping is necessary to produce detectable, periodic, photoacoustic transients.

7.6.4 Fourier Transform IR Spectrophotometry

A second very important category of nondispersive spectrophotometer is the *FTIR system*, shown schematically in Figure 7.88. Some advantages of FTIR spectroscopy include (1) better sensitivity and brightness (more light incident on the sample than dispersive IR spectrometers), (2) high wave number accuracy (to 0.01 cm^{-1}), (3) takes less time to make a spectrogram, and (4) relatively insensitive to stray light. A disadvantage of FTIR spectroscopy is the same as dispersive IR spectroscopy: both H_2O and CO_2 absorb strongly at certain wave numbers, obscuring analyte peaks.

As shown in Figure 7.88, a blackbody IR source such as a Globar rod is used to generate a continuous, broadband, IR spectrum given by Planck's radiation law (cf. Figure 7.89). Collimated radiation from this source is the input to a time-modulated,

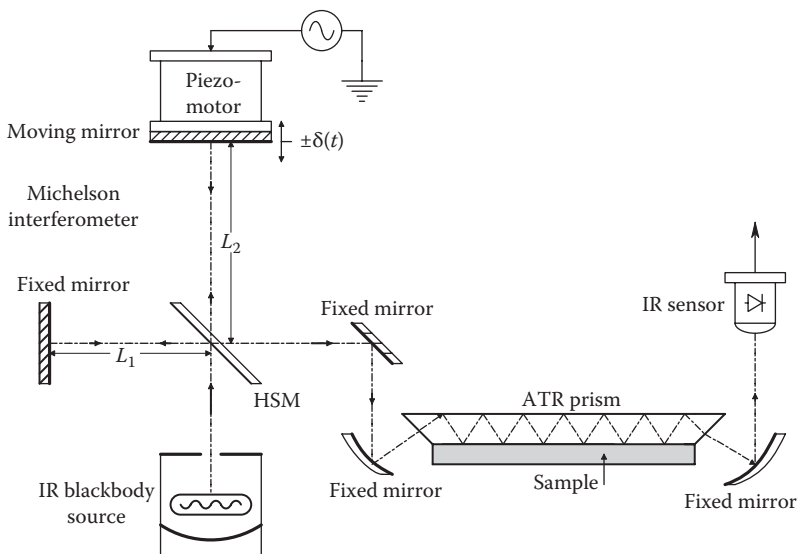


FIGURE 7.88
Diagram of an FTIR spectrometer.

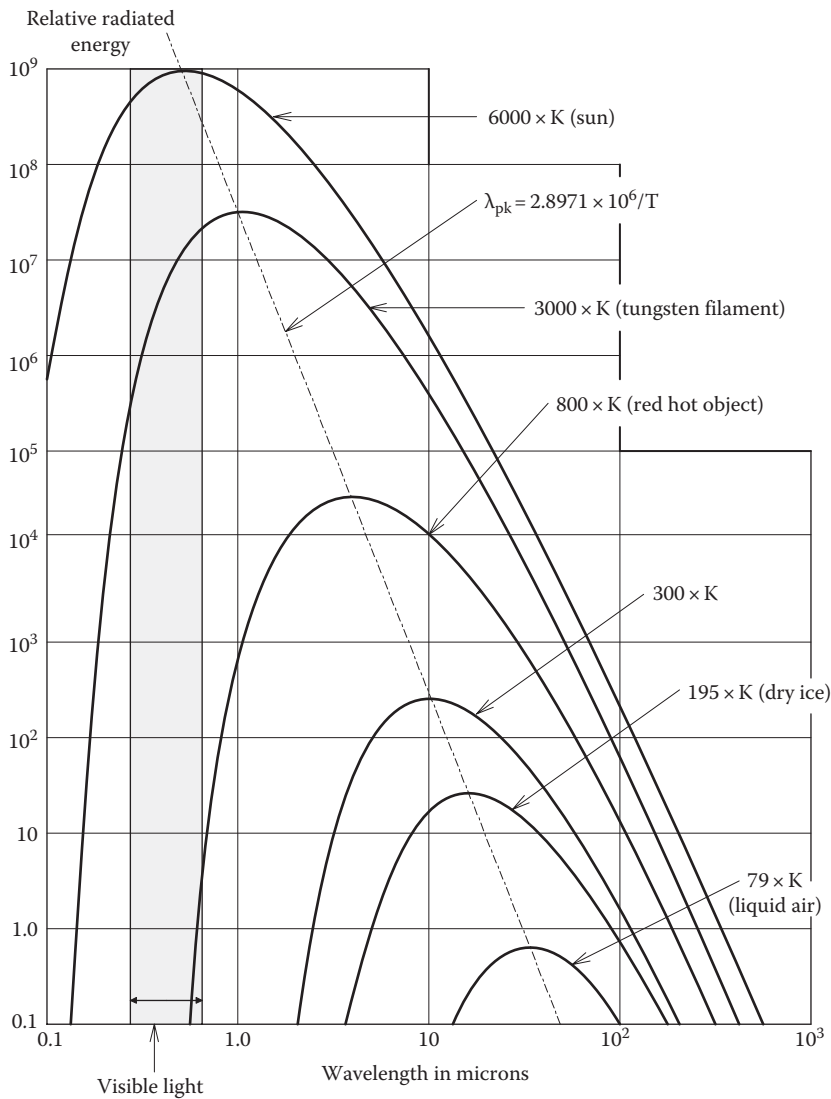


FIGURE 7.89 Ideal blackbody radiation spectra. Note how the peaks shift to longer wavelengths at lower object temperatures.

Michelson interferometer. The interferometer’s mirror is periodically displaced by an amount, $\delta(t)$, around a center position, L_2 . To understand what happens at the interferometer’s output, we need to first consider what happens to a *monochromatic* input ray with wavelength λ_1 . The distances L_1 and L_2 are chosen such that there is maximum *constructive interference* at the interferometer output. Thus, neglecting interfacial losses, the output intensity, I_o , equals the input intensity, I_{in} . Now, let the mirror move $\delta = \lambda_1/4$ away from the beam splitter. The light must now travel a total distance of $\lambda_1/2$ μm more to return to the beam splitter where there is now *destructive interference*, and the output intensity has a null. This extra length the light must travel is called the *retardation*

distance, ρ . In general, $\rho = 2\delta$. We can write an empirical expression for the output intensity as a function of the wave retardation distance, ρ , around L_2 :

$$I_o(\rho, \lambda_1) = I_{in} \left(\frac{1}{2} \right) \left[1 + \cos \left(\frac{2\pi\rho}{\lambda_1} \right) \right] W. \tag{7.244}$$

In terms of wave number, $\xi_1 \equiv \lambda_1^{-1} \text{ cm}^{-1}$, I_o can also be written as

$$I_o(\rho, \xi_1) = I_{in} \left(\frac{1}{2} \right) [1 + \cos(2\pi\xi_1\rho)] W. \tag{7.245}$$

Note that $I_o(\rho, \xi_1)$ is an even function in $\rho\xi_1$. ρ must be in cm in Equation 7.262. When two other beams having wave numbers ξ_2 and ξ_3 are added to the beam of wave number ξ_1 , then the output of the time-modulated Michelson interferometer can be written, using superposition, as

$$I_o(\rho) = \sum_{k=1}^3 I_{ink} \left(\frac{1}{2} \right) [1 + \cos(2\pi\xi_k\rho)] W, \text{ at a given retardation, } \rho \text{ cm.} \tag{7.246}$$

When this even interferogram function is plotted vs. ρ , we see a strong peak at the origin surrounded by lower ripple peaks, as shown in Figure 7.90. In the more general case, when the source intensity input to the interferometer is described by a continuous density distribution in terms of wave number, $\xi \text{ cm}^{-1}$, the interferogram intensity is given by the inverse, cosine FT:

$$I_o(\rho) = \int_{-\infty}^{\infty} W_B(\xi) \left(\frac{1}{2} \right) [1 + \cos(2\pi\xi\rho)] d\xi = I_{W_0} + \int_{-\infty}^{\infty} W_B(\xi) \left(\frac{1}{2} \right) \cos(2\pi\xi\rho) d\xi \tag{7.247}$$

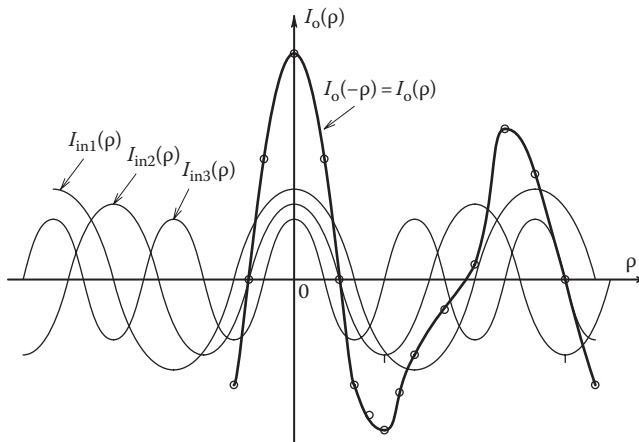


FIGURE 7.90 Superposition of the output intensity of the interferometer as a function of wave number and retardation distance, ρ , which is time modulated. Note that with just three wave numbers, a peak in the interferogram grows around $\rho = 0$.

where $W_B(\xi)$ is assumed to be even in ξ , so $W_B(\xi) = W_B(-\xi)$. Thus,

$$I_{W_0} = \int_{-\infty}^{\infty} W_B(\xi) d\xi. \quad (7.248)$$

The cosine integral can be considered to be the real, *inverse* FT of $W_B(\xi)$. That is,

$$I_o(\rho) = \left(\frac{1}{2}\right) \mathbf{F}^{-1} \{W_B(\xi)\}. \quad (7.249)$$

Note that by the Euler relation, $e^{-j\omega t} \equiv \cos(\omega t) - j \sin(\omega t)$, and one definition of the continuous Fourier transform (CFT) is

$$\mathbf{F}(\omega) = R(\omega) + jX(\omega) \equiv \int_{-\infty}^{\infty} f(t) [\cos(\omega t) - j \sin(\omega t)] dt. \quad (7.250)$$

When $f(t)$ is real and even, then it is well known that

$$F(\omega) \equiv \int_{-\infty}^{\infty} f(t) \cos(\omega t) dt + 0 \quad (7.251)$$

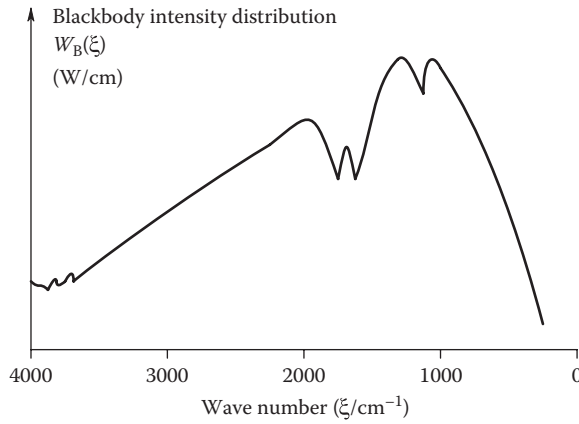
and the continuous, inverse, FT is

$$f(t) \equiv (2\pi)^{-1} \int_{-\infty}^{\infty} F(\omega) \cos(\omega t) d\omega \quad (7.252)$$

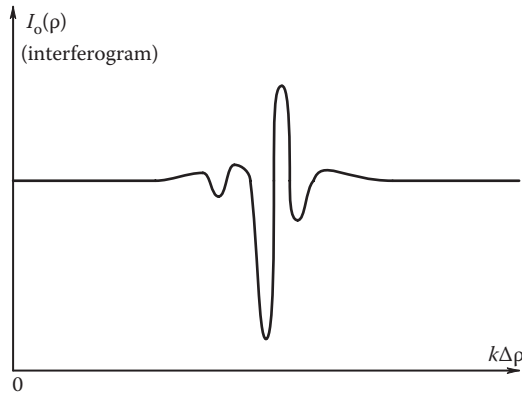
where ω is in r/s and $X(\omega) = 0$. Note that we have defined the blackbody radiation density, $W_B(\xi)$ W/cm⁻¹, as an even function in wave number, ξ cm⁻¹. Let ξ be analogous to lightwave frequency in Hz and the retardation distance, ρ cm, be analogous to time in Equation 7.268; thus, $F(\omega/2\pi) \rightarrow W_B(\xi)$, and we see that the modulation of the source light's spectral distribution, $W_B(\xi)$, by the Michelson interferometer generates an *even, inverse Fourier (intensity) function* of $W_B(\xi)$, $I_o(\rho)$, at its output. Sic,

$$I_o(\rho) = \int_0^{\infty} W_B(\xi) \cos(2\pi\xi\rho) d\xi. \quad (7.253)$$

Now, if we sample the IFT, $I_o(\rho)$, at intervals $\Delta\rho$, and take its DFT, we recover an estimate of the input blackbody intensity distribution, $W_B(\xi)$, as shown in Figure 7.91. This $W_B(\xi)$ estimate is stored in the FTIR system's computer memory. Figure 7.92 illustrates a representative interferogram as a function of the retardation, ρ .

**FIGURE 7.91**

Plot of the blackbody source's intensity spectrum vs. wave number, as determined by taking the FT of $I_o(\rho)$ given by Equation 7.269.

**FIGURE 7.92**

Typical, continuous interferogram output of the photosensor as a function of $k\Delta\rho$.

The next step in the operation of an FTIR spectrometer is to pass the interferometer-modulated IR source radiation through the sample where selective absorption (attenuation) in certain wave number bands occurs. Again, as in Equation 7.262, let us look at one particular wave number, ξ_1 . The intensity of the light emerging from the sample, $I_e(\rho, \xi_1)$, is given by

$$I_e(\rho, \xi_1) = I_{in}(\xi_1)T_s(\xi_1)\left(\frac{1}{2}\right)[1 + \cos(2\pi\xi_1\rho)] \quad (7.254)$$

where

$T_s(\xi_1)$ is the wave number-dependent ($0 \leftrightarrow 1$) transmittance of the analyte and sample
 $I_{in}(\xi_1)$ is the input intensity at $\xi = \xi_1$

If the input is replaced by the BB distribution, $W_B(\xi)$, we can write

$$I_e(\rho) = \left(\frac{1}{2}\right) \int_{-\infty}^{\infty} W_B(\xi) T_s(\xi) [1 + \cos(2\pi\xi\rho)] d\xi = I_{\text{ewo}} + \left(\frac{1}{2}\right) \int_{-\infty}^{\infty} W_B(\xi) T_s(\xi) \cos(2\pi\xi\rho) d\xi. \quad (7.255)$$

Clearly, the DC term is

$$I_{\text{ewo}} = \int_0^{\infty} W_B(\xi) T_s(\xi) d\xi, \quad (7.256)$$

and the integral is the *inverse* FT of the frequency domain product, $\{W_B(\xi) T_s(\xi)\}$. When the output interferogram, $I_e(\rho)$, is DFTd at intervals, $\Delta\rho$, we obtain an estimate of the output spectrum:

$$S_{\text{Sa}}(\xi) = \left(\frac{1}{2}\right) W_B(\xi) T_s(\xi). \quad (7.257)$$

Since we already know $W_B(\xi)$, the desired sample transmittance is easily found by the relation

$$T_s(\xi) = \frac{2S_{\text{Sa}}(\xi)}{W_B(\xi)}. \quad (7.258)$$

It goes without saying that a high-speed computer capable of calculating DFTs and generating plots of $T_s(\xi)$ and absorbance $A(\xi)$ is a necessary component of an FTIR spectrometer.

Discussion: We have seen that all spectrometers pass EMR through a cuvette holding the analyte in solution or use an ATR plate to interface light with the analyte. In some designs, the incident EMR is broadband (*white*), and the emergent radiation from the sample is passed through a monochromator to examine the transmittance or absorbance at particular wavelengths. In other designs, monochromatic light is first passed through the sample. Spectrometers operating in the FIR require the use of special, IR-pass materials for refraction and cuvettes, including ZnSe (0.6–17 μm), ZnS (0.4–12 μm), Ge (2–14 μm), Si (1.2–8 μm), KRS-5 (0.6–50 μm), As_2S_3 (0.8–10 μm), and certain plastics. In nearly all cases, the light beam is chopped in order that phase-sensitive demodulation can be used (equivalent to an LIA) to improve sensitivity and reject noise and stray (DC) light.

Nondispersive spectrometers use narrowband light, two or more LEDs, or LADs to avoid the need for an expensive monochromator. FTIR spectrometers also fall into the class of nondispersive instruments, as do stimulated Raman spectrometers. In the FTIR spectrometer, the modulation is provided by the Michelson interferometer.

All sorts of medically important molecules from body fluids can be quantified by IR spectrophotometry. These include but are not limited to cholesterol, steroid hormones, thyroxine, theophylline, opiates, opioids, and tranquilizers. Knowledge of a drug concentration in blood or urine can be used clinically to adjust the dosage and also to detect drug abuse. Raman spectroscopy is showing promise in the noninvasive diagnosis of certain superficial tumors including skin melanomas, lung cancer, and prostate and bladder cancer in any place an endoscope can reach (Northrop 2002).

7.6.5 Single- and Multiple-Bubble Sonoluminescence

7.6.5.1 Introduction

In this section, we will examine the phenomena of SBSL and MBSL caused by cavitation in liquid due to an intense ultrasound field and consider their potential as chemical analytical tools with possible application to the detection of airborne molecules of highly explosive substances such as C4, Semtex, PETN, and dynamite vapors.

7.6.5.2 Single-Bubble Sonoluminescence

Figure 7.93 illustrates the schematic of a basic, easy-to-construct, experimental, SBSL apparatus (Hiller and Barber 1995). A 100 mL, round, Florence flask was used as a reaction vessel. Three piezoceramic disk transducers were used. Electrical connections to the disk transducers were made using silver epoxy (soldering the leads to the thin silver films on both sides of the transducers can destroy [vaporize] the films if not done carefully and is not recommended).

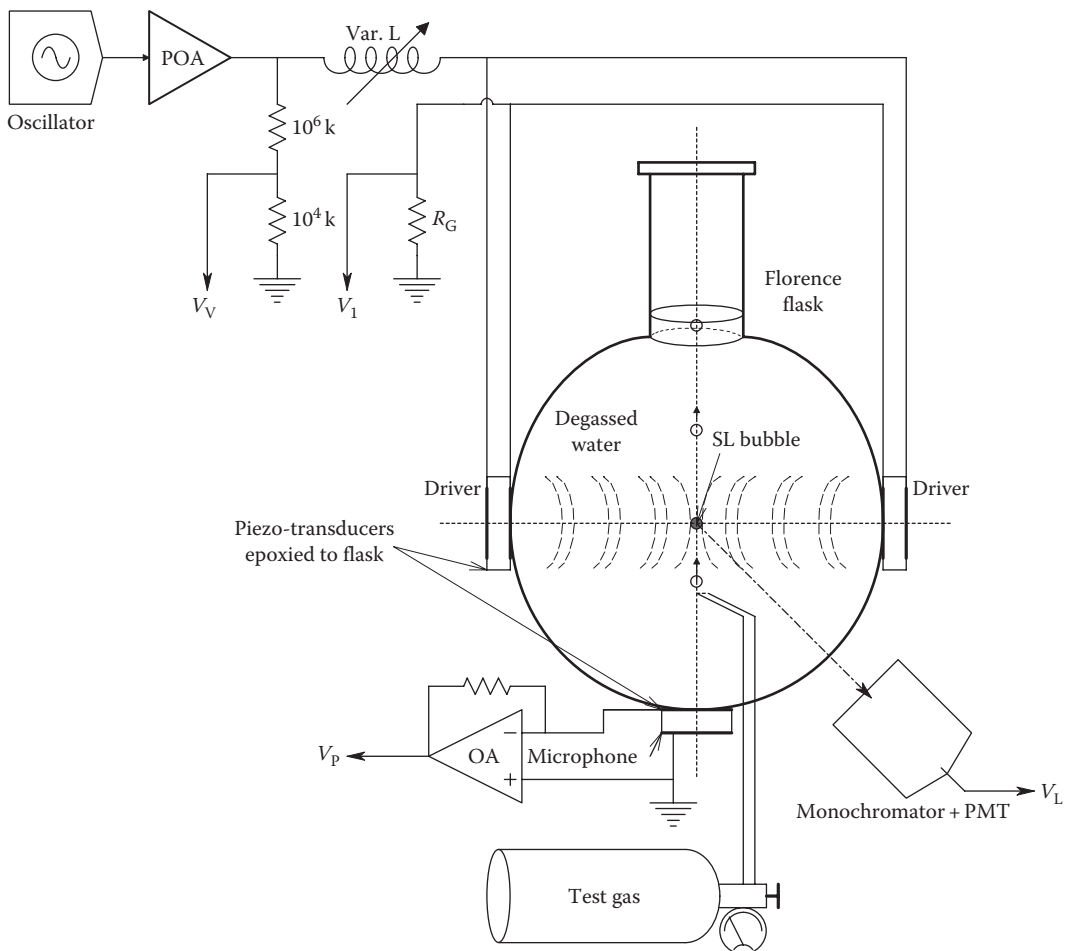


FIGURE 7.93
SBSL system of Hiller and Barber (1995).

The transducers were epoxied to the flask in the positions shown (the flask was suspended from its neck by a clamp, not shown). The two larger transducers (ca. 2 cm diam.) were the ultrasound drivers; the bottom, smaller (ca. 5 mm diam.) transducer was used as a microphone to sense sound in the flask and shock waves from the periodic, rapid collapse of the bubble in the sound field. The microphone signal was conditioned by a simple op amp circuit, and its output, V_p , was sent to an oscilloscope.

A sine wave generator was connected to a power amplifier (PA), whose $32\ \Omega$ output is connected to a variable inductor having an appropriate current rating. The inductor was connected to the two driver transducers in parallel. Below the transducers' *self-resonance* frequency (e.g., 1 MHz), they appeared electrically like leaky capacitors because of acoustic loading by the flask. Their combined capacitance is on the order of 1 nF. It was found experimentally that the *acoustic resonance* frequency, f_A , for the water-filled flask is ca. 20–30 kHz. At f_A , the maximum acoustic energy is coupled to the flask. An easy way to observe f_A is to monitor V_p as the frequency of the function generator is varied *with the inductor shorted out*. At f_A , V_p is maximum. To produce maximum driving power to the transducers, they must be *electrically resonated* by adjusting the variable inductance. L was varied at constant f_A until zero phase shift is seen between V_v and V_i . R_G was made $1\ \Omega$, so V_i equals the current into the transducers. Note that at electrical resonance, the PA sees a real impedance of $(R_G + R_X)$, where R_X is the transducers' equivalent acoustic load into the flask. Small test gas bubbles were introduced into the flask by a thin glass tube connected to a gas supply.

SBSL is an interesting physical phenomenon that was first described in the literature about 1990 (Crum 1994, Lohse 2002). If a single small bubble of gas in water having a resting diameter of ca. $10\ \mu\text{m}$ is trapped in an ultrasound field of appropriate intensity, the bubble undergoes expansion and compression at the ultrasound frequency. When the bubble is compressed by the sinusoidal sound pressure waves, at a critical pressure, its volume abruptly collapses at a speed of about Mach 4 to about $1\ \mu\text{m}$ diameter. Its maximum radius at minimum sound pressure is ca. $100\ \mu\text{m}$. Thus, the volume compression ratio is an enormous 10^6 (Moss et al. 1997a,b). This abrupt compression causes quasiadiabatic heating of the gas volume, causing the internal temperature to rise to an estimated peak of 10^{40}C – 10^{60}C (Crum and Matula 1997). A plasma is created inside the collapsed bubble that emits a pulse of photons ca. 200 ps in duration (Lohse 2002). (See Figure 7.94.) This flash is probably due to trace argon gas in the bubble. Even though the individual flashes are extremely short, the fact that they occur at the ultrasound frequency, f_A (generally from 20 to 30 kHz), causes enough average light flux so that the sonoluminescence is visible to dark-adapted human eyes as a blue point of light.

At the very high temperature and pressure generated when the bubble collapses, a photon-emitting plasma forms. The primary photon emission mechanism appears to be electron–ion or electron–atom bremsstrahlung (Suslick et al. 1999). At the very high temperature and pressure following collapse, various chemical reactions between the nitrogen, oxygen, and other trace molecules in the bubble are postulated to occur. These reaction products are gradually absorbed (rectified) by the water, leaving a bubble containing a high concentration of chemically inert argon from the air. The argon gas sonoluminesces. Hiller et al. (1994) showed that SBSL does not occur with pure N_2 , O_2 , and $\text{N}_2 + \text{O}_2$; Ar (or some other noble gas) is required. On the other hand, dissolved gasses in the water can enter the bubble during its expansion cycle in the aqueous sound wave field.

The emission spectrum of SBSL is largely blackbody, theoretically extending into the far UV because of the high peak plasma temperature. Unfortunately, water and the outer plasma envelope itself absorb the shorter wavelength UV emissions. There are no discernable emission lines from SBSL in water (Suslick et al. 1999, Didenko et al. 2000).

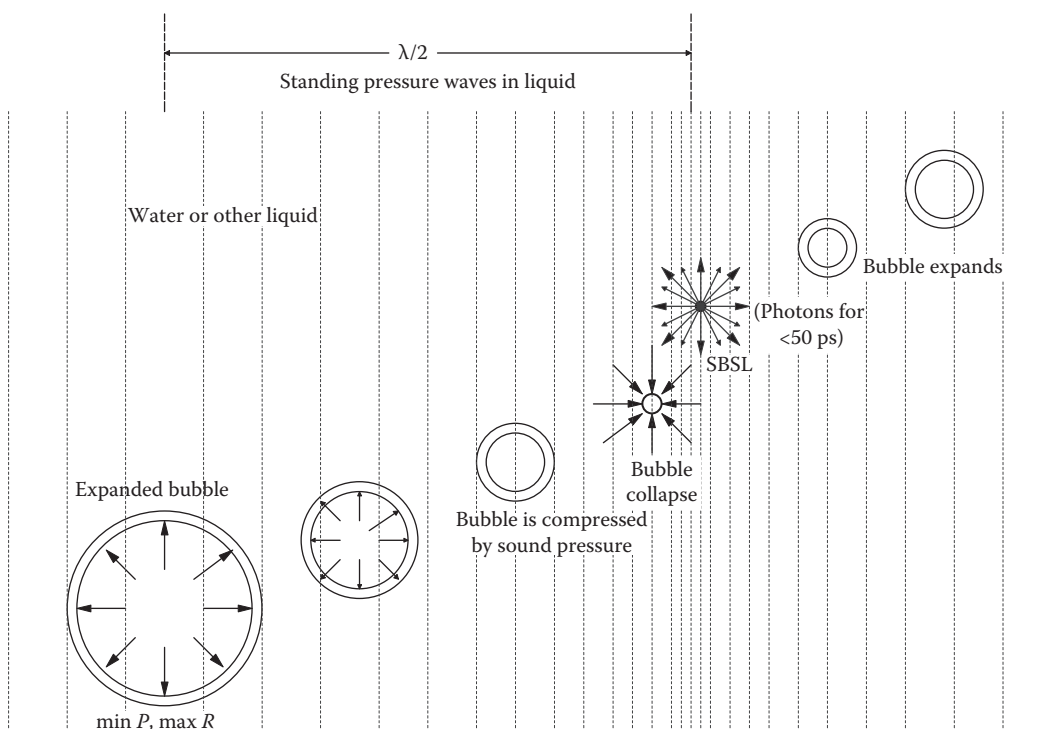


FIGURE 7.94

Steps illustrating the collapse, sonoluminescence, and reexpansion of a gas bubble in an ultrasound field in water.

Didenko et al. speculated that the lineless SBSL emission in water is (1) due to the summation of broadened emissions from multiple molecular and atomic sources and (2) bremsstrahlung emission from partial ionization of the heated gas within the bubble. Didenko et al. studied the SBSL emission spectra from a variety of polar aprotic liquids (cf. Glossary), including *formamide*, *N-methylformamide*, *N,N-dimethylformamide*, *1,2-diaminoethane*, *dimethylsulfoxide*, and *adiponitrile*. They subdivided their results into emissions from bubbles held stationary in the sound field and bubbles that move around the reaction flask's center and emit photons. The spectra from the moving bubbles in *methylformamide* and *adiponitrile* showed a broad emission line at ca. 380 nm that was from $-\text{CN}$ dissociated from the liquid molecules. When DMSO was used, there obviously was no $-\text{CN}$ line to dissociate (no N in the molecule), but there also were no band emission lines from S_2 , $-\text{SO}$, or SO_2 , which were expected. There were no lines from liquids with stationary bubbles, presumably because the intense blackbody bremsstrahlung radiation intensity swamped them.

Yasui (1998), in a modeling study, examined the effect of the surfactant Triton X-100 on the dynamics of SBSL. Yasui concluded that the dissolved surfactant lowered the magnitude of the light pulse because it inhibited condensation of water vapor inside the bubble during collapse. Free water vapor lowered the temperature inside the bubble at collapse due to the larger amount of water vapor that undergoes endothermal chemical reactions at collapse.

Yasui stated: "It is predicted, based on the hot-spot model, that the radiation is not thermalized inside a bubble in the case of SBSL in a solution of the surfactant in water and that

the [light emission] spectrum of SBSL may deviate from the blackbody spectrum and may have some characteristic lines such as the $-OH$ line (@ 310 nm)."

From the work of Yasui and Didenko et al., we can see that if the severe thermodynamic conditions (peak temperature and pressure) inside stationary collapsing bubbles are mitigated, reducing blackbody bremsstrahlung radiation, it may be possible to see individual spectral lines from excited-state emissions during sonoluminescence. For a comprehensive, state-of-the-art description of SBSL, see the 40-page review paper by Brenner et al. (2002).

7.6.5.3 Multiple-Bubble Sonoluminescence

MBSL is also a phenomenon dependent on ultrasonic energy in a liquid. MBSL occurs as a result of sonic *cavitation* at very high sound pressure levels. Quoting Suslick et al. (1999):

If the acoustic pressure amplitude of a propagating acoustic wave is relatively large (greater than ≈ 0.1 MPa), local inhomogeneities in the liquid can serve as nucleation sites for rapid inertial growth into a cavity of macroscopic dimensions, primarily filled with vapor and dissolved gasses. Such a bubble is inherently unstable, and its subsequent collapse can result in an enormous concentration of energy... A normal consequence of this unstable growth and subsequent collapse is that the cavitation bubble itself is destroyed. Gas-filled remnants from the collapse, however, may serve as nuclei for continuation of the process.

Single bubbles, while creating extreme conditions, simply do not contain sufficient material to be useful for driving chemical reactions in any practical amount. In multi-bubble systems ("cavitation clouds"), on the other hand, the interactions between bubbles as they collapse will lead to both substantial asymmetry and the formation of jets during the collapse, which one may well expect to limit the collapse efficiency. Thus the conditions created during multi-bubble cavitation, which is used of necessity for all sonochemical reactions, will be somewhat less extreme than those created during single-bubble sonoluminescence.

MBSL creates its own bubbles by cavitation; an external gas source is not needed to form bubbles. Thus, *any gas-phase reactants must be dissolved in the host liquid if they are to be detected by their spectral signature during MBSL*. MBSL hot spots have been measured at ca. 5050 K, significantly lower than the peak temperatures reached in SBSL (Suslick et al. 1999). The lower temperatures of the MBSL plasma mean far less bremsstrahlung radiation, allowing spectral lines to be seen from the sonochemical reactions occurring inside the cavitation bubbles.

Suslick et al. (1999) illustrated sharp MBSL line spectra from nonaqueous silicone oil containing metal carbonyl compounds such as $Fe(CO)_5$, $Cr(CO)_6$, $W(CO)_6$, and $Mo(CO)_6$. The MBSL emission lines for these compounds were almost identical to the emission lines when the compound was burned in a flame at ca. 5000 K. MBSL in pure water has been shown to give a strong, broad emission line peak at 310 nm from an excited state of $-OH$. When metal salts are dissolved in water or alcohol, the MBSL spectra show the metal line spectra.

7.6.5.4 Summary

In summary, it appears that MBSL is potentially more useful as an analytical tool than is SBSL. Spectral emission lines are lost in SBSL because of the very high plasma temperatures. On the other hand, MBSL is capable of revealing the emission lines of compounds

and metal ions in solution. The question remains to its detection sensitivity for organic, nitrogen-containing explosive vapors. If MBSL were to be used to detect explosive vapors, the liquid would have to be saturated with the very small concentration of the vapor in air by bubbling the vapor-bearing air through the liquid for a long time. Contrast this procedure with concentrating the analyte by passing air through a cold trap, then using a gas chromatograph/mass spectrometer (GC/MS) to identify and quantify the explosive vapor. The latter method is accurate but expensive. MBSL, if it works, would be far less expensive and portable. It is certainly worthwhile to pursue MBSL as a means to detect trace vapors from explosives.

7.7 Other Means of Substance Detection

7.7.1 Mass Spectrometry

The mass spectrometer (MS) is another important analytical instrument used in medical diagnosis, forensic analysis, geochemistry, etc. While instruments like spectrophotometers and gas chromatographs allow investigators to examine intact molecules, the flame photometer and the MS gain their analytical information by breaking molecules apart. In an MS, the sample molecules are bombarded with energetic electrons, moving atoms, or photons from a laser beam to break them into component parts; some parts are positively charged (positive ions or cations) and others can be neutral (i.e., have zero charge; they are lost), and negatively charged ions (anions) can also be produced under the right conditions. An MS can perform analysis of elements (isotopes), compounds, and mixtures. It can use gas, liquid, or solid samples and is fairly rapid, yielding results in seconds rather than 10s of minutes, as for a GC.

A basic magnetic sector MS design is shown in Figure 7.95. The positive ions are first accelerated in a DC electric field, $E_1 = V_1/d_1$. At the exit slit, S_2 , it can be shown that the positive ions having a positive charge magnitude of one electron (q) and a mass m will have a velocity

$$v = \sqrt{\frac{2qV_1}{m}} \text{ m/s.} \quad (7.259)$$

That is, the velocity is inversely proportional to the square root of the ion's mass. Positive ions with a distribution of velocities determined by their masses next enter a *velocity filter* formed by two electrodes parallel to the velocities of the entering ions. The positive ions are attracted to the cathode by a force $\mathbf{F}_e = E_2q = (V_2/d_2)q\mathbf{i}$. (\mathbf{i} is a unit vector pointing at the cathode.) Because there is a \mathbf{B} field perpendicular to the velocity (into page), the moving ions also experience a magnetic Lorentz force, $\mathbf{F}_m = q(\mathbf{v} \times \mathbf{B})$; the direction of the vector \mathbf{F}_m is given by the *right-hand screw rule*. That is, \mathbf{F}_m points in the direction that a normal right-hand screw would advance if rotated in the direction of rotating \mathbf{v} into \mathbf{B} . Thus, \mathbf{F}_m is opposite to \mathbf{F}_e . Because of the narrow separation (d_2) of the velocity filter's electrodes, only positive ions with velocities such that $\mathbf{F}_m \cong \mathbf{F}_e$ emerge through slit S_3 . This selected velocity can be shown to be

$$v_o \cong \frac{V_2}{(d_2B)} \text{ m/s.} \quad (7.260)$$

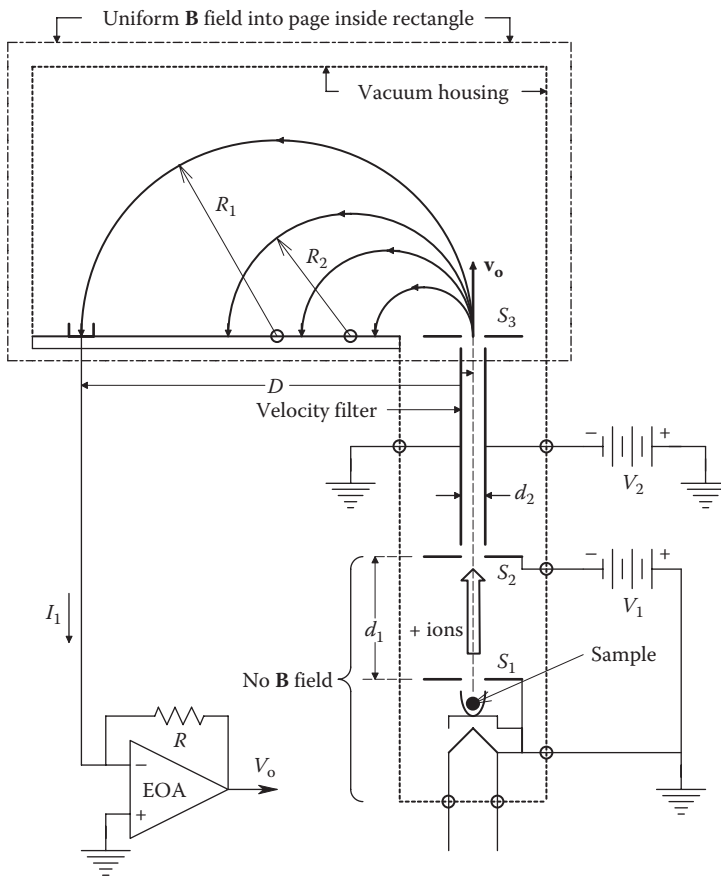


FIGURE 7.95
Diagram of a basic magnetic MS.

Again, the ions with velocity \mathbf{v}_o are acted on by the Lorentz force in the \mathbf{E} field-free chamber of the MS. Because \mathbf{B} is perpendicular to \mathbf{v}_o , the ion trajectories are semicircular. The ions strike the detector electrodes at a distance $D = 2R$ from S_3 . Elementary physics (Sears 1953) tells us that

$$D = 2R = \frac{2v_o m}{qB} = \left(\frac{m}{q} \right) = \frac{2V_2}{d_2 B^2} \text{ m.} \quad (7.261)$$

Thus, the distance D at which the ion beam strikes the collector is proportional to the ratio of mass to charge of the ion exiting the velocity filter. The + ion beam at D can be collected by an electrode (a Faraday cup), and the resulting electron current, I_1 , is converted to a voltage, V_o , by an electrometer OA connected as a transresistor.

Compound identification and quantification can also be carried out on analytes with mass ranges less than 10^3 Da by *negative ion mass spectroscopy* (NIMS). Negative ions are created in a sample by *resonance electron capture* or direct ionization by low-energy electron bombardment of the sample through a buffer gas such as methane. The methane gas slows down the electrons and stabilizes the resultant *anions*. Now, the anions are accelerated

through an electric field toward an anode, thence a velocity filter, thence into the magnet chamber. The anions bend to the right, however, because the Lorentz force has the opposite sign.

Unless the investigator has an a priori knowledge that a certain compound is present in a sample, the identification of a molecular species from its *mass fragment* peaks can often be challenging. Table 7.3 illustrates *commonly lost mass fragments* that lack + charges and thus cannot produce peaks (they cannot be accelerated).

Figure 7.96 shows some common stable mass fragments encountered in mass spectrometry (based on data from Young [1996]).

TABLE 7.3

Commonly Lost Mass Fragments

Approximate Mass	Fragment
15	-CH ₃
17	-OH
26	-CN
28	H ₂ C=CH ₂
29	-CH ₂ CH ₃ , -CHO
31	-OCH ₃
35	-Cl
43	CH ₃ C=O
45	-OCH ₂ CH ₃
91	Benzene-CH ₂

Source: Young, P.R., Basic mass spectroscopy, Organic chemistry online, 1996, <http://people.stfx.ca/tsmithpa/Chem361/labs/spec/MS.htm> (accessed May 09, 2013).

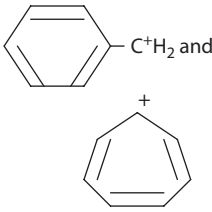
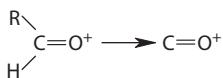
Approximate mass	Common stable cations
43	CH ₃ -C ⁺ ≡O
91	 C ⁺ H ₂ and
m → m - 1	 R C=O ⁺ → C=O ⁺ H

FIGURE 7.96

Some stable mass fragments detected with an MS.

In the analysis of biochemical and drug molecules from respiratory gasses, urine, stool, blood, and tissue samples, the investigator often knows the molecules to expect, and thus an MS system can be set up or *tuned* for a specific molecule in terms of the ionization method and the expected D_s for the expected group of anions. A given analyte molecule can produce as many as 10 or more anions when fragmented and ionized. As in GC, the integral of a peak's area is proportional to the quantity of an ion present.

General-purpose MS instruments with few exceptions are generally large, heavy, and fixed. Smaller, portable MS systems have been designed to measure respiratory gasses such as O_2 , CO_2 , N_2 , and halothane. Ion ambiguities exist so that O_2 and CO_2 cannot be measured when ether ($C_4H_{10}O$) or N_2O is present (Webster 1992). N_2 and carbon monoxide also have the same mass number (28).

The *quadrupole MS* (QPMS) was developed in the mid-1950s by Wolfgang Paul and associates at the University of Bann. The QPMS is a lighter, more portable instrument than conventional MSs because its ion selectivity does not depend on a magnetic field (or the need for heavy magnets). Instead, ion selectivity depends entirely on the interaction of moving ions with combined AC and DC electric fields maintained between four, cylindrical electrodes, as shown in Figure 7.97. The voltage applied to the two electrode pairs is the sum of a DC potential and an AC, *RF* voltage; $V_+ = V_{DC} + V_{rf} \cos(\omega t)$, and $V_- = -[V_{DC} + V_{rf} \cos(\omega t)]$. The QPMS is *tuned* for specific m/q values by linearly increasing the V_{DC} voltage between 0 and 300 V while simultaneously increasing the peak *RF* voltage, V_{rf} , from 0 to 1.5 kV. It can be shown that the complex electric field between the tubes causes the entering ions to effectively *run the gauntlet*. If an ion has an inappropriate mass/charge ratio, it will collide with an attracting tube and lose its \oplus charge. Such noncharged molecular fragments are pumped out by

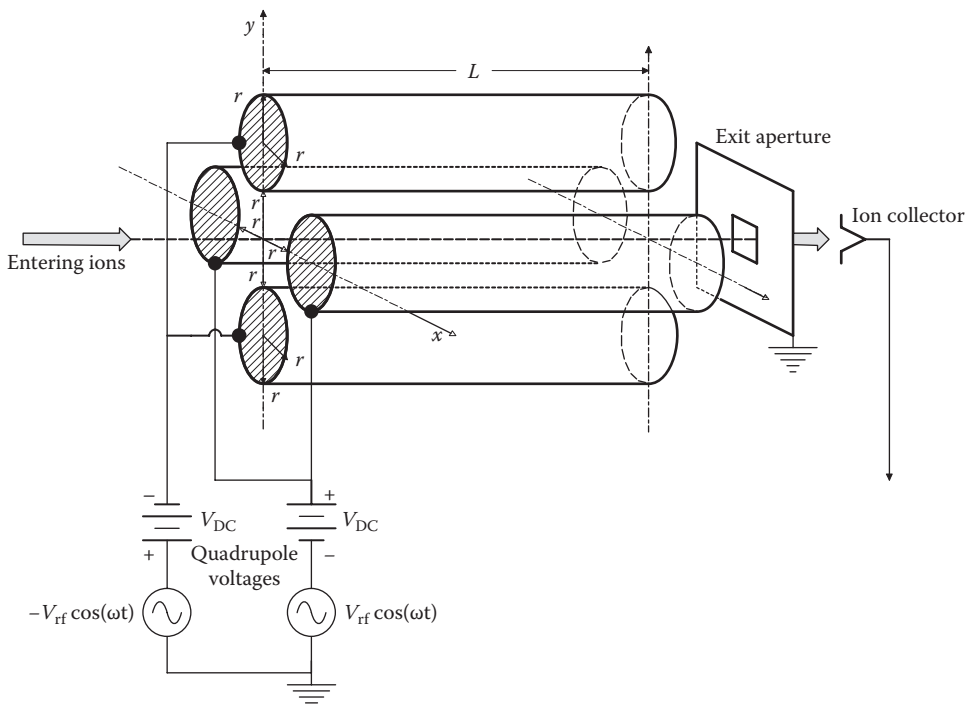


FIGURE 7.97
Quadrupole (electrostatic) MS (QPMS).

the vacuum system. An ion with the *tuned* mass/charge will describe a spiral or corkscrew path between the electrodes, emerging through the exit window and striking the ion collector, its charge contributing to the collector current. The voltages on the QPMS's electrodes are increased linearly in time, providing a swept tuning of the ions' mass/charge ratios. As the voltages increase, larger and larger mass ions can pass through the quadrupole electrodes without collision. Thus, by plotting current peaks from the detector vs. the electrode sweep voltage, one can obtain a rapid mass spectrogram. The computer plots relative ion current vs. m/z . The advantages of QPMSs are that they are relatively inexpensive, smaller, and faster (a scan completed in milliseconds) than conventional electrostatic/magnetic sector MS systems. They also give relative mass readings, have good reproducibility, but do not have the resolution of conventional MSs. Like gas chromatographs, QPMS machines must be calibrated with standard samples. Figure 798 shows a QPMS spectrogram for a methanol sample. Note that the molecular ion has the highest (32) m/z ratio. Relative abundances of the fragments are peculiar to the ionization method used. The very small peaks are due to the natural presence of isotopes of C, H, and O in methanol (see Table 7.4). Figure 799 illustrates a typical QPMS spectrum of exhaled air.

From Table 7.4, we see, for example, that for every 100 ^{12}C atoms, we will typically find 1.11 ^{13}C atoms, while for every 100 ^{79}Br atoms, we find 98 ^{81}Br atoms, a surprisingly large ratio. Thus, a mass spectrogram of a compound containing a bromine atom will exhibit curious double peaks of almost the same size for those ions containing Br.

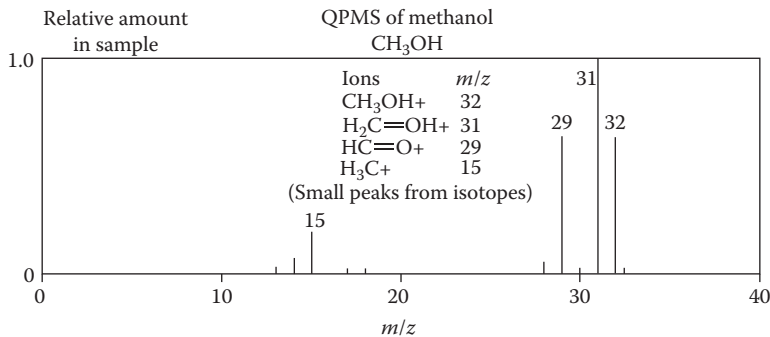


FIGURE 7.98
QPMS spectrogram of methanol.

TABLE 7.4

Relative Abundance of Some Naturally Occurring, Nonradioactive Isotopes

Natural Isotope at 100%	Isotope and Rel. Abundance	Isotope and Rel. Abundance
Carbon: ^{12}C	^{13}C : 1.11	
Hydrogen: ^1H	^2H : 0.16	
Nitrogen: ^{14}N	^{15}N : 0.38	
Oxygen: ^{16}O	^{17}O : 0.04	^{18}O : 0.20
Sulfur: ^{32}S	^{33}S : 0.78	^{34}S : 4.40
Chlorine: ^{35}Cl		^{37}Cl : 32.5
Bromine: ^{79}Br		^{81}Br : 98.0

Source: Northrop, R.B., *Non-Invasive Instrumentation and Measurements in Medical Diagnosis*, CRC Press, Boca Raton, FL, 2002.

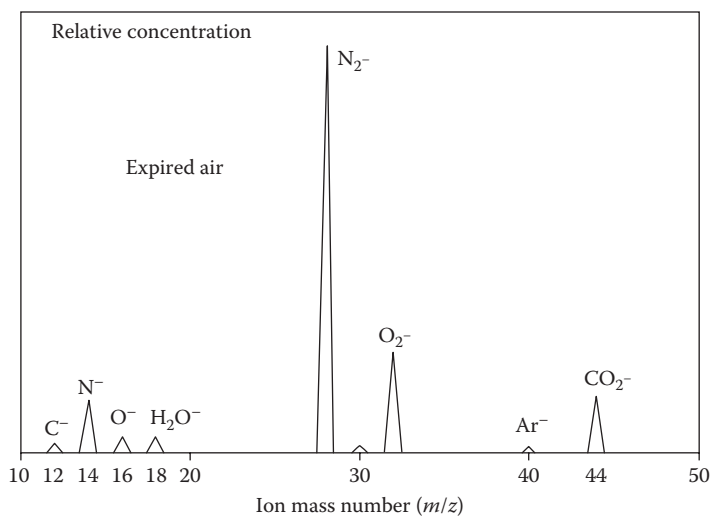


FIGURE 7.99
QPMS of exhaled human breath.

Several other types of MS exist in addition to the QPMS and magnetic sector MS described earlier. These include, but are not limited to, the *time-of-flight MS*, the *ion cyclotron resonance MS*, and the *Fourier transform ion cyclotron resonance (FTICR) MS*.

In summary, we see that compound separation and identification using an MS can be a considerable challenge. An MS gives the total number of each constituent atom in a compound (the empirical formula) but can seldom provide clues about molecular structure that can be given, for example, by an optical spectrophotometer. Often, a molecular formula is found by trial and error from the ion mass peaks, using information from isotope abundance as well. Because gas chromatographs separate pure compounds in time as they traverse the column, feeding the column output into an MS simplifies the analysis of an unknown sample under the assumption that the MS will be analyzing one pure eluent compound at a time, hence the popularity of GC/QPMS systems.

7.7.2 Gas Chromatography

Gas chromatography (GC) is an exquisitely sensitive chemical analytical tool that is very competitive with spectrophotometry in terms of the analytes it can quantify in medically derived samples. If an analyte is stable (does not decompose) when in the gas or vapor phase at temperatures up to 400°C, it probably can be identified and quantified by GC. GC works by the simple principle that the volatile components of a sample injected as a bolus travel through the GC's column at different speeds. Each sample component is adsorbed by the column's stationary phase and then released, forming a continuous, traveling wave of adsorbed component. The speed of a given traveling wave depends on many physical and chemical factors of the GC system. Of particular importance is the fact that there is very little dispersion (broadening) of an individual analyte's peak as it propagates through the column. Thus, the time that the peak concentration of a particular analyte exits the column is peculiar to that analyte, given identical conditions of carrier gas, carrier gas flow rate, column temperature, column length and inside diameter, and column stationary phase adsorber. A GC is well suited to separate mixtures of analytes in samples of complex composition. A GC can

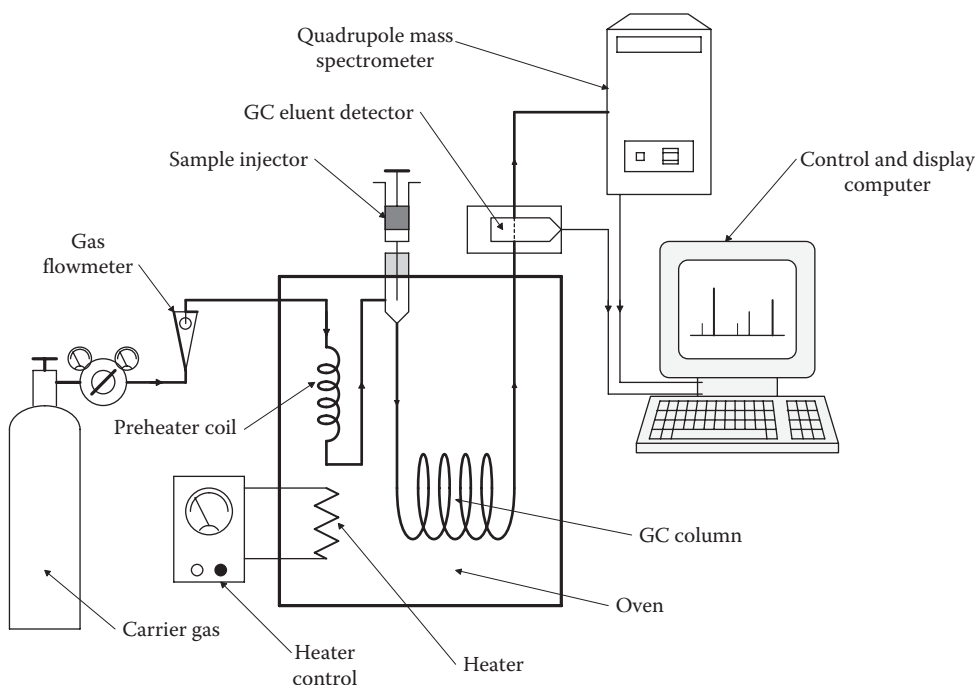


FIGURE 7.100
Diagram of a typical GC/MS system.

be used to identify and quantify biochemicals such as alcohol, acetone, and various steroid hormones and drugs such as various tricyclic antidepressants, theophylline, and various opioids. The schematic of a typical GC/MS system is shown in Figure 7.100.

A basic laboratory bench GC has seven key components (cf. Figure 7.100): (1) A source of *inert carrier gas*, such as dry nitrogen, helium, or argon. The gas literally carries the sample and analyte through the column. (2) A sample injection port. (3) The *capillary column* is a long, thin tube of stainless steel or fused silica coated on the inside with polyimide resin. Column lengths can range from 10 to 60 m, with 20–30 m being most typical. The *resolution* of a GC (ability to separate two nearly coincident eluent peaks) is proportional to the square root of the column length. Capillary column diameter is typically from 0.32 to 0.25 mm. As column diameter decreases, the retention of a given solute will increase, other factors being constant. This means that a smaller-diameter column provides better resolution (wider separation of the eluent peaks). (4) The inside of a GC column is coated with a thin layer of a thermally and chemically stable *stationary phase absorber* (SPA). The thickness of the SPA is another critical parameter affecting GC separation resolution. A thicker SPA will give greater solute retention, hence better resolution of adjacent eluent peaks. SPA thicknesses can range from 0.25 to 1 μm in capillary columns; the SPA coats the inside walls of the column. The materials used for SPAs are described below. (5) The entire column is placed inside an *oven*. The oven temperature can be held constant or programmed to increase (e.g., from 80°C to 280°C at 5°/min). The temperature limits and its rate of increase are set according to the particular analysis. (6) The *detector* senses the changes in eluent gas composition from pure carrier gas when boluses of sample constituents (and the desired analyte) exit the column. There are many kinds of detectors; these are described in detail below. Detector sensitivities allow limits of detection ranging from 100 ppm to 100 ppb (analyte to carrier gas). (7) The detector

output consists of voltage peaks of different heights and widths from sample components that exit the column at different times. A complete GC run may take from 5 to 30 min; GCs are not fast instruments! The GC detector output voltage is digitized, and the GC's computer integrates the area of each peak to determine the concentration of each sample component (eluent). An analyte sample of known concentration is injected into the column for calibration purposes. The time it exits is peculiar to that analyte, and its area is proportional to the concentration, providing it is in the linear range of the column and detector.

There are many types of GC detectors. Some types of detectors respond to any analyte exiting from the column; others are specific for certain chemical classes of eluents, such as chlorinated hydrocarbon insecticides. As in many branches of measurements, the threshold detection concentration and resolution of a GC system are set primarily by noise arising in the detector and its electronic amplifiers. The *minimum detectable amount* (MDA) of an eluent is defined as the concentration or amount of analyte that will produce a minimum output peak voltage twice the output RMS noise voltage. Thus, the practical lower limit of detector operation is set by its MDA. All GC detectors also have a *linear dynamic range* (LDR) of analyte concentration above their MDAs in which their outputs follow the linear relation: $V_o = b + m[A]$ to within $\pm 5\%$, where $[A]$ is the analyte concentration in g/mL or ppm. The names and characteristics of some GC detectors are listed in Table 7.5.

We shall examine the operating mechanisms of two of these detectors in detail. The thermal conductivity detector (TCD) is one of the simplest, general-purpose GC detectors; it is generally used with helium carrier gas. Figure 7.101 illustrates the circuit of this

TABLE 7.5

Summary Characteristics of the Major Detectors Used in GC

Detector Type	MDA (g/mL)	LDR (Decades)	Comments
TCD	10^{-7}	≈ 5	A general detector used with He carrier gas. Temp. differences on Wheatstone bridge resistor arms due to analyte sensed.
Flame ionization detector (FID)	10^{-12}	5–7	A mass-sensitive detector for C–H bonds. H_2 is burned, creating ions from analytes.
PID	10^{-12}	5–6	Sensitive to aromatics and olefins. Uses 10.2 eV UV lamp to ionize analytes.
ECD	10^{-14}	3–4	Uses radioisotopes to ionize halogens, quinones, peroxides, and nitro groups. Used for insecticides and PCBs.
Flame photometric detector (FPD)	10^{-11}	3–5	Used for S- or P-containing analytes. H_2 flame causes atomic emission of S at 394 nm and P at 526 nm. A PMT is used.
Electrical conductivity detector (ELCD)	5–10 pg (halogens)	5–6	Eluents react at high T with reaction gas.
	10–20 pg (S)	4–4	Products dissolved in solvent are passed
	10–20 pg (N)	3–4	through electrical conductivity cell.
Mass spectrometry (MS)	1–10 pg (selected ion-monitoring mode)	5–6	An MS replaces detector. The MS measures the mass/charge ratio of ions fragmented at high T by e-bombardment.
Nitrogen–phosphorus detector (NPD)	1–10 pg	4–6	Similar to FID except uses <i>rubidium bead</i> to enhance sensitivity to N compounds by $\times 50$ and P compounds by $\times 500$.

Source: Northrop, R.B., *Non-Invasive Instrumentation and Measurements in Medical Diagnosis*, CRC Press, Boca Raton, FL, 2002, Table 8.2.7-1.

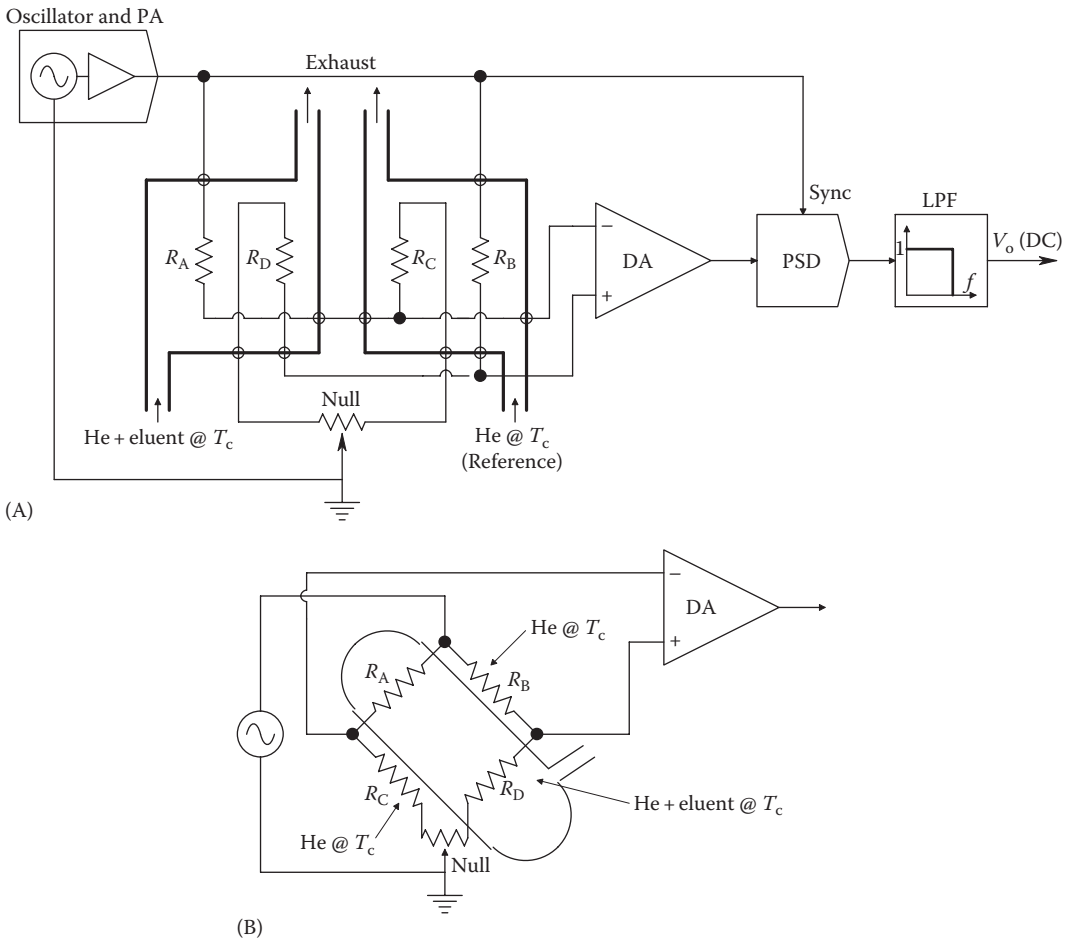


FIGURE 7.101 (A) GC thermal conductivity eluent detector. (B) Thermal detector Wheatstone bridge circuit.

detector. Note that it is a simple Wheatstone bridge; the resistors of two opposite arms are surrounded with carrier plus eluent gases at temperature T_C ; the other two arms are surrounded with pure He at column temperature, T_C . The resistors are platinum or nichrome. Current is passed through the bridge arms so that they heat to a temperature, $T_{B'}$ above T_C . The moving helium gas surrounding the resistors conducts heat away from them, lowering their temperature toward T_C . (Helium is unique in that it has a very high heat conductivity.) The drop in resistor temperature is given by the relation

$$(T_B - T_R) = \Delta T = P_R \Theta_{R\#} > 0, \dagger \quad T_B < T_R < T_C \tag{7.262}$$

where

- T_R is the actual equilibrium resistor temperature in moving He
- T_B is the reference temperature
- P_R is the electrical power dissipated in each resistor
- Θ_R is the thermal resistance seen by a resistor in moving He

In general, $\Theta_R = \Theta_{R_0} + \beta[A]$. That is, the eluent analyte gas *decreases* the ability of the pure He to conduct heat away from the two R_A resistors. Θ_{R_0} and β are positive constants, and $[A]$ is the eluent analyte concentration in the He carrier gas at the TCD. A metal resistor has a positive tempco, that is, its resistance increases as its temperature increases. This property can be approximated by the relation

$$R(\Delta T) \cong R_0(1 + \alpha\Delta T), \quad \alpha > 0. \quad (7.263)$$

R_0 is the resistance at the reference temperature, T_B
 α is the tempco

If we combine relations 7.278 and 7.279, we can write

$$\Delta T = I_B^2 R_0 (1 + \alpha\Delta T) \Theta_R. \quad (7.264)$$

Solving for the equilibrium ΔT , we find

$$\Delta T = \frac{I_B^2 R_0 \Theta_R}{(1 - \alpha I_B^2 R_0 \Theta_R)}. \quad (7.265)$$

I_B^2 is the mean-squared AC current through each resistor (assumed constant). $I_B \cong V_s / 2R_{He}$. Knowing the temperature drop, we can find the resistance from relation 7.279:

$$R(\Delta T) \cong \frac{R_0}{(1 - \alpha I_B^2 R_0 \Theta_R)}. \quad (7.266)$$

For pure He,

$$R_{He} = \frac{R_0}{(1 - \alpha I_B^2 R_0 \Theta_{R_0})}. \quad (7.267)$$

When an eluent analyte is present, Equation 7.282 becomes

$$R_A = \frac{R_0}{[1 - \alpha I_B^2 R_0 (\Theta_{R_0} + \beta[A])]} = \frac{R_0}{[1 - \alpha I_B^2 R_0 \Theta_{R_0} (1 + \beta[A] / \Theta_{R_0})]}. \quad (7.268)$$

Now, the general relation for Wheatstone bridge unbalance voltage with two active arms is

$$V_B = V_1 - V_2 = V_s = \frac{R_A}{R_{He} + R_A} - V_s \frac{R_{He}}{R_{He} + R_A}. \quad (7.269)$$

So

$$\frac{V_B}{V_s} = \frac{R_A - R_{He}}{R_A + R_{He}} = \frac{\frac{R_0}{[1 - \alpha I_B^2 R_0 \Theta_{R_0} (1 + \beta[A] / \Theta_{R_0})]} - \frac{R_0}{[1 - \alpha I_B^2 R_0 \Theta_{R_0}]}}{\frac{R_0}{[1 - \alpha I_B^2 R_0 \Theta_{R_0} (1 + \beta[A] / \Theta_{R_0})]} + \frac{R_0}{[1 - \alpha I_B^2 R_0 \Theta_{R_0}]}}. \quad (7.270)$$

After some algebra, Equation 7.286 can be reduced to

$$\frac{V_B}{V_s} \cong \left(\frac{R_{He}}{2} \right) \alpha I_B^2 \beta [A]. \tag{7.271}$$

That is, the balanced bridge output is zero, and any eluent gas mixed with He will *reduce* the cooling (increase Θ_R), allowing the two R_A resistors to warm slightly, raising their resistance, unbalancing the bridge, and producing an output. An advantage of the TCD is that no chemical change occurs in the eluents. Thus, the eluents can be individually captured by condensation in nearly pure form or passed on to another type of detector for another detailed analysis, such as a QPMS.

The *electron capture detector* (ECD) is an exquisitely sensitive means of detecting chlorinated pesticide residues and PCBs in and on foods and in humans and animals. Mass sensitivities to halogenated hydrocarbons in the mass range from 0.1 to 10 pg are possible with the ECD. Figure 7.102 illustrates a schematic cross section of an ECD. Ionization is

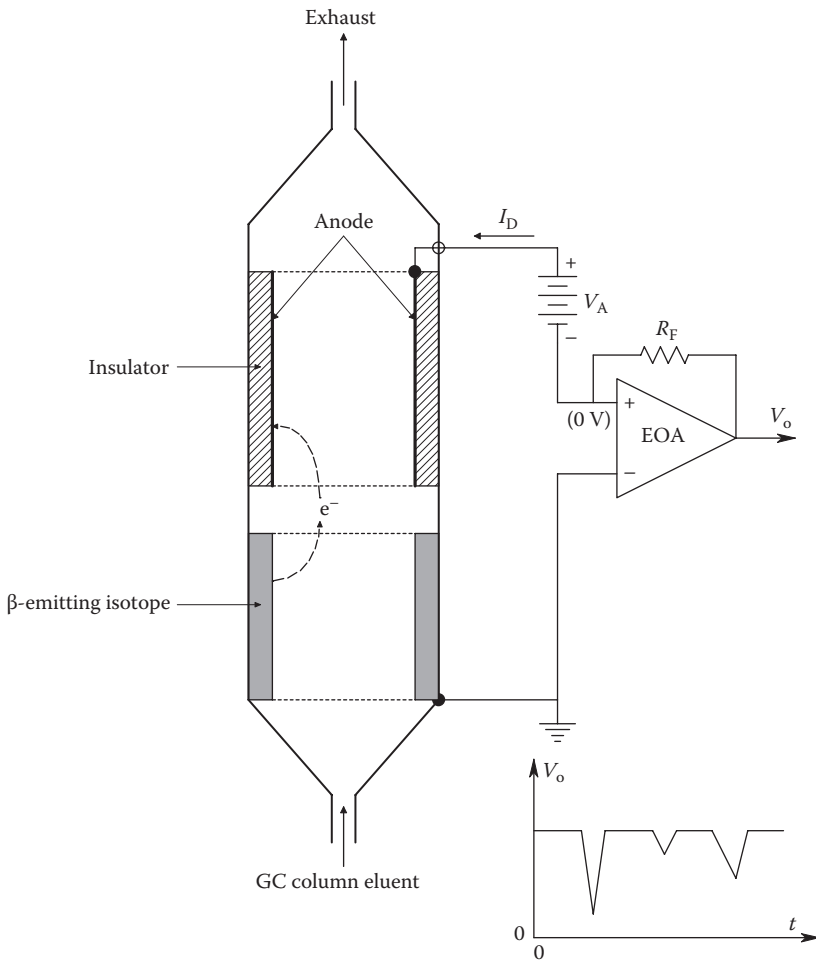


FIGURE 7.102
Electron capture, GC eluent detector.

caused by beta-emitting isotopes of either tritium (^3H) or nickel (^{63}Ni) foil. The electrons are accelerated to and captured by the anode electrode. Chlorinated hydrocarbon molecules in the detector chamber capture some of the radioelectrons and reduce the net anode current, signifying the presence of the analyte. The ECD is insensitive to amines, alcohols, and hydrocarbons. The carrier gasses used with ECD are N_2 or Ar/CH_4 ; gas temperature ranges from 300°C to 400°C .

As you can see from Table 7.5, many other application-specific detectors exist.

Common stationary-phase, GC column coatings (or fillings) range from the mundane zeolite particles (*kitty litter*) to styrene beads, aluminum oxide particles to a host of thermally stable liquid polymers, such as the various polysiloxanes and polyethene glycols used with capillary columns. The solid-phase, porous open layer tubular (PLOT) columns are very retentive and are used to separate analytes whose peaks may be nearly coincident using a conventional capillary column.

There has been research and development directed at producing lightweight, field-portable GCs. Defiant Technologies, Inc., Albuquerque, NM, has developed one such system, called the Frog-4000TM (Frog 2013). This is a 4.8 lb, $10 \times 7.5 \times 14.5$ " field-portable GC that uses a micro-gas chromatography (μGC) column and a photoionization detector (PID). It uses scrubbed ambient air as a carrier gas and can detect VOCs such as benzene, toluene, ethylbenzene, and xylenes in water and soil, in situ. It also can quantify alcohols, ketones, naphthalene, etc. The Frog-4000TM can be connected to a computer with an RS-232 port or Bluetooth and has a cycle time of 5 min. Defiant also offers another handheld, field-portable GC, the Canary-3TM. Like the Frog-4000TM, the Canary-3TM uses scrubbed ambient air for a carrier gas. Eluents from the microcapillary column are detected by two surface acoustic wave (SAW) detectors; this GC has a cycle time of only 2 min. It can be used to quantify chemical weapons, pesticides, meth lab vapors, fragrances, flavorants, and other VOCs. The Canary-3 weighs 3.8 lbs and measures $12" \times 6" \times 4"$.

Agilent Technologies has developed *plug and play* micro GC column modules for specific GC applications that plug into a host GC system (Agilent 2013).

The design, fabrication, and characterization of a MEMS-based micro-gas chromatograph were described in the MS dissertation by Zareian-Jahromi (2009).

7.7.3 Surface Plasmon Resonance

SPR sensors are a relatively new chemical analytical tool. They allow rapid, specific determination of the concentration of a variety of medically, biologically, and environmentally important analytes. For example, specific bacteria, antibodies, AB-specific lymphocytes, nucleic acid oligos, theophylline, caffeine, NO_2 , pesticides, explosives, and controlled substances (opioid drugs) have been sensed with SPR (BioNavis 2012, Biosensing 2012, SensiQ 2012).

Threshold sensitivities for certain analytes had been reported as low as 0.05 ppb, and Zijlstra et al. (2012) recently reported the SPR sensing of single protein molecules using a single, biotin receptor-coated gold nanorod. Zijlstra et al. stated that the sensitivity of their device was ca. "700 times higher than state-of-the-art plasmon sensors, and is intrinsically limited by spectral diffusion of the surface plasmon resonance." They used three different analytes in their single-molecule SPR studies: *streptavidin* (53 kDa), *antibiotin* (150 kDa), and *streptavidin-R-phycoerythrin conjugate* (300 kDa).

SPR technology is relatively simple and inexpensive to implement, compared with analytical systems such as HPLC, mass spectrometry, IR spectrometry, and GC, and it lends itself well to field measurements. For noninvasive measurements, the analytes are

presumably derived from urine or saliva, smears from mucous membranes, etc. There is no reason, however, why SPR sensors cannot be used on blood or respiratory gasses to look for any analyte (chemical, AB, bacteria, etc.) therein.

SPR is a *quantum* phenomenon that occurs when a beam of monochromatic, LPL is reflected off a thin metal film, vapor deposited on one side of a glass prism, or when a beam of LPL is incident at a critical angle on a gold-coated diffraction grating. For illustrative purposes, we will examine in detail the prism SPR system using the so-called Kretschmann geometry. This system is shown schematically in Figure 7.103. A thin (ca. 50 nm) film of conducting metal such as gold or silver is vapor deposited on one face of a prism or on the flat face of a half-round rod. A beam of monochromatic LPL is directed into the prism or rod so that it strikes the gold film face at an angle of incidence, θ_i . The incident beam's E vector must lie in the plane of incidence (be in the transverse magnetic [TM] mode) for SPR to occur. In intimate contact with the other side of the gold film is a thin film of the analyte having a permittivity, ϵ_a . It has been experimentally observed that when the incoming beam's wavelength and angle of incidence have unique, critical values, the intensity of the

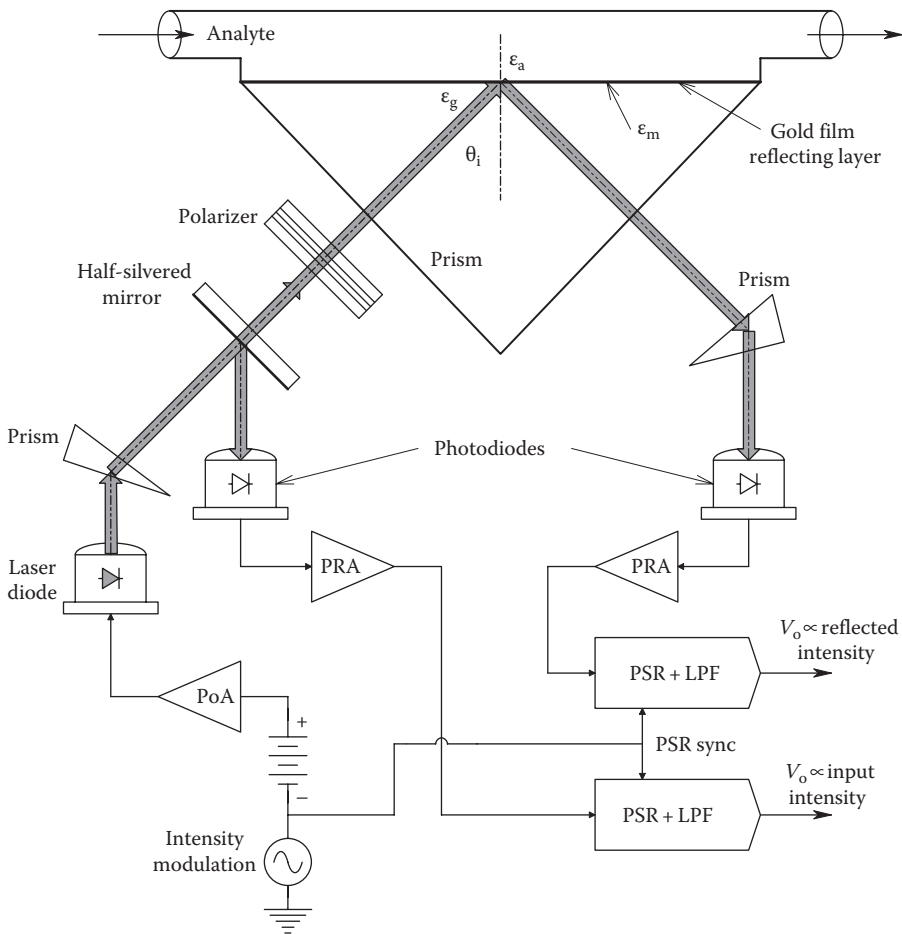


FIGURE 7.103
Diagram of an SPR system using the Kretschmann geometry.

reflected beam reaches a *minimum*. The depth of the null in the output beam intensity is a function of how much energy from the input LPL beam is coupled into the generation of surface plasmons in the metal film. The degree of coupling is a function of λ , θ_i , and, most importantly, the dielectric constant of the analyte material.

Surface plasmons are defined as induced, wavelike fluctuations in the density of conduction band electrons in the thin metal film. These fluctuations exist in both space and time, that is, they can be thought of as traveling waves induced by the incident TM, LPL. The basis for using SPR as an analytical chemical tool is based on the conditions required for resonance. Under conditions of non-SP resonance, the incident beam of LPL reflects off the metallized surface of the prism and exits the prism to a photosensor. The metal film effectively forms a conducting mirror surface, and conventional refraction and reflection optical laws apply.

The velocity of light in vacuo is $c = 2.998 \times 10^8$ m/s. In transparent liquids or glass, light travels more slowly; the ratio of the speed of light in vacuo to the speed of light in the medium is defined as the *refractive index* of that medium, n_m . That is, $n_m \equiv c/v_m$. From EM theory, $c \equiv 1/\sqrt{\epsilon_0\mu_0}$, and $v_m = 1/\sqrt{\kappa_m\epsilon_0\mu_m}$. However, in nonmagnetic materials, the magnetic permeability $\mu_m \cong \mu_0$. Thus, we can write the refractive index as

$$n_m \cong \sqrt{k_m}. \quad (7.272)$$

v_m (and n) are generally functions of frequency. An EM wave in a medium or free space such as light or surface plasmon waves that basically exist in two dimensions can be described by their *wave vector*, \mathbf{k} . \mathbf{k} is directed along the direction of wave propagation, and its magnitude in vacuo for EM waves is

$$k_o = \frac{\omega}{c} = \frac{2\pi\nu}{c} = \frac{2\pi}{\lambda}. \quad (7.273)$$

In the prism glass with refractive index, n_1 , the magnitude of the EM wave vector is (Kraus 1953)

$$k_1 = \frac{\omega}{v_1} = \frac{\omega\sqrt{k_1}}{c}. \quad (7.274)$$

Surface plasmons are generated on the metal film under the condition that the magnitude of the wave vector in the glass incident on the gold film equals the magnitude of the wave vector at the metal/analyte interface. This can be written as

$$k_i = \left(\frac{2\pi}{\lambda}\right) \sin(\theta_i) \sqrt{\kappa_1} \mp \left(\frac{2\pi}{\lambda}\right) \sqrt{\frac{(\kappa_a + |\kappa_m|)}{(\kappa_a + |\kappa_m|)}} \quad (7.275)$$

where

κ_1 is the dielectric constant of the glass prism

κ_a is the dielectric constant of the analyte

$\kappa_m = \kappa_m' + j\kappa_m''$ is the complex dielectric constant of the gold film

All dielectric constants are generally functions of ν (or λ). Canceling like terms and solving for the angle of incidence inside the prism, we have the angle criterion for SPR:

$$\theta_{iR} \cong \sin^{-1} \sqrt{\frac{\kappa_a \kappa_m}{(\kappa_a + \kappa_m) \kappa_1}} \tag{7.276}$$

The resonance condition is due to momentum matching of incident photons with plasmons in the metal. The fact that the permittivity of the analyte layer (typically ca. 250 nm thick) on the other side of the metal affects SPR may be due to the evanescent field expanding through the metal and coupling into SPs at the analyte surface. The resonance angle, θ_{iR} , is exquisitely sensitive to the dielectric constants of the metal and the analyte in contact with it. Thus, any chemical reaction that takes place at the metal surface, such as binding of antibodies to metal-bound antigen, will affect ϵ_a and the value of θ_{iR} . Thus, surface reactions of analyte can be used to sense antibodies, or if the antibodies are bound to the metal, it can sense antigen molecules such as those on bacterial and viral surfaces or analyte molecules in a solution or suspension.

Figure 7.104 illustrates typical SPR curves for a Kretschmann prism system receiving monochromatic light. Note that when the analyte index of refraction increases due to AB bonding at the metal–analyte surface, κ_a also increases to κ_a' , and the SPR curve as a function of incidence angle shifts to the right (a larger reflected angle) and broadens. Note that $n_a \cong \sqrt{\kappa_a}$. Thus, the intensity measured at angle θ_{im} increases, while the intensity minimum moves to θ_{iR}' . Both shifts, either taken together in a formula or as separate phenomena, can be used to quantify the extent of the binding reaction at the metal surface. The depth of the null at SPR depends in part on the thickness of the metal film (Foster et al. 1994).

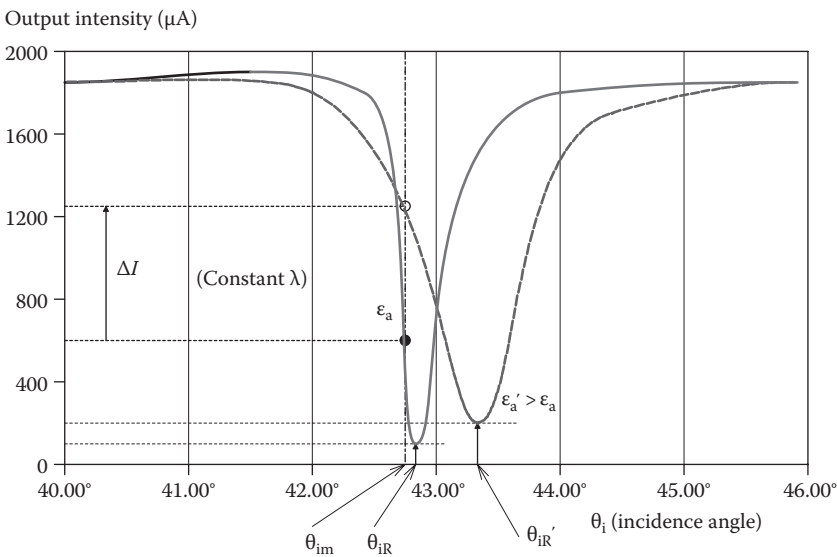


FIGURE 7.104

Reflected light intensity as a function of incidence angle, θ_i , for a Kretschmann SPR system. The solid curve is the system response in the absence of analyte; the dashed curve is obtained with the analyte molecules in intimate contact with the gold film. Monochromatic light is used. Note the peak shift.

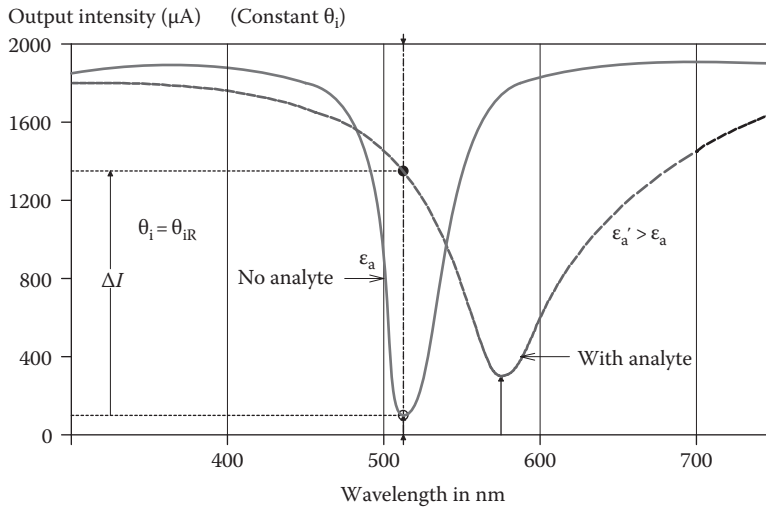


FIGURE 7.105

An alternate mode of operation of the Kretschmann SPR system is to hold the incidence angle constant and vary the input beam wavelength. Again, there is a peak shift with analyte present (dashed curve). At a fixed wavelength (e.g., 512 nm), the difference between the output intensity for no analyte (solid curve) and for analyte is an increasing function of analyte concentration.

Another way of using the Kretschmann system is to set the incidence angle at the SPR null for some standard source λ . The source λ is then varied by a monochromator, and the output intensity is plotted as a function of λ as a surface reaction takes place. Typical intensity vs. λ curves are illustrated in Figure 7.105. Because the permittivities of the glass, metal film, and analyte are functions of λ , we again see a shifting and broadening of the SPR intensity curves with λ as AB binding at the metal surface takes place. While varying the source λ can yield good analytical results, it makes the SPR system more expensive because of the need for a precision monochromator.

Note that the design of SPR devices for chemical analysis is a rapidly growing field. One alternative configuration of SPR sensor places the analyte solution over the surface of a plastic diffraction grating whose surface has been vapor deposited with gold, silver, or aluminum (gold is generally preferred). The plastic top of the grating acts as an ATR prism where light reflected from the grating where the beam strikes it initially is reflected back many times to the grating surface. Such a grating SPR design was proposed by Simon (1998). Figure 7.106 shows a side view of Simon's *long range*, SPR grating system. SPR occurs at a critical beam input angle, θ_i , giving a minimum of output light intensity when the coupling component of the monochromatic, TM-polarized, input beam wave vector satisfies the relation

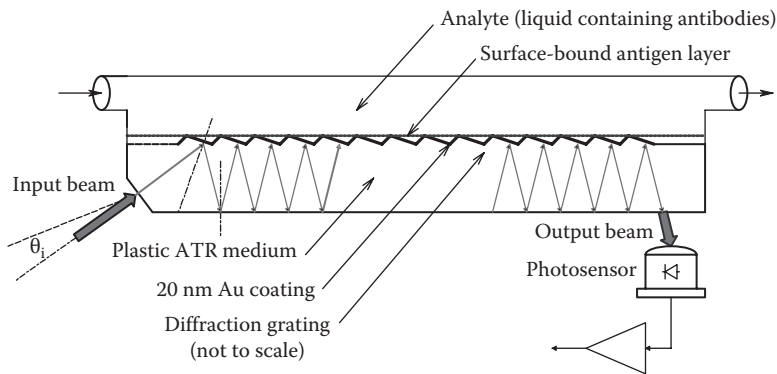
$$\left(\frac{2\pi}{\lambda}\right)\sin(\theta_i) \neq \left(\frac{2\pi}{b}\right) \mp \left(\frac{2\pi}{\lambda}\right)\sqrt{\frac{(\kappa_a\kappa_m)}{(\kappa_a + \kappa_m)}} \quad (7.277)$$

where

b is the grating constant

κ_a is the dielectric constant of the analyte

κ_m is the magnitude of the complex dielectric constant of the metal film

**FIGURE 7.106**

SPR system can be made more sensitive by using a metal film-covered diffraction grating cut into one face of an ATR plate. The metal film is coated with a thin layer of molecules that have affinity to bind with the analyte. SPR systems have been operated with both liquid- and gas-phase analytes. To make the SPR system an effective analytical instrument, it must be possible to remove all bound analyte between readings.

A very thin layer of antigen with high affinity to the AB to be detected is chemically bound to the 20 nm gold film over the grating surface. Antibodies in the analyte liquid bind to the protruding antigen molecules, changing κ_a at the gold–solution interface. This change of dielectric constant changes the interfacial refractive index and *retunes* the SPR to a new input angle, θ_i . Either by changing θ_i to regain a minimal light output or by measuring the *increase* in light output due to $\Delta\kappa_a$, the number or density of bound antibodies can be quantified. A possible disadvantage of Simon's grating SPR configuration is the formation of bubbles on the grating surface and the fact that the measurement is a *one-shot* event; a new, virgin, antigen surface must be reapplied to the gold film and the output nulled again before another AB assay can be done. Simon (1998) claimed that the output signal from the ATR layer increases by approximately fivefold when about 2 nm of antigen–AB complex forms at the gold surface. Also, a refractive index change in the analyte of less than 1% causes a 0.5° shift in θ_i to renul the system.

Other embodiments of the grating SPR system place the analyte over the gold film-covered grating; the monochromatic, TM-polarized light is directed through a thin layer of analyte onto the grating. Jory et al. (1995) reported on an exquisitely sensitive grating SPR system in which an acoustooptical tunable filter (AOTF) element was used to control the wavelength of the incident beam to a precision of 0.0005 nm. They used their system to measure the concentration of NO_2 gas in N_2 (a gas-phase analyte). By depositing a thin layer of *phthalocyanine* over the old coating, a wavelength shift of -0.004 nm renulled the system from zero concentration when 0.01 ppm NO_2 in N_2 was applied. They claimed that the sensitivity of their system allowed detection of changes in the refractive index of the gas of 1×10^{-6} ! The fact that very small concentrations of gas are detectable suggests that the grating SPR technology might be developed in the future to detect medical gas concentrations or even trace gasses emanating from controlled drugs and explosives.

At this writing, trained dogs have the best record for detecting explosives and controlled drugs. For example, dogs can detect *amyl acetate* vapor at concentrations ca. 400 times lower than humans and *alpha-ionone* at concentrations ca. 10^3 – 10^4 lower than humans (Johnston 1999, Olfaction 2010).

The potential application of SPR sensors in medical diagnosis is enormous and is just beginning to be realized (Geddes and Lawrence 1997). Many manufacturers now make SPR systems, for example, GE Healthcare: Life Sciences and Reichert SPR systems. As we have seen, both gas- and liquid-phase sensing are possible. Practically any analyte that can react with a reactant bound to the gold film surface on a prism or grating with a strong affinity can be sensed by SPR. The reaction must cause a change in the refractive index or the permittivity at the gold surface in order to affect the SPR conditions (λ and/or θ_i). SPR antigen–AB reactions can be used to sense specific antibodies, bacteria, viruses, proteins, hormones, cytokines, etc. Note that specific, monoclonal antibodies can also be bound to the gold film in order to sense any protein or molecule for which they can be made specific.

The problem of quick, efficient, SPR sensor regeneration (cycle time) remains to be solved, however. Once a bound surface reactant has combined with the analyte, the analyte must be totally removed before the next measurement without affecting the bound reactant, or the complex must be removed and the surface reactant layer must be renewed or rejuvenated.

7.8 Temperature Measurements

Many physical and chemical phenomena and physical *constants* are found to be functions of temperature and thus can be used to measure temperature. Temperature-dependent properties and constants include resistance, dielectric constant, and the magnetic permeability and susceptibility (of paramagnetic salts). Other temperature-sensitive phenomena include linear and volume expansion of solids and gasses, generation of the Seebeck (thermoelectric EMF) by thermocouples, and the generation of Johnson (thermal) white noise by resistors.

Scientific temperature measurements are generally made using the Celsius (centigrade) or Kelvin scales. Absolute zero (thermodynamic zero) occurs at 0 K or -273.15°C . That is, $\text{K} = ^\circ\text{C} + 273.15$. While most of the civilized world uses the Celsius scale for such mundane things as cooking and weather reports, the use of the Fahrenheit scale is dominant in the United States for these applications. The nominal boiling and freezing temperatures of water were originally taken as the two calibration points for linear temperature scales; 100° and 0° are those respective temperatures in the Celsius scale, and 212° and $+32^\circ$ are boiling and freezing in the Fahrenheit scale. It is easy to derive a conversion formula between $^\circ\text{F}$ and $^\circ\text{C}$:

$$^\circ\text{C} = 0.55556 (^\circ\text{F} - 32). \quad (7.278)$$

Note that $-40^\circ\text{C} = -40^\circ\text{F}$.

In the following discussions, we will examine the details of some of the common means of temperature measurement and secondary standards for temperature sensor calibration.

7.8.1 Temperature Standards

The SI primary standard for temperature is the triple point of pure water. The *triple point* of a pure substance is defined as that temperature and pressure at which all three

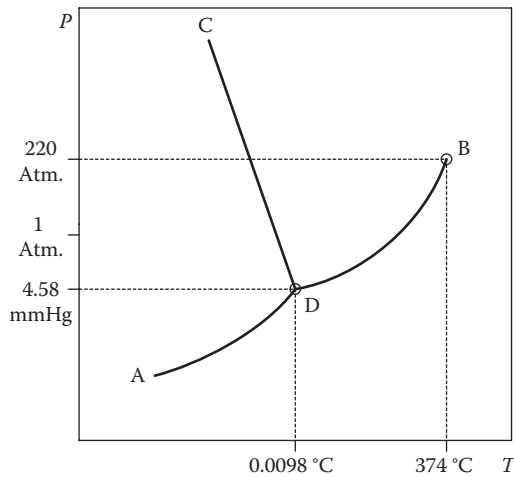


FIGURE 7.107
Phase diagram for water (not to scale), showing the triple point, *D*.

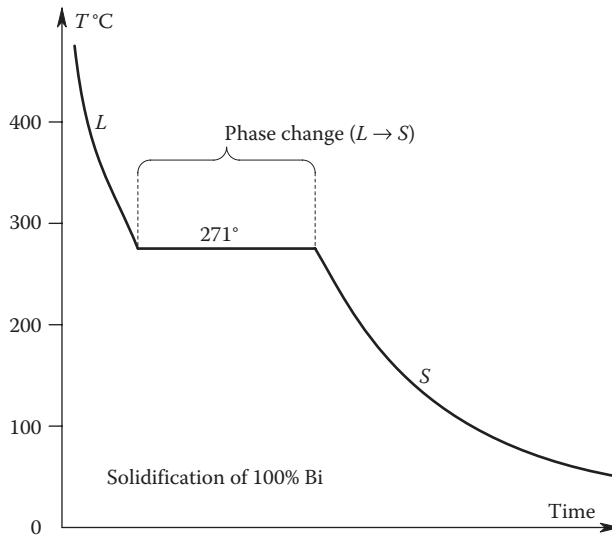
phases (solid, liquid, and vapor) are in equilibrium in a closed vessel. The triple point of pure water occurs at $+0.0098^{\circ}\text{C}$ and 4.58 mm Hg pressure. This is a single, unique point in the P,T phase diagram for H_2O . The transition (melting) temperature between solid (ice) and liquid water above 4.58 mm Hg pressure is a decreasing function of pressure; hence, the melting point of ice is slightly pressure dependent at atmospheric pressures (see Figure 7.107.) The pressure dependence is sufficiently small so that an ice bath made with double-distilled water (both ice and water) can serve as a secondary standard for 0°C .

The phase transition temperatures of other pure substances can be used as secondary temperature standards. In some cases, these transitions can be observed through a microscope that focuses on a single crystal in a capillary tube. A visible change in the index of refraction generally accompanies the phase change; often, there is also a change of color or of shape. When the liquid- to solid-phase change of a large amount of a pure calibrating substance is used, the cooling curve has a plateau at the freezing temperature. This effect is illustrated in Figure 7.108 for the metal bismuth. As a substance freezes, it releases, isothermally, at the freezing temperature, a heat of fusion that prevents the substance temperature from continuing to drop until the phase change is complete. This process is reversed in a melting curve (not shown); the rising temperature plateaus because of the isothermal absorption of the heat of fusion. Some substances that can be used for secondary temperature calibration at atmospheric pressure are listed in Table 7.6 with their melting points.

7.8.2 Some Common Means of Temperature Measurement

7.8.2.1 Mechanical Temperature Sensors

The bending, bimetallic strip is one of the more commonly encountered mechanical temperature-sensing systems. This device is used in many common household thermostats and thermometers and is illustrated in Figure 7.109. At reference temperature, T_o , the

**FIGURE 7.108**

Cooling diagram for liquid bismuth metal. The plateau occurs at the freezing point temperature of Bi due to the release of the heat of fusion.

TABLE 7.6

Melting Points of Some Pure Substances (at Atmospheric Pressure) That Can Be Used for Temperature Calibration

Substance	Melting Temperature (°C)
Oxygen	-218.4
Mercury	-38.87
Water	0
Sulfur	112.8
Bismuth	271.3
Cadmium	320.9
Silver	961.9
Gold	1064.4

strip is perfectly straight and has a length, L . At some higher (or lower) temperature, T , the strip is seen to bend in a section of an arc with radius R , as shown. Beckwith and Buck (1961) show that R is given by

$$R = \frac{t \left\{ 3(1+m)^2 + (1+mn) \left[\frac{m^2+1}{mn} \right] \right\}}{6 (\alpha_B - \alpha_A)(T - T_0)(1+m)^2} \quad (7.279)$$

where

t is the total thickness of the strip

m is the ratio of thickness of the top to the bottom strip

n is the ratio of Young's moduli of the top to the bottom strip

α_A and α_B are the coefficients of linear expansion of the top and bottom strips, respectively

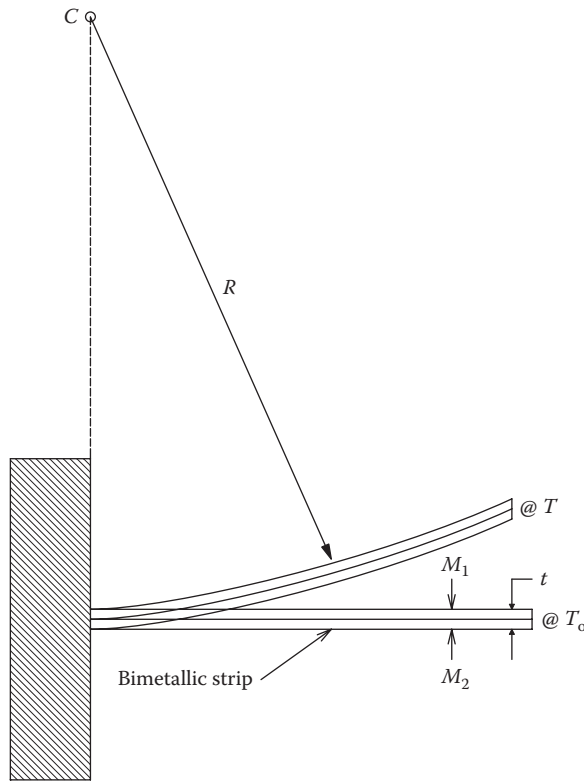


FIGURE 7.109
Bending of a bimetallic strip with increasing temperature.

We assume $\alpha_B > \alpha_A$. If the strips have the same thickness, and their Young’s moduli are the same, Equation 7.295 reduces to

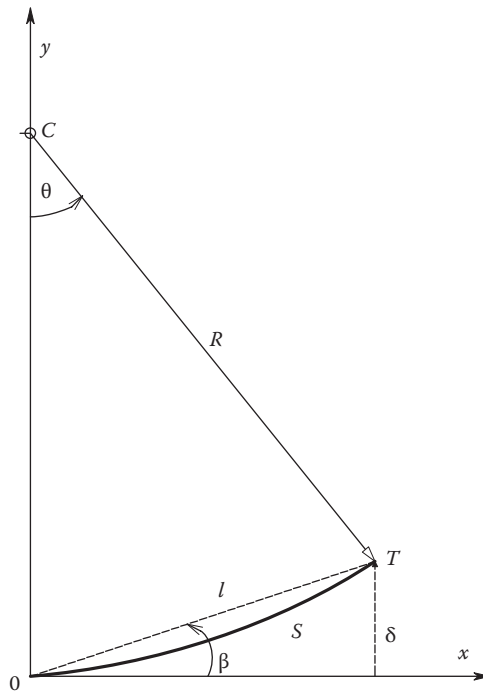
$$R = \frac{2t}{3(T - T_0)(\alpha_B - \alpha_A)}. \tag{7.280}$$

In general, R will be large compared to L . From the system geometry, and a knowledge of $R(T)$, we can predict the deflection of the tip of the bimetallic strip and the angle the tip makes with the equilibrium (horizontal) position of the strip. Refer to Figure 7.110. From simple trigonometry, it can be shown that the angle β that the tip of the bimetallic strip makes with its equilibrium position is given by

$$\beta = \frac{\theta}{2} = \frac{L}{2R}. \tag{7.281}$$

The linear deflection, δx , of the tip T from the x -axis is

$$\delta x = 2R \sin^2\left(\frac{L}{2R}\right). \tag{7.282}$$

**FIGURE 7.110**

Geometrical considerations of the tip deflection of a bimetallic strip thermometer.

And if $\beta < 15^\circ$, then, more simply,

$$\delta x \cong L\beta = \frac{L^2}{2R}. \quad (7.283)$$

Thus, we see that both the linear tip deflection and the deflection angle are linear functions of the temperature difference, $T - T_0$. In some simple mechanical thermometers, the deflection of the tip of the bimetallic strip is converted to rotary motion of a pointer by means of a linkage. In most home thermostat designs, the tip tilts a mercury-filled switch that starts the furnace.

Another commonly encountered mechanical temperature sensor is the mercury or colored alcohol-filled, sealed glass reservoir and capillary tube. In this common design, the filling liquid expands at a greater rate than the glass, so the column of liquid rises. The space above the expanding liquid in the sealed capillary tube is often filled with an inert gas or is under vacuum. Most mercury-filled, glass laboratory thermometers are designed for 76 mm immersion in the medium whose temperature is being measured. Other thermometers may require total immersion. If a mercury-filled, glass capillary thermometer calibrated for 76 mm immersion depth is used at a different immersion depth, then there will be a small error between the true and indicated temperatures. Beckwith and Buck (1961) discussed means for correction of this error.

7.8.2.2 Electrical and Electronic Temperature Sensors

There are several types of electronic temperature-sensing systems. We have already discussed RTDs in Section 6.2.1. Thermocouples and thermopiles were covered in Section 6.3.1. We will describe some of the less well-known electronic temperature-measuring systems in the following paragraphs.

The *resistance noise thermometer* is a system that uses the fact that the PDS of Johnson (thermal) noise from a resistor is proportional to the Kelvin temperature of the resistor (see Section 3.1.3.1). That is,

$$S(f) = 4kTR \text{ MSV/Hz.} \tag{7.284}$$

Figure 7.111 illustrates the schematic diagram of a resistance noise thermometer. AMP is a low-noise, high-gain voltage amplifier; BPF is a unity-gain BPF with an equivalent noise bandwidth of B Hz. TRMS is a sensitive, broadband, true RMS voltmeter. R_1 is the probe resistor at the unknown temperature, T_1 . R_2 is the variable reference resistor at known reference temperature, T_2 . With the switch in position A, the mean-squared meter voltage is

$$N_A = (4kT_1R_1 + e_{na}^2)K_V^2B \text{ MSV.} \tag{7.285}$$

With the switch in position B, the MS meter voltage is

$$N_B = (4kT_2R_2 + e_{na}^2)K_V^2B \text{ MSV.} \tag{7.286}$$

In practice, the resistor R_2 is adjusted until $N_B = N_A$. Equating N_A and N_B , we easily find that at equality,

$$T_1 = \frac{T_2R_2}{R_1}. \tag{7.287}$$

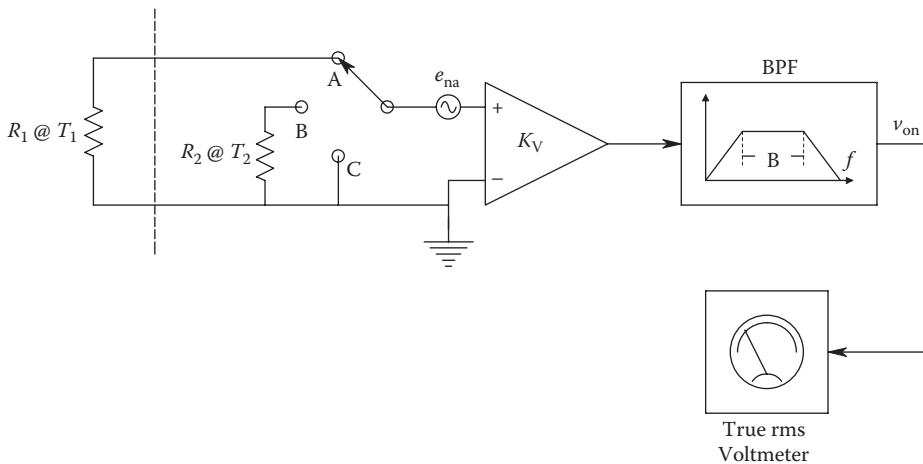


FIGURE 7.111
Schematic of a resistance noise thermometer.

The lower limit of this system's range is set by the e_{na} of the amplifier. The T_1 at which the test resistor's noise is equal to e_{na} is given by

$$T_{1L} = \frac{e_{na}^2}{4kR_1} \text{ K.} \quad (7.288)$$

If we let $e_{na} = 3 \text{ nV}/\sqrt{\text{Hz}}$, $R_1 = 3000 \text{ } \Omega$, and $k = 1.38 \times 10^{-23}$, then $T_{1L} = 54.3 \text{ K}$. The RMS output voltage of the filter when the switch is in position A, $T_1 = 54.3 \text{ K}$, and $K_v = 10^4$, is 13.4 mV, well within the range of a modern, true RMS voltmeter. Since the noise thermometer works by adjusting R_2 so that the thermal noise from R_2 at T_2 equals the noise from R_1 at T_1 , and there is uncertainty in the matching process, system accuracy is improved when the matching is repeated N times, and the mean R_2 is used in Equation 7.303. The resistance noise thermometer should be useful at temperatures as low as 0.5 K. A noise thermometer made with fine tungsten wire has been used to temperatures as high as 1700 K with an accuracy of 0.1% (Lion 1959).

Because the Johnson noise from a resistor is a linear function of resistance (at a given temperature), and because the resistance of metals increases with temperature, the noise output of a noise thermometer increases more rapidly with temperature than given by the simplified relations for N_A and N_B earlier. This nonlinearity in response makes it necessary to calibrate a noise thermometer at several temperatures when it is to be used over a wide range of temperatures. However, the use of special, low-tempco alloys for the resistor R_1 , such as manganin, eliminates the noise thermometer's nonlinearity problem in limited ranges of measurement.

The *resistance thermometer* makes direct use of the fact that a metal's resistance increases with temperature. This effect was discussed in Section 6.2.1 and will be expanded on here. In general, $R(T)$ is a nonlinear function that can be approximated around a given reference temperature, T_o , by a power series:

$$R(T) = R_o(1 + \alpha\Delta T + \beta\Delta T^2 + \gamma\Delta T^3 + \dots) \quad (7.289)$$

where $\Delta T \equiv T - T_o$ and α , β , etc., are first- and second-order tempcos. Generally, when ΔT is small, we neglect the higher-order coefficients and use only α to describe $R(T)$. In general, α is defined by

$$\alpha \equiv \frac{dR(T)/dT}{R(T)}. \quad (7.290)$$

We note that the first-order tempco, α , defined by Equation 7.306, is in fact a function of temperature. Resistance thermometers are called *RTDs*. They are generally used in a Wheatstone bridge circuit that is nulled at the reference temperature (e.g., 0°C), and the bridge unbalance voltage, V_o , is an almost linear function of ΔT . Table 7.7 lists the first-order tempcos of some of the common metals used for RTDs measured at room temperature and the useful range of temperatures over which they can be used.

Platinum is the preferred RTD metal because of several factors: it has a high melting point (1775.5°C), it does not oxidize in air at high temperatures, it is relatively chemically inert, and its $R(T)$ characteristic is quite linear from -190°C to +400°C and has a slight negative second derivative above 400°C. In other words, the high-order coefficients of the power series, Equation 7.305, for platinum are very small. Of course, pure platinum wire is expensive.

TABLE 7.7

Some Metals Used for RTDs [T2R-7.7-2]

Metal	Tempco (α)	Working Range of RTD
Platinum	+0.00392	-190°C to 800°C -264°C to 1000°C with corrections
Nickel	0.0067	-100°C to 300°C
Tungsten	0.0048	-100°C to 400°C
Copper	0.0043	-100°C to 250°C

Source: Northrop 2005.

Sources of error in using an RTD include self-heating and the effect of lead resistances. Self-heating can be avoided by keeping the power dissipation in the RTD very small and constructing the RTD so there is low thermal resistance between the wire of the RTD and the medium whose temperature is being measured. A low thermal resistance also means a rapid response time to changes of temperature in the medium. Means of compensating for lead resistances were illustrated in Figure 6.2 of this text.

In the biological temperature range, platinum RTDs are very precise; a resolution of 0.0001°C is not uncommon. Accuracy decreases with increasing temperature, however. At 450°, accuracy is several hundredths of a degree, and at around 1000°C, accuracy is around 0.1°C (Lion 1959).

Thermistors can be subdivided into negative- and positive-tempco devices. Negative-tempco (NTC) thermistors are sintered, amorphous semiconductor devices whose resistance *decreases* with increasing temperature according to the relation

$$R(T) = R_0 \exp \left[\beta \left(\frac{1}{T} - \frac{1}{T_0} \right) \right] = R_0 \exp \left[\frac{\Delta T}{TT_0} - \beta \right] \quad (7.291)$$

where T and T_0 are in K. From the tempco definition of Equation 7.306, an NTC thermistor's α is given by

$$\alpha = \frac{-\beta}{T^2}. \quad (7.292)$$

β is typically 4000 K, and for $T = 300$ K, $\alpha = -0.044$. This relatively large tempco enables the precise measurement of temperatures in the biological range with 0.0001°C resolution. Consequently, thermistor bridges are used in many applications in ecology and environmental studies, in physiological measurements, and in physical chemistry. Thermistors come in a wide variety of sizes and shapes, including 0.01" diameter, glass-covered balls. As in the case of RTDs, the power dissipation of a thermistor must be kept low enough so that self-heating does not disturb the temperature measurement.

Thermistors can also be used as fluid *flow probes*. Two, matched NTC thermistors are used: one in still fluid for temperature compensation and the other in the moving stream. Both devices are operated at average power dissipations high enough to raise their temperatures well above ambient. The moving fluid extracts more heat from the measurement thermistor, cooling it and causing its resistance to increase. This increase in resistance is a nonlinear function of fluid velocity (see Section 7.2.2.1 for a description of conventional HWAs). We often see thermistors used as frequency-determining resistors in oscillators,

the frequency of which is a function of temperature. Such oscillators have been used in biotelemetry applications. The thermal response time constant of a thermistor depends on its mass, its insulation, and the medium in which it is immersed. Thermistor time constants range from 100 ms to several minutes.

Thermocouples and thermopiles are discussed in depth in Section 6.3.1. Traditionally, a thermocouple system is used with a precision potentiometer to read its EMF, and the temperature is found from a lookup table for the particular couple materials used, such as copper and constantan. Potentiometers are expensive, slow to use, and require a standard EMF and a reference temperature source. Consequently, where possible, it is more economical and timely to use calibrated RTDs or thermistors for temperature measurements formerly done with thermocouples.

Electronic, IC temperature sensors are specialized ICs used for sensing temperatures in the -55°C to $+150^{\circ}\text{C}$ range. Physically, they can be packaged in the form of a small metal (TO-52) can, a flat pack (F-2A), or a plastic (TO-92) case. Analog devices offer the AD590 and AD592, two-terminal, temperature-to-current sensors. The AD sensors operate from a 4 to 30 V_{DC} supply and provide an output current that is a linear function of the sensor's Kelvin temperature. That is, at 0°C , the output current is a nominal $273.15\ \mu\text{A}$, and at 25°C , the output current is $298.15\ \mu\text{A}$. Over the entire operating temperature range, the AD590M has a maximum nonlinearity of $\pm 0.3^{\circ}\text{C}$, maximum repeatability and long-term drift are $\pm 0.1^{\circ}\text{C}$, and the absolute error with the sensor calibrated for zero error at 25°C is no more than $\pm 1.0^{\circ}\text{C}$. The AD temperature-to-current sensors are generally used with an op amp current-to-voltage amplifier that allows generation of a voltage output proportional to the Fahrenheit or Celsius scales. Figure 7.112 shows the internal schematic of an AD590 sensor and its current/voltage/temperature curves. An op amp conditioning circuit for an AD590 is shown in Figure 7.113.

NS (now Texas Instruments) also makes IC temperature sensors; the LM34, LM35, LM134, LM135, LM234, LM235, LM335 series. These sensors are three-terminal devices that produce an output *voltage* proportional to the temperature. For example, the AD34CA has a $10\ \text{mV}/^{\circ}\text{F}$ output voltage gain, a range from -50°F to $+300^{\circ}\text{F}$, and a nonlinearity of $\pm 0.3^{\circ}\text{F}$. The LM35 precision centigrade temperature sensor series has a $10\ \text{mV}/^{\circ}\text{C}$ output voltage gain over a -55°C to $+150^{\circ}\text{C}$ range and $\pm 0.25^{\circ}\text{C}$ nonlinearity. IC temperature sensors are ideally suited for environmental monitoring and control applications within their temperature ranges.

Optical pyrometers provide a no-touch means of estimating the surface temperatures of hot objects in the range of 775°C – 4200°C , such as metals being hot-worked, molten metals, gas plasmas, and furnace interiors. Optical pyrometers make use of the fact that all objects at temperatures above 0 K radiate heat in the form of broadband, EM energy. The range of the EM spectrum generally considered to be thermal radiation lies from 0.01 to 100 μm wavelength. Objects that are radiating heat are characterized by three parameters that describe what happens to long-wave, EM radiation (heat) at their surfaces. This relation is

$$e = a = 1 - r \quad (7.293)$$

where

e is the surface's *emissivity* that is always equal to its *absorptivity*

r is the reflectivity of the surface

An ideal blackbody radiator has $e = a = 1$, and $r = 0$ (i.e., all radiant energy striking its surface is absorbed, and none is reflected). For a nonideal radiator, a finite fraction of the

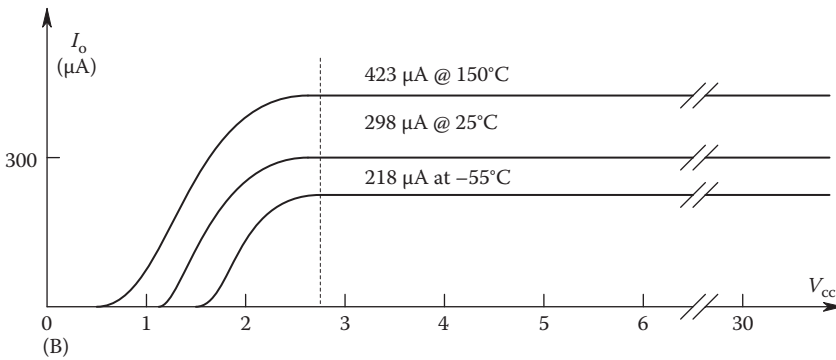
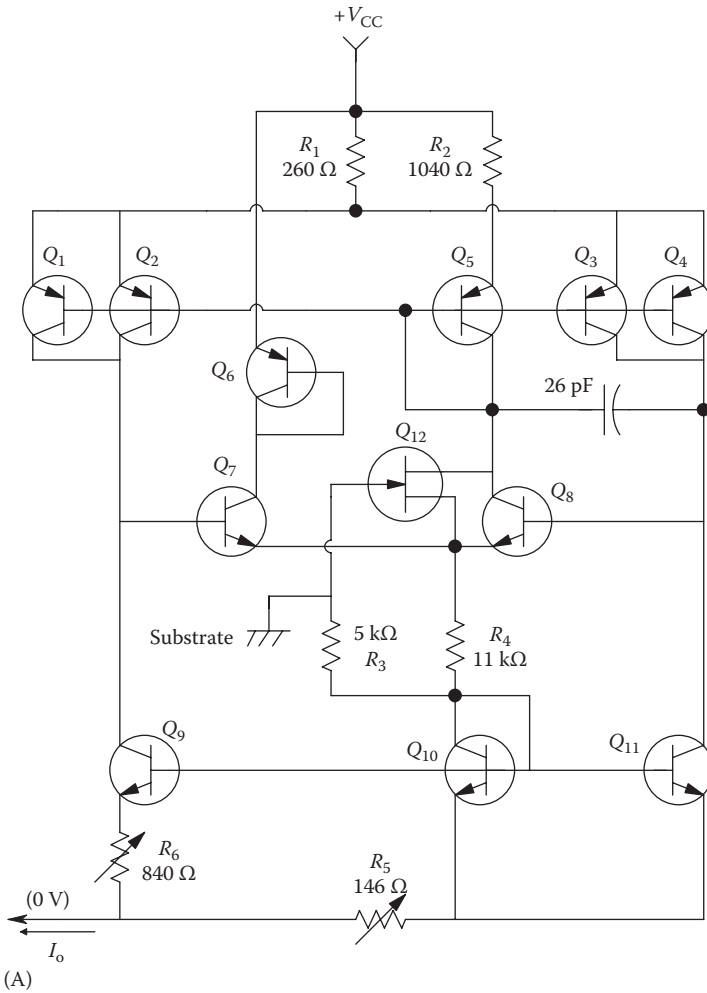
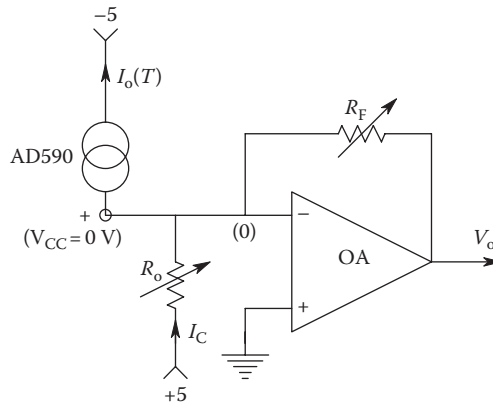


FIGURE 7.112 (A) Simplified schematic of an analog devices' AD590 electronic temperature sensor. (B) Output current vs. applied voltage and temperature for the AD590 sensor.

**FIGURE 7.113**

Simple OA circuit to condition the current output of the AD590 temperature sensor. For DC stability, the OA should be a chopper or autozeroing type. R_o sets $V_o = 0$ at 0°C , and R_F sets the output scale, so, for example, 100°C gives $V_o = 10.0$ V.

incident energy is reflected. The same properties exist for emitted radiation from a blackbody. For a hot object that is not an ideal radiator, the radiated heat output is given by

$$W = eW_{\text{bb}} \quad (7.294)$$

where

W is the total radiant emittance (in MKS units, W/m^2) from the hot surface

e is the emissivity

W_{bb} is the radiant emittance from an ideal blackbody

Max Planck, using the newly developed quantum theory in the early 1900s, developed an expression for the power spectrum of heat radiation from a blackbody (Sears 1949):

$$W_f = \frac{2\pi h f^3}{c^2 \exp(hf/kT) - 1} = \frac{dW_{\text{bb}}}{df} \quad (7.295)$$

where

W_f is in $\text{W}/(\text{m}^2 \text{ Hz})$

c is the speed of light

k is Boltzmann's constant

h is Planck's constant

W_{bb} is the radiant emittance of the blackbody in W/m^2

From an experimental and practical point of view, it is useful to write Planck's equation in terms of wavelength instead of frequency. Substituting $f = c/\lambda$, we find

$$W_\lambda = \frac{C_1 \lambda^{-5}}{\exp(C_2/\lambda T) - 1} = \frac{dW_{\text{bb}}}{d\lambda} \quad (7.296)$$

where

$$C_1 = 2\pi c^2 h = 3.740 \times 10^{20}$$

$$C_2 = hc/k = 1.4385 \times 10^7$$

W_λ is in $\text{W}/(\text{m}^2 \mu\text{m})$

Plots of Equation 7.312 are shown in Figure 7.88. There are two important observations to make about this figure: (1) As the temperature of the blackbody increases, the peak of maximum spectral emittance shifts systematically to shorter wavelengths. (2) As the temperature increases, the area under the spectral emittance curves increases. Subjectively, this means that when we look at a hot object, the color of the object's surface shifts from a dim, dull red to a bright orange to a brighter yellow, etc. The peaks of the spectral emittance curves can be found by differentiating the expression for W_λ and setting the derivative equal to zero. This gives a transcendental equation that can be solved numerically to yield the wavelength at the peak, λ_{PK} :

$$\lambda_{PK} = 2.8971 \times 10^6 / T \quad (7.297)$$

where

λ_{PK} is in millimicrons
 T in Kelvins

This relation is called *Wien's displacement law*

If we integrate the expression for spectral emittance, we obtain an expression for the total radiated power/m² from the blackbody's surface:

$$W_{bb} = \int dW_{bb} = \int W_\lambda d\lambda = \frac{p^4 C_1 T^4}{15 C_2^4} = \sigma T^4. \quad (7.298)$$

The constants have been defined earlier, and $\sigma = 5.672 \times 10^{-8}$ for SI units. Equation 7.314 is known as the *Stefan-Boltzmann equation*.

Practical hot surfaces have nonunity emissivities that are generally a function of wavelength. Thus, the practical, spectral emittance curve for a hot object may have many peaks and valleys. If these irregularities are averaged out, we can often fit a scaled-down, blackbody spectral emittance curve to the practical curve so that their peaks occur at the same temperature, T . Such a scaled blackbody W_λ is called a *gray body curve*.

In one form of *optical pyrometer*, shown in Figure 7.114, a human operator makes a subjective color comparison of a glowing tungsten filament with the hot surface under measurement. The color comparison is made easier by optically superimposing the image of the filament on that of the hot object. When the filament's color, determined by its current, matches that of the object, it disappears on the background of the object, and the filament current is read. Because the filament may be considered to be a blackbody, its spectral emittance closely follows that of the object when its temperature is the same as the object's. Often, a red filter is used to convert the color-matching task to one that involves brightness matching. The filament ammeter must be calibrated in temperature.

In another form of optical pyrometer, shown in Figure 7.115, the filament is run at a constant current and brightness. The intensity of the image of the hot object is then varied with an NDW. Again, a red filter is used to make the null process an exercise in monochromatic intensity matching. The wedge position is calibrated in object temperature, assuming blackbody emission.

Total radiation pyrometers are also known as radiometers and are electronic instruments that measure the integral of the W_λ curve. That is, their output is proportional to the integral of W_{bb} over the area of the instrument's aperture. In one form of radiometer,

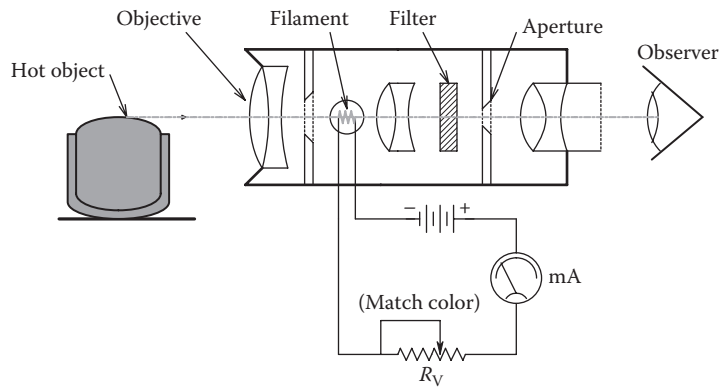


FIGURE 7.114

Diagram of a simple optical pyrometer. The operator adjusts R_V until the color temperature of the lamp filament matches that of the object. The human operator must have reliable color vision in the yellow–red end of the spectrum.

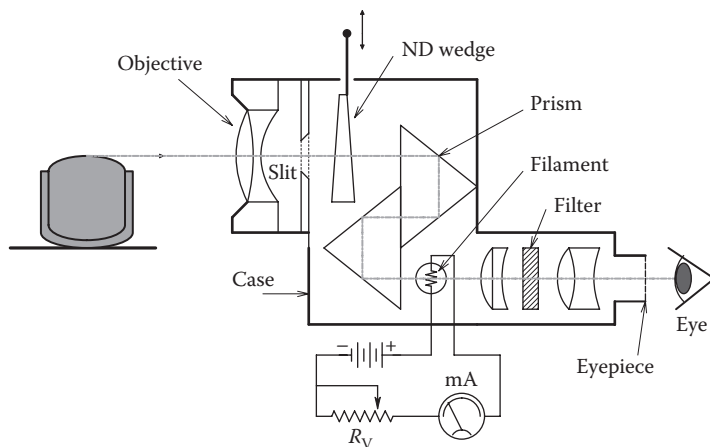


FIGURE 7.115

In this pyrometer design, the operator matches monochromatic intensities of the filament and object. The NDW position is calibrated in object temperature.

the detector is a *thermopile* in which half of the thermocouple junctions must be kept at a reference temperature. The thermopile is designed to be a nearly 100% blackbody absorber. Thus, regardless of the shape of $W(\lambda)$, all of the incoming energy is captured and converted to a temperature rise of the sensing junctions. Total radiation pyrometers can be used with blackbody radiation as well as coherent sources such as lasers.

Other detectors used in radiation pyrometers may include PCs, photodiodes, and *pyroelectric detectors*. Pyroelectric detectors absorb thermal energy and generate electrical signals. One example of a pyroelectric detector material is the polarized polymer film, *polyvinylidene difluoride* (PVDF). PVDF film absorbs strongly in the IR, and a free mounted, 28 μm PVDF film has a response time constant of about 5 s to a step temperature change. All metallic radiation detectors have spectral response characteristics. The broadest

response characteristic is that of the blackbody thermopile; this sensor responds to radiation from UV to far IR (0.25–20 μm). PVDF pyroelectric sensors respond primarily from 6.6 to 66 μm . LiTaO_3 crystals are also used in faster-responding pyroelectric sensors. Most photodiodes and PCs respond maximally in the near IR and are relatively narrowband devices compared with PVDF and thermopiles. Figure 7.116 illustrates the *detectivity* (D^*) of some common photoconductive IR detectors, pyroelectric detectors, thermopiles, and thermistor bolometers. Note that several of the photoconductive sensors are operated at LN_2 temperature (e.g., HgCdTe , GeCu at 77 K) to reduce noise. The detectivity, D^* , is defined as $D^* \equiv \sqrt{(\text{sensor area})/\text{NEP}}$. Its units are $(\text{cm}\sqrt{\text{Hz}})/\text{W}$, and at a given wavelength, a higher D^* is better. NEP is the sensor's NEP. The NEP is the sensor's optical input power in watts required to give an output SNR of unity. The sensor's total noise can be from several sources and is assumed to have a white PDS. NEP is also a function of wavelength, and a low NEP is better.

The optics of pyrometers also present a problem, as conventional glasses do not transmit effectively at wavelengths beyond 2.6 μm . Thus, special materials must be used in IR radiometers for windows, mirrors, and lenses. A broadband, IR radiometer/pyrometer uses front surface mirrors (FSMs) coated with aluminum or gold to focus the radiation on the sensor. A chopper wheel is used to 0,1 modulate the radiation and permit AC amplification

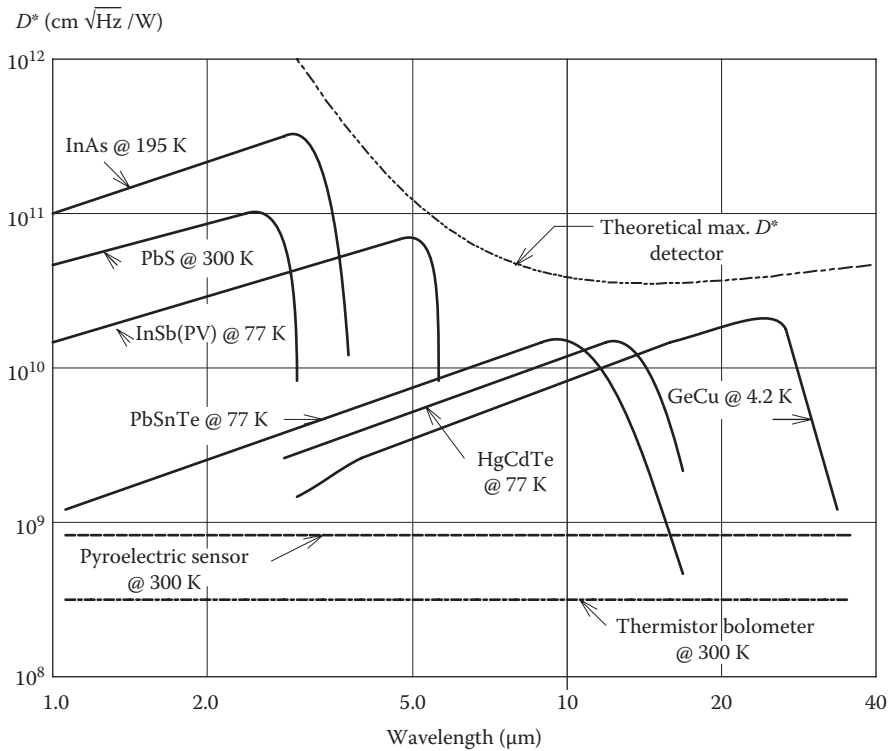


FIGURE 7.116

Spectral sensitivities of some common IR sensor materials. Generally, sensitivity can be improved by chilling the sensor with liquid nitrogen. Note: solid curves = PCs. Dashed line = pyroelectric sensors. Dash-dot line = thermistor bolometers. Vertical axis = D^* , the specific detectivity. (Adapted from Barnes Engineering, *Handbook of Infrared Radiation Measurements*, Barnes Engineering Co., Stamford, CT, 1983.)

of the sensor output signal out of the $1/f$ noise band. Phase-sensitive rectification is then used to recover the amplified signal. A typical IR radiometer system is shown in Figure 7.117. When lenses and windows are used in radiometer design, they must be able to transmit in the IR. Figure 7.118 summarizes the IR transmission characteristics of commonly used IR optic materials. Note that pure germanium is now widely used for IR optics (Barnes Engineering 1983). The passband of germanium is from about 2 to 30 μm . Pure germanium can be considered to be a dielectric and has a high refractive index of $n = 4$. The reflection coefficient of germanium, r , is given by the Fresnel reflection equation:

$$r = \frac{(n-1)^2}{(n+1)^2} = 0.36. \quad (7.299)$$

Thus, assuming zero absorption for a Ge window, only $(1 - 0.36)^2 = 0.41$ of the incident IR would be transmitted (Barnes Engineering 1983).

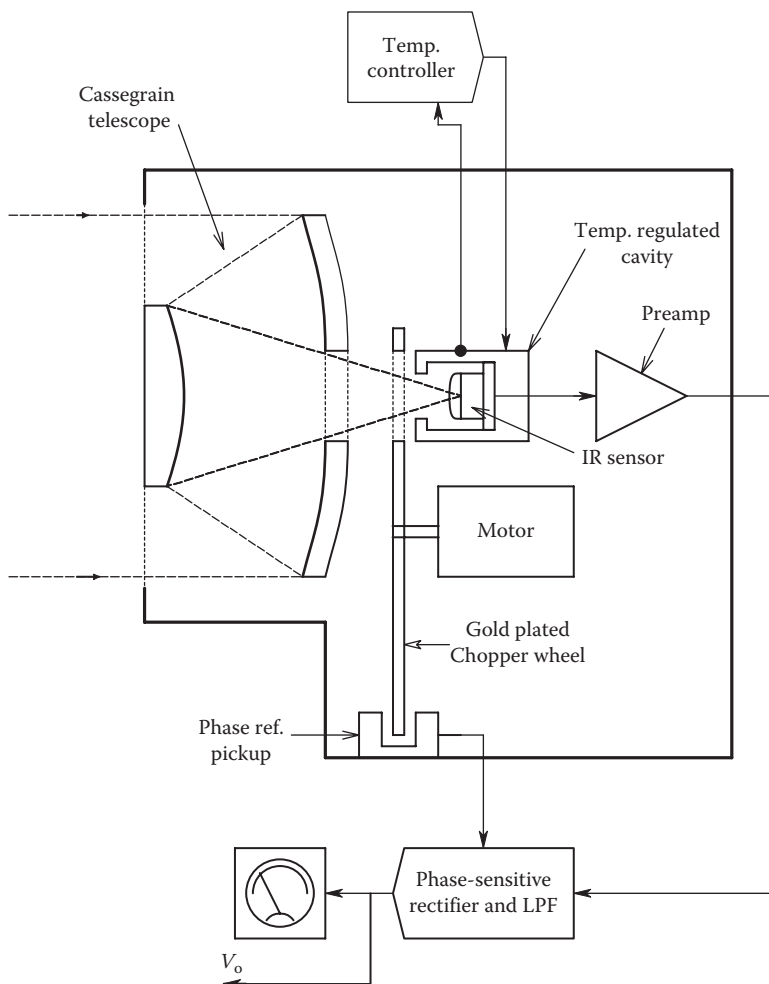


FIGURE 7.117

Diagram of a basic IR radiometer. Note that chopping is necessary to avoid $1/f$ noise from the sensor and preamp. (Courtesy of Barnes Engineering Co., Shelton, CT.)

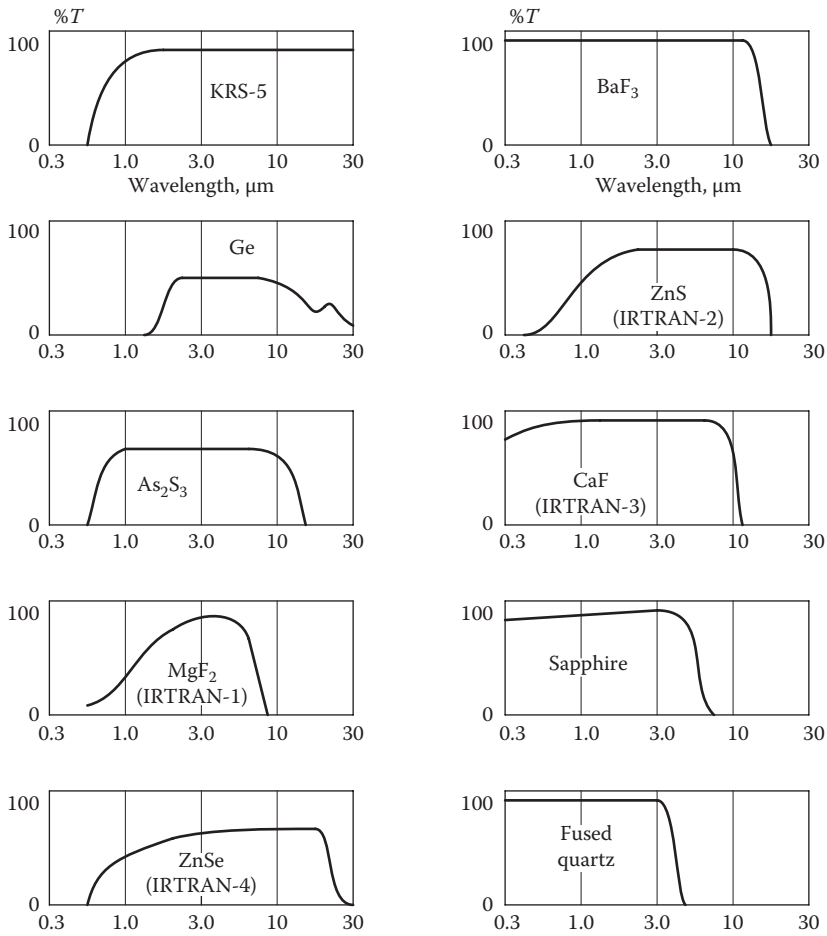


FIGURE 7.118

IR transmission characteristics of some materials used in IR optical systems. Note that germanium metal transmits into the far IR. (Courtesy of Barnes Engineering Co., Shelton, CT.)

The optics of some radiometers make use of a chopping scheme in which the IR sensor is alternately exposed to the radiation under measurement, dark, and a standard blackbody source of known temperature. Thus, calibration and correction for dark response can be done simultaneously by a computer attached to the radiometer.

7.9 Chapter Summary

In this large chapter, we have presented a survey of some of the common (and not so common) means of measuring certain physical quantities, including angular acceleration, velocity, and position; linear acceleration, velocity, and position; force; torque; pressure; vacuum; and temperature.

Mechanical rate gyroscopes (including classical rotating mass, vibrating mass, air jet, and micromachined oscillating disk) were described and analyzed using the Lagrange method. Sagnac FO rate gyros were presented as systems for sensing the inertial angular velocity and position of vehicles. Clinometers using pendulums and liquid filling were also described and their dynamics analyzed.

Various kinds of linear and rotational accelerometers were described, including servonulled, Newtonian moving-mass systems and piezoelectric and micromachined units.

The Doppler effect was introduced as the basis for many instruments using sound or EM radiation to measure the linear velocity of vehicles or particles suspended in a moving fluid. LDV was described.

A new section on the global positioning system (Section 7.3.3.1) and its upgrades was presented as a means for navigation on the Earth's surface and in its atmosphere.

Optical interferometry was introduced and analyzed as a means to measure small displacements, and closed-loop, optical, and ultrasonic phase-lock velocity and ranging and systems developed by the author were described as a means of continuously measuring vehicle motion.

Sensors used to measure force and pressure were seen to be related in that a pressure acting on the area of a diaphragm or piston produces a force. Forces can produce changes in resistance, capacitance, inductance, charge displacement (in piezoelectric transducers), as well as the optical transmission properties of FO waveguides. These physical changes can be sensed as functions of force or pressure. Some torque sensors also make use of sensed mechanical strain caused by the torque.

The principles of spectrometry were introduced in Section 7.6. Dispersive, nondispersive, and FT spectrometers were described as means to quantify chemicals in gas, liquid, or solid states. The usefulness in chemical analysis of single- and multiple-bubble sonoluminescence and SPR was also reviewed.

We saw that many physical processes respond to changes in temperature. These include chemical reactions, conduction in semiconductors, resistance of conductors, dielectric constants, the speed of sound, and refractive index of OFs. In fact, many other physical measurements require that we compensate for temperature changes. Optical pyrometry was introduced as a means of remote sensing the temperature of objects by their blackbody IR emissions.

The basis for many measurement systems and sensors has been found to lie in several, basic physical laws. These laws include the Doppler effect, the Sagnac effect, the piezoelectric effect, the Hall effect, and the piezoresistive effect (strain gauges), to name a few.

Problems

- 7.1 A conventional pendulum clinometer is subject to error in its static tilt angle reading when the case is subject to a linear acceleration component, $\ddot{\mathbf{x}}$, in the $+x$ -direction, perpendicular to the gravity vector, as shown in Figure P7.1. The \mathbf{o} symbol shows the center of mass of the pendulum. In the figure, a $\theta_c = +17^\circ$ is shown. If there were 0 acceleration of the case in the x -direction, the pot wiper (and pendulum)

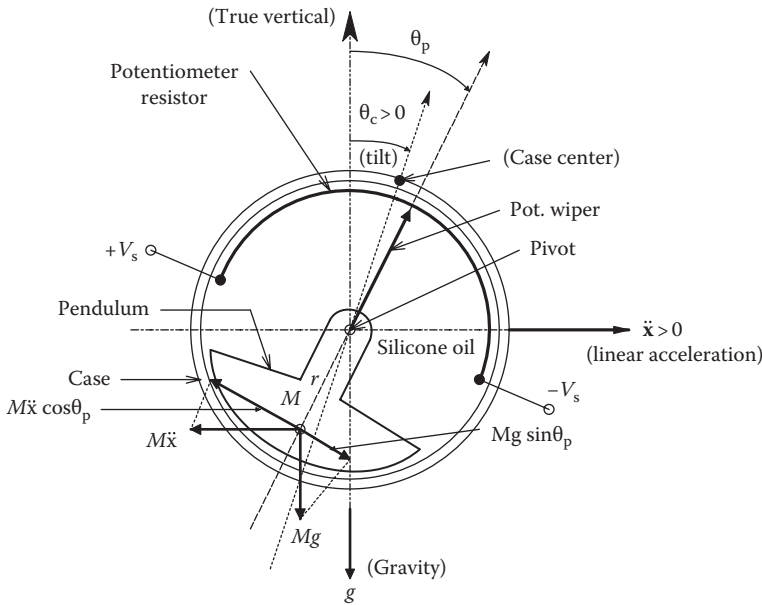


FIGURE P7.1

would be vertical. Because of the acceleration to the right, the pendulum angle is $\theta_p = +26^\circ$, producing a serious error in the tilt reading:

- A. Derive a transfer function for V_o as a function of the linear acceleration, \ddot{x} . Assume $\theta_c = 0$. Assume $|\theta_p| < 15^\circ$, so $\sin(\theta_p) \cong \theta_p$ in radians, and $\cos(\theta_p) \cong 1$. Also, $V_o = K_o(\theta_c - \theta_p)$. The pendulum is subject to a viscous friction torque from the silicone oil, $\tau_f = D(\dot{\theta}_p - \dot{\theta}_c)$ newton meters. τ_f acts to retard the clockwise swing of the pendulum, as shown in the figure.
 - B. Assume a constant linear acceleration, \ddot{x} , as shown. In the SS, how much would the case have to be rotated, and in what direction, so $V_o = 0$?
- 7.2 An active, EM, differential pressure sensor is shown in Figure P7.2. Assume that the net force acting on the diaphragm is given by

$$0 = P_o A - P_i A - F_M - M \ddot{\Delta x}.$$

Forces with a minus sign act upward. Diaphragm deflection is measured with an LVDT whose output is given by $V_x = K_L \Delta x$. The diaphragm stiffness is neglected. The upward EM force from the coil is $F_M = K_M I_C$. From the figure, $I_C = G_M [K_d V_x + K_p V_x]$, and $\Delta P = P_o - P_i$:

- A. Derive an expression for the transfer function, $\frac{V_x}{\Delta P}(s)$, in time-constant form.
 - B. Find expressions for the sensor's SS compliance, $\Delta x / \Delta P$, and the sensor's ω_n .
- 7.3 The circuit of Figure P7.3 is used to generate a DSBSCM square wave at $v_o(t)$ modulated by the ΔC of the differential capacitor sensor. The square wave period T satisfies $RC_o \ll T \ll R_F C_F$. For example, let $RC_o = 10^{-5}$ sec, $T = 0.001$ s, and $R_F C_F = 0.1$ s:
- A. Plot and dimension the SS currents $i_1(t)$ and $i_2(t)$ flowing into the op amp's summing junction.

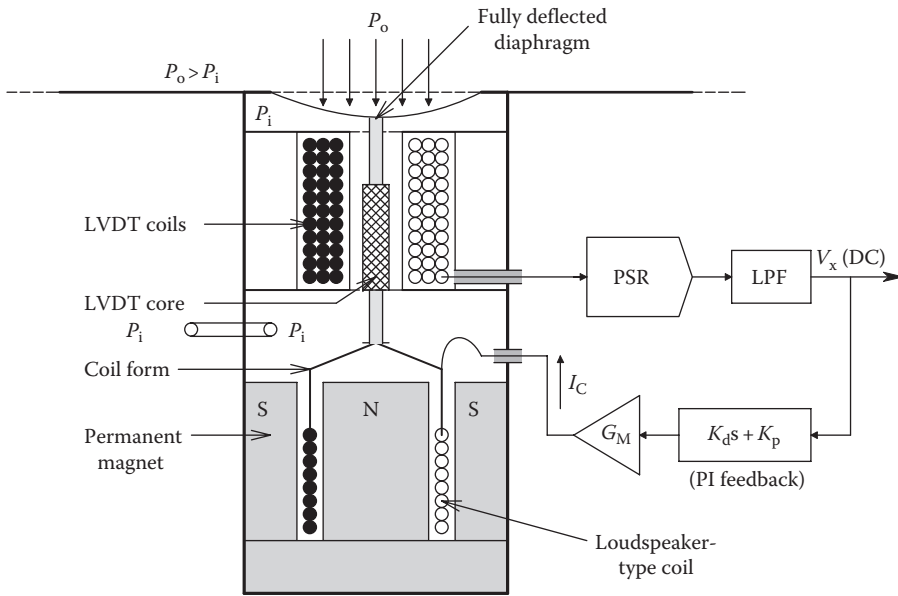


FIGURE P7.2

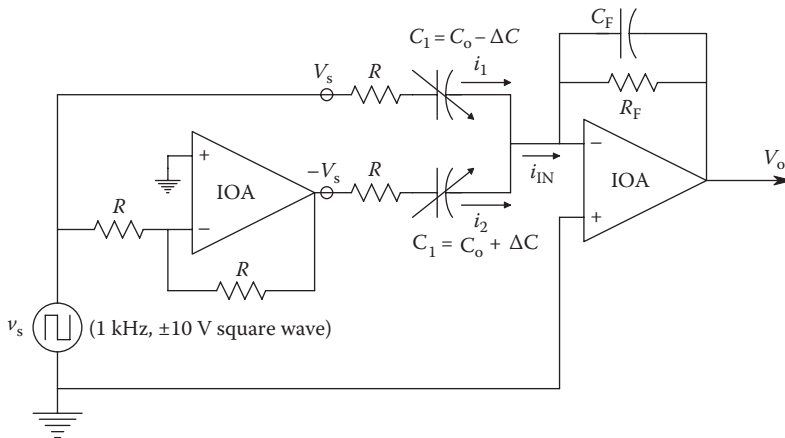


FIGURE P7.3

Give an expression for the area under $i_1(t)$, A_1 , and under $i_2(t)$, A_2 , over a half cycle ($0 \leq t \leq T/2$). Note that both currents effectively $\rightarrow 0$ before $T/2$, so ∞ can be used for the upper limit of integration in finding the areas to simplify the math. Note the summing junction is at virtual ground.

- B. Because the duration of the net input current transient is very short compared to $\tau_F = R_F C_F$, we can consider the inputs to the op amp's SJ to be impulse functions with areas of $+(A_1 - A_2)$ and $-(A_1 - A_2)$ occurring at the positive and negative switching times of the input square wave, respectively. That is, at $t = 0$, $i_{in}(t) = (A_1 - A_2)\delta(t)$, etc. Find, plot, and dimension $v_o(t)$ over two square wave cycles for $\Delta C > 0$. Note $e^{-x} \cong 1 - x$ for $x \ll 1$.

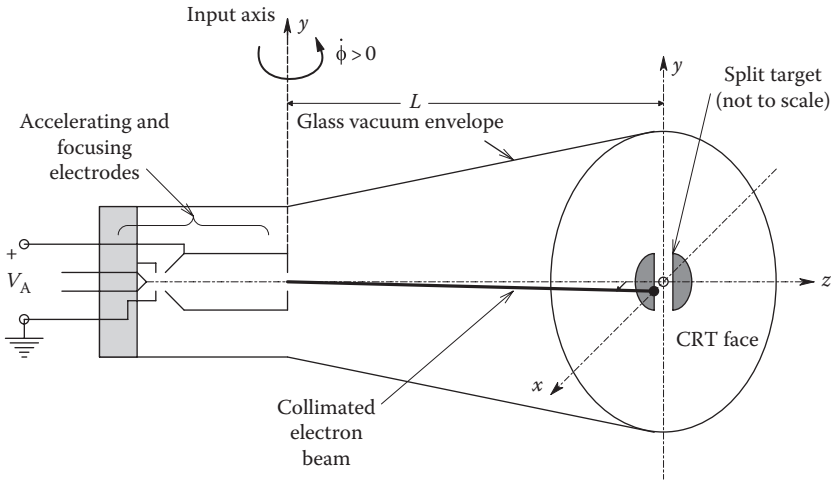


FIGURE P7.4

- 7.4 In Section 7.2.2.4, we examined the operation of the novel *Humphrey air jet rate gyro*. In this problem, we will examine the feasibility of using a modified CRT as a rate gyro. The system is shown schematically in Figure P7.4. Electrons are accelerated through a potential V_A , and then a focused beam of these moving electrons leaves the accelerating/focusing plates with a velocity v_x traveling to the CRT face in vacuo. On the inside of the face is a circular tungsten target, 1 cm in diameter. The target is split into two halves by a narrow, 1 μm gap. With the CRT stationary ($\dot{\phi} = 0$), the beam is focused on the stationary CRT face so one-half its current is collected by each half of the target. Assume the DC accelerating voltage is $V_A = 100 \text{ V}$, the free flight distance is $L = 31 \text{ cm}$, and the electron beam has a circular cross section and is 1 mm in diameter at the target. Assume further that a 1 μm lateral deflection of the beam at the target is just detectable electronically:
- Find the threshold rotation rate, $\dot{\phi}_t$, that will produce this detectable beam deflection.
 - Now, assume Hg^{++} ions form the beam, other conditions being the same as in part (A). Find the threshold rotation rate for a mercury ion beam.
 - Design an analog circuit to convert the differential electron currents to or from the target electrodes to an output voltage proportional to $\dot{\phi}$.
- 7.5 The circuit of Figure P7.5 uses an AD532 temperature-to-current sensor instead of an ice bath as a temperature reference for an iron–constantan thermocouple. For $V_{SS} > 4 \text{ V}$, the DC current sourced by the AD532 IC is given by $I(T) = (T + 273)10^{-6}$ amps. T is in $^{\circ}\text{C}$. Assume the op amp is ideal. $R_F = 9 \text{ k}\Omega$, $R_1 = 1 \text{ k}\Omega$, $S_{\text{Fe}} = 18.5 \mu\text{V}/^{\circ}\text{C}$, $S_{\text{Con}} = -35.0 \mu\text{V}/^{\circ}\text{C}$, $S_{\text{Cu}} = 6.5 \mu\text{V}/^{\circ}\text{C}$:
- Derive a general algebraic expression for $V_o = f(T_m, T_A, \text{Ckt. parameters})$. Use superposition.
 - Find numerical values for R_C and V_2 that will make $V_o = KT_m$ exactly. Find the numerical value of K .

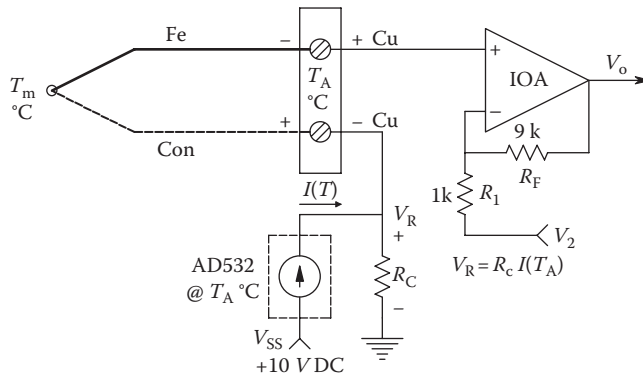


FIGURE P7.5

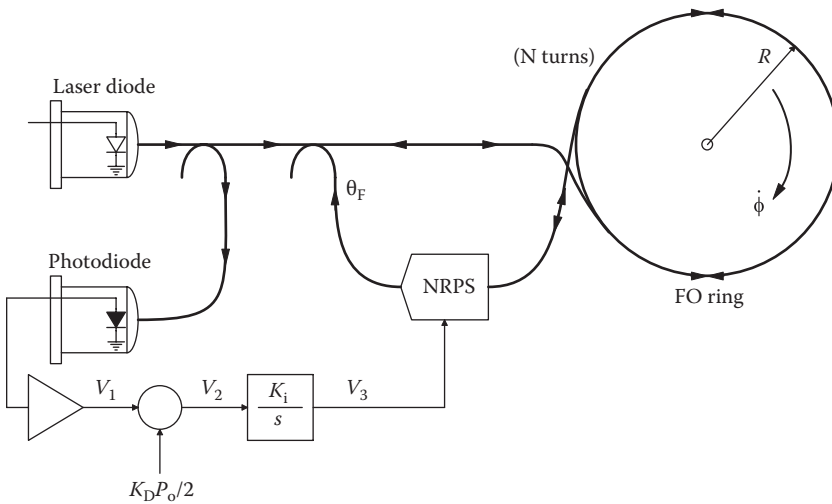


FIGURE P7.6

- 7.6 A Sagnac-effect FOG, shown in Figure P7.6, uses a nonreciprocal optical phase shift (NRPS) device to null the interference fringe shift pattern that is caused by the angular velocity of the FO ring. The system relations are $\theta_s = K_s \dot{\phi}$, $\theta_e = \theta_s - \theta_f$; optical power at the detector is $P_D = (P_o/2)[1 + \sin(\theta_e)]$ W; $V_1 = K_D P_D$; NRPS gain is $\theta_f = K_C V_3$. Under normal operating conditions, $|\theta_e| < 3^\circ$, so $\sin(\theta_e) \cong \theta_e$ radians:
- Draw a linearized block diagram of the system.
 - Find an expression for $\frac{V_o}{\dot{\phi}}(s)$ in time constant form. What advantage does this system have over a simple FOG?

7.7 Instead of using a VPC, a constant phase LADAR system is configured using a VFC in which $f_o = K_v V_o$, as shown in Figure P7.7. The distance to the stationary, reflecting target is L_o meters. In the SS, $V_v = 0$ and $V_\phi = V_{set}$:

- A. Derive an expression for V_o in the SS.
- B. Now, assume the target is receding at constant velocity $v = \dot{L} > 0$. Assume SS conditions so that $\dot{V}_v = 0$. Use the integrator transfer function to derive an expression for the SS V_v . Note that \dot{V}_o can be found by assuming $L(t)$ and differentiating the V_o expression of part (A). Also note that V_v and V_o are both nonlinear functions of L .

7.8 A double-gimbal mechanical gyro is spun up and initially set so $\theta = 0^\circ$ and $\phi = 0^\circ$. The linearized gyro torque equations are

$$M_\theta = J_1 s^2 \Theta(s) - H_r s \Phi(s)$$

$$M_\phi = H_r s \Theta(s) + J_1 s^2 \Phi(s).$$

Assume $M_\theta = 0$ (unrestrained inner gimbal), and let $M_\phi(t) = M_{\phi_0} \delta(t)$ (impulse of torque) applied around the outer gimbal. See Figure P7.8:

- A. Find expressions for $\Theta(s)$ and $\theta(t)$. Sketch and dimension $\theta(t)$.
- B. Repeat (A) for $\Phi(s)$ and $\phi(t)$.

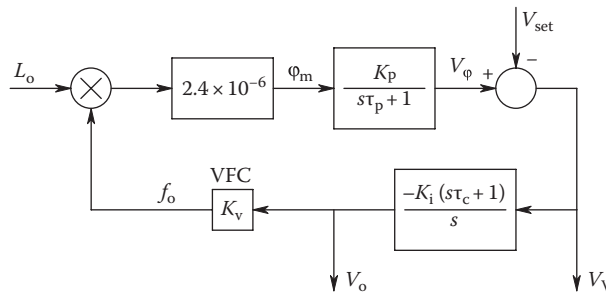


FIGURE P7.7

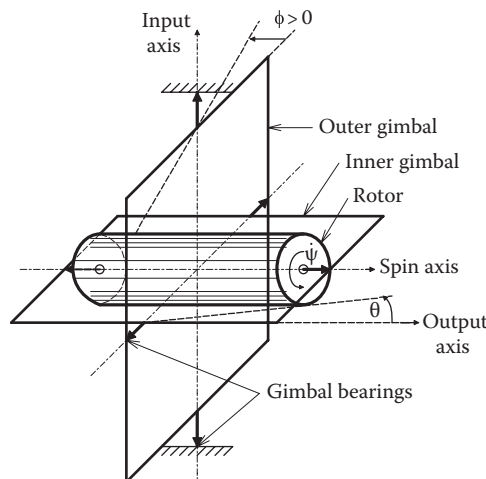


FIGURE P7.8

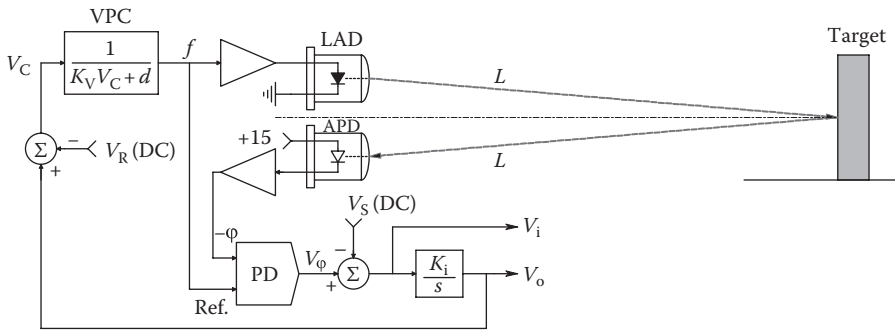


FIGURE P7.9

- 7.9 The system shown in Figure P7.9 is a constant-phase, closed-loop, laser ranging (LAVERA) system. Short, periodic pulses of frequency f are emitted from a collimated, 910 nm LAD source. The light reflects off a target at distance L ; the reflected light propagates back to the receiving optics that focus it on an APD sensor. In the SS, $V_i = 0$:
- Derive a simple expression for the SS phase, φ , between the transmitted and received pulses. Neglect the delays associated with the logic and analog signal processing. In the system, a digital PD converts φ to a voltage: $V_\varphi = K_D \varphi$. V_C is the input to a VPC oscillator in which the period is given by $T = K_V V_C + d$. The oscillator output is narrow pulses that drive the LAD to flash at a rate $f = 1/T$. Assume that in the SS, $dL/dt = V_i = 0$:
 - Derive expressions for φ_{ss} , f_{ss} , and V_{oss} .
 - Now, assume the target is moving at a constant velocity, $\dot{L} > 0$, away from the system. Derive expressions for $V_i = f(\dot{L}, \text{other parameters})$ and $f = g(\dot{L}, L, \text{other parameters})$.

8

Basic Electrical Measurements

8.1 Introduction

Electrical measurements are defined in this chapter as being measurements of the traditional electrical parameters of voltage, current, electric field strength, magnetic fields, resistance, capacitance, inductance, plus the SS AC parameters of impedance and admittance, power, frequency, and phase.

8.2 DC Voltage Measurements

DC voltage measurements can be made over an enormous range: from nanovolts to thousands of kilovolts. The practical limit to low voltage measurement is noise, and the practical limits to high-voltage measurements involve circuit loading and insulation. Of course, the specialized voltmeters that allow measurement of nanovolt potentials are generally unsuitable for high-voltage measurements and vice versa.

All voltmeters, regardless of their specialization, can be represented by an equivalent circuit consisting of a parallel impedance by which the voltmeter loads the circuit under test (CUT) and an *ideal*, infinite impedance voltmeter. Thus, we see that every voltmeter takes some power from the CUT. An ideal voltmeter would take zero power and not load the CUT. The CUT itself, by Thévenin's theorem, can be represented by an open-circuit voltage and an equivalent series impedance. Obviously, connecting a practical voltmeter to the Thevenin equivalent circuit will result in current flowing in the circuit and the (indicated) voltage at the voltmeter's terminals being less than the open-circuit voltage by the amount of the voltage drop across the Thévenin resistor, R_{TH} . This situation is illustrated in Figure 8.1. In mathematical terms, the voltmeter reads

$$V_M = \frac{V_{OC}R_M}{R_M + R_{TH}}. \quad (8.1)$$

Under normal circumstances, $R_M \gg R_{TH}$.

Voltmeter loading of the CUT is related to a parameter called the *voltmeter sensitivity*, η . η is defined as the full-scale voltage of the meter divided by the power dissipated in the meter at full-scale voltage, V_{FS} :

$$\eta \equiv \frac{V_{FS}}{P_{FS}} = \frac{V_{FS}}{V_{FS}^2/R_M} = \frac{R_M}{V_{FS}} = \frac{1}{I_{MFS}} \quad \Omega/V \quad (8.2)$$

where I_{MFS} is the current drawn by the voltmeter when at its full-scale voltage.

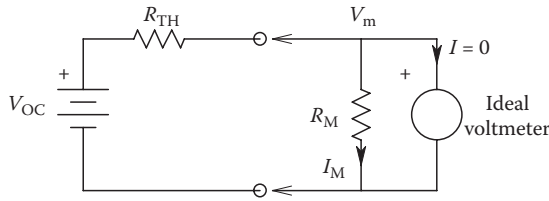


FIGURE 8.1

Simple circuit showing a nonideal voltmeter attached to a Thevenin equivalent source. R_M is the voltmeter’s internal resistance. ($R_M = \text{voltmeter sensitivity } \eta \text{ in ohms per volt} \times \text{meter full-scale voltage.}$) V_m is the voltage at the voltmeter’s terminals.

Typical analog DC voltmeters, such as found in nonelectronic multimeters, have η s of 20,000 Ω/V , which means that the voltmeter draws 50 μA from the CUT at full-scale voltage. The equivalent resistance of any DC voltmeter can always be found by multiplying V_{FS} by η . Most electronic voltmeters have a fixed input resistance, regardless of scale. This resistance is typically 10 or 11 $\text{M}\Omega$, but some specialized electronic voltmeters have R_M s as large as $10^{11} \Omega$ (e.g., electrometer voltmeters).

Figure 8.2 illustrates the role of Johnson (thermal) noise from R_{TH} in limiting the resolution of DC voltage measurements. Recall from Section 3.2.3.1 that the Johnson noise power density spectrum from a resistor is given by $S(f) = 4kTR \text{ MS V/Hz}$. The effective Hz bandwidth, B , of a voltmeter may be considered to be the reciprocal of the time required to take a reading of an applied step of voltage. This response time generally ranges from 0.1 to 10 s; hence, the effective bandwidth ranges from 10 to 0.1 Hz. If we assume that the DC open-circuit voltage, V_{OC} , must be greater than the Johnson noise from R_{TH} , $R_M \gg R_{TH}$, and R_{TH} is the sole source of noise in the circuit, then we can write

$$V_{OC} \geq \sqrt{4kTR_{TH}B} \text{ rms V.} \tag{8.3}$$

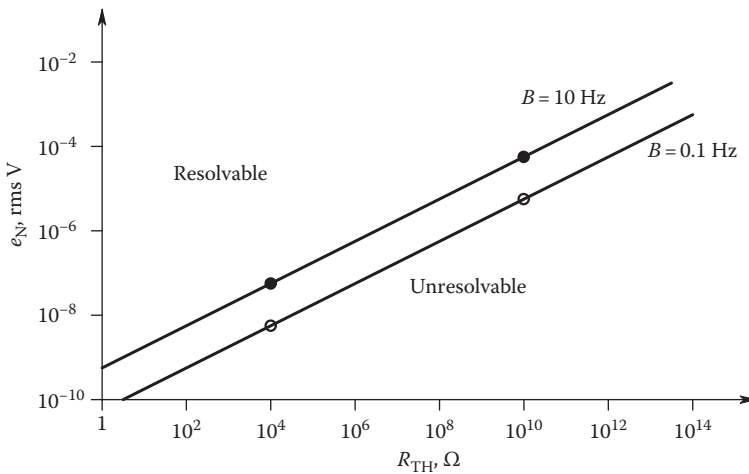


FIGURE 8.2

Graph illustrating the role of thermal noise from the Thevenin source resistor, R_{TH} , in limiting the resolution of low-level DC voltage measurements. 10 and 0.1 Hz noise bandwidths are considered.

This function is plotted for $B = 0.1$ and 10 Hz in Figure 8.2. Note that DC voltages below the lines are generally unresolvable because of noise.

8.2.1 Types of Electromechanical DC Voltmeters

There are many types of DC voltmeters, some in fact are effective as powerline frequency AC voltmeters as well. In this section, we examine the designs of some DC voltmeters, both analog and digital. Included are *D'Arsonval meters*, *dynamometer meters*, *capacitor voltmeters*, *electrometer voltmeters*, *chopper-type nanovoltmeters*, and *thermocouple voltmeters*.

D'Arsonval DC voltmeters, or permanent-magnet moving-coil voltmeters, are the most common design of electromechanical DC voltmeter. A typical *D'Arsonval* microammeter meter movement, illustrated in Figure 8.3, is basically a current-to-angular position transducer. Current flowing through the turns of the rectangular coil suspended in the uniform magnetic field in the air gap generates an orthogonal component of force on each

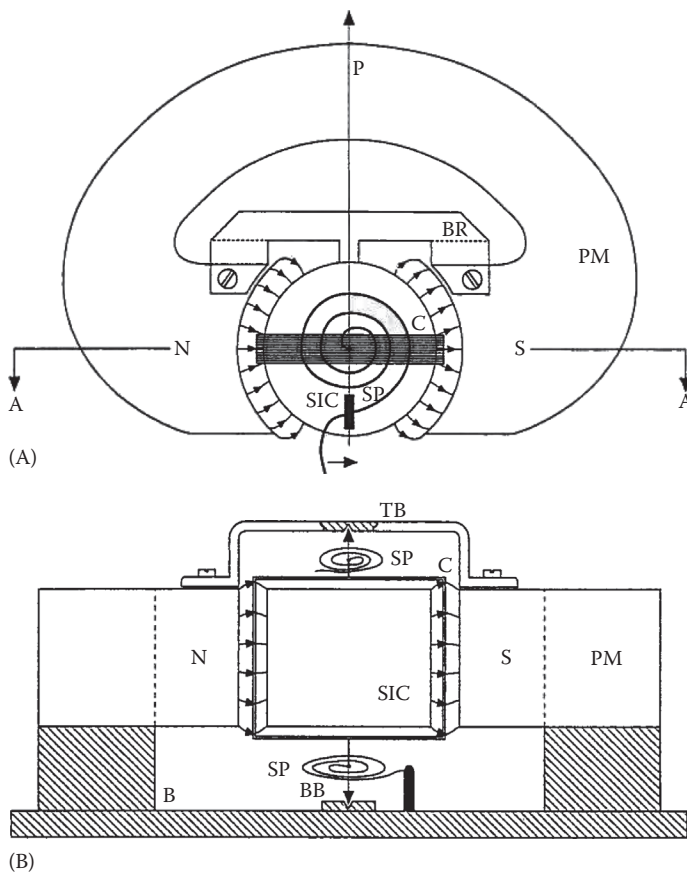


FIGURE 8.3 (A) Face view (without scale) of the innards of a *D'Arsonval* meter movement. *Note:* P = pointer. PM = permanent magnet. C = pivoted rectangular coil. SP = helical spring. SIC = soft iron core to concentrate magnetic flux. (B) Vertical section (A-A) through *D'Arsonval* meter. Shaded material is bakelite (insulator). *Note:* TB and BB = top and bottom coil bearings. B = base.

conductor on each side (but not the top or bottom) of the coil. The magnetic field is generated by a strong permanent magnet. The resultant torque acts against a helical spring, or in some meters a flat torsion spring, used to suspend the coil. When the magnetically produced torque equals the spring torque in the SS, the meter pointer stops moving.

The force per unit length of a current-carrying conductor in a uniform magnetic field is given by the vector cross-product equation:

$$d\mathbf{F} = I(\mathbf{dl} \times \mathbf{B}) \text{ N/m.} \tag{8.4}$$

If the vector $I \mathbf{dl}$ is rotated into the vector \mathbf{B} , the direction of $d\mathbf{F}$ is given by the right-hand screw rule. Accordingly, the magnitude of $d\mathbf{F}$ can be written as

$$d\mathbf{F} = IB \sin(\varphi) \mathbf{dl}. \tag{8.5}$$

Now from Figure 8.4, we see that the length of each side of the coil in the perpendicular magnetic field is L . The angle φ between the current and the \mathbf{B} field is 90° . Also, the radial distance from each side to the pivot is $W/2$, and there are N turns of wire in the pivoted coil. Thus, the mechanical torque generated by passing current through the N turns of the coil in the magnetic field is given by

$$T_M = 2BLNIW/2 = BNAI = K_T I \tag{8.6}$$

where

A is the area of the rectangular coil ($A = LW$)

$K_T = BNA$ is the meter's torque constant

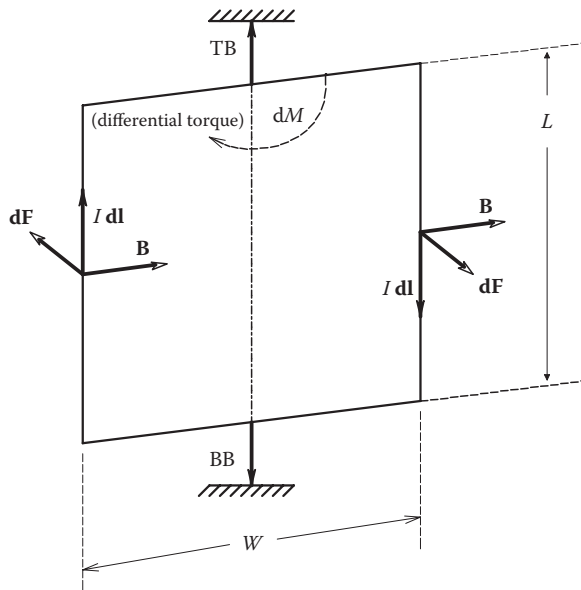


FIGURE 8.4

Magnetic (Faraday) vector forces on one turn of the rectangular, current-carrying coil of a D'Arsonval meter.

Unfortunately, the dynamic behavior of a D'Arsonval meter movement is not simply described by the balance between T_M and the spring torque. When the coil is moving, a back EMF is induced in it, given by

$$E_B = 2NLB(\dot{\theta}W/2) = NBA\dot{\theta} = K_B\dot{\theta} \tag{8.7}$$

where

- θ is the angle of rotation (pointer angle) of the coil
- $\dot{\theta}W/2$ is the tangential (linear) velocity of a side of the coil
- K_B is the coil's back EMF constant

A complete circuit of the D'Arsonval meter coil is shown in Figure 8.5. The self-inductance of the coil, L_C , is generally negligible at DC and low frequencies. From Figure 8.5, we see that the current in the coil is given by

$$I_C = \frac{V_C - E_B}{R_C + R_1} = \frac{V_C - E_B}{R_T} \tag{8.8}$$

The Newtonian torque balance equation can be written as

$$T_M = K_T I_C = J\ddot{\theta} + D\dot{\theta} + K_S\theta \tag{8.9}$$

or

$$K_T \left[\frac{V_C - K_B\dot{\theta}}{R_T} \right] = \{ J\ddot{\theta} + D\dot{\theta} + K_S\theta \}, \tag{8.10}$$

which reduces to

$$V_C(t) = \frac{R_T}{K_T} \{ \ddot{\theta}J + \dot{\theta}[D + K_B K_T/R_T] + \theta K_S \} \tag{8.11}$$

where

- D is the viscous damping torque constant
- J is the moment of inertia of the coil and pointer
- K_S is the torque constant of the torsion spring
- $R_T = R_C + R_1$ is the circuit resistance

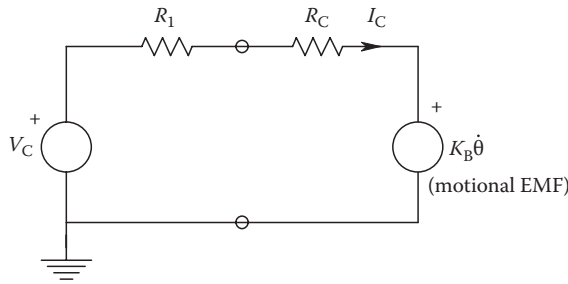


FIGURE 8.5

Equivalent circuit of the D'Arsonval meter's coil. The EMF induced by coil motion in the magnetic field is $V_B = K_B\dot{\theta}$. V_B acts to oppose the current induced by V_C .

When this second-order, linear ODE is Laplace transformed, we finally derive a transfer function relating pointer angle, θ , to applied voltage:

$$\frac{\Theta(s)}{V_C(s)} = \frac{K_T/(R_T K_S)}{s^2 J/K_S + s(D + K_B K_T/R_T)/K_S + 1}. \quad (8.12)$$

Note that the mechanical undamped natural frequency of the D'Arsonval meter movement is

$$\omega_n = \sqrt{\frac{K_S}{J}} \text{ r/s}, \quad (8.13)$$

and its damping factor is given by

$$\zeta = 1/2(D + K_B K_T/R_T) \sqrt{\frac{1}{J K_S}}. \quad (8.14)$$

Note that the electromagnetic component of the damping increases as R_T decreases (I_C increases).

Equation 8.12 earlier is a classic example of the low-pass transfer function of a second-order, linear electromechanical system. The step response of the D'Arsonval meter is easily found by setting

$$V_C(s) = \frac{V_C}{s}, \quad (8.15)$$

and using a table of Laplace transforms. The meter deflection as a function of time is found to be

$$\theta(t) = V_C \frac{K_T}{R_T K_S} \left\{ 1 - \frac{e^{-\omega \zeta t}}{\sqrt{1-\zeta^2}} \sin\left(\{\omega_n \sqrt{1-\zeta^2}\}t\right) + \tan^{-1} \frac{\sqrt{1-\zeta^2}}{\zeta} \right\}. \quad (8.16)$$

The mechanical response of a D'Arsonval meter to a step of applied voltage is shown in Figure 8.6. Note that as the damping factor, ζ , varies from 1 to 0, the response becomes more and more oscillatory. Obviously, it is inconvenient to have to wait for a highly underdamped meter to settle down to its SS reading of

$$\theta_{SS} = \frac{V_C K_T}{R_T K_S}. \quad (8.17)$$

Experience tells us that the best dynamic performance at a fixed ω_n for a second-order system's step response occurs when the damping factor, ζ , lies somewhere between 0.5 and 0.707.

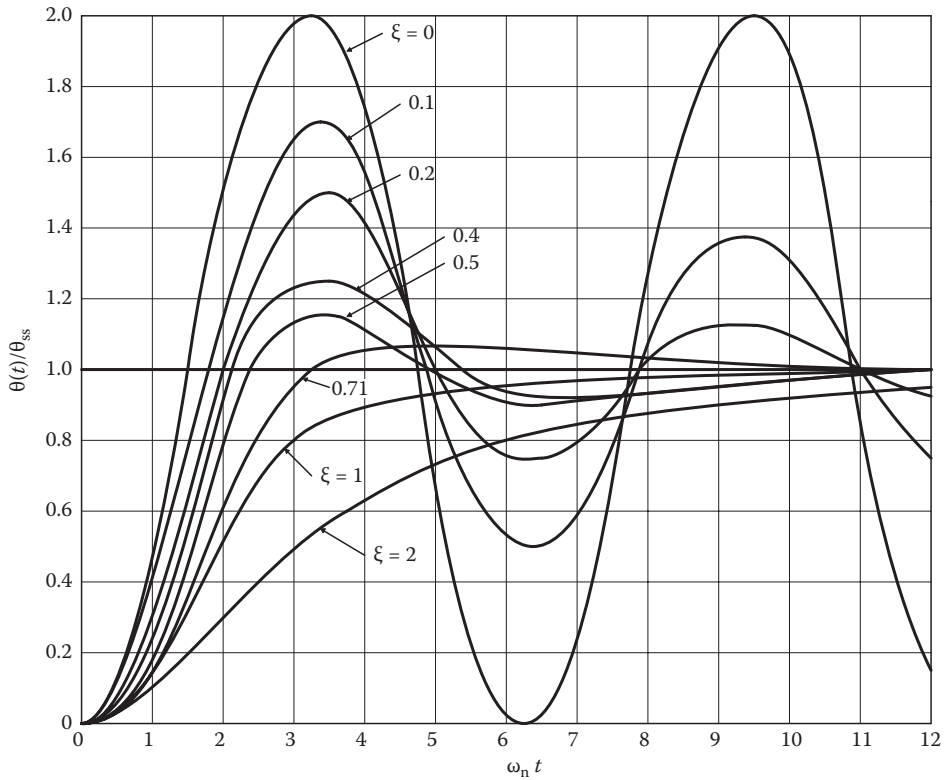


FIGURE 8.6 Mechanical response of a D’Arsonval meter to an applied step of voltage. Different damping factors are shown. $0.7 \leq \xi \leq 1$ gives the best compromise between rise time and no overshoots.

From the earlier analysis, we observe several important properties of D’Arsonval meters used as voltmeters. First, they require an SS current to produce a reading. This current must come from the CUT. Practical full-scale current is typically 20–50 mA, although special, sensitive D’Arsonval meter movements with torsion spring suspensions have been built with full-scale currents of 1 μ A or less. To obtain greater voltmeter sensitivity, we see from Equation 8.17 that we must increase $K_T = BNA$. Meter size and cost are design considerations. B is set by the state of the art and cost of permanent magnets available. A is set by size considerations and N by the maximum tolerable mass and moment of inertia of the coil. Too high a mass will give problems with bearings and coil suspension, a high J will result in an underdamped transient response and long settling time. Likewise, increased sensitivity reached by making the torsion spring constant, K_s , small will also give low damping. Second, the total resistance in the circuit, R_T , is set by the voltmeter’s sensitivity and the desired full-scale voltage. From the relations mentioned earlier, we see that

$$V_{FS} = \frac{R_S + R_C}{\eta} = \frac{R_T}{\eta} = R_T I_{MFS}. \tag{8.18}$$

Obviously, for any given full-scale voltage, meter loading by a D'Arsonval analog DC voltmeter will be less with a meter with a high η . Also, for very large V_{FS} , R_T will be very large. For example, R_T for a 500 VFS with a 20,000 Ω/V movement will be 10 M Ω . Multirange, DC, D'Arsonval voltmeters have a switch that switches various R_S s in series with the microammeter, satisfying Equation 8.18 earlier.

The capacitor or electrostatic voltmeter is an unusual type of electromechanical voltmeter that is best suited for DC or line frequency AC, high-voltage measurements. Capacitor voltmeters have been used from about 0.5 kV to over 200 kV with special insulation. A front view schematic diagram of a capacitor voltmeter is shown in Figure 8.7. The design is very much like a parallel-plate, radio tuning capacitor. Rotor plates pivot on low-friction bearings and are restrained by a linear torsion spring. The capacitance between rotor and stator plates is a function of the rotation angle, θ (pointer angle), and, in a linear capacitor, assumed here for convenience, can be written as

$$C(\theta) = C_o + K_C\theta. \quad (8.19)$$

Here, C_o is the capacitance between the plates at $\theta = 0^\circ$ and K_C is the capacitance constant. $C(\theta_{MAX})$ ranges from about 20 pF up to several hundred pF, depending on design. It can be shown that at a given voltage, there is an attractive torque acting to rotate the rotor into the stator to produce maximum capacitance and maximum stored electric field energy in the capacitance. The energy stored in the capacitor at SS is

$$W = \frac{1}{2}C(\theta)V_C^2. \quad (8.20)$$

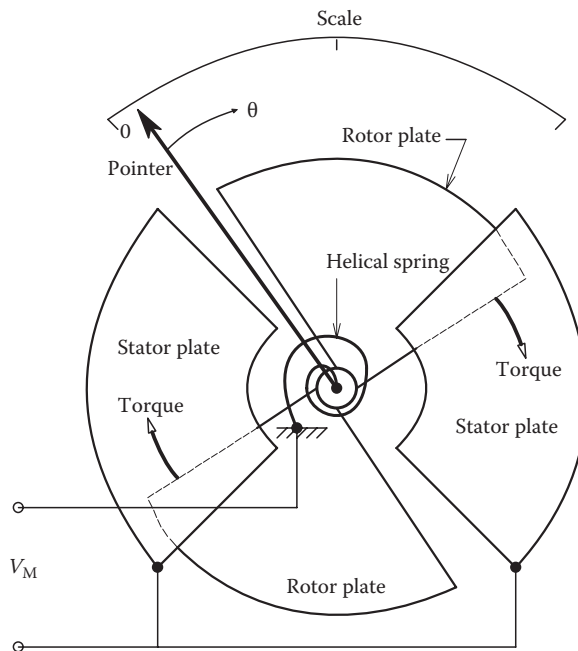


FIGURE 8.7

A transparent face view of an electrostatic, capacitor voltmeter. The electric field between the plates generates the deflection torque that must equal the spring torque in the SS. The meter scale is nonlinear.

The torque acting to rotate the rotor into the stator plates may be shown to be given by

$$T_M = \left. \frac{\partial W}{\partial \theta} \right|_{V_C} = \frac{V_C^2}{2} \frac{dC(\theta)}{d\theta} = \frac{V_C^2}{2} K_C. \quad (8.21)$$

Note that T_M is independent of θ . The electrostatic-derived torque must equal the Newtonian torques:

$$T_M = J\ddot{\theta} + D\dot{\theta} + K_S\theta. \quad (8.22)$$

After substituting the expression for T_M into the second-order ODE mentioned earlier and Laplace transforming, we obtain the second-order, low-pass transfer function:

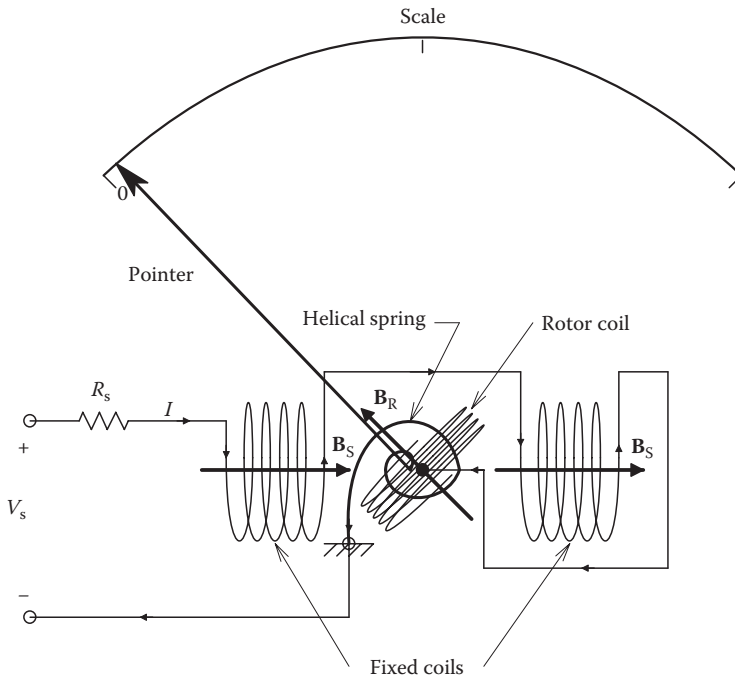
$$\frac{\Theta}{V_C^2}(s) = \frac{K_C/2K_S}{s^2J/K_S + sD/K_S + 1}. \quad (8.23)$$

The mechanical properties of the capacitor voltmeter, similar to those of the D'Arsonval meter movement, act as a mechanical LPF above $\omega_n = \sqrt{K_S/J}$ r/s. Hence, for DC SS conditions,

$$\theta_{SS} = \frac{\overline{V_C^2} K_C}{2K_S}. \quad (8.24)$$

The capacitor voltmeter is an example of a *square-law meter movement*. The scale has nonlinear calibration from Equation 8.24 because $C(\theta)$ is, in general, not linear. After the meter reaches SS deflection, no current flows in the circuit for an applied DC voltage step. However, if a capacitor voltmeter is suddenly placed directly across an energized, high-voltage circuit, it appears initially as a short circuit and heavy current flows into the meter. This may damage the CUT, so often a series resistor is used with the electrostatic voltmeter to limit the initial charging current. A typical electrostatic voltmeter, such as the Sensitive Research model ESD, has a scale calibrated from 1 to 5 kV, with the most sensitive (expanded) scale at the center of the meter ($\theta_{MAX}/2$) at about 2.75 kV. This meter has a guaranteed accuracy of 1% of full-scale voltage, a capacitance of 18 pF at full scale, and an insulation resistance of $10^{15} \Omega$. θ_{MAX} is about 65° for this meter, which is less than the 90° typical of most D'Arsonval meter movements.

The *electrodynamometer meter movement* is another voltmeter design, which, as we will see in the following, is square law and can be used for the measurement DC or power-line frequency AC voltage, current, and power. Here, we will examine the electro-dynamometer as a DC voltmeter. Figure 8.8 shows a diagram of the basic electro-dynamometer movement. Note that there are two symmetrically placed stator coils around a rotor coil, which, as in the case of other analog meters, is constrained by a torsion spring. When an electro-dynamometer meter is connected as a voltmeter as shown, the three coils are wired in series, as shown. Note that, initially, the flux density vectors repel each other and create a torque that rotates the pointer clockwise. Here, as in the case of the D'Arsonval

**FIGURE 8.8**

Face view of an electrodynamic meter movement connected as a voltmeter. The scale is nonlinear.

DC voltmeter, a series resistor is used to limit the meter's full-scale current. The energy stored in the magnetic field of the electrodynamic meter's coils is given by

$$W = i_M^2 [L_S/2 + L_R/2 + M_{SR}(\theta)] \quad (8.25)$$

where

i_M is the current in the coils

L_R is the self-inductance of the rotor coil

L_S is the self-inductance of the two stator coils together

$M_{SR}(\theta)$ is the mutual inductance between the stator coils and the rotor coil

Note that $M_{SR} = M_{RS}$. Because of meter coil geometry, $M_{SR}(\theta)$ may be approximated by

$$M_{SR}(\theta) = K_M (\theta - \theta_{MAX}/2). \quad (8.26)$$

Note that at half maximum deflection, the axis of the rotor coil is perpendicular to the axis of the stator coils, and $M_{SR} = 0$. The electromagnetically produced torque is given by

$$T_M = \left. \frac{\partial W}{\partial \theta} \right|_{i_M} = i_M^2 \frac{dM_{SR}(\theta)}{d\theta} = i_M^2 K_M = \left[\frac{V_M}{R_S + R_R + R_I} \right]^2 K_M. \quad (8.27)$$

As in the previous examples, we equate the electromagnetic torque with the Newtonian reaction torques and form a low-pass transfer function:

$$\frac{\Theta}{V_M^2}(s) = \frac{K_M / [(R_S + R_R + R_1)^2 K_S]}{\frac{1}{2} J / K_S + s D / K_S + 1}. \quad (8.28)$$

In the SS, the meter deflection is

$$\theta_{SS} = \overline{V_M^2} K_M / [(R_S + R_R + R_1)^2 K_S]. \quad (8.29)$$

Here, as in the case of the electrostatic voltmeter, the deflection is square law, that is, proportional to the mean-squared applied voltage. (The inertia of the meter movement effectively averages the squared voltage.) Electrodynamicometer voltmeters are seldom used above powerline frequency because the coil inductive reactance begins to become more than 1% of $(R_S + R_R + R_1)$, producing frequency-dependent calibration errors. Electrodynamicometer voltmeters generally require far more current for SS, full-scale deflection than do D'Arsonval voltmeters. Typical electrodynamicometer voltmeter sensitivities range from 10 to 50 Ω/V , or full-scale currents range from 0.1 to 0.02 A. It appears that this type of meter is truly a *wattsucker* and best suited for measurements on power systems rather than on electronic circuits. Dynamometer voltmeters, ammeters, and wattmeters generally have 80° – 90° arc scales with scale lengths of 6.5"–7.0" and a mirror to eliminate parallax when reading the pointer. They are calibrated for DC or AC voltages or currents ranging from about 25 to 500 Hz, depending on design. Accuracies range from 0.1% to 0.25% of full-scale reading. Because of the mass of the moving coil, the bearings must be rugged and have low friction. They are mechanically fragile.

8.2.2 Electronic DC Voltmeters

In this section, we will consider the design, applications, and limitations of various types of electronic, DC voltmeters, beginning at the low end of the voltage scale with nanovoltmeters, which are designed to work with extremely low DC input potentials seen through the Thevenin resistors on the order of tens of ohms. Figure 8.9 illustrates the design of a *chopper-type nanovoltmeter*. A special, low-noise electromechanical chopper running at 60 Hz is used to convert the low-level DC potential, V_s , to an AC signal, V_2 , which is amplified by an r-c amplifier, A_1 , with a gain of -10^4 . The AC signal's frequency, 60 Hz, is above the $1/f$ noise portion of the r-c amplifier's input voltage noise spectrum, e_{na}^2 . The equivalent short-circuit voltage noise spectrum of the input amplifier is of the form

$$e_{na}^2(f) = \eta + b/f \text{ MS V/Hz}. \quad (8.30)$$

In the flat (white) part of the noise's power density spectrum, η can be as low as 49×10^{-18} MS V/Hz (Toshiba 2SK146 JFET). Following amplification by A_1 , V_3 is synchronously demodulated by a second chopper. The output of the second chopper, V_4 , has a DC component that is conditioned by the OA LPF, A_2 . A block diagram summarizing these operations is shown in Figure 8.10. With the feedback resistors shown, the nanovoltmeter's transfer function can be written as

$$\frac{V_o}{V_s}(s) = \frac{(-10^4)[-100/(10s+1)]}{1 + \frac{1}{10}[-(10^6)(10^{-4})/(10s+1)]} = \frac{9.901 \times 10^3}{s10/101 + 1} \quad (8.31)$$

where

V_o is the indicated voltage

V_s is the applied voltage

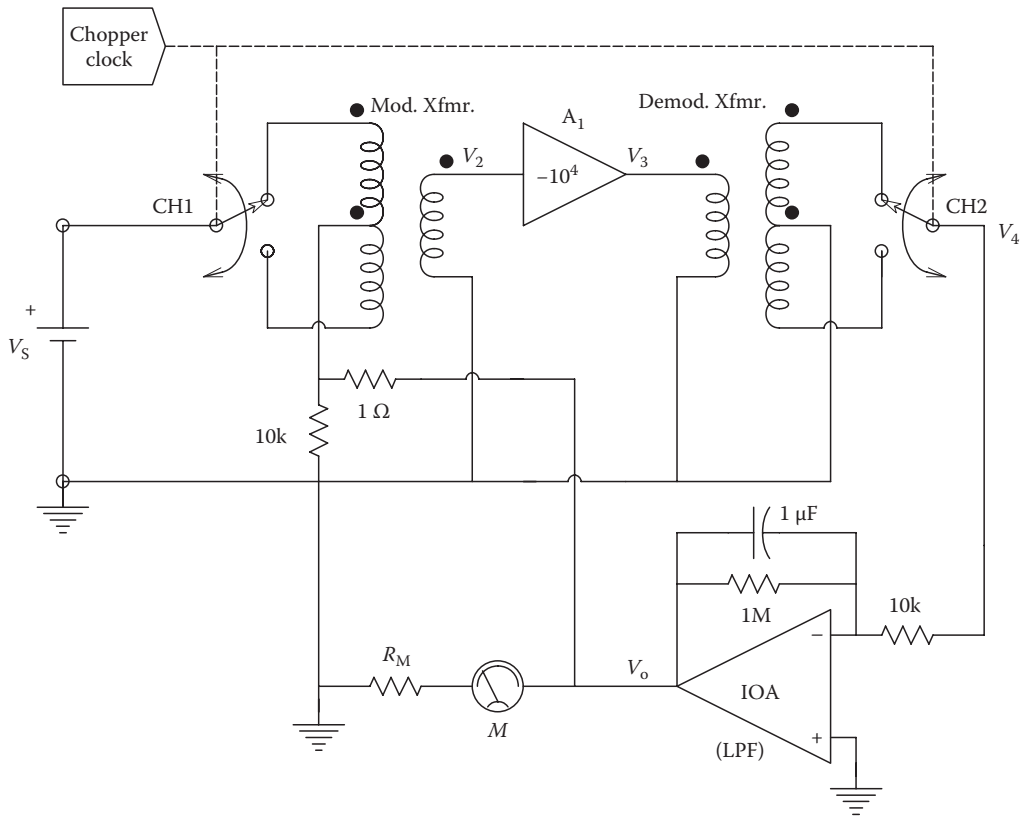


FIGURE 8.9
Schematic diagram of a chopper-type DC nanovoltmeter.

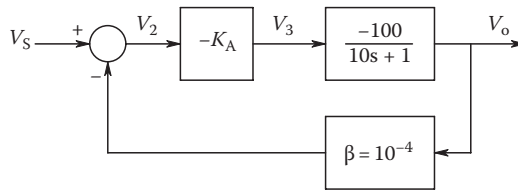


FIGURE 8.10
Systems block diagram illustrating the dynamics of a chopper nanovoltmeter.

Thus, there is low-noise amplification of the DC input voltage by 79.9 dB and single-pole, low-pass filtering with a break frequency at 1.61 Hz. When operating on the lowest range (typically 10 nV full scale), it is necessary to further amplify V_o to a level where an analog, D'Arsonval microammeter, can serve as an output indicator. An additional DC gain stage of 10^2 will boost a 10 nV input to a 10 mV output. Now if a 20 μ A full-scale, DC, D'Arsonval meter is used, it will read full scale if its internal resistance is 500 Ω (or less).

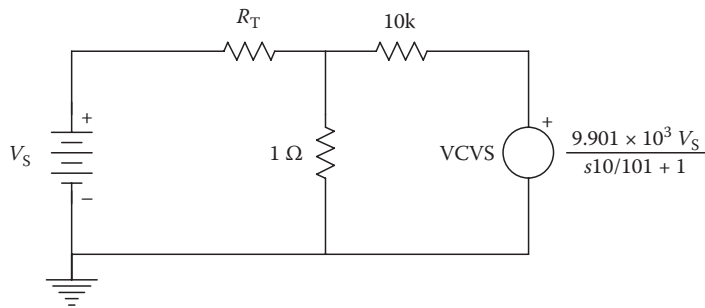


FIGURE 8.11
Equivalent input circuit of the chopper-type nanovoltmeter.

The input resistance of the nanovoltmeter can be calculated using the equivalent input circuit of Figure 8.11. At DC, R_{IN} can easily be shown to be

$$R_{IN} = (R_T + 1\Omega)101.01\Omega, \quad (8.32)$$

where R_T is the DC series resistance of either side of the input chopper's coil.

The DC input resistance of the nanovoltmeter amplifier may be on the order of 1 k Ω because of the action of the negative feedback creating the VCVS.

The specifications of two commercially available nanovoltmeters are outlined below. The now-discontinued Keithley Model 148 nanovoltmeter has a design similar to that described earlier. Its output is read on a center-scale analog microammeter. Full-scale ranges begin at 10 nV and go to 0.1 V in 18 overlapping ranges. On the 10 nVFS range, the input resistance is over 1 k Ω and the DC nV source must not have a Thevenin resistance of over 10 Ω for noise and loading reasons. On the 0.1 V_{FS} range, the input resistance is greater than 1 M Ω , and the source resistance should be less than 10 k Ω . Resolution is better than 1 nV on the 10 nV range, and accuracy is $\pm 2\%$ of full scale on all ranges. The line frequency rejection is 3000:1 on the 10 nV range, and the CMRR is 160 dB at 60 Hz on all ranges.

The now-discontinued Keithley Model 181 nanovoltmeter has a 6 1/2-digit digital read-out and is IEEE-488 bus compatible. It uses a special low-noise JFET headstage to obtain an input resistance of over $10^9\Omega$ on the low ranges. There are seven full-scale ranges, from 2 mV to 1 kV. Meter noise is claimed to be less than 30 nV peak to peak on the 2 mV scale with the output LPF on. The LPF is a 3-pole digital design, with an equivalent time constant of 0.5, 1, or 2 s, depending on meter range. This meter can be run with 5 1/2-digit resolution when inputs are noisy. Keithley's current digital nanovoltmeter is their Model 2182A.

Because of the special challenge in measuring DC voltages in the nanovolt range, meters must be allowed to warm up at least an hour before use to reach thermal equilibrium. DC nanovolt measurements are confounded by factors that are normally neglected in high-level measurements. These include the generation of thermoelectric (Seebeck) EMFs by junctions between dissimilar metals at different temperatures, the induction of LF EMSs by time-varying magnetic fields whose flux passes through the circuit from the meter to the source, LF and DC potentials due to ground loops, LF EMFs due to triboelectric effects on connecting coaxial cables, and piezoelectric EMF artifacts. A good discussion of sources of errors in low-level, DC measurements can be found in the Keithley Handbook on *Low Level Measurements*, 6th edition (Keithley Instruments, Inc., 2004).

In order to operate either an analog meter or a digital interface, the DC voltage under measurement must be amplified to a level useful to drive the analog indicating meter (about 100 mV) or the ADC (1–10 V). This means that microvolt level signals need to be amplified by factors of 10^5 – 10^7 . Inherent in all DC amplification is the addition of noise from the amplifier, generally consisting of a mixture of white noise and $1/f$ noise, DC offset, and drift due to temperature changes. It was seen in Section 3.6 that the noise performance (as well as the DC drift performance) of a high-gain amplifier is set by the headstage that must have a gain greater than five. Thus, the amplifiers used in any sensitive, low-level, DC voltmeter must have low noise and also have very low DC drift and offset voltage. OAs suitable for high-gain, DC signal conditioning include chopper-stabilized and commutating auto-zero amplifiers, the latter being described in Section 2.4.3.

The Intersil ICL7600/ICL7601 commutating auto-zero OA is well suited for DC voltage amplification system headstages. The ICL7600 has a low input offset voltage of $2\ \mu\text{V}$, a very low input offset voltage drift of $0.2\ \mu\text{V}/\text{year}$, a very low input offset voltage tempco of $\pm 0.005\ \mu\text{V}/^\circ\text{C}$, and a DC input bias current of $300\ \text{pA}$. The switching (commutating) frequency of the ICL7600 system capacitors can be set from 160 to 5200 Hz.

Another approach to low-noise, drift-free DC amplification is to use a chopper-stabilized OA, such as the ICL7650. This OA has a DC input offset voltage (V_{OS}) of $1\ \mu\text{V}$ over the operating temperature range, an offset voltage tempco of $\pm 0.01\ \mu\text{V}/^\circ\text{C}$, a DC input bias current of $35\ \text{pA}$, a DC open-loop gain of 134 dB, and a GBWP of 2 MHz. The internal chopping frequency is 200 Hz. Many other manufacturers offer integrated circuit, chopper-stabilized, or auto-zero DC amplifiers. Tempcos of $50\ \text{nV}/^\circ\text{C}$ are pretty much state of the art, and input DC offset voltages range from 1 to $5\ \mu\text{V}$.

An LPF is generally used at the output of the DC amplifier stages, before the ADC or the DVM module. If a D'Arsonval analog meter is used, the meter movement itself acts as an electromechanical LPF (see Equation 8.11). However, additional low-pass filtering may be required to improve the DC microvoltmeter's resolution. A simplified schematic of an analog DC voltmeter with electronic amplification is shown in Figure 8.12. When an analog

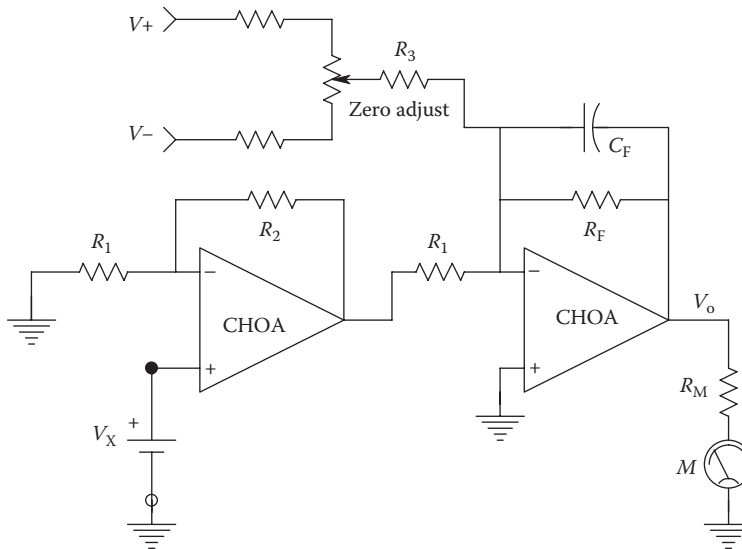


FIGURE 8.12

Schematic diagram of an analog, DC millivoltmeter. Two chopper-stabilized op amps are used for amplification. The second op amp is used as an LPF and for zeroing. The meter, M , is a DC D'Arsonval microammeter.

meter is used, precision is seldom more than $\pm 0.25\%$ of full scale because of the visual uncertainties of reading the pointer position. Properly designed digital meters can commonly reach precisions of $\pm 0.005\%$, or 50 ppm. Noise, DC drift, and quantization error are still the ultimate determinants of resolution in instruments with digital readouts, however.

Electronic DC voltage measurements above the single volt range generally do not require amplification, either for an analog indicating meter or for a digital output indicator. In fact, electronic voltmeters used in the range of 10–1000 V generally make use of internal input voltage dividers to attenuate the voltage under measurement to a value useful to the meter amplifier or ADC. An external resistive voltage divider probe is generally used to measure DC voltages from 1 to 40 kV. Such a probe makes an external voltage divider with the DC voltmeter's input resistance. For example, the Fluke Model 80K-40 high-voltage probe is designed to be used with any DC voltmeter having an input resistance of 10 M Ω . It makes a 1000:1 voltage divider, so 1 kV applied to the probe gives 1 V at the meter. The probe and meter present a 1000 M Ω load to the high-voltage CUT; hence, the probe and meter draw 1 μ A/kV from the CUT.

At DC voltages above 40 kV, special measurement techniques must be used because of insulation problems associated with the resistive voltage divider. At very high voltages, resistors may behave erratically due to the generation of corona discharge or leakage from the absorption of water from the atmosphere. In fixed installations, these problems can be overcome by using proper insulators and couplings and housing the resistors in an inert, insulating gas such as SF₆ or Freon or in high-pressure (≥ 10 atm.), dry air, or nitrogen.

In order to measure DC voltages in excess of 200 kV electronically, use is often made of electrometer-based, electric field measuring instruments. Such techniques make use of the fact that the geometry is fixed in super high-voltage systems; hence, the electric field geometry is fixed, and at any point in the field, the field strength is proportional to the DC voltage generating the field. We discuss instruments used to measure electric fields in the next section.

8.3 Measurement of Static Electric Fields and the Potential of Charged Surfaces

Static (DC) electric fields can arise from a variety of physical and electrical causes that include the presence of bound or mobile surface charges (electrons and ions) on conductors or insulators. The surfaces can become charged by being bombarded with charged particles such as sand or raindrops or from triboelectric effects where charges become separated from the surfaces of insulators when they rub on conductors or other insulators. Of course, static electric fields exist around conductors operating at high DC voltages with respect to ground and also near the faces of certain monitors using large CRTs.

Static charges and fields can create problems ranging from annoying (*static cling* on clothing and paper) to extreme danger (static discharge sparks ignite solvent vapors or combustible dust, causing an explosion). Other problems caused by static electricity include the attraction of fine dust particles to clean, charged surfaces; the attraction of charged dust particles to clean, grounded surfaces; and the physical attraction of charged materials such as paper or plastic film to other objects, and when the electric field strength exceeds a critical value (about 3×10^6 V/m in air at STP), there can be an abrupt discharge of charged

persons or objects to static-sensitive semiconductor devices or electronic equipment, causing damage. Clearly, it is important to be able to measure the DC voltages associated with the production of DC fields, to give warning of potentially damaging or dangerous conditions.

The first method of measuring the DC potential of a charged object field makes use of a DC electrometer voltmeter connected to a special probe as shown in Figure 8.13A. To measure the potential of the object in question, which may be a charged insulating surface

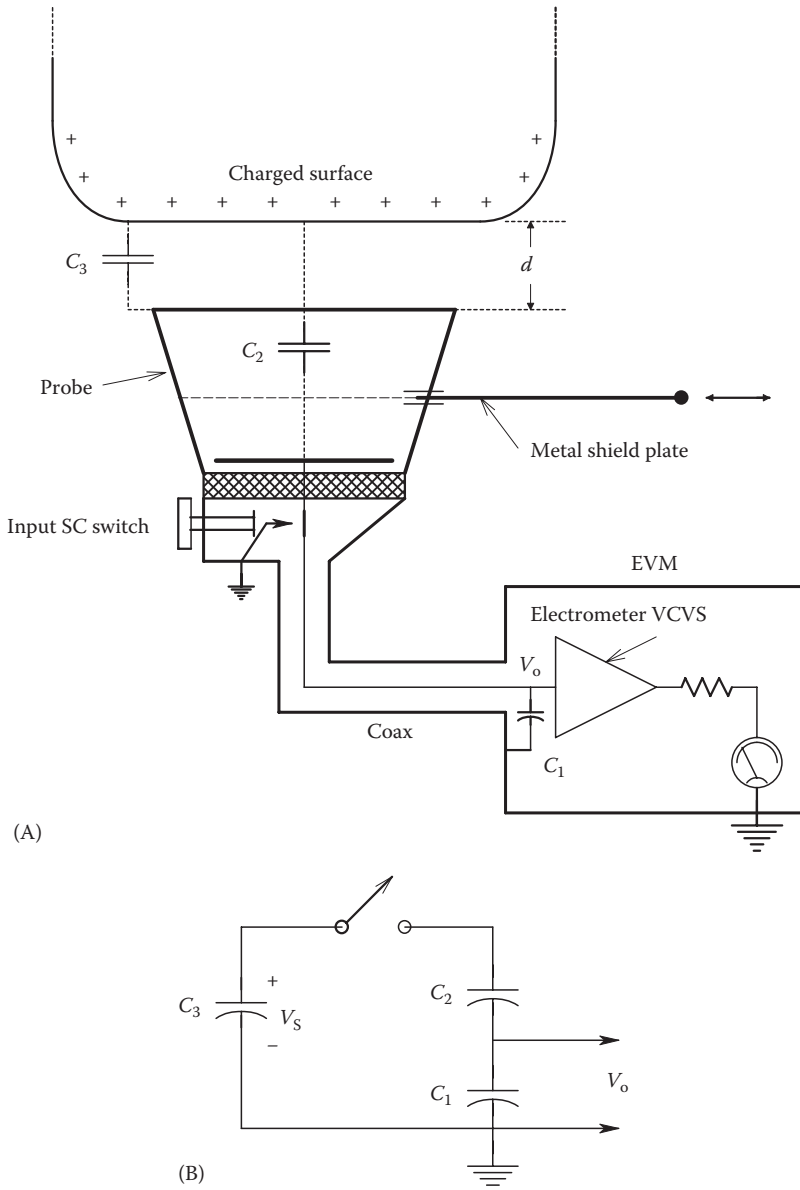


FIGURE 8.13

(A) Section through a capacitive probe used to measure high DC electrostatic potentials on charged objects. The probe acts as a voltage divider. (B) Equivalent circuit of the high-voltage capacitance probe.

or conductor, the probe is brought to a standard distance, d , from the object. The standard distance is required to set the capacitance, C_2 , to a known value. Figure 8.13B shows the equivalent circuit for the DC electrometer field meter. To measure V_S , capacitance C_1 is initially shorted to ground, and a grounded metal plate is used to cover the proximal plate of capacitor C_2 . (C_1 is the input capacitance of the electrometer voltmeter plus the capacitance to ground of the wire connecting the proximal plate of C_2 to the electrometer input. C_2 is the capacitance between the proximal plate of C_2 and the charged object.) Next, the metal plate is slid aside, fully exposing the plate of C_2 . C_2 and C_1 are in series and acquire charge from the field from the potential on C_3 . Initially, if C_3 is isolated and is at potential V_S , some charge flows from C_3 into C_2 and C_1 , causing the voltage on C_3 to drop to V'_S . In the SS, this can be written as

$$(V_S - V'_S)C_3 = V'_S \frac{C_1 C_2}{C_1 + C_2}. \quad (8.33)$$

The new voltage on the isolated object with capacitance C_3 is found to be

$$V'_S = \frac{V_S C_3}{\frac{C_1 C_2}{C_1 + C_2} + C_3}. \quad (8.34)$$

Now because the same current flows in C_3 , C_2 , and C_1 , the charge that resides in C_2 is also present in C_1 .

Thus, from

$$V_o C_1 = (V'_S - V_o) C_2, \quad (8.35)$$

we find that

$$V_o = \frac{V'_S C_2}{C_1 + C_2}. \quad (8.36)$$

If the expression for V'_S in Equation 8.34 is substituted in Equation 8.36 earlier, we finally obtain

$$\frac{V_o}{V_S} = \frac{C_2 C_3}{C_1 C_2 + C_3 (C_1 + C_2)}. \quad (8.37)$$

If $C_3 (C_1 + C_2) \gg C_1 C_2$, then Equation 8.37 reduces to

$$\frac{V_o}{V_S} \cong \frac{C_2}{C_1 + C_2}. \quad (8.38)$$

For example, if $C_2 = 10^{-11}$ F and $C_1 = 10^{-7}$ F, then $V_o = 10^{-4} V_S$. Hence, an initial 10 kV charge on C_3 will result in a 1 V reading on the electrometer.

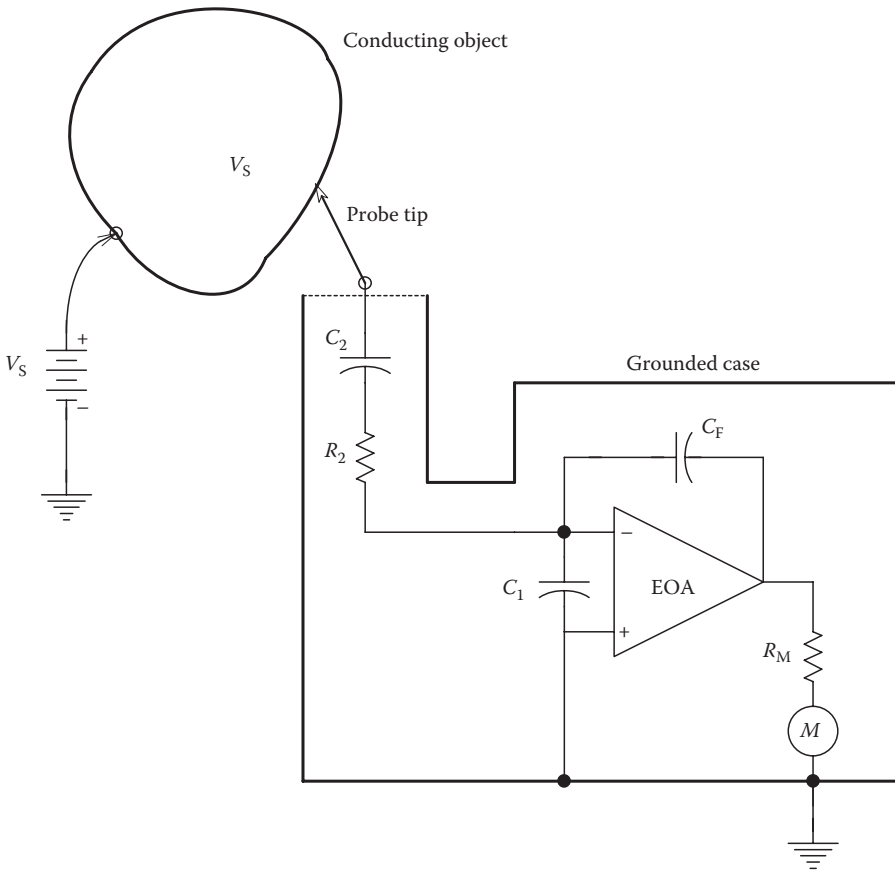


FIGURE 8.14
An EOA is used as a charge amplifier in this circuit to measure high DC voltages.

In the system described earlier, C_1 's value depends on the capacitance to ground of the cable connecting the probe with the electrometer, as well as the input capacitance of the electrometer. In order to eliminate the system's requirement for a fixed input geometry to determine C_1 , the charge amplifier configuration shown in Figure 8.14 can be used. Now, C_1 appears between the summing junction, which is at virtual ground, and actual ground. Hence, negligible current flows through C_1 , and C_F determines the output voltage of the system. R_2 limits the initial charging current through C_2 and C_F so that the EOA will not go into current saturation. Now, it is easy to show that the charge amplifier's output is

$$V_Q = -V_S \left(\frac{C_2}{C_F} \right). \tag{8.39}$$

We wish C_2/C_F to be 10^{-3} or 10^{-4} so the OA will not saturate for large V_S .

Another approach to the problem of measuring DC voltages in the kilovolt range through their fields is to use a time-modulated C_2 and to measure the AC that flows in resistor R , shown in Figure 8.15. In this form of DC high-voltage meter, the modulation of C_2 may be expressed mathematically as

$$C_2(t) = C_{20} + \Delta C \sin(\omega t), \quad \Delta C \ll C_{20}. \tag{8.40}$$

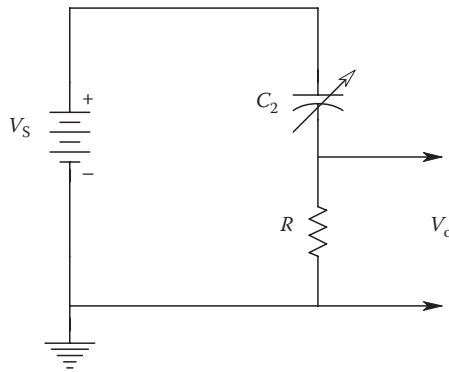


FIGURE 8.15 Equivalent circuit of a field mill, DC voltmeter. The value of C_2 is modulated sinusoidally around an average value (cf. Equation 8.40).

There are several ways in which C_2 can be modulated. One common method is to have the blades of a grounded, rotating chopper wheel alternately cover and uncover the proximal plate of C_2 . Another method physically modulates the distance between the proximal plate and the charged surface. As a result of the modulation of C_2 , an AC of frequency ω will flow through resistor R to ground. This current may be found by writing a loop equation for the system of Figure 8.15:

$$V_S = \dot{q}R + q/C_2(t) \tag{8.41}$$

or

$$V_S = \dot{q}R + q(1 - \Delta C/C_2)/C_{20}. \tag{8.42}$$

The ODE of Equation 8.42 can be put in the standard form:

$$\dot{q} + q \frac{1}{C_{20}R} \left(1 - \frac{\Delta C \sin(\omega t)}{C_{20}} \right) = \frac{V_S}{R}. \tag{8.43}$$

This ODE has the standard solution

$$y \exp\left(\int P dx\right) = \int \exp\left(\int P dx\right) Q dx + c, \tag{8.44}$$

in which $x = t$, $y = q$, $Q = V_S/R$, and $P(t) = [1 - \Delta C \sin(\omega t)]/RC_{20}$. Performing the integrations indicated in the standard solution, we find that

$$q(t) = \frac{V_S C_{20}}{1 - [\Delta C \sin(\omega t)]/C_{20}} + c. \tag{8.45}$$

The SS sinusoidal current in the series circuit is

$$i_{ss}(t) = \dot{q} = V_S \Delta C \omega \cos(\omega t). \quad (8.46)$$

Thus, the AC voltage across R is proportional to V_S :

$$v_o(t) = V_S R \Delta C \omega \sin(\omega t). \quad (8.47)$$

For example, if $\omega = 100$ r/s, $V_S = 100$ kV, $R = 10^5 \Omega$, and $\Delta C = 10^{-11}$ F, then the peak value of v_o is 10 V. Of course, the input resistance of the AC voltmeter used to measure v_o should be over $10^8 \Omega$.

The two methods of measuring the high DC voltage, V_S , described earlier do not require direct contact with the charged surface. However, their accuracy is limited by the accuracy in determining either C_2 or ΔC , which in turn depends on accurate determination of the distance from the proximal plate of C_2 to the charged surface. At least one commercially available *digital field meter*, designed to have a working distance of 4.0" from the charged surface, uses an ultrasonic ranging system similar to that used in popular cameras, which use self-developing, *instant* film, to permit precise adjustment of the working distance. One Desco Industries' handheld, digital static field meter reads DC surface voltage and polarity on objects up to ± 20 kV with $\pm 5\%$ accuracy of the reading at a distance of 1 in. Correct distance is indicated by LEDs. The AlphaLab, Inc. Model SVM2 Surface DC Voltmeter has the following specifications: range at 1" = ± 30 kV, resolution at 1" = 1 V, accuracy = $\pm 2\%$ of reading, and capacitance to ground = 3 nF. LessEMF.com offers a wide range of electric field measurement instruments; their cat. no. A152-D digital electric field meter measures AC fields. This meter reads 1–20,000 V/m with 2% accuracy, from 40 Hz to 20 kHz.

Other field voltmeters require a physical measurement of their working distance, necessarily with an insulated ruler. Voltage ranges of some portable, battery-operated field meters can span from 0–500 V to 0–200 kV. Such meters are typically used to measure static charges accumulated in industrial processes such as paper and plastic sheet manufacturing.

A severe static electricity problem occurs when helicopters fly through charged particles such as sand or water droplets. The metal airframe of the helicopter can acquire a potential of thousands of volts. If not discharged before landing, this charge can be dangerous in fueling and cargo-handling operations. Bradford (1975) described a system that automatically senses the magnitude and sign of the helicopter's charge and automatically discharges the airframe by spraying charged droplets of isopropyl alcohol from the helicopter's fuselage. The helicopter's fuselage has a capacitance of about 2200 pF when on the ground. Its body capacitance when hovering at 25 m is about 500 pF. To measure the helicopter's body potential, Bradford used a corona field sensor manufactured by Dayton Aircraft Products (Cline and Grant 1974). A block diagram of the Dayton sensor is shown in Figure 8.16A. A 400 Hz AC voltage is amplified to 4000 peak volts and applied to a wire brush corona probe that protrudes through a hole in the fuselage into the high DC field region next to the fuselage. As seen in Figure 8.16B, if the fuselage is at zero potential with respect to ground, the AC due to corona flowing from the probe is symmetrical and has zero average. If the helicopter body is at some positive potential, V_S , the DC field from V_S causes extra corona current to flow on the positive cycles and less to flow during the negative cycles, giving the corona current an average (DC) component that is found to be proportional to E , hence V_S . The sensor electronics and static charge–discharge controller were located inside the fuselage where, of course, the electric field is zero.

Calibration of the field sensor was made difficult by the fact that helicopter body capacitance decreases nonlinearly with hover height. As this decrease occurs, the body potential increases if the charge remains constant. Offsetting this increase in V_s is the fact that the field strength in the vicinity of the corona probe tends to decrease from geometrical considerations as altitude increases. Bradford (1975) used a large, parallel-plate capacitor (plate separation of 1 ft) to generate a uniform field to calibrate the corona probe field sensor. He obtained a conditioned sensor (DC) output of about 1 V/kV on the capacitor.

It is generally difficult to measure an actual electric field's strength in volts/meter without disturbing the field with the measurement apparatus. This is true of field measurements in conducting media as well as in free space. If the measurement sensor is a dielectric with a low dielectric constant, and not grounded, the disturbance of the field will be minimal.

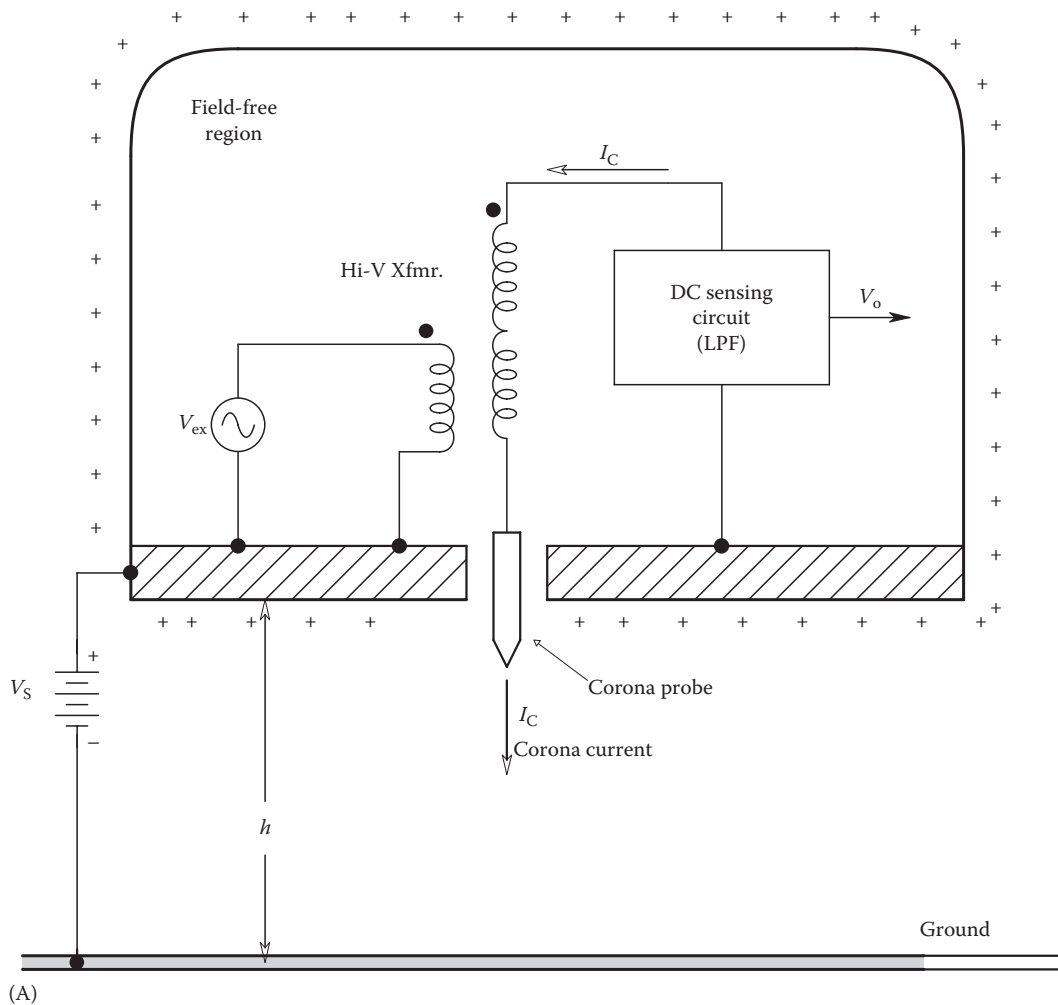


FIGURE 8.16
 (A) Diagram of a Dayton Aircraft corona field sensor, used to measure the potential accumulated on helicopter fuselages.

(continued)

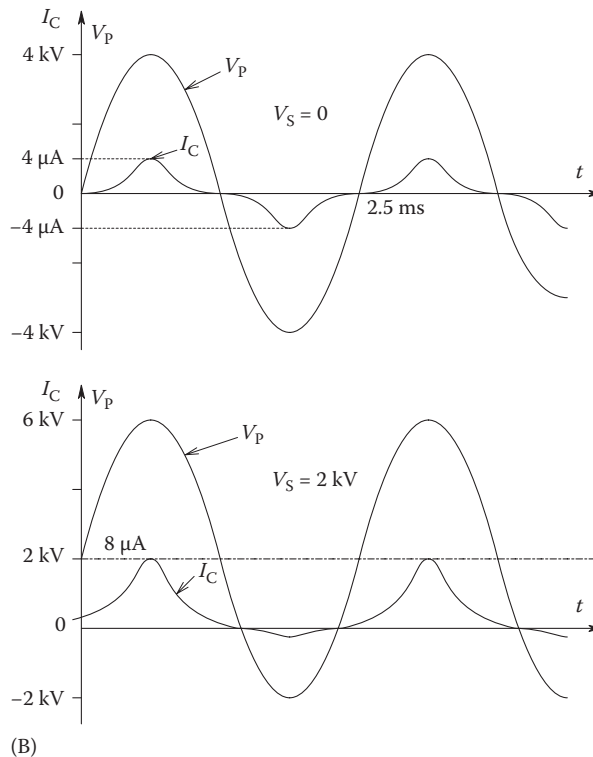


FIGURE 8.16 (continued)

(B) Current waveforms in the corona field sensor. In the top waveform pair, $V_S = 0$, while in the bottom pair, $V_S = 2 \text{ kV}$.

There are several, possible, electro-optical approaches to the problem of measuring DC electric fields. One is to measure the change in the linear polarization of light transmitted through a quartz crystal subject to an electric field applied across a preferred axis. This phenomenon is called the *electrogyration effect* (Rogers 1977). Another electro-optical method proposed for measuring DC electric field strength makes use of the *Pockels effect*, in which the application of an electric field along the *privileged axis* of the crystal causes changes in the refractive indices along two orthogonal axes (x and y). Certain crystalline substances, such as *ammonium dihydrogen phosphate* (ADP), *potassium dihydrogen phosphate* (KDP), *cuprous chloride* (CuCl), *cadmium telluride* (CdTe), and *gallium arsenide* (GaAs), exhibit the Pockels effect.

To examine how the Pockels effect might be used to measure an electric field, we orient a Pockels crystal so that the privileged axis is perpendicular to the gradient of the field to be measured. An internal \mathbf{E} field is set up in the crystal that is proportional to the external field we wish to measure. The internal electrical field, if correctly aligned, causes a symmetric change in the optical index of refraction along the x - and y -axes. Circularly polarized light is directed at the input surface of the crystal and propagates within the crystal along the z -axis. The light is acted on by the birefringence induced by the internal \mathbf{E} field and emerges elliptically polarized. The emergent light is passed through a linear polarizer analyzer plate that is oriented at 45° with respect to the x - and y -axes. The intensity of light from the output of the linear polarizer can be shown to be a linear function of the

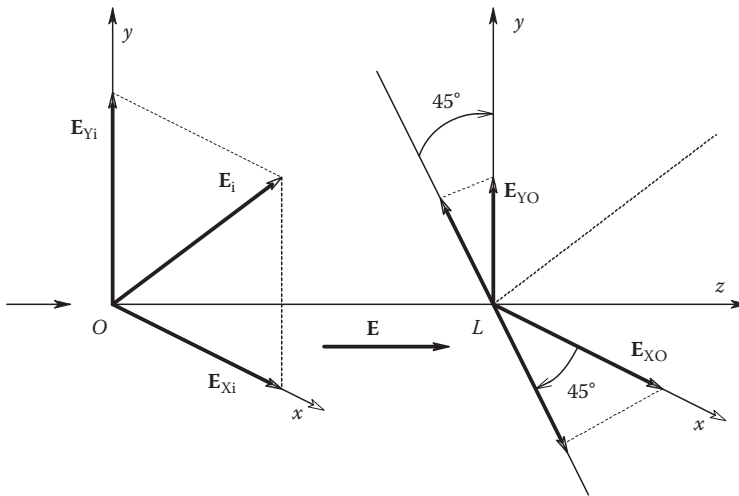


FIGURE 8.17
Vector diagram relevant to the use of a Pockels' cell to measure electric fields.

crystal's internal E field. Referring to Figure 8.17, we see that the x - and y -components of the electric vector of the circularly polarized input light at the entrance face of the crystal can be written as

$$E_{Xi} = E_i \cos(\omega t) \tag{8.48a}$$

$$E_{Yi} = E_i \cos(\omega t - \pi/2). \tag{8.48b}$$

Now, the refractive indexes of the Pockels crystal are affected by its internal electrical field, E , according to the linear model:

$$n_x = n_o + K_p E \tag{8.49a}$$

$$n_y = n_o - K_p E. \tag{8.49b}$$

After propagating a distance L meters through the crystal, the x - and y -components of the exiting light ray's E vector can be written as

$$E_{XO} = E_i \cos(\omega t + \phi_x) \tag{8.50a}$$

$$E_{YO} = E_i \cos(\omega t + \phi_y - \pi/2). \tag{8.50b}$$

The phase lags due to orthogonal changes in propagation velocity are assumed to be linearly proportional to the internal E field and can be written as

$$\phi_x = \frac{2\pi L n_x}{\lambda} = \frac{2\pi L n_o}{\lambda} + \frac{2\pi L K_p E}{\lambda} = \phi_o + \frac{2\pi L K_p E}{\lambda} \tag{8.51a}$$

$$\phi_y = \frac{2\pi L n_y}{\lambda} = \frac{2\pi L n_o}{\lambda} - \frac{2\pi L K_p E}{\lambda} = \phi_o - \frac{2\pi L K_p E}{\lambda}. \tag{8.51b}$$

Propagation through the linear polarizer selects those lightwave \mathbf{E} field components along the line, LAN, defining the polarizer's preferred axis. Thus, the lightwave \mathbf{E} field amplitude along the LAN axis exiting the polarizer can be written as

$$E_{AN} = \frac{\mathbf{E}_{XO}}{\sqrt{2}} - \frac{\mathbf{E}_{YO}}{\sqrt{2}} = \left(\frac{\mathbf{E}_i}{\sqrt{2}} \right) \left[\cos(\omega t + \phi_X) - \cos(\omega t + \phi_Y - \pi/2) \right]. \quad (8.52)$$

We note that the second term in Equation 8.52 is of the form $\cos(\alpha - \pi/2)$, which by trig. identity can be written as $\sin(\alpha)$. It is well known that the *intensity* of an EM plane wave is related to its \mathbf{E} field peak amplitude by

$$I_{AN} = E_{AN}^2 \left(\frac{\epsilon_0 c}{2} \right). \quad (8.53)$$

By substituting Equation 8.52 into Equation 8.53, and assuming that $\sin(x) \approx x$ for small x , we find that the average intensity is given by

$$I_{AN} \cong \frac{E_i^2}{2} \left[1 + \frac{4\pi L K_P E}{\lambda} \right] \left(\frac{\epsilon_0 c}{2} \right). \quad (8.54)$$

In the foregoing analysis, we have assumed that there were zero neutral density and reflection losses in E_i and that $2\pi L K_P E / \lambda \ll 0.25$ r. Other detection schemes are possible, but the important result is that the output intensity varies linearly with the Pockels crystal's internal \mathbf{E} field magnitude. For a numerical example, consider a 1.0 cm crystal of potassium dihydrogen phosphate (KDP, KH_2PO_4): we evaluate KDP's sensitivity to be

$$\frac{4\pi L K_P}{\lambda} = \frac{4\pi \times 0.01 \text{ m} \times 29 \times 10^{-12} \text{ m/V}}{632.8 \times 10^{-9} \text{ m}} = 5.7576 \times 10^{-6} \text{ m/V}. \quad (8.55)$$

Thus, an internal field on the crystal of 10^4 V/m or 10 V/mm would produce a 5.76% increase in the detected intensity. A DC field measuring instrument based on the Pockels effect would need a means of correcting for intensity changes in the laser light source.

8.4 DC Measurements

The technology of analog DC measurements may be traced back to the early nineteenth century (Drysdale et al. 1952). In 1819, Oersted discovered that a wire carrying DC, when held near a compass needle, caused the compass to deflect—implying that the current-carrying wire generated a magnetic field. This effect was put to use in by Lord Kelvin, who, in 1858, developed a four-coil mirror galvanometer. In Kelvin's instrument, shown in Figure 8.18, four coils were arranged in two opposing pairs; their fields acted on two arrays of permanent magnets with opposing polarities attached

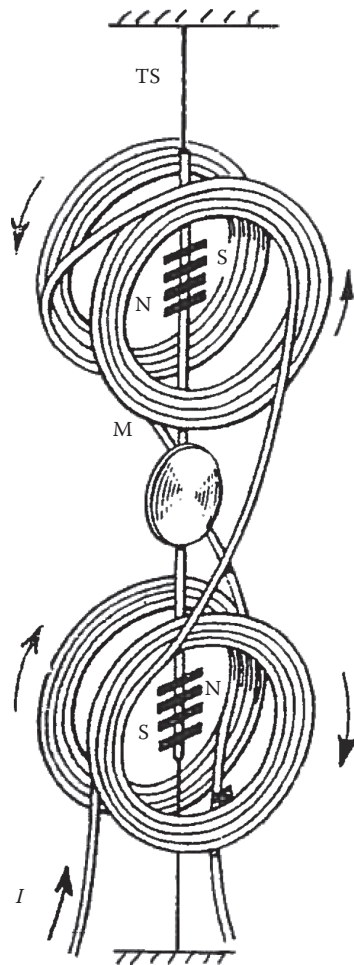


FIGURE 8.18
A drawing of Lord Kelvin's four-coil, permanent-magnet, mirror galvanometer.

to a vertical torsion spring suspension, to which was also attached a small mirror. As the magnetic fields of the current-carrying coils interacted with the magnetic fields of the suspended permanent magnets, torque was developed causing rotation of the suspension and mirror. A collimated beam of light directed at the mirror was reflected onto a distant scale. Even a slight rotation of the suspension is seen as a linear deflection of the spot of light. Galvanometers of this type were used more to detect very small currents (on the order of nA), rather than to measure large currents accurately. Such galvanometers found a major application as null detectors for DC Wheatstone and Kelvin bridges and for DC potentiometers. In 1836, Sturgeon employed a coil suspended in the field of a permanent magnet as a galvanometer. This design evolved, and in 1882, D'Arsonval built a moving-coil, permanent magnet, mirror galvanometer with a torsion spring suspension (see Figure 8.19). Notable in the D'Arsonval design was the use of a soft iron cylinder inside the rectangular coil to concentrate the magnetic flux perpendicular to the sides of the coil. In 1888, Weston developed a similar permanent-magnet, moving-coil microammeter movement in which the coil was suspended

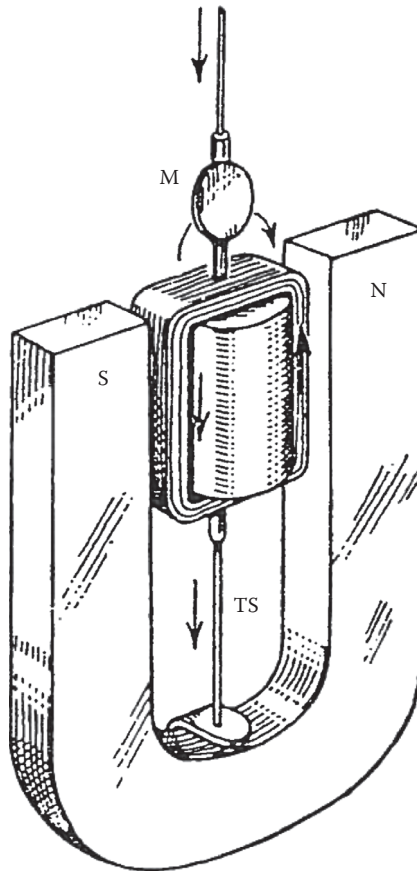


FIGURE 8.19

Diagram of D'Arsonval's mirror galvanometer. The coil was suspended by upper and lower torsion springs. The torsion springs also carried the current to be measured to and from the coil. *Note:* M = mirror. TS = torsion spring.

by pivot bearings and the restoring torque was provided by a pair of helical springs. A pointer was attached to the moving coil. The Weston version of the D'Arsonval meter is what is widely used today as the universal, DC microammeter movement, which has applications in all analog multimeters and in most instruments with analog readouts. D'Arsonval-type DC microammeters using torsion springs, called taut-band meters, are also widely used because of their ruggedness; they have no jeweled bearings that can be damaged by mechanical shocks.

8.4.1 Electromechanical DC Ammeters

Analog DC ammeters are generally of two types, the D'Arsonval permanent-magnet moving-coil design or the electrodynamic movement. Measurement of currents in the range of 1–10,000 A with either dynamometer or D'Arsonval meter movements generally requires an external shunt resistance. A shunt is a precision resistor of very low value designed so that by Ohm's law, when the full-scale current for the meter is flowing, the voltage drop across the resistor and the meter is 50 mV. The lead resistances to the meter

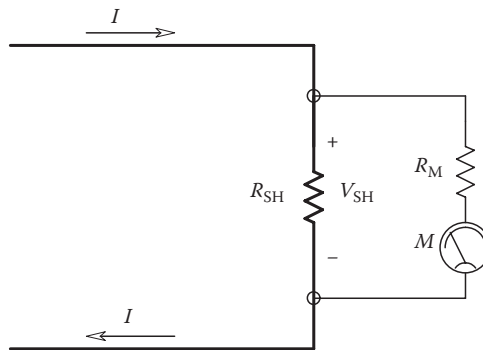


FIGURE 8.20
Circuit of a shunt ammeter. The meter is a DC microammeter.

are generally not important when measuring I_{FS} in excess of 10 A. In cases where I_{FS} is less than 10 A, and an insensitive dynamometer meter movement is used, the meter lead resistance (from shunt to meter) may be important, and a specified set of meter leads with known resistance must be used to obtain readings with rated accuracy. The circuit for an ammeter using a shunt is shown in Figure 8.20.

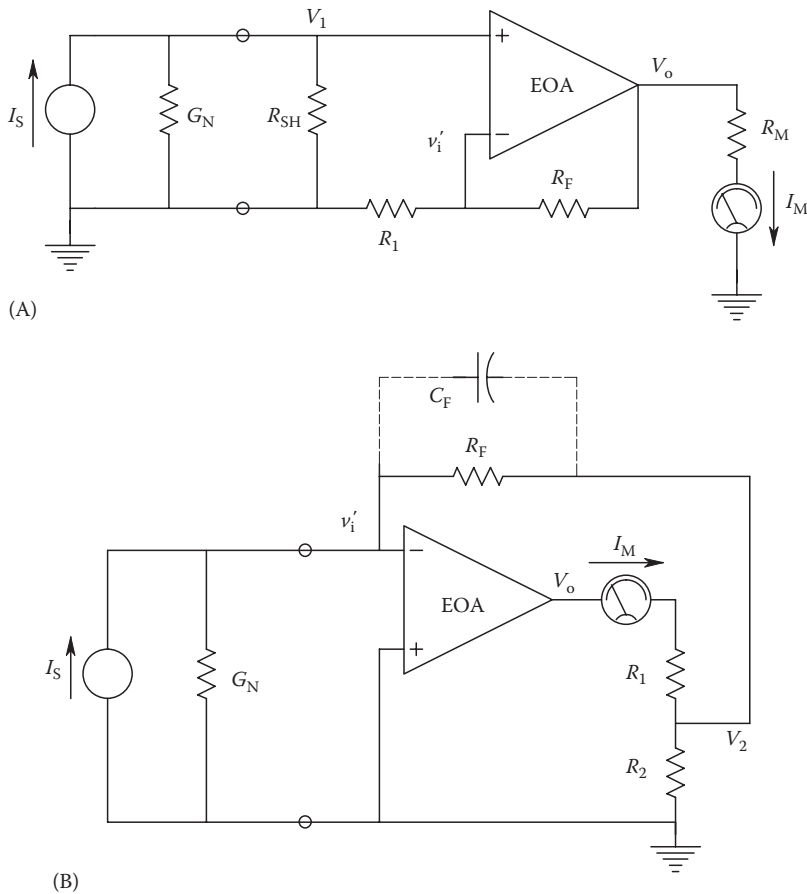
Analog DC ammeters can be subdivided into benchtop (precision) meters and panel meters, which are generally used in instruments, systems, and communications equipment. Some D'Arsonval microammeters have been miniaturized into 1" diameter bodies for use as tuning and signal strength meters in radio equipment and as signal strength meters in tape recorders.

We note that the same basic DC meter movements that can be used for analog DC voltmeters are suitable for use as DC ammeters with appropriate shunts. When full-scale DC less than $20 \mu\text{A}$ are to be measured, it is generally necessary to employ electronic amplification, discussed in the following section.

8.4.2 Electronic DC Ammeters

To make measurements in the range from 1 fA (10^{-15} A) to $50 \mu\text{A}$, it is necessary to employ DC amplification of some sophistication. In general, current measurements in the range from 10^{-8} to 10^{-15} A are made with an *electrometer picoammeter*. Electrometer picoammeters find application in measuring the low-current outputs from photomultiplier tubes and ion chambers, as well as specialized measurements in the semiconductor area. As in the case of low-level voltage measurements, one fundamental factor limiting precision is the thermal noise from the Norton equivalent conductance seen in parallel with the DC source being measured. Noise from the electrometer amplifier also confounds resolution. Both e_{na} and i_{na} input noise root power spectrums can be significant in reducing instrument resolution, as well. In addition to problems with noise, there can be measurement errors caused by instrument DC bias current and offset voltage.

Figure 8.21 illustrates the two types of picoammeter circuits commonly used. In Figure 8.21A, we see a simple shunt picoammeter in which most of the current to be measured flows through R_{SH} . Ordinarily, $R_{SH} \ll 1/G_N$, so $V_1 = I_S R_{SH}$. V_1 is amplified by the EOA connected as a non-inverting, DC millivoltmeter. In Figure 8.21B, the DC under measurement is the input to an operational transresistor circuit, the output of which can easily be shown

**FIGURE 8.21**

(A) Circuit of an electronic DC ammeter. The op amp amplifies the voltage developed across the shunt, R_{SH} .
 (B) An op amp is used as a transresistor. The summing junction appears at virtual ground.

to be $V_o = -I_S R_F$. To obtain greater sensitivity, it would appear that R_F should be increased. There is a practical limit to the size of R_F set by its cost and accuracy. Resistors above $10^8 \Omega$ are expensive, and as their size increases, their accuracy decreases. Also, when $R_F > 10^9 \Omega$, the effects of shunt capacitance act to slow the overall response time of the operational transresistor (the parallel combination of R_F and C_F acts like an LPF).

8.4.2.1 Error Analysis of the Shunt Picoammeter

Three major sources of error in the shunt picoammeter circuit are discussed in this section. First, errors arise because of calibration errors in R_{SH} , R_1 , and R_2 . Second, errors arise because of the EOA's DC bias current, I_B , and offset voltage, V_{OS} . V_{OS} can be nulled to zero at one temperature, but V_{OS} will drift with temperature. I_B cannot be nulled in solid-state EOAs, and we must live with it; fortunately, I_B is very small for modern, IC EOAs, about 40 fA in the AD549L EOA. The VOS tempco is $5 \mu\text{V}/^\circ\text{C}$ for the AD549L. The third source of measurement error comes from amplifier noise and from Johnson (thermal) noise in associated resistors. The short-circuit input voltage noise, e_{na} , for an AD549L EOA is $4 \mu\text{V}$ peak

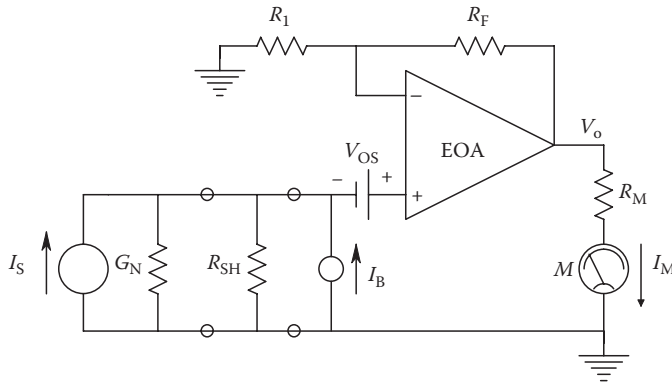


FIGURE 8.22
Circuit showing sources of DC error in the circuit in Figure 8.21A.

to peak in the 0.1–10 Hz bandwidth and 35 nV/√Hz above about 1 kHz. The equivalent input current noise is 0.36 fA ppk in the 0.1–10 Hz bandwidth and can be considered to be 0.11 fA/√Hz white noise from 0.1 Hz to 100 MHz.

Figure 8.22 illustrates the DC sources that can cause errors in reading the unknown DC, I_s . Let us assume that $R_N \gg R_{SH}$. The amplifier gain for the non-inverting input is $K_V = (1 + R_F/R_1)$. Now, the amplifier output voltage due to I_s is just

$$V_{OM} = I_s R_{SH} K_V. \tag{8.56}$$

The output voltage due to V_{OS} , I_B , and I'_B is

$$V_{OE} = V_{OS} K_V + I_B R_{SH} K_V + I'_B R_F. \tag{8.57}$$

Thus, we can write an expression for the DC output signal to DC output voltage error ratio:

$$\frac{V_{OM}}{V_{OE}} = \frac{I_s}{V_{OS}/R_{SH} + I_B + I'_B R_F / (R_{SH} K_V)}. \tag{8.58}$$

Clearly, the meter should be zeroed with input open-circuited to null out V_{OE} .

To examine the effect of resistor thermal noise and amplifier noises on the measurement of the DC I_s , we calculate the shunt picoammeter’s SNR. This calculation is made more difficult by the fact that the $e_{na}(f)$ root spectrum is in its $1/f$ region. Thus, to evaluate the total RMS output noise, we must assume that all noise sources are independent and uncorrelated, and we must integrate the $1/f$ spectrum over the nominal frequency range used to make DC measurements, from 0.1 to 10 Hz. From the published $e_{na}(f)$ curve on the AD549L EOA data sheet, we can write the short-circuit input, noise voltage power density spectrum as

$$e_{na}^2(f) = \eta + b/f = 1.2250 \times 10^{-15} + (4.0833 \times 10^{-17})/f \text{ MS V/Hz}. \tag{8.59}$$

In a JFET input amplifier, $i_{na}(f)$ generally has a flat PDS at low frequencies and can be considered to be white.

Now, the total mean-squared noise output voltage can be found in the 0.1–10 Hz bandwidth:

$$N_O = K_V^2 \int_{0.1}^{10} e_{na}^2(f) df + K_V^2 4kT \frac{R_{SH}R_N}{R_{SH} + R_N} B + K_V^2 \frac{(R_{SH}R_N)^2}{(R_{SH} + R_N)^2} i_{na}^2 B + 4kTR_1 \left(\frac{R_F}{R_1} \right)^2 B + 4kTR_F B \text{ MS V} \quad (8.60)$$

where $B = f_H - f_L = 10 - 0.1 = 9.9$ Hz, $i_{na} = 0.11$ fA/ $\sqrt{\text{Hz}}$, and $K_V = (1 + R_F/R_1)$. Note that we have included Johnson noise from R_1 and R_F .

The mean-squared output signal is

$$V_{OS}^2 = I_S^2 \frac{(R_{SH}R_N)^2}{(R_{SH} + R_N)^2} K_V^2 \text{ MS V.} \quad (8.61)$$

The MS SNR can be written algebraically as

SNR_{OUT}

$$= \frac{I_S^2/B}{\eta(G_{SH} + G_N)^2 + (b/B) \ln(f_H/f_L)(G_{SH} + G_N)^2 + i_{na}^2 + 4kT(G_{SH} + G_N) + 4kTR_F K_V^{-1}(G_{SH} + G_N)^2}. \quad (8.62)$$

A numerical value for the RMS SNR can be found from the expressions earlier. For simplicity, assume $R_N \gg R_{SH}$, $R_{SH} = 10^6 \Omega$, $R_F = 100$ K, $R_1 = 1$ K, and $T = 300$ K. The RMS SNR is found to be

$$\text{SNR}_{(\text{RMS})} = \frac{I_S}{4.2184 \times 10^{-13}} \text{ A/rms A.} \quad (8.63)$$

Thus, noise limits the resolvable I_S to ca. $3 \times 4.22 \times 10^{-13}$ A for the parameters given earlier, creating random errors.

8.4.2.2 Error Analysis of the Feedback Picoammeter

As in the case of the shunt picoammeter, the EOA's offset voltage and bias current cause uncertainty in the measurement of I_S . From the equivalent circuit of Figure 8.23, we can write an expression for the worst-case, DC output signal to DC output error ratio:

$$\frac{V_{OM}}{V_{OE}} = \frac{I_S}{I_B \pm V_{OS}(G_N + G_F)}. \quad (8.64)$$

Here again, the systematic DC error due to V_{OS} and I_B can generally be observed with the input open-circuited and compensated for.

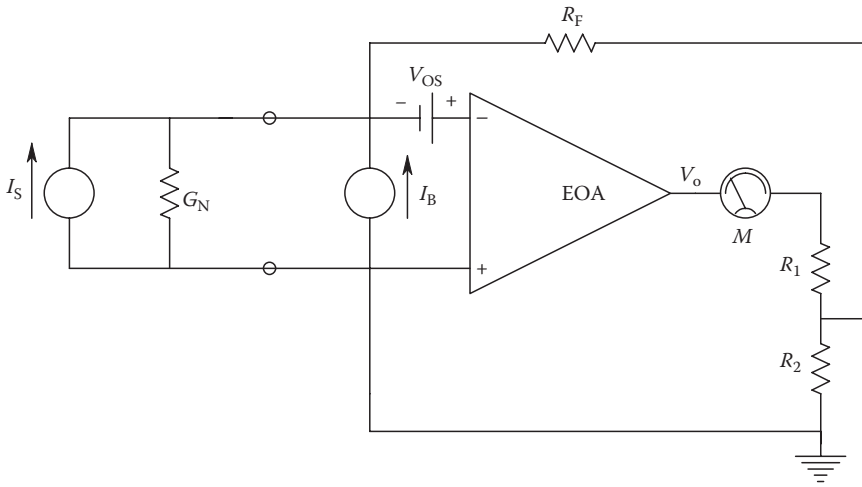


FIGURE 8.23
Circuit showing sources of DC error in the feedback picoammeter of Figure 8.21B.

The error due to random noise includes the thermal noise from R_N and R_F , as well as the amplifier's e_{na} and i_{na} . The output mean-squared SNR can be written as

$$SNR_O = \frac{I_s^2 R_F^2}{\left[i_{na}^2 + 4kT(G_N + G_F) \right] R_F^2 B + \left[\eta B + b \ln(f_H/f_L) \right] (R_N + R_F)^2 / R_N^2} \tag{8.65}$$

Equation 8.65 can be reduced to a more interesting form:

$$SNR_O = \frac{I_s^2}{B \left\{ i_{na}^2 + 4kT(G_N + G_F) + \left[\eta + (b/B) \ln(f_H/f_L) \right] (G_N + G_F)^2 \right\}} \tag{8.66}$$

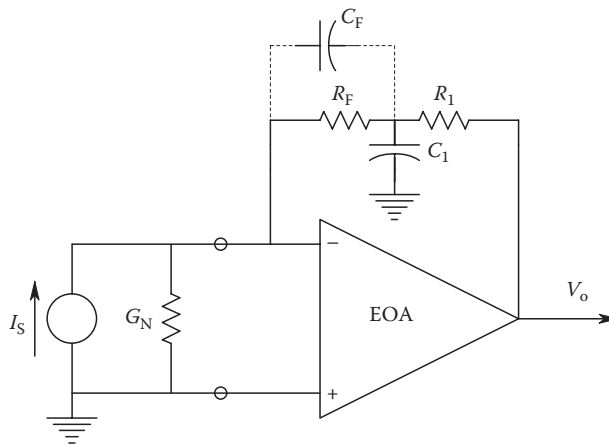
It is instructive to evaluate the MS SNR_O using Equation 8.66 earlier for typical system parameters of $\eta = 1.2250 \times 10^{-15}$ MS V/Hz, $b = 4.0833 \times 10^{-17}$ MS V, $R_F = 10^9 \Omega$, $f_H = 10$ Hz, $f_L = 0.1$ Hz, $i_{na} = 0.11$ fA/ $\sqrt{\text{Hz}}$, $4kT = 1.656 \times 10^{-20}$, and $R_N = 10^6 \Omega$. We find

$$SNR_O = \frac{I_s^2}{1.643 \times 10^{-25}} \tag{8.67}$$

So for reliable, low-noise measurements, the DC I_s should be over three times the equivalent input RMS noise current, or $I_s > 1.216 \times 10^{-12}$ A. See the home problems at the end of this chapter for exercises in calculating noise and DC drift limits to electronic DC picoammeter accuracy.

A problem that often arises in the design of feedback picoammeters is the loss of HF response due to the parasitic shunt capacitance, which appears in parallel with the very large R_F used. For example, if $R_F = 10^{11} \Omega$, a 1 pF stray capacitance in parallel with R_F will produce a transresistance given by

$$\frac{V_o}{I_s}(s) = \frac{-R_F}{sR_F C_F + 1} \tag{8.68}$$

**FIGURE 8.24**

A circuit to speed up the response of the feedback picoammeter.

The time constant is $\tau = 0.1$ s in this case. This would give little problem in the measurement of DC but would be unsuitable for the measurement of AC. A circuit to speed up the response of the feedback picoammeter is shown in Figure 8.24. It is posed as a home problem at the end of the chapter to find the condition on R_1 , C_1 that will cancel the low-pass effect of C_F in parallel with R_F .

8.4.3 Magneto-Optic Current Sensors

The *Faraday magneto-optic effect* has been used in several designs for the measurement of high currents in power distribution systems. Typically, currents can range from tens of amperes to kiloamperes, and the conductors are generally at kilovolts above ground potential, rendering the use of a shunt and millivoltmeter quite impractical. The Faraday magneto-optic effect makes use of the fact that a current-carrying conductor is surrounded by a solenoidal magnetic field whose magnitude is proportional to the current. The magnetic field, in turn, induces changes in the optical polarization properties of glass or fiber-optic cable on which it acts. Thus, by measuring the amount of optical rotation induced on linear polarized light emerging from the optical path in the \mathbf{B} field, we can indirectly sense the current in the conductor. Details of several Faraday magneto-optic effect current sensors are described in Section 6.6.1.

8.5 AC Voltage Measurements

Any discussion of AC voltage measurements necessarily must cover a wide range of applications in which the amplitude and frequency of the voltage to be measured are of critical importance. AC voltage measurements can be divided into a number of frequency ranges: audio (including powerline frequencies), ranging from around 10 Hz to about 30 kHz, and radio frequencies, including the following bands—LF 30–300 kHz, medium

frequency (MF) 300–3000 kHz, HF 3–30 MHz, very high frequency (VHF) 30–300 MHz, ultrahigh frequency (UHF) 300–3000 MHz, super high frequency (SHF) 3–30 GHz, and extra high frequency (EHF) 30–300 GHz. Resolvable amplitudes range from nanovolts to hundreds of kilovolts. AC voltages are generally sinusoidal, but situations arise in which one must measure other periodic waveforms including square waves, triangle and sawtooth waves, and pulse trains. We also need to measure completely random (noise) voltages with various spectral contents and amplitudes. In the following sections, we will first describe various nonelectronic, electromechanical AC voltmeters and then discuss the designs for electronic AC voltmeters.

8.5.1 Electromechanical AC Voltmeters

Most electromechanical AC voltmeters are designed to work at powerline or low audio frequencies and at voltages ranging from about 1 V to tens of kilovolts. As in the case of DC voltmeters, basic electromechanical AC voltmeter design uses a series resistor to limit meter current to give full-scale deflection at full-scale applied voltage at the meter terminals. Almost all AC voltmeters have their scales calibrated in RMS volts of a sine wave, regardless of the deflection mechanism of the meter movement. Only true RMS-reading voltmeters will remain calibrated when a periodic voltage waveform other than a sine wave is the input to the meter. There are five common types of AC, powerline frequency voltmeter movements, some of which we have already described in Section 8.2.1. These include the dynamometer movement, the capacitor (electrostatic) movement, the iron vane movement, the rectifier D'Arsonval design, and the VTC D'Arsonval design. Of these AC voltmeters, only the thermocouple meter is useful at radio frequencies, up to 50 MHz in some designs.

We defined voltmeter sensitivity for DC voltmeters as the reciprocal of the current required for full-scale meter deflection. A similar definition is useful for AC voltmeters, except it is the total meter impedance that is found by multiplying the sensitivity in ohms per volt by the meter's full-scale voltage.

The dynamometer AC voltmeter is generally a large, insensitive, benchtop instrument used to measure sinusoidal voltages with frequencies from about 20 to 133 Hz. Accuracy is typically 0.2% of full-scale voltage, and meter deflection, assuming a linear spring whose torque is proportional to the deflection angle, and a mutual inductance, which varies linearly with deflection angle, are proportional to the mean-squared current in the coils. Typical dynamometer voltmeters can be found with full-scale ranges from 30 to 600 V rms. About 73.34 mA is required for full-scale deflection of the venerable GE type P-3 electro-dynamometer voltmeter. See Figure 8.8 for a schematic diagram of a dynamometer voltmeter.

Capacitor voltmeters are also large, insensitive, benchtop instruments that can be used to measure sinusoidal voltages with frequencies from 15 Hz to 300 kHz. Accuracy is typically 1% of full-scale voltage, and meter deflection, assuming linear spring torque with deflection angle and a linear increase of meter capacitance with deflection angle, is proportional to the mean-squared voltage across the meter. Typical electrostatic voltmeters can be found with full-scale deflections of 300 V rms to 50 kV rms. It should be noted that the current at a given deflection of an electrostatic meter is given by Ohm's law, that is, $I_M = V_M 2\pi f C_M$, where the meter capacitance, C_M , is itself an increasing function of meter deflection. The innards of an electrostatic voltmeter are shown in Figure 8.7. Note that while the deflection angle of the movement may be proportional to the MS V, the scale is calibrated in RMS V.

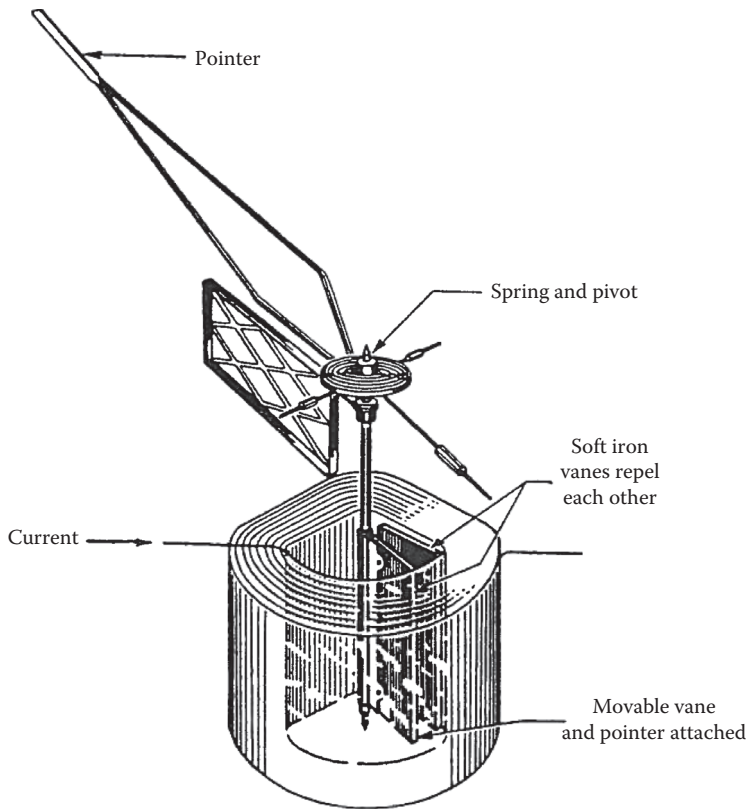


FIGURE 8.25
Diagram of an iron vane, AC meter movement of the repulsion type.

Iron vane voltmeters are found as large, insensitive, benchtop instruments that can be used to measure sinusoidal voltages with frequencies from about 15 to 133 Hz, with accuracies of about 0.5% of full-scale voltage. Iron vane AC panel voltmeters are also found. Iron vane meters can be of the repulsion type, as shown in Figure 8.25, or be of the attraction type, as shown in Figure 8.26. Typical meter efficiency is about $29 \Omega/V$. This means that the meter draws about 34 mA at full-scale deflection. Because of their inefficiency, iron vane voltmeters are generally used to monitor powerline voltages and currents. Typical full-scale voltages for these meters range from 8 to 750 V rms, depending on the application. Iron vane meter scales are generally nonlinear because of the geometry of the vanes and/or coils; the lower quarter of the scale is generally crowded. They are an obsolete design.

Rectifier/D'Arsonval AC voltmeters generally find application in benchtop and pocket volt-ohm-milliammeters (multimeters). Full-scale voltages typically range from 3 V rms to over 1 kV rms. The circuit of a rectifier/D'Arsonval AC voltmeter is extremely simple, as shown in Figure 8.27. A series resistor, R_1 , limits the meter current for a given full-scale voltage. A series capacitor, C_1 , blocks DC from the meter circuit. A bridge rectifier converts AC in R_1 to a full-wave rectified current with a DC average value. The D'Arsonval microammeter movement deflects proportional to the average current through its coil; hence, deflection of the rectifier/D'Arsonval meter is proportional to the full-wave rectified

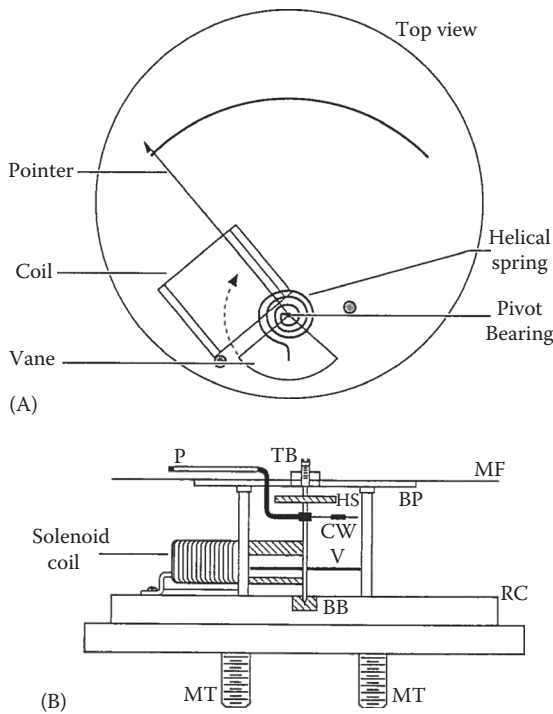


FIGURE 8.26 (A) Face view of an iron vane, AC meter of the attraction type. (B) Side view of the attraction-type, AC, iron vane meter. Magnetic torque pulls the vane into the current-carrying coil.

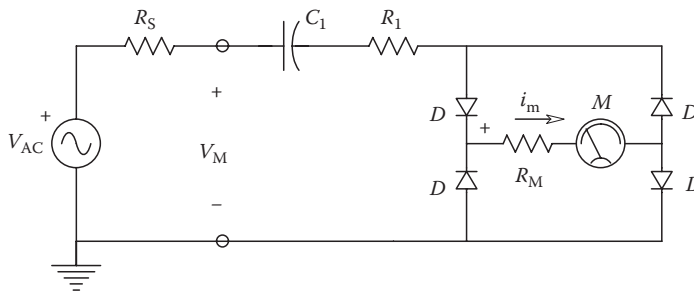


FIGURE 8.27 Schematic of a bridge rectifier AC voltmeter. A DC D'Arsonval microammeter (M) is used. Silicon pn junction diodes (D) are used.

average value of the input voltage. AC voltmeter calibration is, of course, in terms of the rms of a sine wave input voltage. The D'Arsonval meter movement acts as a mechanical LPF (cf. Equation 8.12), its deflection being proportional to the average current through its coil. The average value of a full-wave rectified sine wave is easily shown to be

$$\overline{i_M} = \overline{I_{M(PK)} \sin(\omega t)} = \frac{2}{\pi} I_{M(PK)} = \frac{2\sqrt{2}}{\pi} I_{M(RMS)}. \quad (8.69)$$

If we assume at high input voltages that the rectifier diodes are ideal (i.e., they have zero forward resistance and voltage drop and infinite reverse resistance), then we may write for full-scale meter deflection the following:

$$I_{M(\text{FS})} = \frac{2\sqrt{2}V_{S(\text{RMS,FS})}}{\pi(R_1 + R_M)}. \quad (8.70)$$

For a D'Arsonval microammeter with resistance R_M and full-scale current, $I_{M(\text{FS})}$, we may solve for the R_1 required to make an AC voltmeter with a full-scale voltage reading of $V_{S(\text{RMS,FS})}$, assuming ideal diodes:

$$R_1 = \frac{2\sqrt{2}V_{S(\text{RMS,FS})}}{\pi I_{M(\text{FS})}} - R_M. \quad (8.71)$$

The AC, full-wave rectifier voltmeter's efficiency can be shown to be

$$\eta_{\text{AC}} = \frac{1}{I_{M(\text{RMS})}} = \frac{R_1 + R_M}{V_{S(\text{RMS,FS})}} = \frac{(2\sqrt{2}/\pi)(V_{S(\text{RMS,FS})})/I_{M(\text{FS})}}{V_{S(\text{RMS,FS})}} = 0.9 \eta_{\text{DC}}. \quad (8.72)$$

These relations are generally valid for AC input voltages above 10 V rms. The scales of meters reading below 10 V rms full scale generally must be corrected for the finite forward and reverse resistances of the real diodes used in the meter rectifier circuit. Also, practical diodes exhibit a voltage-dependent depletion capacitance that is effectively in parallel with their reverse resistance. Thus, at high audio frequencies, considerable reverse diode current can flow, decreasing the average current through the meter movement and causing frequency-dependent low meter readings. At powerline frequencies and low audio frequencies, diode capacitance is generally not a problem.

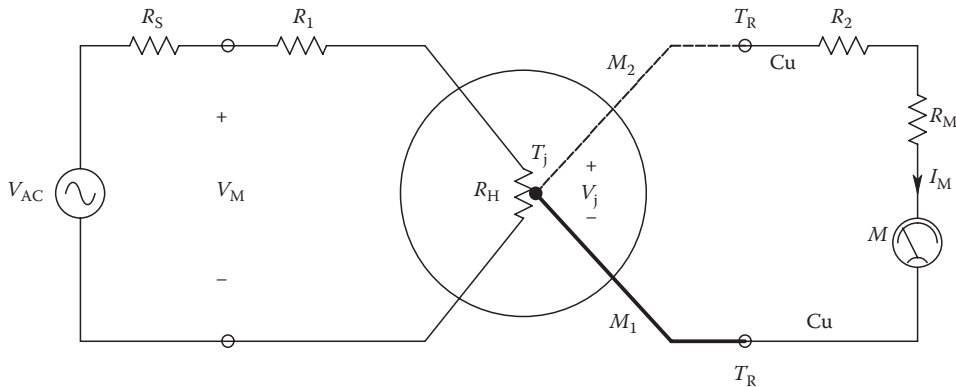
A problem commonly encountered in using AC voltmeters of this sort is how do they respond to nonsinusoidal voltages, such as square waves and triangle waves. To analyze this problem, we note first that any average or DC component in V_S is removed from the voltmeter input current, I_S , by the series capacitor, C_1 . The average (DC) current through the D'Arsonval movement can be written as

$$\overline{i_M(t)} = \frac{\overline{|v_S(t)|}}{R_1 + R_M} = \frac{\overline{|v_S(t)|}}{(2\sqrt{2}/\pi)(V_{S(\text{RMS,FS})}/I_{M(\text{FS})})}. \quad (8.73)$$

The actual meter reading can be expressed as a fraction of the full-scale voltage:

$$V_{\text{READ}} = \frac{\overline{i_M(t)}}{I_{M(\text{FS})}} V_{S(\text{RMS,FS})} = \frac{\overline{|v_S(t)|}}{2\sqrt{2}/\pi} = 1.111 \overline{|v_S(t)|}. \quad (8.74)$$

VTC/D'Arsonval AC voltmeters are the only type of electromechanical AC voltmeter that will remain calibrated for input voltage frequencies from 15–20 Hz to

**FIGURE 8.28**

Schematic of a VTC AC voltmeter. A DC, D'Arsonval microammeter (M) movement is used. Meter deflection is proportional to the mean-squared heater current.

about 50 MHz. The reading of a VTC AC voltmeter is proportional to the true mean square of the voltage being measured.

A schematic of a simple, VTC AC voltmeter is shown in Figure 8.28. The input circuit consists of a series resistor, R_1 , and a heater filament with resistance, R_H . For example, R_H for a Western Electric type 20 K VTC is given as $750\ \Omega$. The maximum safe heater current is given as 5 mA rms. A Western Electric type 20D VTC has a heater resistance of $35\ \Omega$ and a safe current of 16 mA rms. An overcurrent of as little as 50% is likely to burn out the heater wire, so extreme care must be taken to prevent over-range inputs when using VTC/D'Arsonval AC voltmeters. The maximum rated EMF for the Western Electric VTCs mentioned earlier is 15 mV (DC), and the thermocouple resistance is $12\ \Omega$. If the thermocouple is connected to a $0\text{--}20\ \mu\text{A}$ D'Arsonval meter, and the total series resistance in the circuit ($R_{TC} + R_M + R_2$) equals $750\ \Omega$, then the meter will read full scale when the heater current is at its rated maximum (3.6 mA rms in the type 20 K VTC's heater).

The DC from the thermocouple, hence meter deflection, is proportional to the mean-squared heater current (or heater voltage). This is easy to see from Joule's law; the thermocouple EMF is proportional to heater temperature. Heater temperature is proportional to the electrical power dissipated in the heater, and the power, in turn, is proportional to the mean-squared heater current times the heater resistance. In order that the meter be calibrated in RMS volts, a square root scale is printed on the meter face, rather than the linear one commonly found on D'Arsonval meters used in DC applications. Thus, the lower end of the scale is crowded, and calibrations are seldom given below 20 V on a 100 VFS meter. Laboratory quality thermocouple meters commonly have accuracies of 0.5% of full-scale voltage.

8.5.2 Analog Electronic AC Voltmeters

In this section, we will review the more common types of analog AC voltmeters, that is, AC voltmeters, which make use of analog electronic signal conditioning and which use analog output meters. AC voltages can be measured effectively from the microvolt range to tens of kV. Frequencies can range from power line to tens of GHz. No one type of AC

voltmeter is effective over the ranges of frequency and voltage cited. The fundamental limit to the resolution of small AC voltages is noise: thermal noise from resistors, noise from amplifiers, and coherent noise picked up from the environment in spite of our best efforts to shield and guard the circuits being used.

8.5.2.1 AC Amplifier–Rectifier AC Voltmeters

In the first class of AC electronic voltmeter, the AC voltage under measurement is conditioned by a flat band-pass, RC amplifier of accurately known gain. The amplified AC voltage is then converted to DC that causes meter indication of the AC voltage. Most electronic AC analog voltmeters make use of the common D'Arsonval movement as the output indicator, in conjunction either with some type of electronic rectifier to convert amplified AC voltage to a DC or with a VTC to convert the mean-squared value of the amplified AC voltage to a DC to drive the analog meter. A block diagram of the AC amplifier–rectifier type of AC voltmeter is shown in Figure 8.29. The amplifier's gain is generally fixed and large (e.g., 3.33×10^3), and a precision, frequency-compensated attenuator is used to reduce input signals that are above the minimum full-scale voltage (e.g., 3 mV) to a level that will not saturate the amplifier. Input resistance is thus set by the attenuator and is typically 10 or 11 M Ω . Because the amplifier's AC output is rectified to drive the D'Arsonval meter, the rectifier's frequency response generally sets the upper useful frequency of this type of meter. Such meters are typically used at audio frequencies, and some versions are useful up to 10 MHz. To overcome the nonlinearity of the rectifier diodes at low output voltage levels, it is possible to use an operational rectifier circuit, shown in Figure 8.30, instead

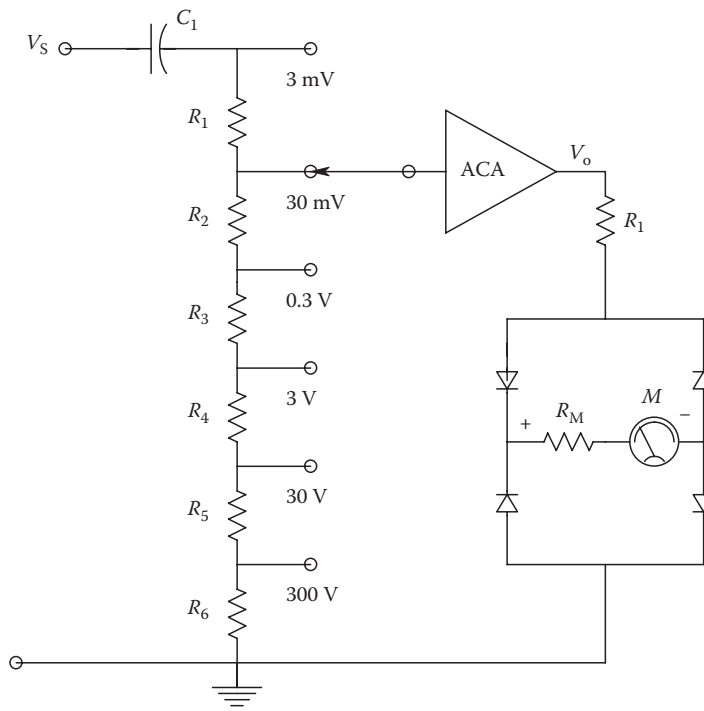
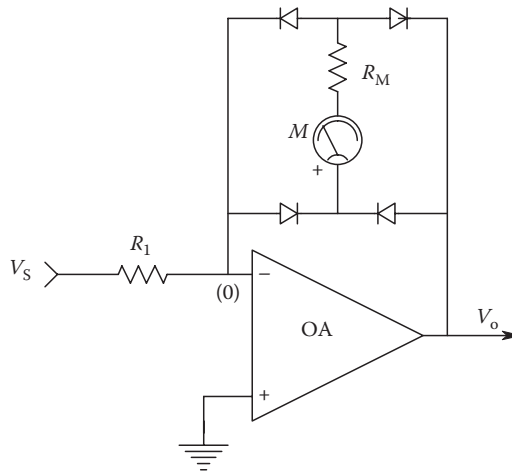


FIGURE 8.29
Schematic of an AC amplifier–rectifier-type AC voltmeter.

**FIGURE 8.30**

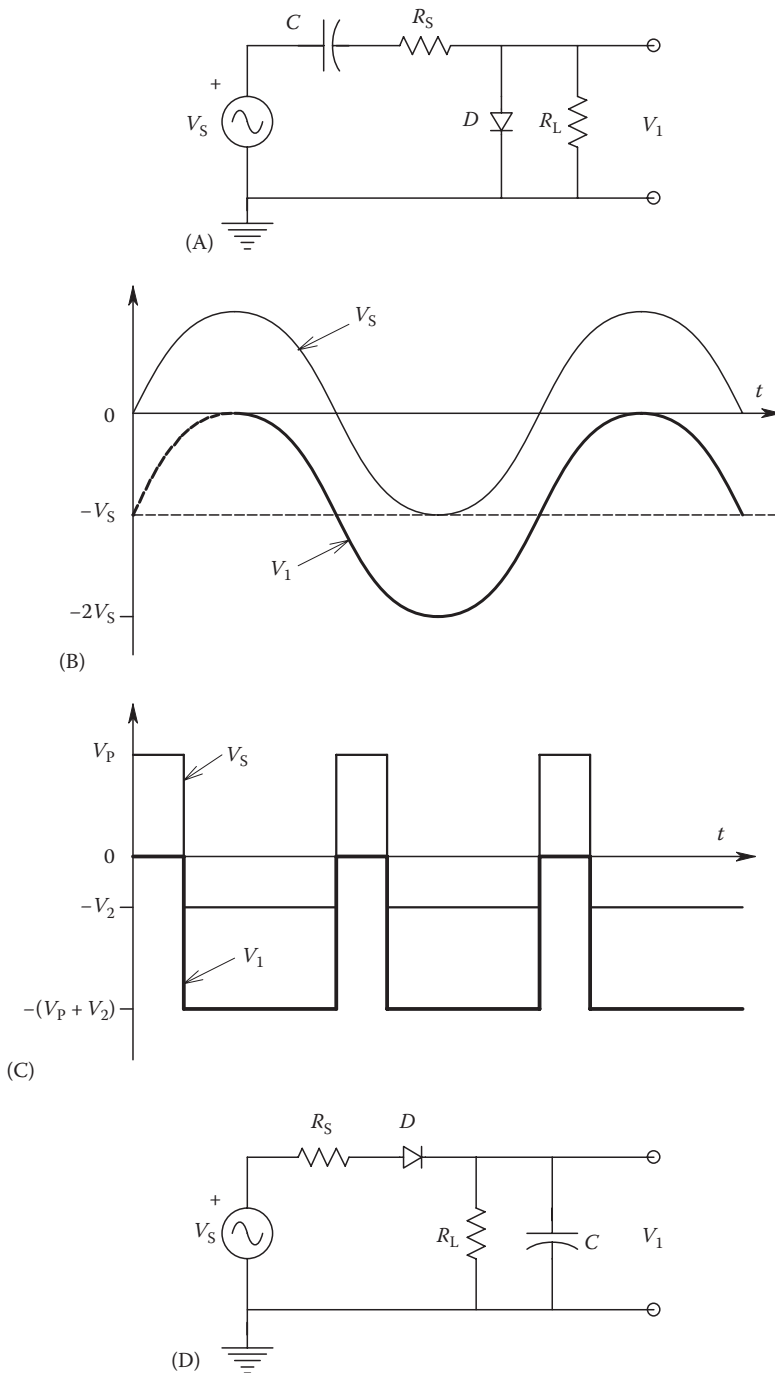
A simple operational rectifier circuit that provides linear, near-ideal rectification of the audio-frequency AC input voltage, even at low (millivolt) values of V_o .

of a passive diode bridge. Such a circuit will improve meter linearity at the low end of the meter scale; however, the reverse capacitance of the diodes and the finite slew rate and gain–bandwidth product of the OA provide fundamental limitations to the operating frequency of the operational rectifier. Graeme (1974) gives several circuit modifications for operational rectifiers that improve their performance.

8.5.2.2 Peak-Reading Electronic AC Voltmeters

The second type of electronic AC voltmeter (peak reading) uses a low-capacitance diode to half-wave rectify the AC voltage under measurement. The rectified AC is smoothed by a capacitor LPF and then is amplified by a drift-free, DC amplifier with known gain, such as a chopper-stabilized or commutating auto-zero circuit. The DC output of the amplifier causes the deflection of the voltmeter's D'Arsonval meter. The rectifier used in the peak-reading AC voltmeter can be a vacuum, thermionic diode, although low-capacitance semiconductor diodes can also be used. The advantage of this design lies in the measurement of RF voltages. Because the AC voltage under measurement is converted to DC at the front end, there is no need for an expensive, high-gain, high-bandwidth linear amplifier to condition the AC voltage. There is need for a low-gain, drift-free, DC amplifier, however. Because of the inherent nonlinearity of semiconductor and thermionic diodes at low voltages, peak-reading AC voltmeters are generally used for AC input voltages above 1 Vpk. Below 1 Vpk, special scale calibration must be used to compensate for diode nonlinearity, which can be shown to be square law for input voltages below 50 mV (Oliver and Cage 1971). Most peak-reading AC voltmeters have input ranges of 1 Vpk full scale to 300 Vpk full scale. Their frequency response ranges from power line to over 300 MHz. Indeed, some special microwave diodes have been used to measure peak voltages in waveguides at frequencies up to 40 GHz.

There are two common forms of peak-detecting AC voltmeter illustrated in Figure 8.31. In Figure 8.31A, we see a DC peak detector in which the capacitor charges through the diode to a DC voltage equal to the positive peak value of $v_s(t)$. Thus, the output voltage across the diode is $v_1(t) = [v_s(t) - V_{PK+}]$. For the case of a sine wave input with zero mean, the

**FIGURE 8.31**

(A) Schematic of a diode, peak-detector AC voltmeter. D is an ideal diode. (B) SS voltage waveforms in the peak detector given a sine wave input, V_S . (C) SS waveforms in the peak detector given a pulse train input. (D) A second type of peak-reading AC voltmeter. D is an ideal diode.

average voltage across the diode is $v_1 = V_{S(PK)}$. This DC level is recovered by low-pass filtering, amplified, and used to drive the D'Arsonval indicating meter. Thus, meter deflection is proportional to the peak value of the zero mean sine wave input. In more general terms, for an arbitrary input $v_s(t)$, the mean value of $v_1(t)$ can be shown to be $v_1(t) = -(+peak\ v_s - \bar{v}_s)$.

To demonstrate this property, consider the pulse train waveform in Figure 8.31C. Its peak value is V_p and its average value is

$$\bar{V}_S = \frac{V_P T_1 - V_2 T_2}{T_1 + T_2} \tag{8.75}$$

So the average value of $v_1(t)$ is the negative peak-above-average value of $v_s(t)$:

$$\bar{V}_1 = - \left[V_P - \frac{V_P T_1 - V_2 T_2}{T_1 - T_2} \right] = - \frac{T_2 (V_P + V_2)}{T_1 + T_2} \tag{8.76}$$

Peak-reading voltmeters are generally calibrated in terms of the rms of a sinusoidal input. The input impedance of the diode probe of a peak-reading voltmeter can be shown to be $Z_{IN} = R_L/3$, where in general $R_S \ll R_L$ (Angelo 1969).

A second type of peak-reading electronic AC voltmeter is shown in Figure 8.31D. In this simple circuit, the capacitor acts as the LPF. The capacitor charges up to the positive peak value of $v_s(t)$. Hence, the average output of the detector is $\bar{v}_1(t) = V_{S(PK)}$. Charge leaks off C through R_L with a time constant adjusted to be about 1/2 s. The input impedance of the peak rectifier can be shown to be $Z_{IN} = R_L/2$ at frequencies, where $1/\omega C \ll R_L$ (Angelo 1969). A peak-to-peak reading AC voltmeter can be made from a positive and a negative peak rectifier connected to a DC differential amplifier, as shown in Figure 8.32.

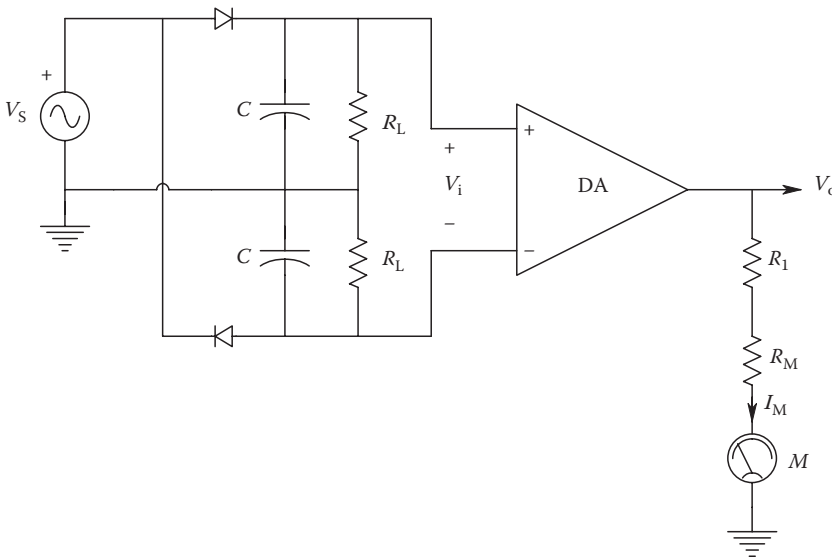


FIGURE 8.32

Schematic of a peak-to-peak reading AC voltmeter using a differential amplifier (DA). A DC, D'Arsonval microammeter (M) movement is used.

8.5.2.3 True RMS AC Voltmeter of the Feedback Type

The circuit for this AC voltmeter is shown in Figure 8.33. In the heart of the circuit are two, matched VTCs. The heater current of the input thermocouple is derived by AC amplification (or attenuation) of the AC input signal, V_s . The heater current of the feedback thermocouple comes from the direct current through the D'Arsonval meter movement. We will show in the following texts that the DC meter current is proportional to the true RMS value of $v_s(t)$ when certain conditions on circuit parameters are met.

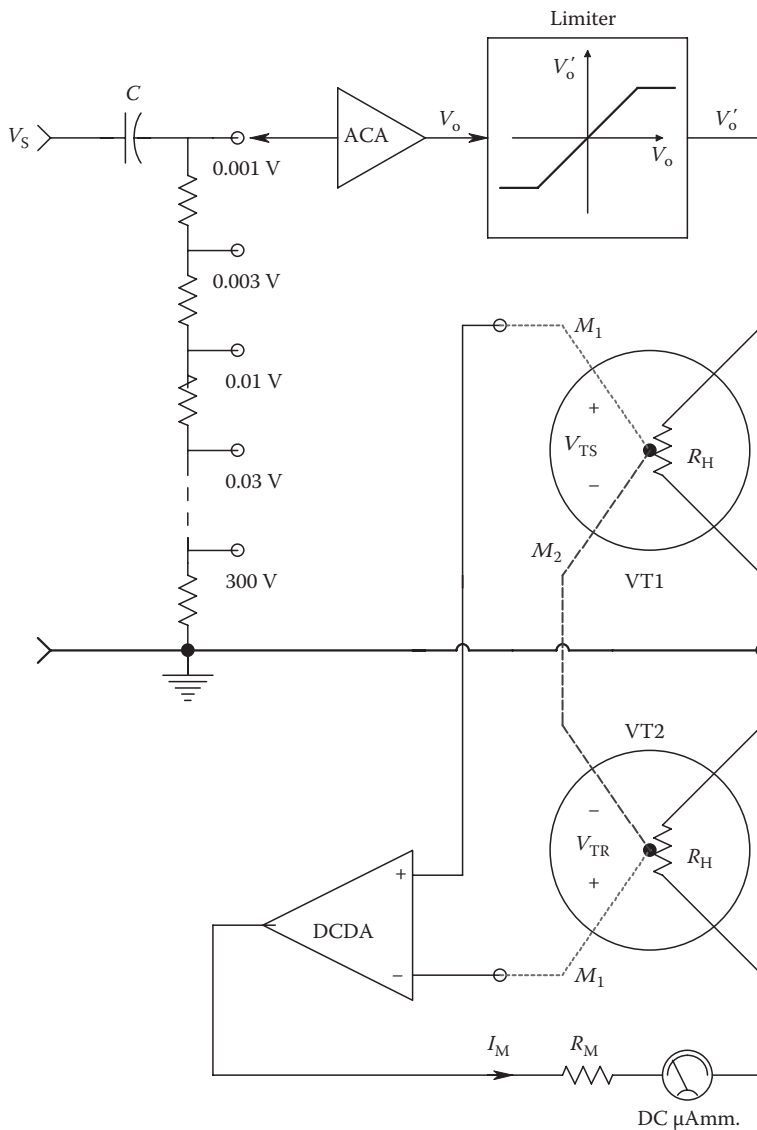


FIGURE 8.33

Simplified schematic of a feedback type, true RMS, AC voltmeter. The feedback causes $V_{TS} - V_{TR} \rightarrow 0$. Thus, the DC I_M is equal to the RMS AC in R_H of VT1.

After attenuation and/or amplification by the linear AC amplifier, an AC signal proportional to $v_s(t)$, v_{HS} , is applied to the input thermocouple heater, causing a rise in its temperature and a resultant DC output voltage, V_{TS} , from the thermocouple junction attached to the input heater. The input to the high-gain DC differential amplifier is the difference between V_{TS} and V_{TR} . V_{TD} is amplified and produces a DC output voltage, V_{OD} . The DC meter current is thus $I_M = V_{OD}/(R_o + R_M + R_H)$. From inspection of the circuit, we can write the following equations:

$$R = R_o + R_M + R_H \quad (8.77)$$

$$V_{TS} = K_T \overline{v_{HS}^2} \quad (8.78)$$

$$V_{TR} = K_T V_{HS}^2 = K_T (R_H I_M)^2. \quad (8.79)$$

In the earlier equations, R_o is the output resistance of the DC differential amplifier, K_D is its gain, R_H is the thermocouple heater resistance, R_M is the DC microammeter's resistance, and K_T is the thermocouple's transfer gain in volts/mean-squared volt. The DC meter current can be written from Ohm's law:

$$I_M = \frac{K_D (V_{TS} - V_{TR})}{R} = \frac{K_D K_T (\overline{v_{HS}^2} - V_{HS}^2)}{R}. \quad (8.80)$$

If we substitute for V_{HS} , we obtain a quadratic equation in I_M :

$$I_M^2 + I_M \frac{R}{K_D K_T R_H^2} - \frac{\overline{v_{HS}^2}}{R_H^2} = 0. \quad (8.81)$$

Solution of this equation is of the form

$$I_M = -\frac{R}{2K_D K_T R_H^2} \pm \frac{1}{2} \sqrt{\frac{R^2}{(K_D K_T R_H^2)^2} + \frac{4\overline{v_{HS}^2}}{R_H^2}}. \quad (8.82)$$

We extract the $4/R_H^2$ factor from the previous Equation 8.82 and write

$$I_M = -\frac{R}{2K_D K_T R_H^2} \pm \frac{1}{R_H} \sqrt{\frac{R_2}{(2K_D K_T R_H)^2} + \overline{v_{HS}^2}}. \quad (8.83)$$

Now, if we make $V_{HS(RMS)} \gg R/2K_D K_T R_H^2$, I_M is approximately given by

$$I_M \cong \frac{V_{HS(RMS)}}{R_H}. \quad (8.84)$$

V_{OD} is given by

$$V_{OD} = \frac{V_{HS(RMS)}}{R_H} (R_M + R_H) \text{ DC V.} \quad (8.85)$$

In a typical VTC, $I_{M(MAX)}$ through the heater is 16 mA, and the heater resistance is 35 Ω . Thus, $V_{SH(RMS, MAX)}$ is 0.56 V. Now, if we let $K_D = 10^6$, $K_T = 3 \times 10^{-2}$, $R_M = 250 \Omega$, and $R_o = 50 \Omega$, then $V_{HS(RMS, MIN)}$ would be around 1 mV. This gives the meter around a 560 to 1 dynamic range over which meter deflection is proportional to the true rms value of $v_s(t)$.

The venerable HP model 3400A true RMS voltmeter used a design similar to the one analyzed earlier.

8.5.2.4 True RMS AC Voltmeters Using the Direct Conversion Approach

In Section 2.5.4, we saw that there are several dedicated ICs available that will perform true rms conversion on AC signals. The DC output of these ICs can be conditioned and used to drive a D'Arsonval DC microammeter. The maximum frequency response for a specified % error in conversion is generally obtained for maximum AC input. For example, the AD536 rms converter has less than 1% error at frequencies up to about 140 kHz when its input is a 7 V rms sine wave and when the input is a 10 mV rms sine wave, less than 1% error at frequencies up to 6 kHz. Thus, to make a true rms voltmeter with ICs of this type, we need to adjust the input gain or attenuation such that the input to the IC is always as large as possible without saturating the circuit. One way of insuring this accuracy and dynamic range is to always make measurements on the top 1/3 of the meter scale.

For some interesting TRMS voltmeter designs and applications of TRMS converter ICs, the reader should consult the Analog Devices' *RMS to DC Conversion Application Guide*, 2nd edn., 1986.

Note that it is possible to build a TRMS analog voltmeter from a wide-bandwidth analog multiplier, an LPF, and a square root circuit. Such a circuit is capable of a wider bandwidth than obtainable with a TRMS converter IC. For example, if a four-quadrant analog multiplier such as an AD834 is used, the input signal bandwidth can range from 5 Hz to over 20 MHz for peak input voltages up to 10 V. The schematic of an analog squarer-averager-rooter TRMS circuit is shown in Figure 8.34. Again, maximum performance in terms of low error and maximum bandwidth is obtained with the use of the full dynamic range of the analog multiplier and square rooter. The analog multiplier used in the square rooter does not need outstanding bandwidth since it is operating on DC, so a high-accuracy analog multiplier such as the AD534 can be used here.

8.5.3 Measurements of Amplifier Noise Voltages, Noise Factor, and Figure

In making noise voltage measurements, one must be certain that the noise being measured is indeed uncontaminated by coherent interference or hum from sources external to the system under measurement. The best way to gain this assurance is to visualize the noise spectrum in question on a modern, sampling, fast Fourier transform (FFT) spectrum analyzer. Such instruments generally display a one-sided, root power spectrum whose units are generally rms volts/ $\sqrt{\text{hertz}}$. The presence of any spike-like peaks generally indicates the presence of contamination of the desired spectrum by coherent interference. Obviously, steps must be taken to remove extraneous coherent voltages from the desired spectrum in order to obtain valid noise measurements.

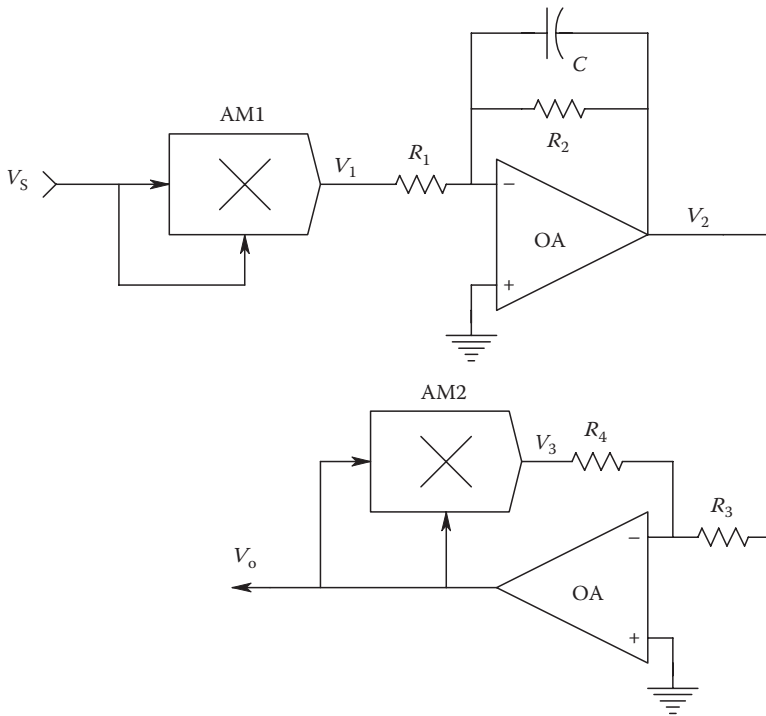


FIGURE 8.34

Schematic of a true RMS converter using an analog multiplier as a squarer, an op amp LPF averager, and a square root. A DC output ($V_o \geq 0$) is obtained.

The quantitative measurement of noise and noise figure requires the use of true rms AC voltmeters, such as we have described earlier. Noise measurements also require a knowledge of amplifier gain and the equivalent Hertz noise bandwidth of band-pass filters conditioning the noise from the noise source(s). Noise measurements can be subdivided into two categories: the use of NBPFS to characterize the power density spectrums of noise sources and broadband noise measurements over the entire bandwidth of a signal conditioning system. We are often interested in measuring the equivalent input noise sources of an amplifier, $e_{na}(f)$ and $i_{na}(f)$. $e_{na}(f)$ is the equivalent, short-circuited input noise of the amplifier. It is a root power spectrum, and its units are rms volts per root Hz. $e_{na}(f)$ generally increases at low frequencies, is flat in a midrange, and rises again at high frequencies. Ideally, in order to obtain a maximum SNR at the amplifier’s output, we would like to condition the input signal over a range of frequencies where $e_{na}(f)$ is a minimum.

Figure 8.35 illustrates a system that can be used to measure the flat, white noise portion of the $e_{na}(f)$ root spectrum. It is often necessary to follow amplifier A_1 with a second, high-gain, low-noise amplifier, A_2 , which has a broader bandwidth than A_1 . That is, the band-pass characteristics of A_1 dominate the system and determine B , the equivalent system Hz noise bandwidth. It can be shown (Northrop 2012) that as long as the midband gain of A_1 is greater than 5, any noise from A_2 is negligible in the measurement process. If we assume $e_{na}(f)$ is flat over B , then we can write an expression for the noise power density spectrum at the system output, $S_{NO}(f)$:

$$S_{NO}(f) = e_N^2 |H_{12}(j2\pi f)|^2. \tag{8.86}$$

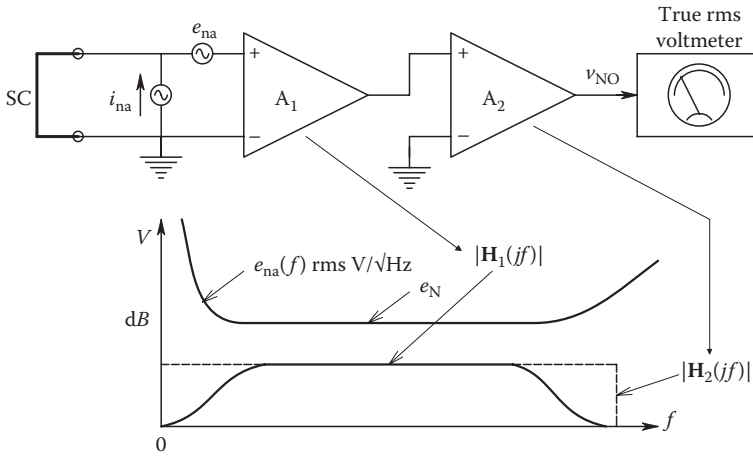


FIGURE 8.35 Circuit used to measure the white noise portion of e_{na} from A_1 . The noise passband excludes the $1/f$ region of the $e_{na}^2(f)$ power density spectrum.

The rms meter must have a flat frequency response that is wider than that of A_1 . The rms noise read by the meter is given by

$$v_{NO(RMS)} = \sqrt{\int_0^\infty e_N^2 |H_{12}(j2\pi f)|^2 df} = e_N K_{V12} \sqrt{B} \text{ rms V} \tag{8.87}$$

where

$$B = \int_0^\infty |H_{12}(j2\pi f)|^2 df / K_{V12}^2 \text{ Hz} \tag{8.88}$$

and K_{V12} is the midband gain of the cascaded amplifiers, A_1 and A_2 , and $H_{12}(j2\pi f)$ is the frequency response of the cascaded amplifiers. If $e_{na}(f)$ is not flat over B , then the rms output noise voltage is given by

$$v_{NO(RMS)} = \sqrt{\int_0^\infty e_{na}^2(f) |H_{12}(j2\pi f)|^2 df} \tag{8.89}$$

To measure $e_{na}(f)$ in its LF, $1/f$ region, we note that the e_{na} power density spectrum can be described by

$$e_{na}^2(f) = b/f + e_N^2 \text{ MS V/Hz} \tag{8.90}$$

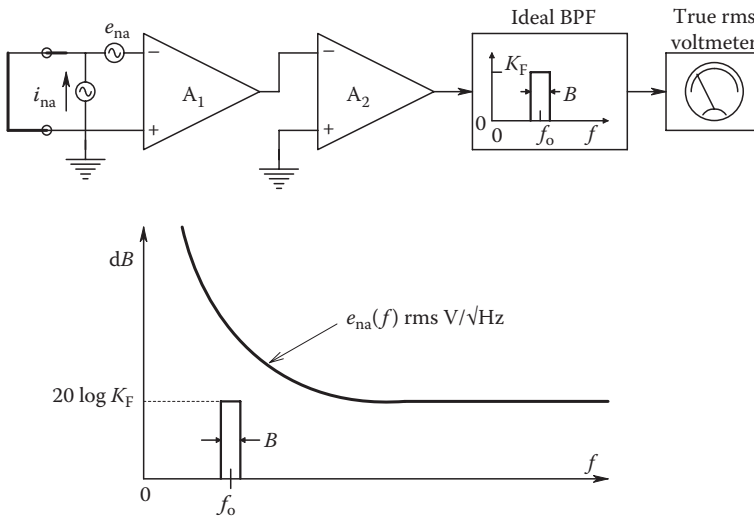


FIGURE 8.36

System used to measure the $1/f$ portion of the $e_{na}^2(f)$ power density spectrum. Note that a band-pass filter with noise bandwidth B Hz is used to select the mean-squared voltage at f_o in the desired $1/f$ region of e_{na} .

Now, we use an NBPF to examine the spectral power of $e_{na}^2(f)$. We assume that the filter has a rectangular band-pass characteristic with peak gain K_F and effective noise bandwidth B Hz. Refer to Figure 8.36. We see that the mean-squared output noise is given by

$$\begin{aligned}
 v_{NO}^2 &= \int_{f_o-B/2}^{f_o+B/2} \left(\frac{b}{f}\right) K_F^2 K_{V12}^2 df + e_N^2 K_F^2 K_{V12}^2 B \\
 &= K_F^2 K_{V12}^2 \left[b \ln\left(\frac{2Q+1}{2Q-1}\right) + e_N^2 B \right]
 \end{aligned}
 \tag{8.91}$$

where $Q \equiv f_o/B$. From Equation 8.91 earlier, we can solve for the b coefficient:

$$b = \frac{[v_{ON}^2 / (K_F^2 K_{V12}^2) - e_N^2 B]}{\ln\left(\frac{2Q+1}{2Q-1}\right)} \text{ Hz.}
 \tag{8.92}$$

In certain modern, low-noise OAs, e_N is about $3 \text{ nV}/\sqrt{\text{Hz}}$, and b is about $27 \times 10^{-18} \text{ MS V}$.

Although in many applications the amplifier's equivalent input noise current, $i_{na}(f)$, has negligible effect on the output SNR, there are instances where its value does have an effect. To measure the effect of $i_{na}(f)$, we replace the input short circuit with a resistor, R , at temperature T K. Now there are three sources of noise voltage at A_1 's input: e_{na} , the

Johnson noise from R (assumed white), and the noise $i_{na}R$. Using the procedures from Chapter 3, we can write an expression for the mean-squared output noise:

$$\overline{v_{NO}^2} = \int_0^\infty [e_{na}^3(f) + 4kTR + R^2i_{na}^2(f)] |H_{12}(j2\pi f)|^2 df \text{ MS V.} \tag{8.93}$$

Now, if we assume both e_{na} and i_{na} are white over B , Equation 8.93 can be simplified to

$$\overline{v_{NO}^2} = [e_{na}^2 + 4kTR + R^2i_{na}^2] K_{V12}^2 B \text{ MS V.} \tag{8.94}$$

Since we know K_{V12} , B , e_{na} , R , T , and B , it is easy to calculate i_{na} :

$$i_{na}^2 = \left(\frac{\overline{v_{NO}^2}}{K_{V12}^2 B} - e_{na}^2 - 4kTR \right) / R^2 \text{ MS A/Hz.} \tag{8.95}$$

Our final topic in this section will be to describe how we can measure an amplifier's *spot noise figure*, NF. Spot noise factor, F , is defined as the ratio of the mean-squared SNR at the amplifier's input to the mean-squared SNR at the amplifier's output. An ideal, noiseless amplifier would have an $F = 1$. An amplifier's noise figure is defined as $10 \log_{10}$ (noise factor). An amplifier's spot noise factor is a useful measure of its noise performance under various conditions of source resistance and frequency.

A circuit that can be used to measure spot noise figure of amplifier A_1 is shown in Figure 8.37. A sinusoidal voltage source is used at the input to enable the measurement of F_{SPOT} . An NBPB with peak gain, K_F , and equivalent Hz noise bandwidth, B , is used to select the range of noise frequencies for the measurement. A broadband, true rms

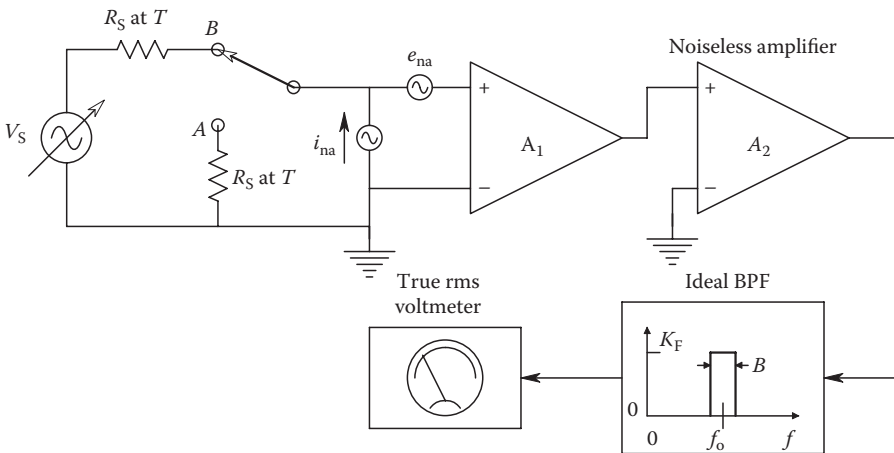


FIGURE 8.37 Circuit used to measure the spot noise figure of amplifier A_1 over a wide range of frequencies. Amplifier A_2 is assumed noiseless. The BPF center frequency f_0 is adjustable. The MS noise input to A_1 includes e_{na}^2 , the white thermal noise from R_S at T K, and $i_{na}^2 R_S^2$.

responding meter is used to read the total rms voltage at the amplifier output. With the switch in position *A*, the meter reading squared can be written as

$$\begin{aligned} \overline{v_{\text{NOA}}^2} &= \int_0^\infty K_{\text{V12}}^2 |\mathbf{H}_F(j2\pi f)|^2 [e_{\text{na}}^2(f) + i_{\text{na}}^2 R_S^2 + 4kTR_S] df \\ &= \int_0^\infty K_{\text{V12}}^2 |\mathbf{H}_F(j2\pi f)|^2 e_{\text{na}}^2(f) df + K_{\text{V12}}^2 K_F^2 B (i_{\text{na}}^2 R_S^2 + 4kTR_S) \text{ MS V.} \end{aligned} \tag{8.96}$$

Note that the bandwidth of the meter and the amplifier are assumed to be greater than and overlap that of the NBPF, so that the filter’s frequency response curve is dominant. Now, the switch is set to position *B*, and the sinusoidal source is adjusted to the center frequency of the NBPF. Its amplitude is adjusted until

$$\begin{aligned} \overline{v_{\text{NOB}}^2} = 2v_{\text{NOB}}^2 &= K_F^2 K_{\text{V12}}^2 \left(\frac{V_S^2}{2} \right) + \int_0^\infty K_{\text{V12}}^2 |\mathbf{H}_F(j2\pi f)|^2 e_{\text{na}}^2(f) df \\ &+ K_{\text{V12}}^2 K_F^2 B (i_{\text{na}}^2 R_S^2 + 4kTR_S) \text{ MS V.} \end{aligned} \tag{8.97}$$

Now, the MS SNR in bandwidth *B* at the amplifier’s input is

$$\text{SNR}_{\text{IN}} = \frac{\overline{v_S^2}}{4kTR_S B} \tag{8.98}$$

The SNR_{OUT} is

$$\text{SNR}_{\text{OUT}} = \frac{\overline{v_S^2}}{\left\{ \int_0^\infty K_{\text{V12}}^2 |\mathbf{H}_F(j2\pi f)|^2 e_{\text{na}}^2(f) df + K_{\text{V12}}^2 K_F^2 B (i_{\text{na}}^2 R_S^2 + 4kTR_S) \right\}} \tag{8.99}$$

Now, the spot noise factor,

$$F_{\text{SPOT}} \equiv \frac{\text{SNR}_{\text{IN}}}{\text{SNR}_{\text{OUT}}}, \tag{8.100}$$

can be written, after some algebra, as

$$F_{\text{SPOT}} = \frac{\int_0^\infty |\mathbf{H}_F(j2\pi f)|^2 e_{\text{na}}^2(f) df / K_F^2 + B (i_{\text{na}}^2 R_S^2 + 4kTR_S)}{4kTR_S B} \tag{8.101}$$

If we combine Equations 8.96 and 8.97, it is easy to show that the numerator of Equation 8.101 is equal to $V_S^2/2$, which is the mean-squared sinusoidal input signal of frequency f_o required to make $v_{\text{NOB}}^2 = 2v_{\text{NOA}}^2$. Thus, we can finally write the expression for the spot noise factor as

$$F_{\text{SPOT}} = \frac{V_S^2/2}{4kTR_S B} \tag{8.102}$$

Note that V_S is the peak value of the sinusoid of frequency f_o required to make the true rms meter reading increase by $\sqrt{2}$ in the B switch position, B is the NBPF's equivalent Hz noise bandwidth, and T is the temperature in K of R_S . F_{SPOT} is often presented as a plot of $10 \log_{10}(F_{\text{SPOT}})$ vs. R_S and f_o values by manufacturers of low-noise amplifiers and transistors.

An alternate means of measuring F_{SPOT} makes use of the same circuit as in Figure 8.37, except the sinusoidal source is replaced with a calibrated, broadband, white noise source with a flat spectrum in the $1/f$ region of e_{na} . The same procedure is followed; first, we measure v_{NOA}^2 and then adjust the noise source level to e_{NS}^2 MS V/Hz so that $v_{\text{NOB}}^2 = 2v_{\text{NOA}}^2$. It is left as an exercise for the reader to show that for the noise source method,

$$F_{\text{SPOT}} = \frac{e_{\text{NS}}^2}{4kTR_S}. \quad (8.103)$$

Note that the center frequency of the NBPF and its noise bandwidth do not appear in this simple expression, but they should be specified in presenting the calculated values for F_{SPOT} . In general, a different value of e_{NS} will be found for each point in f_o , R_S space.

Often, we wish to measure the rms voltage of a noise source in bandwidth B when the noise is of the same order of magnitude or smaller than the noise arising in our signal conditioning amplifiers. Figure 8.38 illustrates a means of measuring such low-level noise voltages. The noise voltage under measurement, e_{NS} , appears in series with a

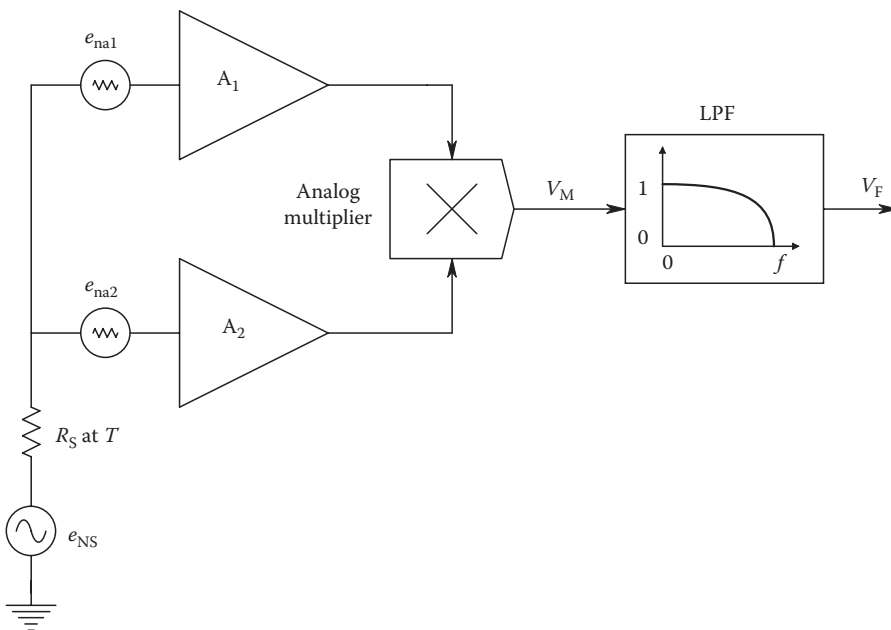


FIGURE 8.38

A correlation detector used to measure the MS voltage of the noise source $e_{\text{NS}}(t)$ and the Johnson noise from R_S at T K over the noise bandwidth B of the LPF. e_{na1} and e_{na2} have no effect on V_F .

Thevenin source resistance, R_S . This source is connected to two identical amplifier channels with gain K_V and noise bandwidth B Hz. Both amplifiers have short-circuit input equivalent voltage noise sources that are equal statistically, have zero means, but otherwise are independent and uncorrelated. The outputs of the two channels are multiplied together using a precision, four-quadrant, broadband, analog multiplier. The output of the multiplier is then passed through an LPF to average it. For simplicity, let us assume that the noises have finite bandwidths less than B . Thus, the multiplier output can be written as

$$\begin{aligned} v_M(t) &= (0.1)\{K_V[v_{NS}(t) + e_{na1}(t)]\}\{K_V[v_{NS}(t) + e_{na2}(t)]\} \\ &= (0.1)K_V^2\left[v_{NS}^2(t) + v_{NS}(t)e_{na2}(t) + e_{na1}(t)v_{NS}(t) + e_{na1}(t)e_{na2}(t)\right]. \end{aligned} \quad (8.104)$$

Low-pass filtering $v_M(t)$ effectively computes $\overline{v_M}$. Because the three voltages have zero means and are statistically independent and uncorrelated with one another, the average of each cross-term product is the product of the averages, which by definition are zero. Thus, the output of the filter is a DC voltage, V_F :

$$V_F = (0.1)K_V^2\overline{v_{NS}^2}. \quad (8.105)$$

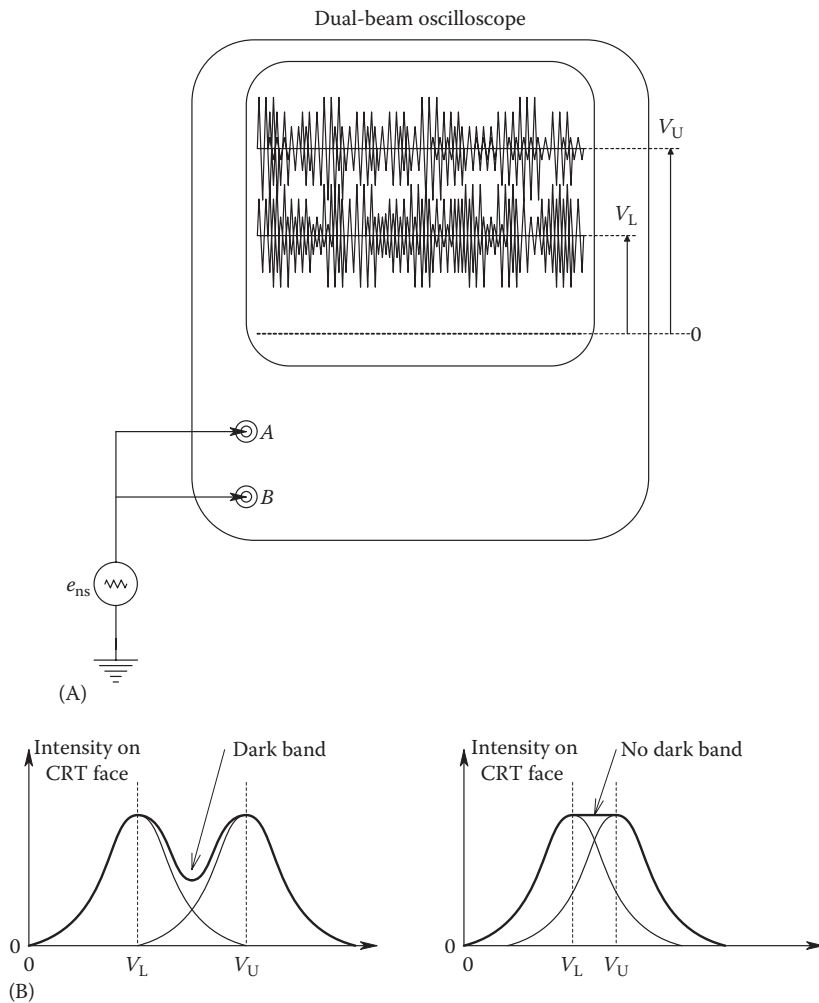
V_F is thus proportional to the mean square of the noise voltage under measurement, taken over bandwidth B . The effects of the amplifier voltage noise drop out; we have assumed the current noise to be negligible (i.e., $i_{na}R_S \ll e_{na}$). Also, because of R_S ,

$$\overline{v_{NS}^2} = (4kTR_S + e_{NS}^2)B. \quad (8.106)$$

An approximate means of estimating the RMS value of a broadband, the Gaussian noise source was described by HP. The *tangential noise measurement method* is extremely simple. It makes use of a dual-beam oscilloscope. The noise voltage is applied to both channels with the same gain. At wide trace separation, we see two, *grassy* noise traces separated by a dark band. The vertical position control of one trace is then adjusted slowly to bring the grassy traces together to a critical separation so there is no dark band between the traces, that is, there is a uniform illumination of the CRT screen between the traces. It can be shown that this condition corresponds to a trace separation of two standard deviations of the Gaussian noise. Referring to Figure 8.39, we see that

$$\sigma_x = \frac{V_U - V_L}{2}. \quad (8.107)$$

If the measurement is repeated five times and the results averaged, the $\overline{\sigma_x}$ is claimed to be accurate to about ± 0.5 dB, or $\pm 6\%$.

**FIGURE 8.39**

Visual noise estimation by the tangential method using a dual-trace oscilloscope using alternate sweeps. (A) Oscilloscope display showing two noise traces separated by a DC voltage ($V_U - V_L$). The DC trace separation is adjusted until the noise traces appear to merge into a solid, noisy band. When this occurs, the noise standard deviation $\sigma_x \cong (V_U - V_L)/2$. (B) Average CRT trace intensity as a function of the average vertical deflection voltage.

8.6 AC Measurements

AC measurements can be made over a wide range of frequencies and amplitudes. At audio and power frequencies, currents can be measured from nanoamperes to hundreds of kiloamperes. Different ranges require different techniques, which we shall describe in the following texts. At radio frequencies, the techniques available are considerably fewer. Regardless of the meter mechanism employed, AC ammeters are generally calibrated in terms of the RMS of a sine wave.

8.6.1 Electromechanical (Analog) AC Ammeters

Most of the electromechanical meter movements we have already discussed in considering AC voltmeters can be used for AC ammeters, with attention to their range and input current frequencies. To measure currents above the full-scale sensitivity of the meter movement, a shunt must be used. For currents in excess of about five amperes, a 50 mV external shunt is generally used. Recall that a 50 mV shunt has a resistance such that when the full-scale RMS current enters it, 50 mV rms is developed across the parallel combination of the shunt and the 50 mV rms voltmeter that reads the current. Shunts are typically used for high current measurements at audio or powerline frequencies using thermocouple, electro-dynamometer, or iron vane millivoltmeters. They are not used with electrostatic or D'Arsonval/rectifier-type AC meters. The thermocouple ammeter can be used at radio frequencies up to 50 MHz or so. The other meter movements, using coils, are only useful at powerline frequencies.

Another means used at powerline frequencies to measure high currents on conductors at high potentials is the *current transformer*. One or more turns of the alternating current-carrying conductor are passed through the center of a high-permeability, toroidal iron core, inducing a sinusoidally changing magnetic flux. Also, wound on the toroidal core are a number of turns of wire that are connected to an AC ammeter, generally having a dynamometer or iron vane movement. Figure 8.40A illustrates a typical

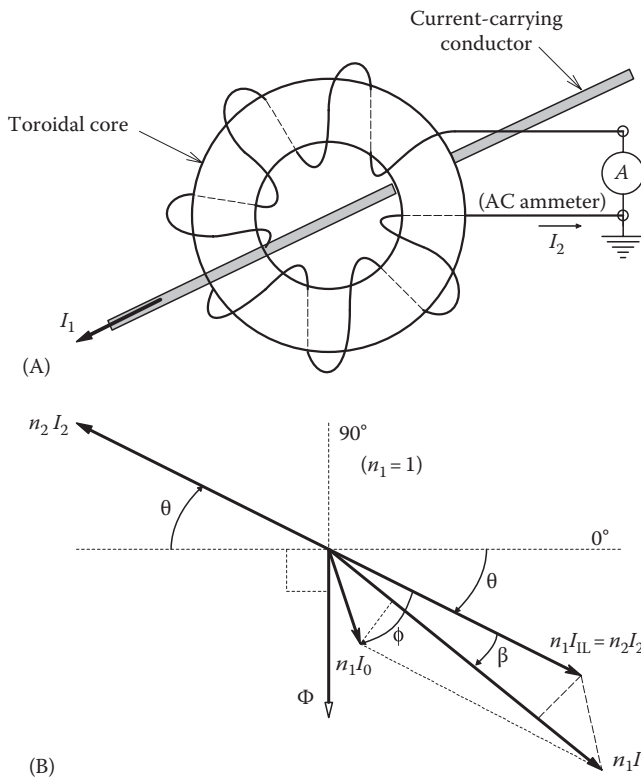


FIGURE 8.40 (A) A current transformer used at powerline frequencies. (B) Phasor diagram showing the current and flux vectors in a typical current transformer.

current transformer circuit. In a Midwest Electric Products, Inc. current transformer, the secondary (meter) winding was found to have a resistance of 0.04Ω , an inductance of 3.66 mH , and a turns ratio of $75/5$. This transformer was rated for 2 VA and works over a $25\text{--}400 \text{ Hz}$ range. Stout (1950) gave a current transfer function for the current transformer:

$$\frac{I_1}{I_2} \cong \frac{n_2}{n_1} \left[1 + \frac{I_o}{I_1} \sin(\theta + \phi) \right] \quad (8.108)$$

where

n_2/n_1 is the ratio of secondary (output) winding turns to primary (input) turns

I_o is the excitation or magnetizing component of the primary current, I_1

θ is the power factor angle between V_1 and I_1 and includes both the effect of the load (burden) and the current transformer

I_2 is the AC in the secondary ammeter

ϕ is the angle between the net magnetic flux in the current transformer's core, Φ_o , and the equivalent sine wave of I_o

A phasor diagram showing currents and flux in a typical current transformer is shown in Figure 8.40B. Current transformers should never be operated with their secondaries open-circuited; dangerous high voltages can exist there.

The *clamp-on ammeter* is a portable version of the current transformer in which the secondary load impedance is high, so sufficient secondary EMF is developed to operate an AC millivoltmeter or a D'Arsonval/rectifier-type AC voltmeter. Another meter movement used with a clamp-on ammeter is the rotating iron vane type used in the venerable Columbia Electric Mfg. Co's. *tong test ammeter*. In the Columbia meter movement, magnetic flux induced in the iron jaws by enclosing a current-carrying conductor acts to draw an iron vane into an air gap at the base of the jaws. This torque acts against a torsion spring, producing a pointer deflection that is a nonlinear function of primary current. Mechanical damping is produced by eddy currents induced in an aluminum damping pan. Ranges were changed in the Columbia clamp-on AC ammeter by plugging in different meter assemblies into the clamp-on jaws' air gap. Full-scale currents from 20 to 1000 A rms were available. The bottom 20% of the Columbia meters' scales is not useful, because of deflection nonlinearity.

Yet another means of measuring AC makes use of a *compensated thermocouple*, illustrated in Figure 8.41. Two large copper blocks serve as attachment points for the input current line and as thermal reference points. The heater element is attached between the blocks, and the hot junction of the thermocouple is spot welded to the middle of the heater. The ends of the thermocouple wires are attached to copper strips that are in intimate thermal contact with the copper blocks but electrically insulated from them by mica sheets. The leads to the D'Arsonval microammeter are attached to the copper strips. The AC (or DC) input current causes a temperature rise in the heating element, which is sensed by the thermocouple. The voltage drop across the thermocouple's heater is designed to be 0.15 V at full-scale current. Hence, this type of ammeter absorbs 150 mW for each ampere of full-scale range. This type of ammeter operates at DC and at AC frequencies up to 50 MHz and for full-scale currents of $5\text{--}50 \text{ A rms}$ (Stout 1950).

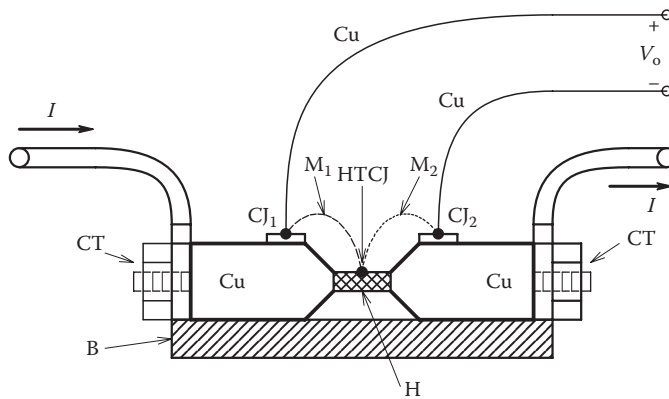


FIGURE 8.41 Section through a compensated thermoelement device used to measure the MS values of DC or ACs. *Note:* B = bakelite base. Cu = copper blocks. CT = current input and output terminals. H = heater element. HTCJ = hot thermocouple junction. CJ₁ and CJ₂ = reference (cold) thermocouple junctions. M₁ and M₂ = metals used in thermocouple. M₁/Cu and M₂/Cu make the reference thermocouples.

8.6.2 Electronic and Magneto-Optical AC Ammeters

To measure AC smaller than the full-scale currents required to operate the inefficient electro-dynamometer and iron vane meter movements (about 50 mA rms), or the full-scale current for VTC AC ammeters (about 3 mA rms), it is necessary to amplify the current or a voltage proportional to the current. The OA transresistor circuit offers a means of measuring very small AC flowing into ground, ranging from picoamperes to about 10 mA. In this case, the OA's summing junction appears as a virtual ground, as shown in Figure 8.42. The AC voltage at the output of the operational transresistor is simply $V_o = -R_F I_S$ and can now be displayed by any of the conventional AC, analog indicating meters we have discussed earlier, or it can be displayed digitally.

Another means of measuring low-level ACs is to use a current transformer followed by an AC amplifier, followed by an AC analog indicating meter or a digital AC voltmeter. Pearson Electronics, Inc. makes a variety of wide-bandwidth current sensors that have a voltage output. For example, the Pearson model 2877 current sensor has a

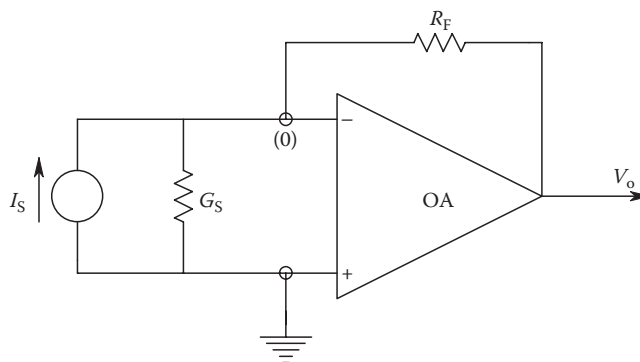


FIGURE 8.42 An operational transresistor circuit is used to convert a small current, I_S , to a voltage, V_o .

50 Ω output impedance, an open-circuit transresistance of 1 V/A, a max peak current of 100 A, a max rms current of 2.5 A, a 2 ns rise time, a 0.2%/ μ s droop, and -3 dB frequencies of 300 Hz and 200 MHz. At the other end of the current measurement range, the Pearson model 1423 has an open-circuit transresistance of 0.001 V/A, a max peak current of 500,000 A, a maximum RMS current of 2,500 A, a 300 ns rise time, a 0.7%/ms droop, and -3 dB frequencies of 1 Hz to 1.2 MHz. Current transformers, such as the Pearson sensors, find application in measuring current transients in insulation breakdown studies and in investigations of high-power pulse circuits using SCRs, hydrogen thyratrons, strobe tubes, klystrons, and magnetrons. Of course, SS sinusoidal currents can also be measured. Since the maximum current sensitivity of the Pearson model 2877 sensor is 1 V/A, it is possible to resolve AC in the microamp range using appropriate low-noise amplifiers and band-pass filters.

Another means of measuring AC in a conductor is to *measure the magnetic field around the conductor*. A sensitive Hall effect sensor probe (see Section 6.2.4.2) can be put next to the conductor to intercept the surrounding, solenoidal B field. The AC output voltage of the Hall sensor is amplified and then used to drive an appropriate meter. For more sensitive current sensing, a flux concentrating, iron C core can be put around the conductor, as shown in Figure 8.43, and the Hall sensor put in its air gap. The ferromagnetic material and the length of the air gap must be chosen so that the magnetic material does not go into saturation at the desired, full-scale, AC to be measured.

The final AC measuring system we will consider is the *Faraday magneto-optic sensor*, discussed in some detail in Section 6.5.1. A good, physical discussion of the Faraday magneto-optic effect can be found in Chapter 7 of the optics text by Möller (1988). Some of the earliest work on fiber-optic Faraday effect current sensors was done by Rogers (1973). As shown in Figure 6.39 in this text, the Rogers system uses a ratio detector to make the system insensitive to amplitude variations in the laser source. Section 6.7.1 describes the operation of the ratio detector and shows that the output of the ratio detector is a sinusoidal voltage at powerline frequency with peak amplitude proportional to the peak amplitude of the conductor by the Verdet constant. After amplification, this line frequency voltage can be displayed by one of the many AC responding meters we have described earlier. Other Faraday magneto-optic current sensor designs have been described by Cease and Johnston (1990) and Nicati and Robert (1988).

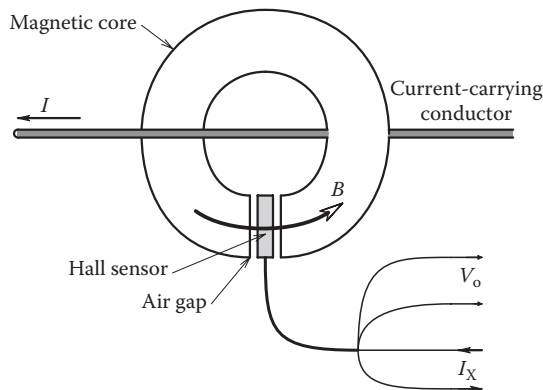


FIGURE 8.43

An analog Hall sensor is used in the air gap of a ferromagnetic ring to measure the AC in the conductor. The Hall output voltage, V_o , is proportional to the product of B and I_x , hence I .

8.7 Magnetic Field Measurements

There are many applications in the design of rotating electric machines (alternators, DC generators and motors, brushless DC motors, stepping motors, induction motors, etc.) where it is useful to be able to measure the magnetic field strength in air gaps in order to verify designs. There are also many applications in physics and in chemistry where it is necessary to measure magnetic fields, for example, in mass spectrometers, in magnetic resonance imaging systems, and in particle accelerators. Recently, powerline frequency magnetic fields have been implicated as possible contributing causes in health problems including birth defects, leukemia, and cancer (*EPRI Journal*, January/February 1990). Thus, measurement of the spatial distribution of 60 Hz B fields in areas containing people, experimental animals, or cell cultures is important for the clarification of the significance of 60 Hz B fields in health problems. The magnetic field vector, \mathbf{B} , is also called the *magnetic induction* or the *magnetic flux density*. Its units are newtons per coulomb divided by meters per second, or equivalently volt-seconds per square meter, or webers per square meter (Wb/m^2). The more common units for \mathbf{B} are gauss ($1 \text{ Wb/m}^2 = 10^4 \text{ G}$) or teslas ($10^4 \text{ G} = 1 \text{ T}$). Magnetic flux is defined as the integral of \mathbf{B} over an area A perpendicular to \mathbf{B} , or

$$\Phi = \int_A \mathbf{B} \cdot d\mathbf{A}. \quad (8.109)$$

The units of magnetic flux are webers (SI), or maxwells (CGS), found by integrating 1 G over 1 cm^2 . The vector \mathbf{H} is called the *magnetic field*. It is related to \mathbf{B} by the well-known relation, $\mathbf{H} = \mathbf{B}/\mu$. The SI units of \mathbf{H} are newtons/weber or amps/meter, and the CGS units of \mathbf{H} are oersteds ($1 \text{ A/m} = 4\pi \times 10^{-3} \text{ Oe}$).

There are a variety of magnetic sensors that can be used to measure DC and AC \mathbf{B} fields in space. The first and most basic means is the solenoidal search coil. If the \mathbf{B} field is homogeneous over the area of the coil, a time-varying \mathbf{B} field will induce an open-circuit EMF given by the well-known relation

$$E_M = N A \dot{B} \cos(\theta) \quad (8.110)$$

where

N is the number of turns in the coil

A is the area of the coil

θ is the angle \mathbf{B} makes with respect to a normal to the coil's area

$$\dot{B} = dB/dt$$

In most powerline \mathbf{B} fields, $B(t) = B_o \sin(\omega t)$, so

$$e_M(t) = N A \cos(\theta) \omega B_o \cos(\omega t). \quad (8.111)$$

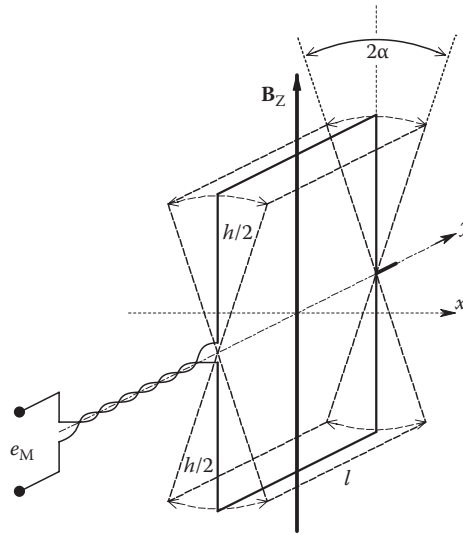


FIGURE 8.44
An oscillating search coil used to measure a DC magnetic field, \mathbf{B}_z .

If the \mathbf{B} field is constant, for example, the Earth’s magnetic field, an EMF may be induced in the coil proportional to B by rotating the coil at an angular velocity of ω_c r/s. In this case, the $\cos(\theta)$ term becomes $\cos(\omega_c t)$, and the induced EMF is given by

$$e_M(t) = NAB\omega_c \sin(\omega_c t). \tag{8.112}$$

Several variations on the rotating coil magnetic field meter were described by Lion (1959). For example, a coil can be mechanically sinusoidally oscillated around a center position in which a constant \mathbf{B} vector is parallel with the plane of the coil, as shown in Figure 8.44. The EMF induced in the coil can be shown to be given by

$$e_M(t) = NAB\alpha_o\omega_r \cos(\omega_r t). \tag{8.113}$$

Here, α_o is the angle over which the coil is oscillated at mechanical radian frequency, ω_r , so $\alpha(t) = \alpha_o \sin(\omega_r t)$. We assume that $\alpha_o < 15^\circ$.

Three orthogonal search coils can be used to determine the magnitude and direction of a powerline frequency \mathbf{B} field. The \mathbf{B} vector can be resolved into three orthogonal, vector components, \mathbf{B}_x , \mathbf{B}_y , and \mathbf{B}_z . Each component may be considered to produce an EMF in its respective search coil. Referring to Figure 8.45, we see that in spherical coordinates, the three components of \mathbf{B} can be written as scalars:

$$B_z = B \cos(\phi) \tag{8.114a}$$

$$B_x = B \sin(\phi) \cos(\theta) \tag{8.114b}$$

$$B_y = B \sin(\phi) \sin(\theta) \tag{8.114c}$$

where

ϕ is the colatitude angle measured between the positive y -axis and \mathbf{B}

θ is the longitude angle, measured between the positive x -axis and the projection of \mathbf{B} on the XY plane

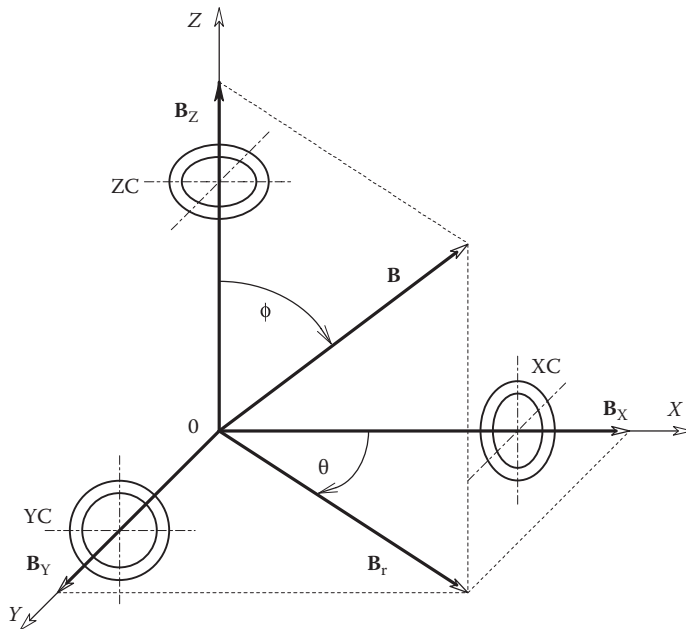


FIGURE 8.45 Three orthogonal search coils used to measure the x , y , and z components of a time-varying magnetic field.

This projection is B_r and is equal to

$$B_r = B \sin(\phi). \tag{8.115}$$

From Equation 8.113 earlier, we can write expressions for the instantaneous EMFs induced in the coils:

$$e_{MX}(t) = NA\omega B \sin(\phi) \cos(\theta) \cos(\omega t) \tag{8.116a}$$

$$e_{MY}(t) = NA\omega B \sin(\phi) \sin(\theta) \cos(\omega t) \tag{8.116b}$$

$$e_{MZ}(t) = NA\omega B \cos(\phi) \cos(\omega t). \tag{8.116c}$$

Now, it is easy to show that if we square each of the coil EMFs with four-quadrant analog multipliers and then average each multiplier output by passing it through an LPF, add together the mean-square voltages, then take the square root, we will obtain a DC voltage proportional to B , regardless of the orientation of the three-coil probe in the B field. This process is illustrated in Figure 8.46. Numerically, B can be shown to be

$$B = \frac{V_o \sqrt{2}}{NA\omega}. \tag{8.117}$$

(Assume all coils to have identical N s and A s.) The angles ϕ and θ can also be found relative to the probe's x -, y -, and z -axes from trigonometric identities; however, this is tedious.

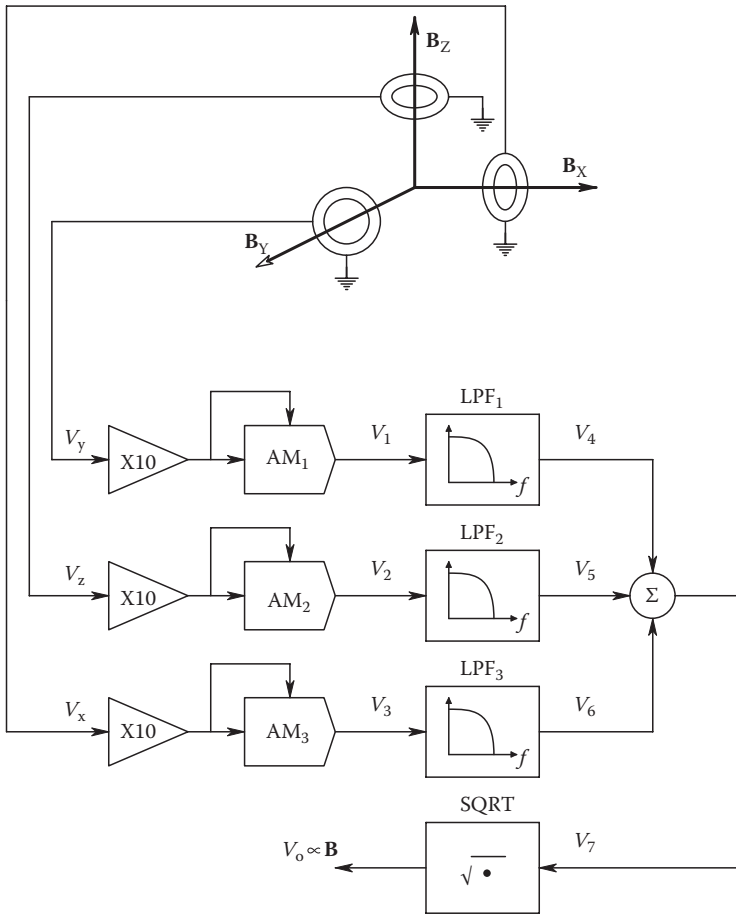


FIGURE 8.46 Electronic circuit used to compute the magnitude of the net $\mathbf{B}(t)$ vector using the Pythagorean theorem. The three orthogonal coils of Figure 8.45 supply the input voltages to the system.

A better way to find the direction of \mathbf{B} is to experimentally orient the probe so that $e_{MX}(t)$ is maximum and $e_{MY}(t)$ and $e_{MZ}(t)$ are minimum. Now, the angle of \mathbf{B} is parallel with the probe's x -axis.

Analog Hall effect sensors provide another major means of measuring either constant or time-varying magnetic fields. The physics governing the behavior of the Hall sensors was described in detail in Section 6.2.4.2. It was shown that the Hall sensor EMF is given by

$$E_H(t) = \frac{B(t) \cos(\theta) I_x(t) R_H}{h} \tag{8.118}$$

where

θ is the angle \mathbf{B} makes with the XZ plane

$I_x(t)$ is the current flowing in the $+x$ -direction (electrons in n -semiconductor drift in the negative x -direction)

R_H is the Hall constant

$R_H = -1/qd$ for n semiconductor and $R_H = 1/qd$ for p semiconductor. q is the magnitude of the electron charge, d is the density of the carrier doping in the semiconductor, and h is the thickness of the Hall slab in the y -direction. In the operation of most analog Hall sensors, I_x is generally DC. However, in the measurement of a low-level, constant \mathbf{B} , a sinusoidal or square wave $I_x(t)$ can be used to permit noise reduction by the amplification of $e_H(t)$ out of the $1/f$ noise region and also the use of synchronous rectification to further improve SNR.

Modern analog Hall sensors can be purchased with built-in, IC DC amplifiers, or as the basic sensor chip. Unamplified Hall sensors have conversion gains ranging from 10 to 55 mV/kG. Through the use of a 9" flux concentrating iron bar, the F. W. Bell model BH-850 Hall sensor has a gain of 18 mV/G. Normal operating current for the Bell BH-850 sensor is 200 mA. The resistance in its current path and its Thevenin output resistance for E_H are both 3.5 Ω , giving good, low-noise performance.

Various companies make gaussmeters for various applications having either analog or digital meter readouts. For example, the Bell model 615 digital gaussmeter, a benchtop instrument, reads 10 – 10^6 G in 5 ranges, has a DC to 2 kHz frequency range, is accurate to $\pm 0.5\%$ of FS, and has a readout resolution of 1 part in 10^3 . The Bell model 4048 is a handheld, portable, digital gauss/tesla meter with three full-scale ranges: 200, 2000, and 2×10^4 G. Accuracy is $\pm 2\%$ of reading for DC fields, $\pm 2.5\%$ for 45–100 Hz AC fields, $\pm 2\%$ for 100 Hz to 3 kHz, and $\pm 6\%$ for 3–5 kHz fields. This instrument is available with transverse and axial Hall probes.

The practical lower limit of Hall sensor sensitivity is set by noise, both from the sensor's thermal and shot noise and from the signal conditioning amplifier's e_{na} and i_{na} . Milligauss resolution is practical with most modern Hall sensors. To achieve higher sensitivity in magnetic field measurements, it is necessary to use sensors that utilize different physical principles.

For example, at microwave frequencies, many materials exhibit a very strong Faraday magneto-optic effect (see Section 6.6.1). The Verdet constant of certain ferromagnetic ferrites at 9 GHz is on the order of $0.1^\circ/(\text{Oe cm})$. By the use of such a ferrite, magnetic flux density variations on the order of 10^{-5} G have been detected (Lion 1959).

The deflection of a collimated, low-energy electron beam by a \mathbf{B} field is also an effective means of measuring low values of \mathbf{B} , ranging from 2×10^{-6} to 2×10^{-4} G (Lion 1959). Figure 8.47 illustrates the geometry of the CRT used for such a measurement. Note that \mathbf{B} must be perpendicular to the electron beam axis and the deflection axis. The electron beam is accelerated through a potential of V volts giving the electrons in the beam an average velocity of

$$\mathbf{v}_x = \sqrt{2V \left(\frac{q}{m} \right)} \mathbf{i}_x \text{ m/s} \quad (8.119)$$

where

q/m is the ratio of electron charge to electron mass

\mathbf{i}_x is a unit vector in the $+x$ -direction

The beam passes through L meters of a uniform magnetic field of strength \mathbf{B}_z . A Lorentz force of

$$\mathbf{F}_y = -q (\mathbf{v} \times \mathbf{B}) = -(qv_x B_z) \mathbf{i}_y \text{ N} \quad (8.120)$$

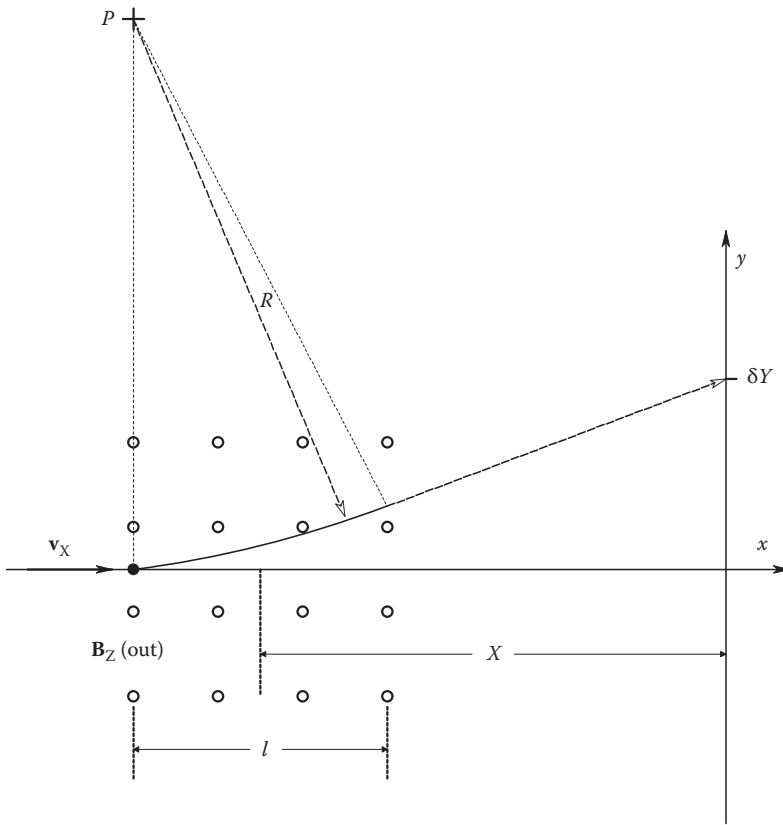


FIGURE 8.47

Diagram showing the behavior of an electron moving at constant velocity, v_x , which enters a region of uniform magnetic field of length l (B_z out of page). The electron crosses the y -axis with a net deflection from the center line (x -axis) of δY . Its trajectory through the magnetic field is curved with radius R . Its trajectory is straight after it exits the magnetic field.

acts on each electron, causing the beam to deflect along a circular path with a radius R . R may be shown to be equal to

$$R = \frac{mv_x}{qB_z} \text{ m.} \tag{8.121}$$

Kraus (1953) showed that the beam deflection, δy , can be used to measure B_z . Referring to Figure 8.47, B_z can be expressed as

$$B_z = \frac{\delta y}{LX} \sqrt{2V \frac{m}{q}} \text{ W/m}^2. \tag{8.122}$$

Lion (1959) illustrates a means of detecting beam deflections, δy , too small to visualize on the phosphor screen of the electron beam magnetometer. The collimated beam strikes a

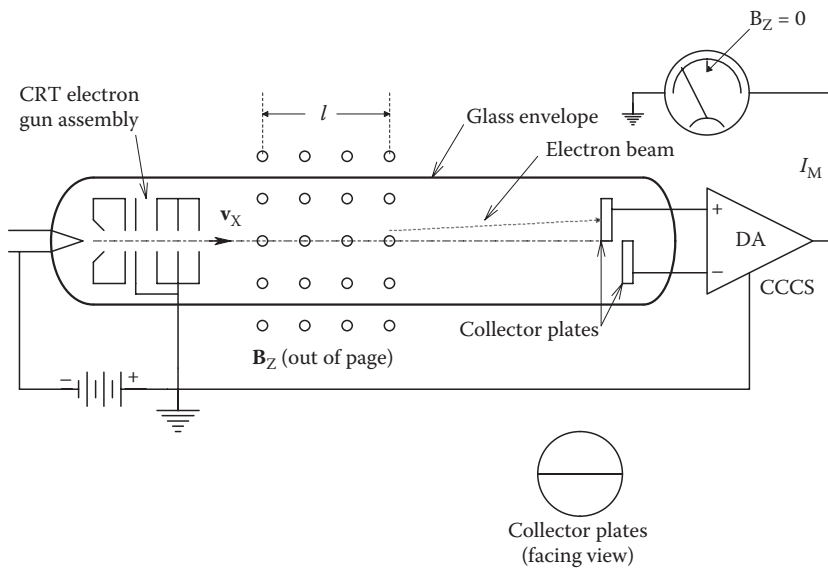


FIGURE 8.48

Use of a CRT to measure a DC B_z (out of page). The Lorentz force on the electrons passing through the orthogonal B field causes a deflection δY .

split anode, shown in Figure 8.48. At zero B_z , the beam is positioned so that exactly half its electrons strike each anode plate. The anode currents are converted to two voltages, which are subtracted by a DC differential amplifier. In the presence of a $B_z > 0$, the beam deflects, and more current strikes one plate than the other, causing an output signal. If this output signal is integrated, and then conditioned to make a current to drive a deflection coil, the effective flux density of the deflection coil will be equal to $-B_z$ in the steady state, giving a null output.

The most sensitive magnetic field sensors are the SQUID magnetometers. SQUID stands for *superconducting quantum interference device*. SQUIDs find application in biophysical measurements in neurophysiology. Nerve action potentials are accompanied by transient flows of electric currents, generally carried by ions. These transient currents generate minute, time-varying, magnetic fields that can be measured and located spatially in the brain by arrays of SQUID sensors (Ribeiro et al. 1988, Northrop 2002). Besides studies in neuro-magnetism, SQUIDS can be used in magnetocardiography and detection of small changes in the Earth's magnetic field in geological studies.

SQUIDS are made from various superconductor materials and are operated at the boiling point of liquid He at atmospheric pressure (4.2 K). There is a possibility that SQUIDS can be made from special, high-temperature semiconductors that can operate at temperatures up to 125 K (Clarke and Koch 1988).

Conventional DC SQUIDS are the most sensitive magnetometers available; they are made from a superconducting ring of niobium or lead and operated in liquid helium. The ring has an inductance, L , in the picohenry range and is joined with two *Josephson junctions*, as shown in Figure 8.49A. The SQUID is really a four-terminal device; two terminals are used to input a DC bias current, I_B , and the same two terminals to monitor the output voltage, V_o . The voltage V_o remains zero until the bias current reaches a

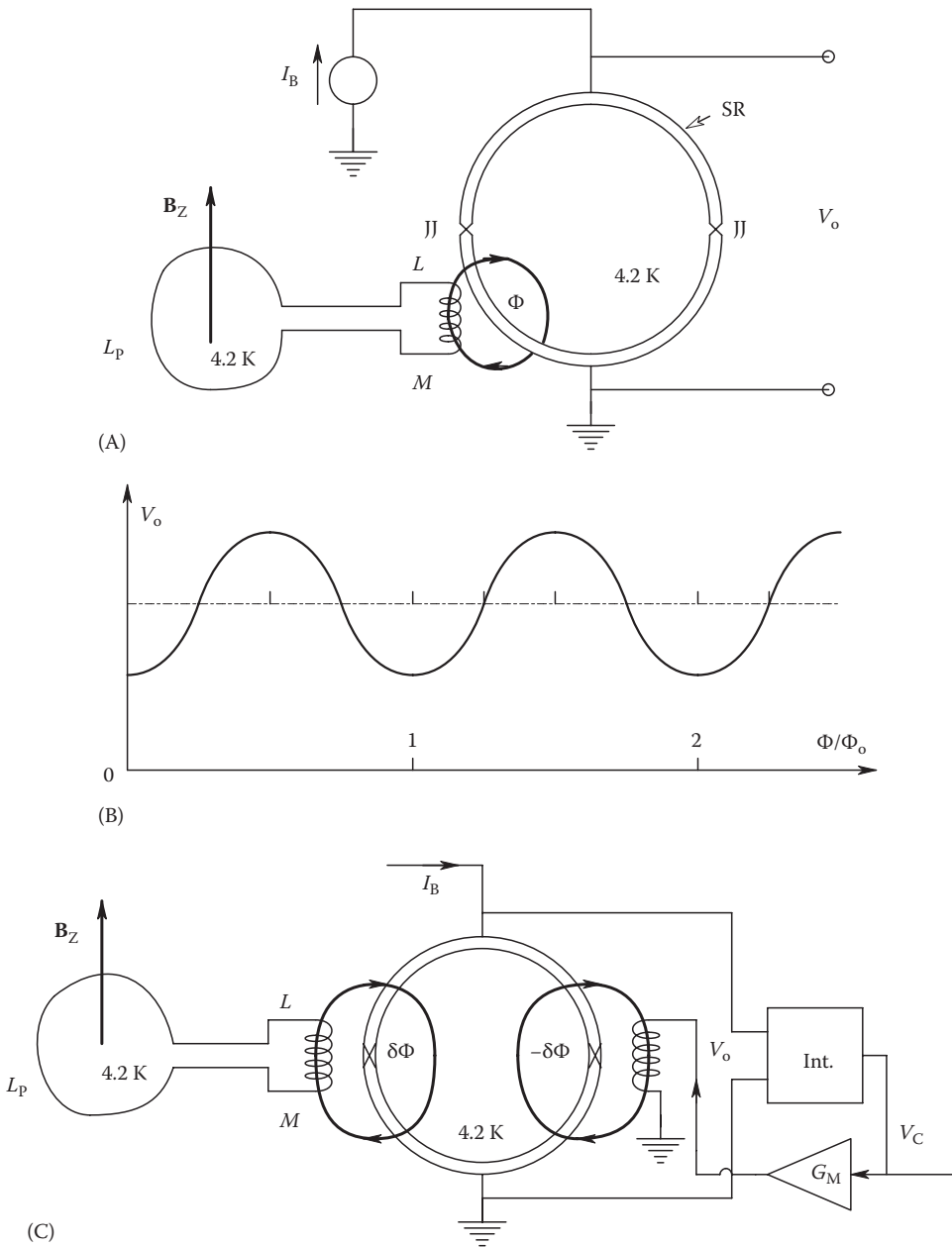


FIGURE 8.49

(A) Schematic of a SQUID device. *Note:* SR = superconducting ring of lead or niobium. JJ = Josephson junction. I_B = DC bias current. L = inductance of coupling coil. M = mutual inductance between coupling coil and the SR. L_p = inductance of probe coil. (B) Variation of SQUID output voltage at constant bias current as a function of the ratio of the applied flux to the flux quantum. (C) A feedback, flux-locked SQUID.

critical value, I_0 . Then the output voltage increases with current and is also a function of the magnetic flux linking the SQUID ring. The DC bias current, I_B , is made greater than I_0 . The superconducting SQUID ring circuit undergoes the phenomenon of *fluxoid quantization* in which the magnetic flux linking the SQUID is given by $n\Phi_0$. n is an integer, and Φ_0 is the flux quantum, equal to $h/2q = 2 \times 10^{-15}$ Wb. If we apply an additional flux, Φ_i , through the SQUID ring, a supercurrent, $I_s = -\Phi_i/L$, is set up in the ring to create a flux that cancels Φ_i . In other words, $LI_s = -\Phi_i$. From Figure 8.49B, we see that at a constant bias current, the SQUID output voltage varies periodically as a function of Φ/Φ_0 . This phenomenon is a lot like the generation of intensity variations due to interference rings in an optical system, such as a Sagnac gyro. The SQUID sensor is often operated as a feedback device, as shown in Figure 8.49C. A bias flux is adjusted so that V_0 is linearly proportional to small changes in input flux, $\delta\Phi$. V_0 is then conditioned and integrated and used to control the current in a feedback coil such that a flux $-\delta\Phi$ is generated to cancel the input flux. This system operates the SQUID as a null-flux detector and is called a *flux-locked SQUID*. This closed-loop operation permits a large dynamic range of input flux to be measured. As shown in Figure 8.49A, a superconducting input coil must be used to couple the flux to be measured into the SQUID toroid. It appears that optimum SQUID sensitivity occurs when the inductance of the pickup loop, L_p , is equal to that of the coupling loop, L_c . Assuming noiseless flux coupling, Clarke and Koch (1988) showed that the equivalent flux density noise of the SQUID can be given by

$$B_N(f) = \frac{2\sqrt{2L_p}}{\pi r_p^2} \sqrt{\frac{\epsilon(f)}{\alpha^2}} \text{ w/m}^2/\sqrt{\text{Hz}} \tag{8.123}$$

where

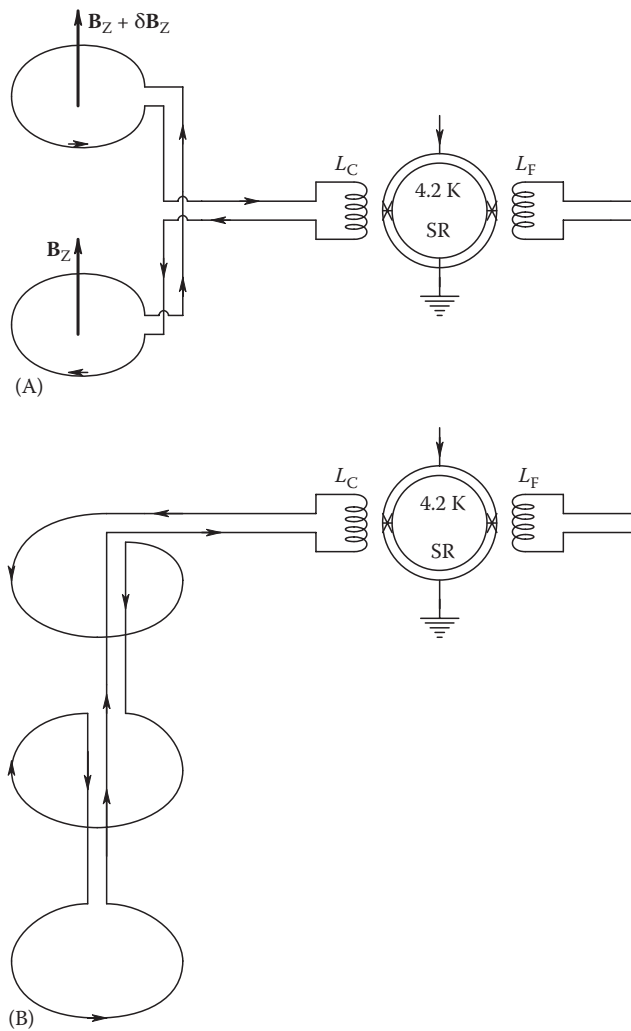
- α is the coupling coefficient between the SQUID and the input coil
- r_p is the radius of the input coil
- L_p is the inductance of the pickup coil
- $\epsilon(f)$ is the white noise energy/Hz of the SQUID

$$\epsilon(f) = \frac{9kTL}{R} \text{ J/Hz} \tag{8.124}$$

where

- k is Boltzmann's constant
- T is in K
- L is the SQUID's inductance
- R is the resistance shunting each Josephson junction in the SQUID ring

For a typical Nb SQUID with $100 < L < 500$ pH, the noise energy is about $1-5 \times 10^{-32}$ J/Hz (Clarke and Koch 1988). With such low noise, it is possible to resolve magnetic flux densities on the order of 10^{-14} T/ $\sqrt{\text{Hz}}$ ($1 \text{ T} = 10^4 \text{ G}$). A very low white noise of 5 fT/ $\sqrt{\text{Hz}}$ has been achieved using thin-film technology. Of course, external magnetic noise pickup can confound sensitive measurements, so SQUID sensors are normally well shielded with layers of Mu metal, aluminum, and even superconducting materials.

**FIGURE 8.50**

(A) A SQUID with a gradiometer flux pickup. (B) A SQUID with a second derivative gradiometer flux pickup.

To further reduce the pickup of unwanted magnetic field noise, use is made of a *gradiometer coil*, shown in Figure 8.50A. This coil responds to spatial nonuniformities in the field under measurement, that is, $\partial B/\partial z$. First-order gradients from a magnetic dipole fall off as $1/r^4$, so the gradiometer coil discriminates against distant sources of interference in favor of local dipoles. Some workers have even used second derivative gradiometer coils that respond to $\partial^2 B/\partial z^2$ to obtain even better resolution of local sources (Ribeiro et al. 1988). A second derivative coil is shown in Figure 8.50B.

A comparison of the magnetic flux density resolution of common magnetometers is shown in Figure 8.51, adapted from Clarke and Koch (1988). Note that the non-superconducting coil gives very high resolution at frequencies above 1 Hz and is generally robust and simple. The extra sensitivity of SQUID sensors is needed, however, for such biomedical applications as neuromagnetism and magnetocardiography.

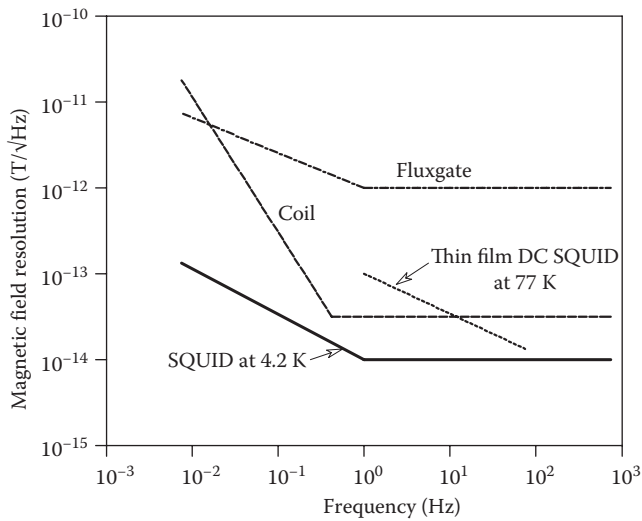


FIGURE 8.51 A comparison of the resolution and bandwidths of various sensitive magnetic field sensors. (From Clarke, J. and Koch, R.H., *Science*, 242, 217, 1988. With permission.)

8.8 Phase Measurements

It is often necessary to measure the phase difference between two periodic signals of the same frequency. In power distribution systems, the power factor angle between the AC supply voltage and the load current is of great economic and practical importance to the power generator and the end user. We have seen that several magneto-optical electric current sensing systems have AC outputs in which the phase difference is proportional to an optical polarization angle change induced by the magnetic field produced by the current under measurement. Also, phase information is important in the description of feedback system performance and filter behavior.

Phase measurements can be classified as being done by analog or digital systems. To illustrate the concept of phase difference, let us define two periodic functions having the same frequency but not coincident in time origin:

$$v_1(t) = v_1(t + T) = v_1(t + kT) \tag{8.125a}$$

$$v_2(t) = v_2(t + T + \Delta T) = v_2(t + kT + \Delta T) \tag{8.125b}$$

where

ΔT is the time difference between v_1 and v_2

T is the period of v_1 and v_2

k is an integer

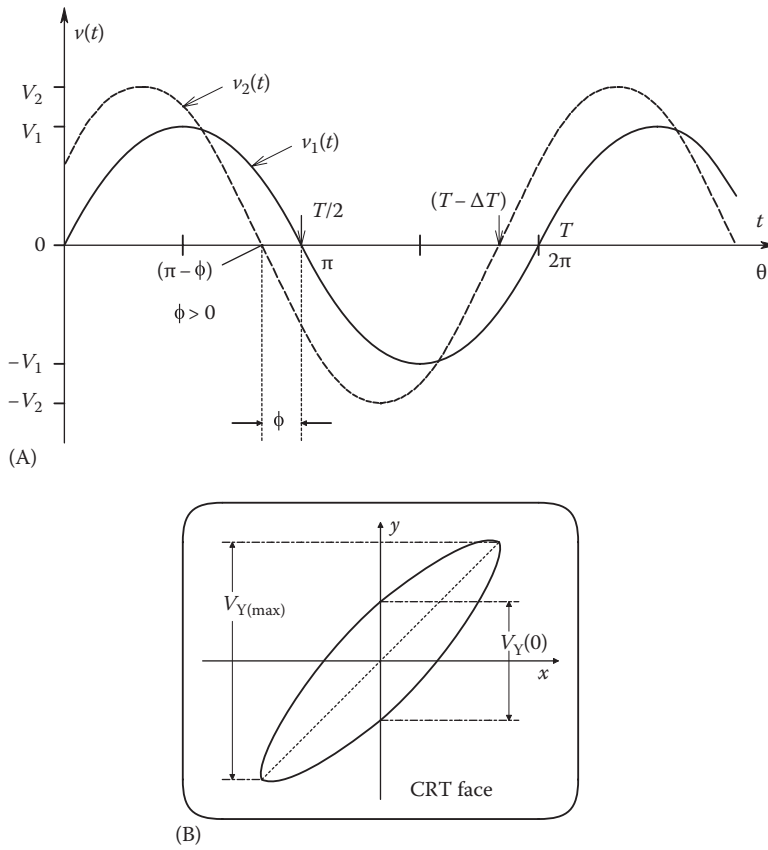


FIGURE 8.52

(A) Figure illustrating the phase difference between two sine waves having the same frequency and different amplitudes. (B) A Lissajous figure displayed on an oscilloscope. Voltage 1 drives the vertical deflection plates of the CRT, and voltage 2 drives the horizontal deflection. See text for analysis.

Figure 8.52A illustrates two sine waves having a phase difference. v_2 is said to *lead* v_1 by a phase difference of ϕ , where

$$\phi = 2\pi \left(\frac{\Delta T}{T} \right) \text{radians,} \tag{8.126a}$$

or

$$\phi = 360 \left(\frac{\Delta T}{T} \right) \text{degrees,} \tag{8.126b}$$

if

$$v_1(t) = V_1 \sin(\omega t) \tag{8.127a}$$

and

$$v_2(t) = V_2 \sin(\omega t + \phi). \tag{8.127b}$$

8.8.1 Analog Phase Measurements

Probably the simplest and easiest direct measurement of phase angle can be made on an oscilloscope screen using the *Lissajous figure* formed when two sine waves, given by Equation 8.127, are input to the X- (horizontal deflection) and the Y-axis (vertical deflection), respectively. In general, we see an ellipse on the CRT screen, as shown in Figure 8.52B. The ellipse must be centered at the origin (center) of the CRT screen. At $t = 0$, $v_1 = 0$, and $v_2 = V_2 \sin(\varphi)$. Thus, it is easy to show that the phase angle is given by

$$\varphi = \sin^{-1} \left[\frac{V_Y(0)}{V_Y(\max)} \right]. \tag{8.128}$$

From Equation 8.128, when the Lissajous figure is a straight line, $\varphi = 0^\circ$, and when it is a circle, $\varphi = 90^\circ$. Accuracy of the Lissajous figure method of phase measurement is poor near $\pm 90^\circ$. At phase angles near integer multiples of 180° , accuracy is better and is limited by the thickness of the oscilloscope trace and one's ability to estimate the distances to its intersections with the vertical axis of the display graticule.

Somewhat more accurate analog phase measurement techniques have been devised based on waveform averaging. Most of these techniques have been used in phase-lock loops to implement the phase detector element (Northrop 1990).

The first analog phase detector we will describe is the analog multiplier/LPF. In this system, shown in Figure 8.53, the input signals must be in quadrature for proper operation. That is,

$$v_1(t) = V_1 \sin(\omega_1 t) \tag{8.129a}$$

and

$$v_2(t) = V_2 \cos(\omega_1 t + \varphi). \tag{8.129b}$$

The output of the analog multiplier is

$$e(t) = \left(\frac{V_1 V_2}{20} \right) [\sin(2\omega_1 t + \varphi) - \sin(\varphi)]. \tag{8.130}$$

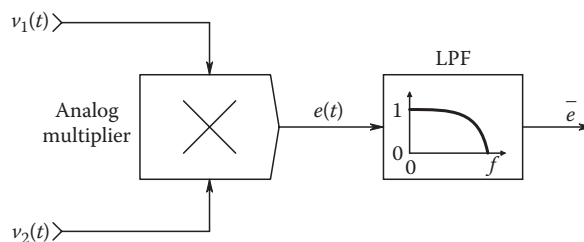


FIGURE 8.53
The analog multiplier (quadrature) phase detector.

After passing through the LPF, we have the DC component of $e(t)$:

$$\bar{e} = -\left(\frac{V_1 V_2}{20}\right) \sin(\phi) \cong -\left(\frac{V_1 V_2}{20}\right) \phi \quad (\phi \text{ in radians for } \phi < 0.25R). \quad (8.131)$$

The sign change is the result of the trigonometrical algebra. Linearity in this phase detector is preserved for $\phi < 12^\circ$; otherwise, an inverse sine nonlinearity must be used to recover a voltage proportional to ϕ .

A NAND gate flip-flop (FF) phase detector that works with TTL signals is shown in Figure 8.54A. The input analog sine waves are converted to TTL square waves by amplitude comparators connected as zero-crossing detectors with hysteresis for noise rejection. The TTL square wave outputs from the comparators are further conditioned by one-shot multivibrators such as 74LS123s, which have narrow complimentary outputs with dwell times generally much less than $T/100$. A simple NAND gate R-S flip-flop produces an output whose duty cycle depends on the phase difference between the input signals. As seen from the waveforms of Figure 8.54B, the FF output has a 50%

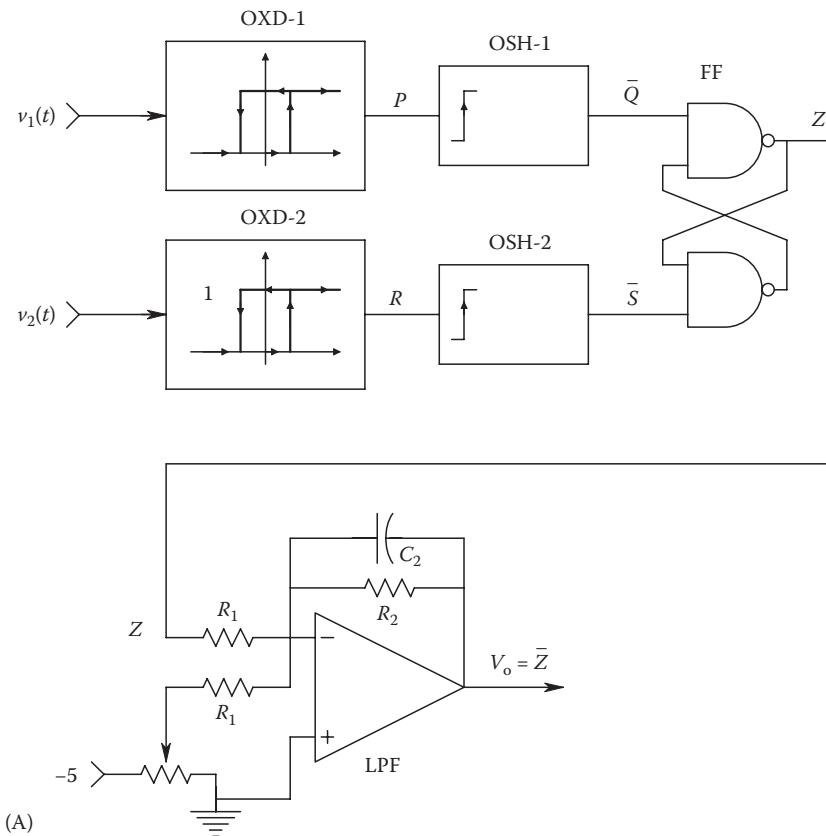


FIGURE 8.54
 (A) Schematic of a NAND gate, R-S flip-flop, analog phase detector. *Note:* OXD = analog/logical zero-crossing detector with hysteresis. OSH = one-shot multivibrator. FF = NAND gate R-S flip-flop. LPF = op amp LPF and DC level setting.

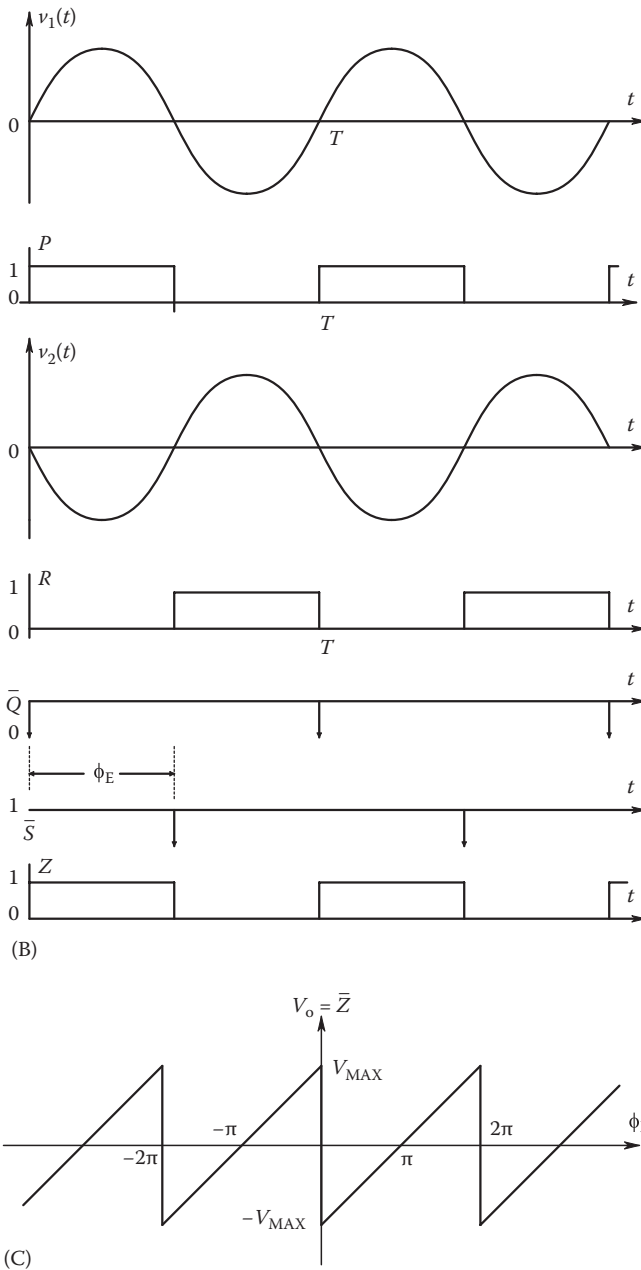


FIGURE 8.54 (continued)

(B) Waveforms in the phase detector. (C) Note the periodic form of this phase detector’s transfer curve. It has zeros at odd multiples of 180° phase difference.

duty cycle when ϕ is 180°. The FF output, $Z(t)$, is inverted and averaged by the OA LPF, and the DC level is adjusted so that when $\phi = 180^\circ$, the OA output voltage \bar{Z} is zero. A plot of \bar{Z} vs. ϕ is shown in Figure 8.54C; note that it is periodic.

An exclusive NOR gate can also be used for an analog phase detector. The input signals are converted to TTL square waves for input to the ENOR gate; the gate’s output is

a double-frequency TTL pulse train whose duty cycle varies linearly with the phase difference, φ . As in the case of the RSFF phase detector, the ENOR's output, $W(t)$, is averaged by low-pass filtering, and the DC level is adjusted. The DC OA output, W , is also a periodic function of φ .

A third logic-based phase detector IC utilizes NAND gates. Figure 8.55A illustrates the simplified circuitry of the *Motorola MC4044* phase detector and shows how its two outputs can be connected to an OA circuit that does low-pass filtering and DC adjustment. The I/O characteristics of the 4044 phase detector are shown in Figure 8.55B. Note that unlike the analog phase detectors described earlier, the output of the 4044 phase detector system is zero when $\varphi = 0$, and the linear range extends to a full $\pm 360^\circ$.

Other analog phase detectors are of the *switched-amplifier design*. Such switched detectors are also of the quadrature type. In effect, the signal $v_2(t) = V_2 \cos(\omega_1 t + \varphi)$ is multiplied by the function, $\text{SGN}[v_1(t)]$. $\text{SGN}[x]$ is the well-known signum function; $\text{SGN}[x < 0] = -1$, $\text{SGN}[x \geq 0] = +1$. Thus,

$$Z(t) = K_A \text{SGN}[v_1(t)] V_2 \cos(\omega_1 t + \varphi). \quad (8.132)$$

From a consideration of the Fourier series for the $\text{SGN}[v_1]$ square wave, it is easy to show that for $\varphi < 12^\circ$, the average of $Z(t)$, $\bar{Z} = K_D \varphi$. An example of a switched-amplifier phase detector is illustrated in Figure 8.56.

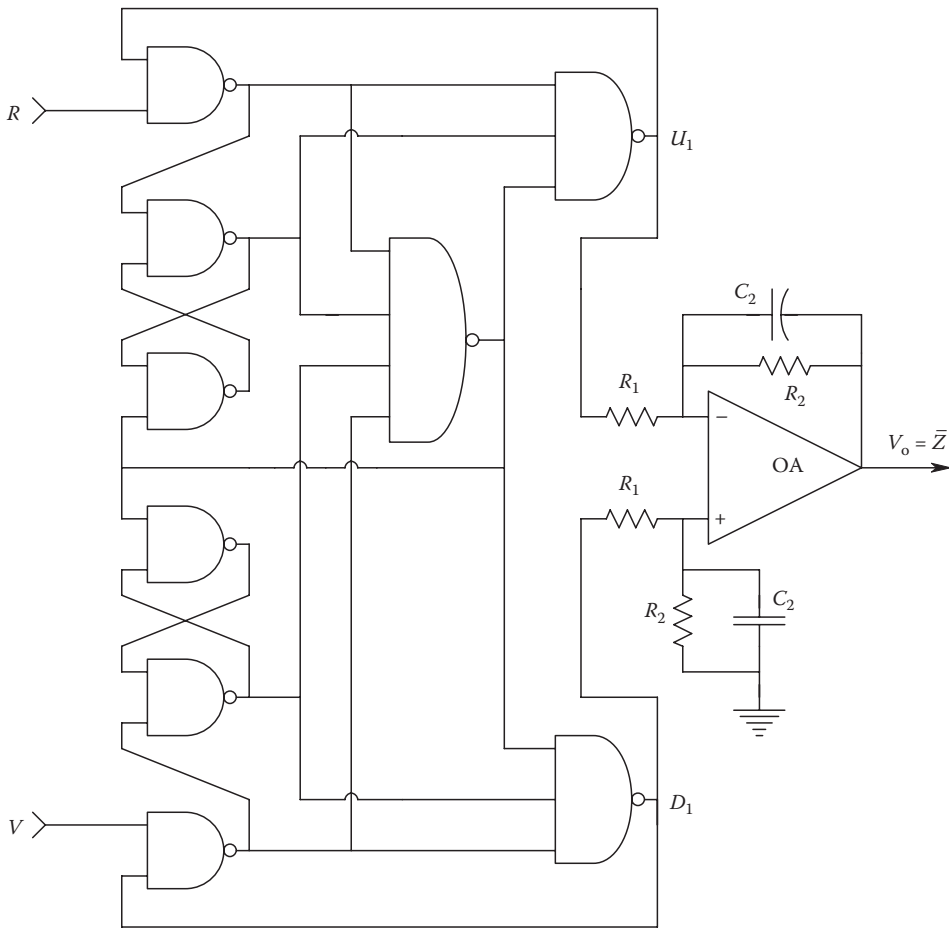
8.8.2 Digital Phase Detectors

We have seen earlier that analog phase detectors all generate waveforms whose average values are proportional to the phase shift, φ . Digital phase meters, on the other hand, measure the time interval, ΔT , as set forth in Equation 8.125b. The time interval is measured by counting clock pulses generated from a precision quartz crystal oscillator between successive, positive-going, zero crossings of v_1 and v_2 . For example, the Stanford Research Systems Model SR620 Universal Time Interval Counter uses a precision 90 MHz clock to count the intervals between the successive positive zero crossings of v_1 and v_2 . It also measures the period, T_1 , of v_1 . An internal microprocessor system then computes and displays

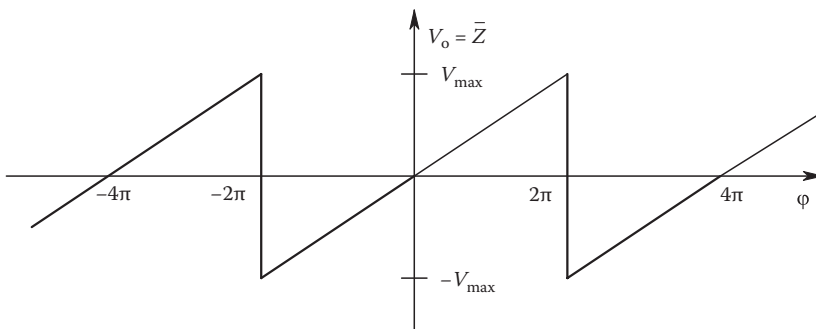
$$\varphi = 360^\circ \Delta T / T_1 \text{ degrees.} \quad (8.133)$$

The SRS SR260 phase meter works from 0 to 100 MHz and has a total range of $\pm 180^\circ$, a resolution of $[25 \times 10^{-12} \times (1/T_1) \times 360 + 0.001]^\circ$, and an error of less than $\pm [10^{-9} \times (1/T_1) \times (360 + 0.001)]^\circ$. The SR620 permits averaging a preset number of input cycles to compute a mean φ for greater accuracy in the presence of phase noise. The SRS SR620 is an example of a state-of-the-art instrument having an IEEE488 instrument/computer interface and internally generated display voltages so that in the absence of a host computer, the instrument output and status can be displayed on any XY lab oscilloscope.

We next describe the millidegree phase meter of Du and Northrop (Du 1993). This system was described in detail in Section 7.3.2 and is illustrated in Figure 7.60. A phase-locked loop was run from the reference input signal as a phase-locked frequency multiplier. Originally designed to be used in an optical system in which the chopping frequency was



(A)



(B)

FIGURE 8.55

(A) Combinational logic used in the Motorola MC4044 CMOS digital phase detector. (B) I/O characteristic of the MC4044 phase detector. Note that its zeros are at multiples of 4π radians and its linear range spans $\pm 2\pi$ radians.

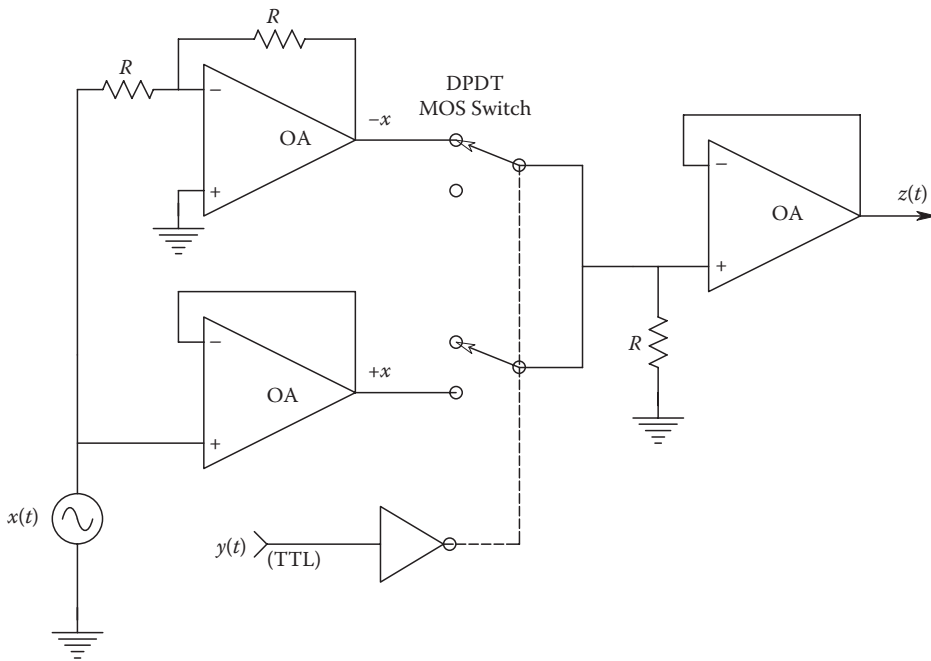


FIGURE 8.56

A switched amplifier phase detector. This is a quadrature detector with zeros at odd multiples of 90° phase difference between the sine wave input at $x(t)$ and the TTL reference input, $y(t)$. The output $z(t)$ must be low-pass filtered to generate a V_o proportional to $\sin(\theta_E)$, where θ_E is the phase difference between a cosine wave applied at x and a sinusoidal square wave at y .

100 Hz, the Du and Northrop system generated a clock frequency that is 360,000 times the input frequency, that is, 36 MHz. Because of the 360,000:1 frequency multiplication of the PLL, there will always be 1 clock pulse per millidegree of input phase, regardless of any small drift in input frequency. The HF PLL clock pulses are gated to a cascade of three, high-speed, TTL counters (74F579) during the interval, ΔT , where $v_1 > 0$ and $v_2 \leq 0$. The total count is obviously equal to ϕ in millidegrees. To improve accuracy in the presence of phase noise, the cumulative counts for $N = 2^k$ intervals, ΔT_{nr} , can be averaged ($k = 0, 1, \dots, 8$, and $n = 1, 2, \dots, N$). The upper bound on the input frequency of the Du and Northrop phase measuring system was set by the PLL VCO (50 MHz for an NE564) or the counters (115 MHz for 74F579s). Also, the range of frequency drift tolerated by the system was set by the lock range of the PLL. Another system limitation was set by counter overflow; a ϕ over 46.6° will cause the counters to exceed their maximum count of 2^{24} .

8.9 Measurements of Frequency and Period (Time)

There are two basic definitions of frequency: average frequency, measured as events (e.g., zero crossings) per unit time, and instantaneous frequency, measured as the reciprocals of the time intervals between successive events. Most modern frequency meters are digital systems that measure average frequency by counting events (cycles) over an accurately determined gate time. However, there are several other less accurate means of measuring frequency that

may be used. Some of these methods compare the unknown frequency with a frequency standard by some means. These include the Lissajous figure viewed on an oscilloscope, the zero-beat heterodyne method, the FFT spectrum analyzer, and the tunable filter.

The oscilloscope can be used to measure frequency by comparing the unknown frequency to an accurate, variable-frequency standard by a direct observation of the Lissajous figure on its CRT. The Lissajous pattern is made by putting the waveform with the unknown frequency on the y -axis (vertical) and the known frequency waveform on the x -axis (horizontal). When the known frequency is adjusted so that the observed pattern is either a stationary straight line, an ellipse, or a circle, the known frequency equals the unknown frequency, exactly. For obvious reasons, the Lissajous method works best at audio and low radio frequencies.

The zero-beat method also makes use of a variable-frequency, standard sinusoidal source that is mixed with the unknown frequency signal. Recall that mixing, in its purest form, is multiplication. Mixing can also be accomplished by adding the two signals together and then passing them through a nonlinear amplitude transfer function, such as $y = a + bx + cx^2 + dx^3$. The output from the mixing process can be shown to contain sine waves having frequencies that are the sum and difference of the two input frequencies. Thus, as the variable, known frequency is adjusted close to the unknown frequency, the frequency of the output difference frequency term approaches zero. This zero-beat phenomenon can be detected by actually listening to the audio beat frequency resulting from mixing the standard frequency source with the unknown. The beat frequency method has an apparent dead zone where the mixed difference frequency lies below the human range of hearing. Visual observation of the beat frequency signal on an oscilloscope or DC voltmeter can partially overcome the dead zone problem.

Another means of estimating the frequency of a coherent signal is to observe its rms spectrum on the display of an FFT spectrum analyzer. In such instruments, the frequency resolution is determined by the sampling frequency and the number of samples. Resolution can be as high as one part in a 1024 point transform and is obviously lower for fewer points.

The *Wien bridge*, which was discussed in Section 5.5.1.5, is a means of measuring capacitance. The Wien bridge also is unique in that its output also exhibits a null at its tuned frequency, behaving like a passive notch filter. When the bridge is configured as shown in Figure 8.57, it is easy to show that the output null occurs when $f = 1/(2\pi RC)$ Hz. Again, the null can be detected at audio frequencies by headphones or by an AC voltmeter. Typically, capacitors $C_1 = C_2 = C$ are varied together to obtain null. Alternately, the resistors $R_1 = R_2 = R$ can be varied together

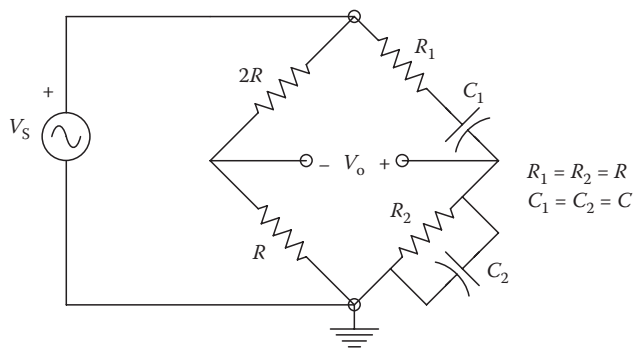


FIGURE 8.57
A Wien bridge used as a frequency-sensitive circuit.

with the Cs constant. The Wien bridge method's accuracy depends on the sensitivity of the null detector and the calibration of bridge components. It is best used at audio frequencies.

Fundamental to all frequency and period measurements is the requirement for a very accurate frequency source. In most benchtop instruments, the frequency reference is obtained from a quartz crystal oscillator operated under conditions of strict temperature control in an oven generally kept to $\pm 0.01^\circ\text{C}$ or better. For example, the HP 105A/B quartz secondary frequency standard has buffered sinusoidal outputs at 5, 1, and 0.1 MHz and a short-term stability measured at 5 MHz of 5 parts in 10^{12} (measured over a 1s averaging time). The *aging rate* is $<5 \times 10^{-10}$ per 24 h for this oscillator.

Accurate, variable, reference frequencies can be derived from accurate, fixed-frequency oscillators by several means. Direct frequency multiplication uses the generation of harmonics of the reference frequency waveform. The reference waveform is passed through an amplitude comparator to generate a TTL or ECL logic waveform of the same frequency. This logic waveform is then passed through a one-shot multivibrator to create a train of narrow pulses having a Fourier spectrum containing many harmonics. The pulse train is then made the input to a high- Q tuned circuit, which has its center frequency equal to that of the desired harmonic. It is practical to recover harmonics up to order 9 with this technique. One can also use synchronous binary logic *rate multipliers*, such as the 7497, to reduce the oscillation frequency to f_R , given by the relation,

$$f_R = f_o \left(\frac{M}{64} \right) \quad (8.134)$$

where

f_o is the oscillator frequency

M is a six-bit binary number, for example,

$$M = D_0(2^0) + D_1(2^1) + D_2(2^2) + D_3(2^3) + D_4(2^4) + D_5(2^5) \quad (8.135)$$

and $D_k = 1$ or 0. More complex frequency synthesis techniques make use of two or more interconnected phase-lock loops, mixers, rate multipliers, and crystal oscillator sources. Figure 8.58 illustrates an example of a vernier loop synthesizer. A detailed treatment of PLL frequency synthesizers may be found in the texts by Egan (1981) and by Kinley (1980). The HP/Agilent Model 8360B series of microwave frequency-synthesized sweep generators makes use of a single, high-accuracy, 10 MHz quartz crystal reference source and uses three PLLs and a feedback scheme to obtain 1 Hz resolution over a range of 2 MHz to 40 GHz. The HP 8360B synthesizer architecture is shown in Figure 8.59. The four subsystems in this synthesizer are a 2–7.8 GHz YIG-tuned oscillator (YTO), a reference PLL, a fractional- N PLL, and a sampler loop. A portion of the RF output from the YTO is sent to a sampler where it is mixed with a high-order harmonic of the 200–220 MHz PLL forming a 20–40 MHz IF signal. A phase detector compares this IF signal to the output of the fractional- N PLL. This phase detector output is summed into the tuning control for the YIG oscillator. The sampler loop can be incremented in 500 kHz steps, and the fractional- N loop can be stepped in 0.001 Hz increments. This mode of operation permits this family of microwave frequency synthesizers to maintain 1 Hz resolution over their entire output range.

The *primary time/frequency standard in current use is the cesium 133 atomic beam clock*. A diagram of the cesium clock is shown in Figure 8.60. This standard was discussed in detail

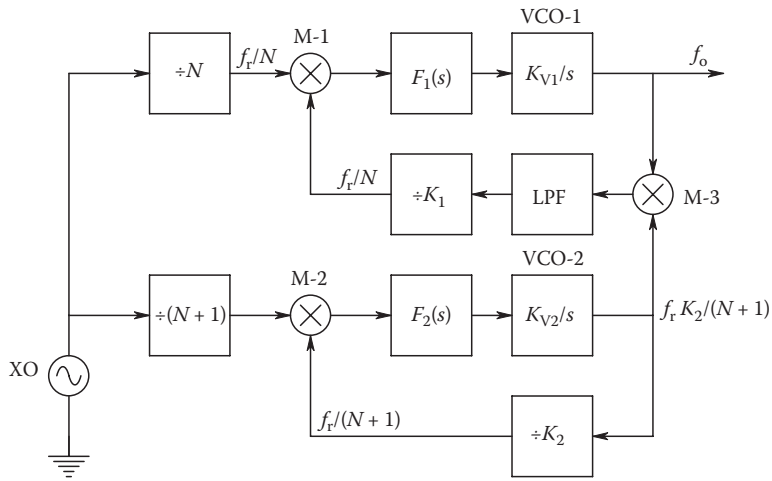


FIGURE 8.58
A phase-lock, vernier loop, frequency synthesizer.

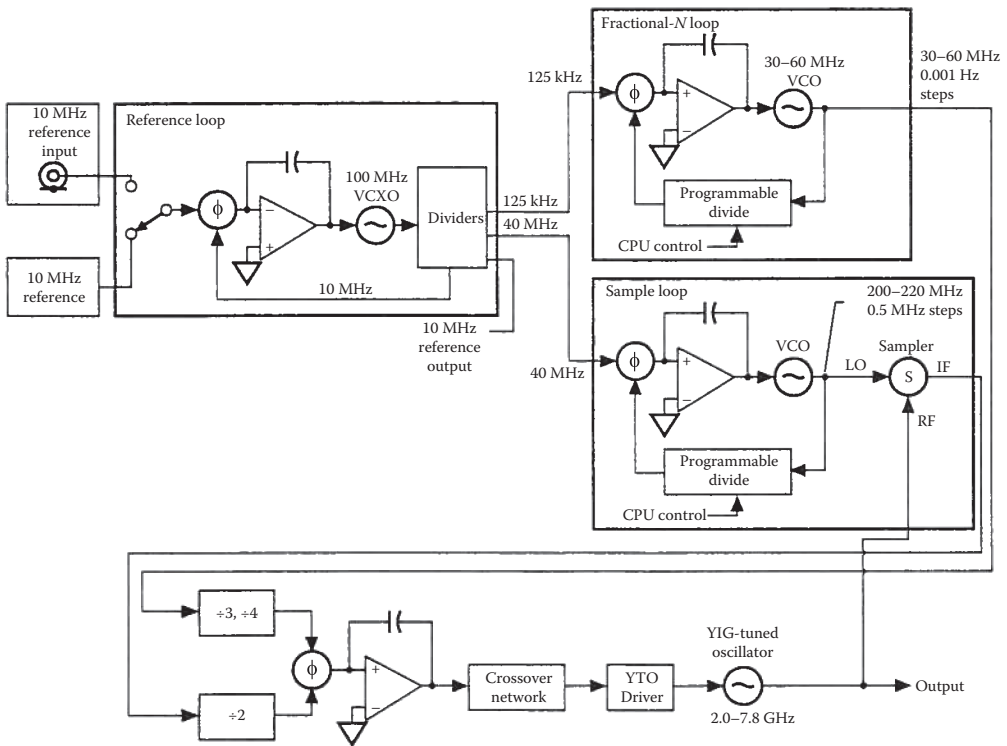


FIGURE 8.59
Systems block diagram of an HP8360 frequency synthesizer. (Courtesy of Hewlett-Packard, Palo Alto, CA.)

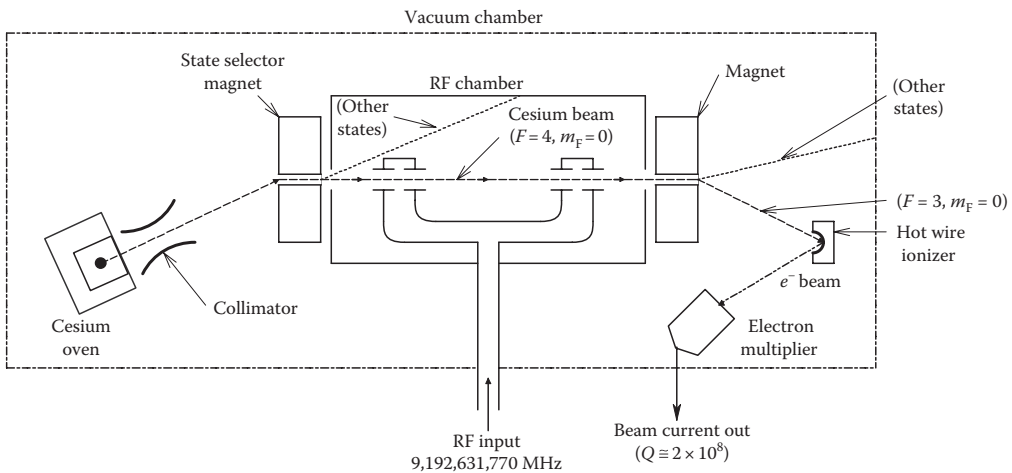


FIGURE 8.60

Diagram of a cesium beam frequency standard clock. (Adapted from Oliver, B.M. and Cage, J.M., *Electronic Measurements and Instrumentation*, McGraw-Hill, New York, 1971. With permission.)

in Chapter 2. To summarize the cesium clock's properties, it oscillates at 9,192,631,770 GHz, having an effective Q of 2×10^8 . *Coordinated universal time* (UTC), whose basic unit is the second, is defined as 9,192,631,770 periods of the cesium 133 beam oscillator. This international standard was adopted in October 1967. The HP/Agilent HP5061B(Opt 004) cesium beam frequency standard has a long-term stability of $\pm 2 \times 10^{-12}$ over the life of the cesium beam tube. Accuracy is also $\pm 2 \times 10^{-12}$. The HP5051B cesium clock has sinusoidal outputs at 10, 5, 1, and 0.1 MHz. HP cesium clocks are used to calibrate and synchronize the SATNAV, Omega LORAN-C, and GPS radio navigation systems for land vehicles, boats, and aircraft. Cesium beam clocks are ordinarily used to adjust rubidium and quartz oscillators used as secondary standards for frequency or period determination.

Rubidium frequency standards are second in the hierarchy of accuracy. Similar to the operation of a cesium beam clock, the atomic resonance of a rubidium vapor cell is used to synchronize a quartz crystal oscillator in a frequency-lock loop. The long-term stability of the rubidium vapor oscillator is $\pm 1 \times 10^{-11}$ /month. It, too, has outputs at 5, 1, and 0.1 MHz.

The two principal modes of operation of universal counter-timer instruments, such as the venerable HP5345A, are frequency measurement by counting events such as positive zero crossings of the signal with unknown frequency over the duration of an accurately determined gate window. The duration of the gate window is determined by logic circuits that count a preset number of internal master clock cycles. The gate dwell can be set from 1000 s to 100 ns in the HP5345A counter. The frequency range on channel A is 0.05 mHz to 500 MHz, and periods can be measured from 2 ns to 2×10^4 s. The master quartz clock in the HP5345A counter-timer runs at 10 MHz and has a long-term stability of $< \pm 5 \times 10^{-10}$ /day and a 1 s stability of $< \pm 1 \times 10^{-11}$. Figure 8.61 illustrates the Hz uncertainty of the HP5345A counter-timer as a function of applied frequency, input noise, and gate time. Note that the measured frequency uncertainty increases with input noise and input frequency and decreases with gate time. Counter-timer instruments can also be operated to measure the time, δT , between positive (or negative) zero crossings of two periodic inputs, thus providing phase information. They also can be run as accurate timers to measure the time between input trigger pulses.

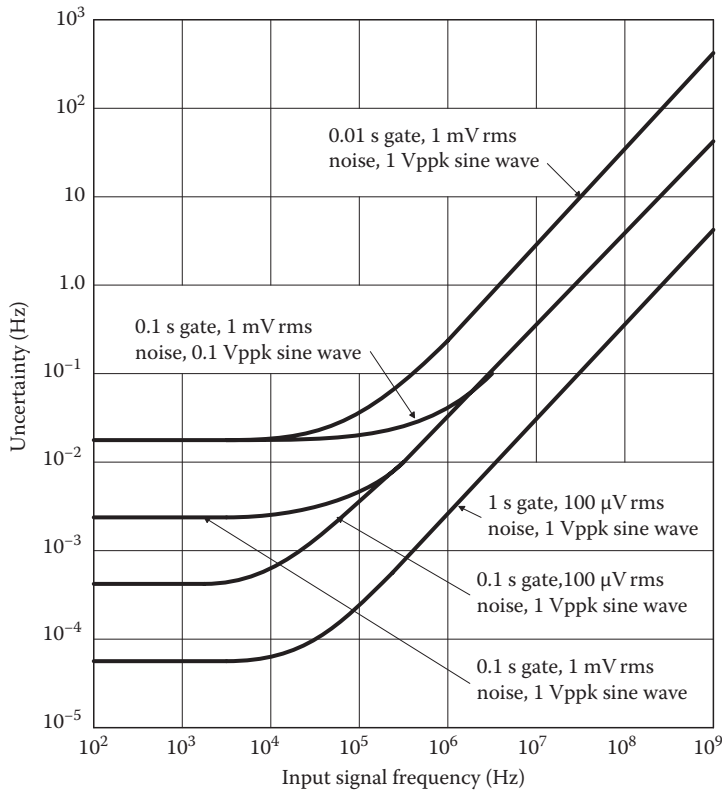


FIGURE 8.61

Frequency resolution error of the HP5345A counter-timer instrument. Noise accompanying the input signal and internal uncertainties affect the accuracy of frequency and period measurements. (Courtesy of Hewlett-Packard, Palo Alto, CA.)

Instantaneous frequency measurements are used to describe the unevenness or *frequency noise* in a periodic signal. An example of such a noisy frequency source is the human electrocardiogram (Northrop et al. 1967). Many physiological factors control and modulate the rate of the heartbeat; these include blood pCO₂, emotions, and exercise. Other applications of instantaneous frequency measurements have been to characterize nerve impulse sequences (Northrop and Horowitz 1966) and the variations in the rotational speed of a turbine.

The IPFD was described in Section 7.2.3, and two circuits for IPFDs are shown in Figures 7.15A and 7.16. The *j*th element of instantaneous frequency, *r_j* is defined as the reciprocal of the *j*th period or interevent interval. That is,

$$r_j \equiv \frac{1}{t_j - t_{j-1}}, \quad j = 2, 3, \dots \tag{8.136}$$

where

t_j is the time of occurrence of the *j*th event

t_{j-1} is the time of occurrence of the next previous event

Any periodic waveform can thus be characterized in the time domain by a number sequence, $\{r_i\}$, which describes the variation of frequency of the waveform on an interval by interval basis. As seen in the following equation, the instantaneous frequency number sequence can be converted to a stepwise analog voltage output, given by

$$q(t) = k \sum_{i=2}^{i=\infty} \left[\frac{1}{t_i - t_{i-1}} \right] [U(t - t_i) - U(t - t_{i+1})] \quad (8.137)$$

where $U(t - a)$ is the unit step function, defined as 0 for $t < a$ and 1 for $t \geq a$.

IPFDs can be of either analog or digital design and have generally found applications in characterizing the irregularities of frequency in audio-frequency waveforms.

In summary, we see that there are a variety of methods that can be used to measure or to estimate a signal's frequency, which can range from 10^{-3} Hz (mHz) to the hundreds of GHz ($>10^{11}$ Hz). Frequency measurements can be made by direct counting of events per unit time or by comparison with an accurate, calibrated oscillator. At very low frequencies, it is more effective to measure the period of the unknown signal, rather than to count its cycles over some fixed interval. Averaging of either period or frequency measurements improves accuracy.

8.10 Measurement of Resistance, Capacitance, and Inductance

In this section, we will examine and review the various means of measuring resistance, capacitance, and inductance. We have already examined null methods of measuring R , C , and L at audio frequencies using various bridge circuits in Chapters 4 and 5. These methods will be referred to but not repeated here. Rather, we will examine various active and passive ohmmeter circuits, means of characterizing linear and nonlinear (voltage-variable) capacitances with DC and HF AC, and means of measuring the properties of inductances at high frequencies.

8.10.1 Resistance Measurements

Techniques have been developed to measure resistances from 10^{-7} to over 10^{14} Ω . Needless to say, at the extreme ends of this range, specialized instruments must be used. We have already seen in Chapter 4 that very accurate resistance measurements are commonly made using DC Wheatstone or Kelvin bridges and a DC null detector such as an electronic nanovoltmeter. The values of the resistances used in these bridges' arms must, of course, be known very accurately. In the following texts, we shall discuss other DC means of measuring resistance.

The voltmeter–ammeter method is probably the most basic means of measuring resistance. It makes use of Ohm's law and the assumption that the resistance is linear. As shown in Figure 8.62, there are two basic configurations for this means of measurement: the ammeter being before R_X , which is in parallel with the voltmeter, and the ammeter being in series with R_X after the voltmeter. In the first case, the ammeter measures the current in the voltmeter as well as R_X ; in the second case, the voltmeter

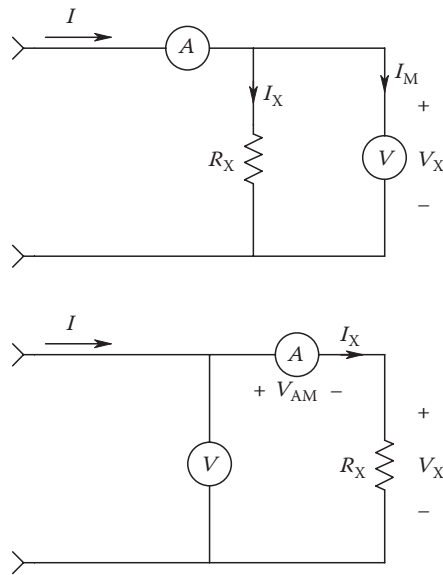


FIGURE 8.62
Two circuits that can be used to measure resistance using a voltmeter and ammeter.

measures the voltage drop across the ammeter plus that across R_X . It is easy to show that in the first case, R_X is given by

$$R_X = \frac{V_X}{I - V_X/R_{VM}} \tag{8.138}$$

where

- V_X is the voltmeter reading
- I is the ammeter reading
- R_{VM} is the resistance of the voltmeter (ideally, infinite)

In general, $R_{VM} = \eta V_{FS}$, where $\eta = \Omega/V$ sensitivity of the voltmeter, and V_{FS} is its full-scale voltage range.

In the second case, we find

$$R_X = V/I_X \mp R_{AM} \tag{8.139}$$

where

- V is the voltmeter reading
- I_X is the ammeter reading (current through R_X)
- R_{AM} is the resistance of the ammeter

The first method is best for measuring low values of R_X where $R_{VM} \gg R_X$, and the second method has less error for $R_{AM} \ll R_X$. Obviously, R_X must be found by calculation, and the accuracy of the result depends on the accuracy of the meters.

The voltmeter–ammeter method of measuring resistance can be used to measure very low resistances, on the order of micro-ohms. An electronic current source is used to pass a known amount of DC through the unknown, low resistance. An electronic, DC nanovoltmeter is connected across R_X . Obviously, V_M is directly proportional to R_X .

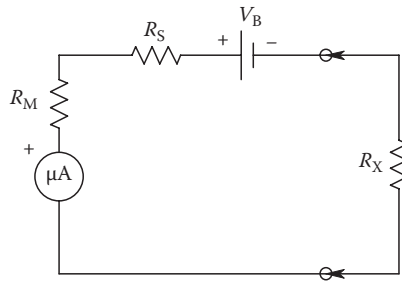


FIGURE 8.63
A simple series ohmmeter.

Common, passive, benchtop volt-ohm-milliammeters (multimeters) use a simple series ohmmeter circuit, shown in simplest form in Figure 8.63, and have a nonlinear (hyperbolic) scale on the D'Arsonval microammeter with zero ohms full-scale and infinite ohms at zero deflection. A typical benchtop VOM series ohmmeter will have five R_H ranges (R_H is the indicated R_X at half-scale meter deflection), ranging typically from 12 Ω to 120 k Ω . Inspection of the simple circuit shows that the DC meter current is given by

$$I_M = \frac{V_B}{(R_M + R_S) + R_X}. \quad (8.140)$$

By Ohm's law, the full-scale meter current is

$$I_{M(\text{FS})} = \frac{V_B}{R_M + R_S}. \quad (8.141)$$

We note that $(R_M + R_S)$ is the Thevenin resistance the resistor R_X sees. It is convenient to define $(R_M + R_S)$ as R_H , the half-deflection resistance. That is, when $R_X = R_H$, $I_M = I_{M(\text{FS})}/2$. Further analysis of the simple series ohmmeter is made simpler by defining the ohmmeter's fractional meter deflection, F , as

$$F = \frac{\theta}{\theta_{\text{FS}}} = \frac{I_M}{I_{M(\text{FS})}} = \frac{R_M + R_S}{R_M + R_S + R_X} = \frac{1}{1 + R_X/R_H}. \quad (8.142)$$

Here, θ is the meter deflection angle, θ_{FS} is the full-scale deflection angle, and the R s are as shown in Figure 8.63.

To find the most sensitive part of the series ohmmeter's scale, we define the ohmmeter's sensitivity, S , as dF/dR_X . This derivative is easily found from Equation 8.142 earlier. It is

$$S = \frac{dF}{dR_X} = \frac{R_H}{(R_H + R_X)^2}. \quad (8.143)$$

S clearly has a maximum with respect to R_H . If we find dS/dR_H and set it equal to zero, the peak in the sensitivity is seen to occur when $R_H = R_X$. That is, the center of the series ohmmeter's scale is the most sensitive (and most accurate) part.

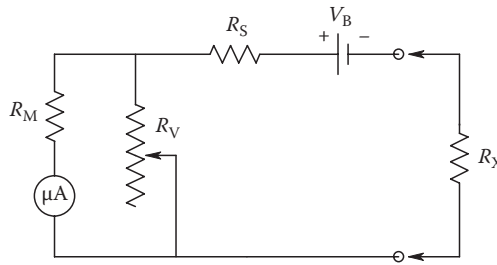


FIGURE 8.64

A practical series ohmmeter circuit. R_V is used to set the ohmmeter to read zero with its input leads short-circuited together to compensate for changes in the value of V_B .

Because the voltage, V_B , of the series ohmmeter’s battery drops as the battery is used and ages, a variable resistor, R_V , is added in parallel with the microammeter to compensate for the drop in V_B . This practical series ohmmeter circuit is shown in Figure 8.64. R_V is adjusted with the meter leads shorted ($R_X = 0$) so $I_M = I_{MFS}$. In this circuit, R_H is still the Thevenin resistance R_X sees; this is

$$R_H = R_S + \frac{R_M R_V}{R_M + R_V}. \tag{8.144}$$

An increase of R_V to compensate for a drop in V_B will thus change R_H from its design value and cause a small meter calibration error.

Shunt ohmmeters are less frequently encountered than are series ohmmeters. Shunt ohmmeters are used to measure low resistances and have R_H s ranging from 0.5 to 50 Ω . They are primarily used to measure resistances associated with coils such as motor armature windings. The circuit of a shunt ohmmeter is shown in Figure 8.65. In this circuit, the battery must supply a substantial current, often in the ampere range. Consequently, a robust battery, such as a lead–acid motorcycle-type battery, is used. As in the case of the series ohmmeter, we can define the half-deflection resistance, R_H , as the Thevenin resistance seen by R_X :

$$R_H = \frac{R_M R_S}{R_M + R_S}. \tag{8.145}$$

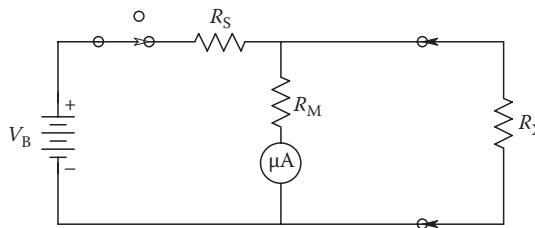


FIGURE 8.65

Circuit of a simple shunt ohmmeter.

Full-scale meter current occurs when $R_X = \infty$ and is simply

$$I_{M(\text{FS})} = \frac{V_B}{R_M + R_S}. \quad (8.146)$$

In general, the meter current is given by

$$I_M = \frac{V_B R_X / (R_X + R_S)}{R_M + R_X R_S / (R_X + R_S)} = \frac{V_B R_X / (R_M + R_S)}{R_X + R_M R_S / (R_M + R_S)}. \quad (8.147)$$

Zero meter deflection occurs for $R_X = 0$ (short-circuited input). The fractional deflection factor of the shunt ohmmeter is found to be

$$F = \frac{\theta}{\theta_{\text{FS}}} = \frac{I_M}{I_{M(\text{FS})}} = \frac{R_X}{R_X + R_H}. \quad (8.148)$$

It can also be shown that the center of the shunt ohmmeter's scale is also the point of greatest sensitivity.

Both series and shunt ohmmeters have relatively low accuracy, generally 2%–5% at center scale, which is adequate for most noncritical applications, such as measuring a motor coil's resistance, checking circuit continuity, or verifying resistor values (for those persons who do not know the resistor color code).

We next examine several *electronic ohmmeter circuits*. The readout of such circuits can be either analog or digital. In general, their accuracy is an order of magnitude better than that obtained with a series or shunt ohmmeter. Electronic ohmmeters can measure resistances ranging from 0.1 $\mu\Omega$ up to $10^{18} \Omega$. Obviously, special techniques and instruments must be used to measure resistances at the extreme ends of the range cited earlier. The first electronic circuit we shall consider is the *normal-mode ohmmeter*, shown in Figure 8.66. The OA can be an IC, electrometer type, having a bias current on the order of 40 fA. A calibrated current source, easily made from a battery and a large resistor, injects a DC, I_S , into the unknown resistance, R_X . The voltage at the OA's non-inverting input is given by KCL:

$$V_1(G_X + sC_i) + I_B = I_S \quad (8.149)$$

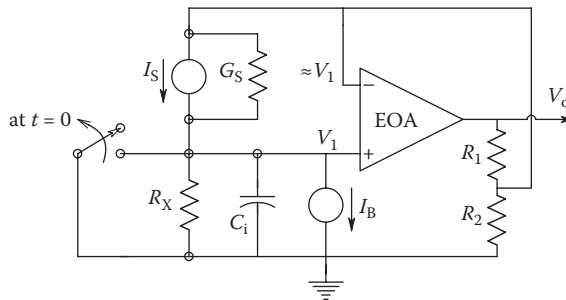


FIGURE 8.66

Schematic of a *normal-mode* electronic ohmmeter. I_S and G_S represent the Norton equivalent circuit of a floating current source. I_B is the electrometer OA's DC input bias current. C_i is the OA's input capacitance.

or

$$V_1 = \frac{(I_S - I_B)R_X}{(1 + sC_1R_X)} \tag{8.150}$$

It is easy to see that the OA's output, $V_o = V_1(1 + R_1/R_2)$, thus

$$V_o = (1 + R_1/R_2) \frac{(I_S - I_B)R_X}{(1 + sC_1R_X)} \tag{8.151}$$

If the switch is opened at $t = 0$, the input may be considered to be $(I_S - I_B)/s$, and $v_o(t)$ will rise exponentially with time constant C_1R_X to an SS value given by

$$v_{o(SS)} = (1 + R_1/R_2)(I_S - I_B)R_X \text{ V.} \tag{8.152}$$

Thus, the OA output voltage is directly proportional to R_X , and a linear ohms scale can be used on an analog output meter. Two problems arise with this circuit when measuring R_X s over $10^{10} \Omega$: the bias current must be kept much less than I_S , and if C_1 is on the order of several hundred pF, the ohmmeter's response time constant can become significantly long (10s or 100s of seconds), requiring excessive time to obtain an SS reading. If we assume that full-scale deflection of the normal-mode electronic ohmmeter is $v_{o(SS)} = 1 \text{ V}$, then full-scale resistances from 1 to $10^{10} \Omega$ can be measured with I_S values ranging from 10 mA to 1 pA, respectively, with $(1 + R_1/R_2) = 100$.

To measure resistors in excess of $10^{10} \Omega$, the *fast-mode electronic ohmmeter configuration*, shown in Figure 8.67, is generally used. Now, the unknown resistor is placed in the feedback loop of the OA. One end of R_X is at virtual ground at the summing junction, and the other end is connected to the V_2 node. The voltage V_2 is clearly equal to $-(I_S + I_B)R_X$. By KCL, we may write the following:

$$V_2(G_2 + G_1) - V_oG_1 - I_S = 0 \tag{8.153a}$$

$$-(I_S + I_B)R_X(G_2 + G_1) - V_oG_1 - I_S = 0 \tag{8.153b}$$

$$-I_S[R_X(G_2 + G_1) + 1] - I_BR_X(G_2 + G_1) = V_oG_1. \tag{8.153c}$$

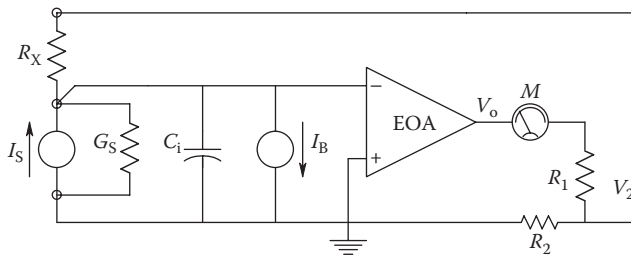


FIGURE 8.67 Schematic of a *fast-mode* electronic ohmmeter. Now R_X is placed in the OA's feedback loop. The summing junction is at virtual ground, and the current source is referenced to ground.

Assuming $R_x \gg R_1, R_2$, we can solve for the output voltage:

$$V_o = -R_x I_S (1 + R_1/R_2) - R_x I_B (1 + R_1/R_2). \quad (8.154)$$

The second term is negligible as long as $I_S \gg I_B$. If the output current is to be read by the microammeter, M , we note that

$$I_o = \frac{V_o - V_2}{R_1} = -I_o R_x (G_2 + G_1) + I_S R_x G_1 = -I_S R_x G_2. \quad (8.155)$$

For example, if $R_x = 10^{14} \Omega$, $I_S = 1 \text{ pA}$, and $R_2 = 10^4 \Omega$, then $I_o = -10 \text{ mA}$.

High-megohm resistors (1 G Ω or higher) are central to the operation of many electrometer instruments, including feedback picoammeters, as described in Section 8.4.2. Such resistors are typically made from metal oxides and are enclosed in sealed, glass tubes. The *voltage coefficient* of this type of resistor, defined as $(\Delta R/R)/V$, is typically less than 5 ppm up to 100 V applied voltage, and the tempcos of high-meg resistors are on the order of 100 ppm/ $^{\circ}\text{C}$ for $10^8 \Omega$ and 1000 ppm/ $^{\circ}\text{C}$ for $10^{11} \Omega$. High-meg resistors are delicate and must be protected from mechanical shock and moisture, and their glass envelopes must be kept free of dirt and fingerprint oils and salt by cleaning with pure alcohol after handling.

Measurement of high-meg resistors with fast-mode electronic ohmmeters is not as simple as just connecting R_x to the meter and taking the reading. The reading can be influenced by static and slowly changing electrostatic fields. Even on humid days, most persons have a DC electric field associated with their bodies, even when cotton clothing is worn. Such fields can influence the measurement of R_x , destroying accuracy. To uncouple R_x from DC electrostatic influences and to prevent pickup of powerline **E** and **B** fields by the measurement apparatus, R_x should be thoroughly shielded both electrostatically and magnetically. (The 60 Hz fields can induce voltages that may saturate the amplifier, giving incorrect DC readings of resistance.)

8.10.2 Capacitance Measurements

We have seen that accurate capacitance measurements can be made at audio frequencies (generally at 1 kHz) using a variety of bridge circuits, described in Section 5.5.1. It is often important to characterize capacitors at radio frequencies and at ultra-low frequencies, the latter range being important in the characterization of capacitor dielectrics and dielectrics used as electrical insulation. There are also voltage-variable capacitances associated with semiconductor *pn* junctions, which are important to measure. These include the capacitance of varactor diodes (reverse-biased *pn* junction diodes); the gate–source capacitance (C_{gs}); gate–drain capacitance (C_{gd}); drain–source capacitance (C_{ds}), C_{iss} , and C_{rss} of FETs; and the C_π and C_μ of BJTs. We will describe in the following some of the nonbridge means to measure capacitors and to characterize their equivalent circuits.

8.10.2.1 Q Meter Used for Capacitance Measurement

At radio frequencies, the equivalent circuit for a capacitor will include not only a resistance to account for dielectric power losses but also losses due to skin effect in the capacitor's leads. Equivalent inductances may also appear at VHF, UHF, and SHFs due to lead geometry and physical layout of the plates and dielectric. Figure 8.68 illustrates one such

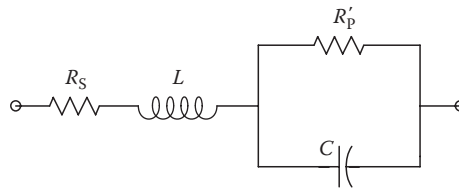


FIGURE 8.68
HF, lumped-parameter equivalent circuit of a practical capacitor.

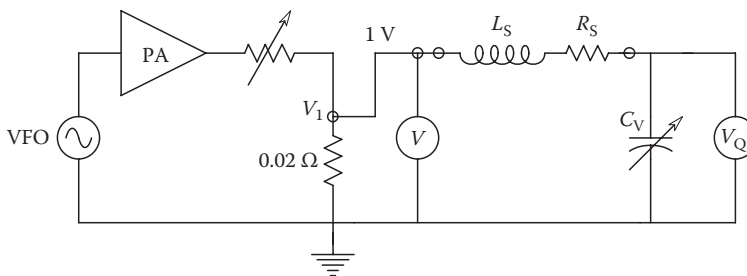
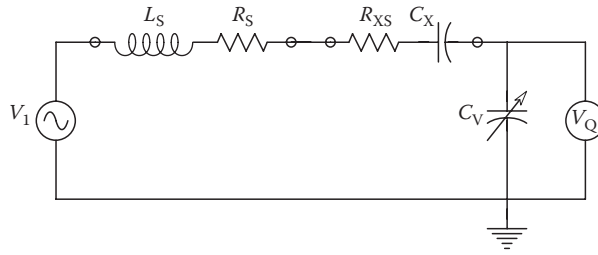


FIGURE 8.69
Circuit of a basic Q meter. *Note:* VFO = variable radio-frequency (RF) oscillator. PA = RF power amplifier. V and V_Q = HF, AC voltmeters.

equivalent circuit. A very practical instrument for measuring the net capacitance of a capacitor at radio frequencies is the Q meter. (The Q meter can also be used to measure inductor Q and inductance, as described in the next section.) There are several commercial Q meters: the venerable Boonton-type 260A covered the frequency range from 50 kHz to 50 MHz and the Boonton-type 190A Q meter covered from 20 to 260 MHz (Boonton was bought by HP in 1959). The venerable Marconi-type TF-1245 Q meter is used with separate voltage sources covering from 40 kHz to 50 MHz (type TF-1246) and from 20 to 300 MHz (type TF-1247). (Marconi was acquired by CMC Electronics, then sold to Esterline Corp. in 2007.) The HP-type HP4342A Q meter operates from 22 kHz to 70 MHz. The Q meter is basically a very simple circuit, consisting of a low-impedance RF voltage source, a variable capacitor, and an RF voltmeter. The Q meter is used with standard coils and capacitors in many of its measurement modes. A basic schematic of a Q meter is shown in Figure 8.69. A variable-frequency RF current is passed through a $0.02\ \Omega$, noninductive resistor. The RF voltage across this resistor appears to be from a nearly ideal RF voltage source, V_1 .

Measurement of an unknown capacitance, C_x , can be done in several ways. In the first method, C_x is placed across the internal capacitance, C_v , and a standard, series inductor of appropriate value is placed in series. The circuit is resonated at f_o and the capacitance of the variable capacitor is recorded as C_1 . Next, C_x is removed, and the variable capacitor again tuned to C_2 to resonate the circuit. It is easy to see that $C_x = C_2 - C_1$. This technique works as long as $C_x < \max C_v$. If C_x is larger than $\max C_v$, then a known, standard capacitor, C_s , can be placed in series with C_x to make the series combination less than $\max C_v$. Once the capacitance of the series combination of C_x and C_s is known, C_x can be calculated.

Cooper (1978) gave a more general method for measuring C_x and its equivalent series resistance, R_{xs} , on a Q meter. Refer to Figure 8.70, in which the unknown capacitor is placed in series with the standard inductor having inductance L_s and series resistance, R_s ,

**FIGURE 8.70**

Series Q meter circuit used to measure an unknown capacitance, C_x , at radio frequencies. See text for analysis.

As in the parallel technique described earlier, two measurements must be made: First, the unknown capacitor is shorted out and the Q meter is resonated with C_v at f_0 . The Q and C_v for this case are recorded as Q_1 and C_1 . Next, the short is removed and resonance is again observed, giving Q_2 and C_2 . Now in the first case, $X_C = X_L$, or

$$\frac{1}{2\pi f_0 C_1} = 2\pi f_0 L_S \quad (8.156)$$

and

$$Q_1 = \frac{2\pi f_0 L_S}{R_S} = \frac{1}{2\pi f_0 C_1 R_S}. \quad (8.157)$$

For the second measurement, it can be shown that the reactance of the unknown capacitor is given by

$$X_X = \frac{1}{2\pi f_0 C_X} = \frac{C_2 - C_1}{2\pi f_0 C_2 C_1}. \quad (8.158)$$

Hence,

$$C_X = \frac{C_1 C_2}{C_2 - C_1}. \quad (8.159)$$

The series resistive component of the unknown capacitance can be shown to be given by (Cooper 1978)

$$R_{XS} = \frac{C_1 Q_1 - C_2 Q_2}{2\pi f_0 C_1 C_2 Q_1 Q_2}. \quad (8.160)$$

The series dissipation factor of the unknown capacitor, $D_{SX} \equiv \omega C_X R_{XS}$, can be written from the earlier relations:

$$D_{XS} = \frac{C_1 Q_1 - C_2 Q_2}{Q_1 Q_2 (C_2 - C_1)}. \quad (8.161)$$

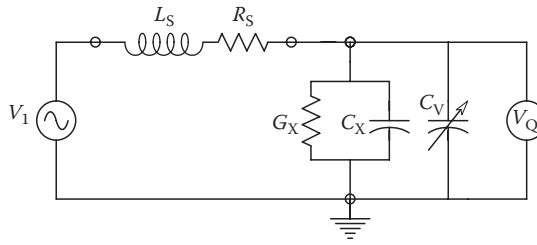


FIGURE 8.71
Parallel connection of the Q meter used to measure an unknown capacitance. See text for analysis.

It is also possible to place the unknown capacitor (impedance) in parallel with C_v , as shown in Figure 8.71. This configuration is used when $X_x > X_{C_1}$. As in the series case treated earlier, we first remove the unknown capacitor and resonate the Q meter by adjusting C_v . At resonance, $X_L = X_{C_1}$; thus,

$$C_1 = \frac{1}{\omega^2 L_S} \tag{8.162}$$

and

$$Q_1 = \frac{\omega L_S}{R_S} = \frac{1}{\omega C_1 R_S}. \tag{8.163}$$

Next, the unknown capacitor is placed in parallel with C_v , and C_v is adjusted to again obtain resonance. At resonance,

$$X_L = \omega L_S = \frac{X_{C_2} X_x}{X_{C_2} - X_x} = \frac{1}{\omega C_1}. \tag{8.164}$$

Equation 8.164 can be manipulated to yield

$$X_x = \frac{1}{\omega(C_1 - C_2)}, \tag{8.165}$$

from which it is easy to write

$$C_x = C_1 - C_2. \tag{8.166}$$

In addition, Cooper (1978) showed that the equivalent shunt conductance of the unknown capacitance can be written as

$$G_x = \omega C_1 \frac{Q_1 - Q_2}{Q_1 Q_2} \text{ S}, \tag{8.167}$$

and the dissipation factor of the parallel, G–C model of an unknown capacitor can be calculated from

$$D_{XP} = \frac{G_x}{\omega C_x} = \frac{C_1(Q_1 - Q_2)}{Q_1 Q_2(C_1 - C_2)} = \frac{C_1}{Q_1 Q_2} \frac{\Delta Q}{\Delta C}. \tag{8.168}$$

The Q meter is seen to be a versatile instrument for measuring the properties of unknown capacitors at radio frequencies. The accuracy is modest, however (1%–2%), and calculations are required to obtain the equivalent circuit parameters.

8.10.2.2 Capacitance Measurement by Q/V

Capacitance is defined by the basic relation, $C \equiv Q/V$. Hence, if one applies a DC voltage, V_C , to an initially uncharged capacitor, and then measures the total charge, Q , accumulated in the SS, the ratio of Q to V_C is, by definition, the capacitance in farads. This basic definition is used in the design of several commercial instruments used to measure capacitors. The Keithley Model 595 Quasistatic CV Digital Capacitance Meter uses a novel, stepwise programming of the voltage applied to the capacitor. A sequence of N steps of voltage, each step of height ΔV , where ΔV is selectable to be 0.01, 0.02, 0.05, or 0.10 V, can be applied to C_X . The maximum range of voltage is ± 20 V. As each step of voltage is applied to the unknown capacitor, an electrometer charge amplifier circuit integrates the current flowing into the capacitor as a result of each voltage step. The Model 595 meter samples the voltage proportional to charge at the integrator output just before each voltage step is applied, and it samples the charge amplifier's output twice at an interval ΔT after its output reaches SS from charging C_X . This sampling is shown in Figure 8.72. The OA's DC bias current may thus be estimated from

$$I_B = \frac{Q_3 - Q_2}{\Delta T} \tag{8.169}$$

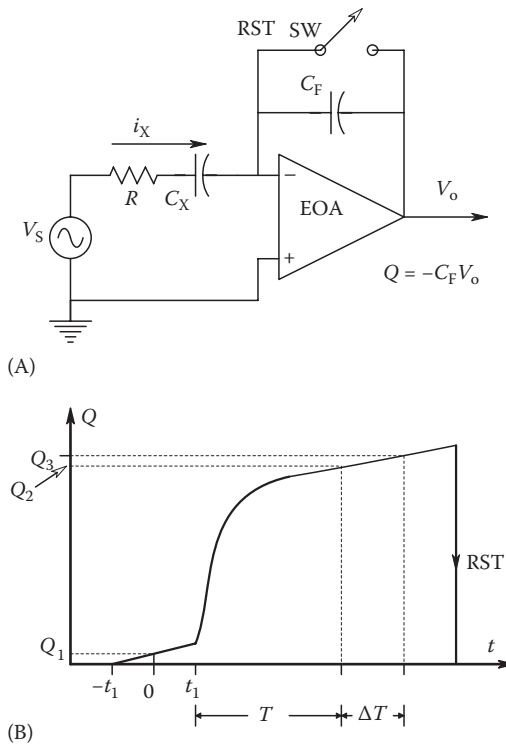


FIGURE 8.72 (A) Electrometer OA charge amplifier circuit used to measure voltage-variable capacitances by the $\Delta Q/\Delta V$ method. (B) Charge amplifier output waveform in response to one input voltage step, showing sampling points used in the Keithley Model 595 CV capacitance meter. See text for analysis.

This bias current must be subtracted from the current charging the capacitor under measurement. The charge amplifier's output can be shown to be

$$v_o(t) = -\frac{I_B}{C_F} - \frac{\Delta VC_X}{C_F} \left\{ 1 - \exp\left[\frac{-(t-t_1)}{RC_X} \right] \right\}. \tag{8.170}$$

Referring again to Figure 8.72 and to Equation 8.170, we see that in the SS, $C_X(V)$ is given by

$$C_X(V) = \frac{(Q_3 + Q_1) - I_B(T + t_1)}{\Delta V}. \tag{8.171}$$

C_X can also be expressed in terms of the charge amplifier's SS output voltage:

$$C_X(V) = \frac{C_F}{\Delta V} [v_{o(SS)} \text{ at } T + t_1] - \frac{I_B(T + t_1)}{\Delta V}. \tag{8.172}$$

The V reported by the Keithley 595 CV Meter is $(V_{LAST} + 0.5 \Delta V)$. If the capacitor is linear, then $C_X(V)$ is constant over the range of V . If the capacitance is a function of voltage, such as the capacitance of a reverse-biased pn junction, then $C_X(V)$ can be plotted vs. V with the Keithley 595 instrument in order to characterize the capacitor's voltage dependency. The Model 595 CV Meter can also be used to measure the junction capacitances of BJTs and FETs, as well as MOS chip capacitors. In spite of its digital display, sampling, microprocessor control, and IEEE-488 bus, the 595 is basically an analog instrument and has 1% accuracy.

Another application of the $C = Q/V$ method to characterize capacitor (and other) dielectrics has been described by Mopsik (1984). The procedure has been called *time-domain spectroscopy*, or TDS. The basic circuit for the *Mopsik system* is shown in Figure 8.73. An electrometer charge amplifier (integrator) is connected to a node between the capacitor and dielectric under investigation and a low-loss, reference capacitor, C_R . C_R is variable and has a dry air dielectric. The input to C_X is a positive step voltage, V_C , from an OA. $-V_C$ is applied simultaneously from another OA to C_R . The purpose of C_R is to permit a charging current equal and opposite to that of C_X to be summed at the charge amplifier's summing junction.

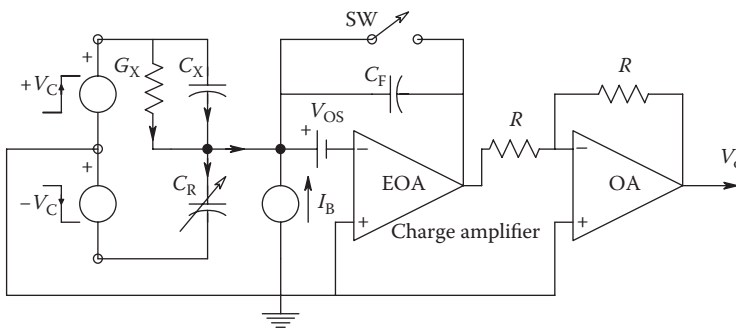


FIGURE 8.73
The Mopsik (1984) circuit for characterization of capacitances in the frequency domain.

In analyzing the Mopsik system, we will consider two cases: *the first*, where $C_X = C_R$, that is, there is perfect compensation, and the second, where $C_X \neq C_R$. The *second* case is most realistic, as perfect matching of C_X and C_R will not ordinarily be obtained in practice.

In the first case, there will be three components to the system output: one due to the DC leakage current through the equivalent conductance, G_X , shunting C_X . G_X describes the behavior of the C_X 's dielectric; G_X may have a purely ohmic component, as well as time- and voltage-dependent nonlinear components. Such behavior may arise from the presence of *free* electrons and holes in the dielectric, as well as mobile positive and negative ions, and impurity centers capable of *trapping* charged particles and ions until sufficient energy releases them. The other two components at the output are due to the OA's bias current, I_B , and offset voltage, V_{OS} . If we assume that C_F has zero charge at $t = 0$ when the steps of $\pm V_C$ are applied to the C_X and C_R , then we may write in the time domain:

$$v_o(t) = \int_0^t \frac{V_C}{C_F} G_X(V_C, t) dt + \frac{I_B}{C_F} t + V_{OS} U(t). \quad (8.173)$$

A plot of these components is shown in Figure 8.74. Note that in the case of a good, low-conductivity dielectric, I_B may exceed $V_C G_X$ in magnitude, making the bias current error a significant component at the output. V_{OS} can generally be nulled out at a particular operating temperature, however.

In the second case, $\Delta C = C_X - C_R \neq 0$, so there will also be a capacitive charging component to the output of the charge amplifier. If we assume that the capacitors charge instantaneously, then the charge amplifier output will be

$$v_o(t) = \int_0^t \frac{V_C}{C_F} G_X(V_C, t) dt + \frac{I_B}{C_F} t + \left(V_{OS} + \frac{V_C \Delta C}{C_F} \right) U(t). \quad (8.174)$$

Note that the ΔC term can be positive or negative.

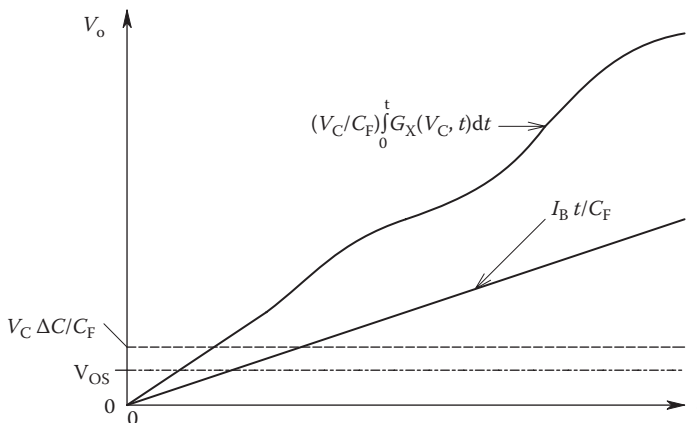


FIGURE 8.74

Plot of the electrometer charge amplifier's output voltage in the Mopsik system. The four components of text Equation 8.174 are shown.

In the SS, for a linear capacitor, we have the well-known relation, $C = Q/V_C$. Here, Q is the total charge accumulated in an ideal capacitor in the SS. Mopsik noted that the charge Q does not really reach an SS but continues to increase after the pure capacitance component, C_x , has charged to the applied DC voltage, V_C . The continuing increase in Q is due to dielectric phenomena. Mopsik defined a *time-variable capacitance* as

$$C(t) = \frac{Q(t)}{V_C} = \frac{v_o(t)C_F}{V_C}. \tag{8.175}$$

In order to characterize the so-called time-variable capacitor given by Equation 8.175 in the frequency domain, Mopsik takes the Fourier transform of the $v_o(t)$ transient, given by Equation 8.174, times C_F/V_C . This Fourier transform can easily be shown to be

$$\mathbf{C}^*(j\omega) = \frac{\mathbf{G}_x(j\omega, V_C)}{j\omega} + \frac{I_B}{V_C(j\omega)^2} + \left(\frac{V_{OS}C_F}{V_C} + \Delta C \right) \frac{1}{j\omega}. \tag{8.176}$$

Note that both $\mathbf{C}^*(j\omega)$ and $\mathbf{G}_x(j\omega, V_C)$ are complex, that is, they have real and imaginary components. Mopsik defined a complex $\mathbf{C}^*(j\omega)$ by Equation 8.177:

$$\mathbf{C}^*(j\omega) = C(\omega)' - jC(\omega)". \tag{8.177}$$

At a given frequency, we can show that the $\mathbf{C}^*(j\omega)$ model for the unknown capacitor is entirely equivalent to the standard, parallel R - C model. Assume an AC voltage, V_C , is impressed across \mathbf{C}^* and the parallel model. Thus, the current is

$$\mathbf{I}^* = V_C j\omega \mathbf{C}^*(j\omega) = V_C [j\omega C'(\omega) + \omega C''(\omega)] = \mathbf{I}_p = V_C [G_p + j\omega C_p]. \tag{8.178}$$

By comparing terms, we find that

$$C_p = C'(\omega) \tag{8.179a}$$

$$G_p = \omega C''(\omega). \tag{8.179b}$$

Referring to Figure 8.75, we see that the loss angle tangent, $\tan(\delta)$, at frequency ω must be the same for both capacitor models:

$$\tan(\delta^*) = \frac{C''(\omega)}{C'(\omega)} = \tan(\delta) = \frac{1}{\omega R_p C_p}. \tag{8.180}$$

$\tan(\delta)$ can also be expressed in terms of complex permittivity, ϵ^* . Here, we assume that

$$\mathbf{C}^*(j\omega) = \frac{A\epsilon^*(\omega)}{d} = \frac{A}{d} [\epsilon(\omega) - j\epsilon''(\omega)]. \tag{8.181}$$

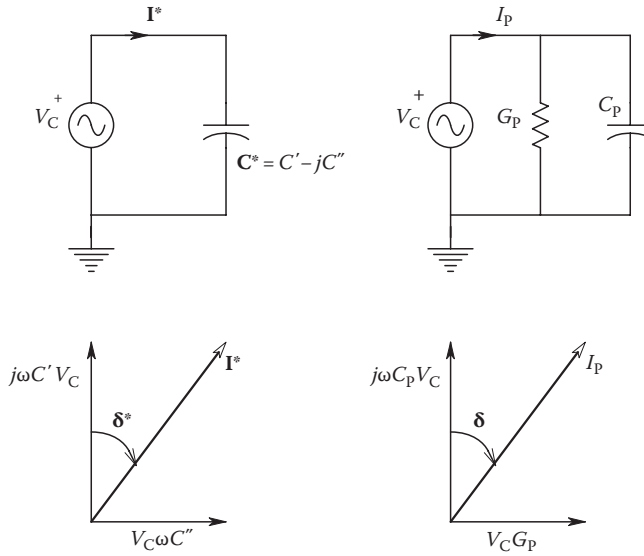


FIGURE 8.75 Equivalence of the loss tangent for the familiar, parallel R-C capacitance model, and the complex capacitance model of Mopsik, where $I_P = I^*$.

Thus,

$$\tan(\delta) = \frac{\epsilon''(\omega)}{\epsilon'(\omega)}. \tag{8.182}$$

The Mopsik system calculates $\tan[\delta(\omega)]$ for the unknown capacitor and dielectric from the ratio of $C''(\omega)$ to $C'(\omega)$. $C'(\omega)$ is the *real part* of $C^*(j\omega)$, which is found by taking the Fourier transform of $v_o(t)C_F/V_C$, as written in Equation 8.181 earlier. $C'(\omega)$ can be expressed as

$$C'(\omega) = RE\{C^*(j\omega)\} = \frac{IM\{G_X(\omega, V_C)\} - I_B / (V_C\omega)}{\omega}. \tag{8.183}$$

The *imaginary part* of $C^*(j\omega)$ is

$$C''(\omega) = IM\{C^*(j\omega)\} = \frac{RE\{G_X(j\omega, V_C)\} + V_{OS}C_F/V_C + \Delta C}{\omega}. \tag{8.184}$$

Equations 8.183 and 8.184 mentioned earlier illustrate the source of artifacts in the Mopsik system. Clearly, I_B can have a major effect on the $\tan[\delta(\omega)]$ result, as can the offset voltage and a nonzero difference between C_R and C_X . When we can assume that $C_R = C_X$, and V_{OS} and $I_B = 0$, then the loss tangent is

$$\tan[\delta(\omega)] = \frac{RE\{G_X(j\omega, V_C)\}}{IM\{G_X(j\omega, V_C)\}}. \tag{8.185}$$

Clearly, the Mopsik approach to characterizing a capacitor and its dielectric is unique. Because the unknown capacitance is assumed to be proportional to the charge, $q(t)$, it too will be a function of time and can be Fourier transformed. Hence, we have a complex capacitance that is a function of frequency and that will have real and an imaginary part. It is doubtful in practice that the $\tan[\delta(\omega)]$ calculated by the Mopsik method is independent of the electrometer's I_B . Such independence would be possible only if the current $V_C G_X \gg I_B$. A typical I_B for an SS, EOA is about 30 fA. Also, the initial step transient artifacts due to V_{OS} and ΔC will contribute error to the HF portion of $C^*(j\omega)$.

8.10.3 Inductance Measurements

Measurement of inductances can be made from subsonic to ultrahigh frequencies with a variety of means. Meaningful measurements can be made over the range of nanohenries to kilohenries but not at the same frequency. Circuits used to characterize inductances were introduced in Section 5.1, and audio-frequency inductance-measuring bridges were described in Section 5.5.2. Audio-frequency bridges are the most accurate means of measuring inductance at audio frequencies. In this section, we will describe two other means of measuring inductance: the AC voltmeter method and the Q meter. Inductances can also be measured with vector impedance meters (VIM), described in Section 8.10 in the following.

8.10.3.1 Voltmeter Method of Estimating Inductance

At powerline frequencies, it is possible to find the inductance of a coil by first measuring the coil's DC resistance, R'_L , with an ohmmeter. The actual real part of the series impedance of the coil at the measurement frequency will generally be slightly higher than R'_L due to magnetic core losses and skin effect at radio frequencies. The effective, AC, series coil resistance is defined to be R_L . A simple series circuit with the coil and a known resistor, R , is used (see Figure 8.76A). The only instrument needed is a high input impedance AC voltmeter. The excitation voltage (V_S), the voltage across the coil (V_L), and the voltage across the resistor (V_R) are measured. By Kirchoff's voltage law, we know that the vector sum of V_L and V_R must equal V_S . Also, we know that the phase angle of the current in a series inductive circuit *lags* the phase of the voltage across the circuit. We can thus draw the voltage vector (phasor) diagram as shown in Figure 8.76B. The three known voltages form the three sides of a triangle. The angle of V_R , $-\theta$, is also the angle of the current in the circuit, I . The angle of V_S is assumed to be zero. By Ohm's law, the magnitude of the current is

$$I = \frac{V_R}{R}. \quad (8.186)$$

The magnitude of the voltage across the coil can be written as

$$V_L = |\mathbf{I}\mathbf{Z}_L| = \frac{V_R}{R} \sqrt{\omega^2 L^2 + R_L^2}. \quad (8.187)$$

Equation 8.187 can be solved for L , since we know ω , R_L , R , V_R , and V_L :

$$L = \left(\frac{1}{\omega} \right) \sqrt{R^2 V_L^2 / V_R^2 - R_L^2}. \quad (8.188)$$

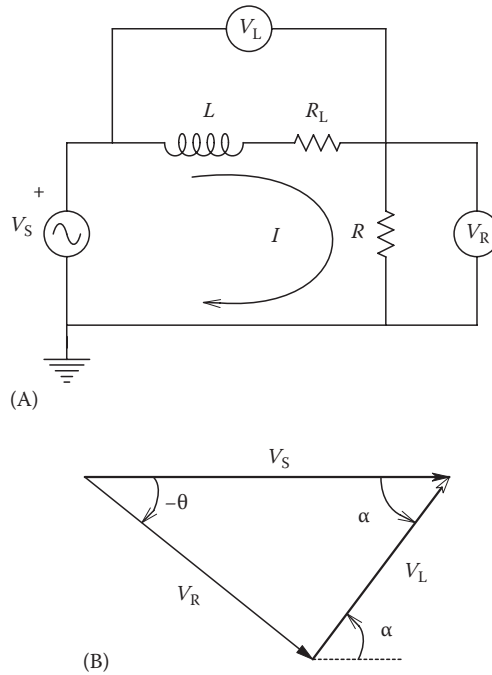


FIGURE 8.76 (A) Circuit of voltmeter method of estimating inductance at audio and powerline frequencies. (B) Vector (phasor) voltages in the circuit of A.

L can also be solved for in terms of the current angle, $-\theta$, ω , R , and R_L and also in terms of V_S , V_R , ω , R , and R_L . The derivation of these expressions is left as exercises at the end of the chapter.

8.10.3.2 Use of the Q Meter to Measure Inductance and Q

We have already seen in Section 8.10.2.1 how a Q meter can be used to measure an unknown capacitor at radio frequencies. Measurement of an unknown series inductance and its Q is quite simple. Referring to the basic Q meter schematic shown in Figure 8.69, we replace the standard inductor with a short circuit; the unknown inductor goes in the Z_S position, and Z_P is made infinite. The frequency of V_1 is set to the desired value, and V_1 is adjusted to be 1.0 V. Next, C_V is varied to obtain series resonance in the circuit. At resonance,

$$\frac{V_Q}{V_1} = \frac{V_C}{V_1} = \frac{IX_C}{IR_L} = \frac{IX_L}{IR_L} = \frac{\omega L}{R_L} = Q_L. \tag{8.189}$$

Hence, V_Q in volts equals the inductor's Q . This simple analysis assumes that the range of C_V is adequate to resonate with L , that is, $f = 1 / (2\pi\sqrt{LC_V})$. If resonance is not possible within the working range of C_V , then, depending on whether we are dealing with an inductor at the lower end of the range (90 nH) or one at the upper end of the range (130 mH),

we can add a known, standard, series inductance in series with the unknown inductor or a known, standard capacitor in parallel with C_v , respectively. The venerable Boonton Model 260A Q meter allowed measurement of inductances $>5 \mu\text{H}$ to $\pm 3\%$ and Q s to $\pm 5\%$.

It should be stressed that a Q meter measures the effective inductance and the Q is determined as the ratio of the effective inductive reactance to the effective series real part of the loop impedance. At HF and VHF, the effective parallel capacitance of the coil acts to reduce the effective Q (see Equation 5.3).

8.11 Vector Impedance Meters

A linear, two-terminal, passive, electrical component may be characterized at a given frequency by its vector impedance or admittance. Impedance, \mathbf{Z} , and admittance, \mathbf{Y} , are defined in rectangular vector form by

$$\frac{\mathbf{V}}{\mathbf{I}} = \mathbf{Z} = R + jX = \frac{1}{\mathbf{Y}} = \frac{1}{G + jB}. \quad (8.190)$$

It is understood that \mathbf{V} is the sinusoidal voltage across the two ports and \mathbf{I} is the vector current through it. Impedance can be measured at a given frequency by computing the vector ratio:

$$\mathbf{Z} = Z \angle \theta = \frac{V \angle \theta_1}{I \angle \theta_2} = Z \angle (\theta_1 - \theta_2). \quad (8.191)$$

If the magnitude of \mathbf{I} is held constant, then measurement of \mathbf{V} and the phase difference between \mathbf{V} and \mathbf{I} will characterize \mathbf{Z} at a given frequency. Figure 8.77 shows the block diagram of a basic, audio- and video-frequency VIM. The VIM is operated in the constant current magnitude mode, although some VIMs can operate in the constant voltage mode. A wideband current transformer is used to sense the current flowing through the unknown impedance to ground. The voltage output of the current transformer is conditioned and used to parametrically adjust the applied voltage to maintain constant current. The microcomputer in the VIM calculates the impedance magnitude from the magnitudes of \mathbf{V} and \mathbf{I} and also samples and presents the output of the phase meter. In the phase meter, the analog \mathbf{V} and \mathbf{I} signals are converted to TTL signals by high-speed comparators, and these signals are inputs to a NAND gate digital phase comparator, such as shown in Figure 8.55. The admittance, \mathbf{Y} , can be easily found by changing the sign of the phase angle of \mathbf{Z} and taking the reciprocal of the magnitude of \mathbf{Z} . VIMs can also give the user impedance in rectangular vector form, that is, R and X , by multiplying $|\mathbf{Z}|$ by $\cos(\theta)$ and $\sin(\theta)$, respectively. Once X is known, and the impedance is known to be inductive, the equivalent L can be found by dividing X by the known $2\pi f$. A similar means can be used to obtain a numerical value of C when the impedance is known to be from a capacitor.

HP/Agilent offers many VIMs covering a wide range of frequencies, useful for a variety of applications, including the screening of critical components in manufacturing processes. We will discuss the specifications of one representative HP VIM to give the

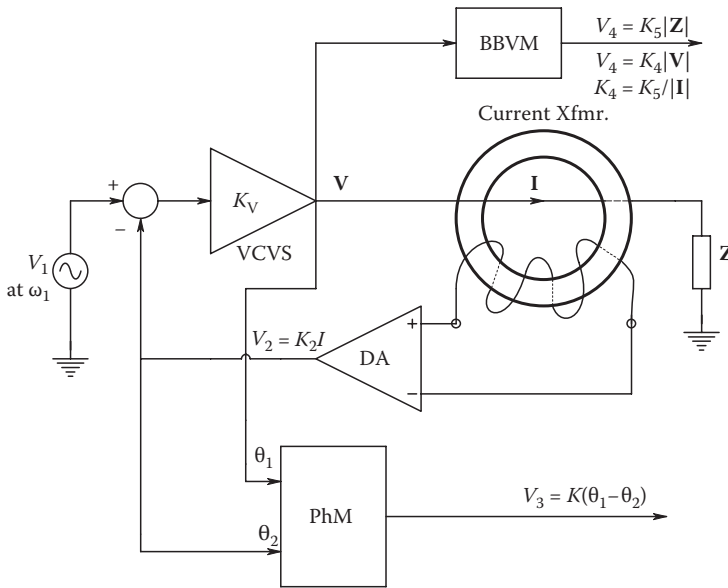


FIGURE 8.77

Diagram of a VIM. *Note:* PhM = phase meter. BBVM = broadband (RF) voltmeter (e.g., a thermocouple type). V_1 = oscillator. VCVS = voltage-controlled voltage source (i.e., a voltage amplifier).

reader a feeling for the state of the art in this type of instrument. The HP/Agilent Model 4192A LF Impedance Analyzer (VIM) operates from 5 Hz to 13 MHz. This instrument can measure 11 impedance-related parameters over its frequency range: $|Z|$, $|Y|$, θ , R , X , G , B , L , C , Q , and D . Outputs are read on two, 4, and one-half digit numerical displays and can be accessed through the HP-IB (IEEE-488) bus. The range of Z , X , and R is from 0.1 m Ω to 1.3 M Ω , and the range of Y , G , and B is from 1 nS to 13 S. Frequency accuracy is ± 50 ppm, and frequency synthesizer steps are 0.001 Hz from 5 Hz to 10 kHz, 0.01 Hz from 10 to 100 kHz, 0.1 Hz from 100 kHz to 1 MHz, and 1 Hz from 1 to 13 MHz. Series or parallel equivalent circuit models may be user selected for impedance and admittance measurements. The complete range of inductance that can be measured is from 10 μ H (at high frequencies) to 1 kH (at low frequencies) with a basic accuracy of $\pm 0.27\%$. Capacitance can be measured from 0.1 fF to 199 mF, with a basic accuracy of $\pm 0.15\%$. $D = 1/Q$ can be measured from 10^{-4} to 20 with an accuracy of 10^{-3} for C measurements and 0.003 for L measurements. Maximum % accuracy in measuring the 11 impedance parameters is a function of frequency and the value of the parameter being measured. The best accuracy generally occurs between 100 Hz and 1 MHz, and it varies inversely with parameter magnitude (e.g., best phase accuracy is obtained when measuring low values of Y , G , or B and when measuring low values of $|Z|$, R , or X ; likewise, best accuracy in measuring Z , R , or X occurs for low ohm values of these parameters).

The HP/Agilent 4191A RF impedance analyzer is a VIM intended to work in the 1–1000 MHz range. In addition to measuring the 11 impedance parameters mentioned earlier, this meter also measures the reflection coefficient, $\Gamma(\omega)$, of a standard transmission line with characteristic impedance, Z_o , terminated with an unknown impedance, Z_L . $\Gamma(\omega)$ is defined as

$$\Gamma(\omega) = \frac{V_r}{V_i} = \frac{Z_L - Z_o}{Z_L + Z_o} = \frac{VSWR - 1}{VSWR + 1} = \Gamma \angle \theta. \tag{8.192}$$

The HP 4191A VIM uses two microwave directional couplers to isolate the input wave, V_i , and the reflected wave, V_r . Hence, $\Gamma(\omega)$ can be calculated from the vector quotient, and knowing Z_o , we can calculate Z_L . In addition to presenting Γ in polar form, the HP 4191A VIM also can give Γ_x and Γ_y over a range of 10^{-4} to 1 with a resolution of 10^{-4} . Typical accuracy in measuring $\Gamma(\omega)$ is $\pm 0.2\%$ from 1 to 100 MHz, rising to about $\pm 0.5\%$ at 1 GHz.

In summary, VIMs are seen to be versatile, modern, microprocessor-based instruments, capable of measuring a wide variety of impedance-related parameters, including inductance, capacitance, Q , and D over a wide range of frequencies and values. Modern VIMs have IEEE 488 bus I/O, which allows them to be directly controlled by and report their measurements to a computer system. They also are relatively expensive.

8.12 Chapter Summary

Chapter 8 has dealt with descriptions of the common means of measuring electrical parameters including DC voltage, DC, electric field strength, AC voltage, AC, phase, frequency, resistance, capacitance, inductance, vector impedance, and vector admittance. The most accurate means of measurement were seen to be null methods, in which precise parametric standards are used for comparison. (Null methods were treated in detail in Chapters 4 and 5.)

The tremendous range of component values used in electrical engineering systems design generally means that no one instrument type, design, or frequency is suitable to make accurate measurements over the entire practical range of a parameter's values. Because of the presence of parasitic capacitance in inductors, and parasitic inductance in capacitors, HF measurement of inductance and capacitance was seen to be more realistically treated as complex impedance or admittance measurements.

Finally, we described the organization of modern, microprocessor-controlled, VIMs that are ubiquitous instruments for the characterization of two-terminal device impedances over wide ranges of frequencies.

Problems

- 8.1 An $\eta = 20,000 \Omega/V$ DC voltmeter is designed to have V_{FS} full-scale volts. It is used to measure the DC voltage between two nodes in a DC circuit. Looking into the nodes, we see a Thevenin equivalent circuit with V_{OC} and R_{TH} as shown. Because the voltmeter draws current, the DC voltage read by the meter, V_m , will be less than V_{OC} by an amount ΔV . That is, $V_m = V_{OC} - \Delta V$. See Figure P8.1.
- Give an expression for the total input resistance of the voltmeter.
 - Derive an expression for the fractional error, $\Delta V/V_{OC}$, in the reading of V_{OC} in terms of V_{FS} , η , and R_{TH} .

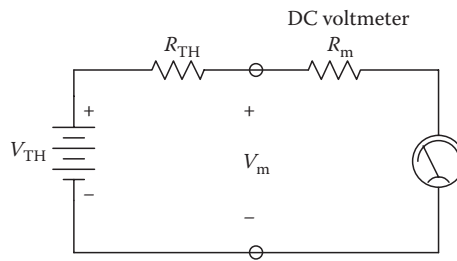


FIGURE P8.1

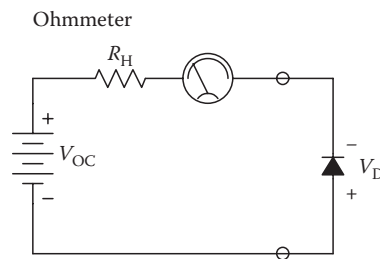


FIGURE P8.2

- 8.2 A series ohmmeter having $R_H = 10^4 \Omega$ and $V_{OC} = 9.0 \text{ V}$ is connected to a silicon diode as shown in Figure P8.2. The diode has $v_T = 0.025 \text{ V}$ and $I_{RS} = 10 \mu\text{A}$. What resistance does the ohmmeter read?
- 8.3 Two series ohmmeters are connected as shown in Figure P8.3 to a $1 \text{ k}\Omega$ resistor. What resistance does each ohmmeter read? Note that $S = I_x/I_{FS} = I_x/(V_{OC}/R_H) = R_H/(R_H + R_x)$.
- 8.4 A charge amplifier, shown in Figure P8.4, is used to measure the charge on a helicopter that has just landed. The helicopter's body capacitance on the ground is $C_H = 1 \text{ nF}$. The SS DC voltage at the OAs' output, neglecting V_{OS} and I_B , is -4.0 V .
- Find the charge in coulombs on helicopter, Q_H . Is Q_H positive or negative?
 - Using superposition, find the error voltage at the OA's output after 10 s due to $I_B = 50 \text{ fA}$, $V_{OS} = 1.0 \text{ mV}$.

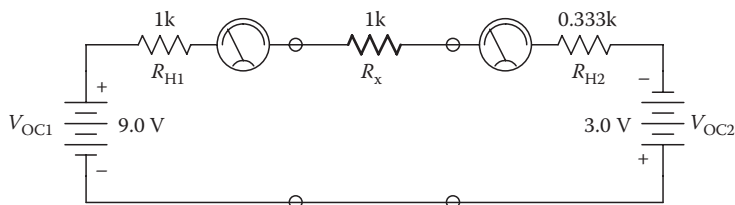


FIGURE P8.3

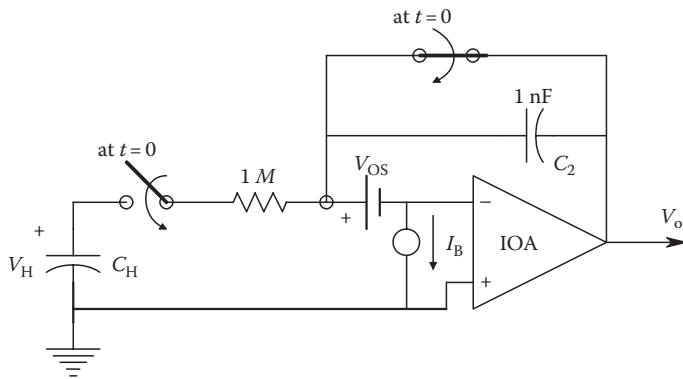


FIGURE P8.4

- 8.5 A Faraday magneto-optical system is shown in Figure P8.5. A conductor carrying I amperes is wound as an N -turn solenoid around a polarization-preserving fiber-optic cable (PPFOC). Current in this solenoid induces a $+\psi$ rotation of LPL in the PPFOC segment AB. Now N' turns of the PPFOC are wound around the current-carrying conductor. N' is chosen so that a net rotation of the LPL of $-\alpha$ degrees occurs.
- Find N'/N required so that $\alpha = \psi$ and the net polarization angle rotation in segment CD of the PPFOC is zero. Assume the Verdet constant, V , of the solenoidal rotation is $3800^\circ/(\text{T m})$; V' of the FO system is $2.68 \times 10^{-4}^\circ/(\text{A turn})$. The average axial B in the solenoid is $(3/4)\mu_0 I/L_c$ T.
 - Find ψ for a current of 100 A. Let $N = 100$ turns.

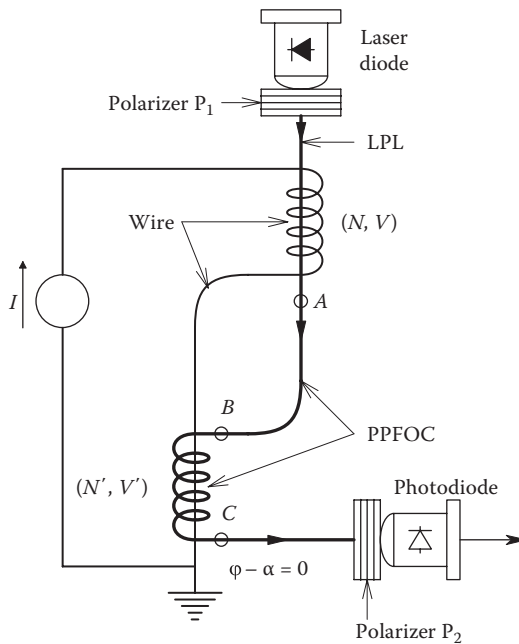


FIGURE P8.5

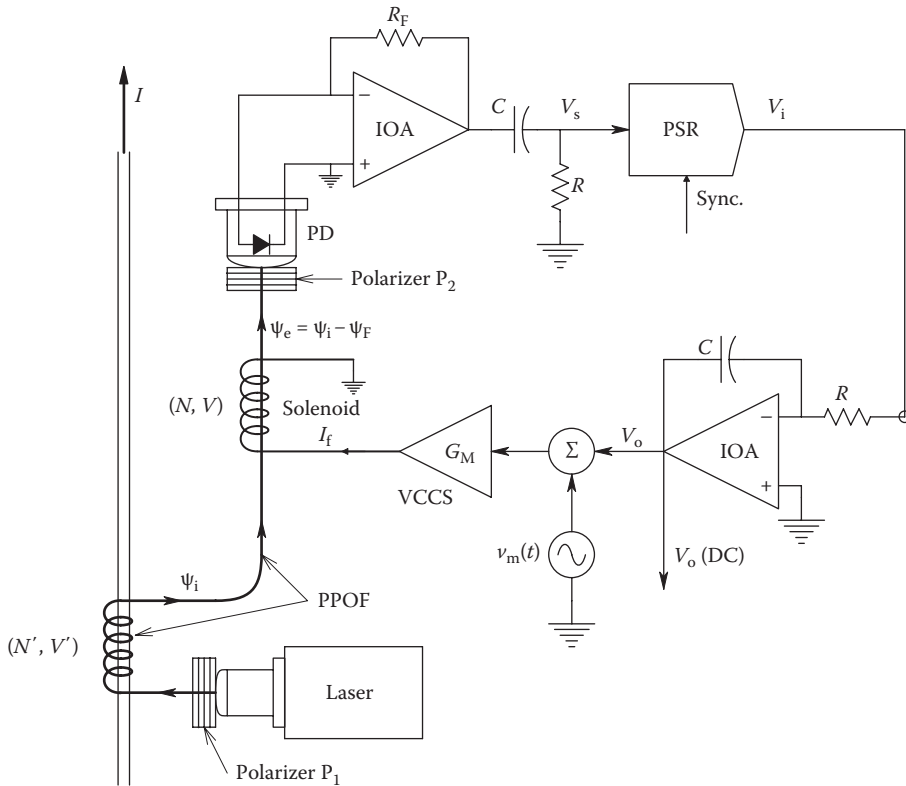


FIGURE P8.6

8.6 The electro-optical feedback system of Figure P8.6 uses a *Gilham polarimeter* to measure the large direct current, I , in the insulated conductor using the *Faraday magneto-optical effect*. The induced polarization rotation in the fiber-optic cable is $\psi_i = V'NI$ degrees. The feedback rotation is $\psi_F = VB I_F$. At null, $\psi_i = \psi_F$. V and V' are Verdet constants.

- A. Derive an expression for V_o in terms of system parameters and I in the cable, assuming an SS null.
- B. Make a block diagram for the system. Assume $V_i = -\psi_e I_4 K_s K_p \theta_m$ V, where I_4 = the laser beam intensity in watts, K_s = detector gain in V/W, K_d = the PSR gain in V/V, and θ_m = peak polarization angle modulation by the current, $G_M v_m(t)$.
- C. Give the transfer function for the system, $\frac{V_o}{I}(s)$. Give the system's corner frequency in r/s.

8.7 A Hall sensor is used to measure a large DC on the order of 100s of amperes. The DC through the conductor acts like a 1-turn coil to induce a DC B_y in the ferrite C core. See Figure P8.7.

The measurement system uses DC feedback to generate a B_f that nulls the net B in the core. That is, $B_e = B_y - B_f$. In order to gain sensitivity, the Hall output voltage is modulated at 1 kHz by passing an AC I_x into the Hall device. That is, $i_x(t) = I_{x0} \sin(2\pi 10^3 t)$. Thus, $v_H(t) = K B_e \sin(2\pi 10^3 t)$. B_e in turn is proportional to $(I_{DC} - NI_F)$. N is the number of turns of the feedback coil on the core, and I_F is the DC feedback current. At null, $v_H(t)$ and the DC $V_2 \rightarrow 0$, and $I_{DC} = NI_F$. $I_F = G_M V_o$. Derive an expression for I_{DC} in terms of the integrator DC output voltage, V_o .

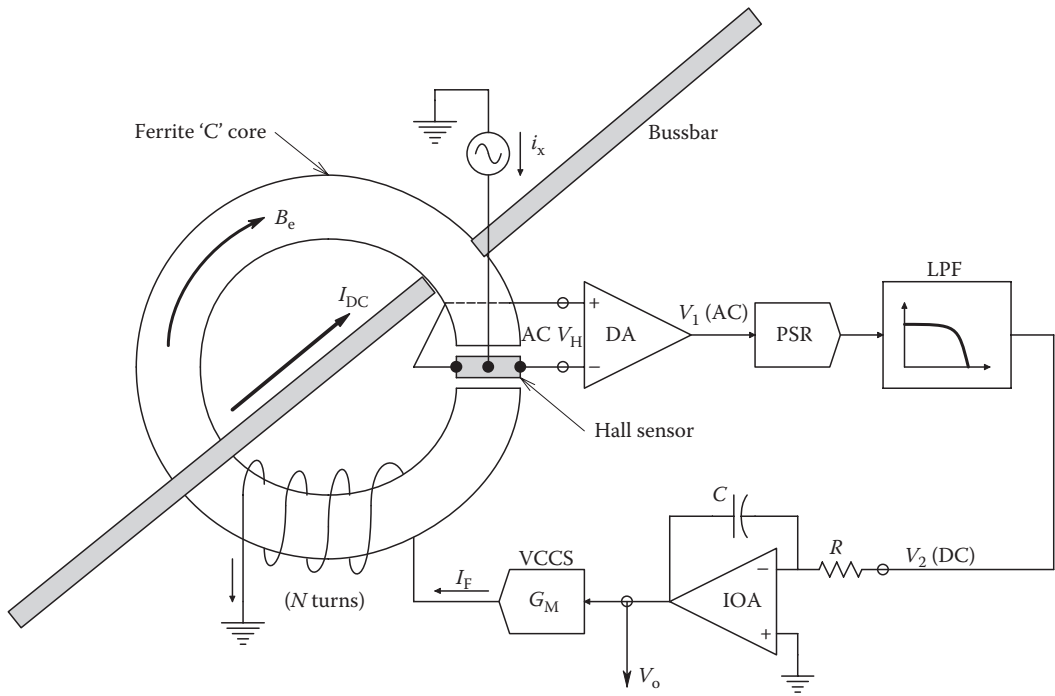


FIGURE P8.7

8.8 A Hall sensor is used to make an AC milliammeter. A solenoid coil makes an axial flux density of

$$B_y = \mu_0 N i_s(t) / (2b) \text{ w/m}^2$$

where

$N = 100$ is the number of turns

$b = 0.01$ m is the length of the coil

$\mu_0 = 4\pi \times 10^{-7} \text{ N} \cdot \text{s}^2 / \text{Cb}^2$. The Hall sensor is connected as shown in Figure P8.8.

$I_x = 0.01$ A (DC), $R_H = 62.5$ (Hall constant), and $h = 0.0001$ m (sensor thickness).

The unknown current is sinusoidal: $i_s(t) = I_s \sin(\omega t)$

- A. Find an expression for the Hall voltage, $v_H(t)$, in terms of system parameters and I_s .
 - B. $v_H(t)$ is amplified by a DA with gain $K_V = 100$. The AC voltage, $v_1(t)$, is full-wave rectified and low-pass filtered to form a DC output, V_o , proportional to I_s . Both the LPF and the rectifier have unity gain. Find I_s to make $V_o = 1.0$ V.
- 8.9 A Hall sensor is used to make a DC wattmeter. In order to gain sensitivity, I_x is made an AC proportional to V_L by using an analog multiplier. $v_s(t) = V_S \sin(\omega t)$, $i_x(t) = (G_M V_L V_S / 10) \sin(\omega t)$. Thus, the Hall voltage, V_H , is an AC signal whose peak value is proportional to $P_{DC} = V_L I_L$. The magnetic flux density is given by $B_y = I_L \mu_0 N / 2b$ T. See Figure P8.9.
- A. Find an expression for the Hall voltage, $v_H(t)$.
 - B. Plot and dimension the output of the full-wave rectifier.

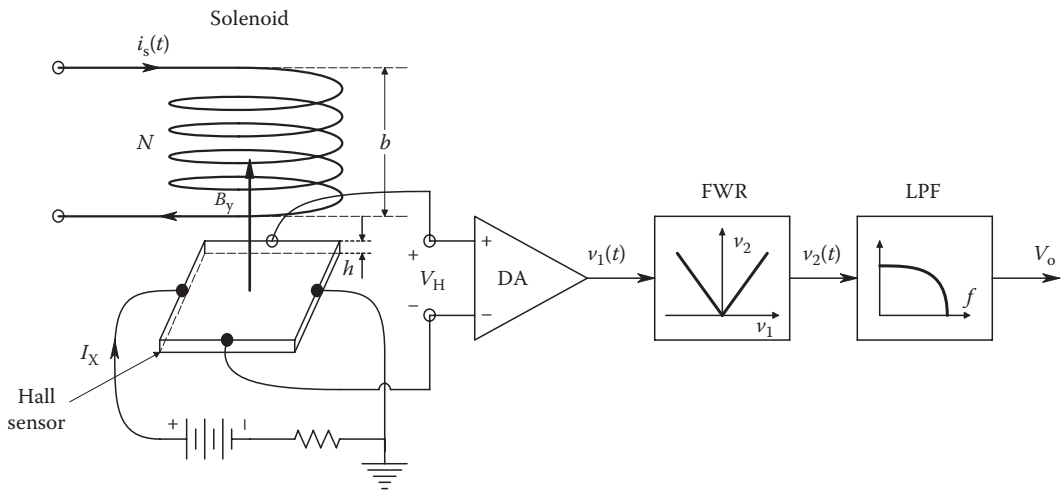


FIGURE P8.8

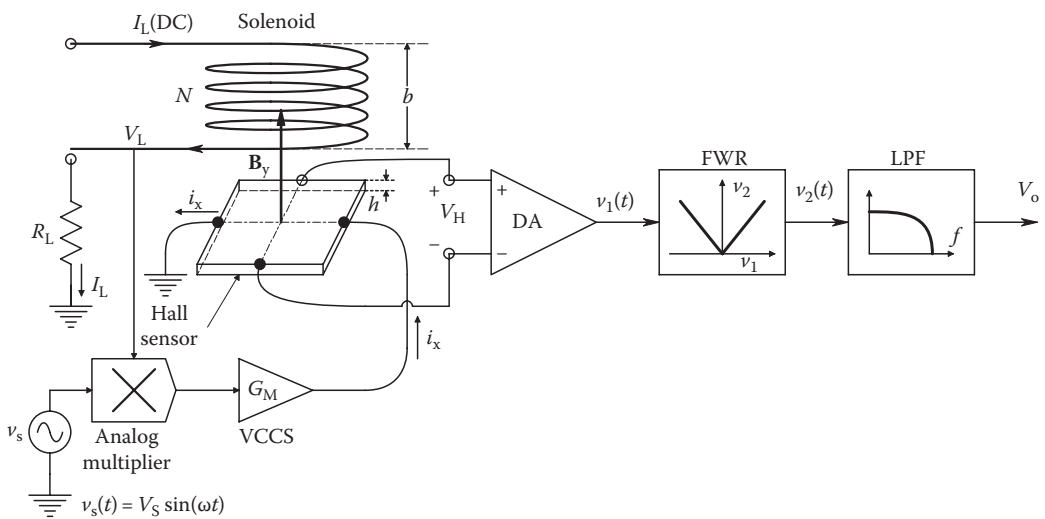


FIGURE P8.9

8.10 Figure P8.10 illustrates an AC wattmeter that measures the real, average power dissipated in the load, Z_L . Note that $v_L(t) = V_L \sin(\omega t)$ and $i_L(t) = I_L \sin(\omega t + \phi)$. Refer to Problem 8.9 for relations governing the Hall sensor's operation.

- A. Find an expression for $v_H(t)$.
- B. Find an expression for V_o .

8.11 Figure P8.11 shows the circuit for a true rms voltmeter based on two, matched, VTCs. It uses a type 1 feedback loop in which \bar{V}_2 in the SS = 0. The DC integrator output causes a DC, I_{T_r} to flow in the reference thermocouple heater such that its temperature

equals the temperature in the input thermocouple heater. Thus, for $V_2 = 0$, the TC voltages, $V_M = V_R$.

- A. Derive an expression for the DC output, V_{M_r} , of the input vacuum TC.
- B. Derive an expression for the rms v_{σ} , assuming SS conditions.

8.12 See Figure 8.12A and B. A manufacturing process for sheet Teflon™ winds the finished product at high speed, under tension, on a roll. Although steps are taken to control surface static charges, some charge may accumulate on the roll, which is a dielectric. The manufacturer wishes to sense the magnitude and sign of the surface charge density in order to adjust processing steps to null it. Assume the sheet winds on the roll at a constant linear velocity, $v = 10$ m/s. The top surface is assumed to have an average charge density of ρ Cb/m². A thin, soft wire brush of width $W = 0.5$ m makes contact with the top of the moving plastic sheet and sweeps the surface it touches clear of loose charges. Two circuits are proposed to measure ρ : In A, the switches periodically switch to the measurement (M) position for T_m seconds, then to the G position for T_g seconds in order to zero V_o and the charge on C_F . V_o is read at the end of each T_m second interval by a sample-and-hold amplifier (not shown). In circuit B, the current from the brush is converted continuously to output voltage V'_o . V'_o is low-pass filtered by a real pole filter with a 1 s time constant. The OA's $e_{na} = 10$ nV/ $\sqrt{\text{Hz}}$, $i_{na} = 1$ fA/ $\sqrt{\text{Hz}}$, $R_F = 10^6 \Omega$ at 300 K. All noises are white.

- A. For the A circuit, derive an expression for ρ from V_o at $t = T_m$.
- B. For the B circuit, derive an expression for ρ in terms of V'_o and system parameters.
- C. Find a numerical value for ρ_{\min} that will give a mean-squared output SNR at V'_o of 10.

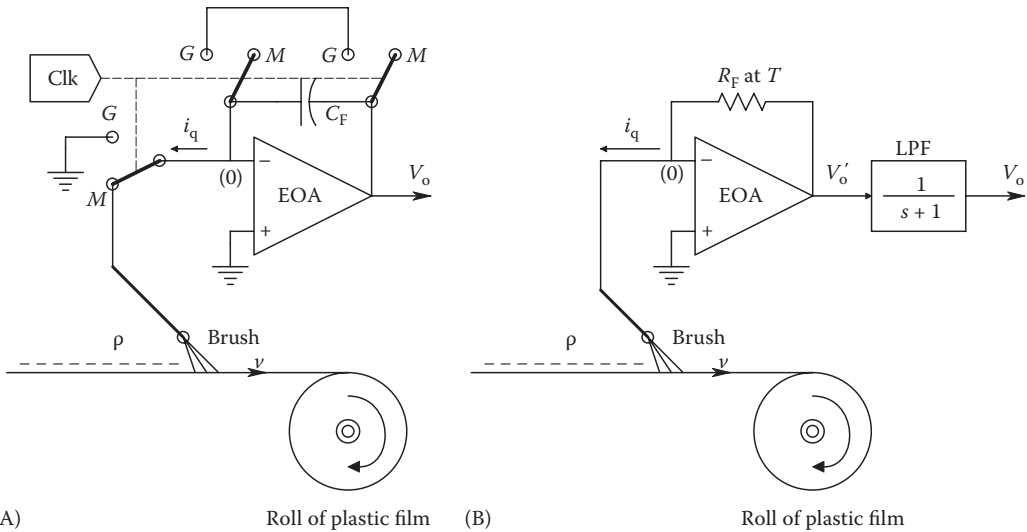


FIGURE P8.12

9

Digital Interfaces in Measurement Systems

9.1 Introduction

As we have seen in Figure 1.1, a modern instrumentation system generally includes a computer that is used to supervise, coordinate, and control the measurements and that is also used to store data (data logging), condition data, and to display it in a meaningful, summary form on a monitor. In this chapter, we will describe the hardware associated with the conversion of analog information to digital formats and noise and resolution problems associated with the ADC process.

We also examine various means of digital communication with sensors and instruments, including the IEEE-488.2 bus, other parallel bus architectures used with PCs, and various serial data transfer protocols (e.g., USB). Commercially available interface cards are described, and problems associated with sending both analog and digital data long distances at high data rates on cables are analyzed.

In describing data conversion interfaces, it is expedient to first consider DACs because these systems are used in the designs of several designs for ADCs. Data conversion from A/D form or digital to analog form is often done continuously and periodically. Continuous, periodic data conversion has a great effect on the information content of the converted data, as we will show in the next section. A noise-free analog signal sample theoretically has infinite resolution. Once the analog signal sample has been converted to digital form, it is represented by a digital (binary) number of a finite number of bits (e.g., 12), which limits the resolution of the sample (one part in 4096 for 12 bits). This obligatory rounding-off of the digital sample is called *quantization*, and the resulting errors can be thought of as being caused by a *quantization noise*. It is shown in Section 9.4 that adding broadband noise (*dithering*) to the analog signal being digitized mitigates the effect of quantization noise on signal resolution.

IC ADCs and DACs are available in a wide spectrum of specifications. They are available with quantization levels ranging from 6 to over 20 bits and with conversion rates ranging from low audio frequencies to less than 1 ns per sample. Most ADCs and DACs have parallel digital outputs or inputs, respectively, although a few are designed to use serial data I/O protocols.

Before discussing data conversion hardware, we will first discuss the linear dynamics of data conversion, quantization noise, and dithering.

9.2 Sampling Theorem

The sampling theorem is important because it establishes a criterion for the minimum sampling rate that must be used to digitize a signal with a given, low-pass, power density spectrum. The relation between the highest significant frequency component recoverable in the signal's power density spectrum and the sampling frequency is called the *Nyquist criterion*; the implications of the Nyquist criterion and *aliasing* are described in the succeeding text.

Because ADC is generally a periodic process, we will first analyze what happens when an analog signal, $x(t)$, is periodically and ideally sampled. The ideal sampling process generates a data sequence from $x(t)$ only and exactly at the sampling instants, when $t = nT_s$, where n is an integer ranging from $-\infty$ to $+\infty$ and T_s is the *sampling period*. It is easy to show that an ideal sampling process is mathematically equivalent to *impulse modulation*, as shown in Figure 9.1. In Figure 9.1B, a continuous analog signal $x(t)$ is multiplied by an infinite train of unit impulses or delta functions that occur only at the sampling instants. This multiplication process produces a periodic number sequence, $x^*(t)$, having infinite resolution, at the sampler output. In the frequency domain, $\mathbf{X}^*(j\omega)$ is given by the complex convolution of $\mathbf{X}(j\omega)$ with the FT of the pulse train, $P_T(j\omega)$. In the time domain, the pulse train can be written as

$$P_T(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s). \tag{9.1}$$

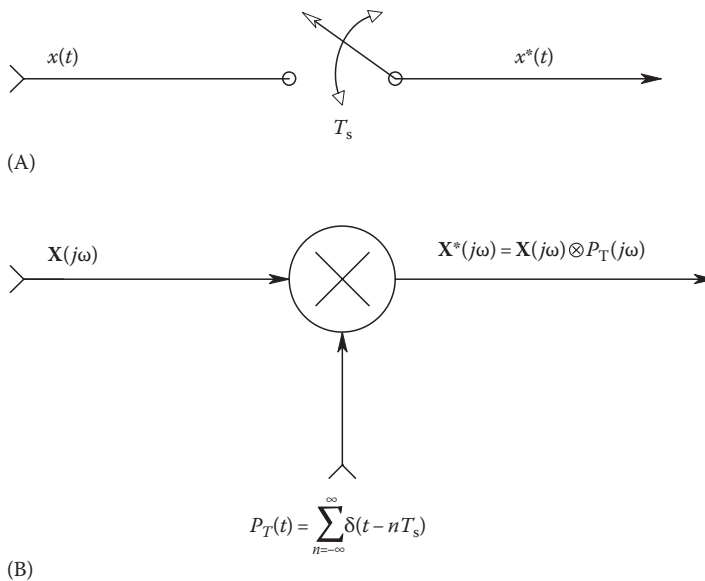


FIGURE 9.1

(A) Symbol for an ideal sampling process (ADC). T_s is the interval between successive samples. $x^*(t)$ exists only at sampling instants where $x^*(t) = x(t = nT_s)$. (B) The impulse modulation process is equivalent to ideal sampling. $\mathbf{X}^*(j\omega)$ = FT of the sampled signal. $\mathbf{X}^*(j\omega)$ can be expressed as the complex convolution of the FTs, $\mathbf{X}(j\omega)$ and $P_T(j\omega)$. (From Northrop, R.B., *Analog Electronic Circuits: Analysis and Applications*, Addison-Wesley, Reading, MA, 1990. With permission.)

Here, $\delta(t - nT_s)$ is a unit impulse (delta function) that is defined only at the point, $t = nT_s$. That is, it has infinite height, zero width, and unity area at $t = nT_s$ and is zero for nonzero argument. The periodic function, P_T , can also be represented in the time domain by a *Fourier series* in complex form:

$$P_T(t) = \sum_{n=-\infty}^{\infty} C_n \exp(-jn\omega_s t), \tag{9.2}$$

where

$$\omega_s \equiv \frac{2\pi}{T_s} \text{ r/s.} \tag{9.3}$$

The complex-form Fourier series coefficients are given by

$$C_n = \frac{1}{T_s} \int_{-T_s/2}^{T_s/2} P_T(t) \exp(+jn\omega_s t) dt = \frac{1}{T_s}. \tag{9.4}$$

Thus, the complex Fourier series for the pulse train is found to be

$$P_T(t) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} \exp(-jn\omega_s t). \tag{9.5}$$

The sampler output is the time-domain product of Equation 9.5 and $x(t)$:

$$\begin{aligned} x^*(t) &= x(t) \frac{1}{T_s} \sum_{n=-\infty}^{\infty} \exp(-jn\omega_s t) \\ &= \frac{1}{T_s} \sum_{n=-\infty}^{\infty} x(t) \exp(-jn\omega_s t). \end{aligned} \tag{9.6}$$

The Fourier theorem for complex exponentiation is

$$F\{y(t) \exp(-jat)\} \equiv Y(j\omega - ja). \tag{9.7}$$

Using this theorem, we can write the FT for the sampler output

$$X^*(j\omega) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} X(j\omega - jn\omega_s). \tag{9.8}$$

Equation 9.8 for $X^*(j\omega)$ is the result of the *complex convolution* of $X(j\omega)$ and $P_T(j\omega)$; it is in the *Poisson sum form*, which helps us to visualize the effects of (ideal) sampling in the frequency domain and to understand the phenomenon of aliasing. In Figure 9.2A, we plot

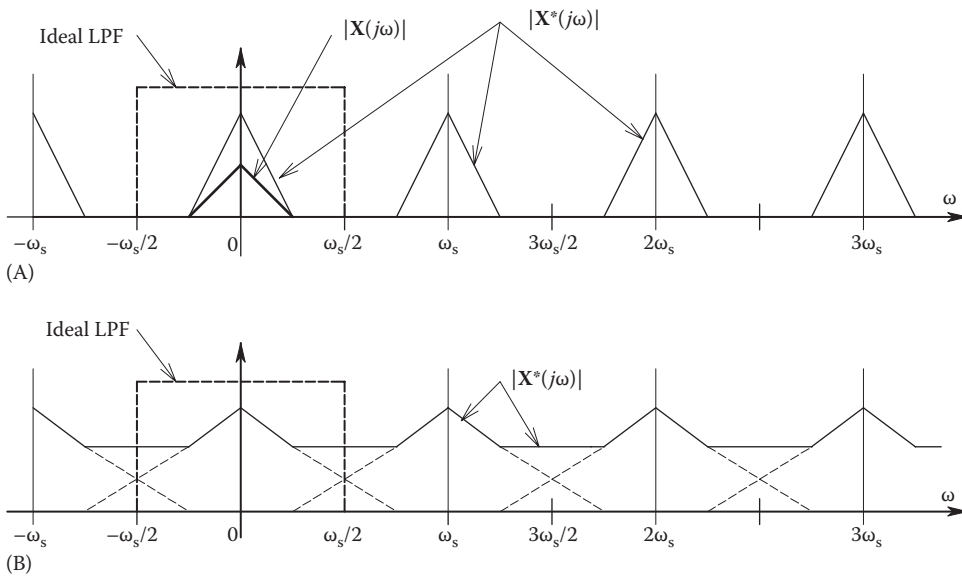


FIGURE 9.2

(A) Magnitude of the FTs of the sampler I/O. Note that the output spectrum, $X^*(j\omega)$, is periodic at the sampling frequency, $\omega_s = 2\pi/T_s$ r/s. The input spectrum, $X(j\omega)$, contains no spectral power beyond the Nyquist frequency, $\omega_N \equiv \pm\omega_s/2$. (B) Magnitude of the FT of the sampler output when the input signal's spectrum extends beyond ω_N . This condition produces *aliasing*, in which there is overlap of the sampled signal's spectra. See text for analysis.

the magnitude of a typical $X(j\omega)$; note that $X(j\omega)$ is assumed to have negligible power above the Nyquist frequency, $\omega_s/2$ r/s. When $x(t)$ is sampled, we see that $X^*(j\omega)$ has an infinite, *periodic spectrum*.

Note that $x(t)$ can be recovered from the sampler output by passing $x^*(t)$ through an ideal LPF, as shown in Figure 9.2A. In Figure 9.2B, we have assumed that the baseband spectrum of $x(t)$ extends beyond $|\omega_s/2|$ r/s. When such an $x(t)$ is sampled, the resultant $X^*(j\omega)$ is also periodic, but the high-frequency corners of the component spectra overlap. An ideal LPF thus cannot uniquely recover the baseband spectrum and thus $x(t)$. This condition of overlapping spectral components in $X^*(j\omega)$ is called *aliasing*, and it can lead to serious errors in digital signal processing.

Because of the problem of aliasing, all properly designed ADC systems used in FFT spectrum analyzers and related equipment must operate on input signals that obey the Nyquist criterion, that is, the power density spectrum of $x(t)$, $S_{xx}(f)$, must have no significant power at frequencies above one-half the sampling frequency. One way to ensure that this criterion is met is to use properly designed, analog low-pass, AAFs immediately preceding the sampler. AAFs are generally high-order, linear phase, LPFs that attenuate the input signal by at least 40 dB at the Nyquist frequency.

Many designs are possible for high-order anti-aliasing filters. For example, Chebyshev filters maximize the attenuation cutoff rate at the cost of some passband ripple. Chebyshev filters can achieve a given attenuation cutoff slope with a lower order (fewer poles) than other filter designs. It should be noted that in the limit as passband ripple approaches zero, the Chebyshev design approaches the Butterworth form of the same order, which has no ripple in the passband. Chebyshev filters are designed in terms of their order, n , their cutoff frequency and the maximum allowable peak-to-peak ripple (in decibels) of their passband.

If ripples in the frequency response stopband of an anti-aliasing filter are permissible, then elliptical or Cauer filter designs may be considered. With stop band ripple allowed, even sharper attenuation in the transition band than obtainable with Chebyshev filters of a given order can be obtained. Elliptic LPFs are specified in terms of their order, their cutoff frequency, the maximum peak-to-peak passband ripple, and their minimum stopband attenuation.

Bessel or Thomson filters are designed to have linear phase in the passband. They generally have a “cleaner” transient response, that is, less ringing and overshoot at their outputs, given transient inputs.

As an example of anti-aliasing filter design, Franco (1988) shows a sixth-order Chebyshev antialiasing filter made from three, concatenated, Sallen and Key (quadratic) low-pass modules. This filter was designed to have an attenuation of 40 dB at the system’s 20 kHz Nyquist frequency, a corner frequency of 12.8 kHz, a -3 dB frequency of 13.2 kHz, and a ± 1 dB passband ripple.

Quoted from Northrop (1990), with permission

One problem in the design of analog AAFs to be used with instruments having several different sampling rates is adjusting the filters to several different Nyquist frequencies. One way of handling this problem is to have a fixed filter for each separate sampling frequency. Another way is to make the filters easily tunable, either by digital or analog voltage means. Some approaches to the tunable filter problem are described in Section 7.4 in Northrop (2012).

Antialiasing LPFs are generally not used at the inputs of digital oscilloscopes, because one can see directly if the sampled waveform is sampled too slowly for resolution. An optional analog LPF is often used to cut high-frequency interference on digital oscilloscope inputs; this, however, is not an AAF.

9.3 Quantization Noise

When a band-limited analog signal is periodically sampled and then converted by an N -bit ADC to digital form, an uncertainty in the digitized signal level exists that can be considered to be equivalent to a broadband, quantization noise added to the analog signal, which is digitized by an infinite-bit ADC. To illustrate the properties of quantization noise, refer to the quantization error generating model of Figure 9.3. A Nyquist-limited analog signal, $x(t)$, is sampled and digitized by an N -bit ADC. The ADC’s digital output is the input to an N -bit DAC. The quantization error is defined at sampling instants as the difference between DAC output and the analog input signal.

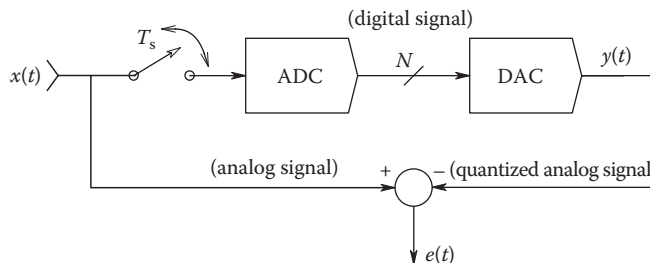


FIGURE 9.3

Block diagram of the system used to observe quantization error in the A/D process.

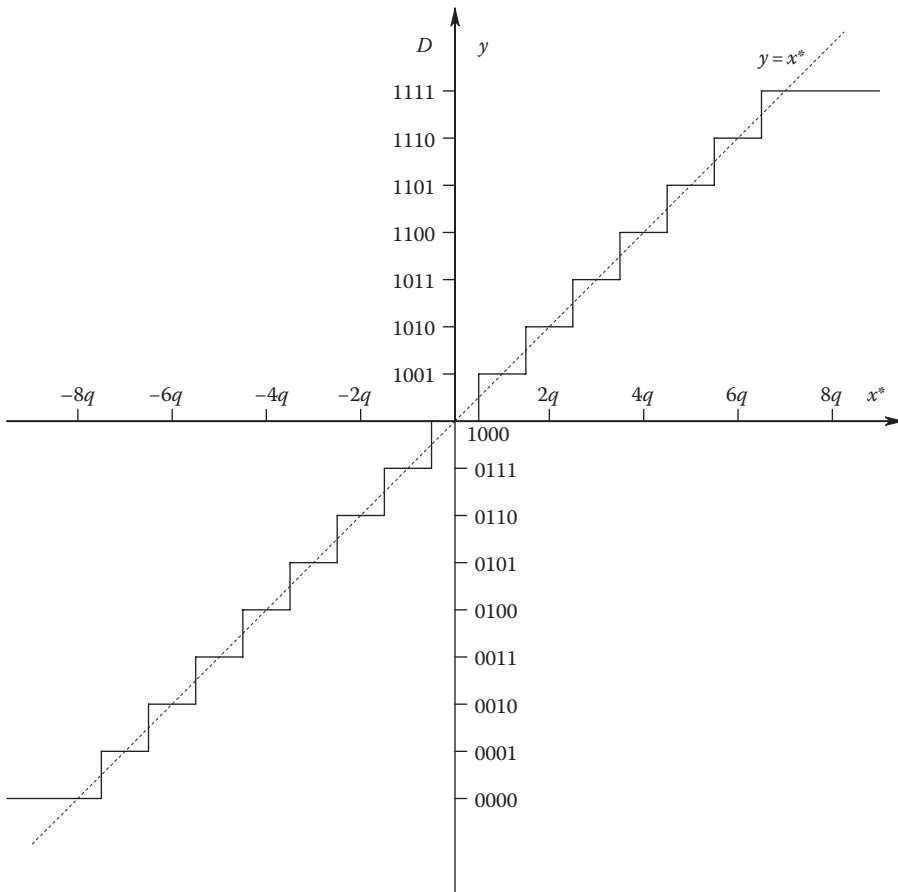


FIGURE 9.4

I/O diagram of a 4-bit quantizer (the ADC/DAC in Figure 9.3). Note that there are $2^4 = 16$ levels and $(2^4 - 1) = 15$ steps.

Figure 9.4 illustrates the nonlinear function relating sampler input, $x^*(t)$, to the DAC output, $y(t)$. In this example, $N = 4$. The quantizer has (2^N) levels and $(2^N - 1)$ steps. Compared to the direct path, the error $e(nT_s)$ can range over $\pm q/2$, where q is the voltage step size of the ADC. It is easy to see that

$$q = \frac{V_{MAX}}{2^N - 1} V, \tag{9.9}$$

where V_{MAX} is the maximum (peak-to-peak) range of the ADC/DAC system. For example, if a 10-bit ADC is used to convert a signal ranging from -5 to $+5$ V, then by Equation 9.9, $q = 9.775$ mV. If the $x(t)$ has zero mean, and its PDF has a standard deviation, σ_x , which is large compared to q , then it can be shown that the PDF of $e(nT_s)$ is well modeled by a uniform (rectangular) density function over $|e| \leq q/2$. This rectangular PDF is shown in Figure 9.5. The mean-squared error voltage is found from

$$E\{e^2\} = \overline{e^2} = \int_{-q/2}^{q/2} e^2 \left(\frac{1}{q}\right) de = \frac{(1/q)e^3}{3} \Big|_{-q/2}^{q/2} = \frac{q^2}{12} \text{ MS V.} \tag{9.10}$$

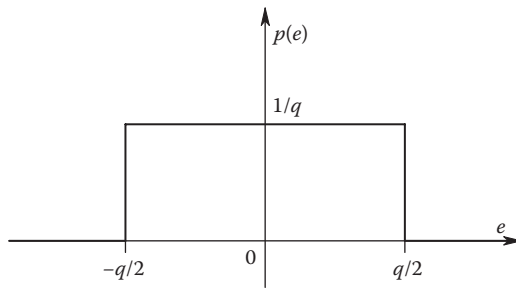


FIGURE 9.5 PDF of the quantization error of the system of Figure 9.3. q is the voltage step size of the ADC.

Thus, it is possible to treat quantization error noise as a zero-mean, broadband noise with standard deviation of $q/\sqrt{12}$ volts, added to $x^*(t)$.

In order to minimize the effects of quantization noise for an N -bit ADC, it is important that the analog input signal, $x(t)$, uses nearly the full dynamic range of the ADC. In the case of a zero-mean, time-varying signal that is Nyquist band limited, gains and sensitivities should be chosen so that the peak expected $x(t)$ does not exceed the maximum voltage limits of the ADC. Also, if $x(t)$ has a Gaussian PDF, the dynamic range of the ADC should be about ± 3 standard deviations of the signal. Under this condition, it is possible to derive an expression for the mean-squared SNR of the ADC and its quantization noise. Let the signal have an rms value of σ_x volts. From Equation 9.9, we see that the quantization step size can be written as

$$q \approx \frac{6\sigma_x}{(2^N - 1)} \text{ V.} \tag{9.11}$$

This relation for q can be substituted into Equation 9.10 for the variance of the quantization noise. Thus, the output noise is

$$N_o = \frac{q^2}{12} = \frac{36\sigma_x^2}{12(2^N - 1)} \text{ MS V.} \tag{9.12}$$

Thus, the mean-squared SNR of the quantizer is

$$\text{SNR}_q = \frac{(2^N - 1)}{3} \text{ MS V/MS V.} \tag{9.13}$$

Note that the quantizer SNR is independent of σ_x as long as σ_x is held constant under the dynamic range constraint described earlier. Table 9.1 summarizes the SNR_q of the quantizer for different bit values.

Sixteen- to 20-bit ADCs are routinely used in modern digital audio systems because of their low quantization noise.

TABLE 9.1
SNR Values for an N -Bit ADC
Treated as a Quantizer (total input range is assumed to be $6\sigma_x V$)

NdB	SNRq
6	31.2
8	43.4
10	55.4
12	67.5
14	79.5
16	91.6

9.4 Dithering

Dithering is a method for statistically reducing the quantization errors and harmonic distortion inherent in the ADC process. In dithering, a signal uncorrelated with the input analog signal, $x(t)$, is added to it, and the sum is sampled and converted. Figure 9.6 illustrates an externally dithered ADC. The dither signal can be a narrowband (almost coherent) waveform with a power spectrum centered just below the Nyquist frequency, $f_s/2$, or a broadband noise whose spectrum covers the signal spectrum, as shown in Figure 9.7. Note that dither noise is subtracted digitally from the ADC's output after conversion.

Adding dither to a converter's analog input can provide significant improvements to the ability to extract signals below the resolution of the converter ($< \pm 1/2$ LSB) and in linearizing an ADC's performance. The SNR improvements depend on the signal characteristics and the amount and type of dither. Combined with the time averaging property of subsequent digital low-pass filtering, dithering effectively removes or smoothes the quantization noise

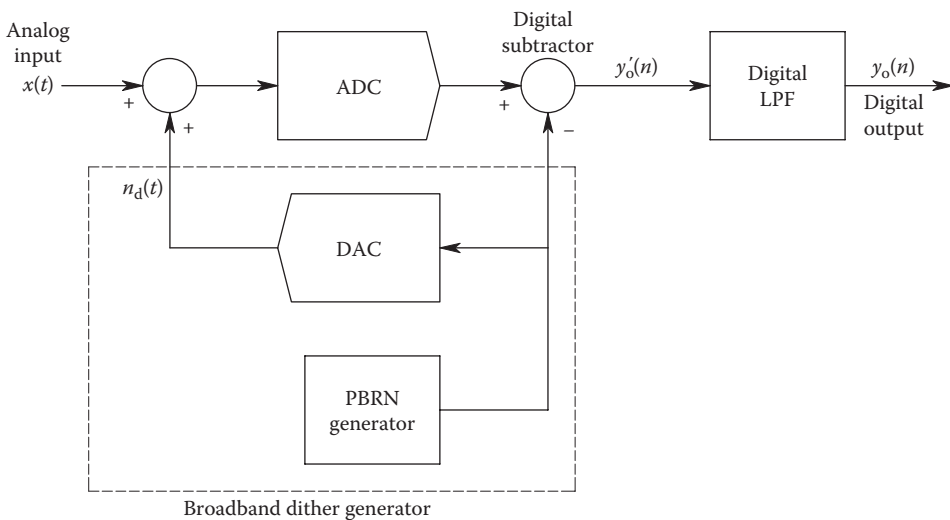


FIGURE 9.6
An externally dithered ADC.

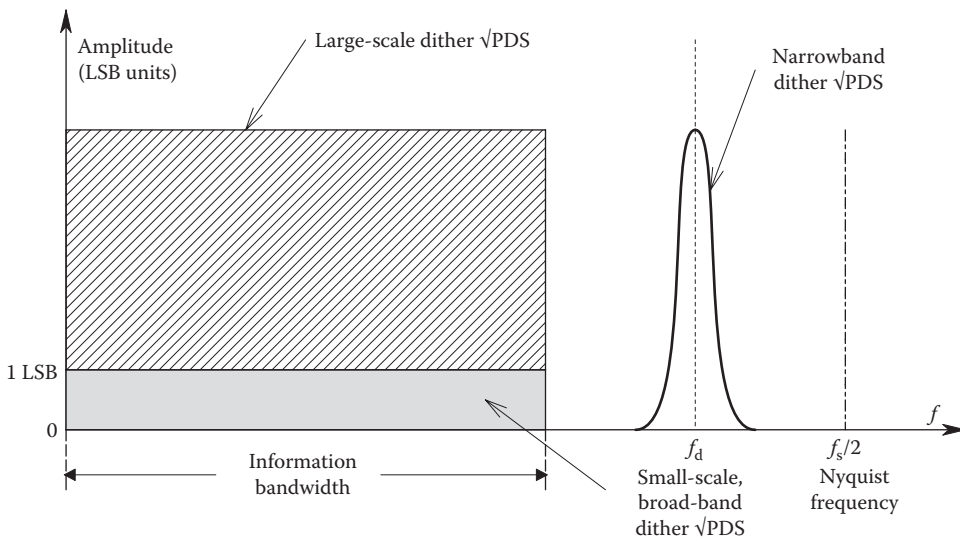


FIGURE 9.7 RMS spectra of signal, broad- and narrowband dithering signals.

inherent in the stepwise transfer function of an ADC. To illustrate this property heuristically, let us consider the expected value of the dither noise at the output of the digital LPF. In general, the expected or probability averaged value of the error transfer function, $e(n_d)$, is given by the well-known expression

$$\bar{e} = \int_{-\infty}^{\infty} e(n_d) p(n_d) dn_d. \tag{9.14}$$

Here, $p(n_d)$ is the PDF of the random variable, n_d , the dither noise. When dithering occurs, the analog signal of amplitude x is added to the RV, n_d . Now the expected transfer function value can be written as

$$\overline{e(x)} = \int_{-\infty}^{\infty} e(n_d) p(n_d + x) dn_d. \tag{9.15}$$

Equation 9.15 is mathematically a form of real convolution of $p(n_d)$ with $e(n_d)$ (the quantization error as a function of noise input). The convolution process is illustrated in Figure 9.8C. Note that the shift is x , and when the rectangular PDF of n_d is given the limits of $\pm q/2$, the integration result is zero for any $x \geq 0$.

According to HP (Bartz 1993), "The high-order distortion terms associated with the polynomial expansion of the quantized transfer function are effectively removed by the dithering. This result is true for dither amplitudes that are integral multiples of the ADC LSB [q]. The third-harmonic distortion is significantly improved by large-scale dithering."

In summary, we note that dithering provides a means to improve output SNR and reduce total harmonic distortion in a digital signal acquisition system.

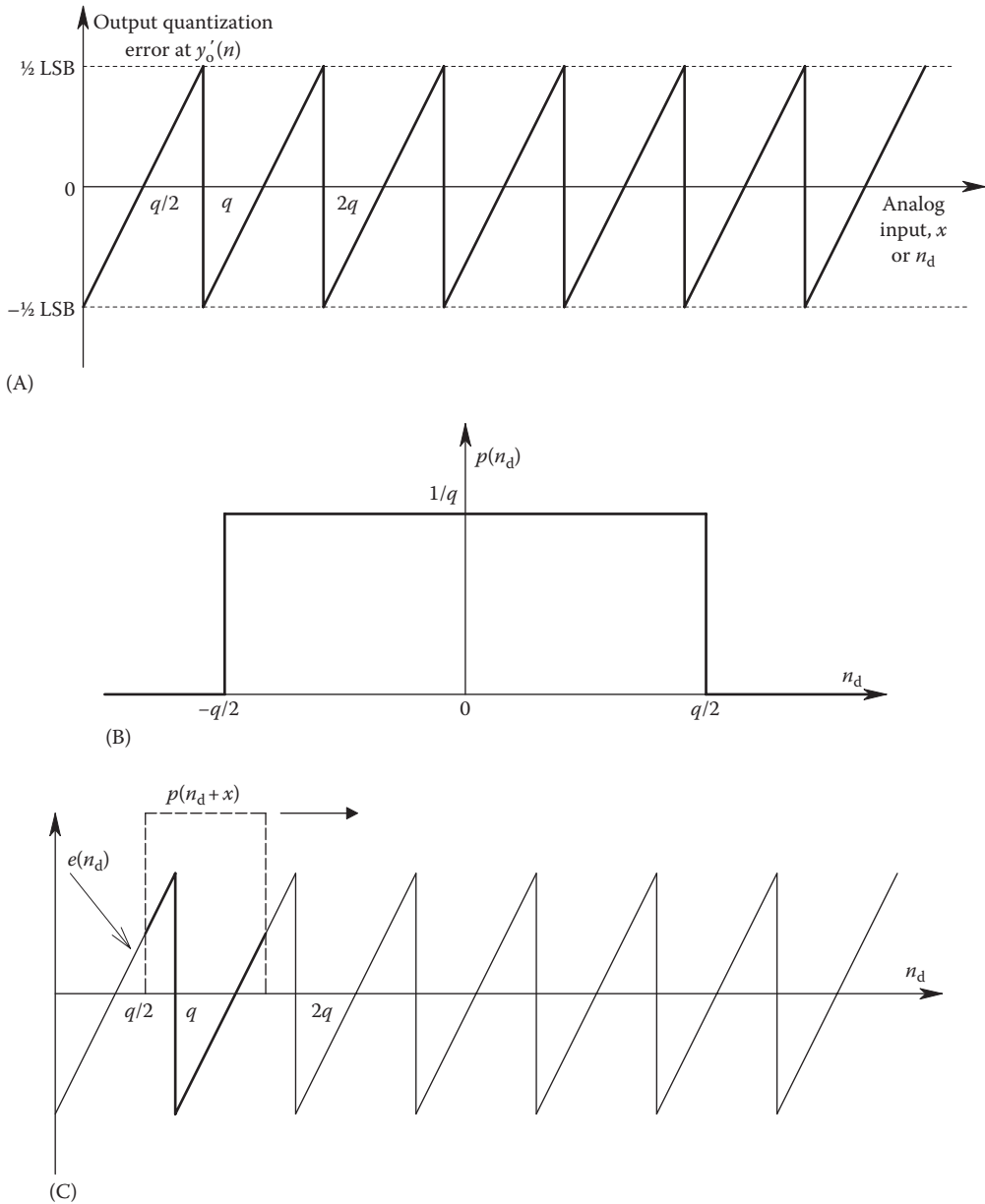


FIGURE 9.8 (A) Output quantization error vs. input x (cf. Figure 9.6). (B) PDF, $p(n_d)$, of the random variable, n_d , the dither noise. (C) Illustration of the convolution of $p(n_d)$ with the quantization error function, $e(n_d)$. The convolution acts to smooth and attenuate the amplitude of the quantization error.

9.5 DACs

A DAC is an IC device that converts an N -bit digital word to an equivalent analog voltage or current. DACs can operate at very high sampling rates or can act as static, DC voltage sources. A DAC allows digital information that has been processed and/or stored by a digital computer to be realized in analog form. DACs are essential components in the design of CRT terminals, modern control systems digital audio systems, and waveform synthesizers, to mention a few applications. DACs can be classified by whether they use BJT switches or MOS transistor switches.

As we mentioned in the introduction to this chapter, it is necessary to understand DACs before we examine the designs of certain ADCs, which use DACs as components. Most modern DACs are designed to accept straight binary inputs, although some units have been built that require BCD or gray code inputs. Most DACs use MOS or BJT switches to pass current through selected resistors in an R - $2R$ resistance ladder circuit. Either voltage or current reference sources are used to power the ladder. Either internal or external op amps are used to condition the DAC's output currents. DACs can be configured to have unipolar or bipolar outputs, depending on their application.

All DACs have a certain settling time required for the analog output to reach SS value following the digital input. This settling (conversion) time can be as long as 100 ms or as short as 10 ns. Immediately following a change in digital input, there is a DAC output transient, or glitch. The glitch can arise from the transient response of the op amp(s) used to condition the output voltage and also from internal switching transients in the DAC.

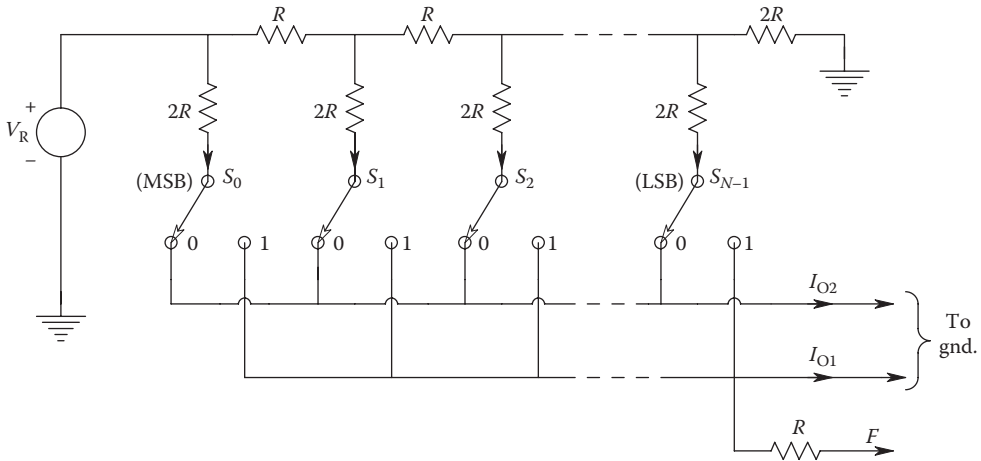
There are many designs for DACs that have evolved over the past 20 years or so. DACs can be classified by whether they use BJT switches or MOS transistor technology. The following DAC architectures are generally implemented with MOS technology:

- Binary-weighted resistor DAC
- R - $2R$ ladder
- Inverted R - $2R$ ladder
- Inherent monolithic ladder
- Switched-capacitor DAC

The following DAC designs generally use BJT technology:

- Binary-weighted current sources
- R - $2R$ ladder using current sources

At present, most DACs use the current-mode, R - $2R$ ladder design, as shown in Figure 9.9. This design has low switching transients and is relatively fast. The switches are typically MOS transistors with low ON resistance. V_R is the DAC's DC reference voltage, generally +5 or +10 V, although in some designs V_R can be a bipolar, time-varying signal, allowing two-quadrant multiplication of V_R times the digital input. A DAC is called a *multiplying DAC* (MDAC) when it is operated in this mode. The DAC of Figure 9.9 has two output lines in which current must flow to ground or an op amp's summing

**FIGURE 9.9**

An R - $2R$ current-scaling DAC ladder. S_0, \dots, S_{N-1} are SPDT MOS switches actuated by the binary input signal to the DAC. (From Northrop, R.B., *Introduction to Instrumentation and Measurements*, 2nd edn., Taylor & Francis, Boca Raton, FL, 2005.)

junction virtual ground at all times. Because of the grounded output lines, the current in the switched $2R$ resistors remains constant; hence, the node voltages, $V_1 \dots V_N$, remain constant, and there are minimum glitches caused by charging distributed capacitances from IC circuit elements to the IC substrate. Note that the V_R source sees a resistance of R ohms looking into the DAC, and thus the MSB current through S_1 is simply $V_R/2R$. Each successive switch current is $1/2$ of that through the preceding switch. Thus, the maximum output current, $I_{o(1)MAX}$ is given by

$$\begin{aligned} I_{o(1)MAX} &= \left(\frac{V_R}{R} \right) \left(1/2^1 + 1/2^2 + 1/2^3 + \dots + 1/2^N \right) \\ &= \left(\frac{V_R}{R} \right) \frac{2^N - 1}{2^N}. \end{aligned} \quad (9.16)$$

In general, the current outputs of this DAC are

$$I_{o(1)} = \left(\frac{V_R}{R} \right) \sum_{k=1}^{k=N} \left(\frac{D_k}{2^k} \right) \quad (9.17a)$$

$$I_{o(2)} = \overline{I_{o(1)}} = \left(\frac{V_R}{R} \right) \left[\frac{2^N - 1}{2^N} - \sum_{k=1}^{k=N} \left(\frac{D_k}{2^k} \right) \right]. \quad (9.17b)$$

Here, D_k is the logic level controlling the k th switch, either 0 or 1; $k = 1$ is the *most significant bit* (MSB); and $k = N$ is the LSB of the input word.

Figure 9.10 illustrates how op amps are used to convert the output currents given by Equations 9.17a and b to voltages. In the DAC circuit of Figure 9.6, note that there is an extra resistor of resistance R connected to the $I_{o(1)}$ node. This resistor is used as the feedback resistor for the op amp transresistor of Figure 9.10A. Negative unipolar voltage output is obtained because $V_o = -RI_{o(1)}$, where $I_{o(1)}$ is given by Equation 9.17a. (A second inverter can be used to obtain a positive V_o .) Figure 9.10B illustrates the same DAC

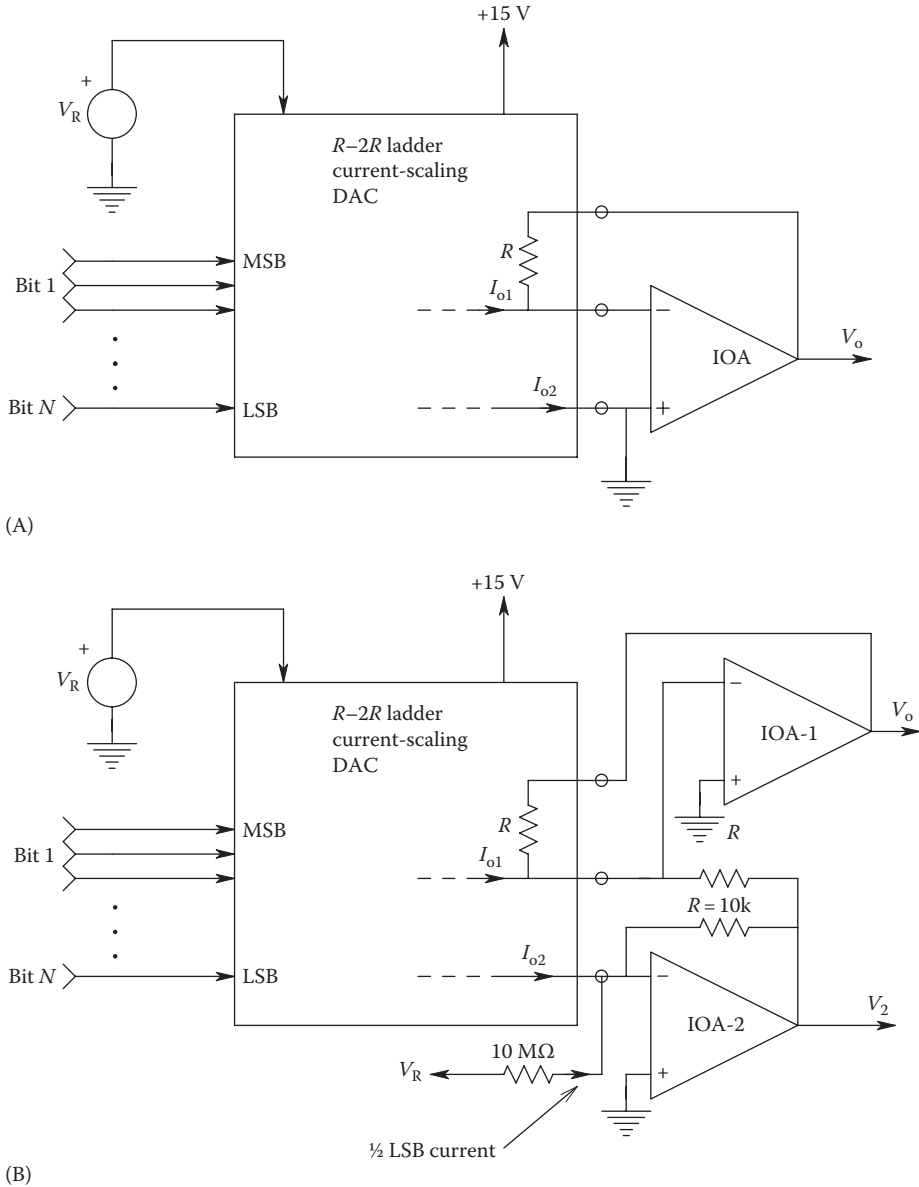


FIGURE 9.10 (A) A unipolar binary DAC circuit. The box contains the circuit of Figure 9.9. (B) Circuit for an offset binary (bipolar) DAC. $R = 10\text{ k}\Omega$. (From Northrop, R.B., *Introduction to Instrumentation and Measurements*, 2nd edn., Taylor & Francis, Boca Raton, FL, 2005.)

TABLE 9.2

Coding and Output for a 10-Bit, Offset Binary DAC (the $\frac{1}{2}$ LSB offset voltage has been neglected for simplicity)

D	V_o
11 1111 1111	$-V_R (511/512)$
10 0000 0001	$-V_R (1/512)$
10 0000 0000	0
01 1111 1111	$+V_R (1/512)$
00 0000 0000	$+V_R (511/512)$

given a bipolar output so V_o ranges over approximately $\pm V_R$. In this case, the input word must be in offset binary format, as illustrated in Table 9.2. In the general case, the output of the offset binary DAC can be written as

$$\begin{aligned}
 V_o &= -RI_{o(1)} - V_2 = -RI_{o(1)} - (-RI_{o(2)}) + \frac{\Delta V_o}{2} \\
 &= -V_R \sum_{k=1}^N \frac{D_k}{2^k} + V_R \left(\frac{2^N - 1}{2^N} - \sum_{k=1}^N \frac{D_k}{2^k} \right) + \frac{\Delta V_o}{2} \\
 &= -2V_R \sum_{k=1}^N \frac{D_k}{2^k} + V_R \left(\frac{2^N - 1}{2^N} \right) + \frac{\Delta V_o}{2}. \tag{9.18}
 \end{aligned}$$

The $\Delta V_o/2$ term is a $\frac{1}{2}$ LSB step used to make the bipolar DAC's transfer function an odd function to minimize quantization error.

Many other DAC designs exist. In many practical R - $2R$ ladder circuits, especially those used for high-speed operation, BJT current sources are used to drive the ladder nodes. Switching of these current sources is often accomplished by differential current switches that allow the currents through the BJT current sources to remain constant whether or not they flow into the ladder nodes or ground. This design is illustrated in Figure 9.11. Switched-capacitor or charge-scaling DACs that use MOS IC technology are available. Instead of a ladder geometry, a capacitive voltage divider is used with the upper, switched capacitors having the values, $C, C/2^1, \dots, C/2^{N-1}$. The lower capacitor of the divider is C pF. This circuit is shown in Figure 9.12. It can be shown (Northrop 1990) that the voltage output of the charge-scaling DAC is given by

$$V_o = \frac{V_R}{C_T} \sum_{k=1}^N D_k \left(\frac{C}{2^{k-1}} \right) \tag{9.19}$$

where C_T is the total capacitance of the switched array, given by

$$C_T = \sum_{k=1}^N \left(\frac{C}{2^{k-1}} \right) + \frac{C}{2^{N-1}} = 2C. \tag{9.20}$$

It is not practical to build switched-capacitor DACs for N s much above 8 because of the large range in sizes required between the MSB and LSB capacitors.

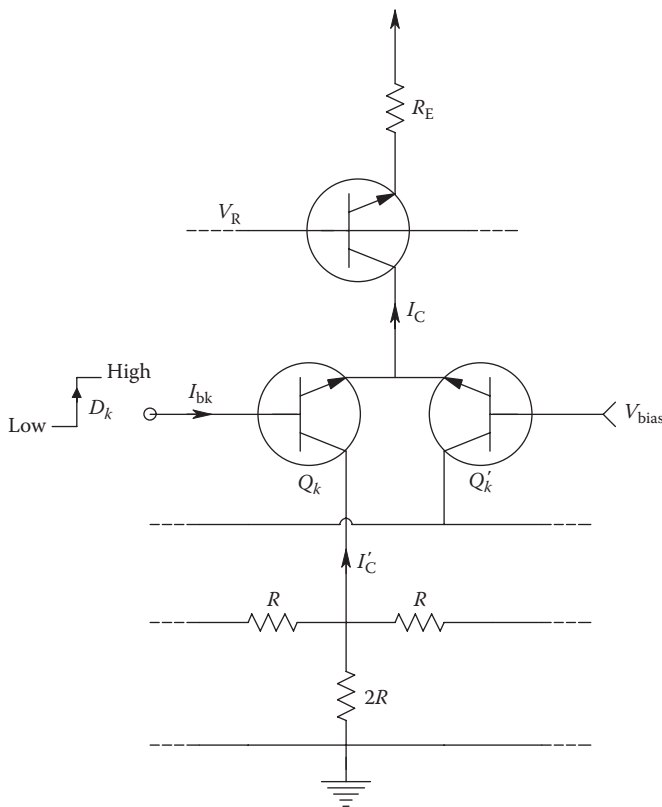


FIGURE 9.11 A differential BJT current switch commonly used in R-2R current scaling DACs. $D_k = k$ th digital input. The actual k th ladder current is $I'_C = I_C - I_{bk}$. (From Northrop, R.B., *Introduction to Instrumentation and Measurements*, 2nd edn., Taylor & Francis, Boca Raton, FL, 2005.)

The reader who is interested in the details of DAC designs and how DACs are specified should consult Section 10.3 in Northrop (2012), Chapter 11 in Franco (1988), Chapter 4 in Tompkins and Webster (1988), and Chapter 2 in Zuch (1981).

9.6 Hold Operation

Just as the sampling process can be interpreted in the frequency domain, so can the process of sample and hold of analog signals from a DAC. Note that the digital input to a DAC is generally periodic with period T_s . The DAC's analog output from the n th digital input is generally held constant until the $(n + 1)$ th input updates it. This process generates a step-wise output waveform if $\{b_k\}$ is changing. This process can be viewed as linear filtering operation. The impulse response of the zero-order hold (ZOH) filter to a unit input word at $t = 0$ and zero inputs at sampling instants thereafter is illustrated in Figure 9.13 and can be expressed mathematically as

$$h_o(t) = U(t) - U(t - T_s), \tag{9.21}$$

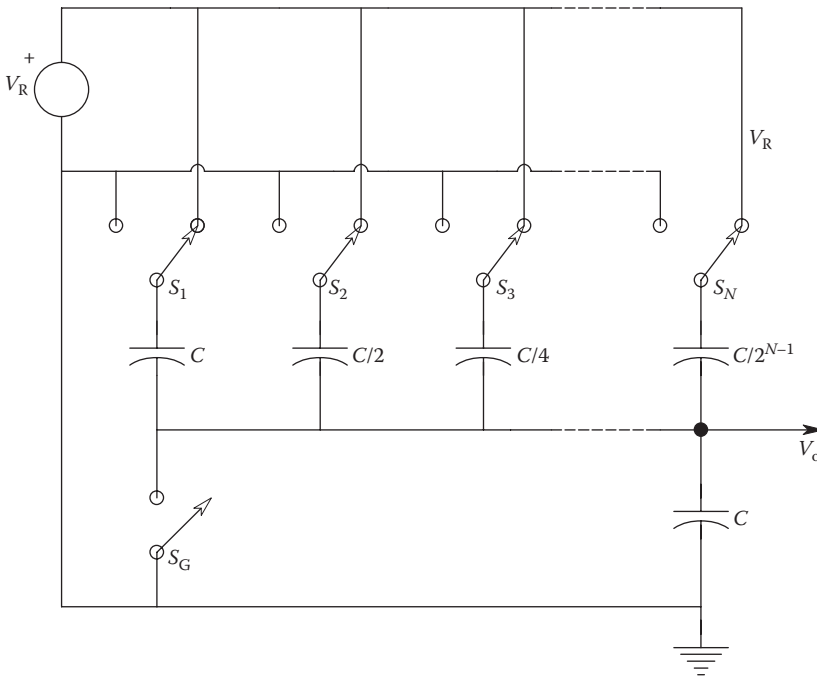


FIGURE 9.12 Simplified circuit of a switched-capacitor or charge-scaling DAC. The entire circuit is realized with MOS IC technology. (From Northrop, R.B., *Introduction to Instrumentation and Measurements*, 2nd edn., Taylor & Francis, Boca Raton, FL, 2005.)

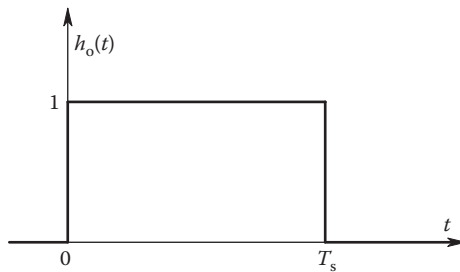


FIGURE 9.13 Impulse response of a ZOH circuit.

where $U(t)$ is the unit step function, defined as zero for all negative argument and 1 for all positive argument. $h_o(t)$ is thus a pulse of unit height and duration $(0, T_s)$, else zero. The Laplace transform of $h_o(t)$ is

$$H_o(s) = \frac{1}{s} - \frac{1}{s} e^{-sT_s} = \frac{1 - e^{-sT_s}}{s}. \tag{9.22}$$

To find the frequency response of the ZOH, we let $s \rightarrow j\omega$ and use the Euler relation for $e^{j\theta}$. The ZOH's frequency response is easily shown to be

$$H_o(j\omega) = T_s \frac{\sin(\omega T_s/2)}{\omega T_s/2} e^{-j\omega T_s/2}. \tag{9.23}$$

Note that the zeros in $H_o(j\omega)$ occur at $\omega = n2\pi/T_s$ r/s, where $n = 1, 2, 3, \dots$, and $2\pi/T_s$ is the radian sampling frequency of the system. Thus, the overall process of sampling an analog signal, $x(t)$, and reconverting to (held) analog form by a DAC can be written in the frequency domain (neglecting quantization) as (Northrop 1990)

$$X(j\omega) = \sum_{n=0}^{\infty} \frac{\sin(\omega T_s/2)}{\omega T_s/2} X(j\omega - jn\omega_s) e^{-j\omega T_s/2}. \quad (9.24)$$

Note that other more sophisticated holds exist, their realization is at the expense of some circuit complexity, however. For example, the *first-difference extrapolator hold*, which generates linear slope transitions between sampling instants (instead of steps), can be realized with three DACs, a resettable analog integrator, and an analog adder (Northrop 1990). Its transfer function can be shown to be

$$H_c(s) = \frac{1 - e^{-sT}}{s} + \frac{1 - 2e^{-sT} + e^{-sT}}{s^2} - \frac{(1 - e^{-sT})e^{-sT}}{s}. \quad (9.25)$$

It is left as an exercise for the reader to sketch and dimension $h_c(t)$.

9.7 ADCs

Various types of ADCs have evolved to meet specific applications in instrumentation, communications, control, and audio. Obviously, the speed of the ADC rate is not critical when digitizing a DC measurand; however, the resolution may be very important, so a large number of bits may be used. For example, a 16-bit ADC has resolution limited by $\pm\frac{1}{2}$ LSB, or ± 7.63 ppm, while a 20-bit ADC can resolve to ± 0.477 ppm. Commercial, plug-in, data acquisition cards for PCs of various types generally do not use ADCs of greater than 16 bits. Thus, custom ADC interface systems must be used if resolution better than ± 7.63 ppm is desired. The fastest ADCs are the *flash converters*, described in the following, which can convert 8 bits with a sampling period of less than 1 ns. Such speed is useful when measuring transient phenomena such as partial discharges in power cable insulation, other insulation breakdown phenomena, and transient events in particle physics and lasers.

One might wonder why it is necessary to have 16-bit resolution when the display resolution of terminals and graphics plotters rarely exceeds 10 bits. The answer to this question lies in the lower quantization noise of the 16-bit systems and the fact that the statistics computed on the measurements have lower variances when calculated using 16-bit data. We will describe the organization, features, and applications in measurements of five major types of ADCs: (1) *successive-approximation (SA) converters*; (2) *tracking (servo) types*; (3) *dual-slope, integrating ADCs (DSI ADCs)*; (4) *flash (parallel) converters*; (5) *dynamic range, floating point converters (DRFP)*; and (6) *sigma-delta converters*.

9.7.1 Successive-Approximation ADCs

SA ADCs are probably the most widely used class of ADC, with accuracies ranging from 8 to 16 bits. They are available as single, LSI, digital/analog ICs. Their advantages include low cost and moderate conversion speed; however, they can have missing output codes, require a sample-and-hold input interface, and are difficult to auto-zero.

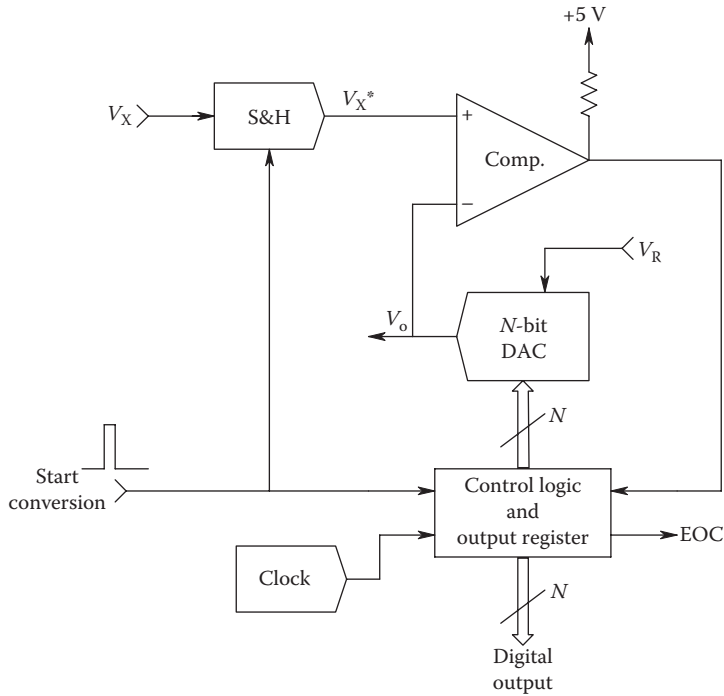


FIGURE 9.14

Block diagram of an SA ADC. S&H is an analog sample-and-hold (or track-and-hold) circuit used to freeze V_X during the ADC process. (From Northrop, R.B., *Introduction to Instrumentation and Measurements*, 2nd edn., Taylor & Francis, Boca Raton, FL, 2005.)

A block diagram of an SA ADC is shown in Figure 9.14. Note that this type of ADC uses a DAC and an analog comparator in a logical feedback loop to control its conversion algorithm. SA ADCs are fast enough for audio-frequency applications, with conversion times ranging from about 2 to 20 μs . Conversion time, as we will see, depends on the digital clock of the SA ADC and its number of bits.

The conversion cycle of an 8-bit SA ADC begins with the analog input signal being sampled and held at $t = 0$. Simultaneously, the output register is cleared and all D_k are set to 0. At the next clock cycle, D_1 is set to 1 with all other $D_k = 0$. This makes the DAC output $V_o = V_R(128/256) = V_R/2$. The comparator performs the operation $\text{SGN}(V_X - V_R/2)$. If $\text{SGN}(V_X - V_R/2) = 1$, then D_1 is kept 1; if $\text{SGN}(V_X - V_R/2) = -1$, then D_1 is set to 0. This completes the first (MSB) cycle in the conversion process. Next, D_2 is set to 1 (D_1 remains the value found in the first cycle). The comparator output signals if $V_X > [D_1(V_R/2) + V_R/4]$. If yes, then D_2 is set to 1; if no, $D_2 = 0$, completing the second bit's conversion cycle. This process continues until all N bits are converted and then stops and signals *data ready*. It is easily seen that N clock cycles are required to convert V_X to an N -bit digital word. Note that the final conversion error is less than $\pm 1/2$ LSB. The SA ADC process is shown in a flowchart form in Figure 9.15.

SA ADCs are designed to work at audio frequencies and higher, with their conversion cycles under computer control. Often, their output registers use tristate logic, so that an ADC can be used with other tristate SA ADCs and DACs on the same, bidirectional data bus. Note that if N , N -bit SA ADCs are on the same data bus, and their starting times are staggered one after another, an ADC output can be read into the computer every clock cycle.

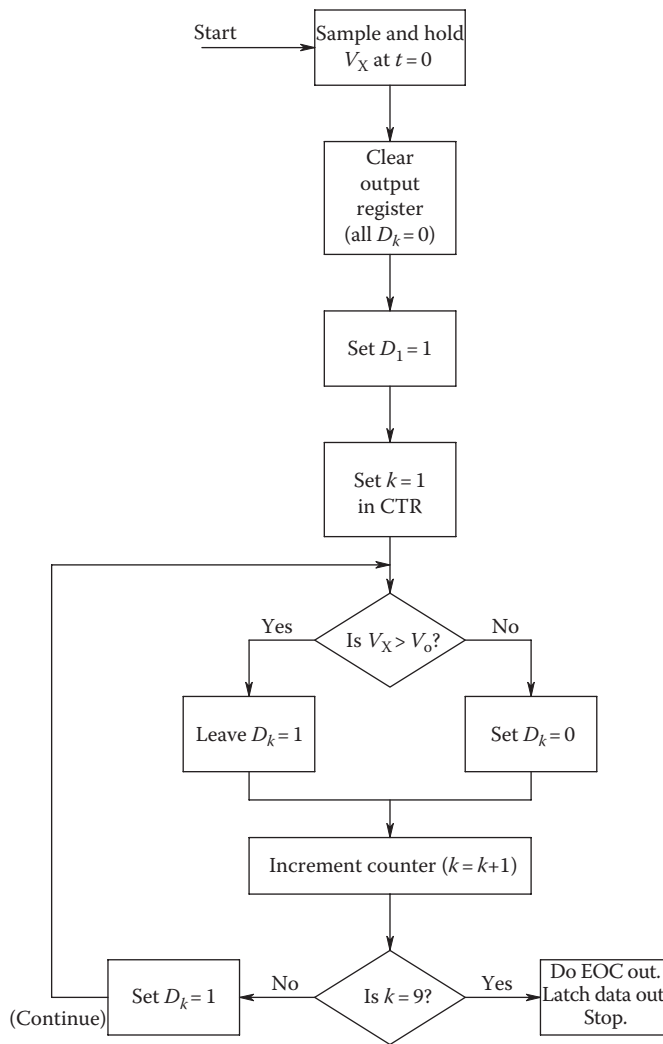
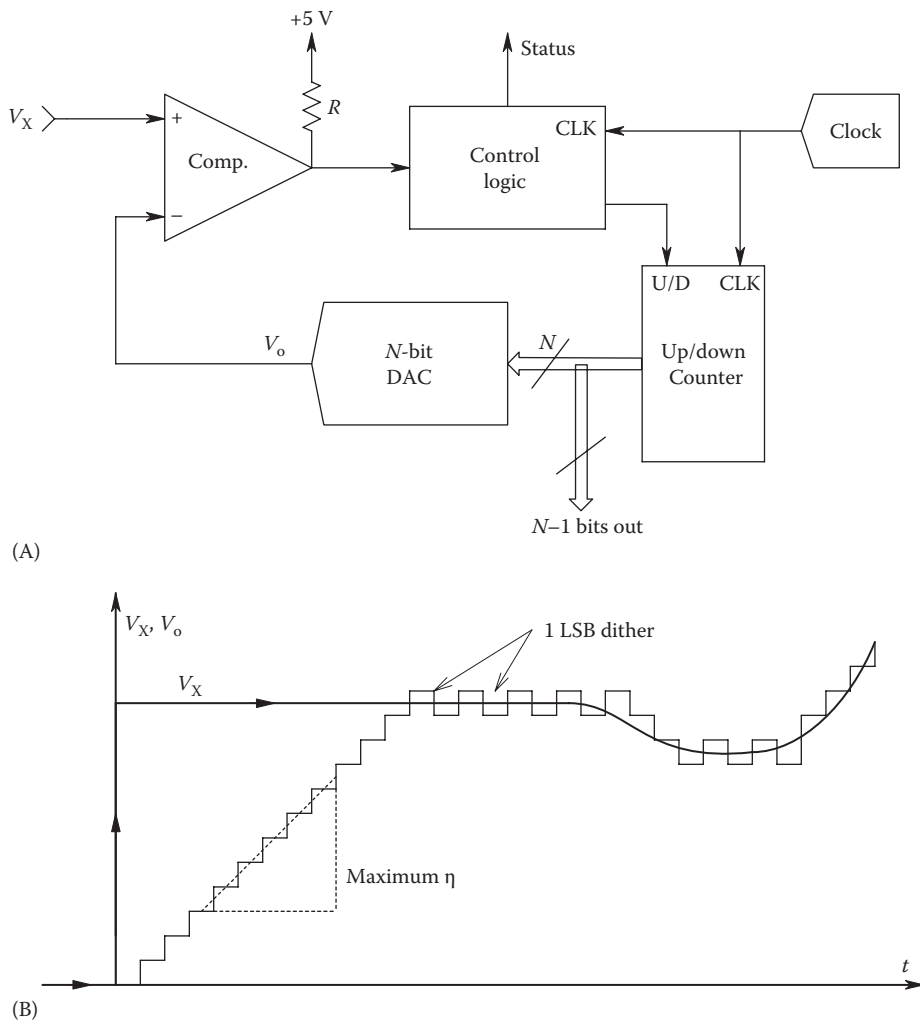


FIGURE 9.15 Flowchart for the operation of a typical, 8-bit, successive approximation ADC. Note: D_1 =MSB. D_8 =LSB. (From Northrop, R.B., *Introduction to Instrumentation and Measurements*, 2nd edn., Taylor & Francis, Boca Raton, FL, 2005.)

9.7.2 Tracking or Servo ADCs

The tracking ADC (T-ADC) is a relatively slow ADC, best suited for the digitization of DC and low audio-frequency measurands. Figure 9.16A illustrates the block diagram of a T-ADC. Figure 9.16B illustrates the input voltage, $V_x(t)$, and the DAC output, $V_o(t)$, in a T-ADC when it is first turned on. Note that the DAC output slews up to V_x at a rate set by the DAC's clock. When V_o exceeds V_x , the comparator's output goes low, which causes the counter to count down one clock cycle to bring $V_o < V_x$. Then the comparator goes high, signaling an up count. In the following, a very slow AC or DC V_x waveform, V_o , is seen to have a ± 1 LSB limit cycle around the true value of V_x . This means that in realizing a digital output, the true LSB digit must be dropped because of the limit cycle oscillation. Thus, an 11-bit DAC is required to realize a 10-bit output T-ADC.

**FIGURE 9.16**

(A) Block diagram of a tracking (servo) ADC. (B) Waveforms in a T-ADC. (From Northrop, R.B., *Introduction to Instrumentation and Measurements*, 2nd edn., Taylor & Francis, Boca Raton, FL, 2005.)

A T-ADC has a maximum slew rate, η , in volts per second, at which it can follow a rapidly changing $V_X(t)$ without gross error. This is

$$\eta = \frac{V_{o(\text{MAX})}}{2^N T_C} \text{ V/s.} \quad (9.26)$$

As an example, let $N = 10$ bits, $V_{o(\text{MAX})} = 10 \text{ V}$, and $T_C = 1 \mu\text{s}$. η is found to be $9.766 \times 10^3 \text{ V/s}$. Thus, $dV_X(t)/dt$ must be less than this η for no gross error. The T-ADC, in addition to slew-rate limitations, performs poorly in the presence of high-frequency noise on V_X . Thus, low-pass filtering of V_X before it is seen by the comparator improves T-ADC performance. This type of ADC is suitable for the design of low-cost, digital, DC voltmeters and ammeters.

9.7.3 Dual-Slope Integrating ADCs

DSI ADCs have the advantages of high inherent accuracy (up to 22 bits output) and excellent high-frequency noise rejection, they do not require a sample and hold at their inputs, and there is no possibility of missing output codes. They are widely used in inexpensive, DC, digital instruments. Their conversion times are long, ranging from about 10 ms to 0.333 s. They generally have an auto-zeroing mode, discussed in the following. Figure 9.17 illustrates a DSI ADC that converts a unipolar, DC input. Note that a DAC is not used in this ADC design, but a high-performance analog integrator is required. The control logic operates MOS analog switches, S_1 and S_2 . The conversion cycle of a DSI ADC begins with the control logic first clearing the counter and setting the charge on the integrator capacitor equal to zero with switch S_2 . Next, $V_X \geq 0$ is integrated for 2^N clock cycles. At the end of this integration time, T_1 , the integrator output, V_2 , can be expressed as

$$V_2 = -\frac{1}{RC} \int_0^{T_1} V_X(t) dt \quad T_1 = -\frac{T_1}{RC} \langle \overline{V_X} \rangle T_1, \tag{9.27}$$

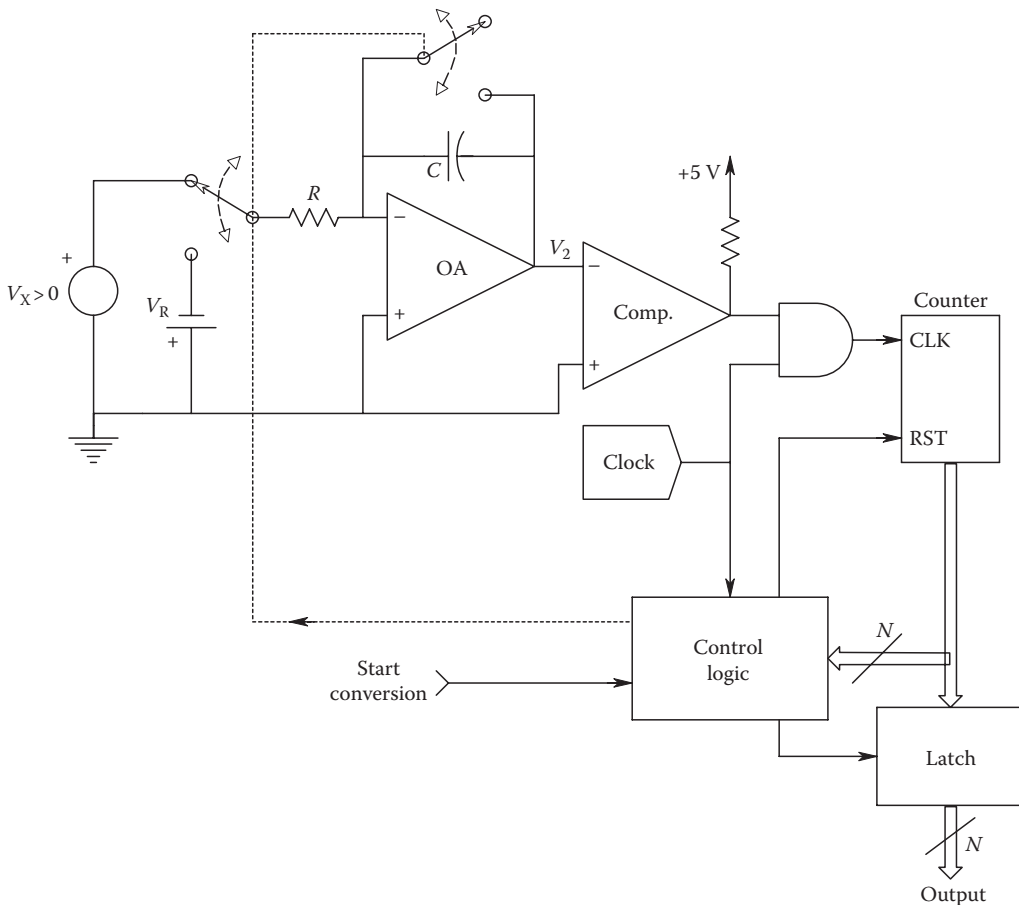


FIGURE 9.17 Diagram of a unipolar, DSI ADC. (From Northrop, R.B., *Introduction to Instrumentation and Measurements*, 2nd edn., Taylor & Francis, Boca Raton, FL, 2005.)

where $T_1 = 2^N T_C$. Switch S_1 now connects the integrator input to $-V_R$. V_2 now ramps linearly toward 0 V with positive slope. The counter is made to count clock cycles from T_1 to T_F when V_2 reaches 0 V. At T_F the counter count, M , is latched into the output register. At T_F we can write

$$V_2 = 0 = -\left(\frac{T_1}{RC}\right)\langle\overline{V_X}\rangle T_1 + T_2 V_R / RC, \quad (9.28)$$

where $T_2 = T_F - T_1$. However,

$$T_2 = MT_C, \quad (9.29)$$

so we can write

$$\frac{MT_C V_R}{RC} = \frac{2^N T_C \langle\overline{V_X}\rangle_{T_1}}{RC}, \quad (9.30)$$

which leads to

$$M = \frac{\langle\overline{V_X}\rangle_{T_1} 2^N}{V_R}. \quad (9.31)$$

Thus, we see that the numerical count, M , is proportional to V_X averaged over T_1 seconds.

Calibration of the DSI ADC is seen to be independent of values of the clock period, R , and C . It *does require* a temperature-stabilized V_R and an integrator op amp and comparator with low offset voltage drift. In addition, the integrator op amp must have negligible DC bias current. DSI ADCs are widely used in digital multimeters. They are available as LSI ICs and generally have an auto-zeroing mode. In one strategy, the output count, M_o , measured with the input shorted to ground ($V_X = 0$), is read every other conversion cycle and then subtracted from M obtained with V_X connected. Another auto-zeroing strategy is analog based in which a special auto-zero capacitor is charged to the net offset voltages of the buffer amplifier, integrator op amp, and the comparator during a special auto-zero cycle with inputs shorted to ground. The auto-zero capacitor, charged to the net offset voltage, is then switched in series with the input during the measurement cycles to effectively reduce the net ADC offset voltage to less than 10 μV (this auto-zero architecture is used in the Intersil ICL7106/7107 DSI ADC) (see Intersil AN023).

It can be shown that if T_1 is made an integer number, K , of power line frequency periods, that is, $T_1 = 2^N T_C = K/60$ s, then power line hum contaminating V_X can be rejected by as much as 70 dB. Other modifications of the DSI ADC include a connection for bipolar input signals, which gives an offset binary output code, and charge-balancing ADCs in which a reference current source switched into the integrator's summing junction replaces V_R in the DSI ADC previously described (Northrop 1990).

Some commercially available, dual-slope ADCs include the Analog Devices' 22-bit AD1175K; the Intersil ICL7106/7107, 3½ digit, auto-zeroing ADC for LCD display digital multimeters; the Intersil ICL7129, 4½ digit DMM DSI ADC; and the Texas Instruments ICL/TLC 7135C, 4½ digit DMM DSI ADC.

9.7.4 Flash (Parallel) ADCs

Flash ADCs (FADCs) are the cornerstone of modern, ultrahigh-speed, digitizing front ends of certain digital storage oscilloscopes (DSOs), such as manufactured by LeCroy, Nicolet, HP, and Tektronix. Eight-bit digitizing rates in excess of 1 gigasamples/s (10^9 samples/s) are currently available for specialized DSOs. The digitized outputs of such FADCs must be buffered by ECL or proprietary logic registers, which store and download the digital data rates acceptable by computer bus structures, logic, and RAM.

An FADC can be realized as a hybrid or LSI IC. From Figure 9.18, we see that an FADC is composed of four subsystems: (1) an analog track-and-hold or sample-and-hold circuit that is used to sample and *freeze* the input voltage and present that sample to the inputs of the high-speed, analog, amplitude comparators; (2) a stable, DC reference voltage source and a voltage divider ladder that supplies the switching reference voltages to the analog comparators; (3) 2^N , high-speed, analog voltage comparators having negligible DC offset voltages; and (4) a combinational logic circuit composed of OR, NOR, and inverter gates that accepts the logic level outputs of the 2^N comparators and that generates a parallel digital word at the FADC output.

Practical considerations in the design of IC FADCs having the architecture of Figure 9.18 limit the number of bits to eight. Problems exist with the input capacitances of the 28 comparators loading the track-and-hold circuit, signaling propagation delays, and maintaining low DC offset voltages on the 256 comparators.

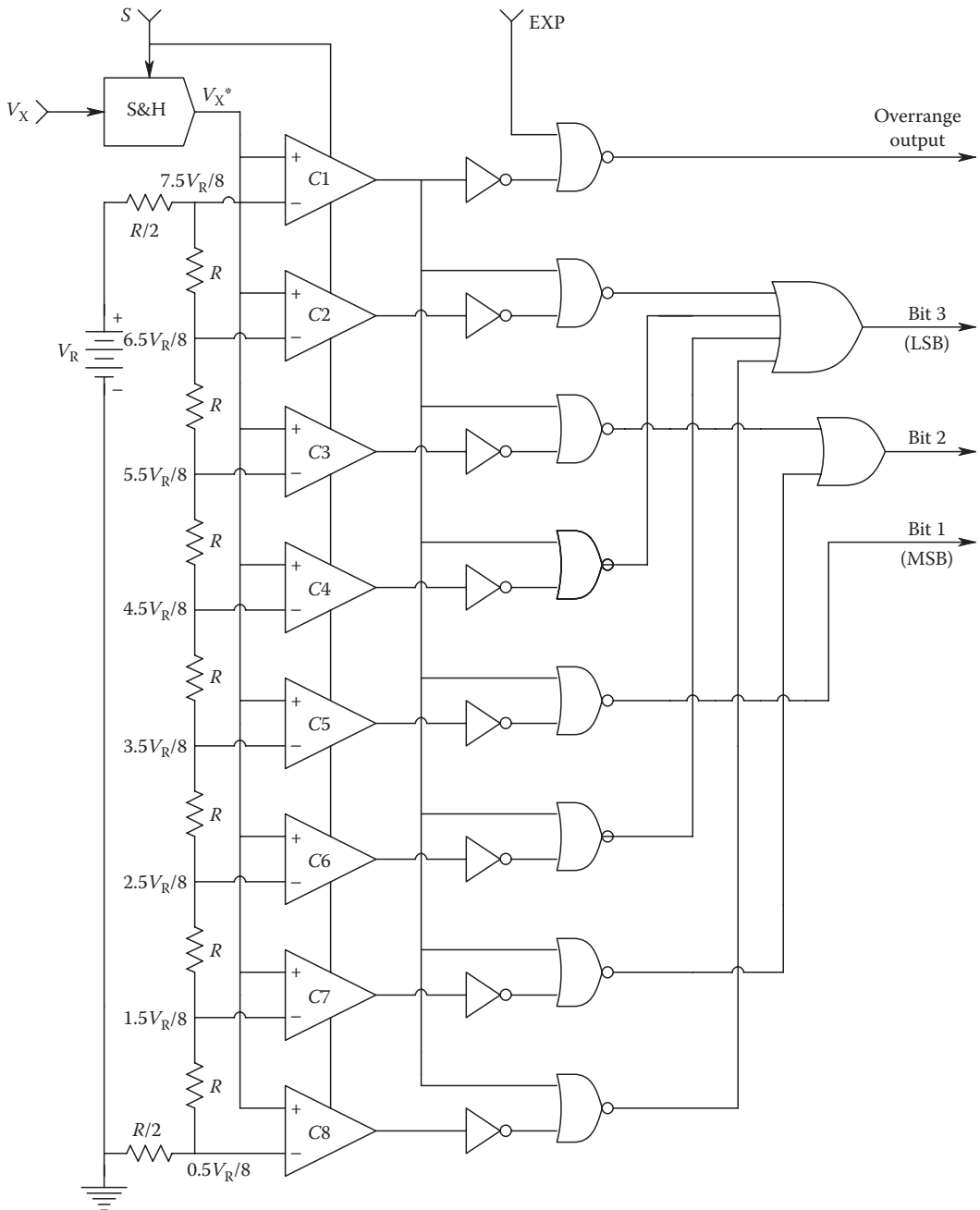
A design for an FADC having a 12-bit output is shown in Figure 9.19. This system uses two 6-bit FADCs, a very fast 6-bit DAC, and two sample-and-hold circuits. Conversion speed of this system will be slower than 6-bit FADC alone because of the conversion lag and settling time of the DAC and $K_V = 2^6$ amplifier. A conversion cycle of the 12-bit, two-stage FADC operates as follows: S&H-1 samples the analog input signal, giving an analog output V_X . V_X is flash converted by the first 6-bit FADC and simultaneously reconverted to analog form by the 6-bit DAC. The DAC output, V_1 , is subtracted from V_X , giving the analog quantization error, V_E , of the first ADC. V_E is amplified by a factor of $2^6 = 64$, again sampled and held, and then converted by the second FADC to give the 6 LSB output. In mathematical terms,

$$V_1 = V_{X(\text{MAX})} \sum_{n=1}^6 \frac{D_n}{2^n} \quad (9.32)$$

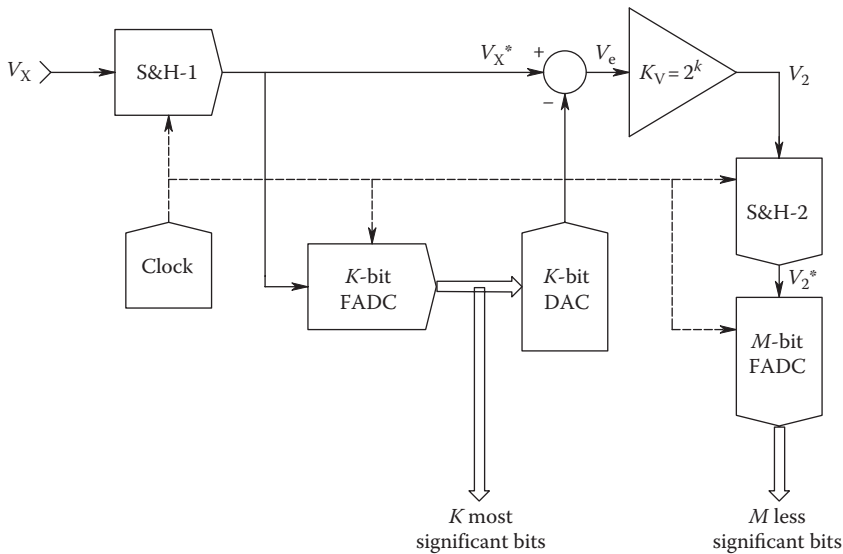
$$V_2 = 2^6 V_E = 2^6 (V_X - V_1) = 2^6 \left(V_X - V_{X(\text{MAX})} \sum_{n=1}^6 \frac{D_n}{2^n} \right) \quad (9.33)$$

In general, V_E can be as large as $V_{X(\text{MAX})}/2^6$; hence, the second 6-bit converter generates a binary code on the quantized remnant, V_E . Internally, the two 6-bit FADCs each requires $(2^6 - 1)$ comparators and thus is far less complex than even an 8-bit FADC. It can be shown that the total coded output of this system can be written as

$$V_X = V_{X(\text{MAX})} \sum_{k=1}^{12} \frac{D_k}{2^k} \quad (9.34)$$

**FIGURE 9.18**

Schematic of a 3-bit FADC with binary output. (From Northrop, R.B., *Introduction to Instrumentation and Measurements*, 2nd edn., Taylor & Francis, Boca Raton, FL, 2005.)

**FIGURE 9.19**

Block diagram of a two-step, FADC with $K + M = N$ output bits. (From Northrop, R.B., *Introduction to Instrumentation and Measurements*, 2nd edn., Taylor & Francis, Boca Raton, FL, 2005.)

In general, FADCs are expensive, specialized ADCs whose principal application is the digitization of analog signals with frequencies from 10 MHz to over 1 GHz. They can, of course, operate at slower sampling rates and are widely used in the front ends of DSOs, which operate on signals having frequencies from mHz to 500 MHz. Eight-bit resolution is typical.

Some examples of FADC ICs include the Burr-Brown/TI ADC600, a 12-bit, 10 megasamples/s (MSPS) IC that uses the two-stage architecture described earlier; the Analog Devices AD770 8-bit, 200 MSPS IC; and the Sony CXA1096P 8-bit, 20 MSPS converter. The state of the art in flash converters lies close to the proprietary NIM BIN/CAMAC digitizing system offered by LeCroy. The venerable LeCroy Model 6880B waveform digitizer system could sample at a fixed 1.35 GSps, giving either an 8-bit or 11-bit output, stored in a 10,016-word buffer memory holding 7.42 μ s of sampled data.

9.7.5 Dynamic Range, Floating Point ADCs

Figure 9.20 illustrates a block diagram of the Micro Networks model MN-5420, DRFP ADC. This ADC has an accuracy better than 2 ppm and can do 3.2×10^5 conversions/s. It has a 16-bit output consisting of a 12-bit mantissa and a 4-bit exponent. The 4-bit FADC codes the exponent, and its output sets the gain of a programmable gain amplifier (PGA) having gains $K_V = 2^k$, $k = 1, 2, 4, \dots, 128, 256$. The settled amplifier output goes through a track-and-hold circuit and is converted by a fast, 1 μ s, 12-bit, SA ADC. When the analog input signal is small, the gain of the PGA is high. For example, for $V_X < 19.5$ mV, the gain is 256. The value of a mantissa LSB at this gain is 9.5 μ V. As designed, the maximum input voltage of the MN-5420 DRFP-ADC is ± 5 V. Table 9.3 gives this ADC's gain switching points and the mantissa's LSB values. When V_X goes negative, the exponent coding is the same as for positive V_X ; however, the MSB of the 12-bit mantissa goes from 0 to 1.

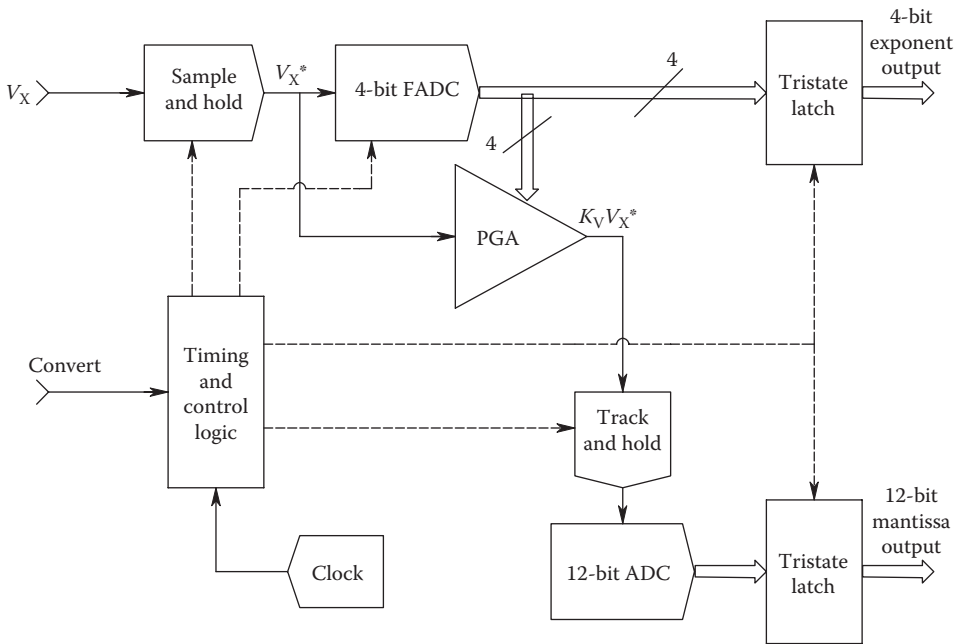


FIGURE 9.20 Block diagram of the MN-5420, DRFP ADC. The programmed gain amplifier (PGA) has an auto-zeroing feature (not shown).

TABLE 9.3

MN-5420, DRFP ADC PGA Switching Points and Exponent Coding for $V_X > 0$ (V_X range is ± 5 V)

PGA Switching Voltage, V_X	PGA Gain	Exponent	LSB Voltage
$+5 \leq V_X \leq 2.5$ V	1	1 0 0 0	2.44 mV
$2.5 < V_X \leq 1.25$ V	2	0 1 1 1	1.22 mV
$1.25 < V_X \leq 0.625$ V	4	0 1 1 0	610 μ V
$0.625 < V_X \leq 0.3125$ V	8	0 1 0 1	305 μ V
$0.3125 < V_X \leq 0.15625$ V	16	0 1 0 0	153 μ V
$0.15625 < V_X \leq 0.078125$ V	32	0 0 1 1	76 μ V
$0.078125 < V_X \leq 0.039063$ V	64	0 0 1 0	38 μ V
$0.039063 < V_X \leq 0.019531$ V	128	0 0 0 1	19 μ V
$0.019531 < V_X \leq 0$ V	256	0 0 0 0	9.5 μ V

The MN-5420 DRFP-ADC design is an ingenious approach to high-resolution, high-speed ADC. The heart of the design is the precision, auto-zeroing, PGA. The overall ADC speed is limited by the speed at which gains can be switched and the speed of the 12-bit SA ADC. Note from Table 9.3 that the resolution of the DRFP ADC is 1.9 ppm on its most sensitive auto range.

9.7.6 Delta-Sigma ADCs

Delta-sigma ADCs are also called *oversampled* or *noise-shaping converters*. Oversampling ADCs are based on the principle of being able to trade-off accuracy in time for accuracy in

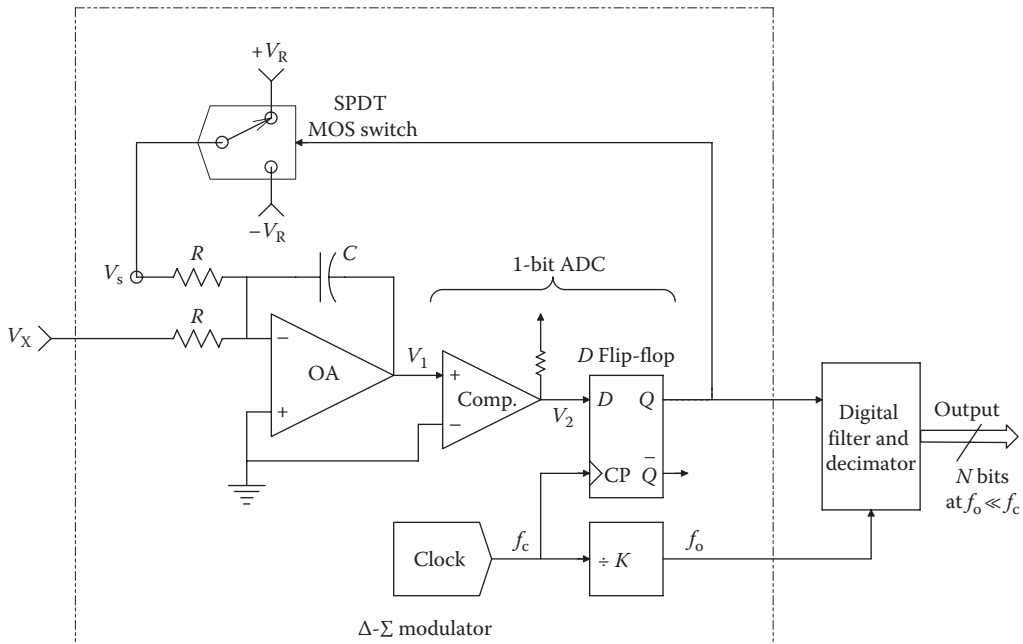


FIGURE 9.21
Circuit of a first-order Δ - Σ ADC. V_x is the input voltage.

amplitude. Figure 9.21 illustrates the circuit of a *first-order*, Δ - Σ ADC. The circuit in the box is a first-order Δ - Σ modulator. The digital filter and decimator produces the N -bit, digital output.

We will first analyze how the Δ - Σ modulator operates in the time domain. First, let us set $V_x \rightarrow 0$. Let the MOS switch apply $-V_R$ to the integrator so that its output, V_1 , goes positive. The comparator output, V_2 , applied to the D input of the D flip-flop (DFF) will be HI. At the rising edge of the next clock pulse, HI is seen at the Q output of the DFF. This HI causes the MOS switch to switch the integrator input to $+V_R$, so the integrator output V_1 begins to ramp down. When V_1 goes negative, the comparator output V_2 goes LO. Again, when the clock pulse goes HI, the LO at D is sent to the DFF Q output. This LO again sets the MOS switch to $-V_R$, causing V_2 to ramp positively until it goes positive, setting V_2 HI, etc. Figure 9.22 shows the Δ - Σ modulator's waveforms for $V_x = 0$. Q is the DFF output that is the comparator output, V_2 , latched in when the CP goes high. The peak-to-peak height of the SS triangle wave oscillation seen at V_1 can easily be shown to be equal to

$$2V_{1m} = \frac{V_R T_c}{RC}, \tag{9.35}$$

where T_c is the DSDAC's clock period and the other parameters are given in Figure 9.22. The average value of $V_s = 0$.

Now, let us set $V_x = +V_R/4$ and examine the Δ - Σ ADC's waveforms. Refer to Figure 9.23. Let the system start with $V_1 = 0$ at $t = 0$, and begin to integrate V_x and $V_s = +V_R$. Initially, the integrator output V_1 ramps negative as $v_1(t) = -(5V_R/4RC)t$. At the first positive-going

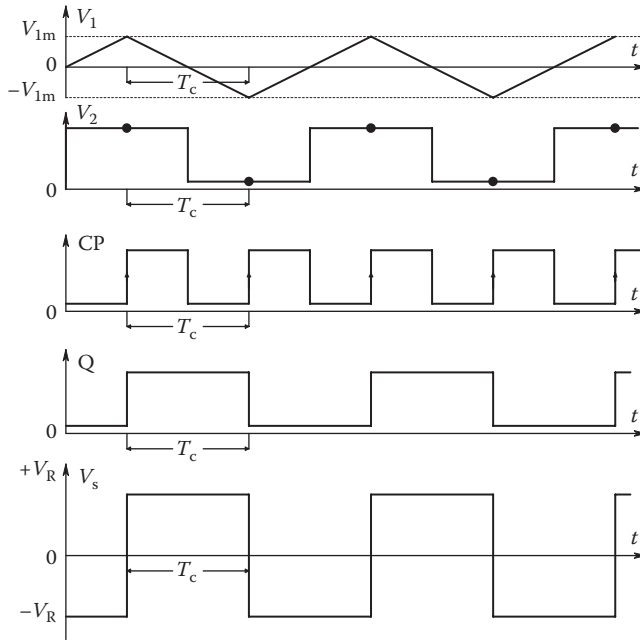


FIGURE 9.22
Waveforms in the Δ - Σ ADC of Figure 9.21 when $V_x = 0$.

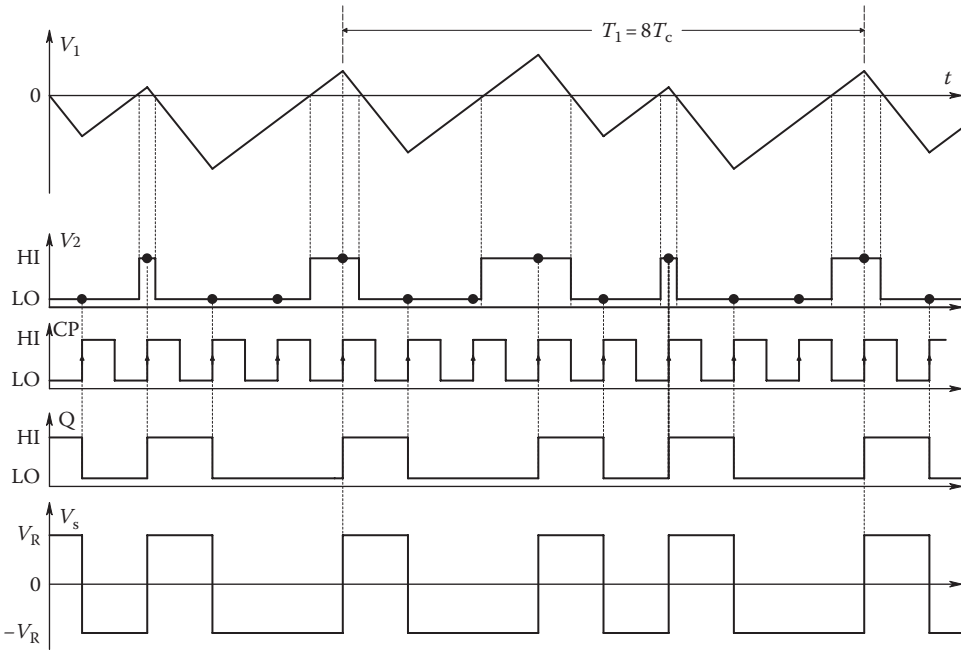


FIGURE 9.23
Waveforms in the Δ - Σ ADC of Figure 9.21 when $V_x = V_R/4$.

clock transition, the LO comparator output, V_2 , is latched to the DFF's Q output, setting it low. When Q goes low, the MOS switch makes $V_s = -V_R$. Now, the integrator output ramps upward with slope $+3V_R/4RC$. When V_1 goes positive, the second positive clock transition latches the HI V_2 to Q, setting V_s to $+V_R$. Now the integrator output, V_1 , ramps down again with slope $-5V_R/(4RC)$. When $V_1 < 0$, V_2 goes LO. The third positive clock transition latches this LO to the Q output of the DFF. As before, this sets V_s to $-V_R$, and the integrator output again ramps positive and the cycle repeats. Note that there is an SS oscillation in V_1 , V_2 , Q, and V_s with a period of eight clock periods. The SS duty cycle of Q is 37.5%, and $\bar{V}_s = -(1/4)$.

It is also of interest to examine the Δ - Σ ADC's waveforms when $V_x = +V_R/2$. Now, let the system start with $V_1 = 0$ at $t = 0$, and begin to integrate V_x with $V_s = +V_R$. Assuming V_x is constant, V_1 ramps negative as $v_1(t) = -3V_R t/(2RC)$, causing the comparator output, V_2 , to be LO. At the first positive-going clock pulse, the DFF latches the low D input to Q, causing the MOS switch to make $V_s = -V_R$. Now, the net input to the integrator is $-V_R/2$, and the integrator output begins to go positive with slope $+V_R/(2RC)$, as shown in Figure 9.24. When V_1 first goes positive, V_2 goes HI. At the third clock HI transition, the DFF output goes HI, setting the MOS switch to $+V_R$. Now, the net input to the integrator is $+3V_R/2$, and V_1 begins to ramp down again with slope $-3V_R/(2RC)$, and the ADC process repeats itself indefinitely until V_x changes. From the waveforms in this figure where $V_x = V_R/2$, we see that the duty cycle of Q is 25%, and $\bar{V}_s = -(1/2)$. From the three sets of V_s waveforms, we see that in general, $V_x = -V_R V_s$.

Now, let us consider the operation of first-order Δ - Σ ADC. If the number of ones in the output data stream (i.e., the DFF's Q output) is counted over a sufficient number of samples (clock cycles), the counter's digital output will represent the digital value of the analog input, V_x . Obviously, this method of averaging will only work for DC or low-frequency V_x .

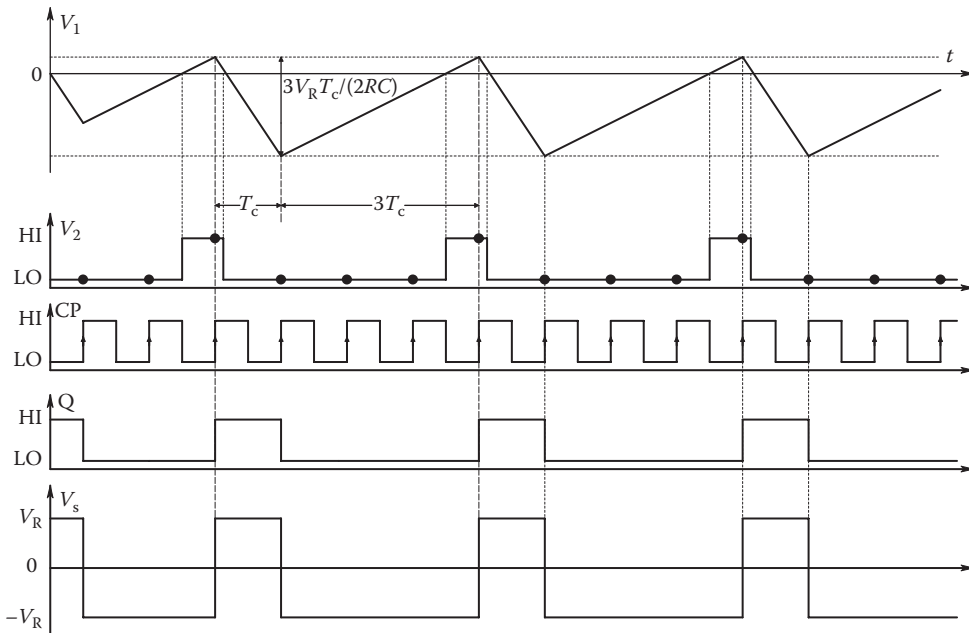


FIGURE 9.24
Waveforms in the Δ - Σ ADC of Figure 9.21 when $V_x = V_R/2$.

In addition, at least 2^N clock cycles must be counted in order to obtain N -bit effective resolution. In the Δ - Σ ADC, the modulator's DFF Q output is the input to the block labeled *digital filter and decimator*. The digital LPF precedes the decimator. The digital LPF serves two functions: (1) It acts as an AAF for the final sampling rate, f_o , and (2) it filters out the higher-frequency noise produced by the Δ - Σ modulator.

Quoting Analog Devices' AN-283 (*sigma-delta ADCs and DACs*):

The final data rate reduction is performed by digitally resampling the filtered output using a process called decimation. The decimation of a discrete-time signal is shown in Fig. 6.12, where the sampling rate of the input signal $x(n)$ is at a rate which is to be reduced by a factor of 4. The signal is resampled at the lower rate (the decimation rate), $s(n) [f_o]$. Decimation can also be viewed as the method by which the redundant signal information introduced by the oversampling process is removed.

In sigma-delta ADCs it is quite common to combine the decimation function with the digital filtering function. This results in an increase in computational efficiency if done correctly.

Figure 9.25 illustrates a heuristic, frequency-domain block diagram approximation for a first-order Δ - Σ ADC. To appreciate what happens in the system in the frequency domain, we can find the transfer function for the analog input, V_x , and the internal quantization noise, n_q (see the following section). Assuming a linear system, we find

$$V'_s = \frac{V_x K_i}{j\omega + K_i} + \frac{j\omega n_q}{j\omega + K_i}. \tag{9.36}$$

Thus, the 1-bit quantization noise in the output is boosted at high frequencies so $V'_s \cong n_q/K_i$. The signal component in V'_s rolls off at -6 dB/octave above $\omega = K_i$ r/s. At low frequencies, the noise in V'_s is negligible and $V'_s \cong V_x$. Because the Δ - Σ modulator is really a sampled system at frequency f_c , the noise in V_s and Q has the root power density spectrum shown in Figure 9.26. Note that the broadband quantization noise is concentrated at the upper end of the root spectrum. The range from $0 \leq f \leq f_b$ is relatively noise free. Operation in this range is achieved by first having the FIR LPF operate on the data stream from Q at clock rate f_c , then decimation so $f_b = f_o/2$, the Nyquist frequency of the decimation frequency. Lastly, the counter counts the decimated data for $2^N f_o$ clock periods.

A state-of-the-art 16-bit, sigma-delta ADC, the AD7709, is offered by Analog Devices (2003). "The AD7709 is a complete analog front-end for low frequency measurement

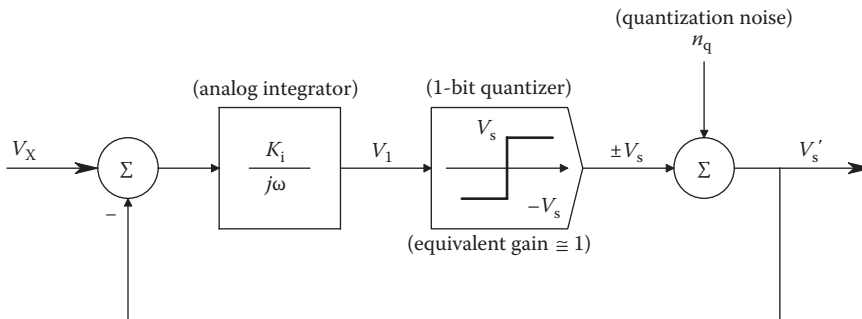


FIGURE 9.25 A frequency domain block diagram approximation describing a first-order Δ - Σ ADC.

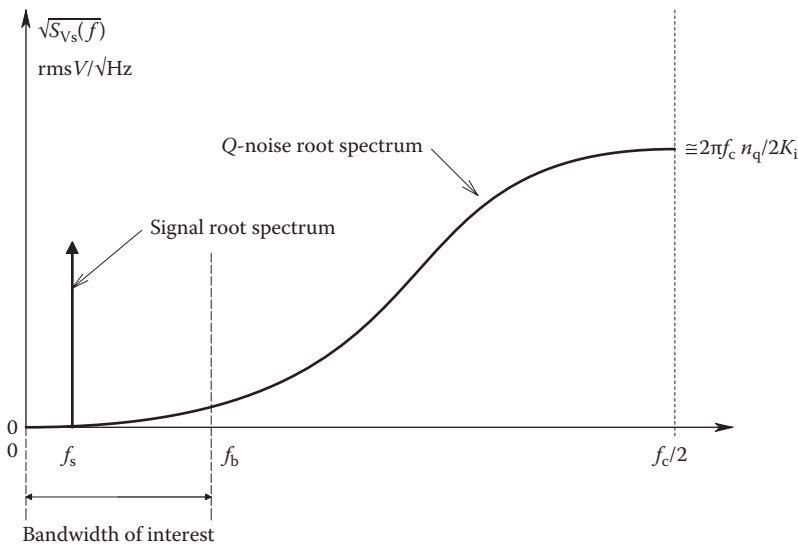


FIGURE 9.26 Root power density spectra of signal and noise in a first-order Δ - Σ ADC.

applications. The AD7709 contains a 16-bit sigma delta ADC with PGA and can be configured as 2 fully-differential input channels or 4 pseudo-differential input channels. Inputs signal ranges from 20 mV to 2.56 V can be directly converted using the AD7709."

9.7.7 Data Acquisition Cards for PCs

It is evident that PCs and laptop computers are central to most modern I&M systems. Consequently, many manufacturers make plug-in interface boards that allow one or more channels of external analog data to be sampled and digitized. In fact, the ubiquity of the modern PC allows it to serve as a DSO; a single specialized card is required. PC Instruments of Akron OH made a series of PCI-420 and PCI-430 digital oscilloscope cards. These cards, depending on the model, have one or two 8-bit channels and bandwidths ranging from 100 to 300 MHz. In the real-time mode, maximum resolution is 80 ns/samples. In the equivalent sampling mode, maximum resolution is 5 ps/samples, or an equivalent sampling rate of 200 gigasamples/s.

GaGe Applied Sciences of Montreal offers a line of superfast, PCI Express Bus™ waveform digitizers and virtual oscilloscopes. Their Model 85G CompuScope™ is their fastest, sampling 8 bits at 5 gigasamples/s. The GaGe Model 2125/ETS PCI Express CompuScope™ does real-time sampling at 250 megasamples/s on one channel and has 2 gigasamples/s equivalent time sampling on two, 8-bit channels. These plug-in digital oscilloscope cards operate as the input elements for sophisticated virtual instruments (VIs); all the normal knobs and switches found on the front of a stand-alone DSO are displayed on the CRT and are mouse operated. The GaGe model CobraMax© 8-bit digitizer board has a maximum sample rate of 4.0 gigasamples/s, 1 or 2 channel input, and a 1.5 GHz bandwidth.

Many manufacturers make A/D boards for PCs that are not part of a virtual DSO system. For example, Delphi Engineering Group in Costa Mesa, CA, makes an 8-bit, 1 GHz PMC ADC digitizer board. This board has an analog -3 dB bandwidth of 1 GHz. Its sampling rate is up to 1 gigasamples/s. SuperLogics, Inc. offers an extensive line of PCI data q boards

having 12-, 14-, and 16-bit resolution on two channels, also 32-bit digital I/O. Sampling rates range from 150 kilosamples/s to 1.25 megasamples/s, depending on the board. Many other companies, including CyberResearch® Branford, CT, National Instruments Corp™ (NI), Keithley Instruments (a Tektronix Company), offer extensive lines of analog I/O boards. One has the option of specifying the number of input channels, single-ended or differential input, the number of bits, the size of the on-board buffer memory, and, of course, the sampling rate. The user designing an I&M system has the option of writing their own data-management software, but most board vendors also sell software packages for their boards. Most notable is the versatile LabVIEW® software sold by NI.

9.8 IEEE-488 Instrumentation Bus (GPIB)

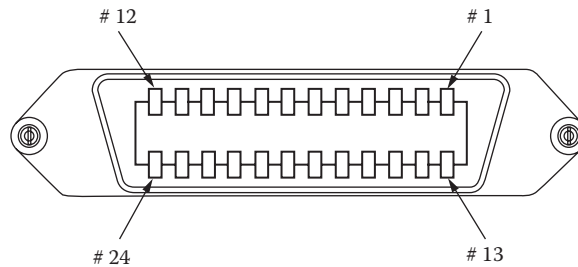
In computer nomenclature, a *bus* is “a collection of unbroken signal lines that interconnect computer modules (the connections are made by taps on the lines)” (Stone 1982). In general, a bus contains data lines, control and signaling lines, and, in some cases, address lines. There are many *parallel bus architectures* and protocols that have been developed since the advent of the microcomputer. These include, but are not limited to, the Intel *Multibus* (IEEE-796 bus), the *VMEbus*, the *Centronics printer bus*, the *DEC Unibus*, the *LSI-11 bus*, the *IBM PC bus* (IBM PC/XT/AT), the *Macintosh II NuBus*, the *MicroVAX Q-bus*, the *IBM PS2 Micro Channel Architecture (MCA) bus*, the *Intel iSBX bus*, the *ISA and EISA PC buses*, and, of course, the *IEEE-488 Instrumentation Bus (GPIB)*, which we describe in the following texts.

The IEEE-488 bus was developed by HP in the early 1970s as a standard, 8-bit, bidirectional, asynchronous bus (HP-IB) to enable a number of HP-IB-compatible instruments to communicate with a controlling computer and with each other. Not only can measured data be sent to the host computer for storage and processing, but in certain GPIB-compatible instruments, the computer can be used to set the instrument’s front panel controls and to control the measurements. The original IEEE-488 standard was defined by the IEEE Standards Committee in 1975, and it was revised in 1978 and again in 1987 as IEEE Std. 488.2-1987. The IEEE 488.2 standard is also called the *Standard Commands for Programmable Instruments* (SCPI). The SCPI Consortium voted to become part of the IVI™ Foundation in late 2002, and the IVI Board of Directors voted to accept the merger in the spring of 2003. IVI assumed control of the SCPI specifications. (IVI stands for *Interchangeable Virtual Instruments*.) (See www.ivifoundation.org/ [accessed April 23, 2013].) The latest IVI specifications may be found at www.ivifoundation.org/specifications/default.aspx/ (accessed April 23, 2013).

The GPIB is unique in that it uses unique, stackable (hermaphrodite) connectors on its cables, as shown in Figure 9.27. Many semiconductor manufacturers offer LSI IC interfaces for the GPIB. Some of these include the three Intel i829X series ICs (the i8291 talker/listener, the i8292 GPIB controller, and the i8293 bus transceiver–driver), the Texas Instruments TMS9914, the Motorola 68488, and the Signetics HEF4738.

9.8.1 Summary of the GPIB Bus Structure

The GPIB consists of eight, tristate, bidirectional data lines and eight control lines that select, deselect, and otherwise coordinate the asynchronous communications between the host computer and satellite instruments. The maximum practical data transmission rate in



Pin signal	Pin signal
1. Data 1	13. Data 5
2. Data 2	14. Data 6
3. Data 3	15. Data 7
4. Data 4	16. Data 8
5. EOI	17. REN
6. DAV	18. Gnd
7. NRFO	19. Gnd
8. NDAC	20. Gnd
9. IFC	21. Gnd
10. SRQ	22. Gnd
11. ATN	23. Gnd
12. Shield	24. Logic ground

FIGURE 9.27

Standard IEEE-488 instrumentation bus stackable connector and its pin assignments. (From Helfrick, A.D. and Cooper, W.D., *Modern Electronic Instrumentation and Measurement Techniques*, Prentice Hall, Englewood Cliffs, NJ, 1990.)

most GPIB systems is about 250 kB/s, equivalent to about 2 Mb/s. Practical considerations on loading a typical IEEE-488 bus limit the number of devices on the bus to about 15, all which should be located within 3 m or so of the host computer. All lines on the GPIB use a complimentary TTL logic protocol. An active LOW is a voltage less than 0.8 V and may be treated as a logical TRUE or 1. An active HI is a voltage greater than 2.5 V and is a logic FALSE or 0. In the tristate, Hi-Z condition, pull-up resistors tied to +5 V make the data lines assume a high-impedance, HI voltage. All commands and most data on the 8-bit data I/O lines are generally sent using the 7-bit ASCII code set, in which case the eighth bit is used for parity or is unused.

The eight control lines and their acronyms and functions are described in the following text; they may be subdivided into *handshake lines* and *interface management lines*.

Handshake lines

1. **DAV** (data valid): When a selected device on the GPIB supplies an 8-bit word to the data lines (i.e., is a talker), DAV is set LOW (or TRUE) to indicate to all on the bus that the data byte on the bus is ready to be read to a listener device. DAV is one of the three *handshaking* lines that control data transmission on the GPIB.
2. **NRFD** (not ready for data): This second handshaking line is pulled LOW by a selected listener to indicate it is ready to accept data (NRFD would be better called RFD).
3. **NDAC** (not data accepted): This third handshaking line is set LOW by the selected listener device when the data byte transmitted has been accepted and that new data may now be supplied.

Interface management lines

4. *ATN* (attention): This line is pulled LOW by the bus controller (computer) to signal that it is sending a command. It drives ATN HI to signal that a talker can send it data messages.
5. *IFC* (interface clear): The GPIB controller drives this line LOW to reset the status of all other devices on the GPIB and to become controller in charge (CIC).
6. *SRQ* (service request): This is the GPIB equivalent of an interrupt line.
Any device on the bus (except of course the controller) can pull the SRQ line LOW, signaling to the controller that it requires *service*. Devices can be programmed to signal SRQ for various reasons, such as being overloaded and out of paper and having completed an internal task such as computing an FFT spectrum. The controller, upon sensing SRQ LOW, must poll the devices on the GPIB to determine which one sent the SRQ and then take appropriate control action (service).
7. *REN* (remote enable): This controller output line is set LOW to allow the controller to take over front-panel controls of a selected instrument on the GPIB (if that instrument has the capability of having its front-panel settings taken over by the controller).
8. *EOI* (end or identify): This line has two functions—a talker uses EOI LOW to mark the end of a multibyte data message. EOI LOW is also used by the bus controller to initiate a parallel poll. When the controller sets both EOI and ATN LOW, a parallel poll causing devices configured for a parallel poll to present status bits on the data bus, which are read by the controller.

9.8.2 GPIB Operation

As we have already seen, the 8-bit data bus can carry data or commands. The controller pulls the ATN line LOW, to signal it is sending commands. There are *four types of command: addressed, listen, talk, and universal*. Commands are generally represented by 7-bit ASCII characters on the bus. The type of command being sent is signaled by bits 5, 6, and 7 on the data bus. If bits 5, 6, and 7 are HI voltages (logic 0), then the command is an addressed command. If bit 5 is LOW and 6 and 7 are HI, then the command is universal. If bit 6 is LOW and 7 is HI, then the command is a listen command, and if bit 6 is HI and 7 is LOW, then the command is a talk command. If both bits 6 and 7 are LOW, the command byte is a *secondary command*. Secondary commands are used for sending secondary addresses or setting up devices for a parallel poll.

As suggested by their name, *universal commands* affect all devices on the GPIB. *There are five universal commands*: LLO (local lockout) is represented by 11 h on the bus; it is used to disable the front-panel controls of all devices on the bus. Under this command, the controller will have complete control of instrument settings. DCL (device clear) resets all devices on the GPIB. It is 14 h on the data bus. A PPU (parallel poll unconfigure) command (15 h on the data bus) resets the parallel poll responses of all devices on the GPIB, allowing new parallel poll responses to be specified. SPE (serial poll enable) responds to an SRQ by sending out an SPE (18 h) on the data bus. The SPE causes each device on the GPIB to prepare a single status word to be sent back to the controller when it is sequentially interrogated by the controller. When the controller finds the device that sent the SRQ, it sends out an SPD command. SPD (serial poll disable) is sent out by the controller (19 h) on the data bus following the SRQ identification. The SPD command resets all the devices' buses to the normal mode following the SPE so that the SRQ problem can be serviced.

Addressed commands are sent to those certain devices that have been sent listen commands. *There are five addressed commands* in the GPIB protocol; they, too, are represented by hex words on the data bus:

1. GTL (go to local) (01 h) cancels the universal command and LLO and restores local, front-panel control to addressed devices.
2. SDC (selected device clear) (04 h) clears those devices that have received listen commands.
3. PPC (parallel port configure) (05 h) is sent to a specific device to tell it which data bit it will use to give its status. A secondary word is sent following the addressed PPC command to specify which data line will have the status signal on it and whether a 1 or a 0 will be used to signal that the device needs service.
4. GET (group trigger) (08 h) synchronizes the operation of several instruments on the GPIB, such as starting a group of related measurements.
5. TCT (take control) (09 h) enables the main controller to pass control of the GPIB to a secondary controller. The device that is to become the new controller is first sent a listen command, followed by TCT.

The interested reader can explore the GPIB in greater depth in the texts by Stone (1982), Tompkins and Webster (1988), and Helfrick and Cooper (1990).

Many manufacturers of interface cards for PCs and MACs offer IEEE-488.2-compatible cards and controlling software. Some of the newer GPIB systems have hardware and software that permit one to make a very fast and flexible measurement control system. For example, the NI's NB-GPIB interface board for the Macintosh II has the latest IEEE-488.2 functions and can read and write data at a sustained 800 kB/s. The NI's LabVIEW 2[®] software for the Mac II is a powerful, graphical programming language that permits the creation of VIs on the monitor whose settings can be altered by the mouse and sent to the actual instrument by GPIB. The NB-GPIB card is compatible with NuBus DMA operations and the RTSI bus. The NI's GPIB-SE card for the Mac SE is another advanced GPIB card that claims data bus transfer rates up to 1 MB/s. This card, too, is compatible with the LabVIEW[®] software. Metrabyte also offers IBM PC-compatible IEEE-488 cards and handling software, the MBC-488 card, and the IE-488 card. These cards support up to 15 devices on the bus and do DMA data transfers at 450 kB/s.

9.9 Serial Data Communications Links

While the GPIB is eminently suited for managing a large measurement system, there is often a need to couple the output of a single instrument to a computer interface over long distances using few electrical conductors. Serial data transmission has well-established protocols and historically predates the GPIB by at least 18 years. The RS-232 serial interface was originally developed in the early 1960s to send data to a CRT terminal or teletypewriter over telephone wires. Its low maximum bit transfer rates now makes the RS-232 serial interface generally obsolete for the control of the modern equivalents of terminals, printers, and modern instruments. The RS-232 interface currently finds limited use in coupling certain mice, trackballs, joysticks, and other slow input devices to computers, and computers to certain plotters and nongraphics printers. Very few measurement instruments are designed to use the RS-232 interface; it lacks the speed and flexibility of the

IEEE-488.2 bus or the USB. However, for specific applications, such as reading in the data from remote DC sensors and sending data to loggers, it can be useful.

Other, more modern, serial, asynchronous data transmission protocols include the RS-422, RS-423, RS-449, RS-485, and the USB. We will first describe the RS-232 serial interface, as it is the oldest serial bus, it has been widely used, and its protocol is used with little variation in the later serial bus designs. (Note that RS stands for *recommended standard* [of the Electronic Industries Association, Washington, DC].)

9.9.1 RS-232C and D Interfaces

In describing the venerable RS-232C interface, we encounter some new acronyms. RS-232C serial communications are sent by *data terminal equipment* (DTE) and *data communications equipment* (DCE). DTE include such things as computer terminals, teleprinters, computers, and digital instruments. DCE are devices (modems) that encode the serial digital signals into low-bandwidth (sinusoidal) formats compatible with transmission on voice band telephone lines. The rate that modems transmit data is given in baud, which stands for *bit rate audio*. (Modern modems used for serial telephone line communications between PCs and instruments typically run between 2,400 baud and 56,000 baud, depending on the limiting bandwidth of the transmission system. FAX modems typically run at 24,400 baud.) Just as the GPIB has a unique, standard connector, the RS-232C interface uses a standard male or female 25-pin, DB25P (male) or DB25S (female) connector, shown in Figure 9.28.

The RS-232C interface operates either in the simplex, half-duplex, or full-duplex modes. In the simplex mode, data transmission is unidirectional, for example, from the computer to a printer. In the half-duplex mode, serial data can be sent in both directions but in only one direction at a time. Full-duplex operation permits simultaneous, bidirectional, serial data transmission.

Data are transmitted as 8-bit, ASCII words signaled by high or low logic voltages on the transmitted data (TD) line. An example of this process is shown in Figure 9.29. In the idle state, the TD line is held high. At the beginning of data word transmission, the TD line goes low and stays low for one clock period. This is the start bit, which is always low. The receiving equipment senses the high–low transition of the start bit and, to verify start, samples the received data (RD) one-half clock period later. If low, the start bit is verified. The state of the RD line is then sampled eight times at intervals of one clock period. The last (eighth) sample is the MSB and is called the parity bit. In setting up an RS-232D interface, the user can specify the use of odd, even, or no parity in the data transmission process. If even parity is used, the parity (8th) bit is set so that the total number of logical 1s in the transmitted word, including the parity bit, is even. If the parity check algorithm in the receiver counts an odd number of highs in a received word, it declares a parity error. The parity error can terminate transmission or require that the same character be sent again. Similarly, if an even number of 1s is counted in a received word when using odd parity, an error is declared. Parity is generally used when serial data are transmitted under noisy conditions over long distances. It is normally not used in situations where the receiver is connected by a short cable to the transmitter in a low-noise environment.

At the end of data transmission, after the parity bit is sent, one or two high stop bits are sent before the TD line is declared idle and is ready to transmit the next word (ASCII character). Note that there are a number of handshaking functions on the RS-232C lines that accompany the serial data transmission and receiving lines. These include *ring indicator*, *data terminal ready*, *carrier detect*, *signal ground*, *dataset ready*, *clear to send*, and *request to send*. For more detailed descriptions of the operation of the venerable RS-232C interface, the reader can consult Section 5.2 in Stone (1982), Chapter 18 in Wolf (1983), Chapter 6 in

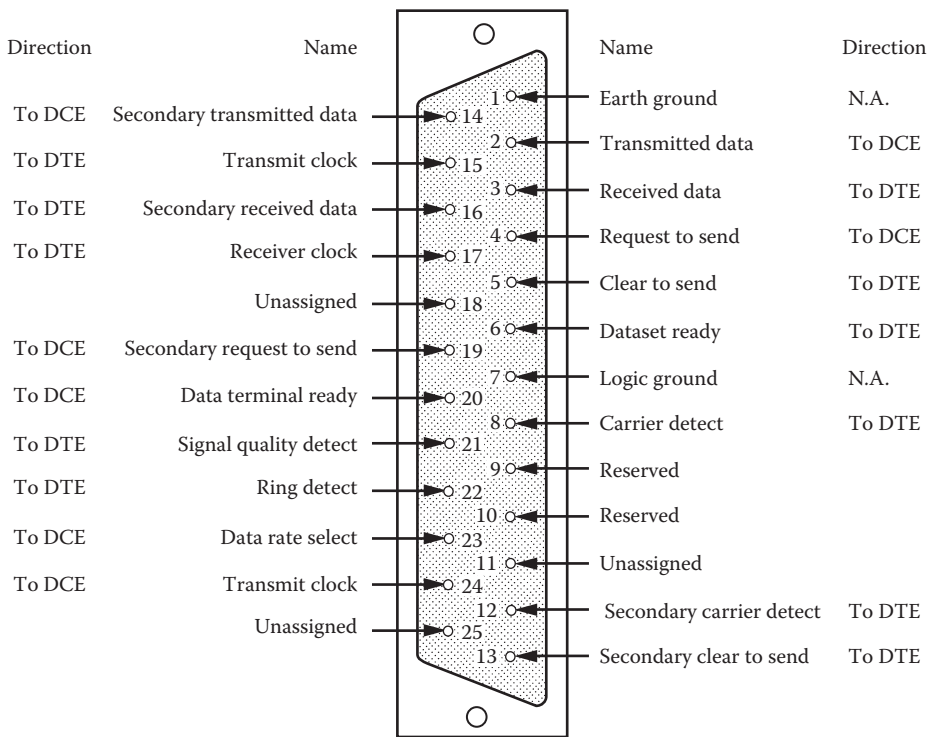


FIGURE 9.28 Standard pin connections for the obsolete RS-232C serial data connector. Pin nomenclature is based on the connections in DTE equipment. (From Wolf, S., *Guide to Electronic Measurements and Laboratory Practice*, 2nd edn., Prentice-Hall, NJ, Chapter 16: Interference signals and their elimination or reduction, 1983.)

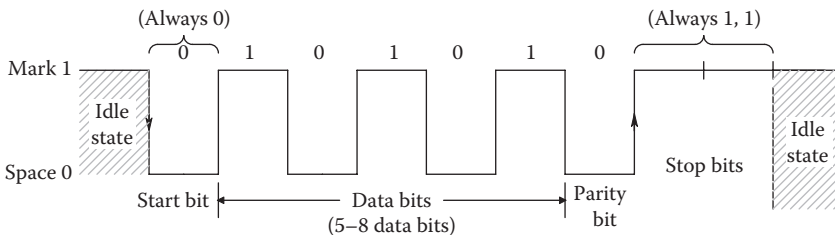


FIGURE 9.29 Example of an 8-bit, serial data signal sent on pin 2 of an RS-232C DTE.

Tompkins and Webster (1988), and Motorola’s App. Note AN-781A. This App. Note also describes the RS-422, RS-423, RS-449, and RS-485 standards (Morgan 1978).

In the earliest days of RS-232 use, it was common for computers to be connected to electromechanical teletypewriters such as the Western Electric ASR33. The ASR33 used no IC electronic components; all logical and switching operations were done electromechanically at very low baud rates. Modern serial data communications use UART IC chip sets that handle the parallel to serial conversion of data for transmission and, at the receiver, convert the RD to parallel form.

The latest RS-232 embodiment uses the *D* designator. It is essentially the same as the RS-232C in its specifications and protocols; however, it uses an 8-pin, telephone-type

connector, the RJ-45 (similar to those used in 10Base-T Ethernet lines). The RJ-45 pin signal descriptions and their acronyms are as follows:

1. DCE ready/ring indicator (DSR/RI)
2. Received line signal detector (DCD)
3. DTE ready (DTR)
4. Signal ground (SG)
5. Received data (RxD)
6. Transmitted data (TxD)
7. Clear to send (CTS)
8. Request to send (RTS)

Like the RS-232C, the RS-232D is used for low-speed, asynchronous communication between a computer and a single, slow, peripheral device. It is seldom found in new instrument designs.

The present state of the art for the RS-232 C&D protocol limits data transmission rates to less than 20 kbaud, and wire cables are seldom effective over 15 m (50 ft.) in length. Also, RS-232 protocol restricts the dV/dt of the TxD line data to a maximum of 30 V/ μ s. To circumvent these problems, other serial data transmission standards have been developed, making use of improvements in hardware to achieve higher baud rates and longer transmission lines. These interfaces are described in the next sections.

9.9.2 TRS-422, RS-423, and RS-485 Interfaces

These interfaces make use of our knowledge about the transmission properties of transmission lines for transient signals and improved IC designs to realize improved data transmission rates over longer distances than possible with the now-obsolete, 50+-year-old technology of the RS-232 interface. Their properties are summarized in the following paragraphs and are illustrated in Figure 9.30.

The *RS-422A interface* uses a balanced, twisted-pair transmission line, terminated in the characteristic impedance of the line. A balanced or differential amplifier line driver as well as a differential line receiver is used. The RS-422A interface can transmit data at up to 10 Mbaud and can have lengths of up to 1200 m. There is a trade-off between baud rate and cable length: 10 Mbaud is possible on a 12 m cable, 1 Mbaud on a 100 m cable, 100 kbaud on a 1 km cable, and up to 80 kbaud on a 1.2 km cable. The TI 9636 and 9637 ICs, as well as the Motorola MC3487 differential receiver and the MC3486 differential driver, can be used to realize an RS-422A link.

The mechanical connections for the RS-422A interface are specified by the RS-449 standard, which specifies a 37-pin connector supporting the mandatory twisted pair lines for receive ready, test mode, data mode, request to send, clear to send, receive data, and send data (Stone 1982).

The *RS-423A interface* uses an unbalanced single line, similar to the RS-232D link. Even so, the RS-423A link is faster than the RS-232C; it was designed, according to Stone (1982), to provide a linkage between the old RS-232C interface and the RS-422A interface. Its operational protocol is very similar to RS-232, although improvements in the driver and receiver electronics allow it to operate at rates up to 100 kbaud over short cables (<30 m) and at significantly slower rates over cables up to 1.2 km.

The *RS-485 interface* is a balanced (twisted pair) *party line* on which a number of secondary receivers and transmitters can operate. In this respect, it is effectively a data bus.

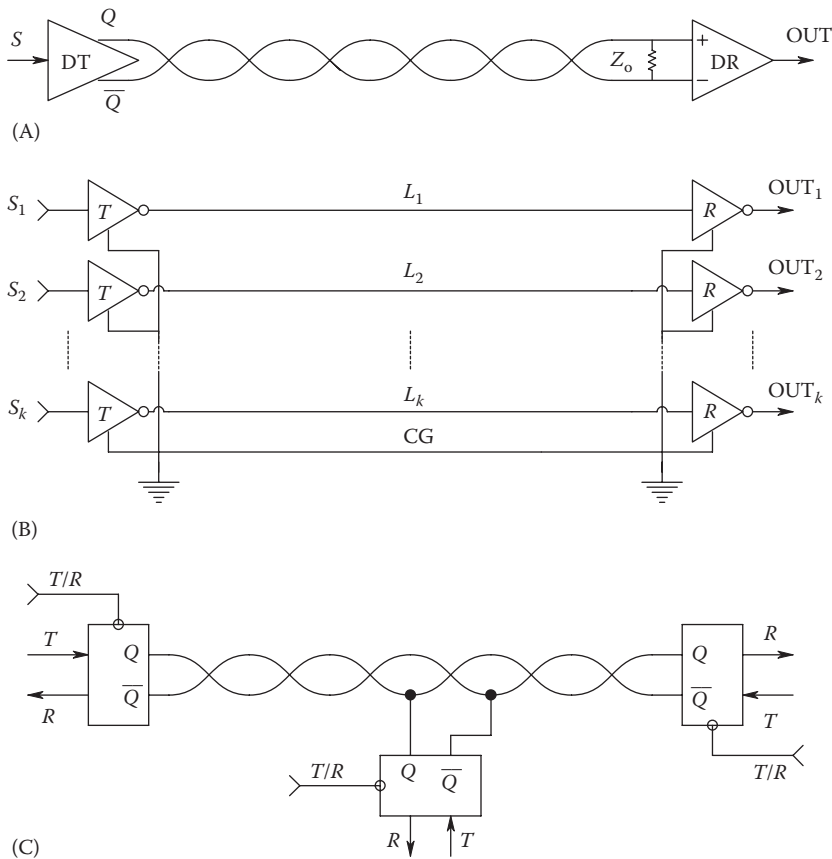


FIGURE 9.30 (A) An RS-422A, balanced, twisted-pair transmission line. *Note:* DT = differential transmitter. DR = differential receiver. Z_o = characteristic impedance of the transmission line. (B) RS-423A, unbalanced serial data transmission lines. *Note:* CG = common ground. (C) RS-485, balanced, bidirectional, twisted pair serial data transmission system. Differential, tristate transceivers are used. *Note:* T/R = transmit/receive signal.

Although the RS-485 interface can transmit data up to 10 Mbaud, practical considerations of line resistance, loading, and terminations set practical limits on line lengths. SN75172 and SN75174 differential drivers and SN74173 and SN75175 differential receivers are used to implement the RS-485 protocol.

Table 9.4 summarized some important specifications for the RS-series serial data lines.

9.9.3 Universal Serial Bus

9.9.3.1 Introduction

The modern USB has rapidly replaced the RS-232 C&D and other serial interfaces as a means for a computer to communicate with medium-speed peripheral devices such as scanners, mice, printers, PDAs, and digital cameras. Practically all PCs and laptops made in the past 18 years have two or more USB ports. Reasons for USB's popularity include its simple connection, plug and play feature, hot insertion capability, speed, and versatility.

USB v.1.0 was released in January 1996. It permitted low-speed (1.5 Mb/s) and medium-speed (12 Mb/s) data transfer rates. (*Note:* Mb/s refers to *megabits* per second,

TABLE 9.4

Specifications of RS-Serial Data Lines

Specifications	RS-232 C&D	RS-423	RS-422	RS-485
Mode of operation	Single ended	Single ended	Differential (twisted pair)	Differential (twisted pair)
Allowed number of Tx and Rx	1 Tx, 1 Rx	1 Tx, 10 Rx	1 Tx, 10 Rx	32 Tx, 32 Rx
Maximum data rate, bps	20k	100k/10M	100k/10M	100k/10M
Normal driver output voltage	± 5 to ± 15 V	± 3.6 V	± 2 V	± 1.5 V
Maximum driver output	± 25 V	± 6 V	± 6 V	± 6 V
Tx load impedance, Ω	3–7k	≥ 450	100	54
Rx input sensitivity	± 3 V	± 200 mV	± 200 mV	± 200 mV
Rx input voltage range	± 15 V	± 12 V	± 7 V	-7 to +12 V
Maximum Rx input resistance, Ω	3–7k	<4k	<4k	≥ 12 k
Maximum cable length, ft	50	4000	4000	4000

not megabytes [MB].) The actual USB data rate is slower because of USB protocol (operating) data sent to and from a peripheral device. USB v.1.1 was released in September 1998. This edition fixed many of the problems found in release v.1.0. USB v.2.0 (high speed) was released in the early 2000 and has a maximum bit clock rate of 480 Mb/s. It is downward compatible with v.1.1 and v.1.0. USB 3.0 (super speed) was released in November 2008. It has a usable data rate of 4 Gb/s and is backward compatible with USB 2.0. A January 2013 USB group press release revealed plans to update v.3.0 to run at 10 Gb/s by mid-2013.

USB data are clock encoded and use NRZI with *bit stuffing*. A *transaction* equals three packets, with flexible packet size of up to 1023 bytes. A *start-of-frame* marker occurs every 1 ms, and it has 125 ms microframes in SOF mode. A USB system sets up a unique *pipe* (aka data pipeline) between a host and a device, which can be assigned bandwidths. It also has the flexibility to implement split USB transactions for a 12 Mb/s operation with a 480 Mb/s (v.2.0) hub.

USB 1.0 cabling is electrically simple; it uses four conductors: pin 4 = ground, pin 1 = +5 V, and a differential pair for bidirectional data transfer (pin 2 = data -, pin 3 = data +). There are two connectors on a USB cable, A and B. The A connector mates with the host computer's USB port. It consists of a simple, rectangular metal sheath, ca. 12 × 4 mm, covering a plastic insert holding four gold-plated wiper strips. The plastic insert serves to polarize the host connector so it cannot be inserted inverted. The B connector on the peripheral end of the USB cable can be smaller than the host connector and is unique to that peripheral (digital camera, scanner, external CD-ROM reader/burner, etc.). Cable length is limited to a maximum of ca. 5 m (Simpson 2003). USB 2.0 has five conductors: pin 1 = +5 V, pin 2 = data -, pin 3 = data +, pin 4 = ID, and pin 5 = ground.

USB 3.0 connector design allows the type A plugs to fit host computer receptacles for USB 3.0 and USB 2.0 only. The USB 3.0 type B receptacles are larger than required for a USB 1.0 or USB 2.0 type B plugs. The larger dimension for a USB 3.0 type B receptacle permits connection of either the larger USB 3.0 type B plug or the smaller USB 2.0 or 1.0 type B plug to a newer USB 3.0 type B receptacle. Thus, a USB 3.0 type B receptacle on a peripheral device can be connected using the corresponding plug end of a USB 2.0 type B cable. A USB 3.0 type B plug cannot mate with a USB 2.0 or 1.0 receptacle.

Power requirements on the USB are such that a peripheral device that requires less than 100 mA can be powered from the host's hub. Devices such as external CD-RW drives that draw more than 100 mA from the host are generally self-powered (batteries or isolated plug-in DC power supply). When a peripheral USB device is first connected,

it identifies itself to the host and tells the host how much current it draws. A battery-powered device, such as a digital camera, can draw power from the host if less than 100 mA is required, then switch back to its internal batteries when it is disconnected from the USB or the host is shut down. When a device is made dormant (suspend mode), it cannot draw more than 500 μ A from the bus if it is bus powered.

By using *external expanders*, it is theoretically possible to connect a total of 127 peripheral devices to one host computer's USB port. If two host USB ports are used fully, you can have an amazing 254 peripherals attached to the computer. In practice, such a great number of USB peripherals would probably run out of USB port bandwidth and require external power. At present, two computers cannot be connected directly, USB to USB. Instead, one needs to use an external *USB bridge*, aka a USB to USB adapter, between them. Since modern laptops no longer come with RS-232 ports, it is possible to transform a USB port to an RS-232 port with an external converter, such as the CyberResearch® USBS-221 device; CyberResearch also makes a series of USB adapters for RS-422/485 interfaces. CyberResearch also offers a type II, PCMCIA laptop card that has four USB 2.0/1.1 ports.

9.9.3.2 How USB Works: An Overview

The digital communication and control of USB peripheral devices is incredibly complicated. What simplicity USB has electrically is more than offset by its digital protocols. All peripheral USB devices have addressable *endpoints* that are memory buffers. An endpoint can be as simple as one register, or a device can have multiple endpoints with which to store incoming and outgoing data (endpoints 0, 1, 2, ..., N). All endpoints must be able to both send and receive data but obviously can only communicate in one direction at a time.

A logical connection link must be set up between the host and a peripheral device before a data transaction can take place. This connection is called a *pipe*. The pipe is set up as soon as the host recognizes the device connected through the USB. When the host responds to a connect signal from the device, the device sends its requirement for data transfer type and speed. The host can refuse to establish a pipe if it lacks the power needed by the device or cannot meet the device's bandwidth requirement. A *smart* device meeting power limitations can lower its requested data rate and again try to negotiate a pipe. Pipes can be explicitly managed; bandwidth and capacity can be assigned to each pipe by the host software. This flexibility guarantees adequate USB bandwidth for operations such as bulk and isochronous data transfer and provides an easy mapping from the host's application software to a specific device function interface.

Pipes connect through the second and third layers of the USB protocol. (The USB protocol layers are (1) *function layer*, (2) *USB device layer*, and (3) *USB interface layer*.) Each device has a control endpoint (*endpoint 0*) linked through a pipe at the device level (second layer). All other endpoint connections are at the third layer. Furthermore, there are two types of pipes: stream and message. Stream pipes are unidirectional and unstructured to pass data streams; message pipes are two-way structured.

The USB has four types of data transfer modes: (1) *control*, (2) *interrupt*, (3) *bulk*, and (4) *isochronous*.

Control mode is initiated by the host and is used to initialize the peripheral device. Bidirectional, *handshaking* traffic occurs in one direction at a time. In *interrupt mode*, the host queries devices to see if they need servicing. *Bulk mode* is used when data transfer accuracy is essential, as when reading or writing to a peripheral CD-RW. *Isochronous mode* sacrifices data accuracy for data timing. Uses include USB loudspeakers and precision motion generation.

Quoting *Simpson 2003*,

The following describes a typical data flow for a device when it is initially plugged into a host's bus when the host is active. Remember here that the host has an internal USB hub, and additional hubs may be connected downstream from the host's hub.

1. The host recognizes that a device has been attached to one of its USB hubs. It realizes this by a simple resistive divider that is connected to the differential data pair of wires in the USB bus. The resistors are inside the USB hubs and devices.
2. The host sends a *Get_Port_Status* request to the hub to find out more about what has been plugged in. It could be another hub, a device connected directly to the host's hub, or a device that has been plugged into one of the downstream hubs.
3. After receiving a response from the hub, the host issues a *Set_Port_Feature* command in which the hub issues a *Reset* over the data pair but only to the newly-connected device on the USB bus.
4. The host then checks to see if the device has come out of the reset state by issuing a *Get_Port_Status* command to the hub. After reset, the device is in the *Default State*, and can only draw a maximum 100 mA. In *Default State*, the device can communicate with the host through *Endpoint 0*.
5. The hub now detects the device's speed by using the resistive dividers that are attached to the USB bus. The hub sends the speed of this device back to the host.
6. The host then sends a *Get_Descriptor* command to the hub in which the hub gets the packet size needed from this particular device and sends the result back to the host.
7. The host then sets a *Set_Address* command to the hub which sends this information to the device. The device in turn acknowledges the command back through the hub to the host and sets up this address internally.
8. To learn more about this device, the host sends a *Get_Descriptor* command to the address that the device has been given. The information that is returned to the host consists of various details of the device that the host needs to know for its operation. These queries by the host continue two more times to retrieve all the information needed.
9. Based on the information received from the device, the host determines the best [logical] device driver to use for communication with it.
10. The device driver in the host now takes over by requesting a *Set_Configuration* command. There may be several configurations for one device, and the device driver determines which to use based on the information received from the device in response to the [previous] *Get_Descriptor* command.
11. The device is now ready to use.

Clearly, plugging in and setting up a peripheral USB device is a user-friendly process. However, the setup protocol is complex, as shown earlier. This complexity is necessary to allow *hot* connection and automatic device ID and initialization. It also allows the host to keep track of the different USB devices that are on the bus and operate them correctly. Included in the host's Microsoft® USB software are a spectrum of device drivers, including *human interface device* (HID) drivers. HID drivers are available for such devices as keyboards, mice, game controls, remote controls, and joysticks. As use of the USB is rapidly expanding, makers of various new peripheral devices include driver software with their products.

Unplugging a USB device, such as a flash drive, requires the host computer *permission*, to ensure that data are not being transferred or checked. Most computers *scold you* if you unplug hot.

9.9.3.3 Summary

The USB architecture is evidently here to stay; more and more computer accessory devices are designed around it: scanners, printers, digital cameras, digital telephones, wireless Internet adapters, PDAs, outboard bulk storage devices (ZIP [flash] drives, GPS Dongles, TEAC® CD-RW drives, and Lexar Media™ JumpDrive™ memories), computer game devices, joysticks, flight simulator hardware, etc., all use the USB system. A major advantage of the USB is that devices can be *hot plugged* into the host computer, which will automatically recognize and initialize them for use. When using a flash drive memory, disconnection requires host computer permission to avoid data loss. Much useful detailed information on USB configuration and operation can be found in the 34-page Wikipedia web article http://en.wikipedia.org/wiki/Universal_Serial_Bus (modified May 12, 2013) (accessed May 13, 2013).

9.10 CAMAC (IEEE-583) Modular Instrumentation Standard and the VXI Modular Instrumentation Architecture

The IEEE-583, IEEE-595, and IEEE-596 CAMAC standards define a hardware and data transmission system that is used to house, support, and communicate with various compatible instrumentation modules. A typical CAMAC *crate*, illustrated in Figure 9.31, is a physical package that has 25 slots or powered stations for compatible, plug-in instruments. Generally, the two right-hand slots are used to house the CAMAC controller module. The controller module allows the individual modules to be coordinated and controlled by an external computer using the IEEE-488.2 bus or, in some cases, the obsolete RS-232C interface. Commands and data transfers to and from the modules and the controller are made over the CAMAC DATAWAY, which use 24-bit data words. Figure 9.32 illustrates the (internal) DATAWAY connections to three modules in a CAMAC system. The read and write lines are each 24 bits wide. Two lines are used for the timing strobes; there are three common control lines, three status lines, four address bus lines, five function lines, and one station number line per module, for a total of 66 lines/module!

Lack of space prevents going into the details of the CAMAC DATAWAY control and command operation protocols. The interested reader can find their descriptions in the Teledyne LeCroy Instruments Catalog (online at www.teledynelecroy.com). The major advantage of CAMAC-housed instrumentation is its ease of use through the GPIB and its speed in local data manipulation. Individual instrument modules are managed by the CAMAC controller over the DATAWAY, and only one GPIB bus is required. CAMAC systems represent top-of-the-line instrumentation and as such are expensive. They are used extensively in large industrial applications and for measurements in nuclear and plasma physics.

The VXIbus measurement systems architecture, adopted in 1987, is similar in many respects to the CAMAC system but improves on certain aspects of CAMAC system design.

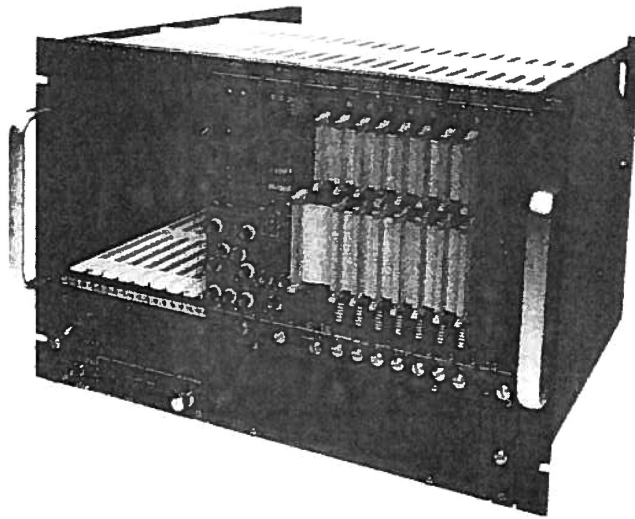


FIGURE 9.31
A partially filled CAMAC crate. (Courtesy of LeCroy Corp., Chestnut Ridge, NY.)

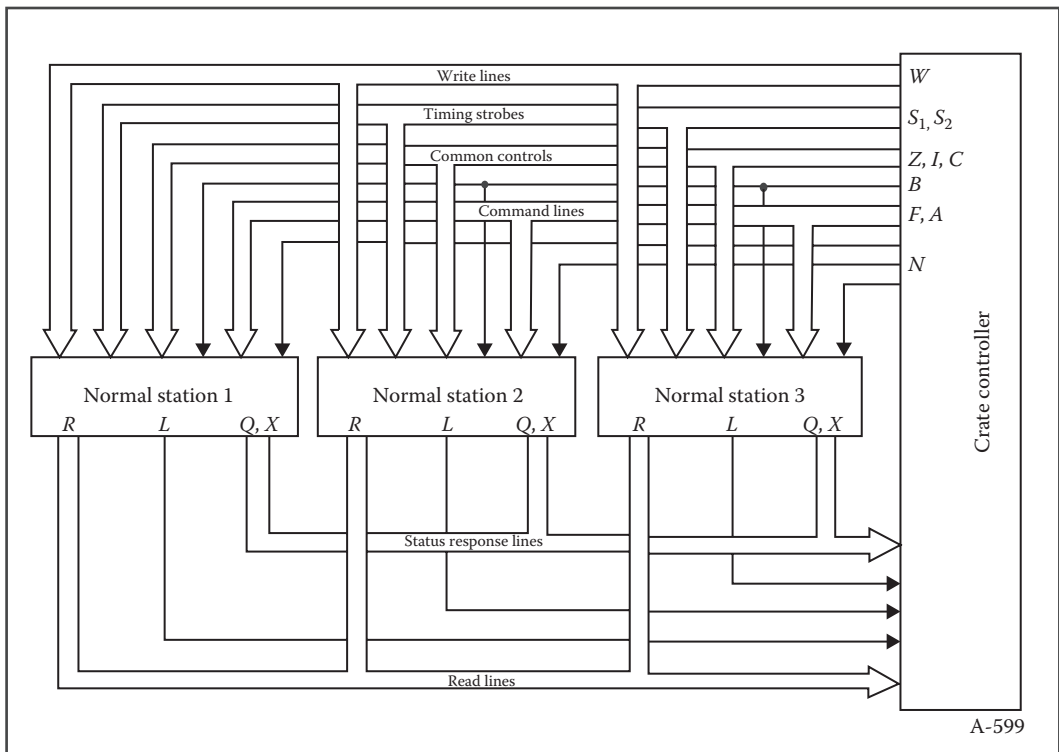


FIGURE 9.32
DATAWAY™ signal paths to three CAMAC modules from a crate controller. (Courtesy of LeCroy Corp., Chestnut Ridge, NY.)

VXI stands for *VMEbus extensions for instrumentation*. The VXIbus is a nonproprietary, open architecture bus for modular instrumentation systems that fully incorporates the VME bus standard. It includes not only mechanical and electrical interfacing but also communications protocols between compatible devices and an external controller. The VXIbus architecture was developed by a consortium of instrumentation companies, including Tektronix, HP, and NI. As in CAMAC systems, special VXI instrument cabinets are used to support a wide variety of modular instruments. All members of the HP 75,000 VXIbus instrument family use the SCPI, an industry standard now set by the IVI Foundation, for command, control, and communication. SCPI conforms to the IEEE-488.2 bus standard. In addition, all HP/Agilent 75,000 products support the HP Interactive Test Generator (HP ITG) software. There are many VXIbus instrumentation modules available from HP and other companies. All of these modules are characterized by having I/O connectors but no knobs, dials, pushbuttons, or switches. All settings and control operations are accomplished by software over the VXIbus. Some of the VXIbus module functions available from HP include command modules and controllers, digital multimeters, counters, function generators, amplifiers, D/A converter, digital I/O, digitizing oscilloscopes, power meter, various switches and multiplexers, a serial port adapter for RS-232C/422 protocols, and a VXIbus to MXIbus extender to tie VXI mainframes together and to an MXIbus computer interface (MXI stands for *multisystem extension interface*; it was developed by and announced as an open industry standard in April 1989 by NI).

Tektronix in 1990 gave a good summary of VXI communications protocols:

Two communications protocols are defined in VXI for Message Based Devices: Word Serial Protocol (WSP) and Shared Memory Protocol (SMP). The other type of communication possible in VXI, is low-level, binary data transfers which is defined for Register Based Device communication. Communication in a VXIbus system has a hierarchical commander/servant structure which is established by the system Resource Manager when the system powers up. A Commander is any device with one or more lower level devices (servants). The Commander initiates communication with its servants. A servant is a device which operates under the control of a Commander. All message based devices must respond to a WSP, but can also respond to the more advanced communication Shared Memory Protocol. With WSP, the required communication is 16 bit words transferred serially by reading or writing to one of the communication registers resident on a servant device. There is a WSP command set defined extensively in the VXIbus Specification.

SMP defines the operation of a shared memory channel established between two communicating devices with shared memory space. A channel can be asynchronous or synchronous. Asynchronous channels are used for high priority communications such as warnings, events and commands which must be immediately executed. The data transfer is accomplished over the ComputerBus in large data blocks. SMP is an efficient method of transferring large amounts of data in a VXI system whether the information is command strings or results obtained from performing a test.

Similar to CAMAC systems, VXIbus equipment is expensive and best suited for an industrial, ATE environment. For further details of the history and technical details of VXIbus instrument systems and the VME and VXS buses, the reader should consult the *Introduction to VME/VXI/VXS Standards* document by Wiener (2013) and also catalogs from NI, Tektronix, and HP and the IVI Foundation documents.

9.11 How Transmission Lines Affect the Transfer of Digital Data

Whenever we connect a computer to an instrument through a parallel or serial interface cable, we face problems associated with the transmission of digital data on a transmission line. Similar problems exist on printed circuit boards when we couple data and address lines between ICs, whenever the clock speed is greater than about 10 MHz. The main sources of difficulty come when the transmission line properties of the high-speed data paths are neglected in the design of interfaces. We can observe delays, attenuation, and rounding of pulses, as well as glitches caused by signal reflections at branches and terminations of the lines caused by improper terminating resistances and impedance mismatches at signal branch points on the lines. All of these problems serve to limit the rate of data transfer on transmission lines.

In this section, we will examine some of the major properties of transmission lines used in computer interfaces. Such lines include coaxial cables, shielded twinax cables, twisted pairs of wires, and ribbon cables. It will be seen that all wire transmission lines can be described by a common model.

9.11.1 Transmission Line Model

In our treatment of transient signal propagation on short transmission lines, we will first assume that the lines are lossless (i.e., $R\Delta x = G\Delta x = 0$ in Figure 9.33). It can be shown (Lathi 1965) that the characteristic impedance of the lossless transmission line is given by

$$Z_o = \sqrt{\frac{L}{C}} \Omega. \tag{9.37}$$

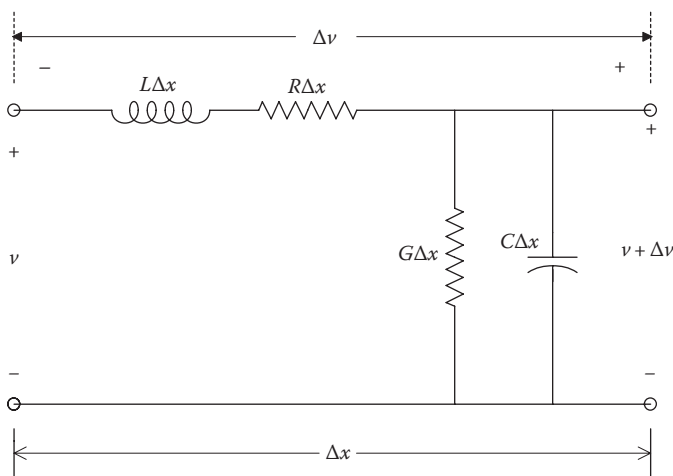


FIGURE 9.33 Lumped-parameter equivalent circuit of a differential length, Δx , of a transmission line.

The velocity of wave propagation on this line is

$$v = \frac{1}{\sqrt{LC}} \text{ m/s,} \tag{9.38}$$

where

- L is the inductance per unit length of the uniform line
- C is the capacitance per unit length between conductors

Most coaxial cables are made to have either 50 or 75 Ω characteristic impedances, and they come in a variety of physical sizes and voltage ratings. For example, a coaxial cable commonly encountered in instrumentation systems is the ubiquitous RG58A/U, which has $Z_o = 50 \Omega$, $C = 29.5 \text{ pF/ft}$, and $v = 0.649c$ (c is the speed of light in vacuo). In the case of ribbon cables having alternate conductors grounded, wires running parallel to a ground plane, and parallel-wire, twin-lead antenna wire, Z_o s generally range from 100 to 300 Ω or more. Figure 9.34 illustrates the end views of three common transmission line geometries. From electrostatics, the capacitances per unit length of the lines can be shown to be (Taub and Schilling 1977)

$$C = \frac{2\pi\epsilon}{\ln(R/r)} \text{ F/m (for coaxial cables)} \tag{9.39}$$

$$C = \frac{2\pi\epsilon_o}{\ln(2h/r)} \text{ F/m (for round wire over a ground plane)} \tag{9.40}$$

$$C = \frac{\pi\epsilon_o}{\ln(D/r)} \text{ F/m (for parallel wires in space),} \tag{9.41}$$

where R , r , h , and D are defined in Figure 9.34. Units are meters.

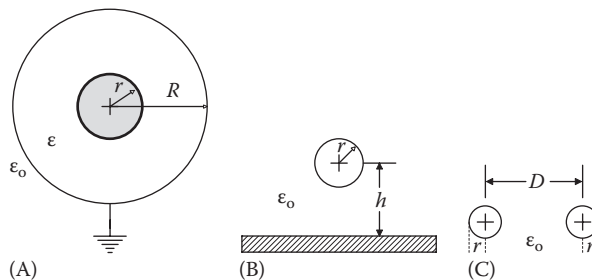


FIGURE 9.34

End views of three common transmission line geometries: (A) a coaxial cable, (B) a wire over a ground plane, and (C) a parallel balanced line (e.g., TV *twin lead*).

The inductance of these models can also be calculated (Kraus 1953, Motorola AN-270) (results are good approximations when h/r and $D/r \gg 1$):

$$L = \frac{\mu}{2\pi} \ln\left(\frac{R}{r}\right) \text{Hy/m (for coaxial cables)} \quad (9.42)$$

$$L = 4.61 \log_{10}\left(\frac{2h}{r}\right) \text{nH/cm (for a round wire in space over a ground plane)} \quad (9.43)$$

$$L = \frac{\mu_0}{\pi} \left[\frac{1}{4} + \ln\left(\frac{D}{r}\right) \right] \text{Hy/m (for parallel wires in space).} \quad (9.44)$$

As an example, we use Equations 9.37, 9.39, and 9.42 to find an expression for the characteristic impedance of coaxial cables:

$$Z_o = \frac{\ln(R/r)}{2\pi} \sqrt{\frac{\mu}{\epsilon}} \Omega. \quad (9.45)$$

9.11.2 Reflections on an Improperly Terminated, Lossless Transmission Line

As you will see in the following, the result of sending digital (pulsatile) data waveforms over improperly terminated transmission lines is the creation of transient glitches at the receiving end, which, along with other noise on the line, can cause errors in data transfer. A transient waveform will propagate without distortion or attenuation on a lossless transmission line with a velocity $v = 1/\sqrt{LC}$ m/s. On most coaxial cables used for digital data transmission, $0.65 \leq v/c \leq 0.83$, depending on cable design. However, when the transient voltage propagating on the line reaches the terminating or receiving end, a reflection or backward-propagating wave can arise at this end to satisfy certain boundary conditions. This first reflection propagates back to the source end of the line with velocity v , where a second reflection can arise and propagate at velocity v to the terminating end. This process can be described mathematically in some detail; the interested reader is referred to the text by Lathi (1965) for an excellent derivation of the relations we use.

Refer to Figure 9.35, showing a lossless transmission line of characteristic impedance, Z_o . To understand the dynamics of reflections on the line, we consider a voltage wave (step) that has propagated to the end of the line. A_d is the instantaneous voltage of the propagating wave just before it sees Z_L , and A_r is the instantaneous voltage of the back-propagating, reflected wave originating at Z_L to satisfy Ohm's law boundary conditions. Thus, at the Z_L termination, at $t = \tau = L/v$, we have the following:

$$V_L = V_d + V_r \quad (9.46)$$

$$I_L = \frac{A_d - A_r}{Z_o}. \quad (9.47)$$

The boundary condition at Z_L will be satisfied if

$$\frac{V_L}{I_L} = Z_L = Z_o \frac{A_d + A_r}{A_d - A_r}. \quad (9.48)$$

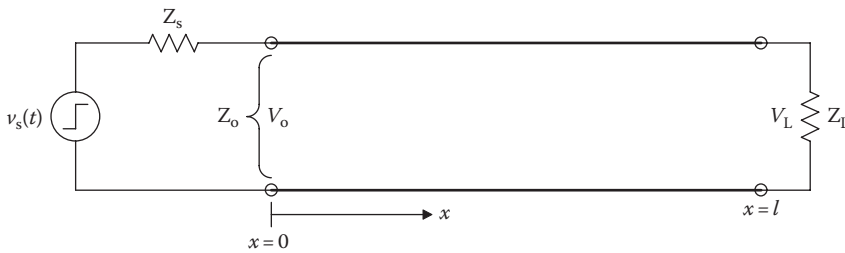


FIGURE 9.35
Schematic of a lossless transmission line of characteristic impedance Z_o and length l .

Solving Equation 9.39 for the ratio of A_r to A_d , we obtain the reflection coefficient, $\rho_L(s)$, defined for the *line's termination*:

$$\rho_L(s) = \frac{Z_L(s) - Z_o(s)}{Z_L(s) + Z_o(s)} = \frac{A_r}{A_d} \tag{9.49}$$

A similar reflection coefficient can be found for the *source end of the line*:

$$\rho_S(s) = \frac{Z_S(s) - Z_o(s)}{Z_S(s) + Z_o(s)} \tag{9.50}$$

where Z_s is the Thevenin source impedance of the line.

In example 1, let us examine what happens on a mismatched line when a positive 5.0 V step is generated by $v_s(t)$ at $t=0$. We assume zero initial conditions on the line, and let $Z_o=75 \Omega$, $Z_S=R_S=75 \Omega$, and $Z_L=R_L=\infty$ (open circuit). The step voltage is acted on by a voltage divider at $x = t = 0$. The source-end voltage divider attenuates by a factor of $75/(75 + 75) = 0.5$. Hence, the initial propagating voltage, V_o , is 2.5 V. The voltage at R_L is 0 until $t = \tau = L/v$. At $t = L/v$, the wave reaches R_L , and boundary conditions require that a reflected wave be generated. The termination reflection coefficient is found to be

$$\rho_L = \frac{R_L - Z_o}{R_L + Z_o} = \frac{-75}{+75} = +1. \tag{9.51}$$

The source reflection coefficient is

$$\rho_S = \frac{R_S - Z_o}{R_S + Z_o} = \frac{75 - 75}{75 + 75} = 0. \tag{9.52}$$

Hence, the first reflected wave has a value of

$$V_r(1) = 1.000 \times 2.5 = 2.5 \text{ V}. \tag{9.53}$$

The voltage at the load is thus $V_L = 2.5 + 2.5 = 5.0 \text{ V}$, over $1/v \leq t \leq 3l/v$. At $t = 2L/v = 2\tau$, the reflected wave of 2.5 V reaches the source end of the line. Then, there will be no reflection, because $\rho_S = 0$. The SS has been reached. In the SS, we expect the voltage everywhere on the line to be 5.0 V, as no current flows on the line. The waveforms for V_o and V_L are shown in Figure 9.36A.

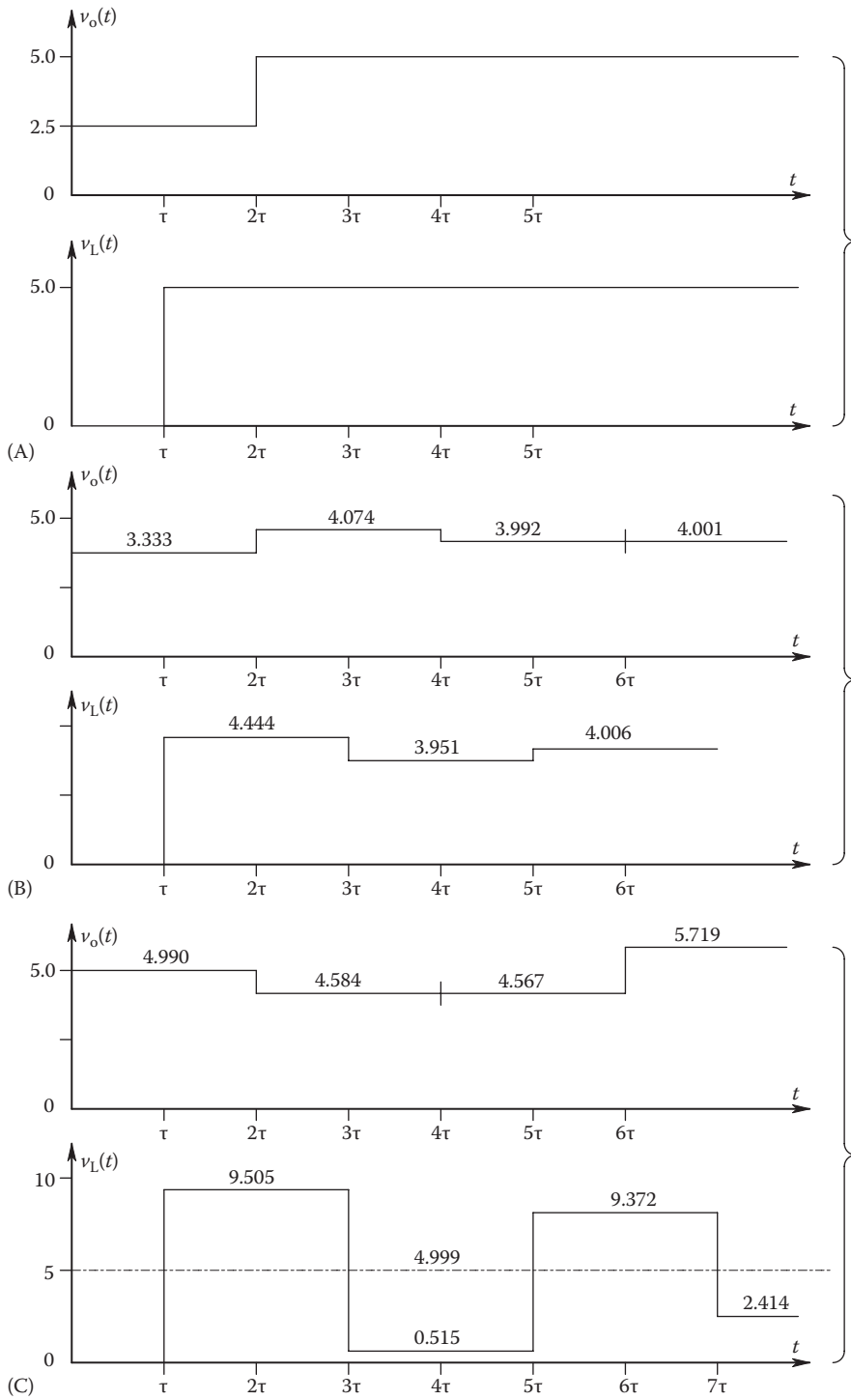


FIGURE 9.36

(A) Voltages at the source (v_o) and terminal (v_L) ends of a transmission line. See text Example 1 for conditions. (B) Transients on the transmission line of Example 2. (C) Transients on the transmission line of Example 3.

In example 2, there are reflections at both ends of the line. We let $Z_o = 50 \Omega$, $R_s = 25 \Omega$, and $R_L = 100 \Omega$, giving a moderately mismatched and underdamped line. Now, the reflection coefficients are $\rho_L = (100 - 50)/(100 + 50) = 1/3$ and $\rho_S = (25 - 50)/(25 + 50) = -1/3$. The amplitude of the initial wave at $t = 0$ is $V_o(0+) = 5 \times 50/(50 + 25) = 3.333$ V. The 3.33 V step travels at v m/s to the load end of the line, arriving at $t = \tau = L/v$ s. There, to satisfy boundary conditions, a reflected wave, $V_r(1)$, is generated, propagating toward the source. Its amplitude is given by

$$V_r(1) = 3.333 \times \rho_L = \frac{3.333}{3} = 1.111 \text{ V.} \tag{9.54}$$

Now, the voltage across R_L is $V_L = 3.333 + 1.111 = 4.444$ V for $\tau \leq t \leq 3\tau$. $V_r(1)$ arrives at the source end of the line at $t = 2L/v$. A second reflection originates at $t = 2L/v$ at the source end of the line, propagating toward R_L . The process of waves of diminishing amplitude being reflected from end to end of the transmission line continues, until practically a DC SS is reached.

Table 9.5 illustrates a systematic way to find the amplitudes of the reflections at the ends of a mismatched, lossless line. It is seen that a type of power series in the reflection coefficients describes the line's behavior. In the table, $v_s(t)$ is the Thevenin (open-circuit) source voltage, V_o is the voltage at the source end ($x = 0$) end of the line, and V_L is the voltage at the load end ($x = L$) of the line. $v_s(t)$ is assumed to be a 5 V step for simplicity. Other numerical

TABLE 9.5

Example of Systematic Means to Calculate the Reflections on a Mismatched, Lossless Transmission Line (parameters are from Example 2)

Time $\tau = L/v$	V_o	V_L
0	$V_o(1) = V_s(0+) \frac{Z_o}{Z_o + Z_s} = 5 \frac{50}{50 + 25}$	0
τ	3.33333 = 3.33333 [1.33333]	$V_L(1) = V_o(1) [1 + \rho_L] = 4.44444$
2τ	$V_o(2) = V_L(1) + V_o(1) [\rho_L^1 \rho_S^1]$ = 4.44444 - 3.33333 [0.11111] = 4.07407	4.44444
3τ	4.074074	$V_L(3) = V_o(2) + V_o(1) [\rho_L^2 \rho_S^1]$ = 4.07407 - 3.33333 [0.037037] = 3.9506173
4τ	$V_o(4) = V_L(3) + V_o(1) [\rho_L^2 \rho_S^2]$ = 3.95062 + 3.33333 [0.0123457] = 3.9917696	3.95062
5τ	3.99177	$V_L(5) = V_o(4) + V_o(1) [\rho_L^3 \rho_S^2]$ = 3.99177 + 3.33333 [0.0041152] = 4.0054870
6τ	$V_o(6) = V_L(5) + V_o(1) [\rho_L^3 \rho_S^3]$ = 4.00549 - 3.33333 [0.0013717] = 4.00091	4.00549
∞	$V_o(SS) = V_L(SS) = V_s(0+) \frac{Z_L}{Z_L + Z_s} = 5.0 \frac{100}{100 + 25} = 4.00000$ V	

parameters are from Example 2. Data from Table 9.5 are plotted in Figure 9.36B. Note that both $v_o(t)$ and $v_L(t)$ closely reach SS in three reflections.

In a third example, we examine the troublesome situation where logic signals are sent to a receiver along a long, severely mismatched transmission line. We assume a coaxial line is driven from an op amp buffer having a very low output resistance; for example, $R_s = 0.1 \Omega$. The signal is received by a linear amplifier with an input resistance of $R_L = 1000 \Omega$. The characteristic impedance of the coaxial cable, Z_o , is 50Ω , its length is 3 m, and the propagation delay at 1 MHz is known to be 5.136 ns/m. The reflection coefficients are found to be $\rho_L = 0.90476190$ and $\rho_S = -0.9960798$. The amplitude of the initial wave is

$$V_o = 5.0 \frac{50}{50 + 0.1} = 4.99002 \text{ V.} \quad (9.55)$$

Following the procedure used in Table 9.5, we can calculate and plot the voltage transient at the load end of the cable, $V_L(t)$. The time for a wave to travel down the cable is $\tau = L/v = 15.408$ ns. For $0 \leq t \leq \tau$, $v_L(t) = 0$, as the wave has not yet reached R_L . At $t = \tau$, the V_o wave reaches R_L , and the first reflection is generated. Table 9.6 summarizes the V_o and V_L values for this cable.

Of particular importance in this severe case of transmission line mismatch is the fact that the received voltage goes low 46.2 ns after going high. Such behavior will render data transmission virtually impossible or at least severely limit its rate to well below the reciprocal of the time it takes $v_L(t)$ to reach values above the logic HI level at the receiver.

By the simple expedient of making $R_s = R_L = Z_o$, all reflections on the line are eliminated, and the only problem is the 15.4 ns initial delay in the signal reaching R_L . Of course, the open-circuit V_s now should be made 10 V because the SS voltage at R_L will be $V_s/2$ by the R_s , R_L voltage divider.

In the earlier examples, we have assumed a short (<10 m), lossless transmission line. If longer lines are used, in addition to possible reflection problems due to mismatched terminating and source impedances, there is noticeable attenuation of high frequencies due to cable losses and a resulting rounding and attenuation of pulses as they travel along the line. The longer the line, the more severe the attenuation and loss of high frequencies and the slower pulsed data must be sent in order to resolve individual pulses. Voltage and current transfer functions for lossy transmission lines have been derived by Lathi (1965). They can be expanded into power series and then inverse Laplace

TABLE 9.6
Reflections on the Cable of Example 3
(see Figure 9.36C for plots)

t	V_o	V_L
0	4.99002	0
τ	4.99002	9.50480
2τ	4.58415	9.50480
3τ	4.58415	0.51537
4τ	4.56762	0.51537
5τ	4.56762	9.37157
6τ	5.71910	9.37157
7τ	5.71910	2.41447
∞	4.99950	4.99950

transformed to obtain the voltage and current at any distance along a finite line, that is, $v(x, t)$ and $i(x, t)$, given an input transient at $t = 0$.

To model the behavior of transmission line in the time domain, two approaches are commonly used. The first approach uses a series of lumped parameter, per unit length R - L / C - G circuits as shown in Figure 9.33. Their behavior is solved with an ECAP such as MicroCap™ or SPICE™. The second approach is to use FFT techniques with the voltage transfer function, $H_v(x, j\omega)$, to directly find the voltage on the line, $v(x, t)$.

9.12 Data Transmission on Fiber-Optic Cables

Other than radio frequency and EM transmission, FOCs are the means for very-high-speed, serial data and video transmission. They have rapidly replaced conventional wire transmission lines in point-to-point applications. FOC systems are used in increasing numbers in instrumentation systems because of their low cost, high reliability, low loss, immunity from EMI, and wide bandwidth.

FOCs are dielectric waveguides in which information is generally transmitted as 0,1 amplitude-modulated electromagnetic waves of lightwave frequencies. One major application of FOC systems is in robust telecommunications. The channel capacity of fiber-optic communications links, however calculated, far exceeds that of conventional wire communications (e.g., coaxial cables, twisted pairs) and present satellite microwave transponders. The present bandwidth of FOC data transmission systems is limited by the power, wavelength and speed of the LED or LAD sources, and the sensitivity and response time of the PIN or APD detectors coupled to the FOC. Theoretically, FOCs can support data bandwidths approaching a terahertz (THz); *the present state of the art, limited by terminal photonic devices, is in the low GHz range*. Such high bandwidths are seldom needed in instrumentation systems, where 5 Hz to 125 MHz signal bandwidths are more commonly used.

As in the case of transmission lines, an FOC behaves in the large like an LPF with a transport lag. The longer a given FOC, the greater the transmission lag and the attenuation at the receiver and the lower the system bandwidth for data transmission. A practical criterion for FOC bit rate can be established by considering how close in time two input current pulses to the LED or LAD source can be in order to just resolve the two pulses as a 101, instead of a 111 or 110, at the photodiode output.

FOCs are made from glasses or plastics. Fiber diameters range widely with the cable application. In telecommunications, FOCs are rapidly replacing copper twisted pairs for all but local subscriber lines. The reasons are as follows: FOCs are far less expensive than wire cables, they are immune to electrical and electromagnetic interference, and they have bandwidths greater than conductive transmission lines. A basic FOC serial communications channel is shown schematically in Figure 9.37. A cross section of an HP HFBR-2204 fiber-optic receiver is shown.

9.12.1 FOC Basics

Figure 9.38 illustrates a typical step-index FOC. The core of a step-index FOC has a uniform (in x , y , and z coordinates) optical index of refraction, n_1 . The FOC core carries the light energy used to signal data. The light energy propagates in the core in a variety of modes, as we will see. To understand how light is entrained in the core, we need to review

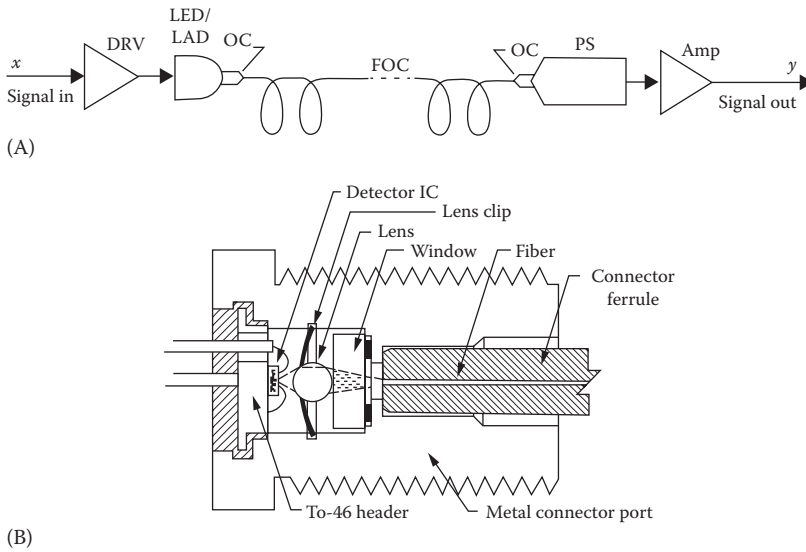


FIGURE 9.37

(A) Organization of a single, fiber-optic, serial communications channel. *Note:* OC = optical coupler, FOC = fiber-optic cable, DRV = driver amplifier, LED/LAD = modulated light source, PS = photo sensor (fiber-optic receiver). (B) Section through an HP model HFBR-2204 fiber-optic receiver. A spherical lens concentrates the light from the OF on the photosensor. (Courtesy of Hewlett-Packard, Palo Alto, CA.)

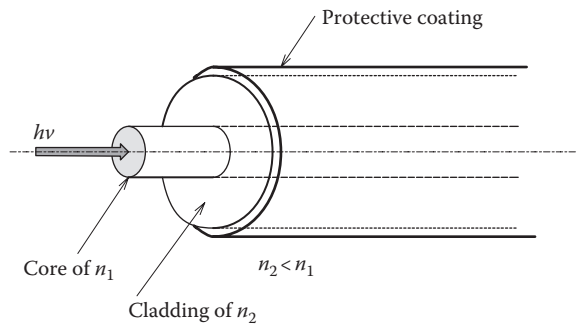


FIGURE 9.38

Diagram of the cutback end of a step-index FOC.

some elementary optical theory governing refraction and reflection at plane interfaces of two dielectric media having different indices of refraction. The index of refraction of an optical medium is the ratio of the speed of light in vacuo (free space) to the speed of light in the medium, v_1 . The refractive index of the FOC core is thus

$$n_1 = \frac{c}{v_1}. \tag{9.56}$$

The refractive index of the cladding material (e.g., glass, plastic) surrounding the core is

$$n_2 = \frac{c}{v_2}. \tag{9.57}$$

Obviously, n_o for air $\cong 1$. In an FOC, $n_1 > n_2 > 1$ by design. Figure 9.39A through C shows what happens to light rays in the core incident on the core/cladding interface, here modeled as a plane surface for simplicity. A critical angle, θ_c , exists such that if the angle of incidence $\theta_1 < \theta_c$, then some fraction of light energy enters the cladding and is trapped and dissipated in it. The angle relations for this case are given by the well-known Snell's law:

$$\frac{n_1}{n_2} = \frac{\sin(\theta_2)}{\sin(\theta_1)} \tag{9.58}$$

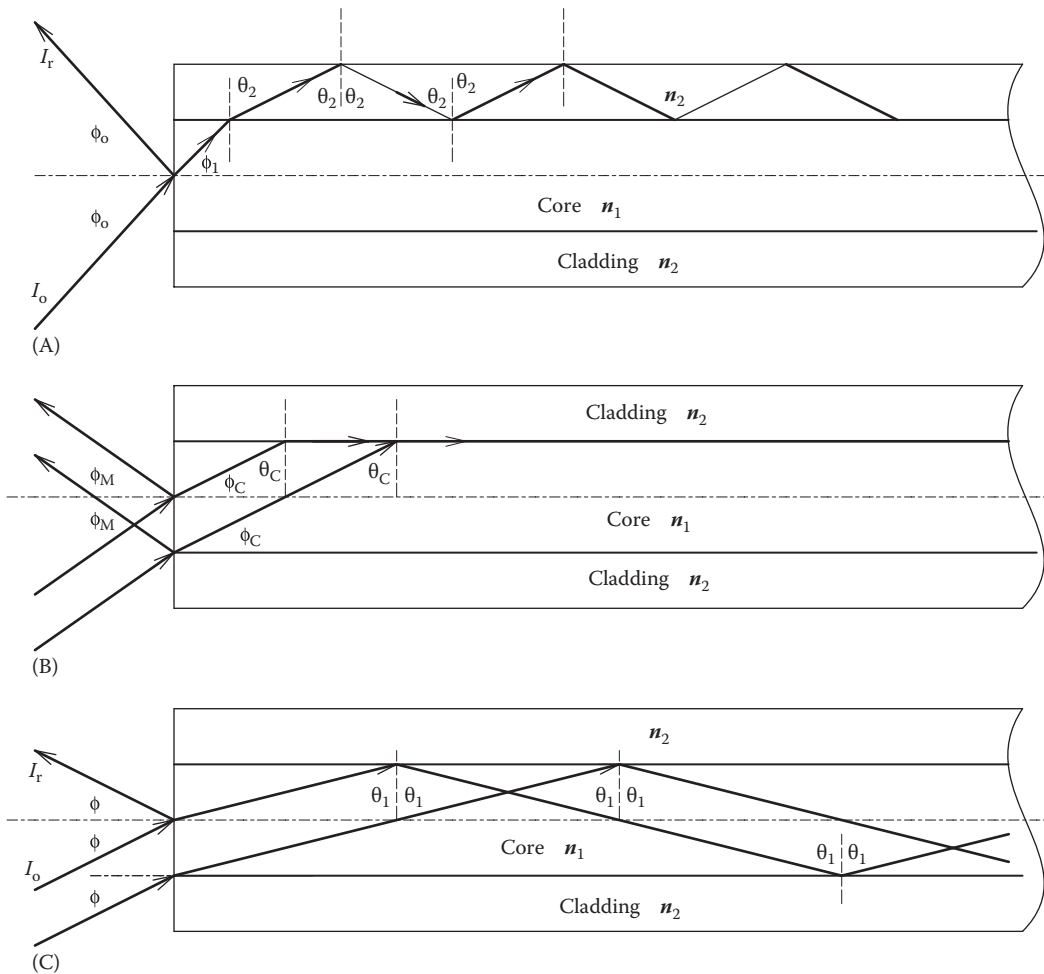


FIGURE 9.39

(A) Lateral view of a step-index FOC. Here, an incident ray enters the cable core at a very steep angle, ϕ_o . The angle of incidence of the internal ray on the core/cladding boundary is less than the critical angle, ϕ_c , and most of its energy enters the cladding and is trapped there. (B) Critical angle condition for the core/cladding boundary. (C) Now, the incident ray is sufficiently normal to the end of the core so that most of its energy propagates down the core by alternately reflecting off the core/cladding boundary. (The situation is actually more complex when one considers the 3D, cylindrical geometry, and incident rays not in the plane of the page.)

A critical angle, $\theta_1 = \theta_c$, exists such that $\theta_2 = 90^\circ$. In this case, Equation 9.58 gives us

$$\theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right). \quad (9.59)$$

Rays in the core with $\theta_1 > \theta_c$ experience *total reflection* at the core/cladding boundary, and thus most of their energy remains within the core, bouncing back and forth off the boundary as they propagate to the receiver end of the FOC. Such entrained rays are called *meridional rays* because they traverse the axis of the fiber.

It can be shown from detailed consideration of the Maxwell equations for EM waves incident on plane dielectric boundaries that some portion of an incident ray's energy will be reflected, even if $0 \leq \theta_1 \leq 90^\circ$, and some portion will be transmitted. The proportions reflected and refracted can be shown to depend on the polarization state of the incident light, on the indices of refraction of the two media, and on the angle of incidence. For example, the reflected fraction of unpolarized light normally incident on the flat, polished end of a step-index FOC is given by (Sears 1949)

$$I_r = I_i \frac{(n_1 - 1)^2}{(n_1 + 1)^2}. \quad (9.60)$$

For $n_1 = 1.5$, $I_r/I_i = 0.040$, or 4% of the normally incident light energy does not enter the cable core. In passing from a higher to a lower refractive index medium, there is a very steep increase in I_r/I_i as the angle of incidence approaches θ_c . This behavior is why light traveling in the FOC core tends to remain in it and not enter the cladding. On the other hand, there is a slow increase in I_r/I_i as the angle of incidence approaches and passes through θ_c as unpolarized light traveling in a low refractive index medium is incident on a plane boundary with a high refractive index medium. Thus, some fraction of the light trapped in the cladding can reenter the core. The reflection vs. refraction relations (Fresnel's formulae) are given in detail in Sears (1949)

For the special case of light entering on the z -axis of the FOC ($\phi = 0^\circ$, $\theta_1 = 90^\circ$), it takes the axial ray $\tau_a = Ln_1/c$ seconds to reach the end of the FOC at $z = L$ meters. From simple geometry, photons following a bouncing path can be seen to take

$$\tau_{\theta_1} = \frac{Ln_1}{c \sin(\theta_1)} \text{ s} \quad (9.61)$$

to reach the end of the cable. The difference between τ_{θ_1} and τ_a is called the *modal dispersion*. Modal dispersion causes broadening of transmitted pulses and thus is a factor reducing the transmission bandwidth of multimode, step-index FOCs. The reader who is interested in a rigorous treatment of modal propagation in plane (rectangular) waveguides using the Maxwell equations should consult texts on FOCs by Cheo (1990), Gowar (1984), or Keiser (1983). The mathematical treatment of modal propagation in cylindrical waveguides (FOCs) is complicated by the fact that all field components are coupled. The text by Cheo (1990) gives such analyses.

Another geometrical optical consideration of FOCs is the *numerical aperture* (NA) of a cable. The NA of an FOC is related to the largest value of the ray entrance angle to the core

from space, ϕ_M , that will permit a ray to propagate down the core. In other words, a ray entrance angle greater than ϕ_M results in $\theta_1 < \theta_C$, and nearly all its energy is coupled into the cladding and lost. Total internal reflections occur for $\phi < \phi_M$, and modal propagation occurs. At the critical entrance angle, we have, by Snell's law,

$$n_0 \sin(\phi_M) = n_1 \sin(\phi_C). \quad (9.62)$$

Also, at the critical angle in the core, from Equation 9.59, we can write

$$\sin^2(\theta_C) = \left(\frac{n_1}{n_2} \right)^2. \quad (9.63)$$

By trigonometric identity,

$$\cos(\theta_C) = \left[1 - \sin^2(\theta_C) \right]^{1/2}, \quad (9.64)$$

and from the right triangle,

$$\sin(\phi_C) = \cos(\theta_C). \quad (9.65)$$

Substituting Equations 9.62, 9.63, and 9.65 into Equation 9.64, we obtain an expression for the step-index FOC's NA (note $n_0 = 1$):

$$\text{NA} = \sin(\phi_M) = \left(n_1^2 - n_2^2 \right)^{1/2}. \quad (9.66)$$

Figure 9.40 shows a geometrical interpretation of NA in terms of θ_C , n_1 , and n_2 . Note that the FOC NA is not a function of the core diameter.

FOCs can be subdivided into *multimode* and *single-mode* fibers. Multimode fibers generally have core diameters ranging from 50 μm to as great as 1000 μm . It is physically easier to couple light into and out of multimode fibers because of their greater core diameters. Single-mode FOCs have core diameters from 3 to 12 μm , and, consequently, I/O coupling

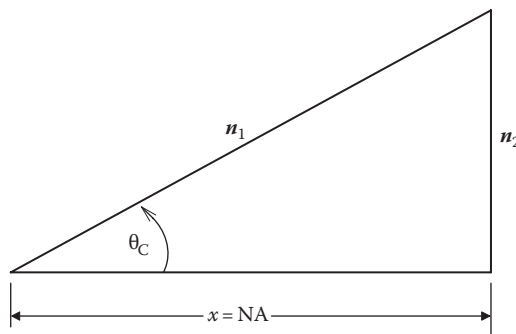


FIGURE 9.40

A geometrical interpretation of a fiber's NA.

requires great precision. The maximum number of modes in a planar, dielectric waveguide (as a first approximation to a cylindrical FOC) can be shown to be approximated by (Stark et al. 1988)

$$M_{\text{MAX}} = 2d (\text{NA}) / (\lambda n_1) - 1, \quad (9.67)$$

where

d is the core diameter

λ is the wavelength of light on the core

NA is the fiber's NA, given by Equation 9.66

From Equation 9.61, the maximum modal dispersion can be shown to be

$$\Delta\tau = (\tau_c - \tau_a) = \frac{Ln_1}{cn_2} (n_1 - n_2) \text{ s.} \quad (9.68)$$

While single-mode fibers do not suffer from modal dispersion, they, as well as multi-mode fibers, exhibit material dispersion, which is caused by the fact that n_1 is not a constant but is, in fact, a function of the wavelength of the light in the core. Cheo (1990) showed that the per unit length difference in group delay, $\Delta\tau_G$, for transmitted components separated by $\Delta\lambda$ in wavelength is given by

$$\Delta\tau_G = \frac{\lambda\sigma_s}{c} \frac{d^2n}{d\lambda} \text{ s/m,} \quad (9.69)$$

where σ_s is the RMS spectral width of the source, given by

$$\sigma_s = \int_0^\infty (\lambda - \lambda_o)^2 S(\lambda) d\lambda \quad (9.70)$$

$$\lambda_o = \int_0^\infty \lambda S(\lambda) d\lambda, \quad (9.71)$$

where $S(\lambda)$ is the spectral distribution of the source.

Thus, the material dispersion is proportional to the source's spectral width and the second derivative of $n_1(\lambda)$ around the mean wavelength, λ_o . Interestingly, $\Delta\tau_G \rightarrow 0$ at λ around 1300 nm for silica fiber cores. Hence, sources operating around this wavelength offer improved bandwidth for single-mode FOCs.

A detailed, theoretical derivation of FOC bandwidth, based on a two pulse, intersymbol interference under noise-free conditions, was given by Stark et al. (1988). These authors assumed an NIR source λ_o of 1000 nm, the second derivative of the phase constant, $\beta''(\lambda_o) = 3 \times 10^{-26} \text{ s}^2/\text{m}$, and an interpulse interference ratio, $\delta = 0.2$, at a point midway between the two received unit pulses. (The tails of the unit height pulses sum to 0.2 half-way between them.) Their cable bandwidth, B , defined as the reciprocal of the interpulse interval required to make $\delta = 0.2$, increases with the spectral purity of the source and decreases with cable length. Figure 9.41, adapted from Stark et al., illustrates FOC bandwidth, B , as a function of source bandwidth, $\Delta\lambda$ around 1000 nm, and cable length.

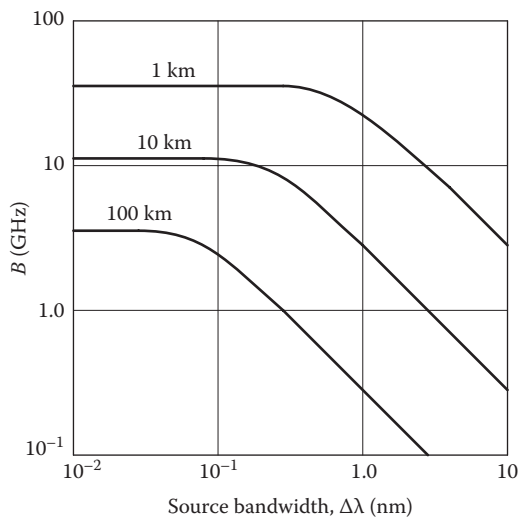


FIGURE 9.41

Plots of FOC signal bandwidth, B , in GHz as a function of FOC length and the spectral width of the 1000 nm source, $\Delta\lambda$. Note that as the optical source’s bandwidth increases, the permissible transmission rate decreases, underscoring the value of IR LADs as FOC system sources. (Adapted from Stark, H. et al., *Modern Electrical Communications*, 2nd edn., Prentice-Hall, Englewood Cliffs, NJ, 1988. With permission.)

The *graded-index fiber* (GRIN) represents a design that mitigates the effect of modal dispersion found in step-index fibers and yet keeps the large core diameter permitting simpler I/O connections. Figure 9.42A and B illustrates a median section of a graded index FOC and its refractive index profile. There are three principal GRIN core index profiles: *parabolic profile* (Equation 9.72), α *profile* (Equation 9.73), and *hyperbolic* or *selfoc* profile (Equation 9.74):

$$n(r) = n_{1(\text{MAX})} (1 - \epsilon r^2)^{1/2}, \quad 0 \leq r \leq r_1 \tag{9.72}$$

$$n(r) = n_{1(\text{MAX})} \left[1 - 2\Delta \left(\frac{r}{r_1} \right)^\alpha \right]^{1/2} \tag{9.73}$$

$$n(r) = n_{1(\text{MAX})} \operatorname{sech}(r\Delta r_1) \tag{9.74}$$

$$\Delta \equiv \frac{n_{1(\text{MAX})} - n_2}{n_{1(\text{MAX})}}. \tag{9.75}$$

Modal dispersion is greatly reduced in GRIN FOCs because instead of reflecting at the core/cladding boundary of a step-index fiber, oblique rays are bent so that they follow a smooth, periodic pathway down the core, as shown in Figure 9.42B. An axial ray travels down the shortest path in the center of the core at the slowest velocity, $v = c/n_{1(\text{MAX})}$. An oblique ray travels a longer path, most of which lies in the higher-velocity periphery of the core. Thus, both oblique and direct rays tend to arrive at the receiving end at the same time, minimizing modal dispersion.

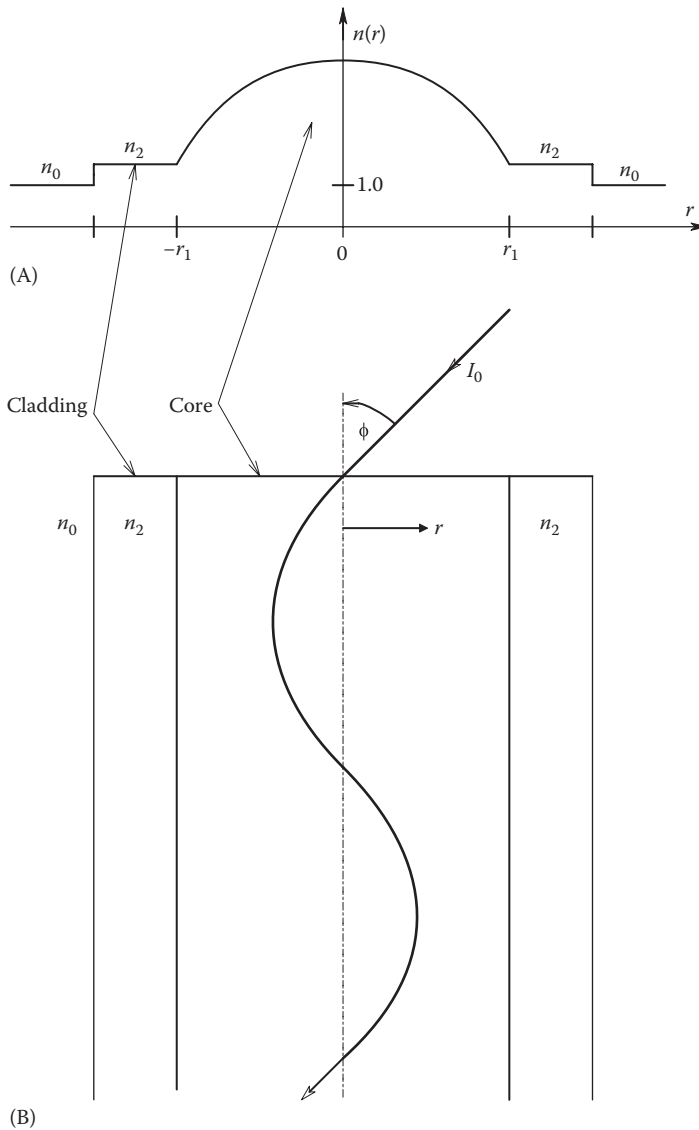


FIGURE 9.42

(A) Plot of the refractive index of a graded index (GRIN) OF vs. fiber radius. (B) 2D path of an obliquely entering ray as it propagates down the GRIN FOC.

The differential delay between the fastest and slowest modes in an α profile, GRIN fiber, for a monochromatic (LAD) source can be shown to be

$$\Delta\tau_G = \frac{n_{1(\text{MAX})}\Delta^2}{8c} \text{ s/m}, \tag{9.76}$$

where

$$\alpha = 2(1 - 1.2\Delta) \tag{9.77}$$

for minimum intermodal dispersion (Stark et al. 1988).

9.12.2 Semiconductor Sources and Detectors Used in FOC Data Transmission

Two principal sources are used for fiber-optic systems, LEDs and LADs. LEDs used in FOC systems are generally double heterojunction designs, as shown in Figure 9.43. A thin recombination layer of *p*-doped GaAs is sandwiched between thicker layers of *p*-type $\text{Al}_x\text{Ga}_{(1-x)}\text{As}$ and *n*-type $\text{Al}_x\text{Ga}_{(1-x)}\text{As}$. For example, the HP HFBR-1119T high-efficiency, fiber-optic transmitter is a 1300 nm, surface-emitting InGaAsP LED. It is designed for 50 or 62.5 μm core multimode OF. The HP HFBR-2119T receiver uses an InGaAs PIN photodiode coupled to a custom silicon transimpedance amplifier IC. These fiber-optic data links support a maximum signaling rate of 266 MBd over 1.5 km. Modulation of the light intensity is accomplished by switching the diode's current on and off.

LADs are used in single-mode FOC systems designed to operate at extremely high data rates. An example of a double heterostructure LAD is shown schematically in Figure 9.44. As long as the LAD's supply current is held constant along with its temperature, its light output remains constant and of constant spectral purity. The envelope of an LAD's output line spectrum narrows with increased output power, and its peak shifts to a slightly longer wavelength, for example, a spectrum peak at 830 nm at 4.6 mW output shifts to a peak at 833 nm at 12 mW output. This particular LAD had about 2.5 spectral lines/nm (Cheo 1990). Cheo described a constricted double heterostructure LAD that produces a single, narrow, spectral output peak at about 837.5 nm at an input current of 100 mA. The two major types of semiconductor photodetectors used in FOC systems are the PIN photodiode and

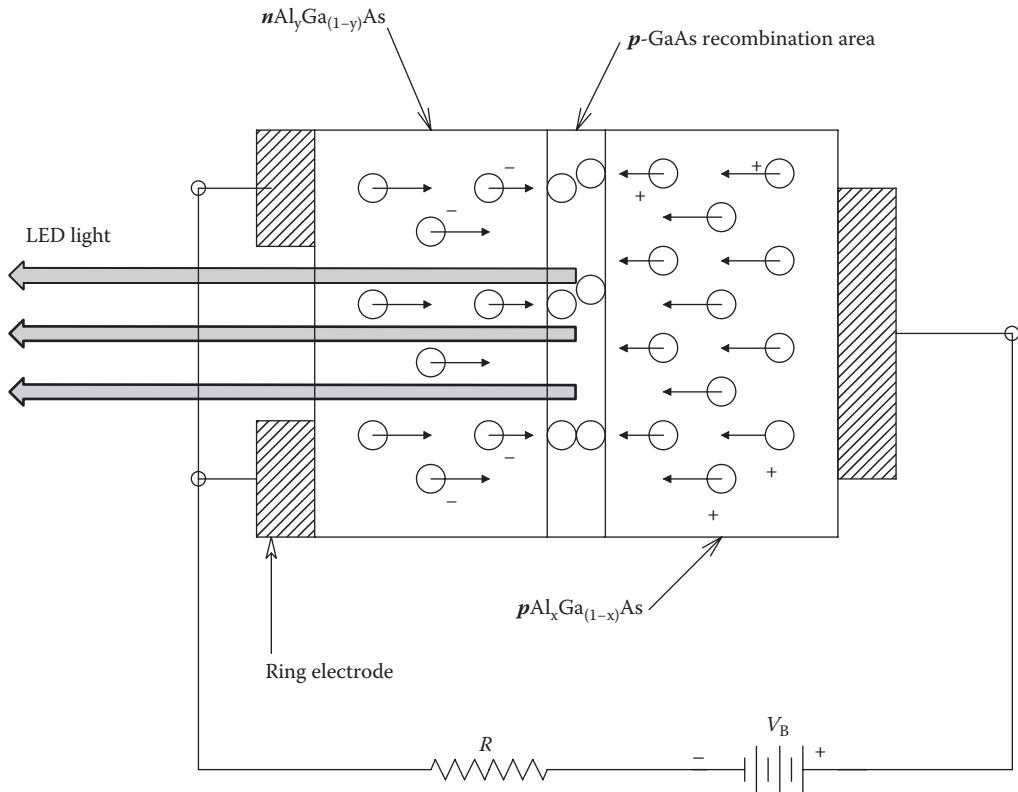


FIGURE 9.43 Layer cake diagram of a double heterojunction AlGaAs LED source.

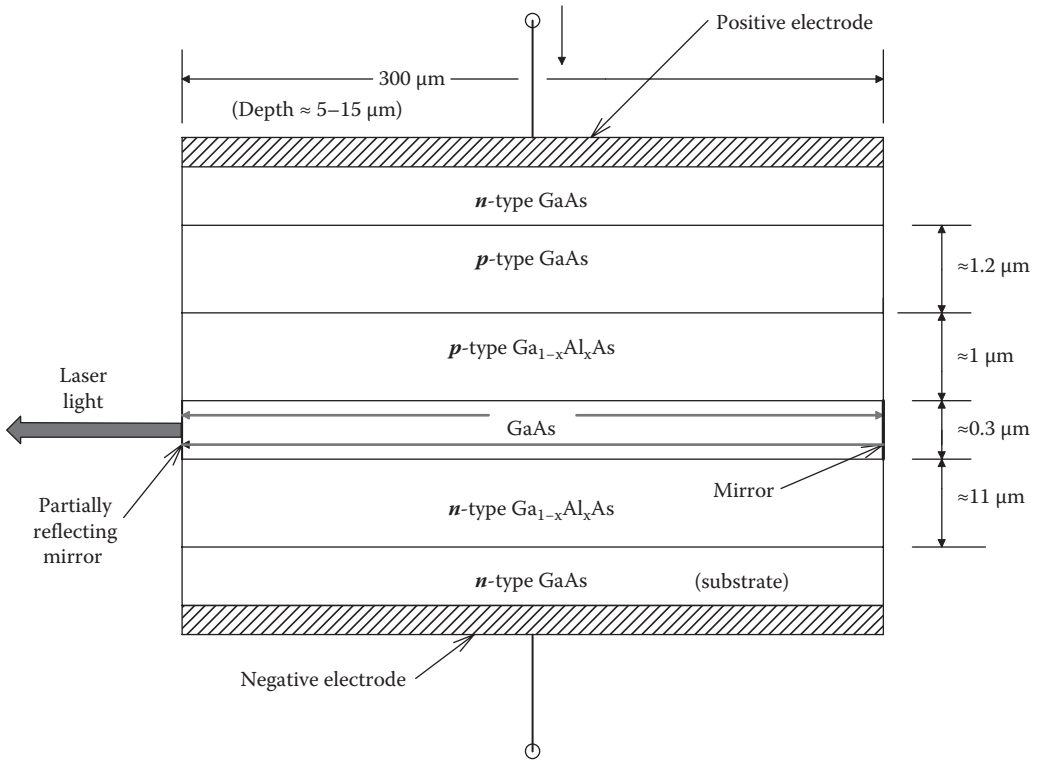


FIGURE 9.44

Layer cake model of a stripe geometry, double heterojunction, GaAs LAD. (From Stark, H. et al., *Modern Electrical Communications*, 2nd edn., Prentice-Hall, Englewood Cliffs, NJ, 1988.)

the APD. All photodetectors must have high quantum efficiency at the operating wavelength, high responsivity, rapid response time, and small size. The quantum efficiency, η , is defined as the number of carrier pairs (hole + electron) generated per incident photon. Ideally, $\eta = 1$. In practice, we see $0.3 \leq \eta \leq 0.95$. The responsivity, $R(\lambda)$, is defined as the ratio of primary photocurrent to incident optical power:

$$R(\lambda) = \frac{\eta e}{h\nu}. \tag{9.78}$$

For an ideal photodiode, $\eta = 1$ and $R = \lambda/1.24 \text{ A/W}$. Typical values for $R(\lambda)$ are $0.65 \mu\text{A}/\mu\text{W}$ for Si at 800 nm and $0.45 \mu\text{A}/\mu\text{W}$ for Ge at 1300 nm (Stark et al. 1988).

The *I* in PIN stands for *intrinsic* semiconductor material, sandwiched between a *p*- and *n*-layer. The PIN photodiode is generally operated in reverse bias (see Figure 9.45). The photocurrent, I_p , is added to the dark (leakage) current, I_L . The objective is to resolve a small change in the total diode reverse current (due to a pulse of absorbed photons) above the shot noise and thermal noise in the diode circuit. PIN photodiode responsivity is a function of the materials from which it is made. Silicon PINs operate effectively from 200 to 1000 nm, with response rising with shorter wavelengths. Germanium PINs are useful in the range from 1000 to 1800 nm. PINs made from $\text{Al}_x\text{Ga}_{(1-x)}\text{AsSb}$ have η s as high as 80% in the 1.8–2.5 μm range (Cheo 1990). The HP HFBR-2100, 1300 nm, PIN, FOC detector has 1 ns output rise and fall time, enabling it to be used in a 1300 nm, 200 MBd, FOC system.

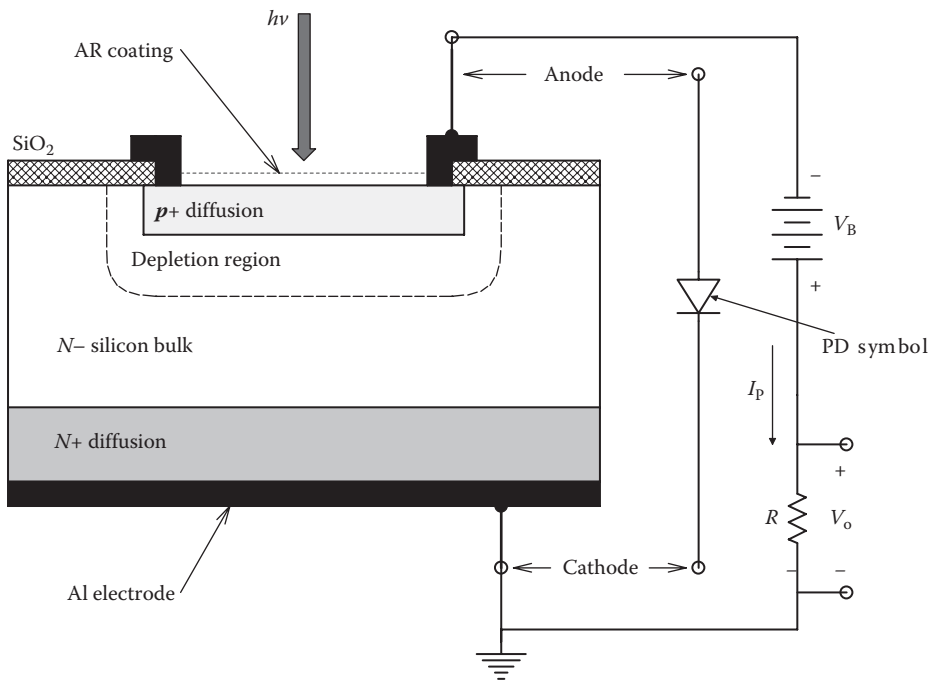


FIGURE 9.45
Layer cake diagram of a PIN photodiode.

The APD is a $p+ipn$, four-layered structure, shown schematically in Figure 9.46. Its increased sensitivity results from an internal, intrinsic gain mechanism that greatly increases the number of mobile carriers above the original photo-induced holes and electrons. The DC electric field conditions in the reverse-biased APD are such that the depletion region reaches into the i -region. The electric field at the pn junction is only 5%–10% lower than that necessary for avalanche breakdown of the junction. Mobile photoelectrons and holes gain enough kinetic energy in the high E field region of the p layer to produce additional mobile electrons and holes through inelastic collisions, which impart enough energy to promote new carriers to the conduction band. These secondary carriers, in turn, can gain enough kinetic energy to cause further ionization. Thus, n primary photoelectrons can give rise to αn electrons at the diode output. n primary photo-holes result in αn holes contributing to I_p . The net APD current multiplication factor, M , is a function of the internal electric field distribution, diode geometry, doping, and the wavelength of the incident photons. Cheo (1990) gave a good analytical treatment of APD behavior and noise. APDs are made from Si, Ge, InGaAsP, etc., depending on the desired operating wavelength. At wavelengths below 1 μm , Si APDs have been built with net current gains of about 100, quantum efficiencies approaching 100%, and response times around 1 ns. In general, APDs exhibit faster rise times than PIN diodes; however, each requires careful setting of its operating voltage and constant operating temperature.

9.12.3 FOC Systems

The instrumentation system designer who wishes to interconnect groups of instruments, such as a VXI cluster, over a long distance to a controlling computer by fiber-optic link

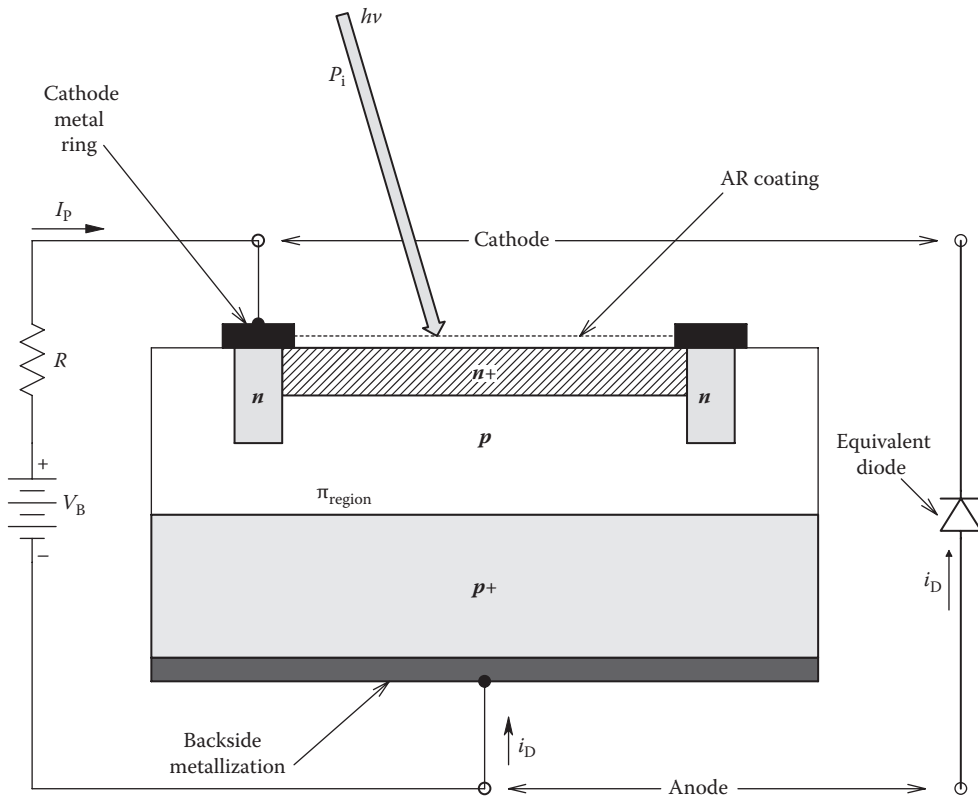


FIGURE 9.46
Layer cake diagram of a silicon APD.

no longer has the burden of assembling custom components to build the link. HP markets a wide variety of plastic and glass FOC components, including single and double (duplex) FOCs, LED transmitters and PIN receivers, and signal conditioners. HP also offers FOC system evaluation kits for designers to test feasibility with 1000 μm core plastic cable and 5 MBd, TTL-compatible transmitters and receivers. HP FOC link equipment ranges from 125 m, 40 kbd, TTL systems to 2 km, 200 MBd, ECL systems. This FOC equipment is described in the *HP Optoelectronics Designer's Catalog*, v. December 31, 2012.

9.13 Virtual Instruments

9.13.1 Introduction

With the development of high-speed bus architectures, and special A/D plug-in interface cards for PCs and laptop computers, there has been a trend to use the power of computer software to develop VI whose controls and displays are presented on a computer screen and whose operation and settings are controlled by the computer's keyboard and mouse. Rather than purchase several, expensive, dedicated, stand-alone instruments, it is more

economical to purchase a computer with a multichannel A/D interface card of appropriate speed and bit resolution or an outboard PXI system and the VI software to control, process, and display the measured quantities.

NI company has pioneered the development of computer-operated VIs and certain digital data processing algorithms used with them. Interface A/D cards, and outboard signal conditioning and A/D systems such as PXI modules connected to the computer through RS-232D, IVI-C, or GPIB buses, are offered by several vendors, including CyberResearch®, NI, and Keithley Instruments, to mention a few.

An example of a typical VI is a true RMS voltmeter. A time-varying analog signal, such as noise, is first antialias filtered and then A/D converted at a suitable rate; then M digital samples in each of N data sample epochs are each numerically squared and averaged; and the average of the squares is square rooted to generate numerical estimates of the true RMS value of the signal. For noise and noisy signals, the N averages of the squares are averaged and then square rooted to reduce the variance of the RMS estimate. The RMS voltage is presented on the computer screen as the display of a VI RMS voltmeter shown in a window. The operator can adjust the VI meter's range, the AAF's cutoff frequency, its sampling rate, the epoch length, M , and the averaging algorithm by setting the knobs and switches on the VI display with the mouse. Note that such a virtual RMS noise voltage is the value over the noise bandwidth of the AAF.

Another example of a user-implemented VI instrument is a *pulse-height window*. Such windows are used in experimental neurophysiology and nuclear physics. The VI software displays the upper and lower DC thresholds, φ_u and φ_l , respectively, and the signal with pulses to be discriminated as a moving (strip-chart type) display. The analog signal is antialias low-pass filtered, sampled, $A \rightarrow D$ converted, and digitally low-pass filtered to remove some noise. The window output is an event time record when pulses on the signal cross φ_l , do not cross φ_u , and then cross below φ_l . The VI operator can adjust the thresholds, the AAF cutoff frequency, the sampling rate, and the digital LPF parameters. The window output point process, $p_w(t) = \dagger \sum_{k=1}^{\infty} \delta(t - t_k)$, is an array of times when pulses lie within the window. $p_w(t)$ can be further processed to generate a stepwise, *instantaneous frequency record*. (t_k is the time of occurrence of the k th window-discriminated pulse relative to a local time origin.) From the point process, $p_w(t)$, the computer calculates the stepwise, continuous variable, $F_1(t)$. $F_1(t)$ is called the *demodulated instantaneous pulse frequency* of the discriminated pulses; the process defined by Equation 9.79 is called *instantaneous pulse frequency demodulation* (IPFD):

$$F_1(t) = \sum_{k=2}^{\infty} r_k [U(t - t_k) - U(t - t_{k+1})], \quad k = 1, 2, 3, \dots, \infty. \dagger (t_0 \text{ occurs at } t = 0.) \quad (9.79)$$

where $r_k \equiv 1/t_k - t_{k-1}$ is the k th element of instantaneous frequency of $p_w(t)$, $k = 2, 3, \dots, \infty$, and $U(t - t_k)$ is the *unit step*: 0 for $t < t_k$ and 1 for $t \geq t_k$. Note that t_1 is the time the first pulse occurs after $t = 0$, and $(t_2 - t_1)$ is required to define the initial element of instantaneous frequency in $F_1(t)$.

NI's LabVIEW™ offers a number of prepackaged VI modules: signal processing VIs include *FFTs*, *cross- and auto-power spectral analysis*, *Hilbert transforms*, and *joint time-frequency analysis (JTFA)* (Northrop 2003). Time series analysis VIs include *real convolution*, *deconvolution*, *auto- and cross correlation*, and *derivative and integral filtering*. *Oscilloscopes* and *signal generation* VIs are also available. VI modules can be user linked to perform complex signal analysis and real-time control tasks, including motion control. External PXI input

signal conditioning modules exist for RF signals, accelerometers, velocimeters, tachometers, thermocouples, pressure sensors, force and torque sensors, phase-sensitive rectifiers for carrier systems such as Wheatstone bridges, proximity sensors, photosensors, microphones, and other sensors, permitting custom VI systems to be built that allow sound and vibration analysis of machinery and appliances, acoustics signal processing, correlation studies on noise and vibrations, etc.

As if there weren't enough computer buses to consider in instrumentation, the NI-sponsored IVI Foundation, founded in 1998, has defined two unified bus architectures for VI drivers, the IVI-C and the IVI-COM. IVI-C is based on ANSI-C and can drive over 300 compatible instruments from NI, Keithley, Pickering, etc. IVI-C source code is generally available. IVI-COM is based on Microsoft COM™ technology and can drive ca. 50 instruments, mainly from Agilent. IVI-COM source code is not generally available. Both IVI architectures permit remote operation of compatible instruments; their input settings are managed as VI images on the host computer, and the data from the instruments are sent to the host for display, processing, and logging. Thus, any analog filtering or signal processing operations, AAF, $A \rightarrow D$ function, and peripheral DSP is done remotely, and each individual IVI-compatible instrument communicates its output to the host (NI 2003a–c).

9.13.2 PXI Systems

PXI stands for *PC extensions for instrumentation*. The PXI standard defines a modular instrumentation platform (hardware modules and operating software) designed specifically for robust measurement and automation applications. PXI peripheral modules are used as remote computer-controlled front ends for VIs. PXI modules generally have only input signal connectors and status LEDs on their narrow front panels; there are generally no control knobs or switches.

Computer communication with the PXI box and its modules can be through the 132 MB/s standard *PCI bus*. The PXI system has over 1000 products currently available from various vendors, including NI. These products include *crates*, sensor interfaces, power supplies, analog signal conditioners, DSP, and motion control driver output modules. The mechanical aspects of PXI and CompactPCI (connectors) are governed by Eurocard specifications (ANSI 310-C, IEC-297, and IEEE 1101.1), which have a long history of application in industrial environments (the reader interested in PXI hardware should consult the PXI Systems Alliance document, *PXI-1 Hardware Specifications* (PXI 2005), and *What is PXI?* (National Instruments 2013)).

NI has also developed their own transparent MXI-3 bus protocol to communicate with PXI peripheral modules. A PCI board is in the desktop computer, and slot 1 of the PXI crate has the peripheral MXI-3 interface card; copper or fiber-optic communication is used. Data rates of 132 MB/s peak and 84 MB/s are possible using the PCI/MXI-3 bus. Ethernet/Internet, SCSI, GPIB, VXI, VME, RS-232D, and wireless communication and control of PXI peripherals are also supported by NI. Two or more PXI crates with peripheral cards can be linked by PXI-to-PXI MXI-3 connections. The PCI specification allows for up to 254 buses connected in a system via PCI-to-PCI bridges such as MXI-3.

9.13.3 Summary

VIs have found great application in science, industry, aerospace, and medicine. They permit remote signal acquisition and conditioning modules to be located near the sensors and managed remotely by a host computer. The host presents data in a VI window

display that allows the operator to adjust input signal conditioning parameters and observe results on real time. In addition, the host can store incoming data on a hard disk or a CD-RW ROM disk. Links to and from the peripheral modules can be copper, fiber optic, or RF. A variety of bus protocols are available for communicating settings and data to and from modules.

VI technology has a large and growing application in engineering education. It offers an economical means to implement oscilloscopes, meters, bridges, signal generators, spectrum analyzers, and control signals without the expense of purchasing complete, dedicated instruments. The host computer can do spectral analysis on signals by FFT and other DSP routines like cross correlation of two signals.

Special VI systems are used in biomedical engineering labs for the same reasons: economy and simplicity. Students can observe ECG, EEG, and EMG waveforms and other physiological signals such as muscle force, blood pressure, and pulse oximetry signals, using medical isolation-grade peripheral signal conditioners.

9.14 Chapter Summary

In this chapter, we have introduced the primary limitations of converting signals from analog to digital format, namely, the introduction of an equivalent broadband noise due to quantization and the loss of high-frequency information due to aliasing. Quantization was seen to be caused by the numerical roundoff in representing analog data samples with digital numbers of finite bit length. Aliasing was shown to be the result of sampling a bandwidth-limited signal too slowly, that is, at a rate below twice the highest frequency in the signal's power density spectrum. Dithering was shown to be an effective way of mitigating the effect of quantization noise.

The R-2R DAC architecture was introduced as representative of most modern DACs. Hold circuits were described as a means of *freezing* a time-varying input signal to an ADC while conversion takes place and also of maintaining a DAC output voltage until the output is updated. Hold dynamics were derived.

Six types of ADCs were described, and their applications were discussed relative to instrumentation systems. Mention was made of PC-compatible data acquisition cards (containing ADCs, multiplexers and buffer memories).

Digital data buses were reviewed, and the IEEE-488, GPIB, parallel instrumentation bus was described in some detail. Serial data communication links, including the venerable RS-232D bus and the ubiquitous versatile USB bus, were also described and their limitations discussed. The CAMAC instrumentation packaging protocol and its related internal bus were covered, as well as the newer PXI system architecture.

In Sections 9.11 and 9.12, we discussed transmission lines used for digital data transfer and showed how improper impedance termination can lead to voltage reflections on the line that cause errors and significantly slow data transmission rates. Long transmission line poses an extra problem because line losses cease to be negligible, and there is not only delay in transmission, but transmitted pulses become rounded and attenuated as they traverse the line. The same problems of delay, attenuation, and distortion are also seen on FOCs, but FOCs have signal bandwidths that far exceed those of conventional, low-loss, coaxial cables. FOCs also are immune to various forms of electromagnetic interference because of their dielectric waveguide properties.

Problems

- 9.1 An 18-bit ADC is used to convert an audio signal ranging over ± 1 V.
- Find the size of the quantization step, q , required.
 - Find the RMS quantization noise associated with full-scale operation of this ADC.
 - Let $v_{in}(t) = 1 \sin(\omega_o t)$. Find the RMS SNR at the ADC output in decibels.
- 9.2 A 4-bit, R - $2R$, offset binary DAC is shown in Figure P9.2. Calculate the transfer characteristic between the input digital input, $\{d_3 d_2 d_1 d_0\}$, and V_o ($d_k = 0, 1$).
- 9.3 The simplified circuit of a 3-bit, bipolar, *flash* DAC is shown in Figure P9.3. It is shown with the input, 011 active.
- Make a table of V_o vs. $\{d_2 d_1 d_0\}$.
 - Discuss the source of output "glitches" in a practical implementation of this DAC when $\{d_2 d_1 d_0\}$ changes states.
- 9.4 Figure P9.4 shows the schematic of a design for a 3-bit, MOS, SA ADC. Note that this complex design requires 2^N resistors and $[2^{(N+1)} - 2]$ MOS switches. The switches are closed for $d_k = \text{HI}$ or $\bar{d}_k = \text{HI}$. $V_R = 5$ V. At the start of the SA cycle, the MSB, d_2 , is HI and d_1 and d_0 are low. Show which switches are successively closed (and opened) in a conversion cycle for $V_s = -3$ V and $+4$ V.
- 9.5 An FM VFC is shown in Figure P9.5. The one-shot output is periodic, narrow ($1 \mu\text{s}$), pulses that are counted by a digital counter (not shown) over 1 s. Each output pulse resets the integrator capacitor voltage to 0 V. Let $C = 1$ nF, $R = 10$ k Ω , $V_\phi = -5$ V.
- Plot and dimension several cycles of $v_i(t)$ for $V_s = 1$ V DC and $V_B = 0$.

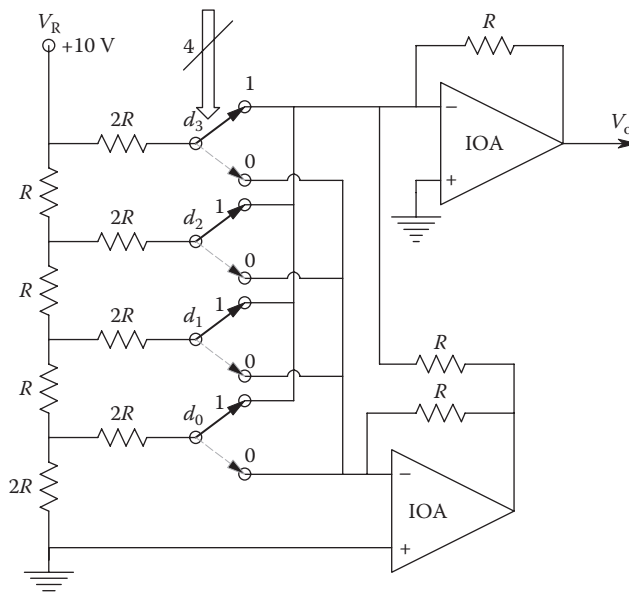


FIGURE P9.2

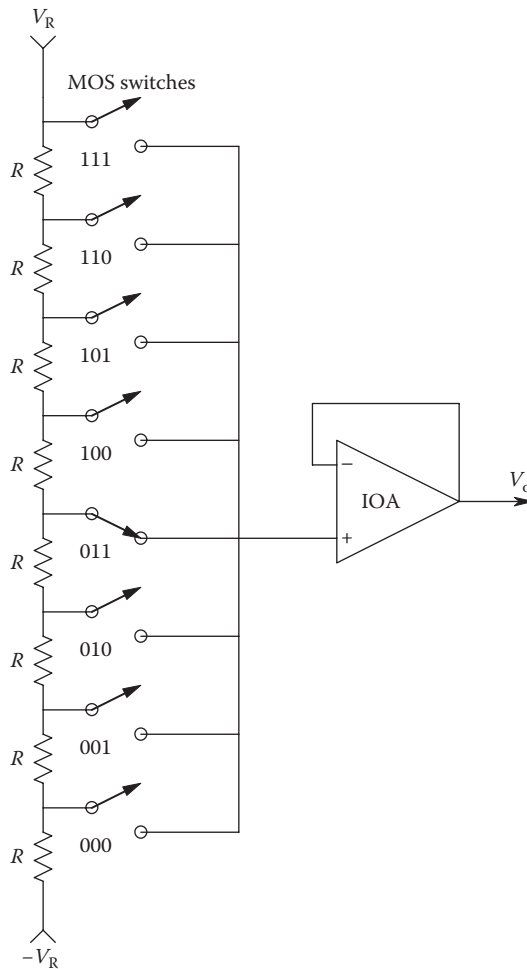


FIGURE P9.3

- B. Let $V_B = +5\text{ V}$ and plot and dimension the output frequency of the one-shot vs. $0 \leq V_s \leq +10\text{ V}$. Do not neglect the $1\ \mu\text{s}$ width of the output/reset pulses. Is $f = g(V_s)$ linear? What is $\Delta f/\Delta V_s$ at $V_s = +5\text{ V}$?
 - C. Now with $V_B = 0$, determine the counter enable time, T_c , such that when $V_s = +10\text{ V}$, the total count will be decimal 10^4 .
- 9.6 A four-bit DAC using MOS-switched, NPN, BJT current sources is shown in Figure P9.6. Derive an expression for $V_o = f[(d_3, d_2, d_1, d_0)$ and other system parameters]. Let $R = 10.1\text{ k}\Omega$, $R_E = 14.4\text{ k}\Omega$, BJT $\beta = 100$, $V_R = +5\text{ V}$, $-V_{EE} = -10\text{ V}$.
- 9.7 A unipolar, DSI ADC is used in a DC digital voltmeter. Figure P9.7 illustrates the system. At the convert command, the control logic resets the integrator and counter, then at local time $t = 0$ allows the integrator to begin integrating the positive DC input voltage, V_x . At time $t = 2^N T_c$, the counter is reset to zero. It is easy to see that at this time, $V_2 = V_x 2^N T_c / RC$ volts. The input switch is now connected to $-V_R$. The integrator now integrates $-V_R$ and ramps up to 0. When $V_2 = 0$, the comparator output Q

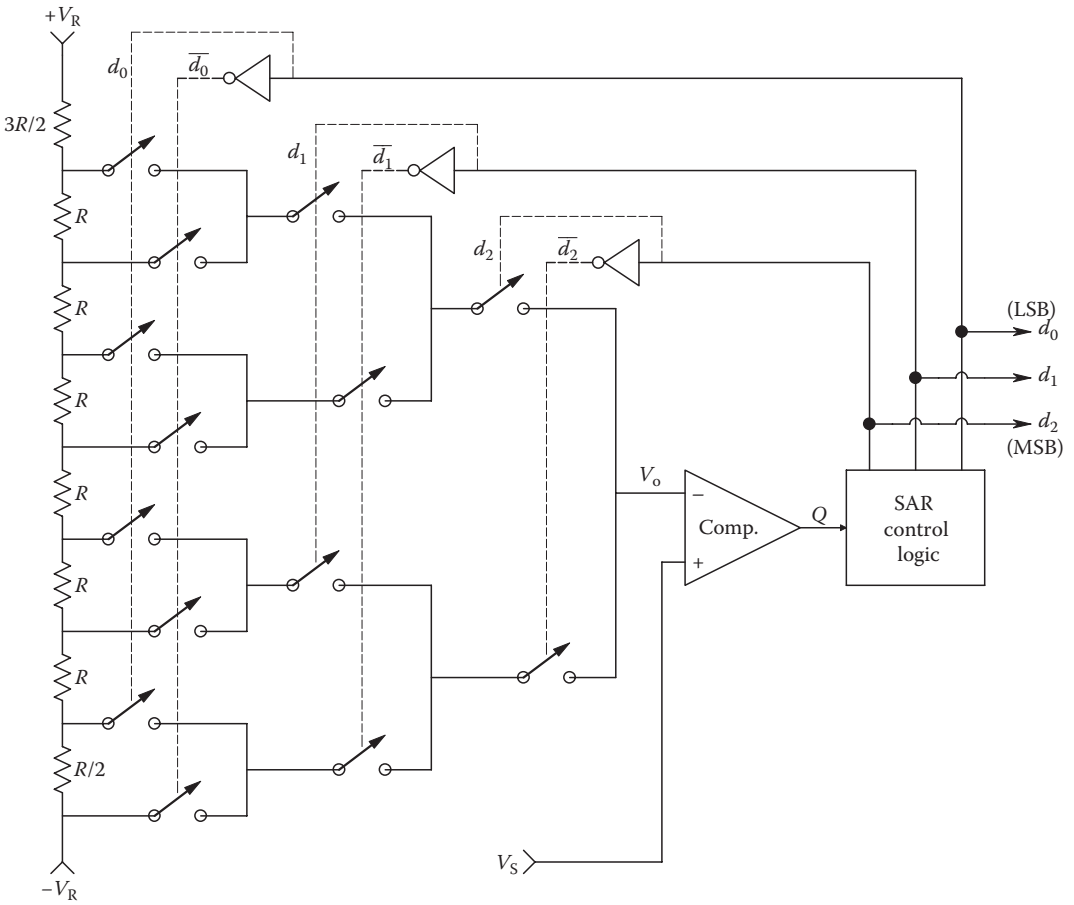


FIGURE P9.4

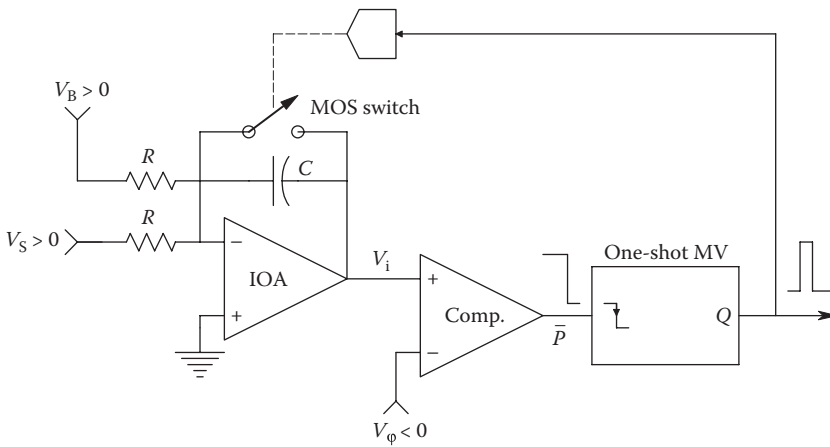


FIGURE P9.5

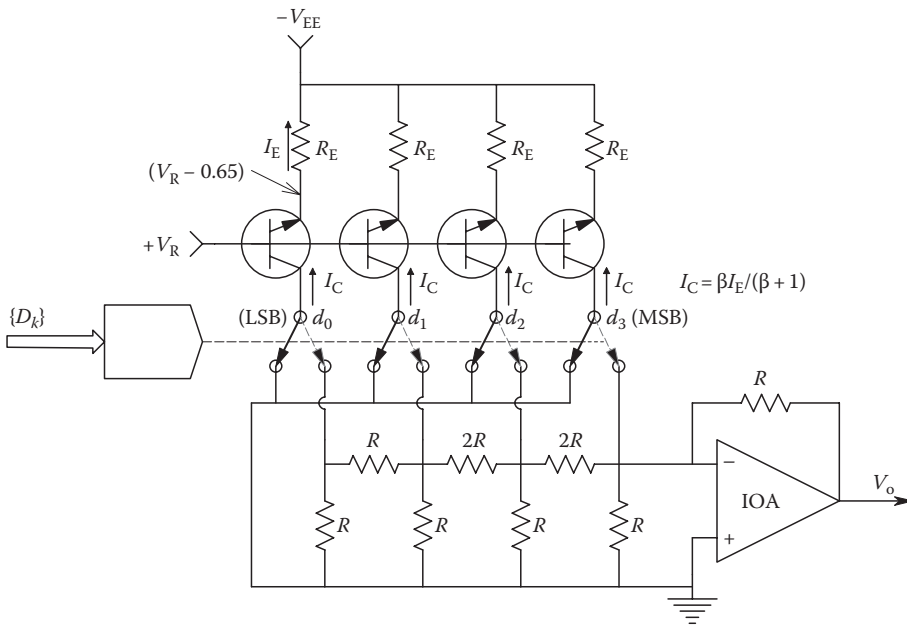


FIGURE P9.6

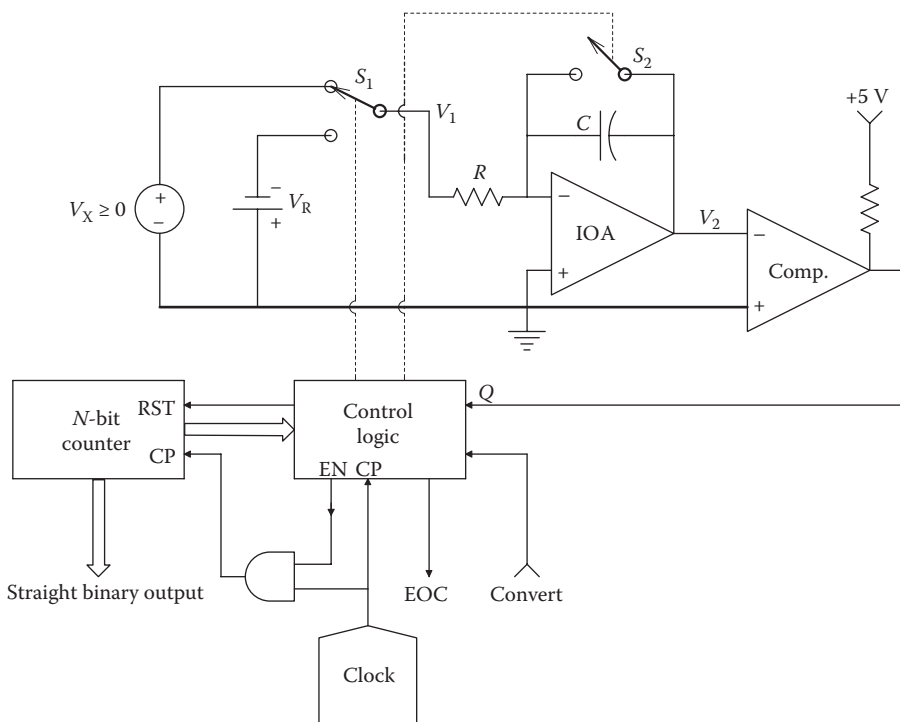


FIGURE P9.7

goes from HI to LO, latching the counter count, M , to the output. Considering V_2 , it is clear that for one conversion cycle

$$\frac{V_X 2^N T_c}{RC} = \frac{V_R M T_c}{RC}$$

or $M = V_X 2^N / V_R$. T_c is the clock period.

Design the system (choose V_R , R , C , T_c) such that when $V_X = +10$ V, $M = 1000$ decimal. The total conversion time $(2^N + M)T_c$ should not exceed 1 s. Let $N = 10$.

- 9.8 Figure P9.8 illustrates a bipolar, DSI ADC for DC voltages. Note that $V_0 = -\frac{1}{2}(V_R + V_X)$. At the convert command, the capacitor voltage is set to zero by switch S_2 and the N -bit counter is set to zero. At local $t = 0$, S_2 opens, and S_1 connects the integrator to the negative voltage, V_0 , for a fixed interval, $T_1 = 2^N T_c$ seconds. V_2 ramps positive causing $Q = \text{LO}$. At T_1 , the integrator output equals

$$V_2(T_1) = + \frac{(V_R + V_X) 2^N T_c}{RC} \text{ V.}$$

When the counter reaches 2^N at T_1 , the counter is zeroed and S_1 connects the integrator to $+V_R$. The integrator output now ramps down until $V_2 = 0$. Then the comparator output Q goes HI, causing the AND gate to block the clock pulses. Also when

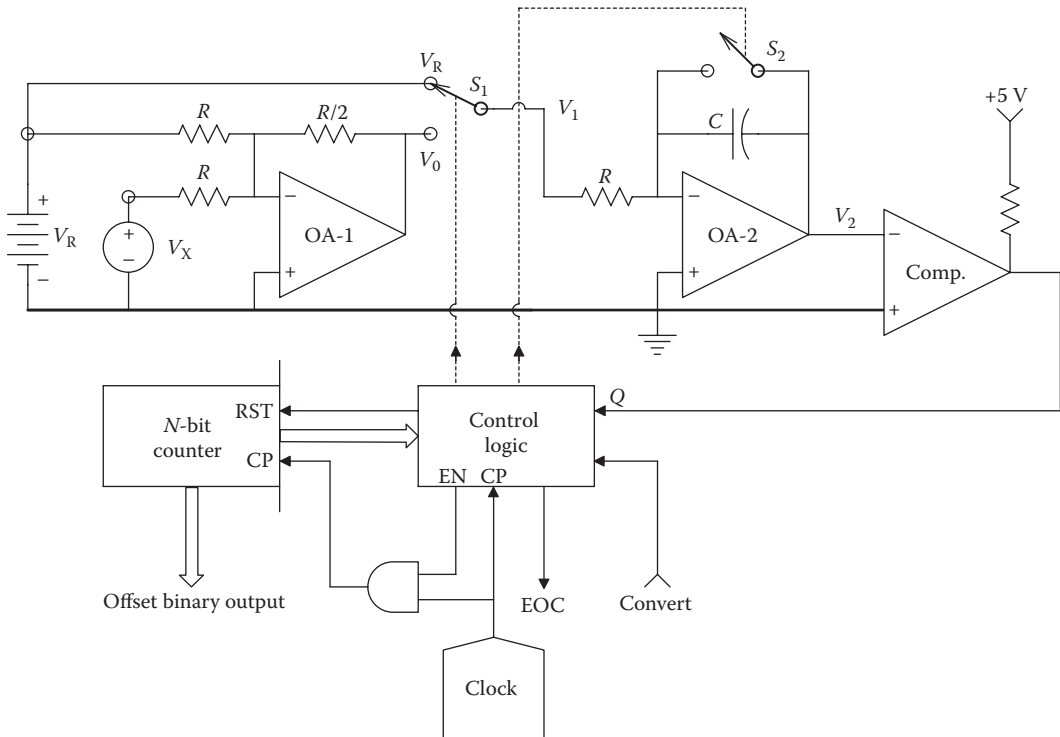


FIGURE P9.8

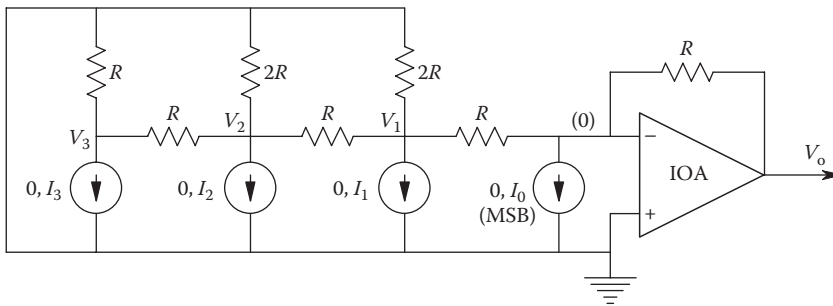


FIGURE P9.9

- $Q \rightarrow \text{HI}$, the counter output, M , is latched to the output and the counter is reset to await the next convert command. The time it takes $V_2(T_1)$ to ramp down to zero is $T_2 = MT_c$ s.
- A. Derive an expression for $M = f(N, V_x, V_R)$.
 - B. What kind of binary output code is produced by this ADC?
- 9.9 Consider the 4-bit resistive ladder DAC shown in Figure P9.9.
- A. Use superposition to find an expression for $V_o = f(I_0, I_1, I_2, I_3)$. Let $R = 1 \text{ k}\Omega$, $I_0 = 1 \text{ mA DC}$.
 - B. What values should I_1, I_2 and I_3 have to make a binary DAC? Find the maximum V_o .

10

Introduction to Digital Signal Conditioning in Instrumentation

10.1 Introduction

As we have seen in the preceding chapter, it is common practice to convey the conditioned analog output of a sensor to an analog-to-digital converter and then pass that sampled data to a digital computer for further processing, display, and storage. In this chapter, we introduce some of the basic DSP operations done on sampled data and describe how they are accomplished. They include, but are not limited to, *smoothing or low-pass filtering, integration, differentiation, computation of the rms value of a data sequence, notch or band-reject filtering* (to eliminate coherent interference in the signal bandwidth), *computation of the discrete Fourier transform (DFT) of a data sequence* (to characterize it in the [discrete] frequency domain), and *interpolation and extrapolation of data sequences*. Some of the DSP algorithms we have just mentioned exist as options in specialized data acquisition software packages, and unless an instrumentation engineer is developing a system from scratch, there is often no need to write custom software for a specific, dedicated DSP task; one simply uses a packaged DSP routine on an input sequence to obtain the conditioned output sequence.

10.2 Digital Filters and the z-Transform

Synchronous operations on sampled data can be described in the time domain as difference equations or in the frequency domain using transfer functions in the complex variable, z . To illustrate how discrete DSP operations are described using the z -transform, we first consider a well-behaved, continuous analog signal, $x(t)$, defined for $0 \leq t \leq \infty$. For example, let $x(t) = Ae^{-bt}$. As we have seen in Chapter 9, sampling $x(t)$ is equivalent to the impulse modulation of $x(t)$. We can write the sampled $x(t)$, $x^*(t)$, in a more compact form, where T is the sample period, n is the sample number, and $x(nT)$ is the area of the n th impulse:

$$x^*(t) = \sum_{n=0}^{\infty} x(nT)\delta(t - nT). \quad (10.1)$$

If we take the Laplace transform of $x^*(t)$, we get

$$X^*(s) = \sum_{n=0}^{\infty} x(nT) e^{-snT}, \quad (10.2)$$

where

$x(nT)$ is the value of $x(t)$ at $t = nT$

s is the Laplace complex variable, $s \equiv \sigma + j\omega$

Here, $x(nT)$ is treated as the area of the n th delta function, and the Laplace transform of the n th delta function is simply e^{-snT} . Equation 10.2 is easily changed to the open-sequence form of the z -transformed $x(t)$ by letting the complex variable, z , be defined as

$$z \equiv e^{sT}. \quad (10.3)$$

Hence, we can write

$$X(z) = \sum_{n=0}^{\infty} x(nT) z^{-n}. \quad (10.4)$$

For the exponential example, we can write

$$X(z) = A \left[1 + e^{-bT} z^{-1} + e^{-2bT} z^{-2} + e^{-3bT} z^{-3} + \dots \right]. \quad (10.5)$$

An infinite power series of this form can be expressed in closed form as the rational function:

$$X(z) = \frac{A}{1 - e^{-bT} z^{-1}} = \frac{Az}{z - e^{-bT}}. \quad (10.6)$$

A general, closed-form definition of the z -transform of any Laplace transformable $f(t)$ can be derived from a consideration of the impulse modulation process, which is multiplication of $f(t)$ in the time domain with a periodic, unit delta function train, $P_T(t)$, beginning at $t = 0$. This multiplication gives $f^*(t)$. In the frequency domain, $F^*(s)$ is found by the complex convolution of $F(s)$ with the Laplace transform of the pulse train, $P_T(s)$. $P_T(s)$ can be written as

$$P_T(s) = 1 + e^{-sT} + e^{-2sT} + e^{-3sT} + \dots = \sum_{n=0}^{\infty} e^{-nsT}. \quad (10.7)$$

In closed form, $P_T(s)$ can be shown to be

$$P_T(s) = \frac{1}{1 - e^{-sT}} \rightarrow \frac{1}{1 - z^{-1}} = \frac{z}{z - 1}. \quad (10.8)$$

Now, by complex convolution,

$$F * (e^{-sT}) = \frac{1}{2\pi j} \oint \frac{F(p)}{1 - e^{-sT} e^{pT}} dp. \tag{10.9}$$

We use the aforementioned definition of z and treat the complex variable, z , as a constant in the convolution integral:

$$F_Z(z) = \frac{1}{2\pi j} \oint \frac{zF(p)}{z - e^{pT}} dp. \tag{10.10}$$

$F(p)$ is the Laplace transform of $f(t)$ with $s = p$. Thus, the *contour integral*, Equation 10.10, can be used to find the z -transform of any $f(t)$ having a Laplace transform, $F(s)$. Table 10.1 shows the Laplace and z -transforms of some typical deterministic time functions defined for $t \geq 0$.

A *digital filter* is a linear, discrete operator that alters a periodic input number sequence, $\{x(k)\}$, producing a different output number sequence, $\{y(k)\}$. Digital filters are generally described as rational polynomials in the complex variable, z . They may be subdivided into two major categories: (1) *Recursive filters*, in which the present filter output, $y(t)$, depends not only on the present and past *input* samples, $[x(0), x(T), x(2T), \dots, x(kT), \dots]$, but also on filter *output* samples, $[y(0), y(T), y(2T), \dots, y(kT), \dots]$; and (2) *Moving average, transversal, or finite duration impulse response (FIR) filters*. FIR filters use only the present and past values of $x^*(t)$; their impulse responses are truly finite, that is, after M sample periods following a 1 input at $t = 0$ only, their output is identically zero. The impulse response of a recursive

TABLE 10.1
Some Common Time Functions and Their Laplace and z -Transforms

Time Function ($t \geq 0$)	Laplace Transform	z -Transform
$\delta(t)$ (unit impulse)	1	1
$\delta(t - nT)$ (delayed impulse)	e^{-snT}	z^{-n}
$U(t)$ (unit step)	$1/s$	$\frac{z}{z-1}$
$tU(t)$ (unit ramp)	$1/s^2$	$\frac{Tz}{(z-1)^2}$
$e^{-bt}, t \geq 0$	$\frac{1}{s+b}$	$\frac{z}{z-e^{-bT}}$
$te^{-bt}, t \geq 0$	$\frac{1}{(s+b)^2}$	$\frac{Tze^{-bT}}{(z-e^{-bT})^2}$
$\sin(\omega t), t \geq 0$	$\frac{\omega}{s^2 + \omega^2}$	$\frac{z \sin(\omega T)}{z^2 - 2z \cos(\omega T) + 1}$
$\cos(\omega t), t \geq 0$	$\frac{s}{s^2 + \omega^2}$	$\frac{z^2 - z \cos(\omega T)}{z^2 - 2z \cos(\omega T) + 1}$
$e^{-bt} \sin(\omega t), t \geq 0$	$\frac{\omega}{(s+b)^2 + \omega^2}$	$\frac{e^{-bT} z \sin(\omega T)}{z^2 - 2ze^{-bT} \cos(\omega T) + e^{-2bT}}$
$e^{-bt} \cos(\omega t), t \geq 0$	$\frac{s+b}{(s+b)^2 + \omega^2}$	$\frac{z^2 - ze^{-bT} \cos(\omega T)}{z^2 - 2ze^{-bT} \cos(\omega T) + e^{-2bT}}$

filter generally approaches zero asymptotically as k approaches infinity. In the frequency (z) domain, the z -transformed output of a linear digital filter can be written as the product of the filter's transfer function, $G_Z(z)$, and the z -transform of the filter's input, that is,

$$Y(z) = X(z)G_Z(z). \tag{10.11}$$

In the time domain, the k th output sequence member from a digital filter can be found by the process of real, discrete convolution. This can be written as

$$y(kT) = y_k = \sum_{i=-\infty}^{\infty} g_i x_{k-i}. \tag{10.12}$$

Here, $\{g_i\}$ are the elements of the discrete filter's unit impulse response, and $\{x_{k-i}\}$ are the elements of $\{x_i\}$ reversed in time and shifted k sample periods. The real convolution process is illustrated using a simple example in Figure 10.1.

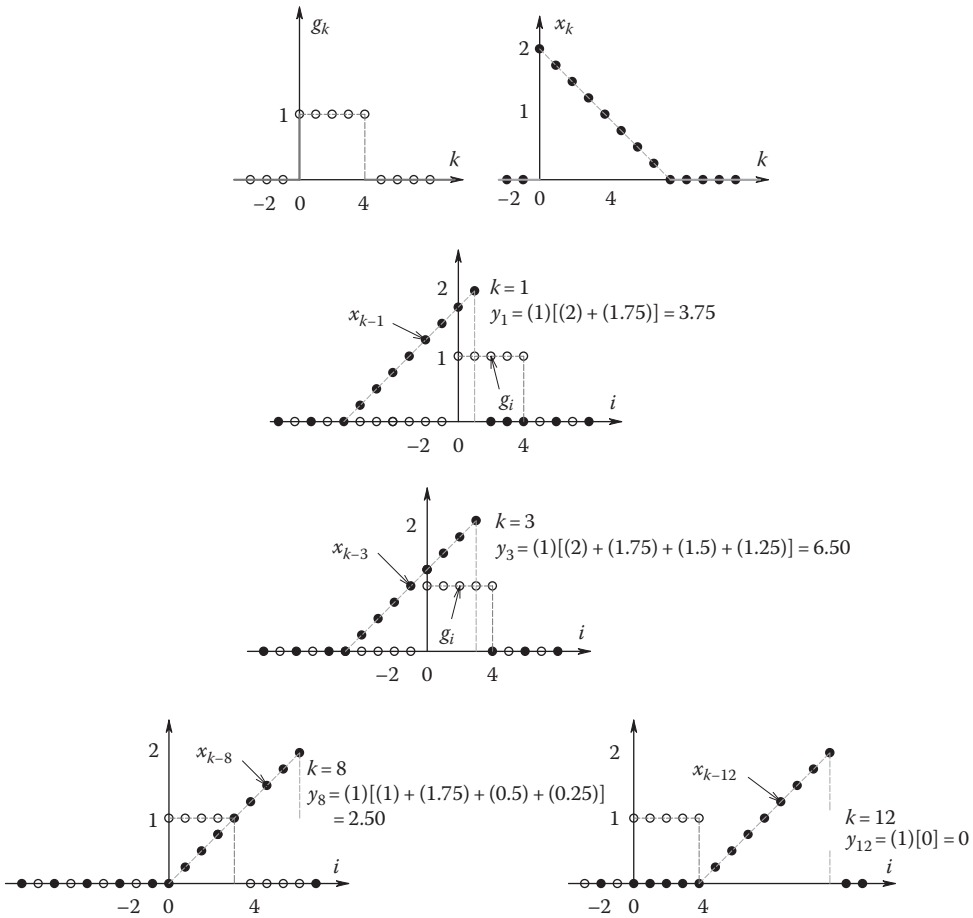


FIGURE 10.1 Illustration of the process of discrete, real convolution. A sampled rectangular pulse, $g_{k'}$, and a sampled finite ramp waveform, $x_{k'}$, are shown on the top row. In the next three rows, the steps in the real convolution operation given by Equation 10.12 are shown. g_k becomes $g_{i'}$ and x_k becomes $x_{k-i'}$ that is, $x_{k-i'}$ is x_k reversed in time i and shifted by constant k .

Recursive filters used in linear DSP operations can be expressed as rational polynomials in the complex variable, z . For example, $G_Z(z)$ can be written in a general, unfactored, polynomial form as

$$\frac{Y(z)}{X(z)} = G_Z(z) = \frac{a_m z^m + a_{m-1} z^{m-1} + \dots + a_1 z^1 + a_0 z^0}{b_n z^n + b_{n-1} z^{n-1} + \dots + b_1 z^1 + b_0 z^0}, \quad n \geq m. \quad (10.13)$$

Both the numerator and denominator polynomials may have roots at the origin of the z -plane; their roots may occur in complex-conjugate pairs or lie on the negative or positive real axis in the complex z -plane. The coefficients, $\{a_k\}$ and $\{b_k\}$, are real numbers. The poles (denominator roots) of a stable $G(z)$ must lie inside the unit circle in the z -plane. Recursive filters are also known as infinite impulse response (IIR) filters because although their impulse response magnitudes die out as $k \rightarrow \infty$, the terms go on forever.

To appreciate how a recursive $G_Z(z)$ can be implemented on a computer in real time, we change $G_Z(z)$ to be in powers of $1/z$ or e^{-sT} , the delay operator. Simple algebra gives

$$\frac{Y(z)}{X(z)} = G_Z(z) = \frac{(a_m + a_{m-1} z^{-1} + \dots + a_1 z^{-m+1} + a_0 z^{-m}) z^{-(n-m)}}{(b_n + b_{n-1} z^{-1} + \dots + b_1 z^{-n+1} + b_0 z^{-n})}, \quad n \geq m. \quad (10.14)$$

From Equation 10.14, it is now easy to write

$$Y(z) (b_n + b_{n-1} z^{-1} + \dots + b_1 z^{-n+1} + b_0 z^{-n}) = X(z) (a_m z^{-(n-m)} + a_{m-1} z^{-(n-m+1)} + \dots + a_1 z^{-(n-1)} + a_0 z^{-n}). \quad (10.15)$$

To simplify the notation, let the k th past sample in $x^*(t)$, $x(t - kT)$ be written as x_{-k} and the p th past sample in $y^*(t)$ be written as y_{-p} . Now we can write Equation 10.15 in the time domain as

$$b_n y + b_{n-1} y_{-1} + \dots + b_1 y_{-n+1} + b_0 y_{-n} = a_m x_{-n+m} + a_{m-1} x_{-n+m-1} + \dots + a_1 x_{-n+1} + a_0 x_{-n}. \quad (10.16)$$

Equation 10.16 can easily be solved for the present filter output, y , in terms of present and past weighted output and input samples. To illustrate this process, we use a simple third-order example:

$$G_Z(z) = \frac{a_2 z^2 + a_1 z^1 + a_0}{b_3 z^3 + b_2 z^2 + b_1 z^1 + b_0}, \quad n - m = 1. \quad (10.17)$$

Following the procedure described earlier, the present output of $G_Z(z)$ can be written as

$$y = \left(\frac{1}{b_3} \right) (a_2 x_{-1} + a_1 x_{-2} + a_0 x_{-3} - b_2 y_{-1} - b_1 y_{-2} - b_0 y_{-3}). \quad (10.18)$$

Implementation of this recursive filter example is shown in Figure 10.2. Note that in this particular example, no *present* value of x is used in computing the present value of y .

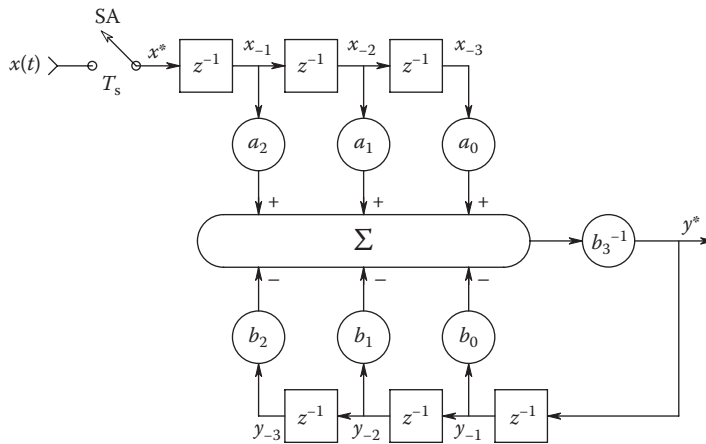


FIGURE 10.2

Block diagram showing the implementation of a third-order recursive filter. z^{-1} represents e^{-sT_s} , the unit delay operator. T_s is the system sampling period; SA is the ideal sampler.

The impulse response of a digital filter, $\{g_k\}$, is simply its output number sequence, given an input $x = 1$ at $t = 0$. $\{g_k\}$ can also be found by long division of the numerator of the filter's $G(z)$ by its denominator. This division generally gives a sequence of terms with coefficients of the form, z^{-k} . In the time domain, this is

$$\{g_k\} = \sum_{k=0}^{\infty} g_k \delta(t - kT). \tag{10.19}$$

For stable transfer functions, $g_k \rightarrow 0$ as $k \rightarrow \infty$.

An n th-order FIR digital filter's transfer function can be written in the general form:

$$\begin{aligned} \frac{Y(z)}{X(z)} &= \frac{a_n z^n + a_{n-1} z^{n-1} + \dots + a_1 z^1 + a_0}{z^n}, \quad m = n, b_n = 1 \\ &= a_n + a_{n-1} z^{-1} + \dots + a_1 z^{-n+1} + a_0 z^{-n}. \end{aligned} \tag{10.20}$$

By taking the inverse z -transform of Equation 10.20, we obtain the difference equation

$$y = a_n x + a_{n-1} x_{-1} + \dots + a_1 x_{-n+1} + a_0 x_{-n}. \tag{10.21}$$

It is easy to see that if a 1 is input at $t = 0$, the FIR filter's impulse response will have, including the present sample, $M = (n + 1)$ terms.

Before going on to consider some typical applications of digital filters, we remind the reader that in using DSP routines, we generally assume that the analog input signal, $x(t)$, is bandwidth limited by an antialiasing LPF so that it contains no significant spectral energy above the Nyquist frequency (one-half the sampling frequency).

10.3 Some Simple DSP Algorithms

The first, simple DSP example we consider is the *FIR, moving average, smoothing*, or LPF. This type of filter is often used to *clean up* or remove high-frequency noise accompanying a slowly varying or DC signal. Moving average, FIR filters generally have an odd number of polynomial terms and have a unity DC response. For example, a three-term, Hanning LPF is described in the time domain by

$$y(nT) = \left(\frac{1}{4}\right) [x(nT) + 2x(nT - T) + x(nT - 2T)]$$

or

$$y = \left(\frac{1}{4}\right) (x + 2x_{-1} + x_{-2}). \quad (10.22)$$

We can write the z -transform for this filter as

$$G_Z(z) = \frac{Y(z)}{X(z)} = 0.25(1 + 2z^{-1} + z^{-2}) = \frac{0.25(z^2 + 0.5z + 1)}{z^2}. \quad (10.23)$$

The numerator of $G_Z(z)$ has two real roots (zeros) at $z = -1$. A z -plane plot of the roots of $G_Z(z)$ is shown in Figure 10.3A. To examine the frequency response of this $G_Z(z)$, it is convenient to substitute $z = e^{sT} = e^{j\omega T}$ into Equation 10.23 and vary ω :

$$\mathbf{G}_Z(j\omega T) = 0.25(1 + 2e^{-j\omega T} + e^{-j2\omega T}) = 0.5[1 + \cos(\omega T)]e^{-j\omega T}. \quad (10.24)$$

The radian sampling frequency is $\omega_s = 2\pi/T$, and the Nyquist frequency is $\omega_N \equiv \omega_s/2 = \pi/T$. Since we assume $x(t)$ is analog antialias filtered before sampling, we need only to consider $0 \leq \omega \leq \omega_N = \pi/T$. The magnitude and phase of $\mathbf{G}_Z(j\omega T)$ are plotted in Figure 10.3B. Note that the phase is linear in ω , and the DC magnitude response of the Hanning filter is unity and decreases to zero at $\omega = \omega_N$ and then increases again.

Many other higher-order, FIR low-pass or smoothing filters can be designed. In general, the higher the order of the FIR filter, the sharper the attenuation can be made. A number of criteria exist for the design of FIR smoothing filters. For example, Tompkins and Webster (1981) derived a five-term, FIR, LPF based on a least squared error criterion to fit the five data samples used in the filter with the best parabola. This gave a time-domain filter function of

$$y = \left(\frac{1}{35}\right) (-3x + 12x_{-1} + 17x_{-2} + 12x_{-3} - 3x_{-4}). \quad (10.25)$$

In the z -domain, this FIR algorithm becomes

$$\frac{Y(z)}{X(z)} = G_Z(z) = \left(\frac{1}{35}\right) (-3 + 12z^{-1} + 17z^{-2} + 12z^{-3} - 3z^{-4}). \quad (10.26)$$

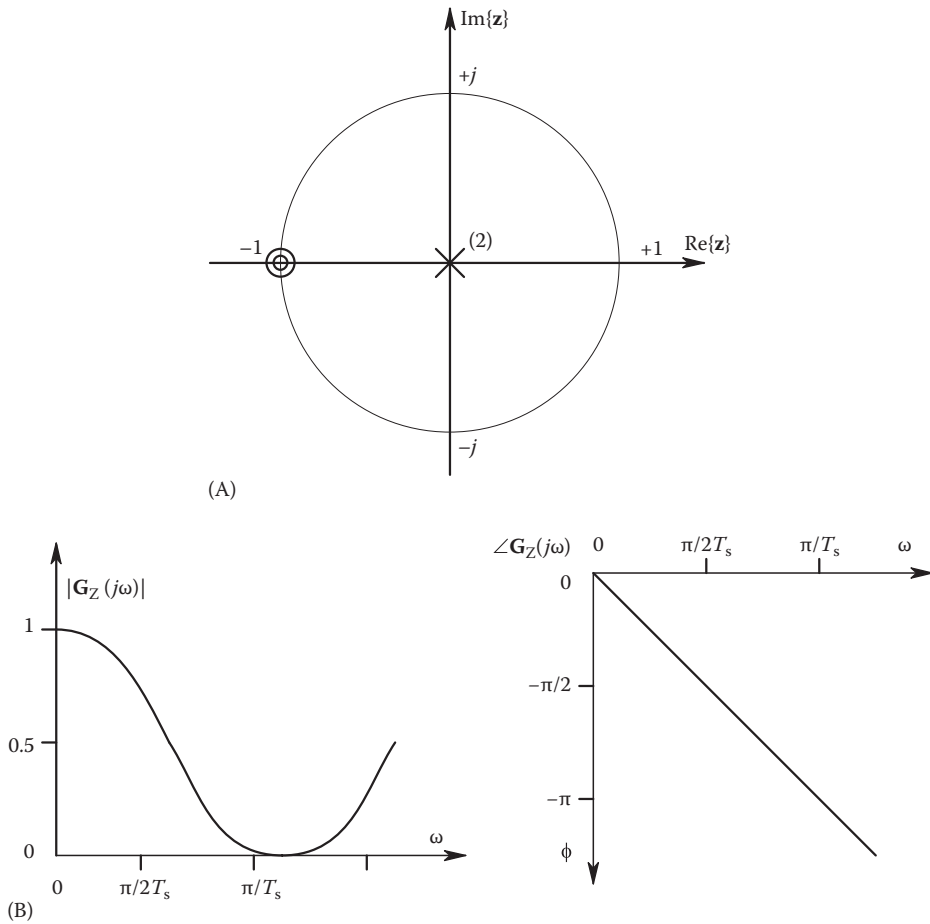


FIGURE 10.3 (A) A z-plane plot of the poles and zeros of the Hanning LPF of Equation 10.23. (B) The magnitude and phase of $G_z(j\omega T_s)$ of the Hanning filter.

The frequency response magnitude of this $G_z(j\omega T)$ and several other FIR LPFs is illustrated in Tompkins and Webster (1981). These authors also derived a five-point FIR LPF that satisfies the following criteria:

$$G_z(j0) = 1 \tag{10.27a}$$

$$\frac{dG_z(j0)}{d\omega T} = 0 \tag{10.27b}$$

$$\frac{d^2G_z(j0)}{d(\omega T)^2} = 0. \tag{10.27c}$$

That is, the filter's frequency response is maximally flat at DC. This filter was shown to be implemented by the difference equation

$$y = \left(\frac{1}{16} \right) (-x + 4x_{-1} + 10x_{-2} + 4x_{-3} - x_{-4}). \quad (10.28)$$

Its frequency response magnitude is

$$\mathbf{G}_Z(j\omega T) = \frac{5}{8} + \frac{4}{8} \cos(\omega T) - \frac{1}{8} \cos(2\omega T). \quad (10.29)$$

Clearly, $\mathbf{G}_Z(j0) = 1$, and $\mathbf{G}_Z(j\pi) = 0$ for this filter.

Often, in instrumentation applications, we wish to eliminate a specific, coherent frequency, ω_i , from a low-frequency signal. We assume that $\omega_i < \omega_N$. To create a *digital notch filter*, we need to place a pair of complex-conjugate zeros of $G(z)$ on the unit circle in the z -plane. This can be done most simply with a quadratic FIR filter of the form

$$G_Z(z) = 0.5(1 + z^{-2}), \quad (10.30)$$

which is implemented as

$$y = 0.5(x + x_{-2}). \quad (10.31)$$

Substituting $e^{j\omega T}$ for z in Equation 10.30 gives us the frequency response of this simple notch filter:

$$\mathbf{G}_Z(j\omega T) = \cos(\omega T)e^{-j\omega T}. \quad (10.32)$$

It is easy to see that $\mathbf{G}_Z(j0) = \mathbf{G}_Z(j\omega_N T) = 1$, and $\mathbf{G}_Z(j(\omega_N/2)T) = 0$, that is, the notch frequency, ω_i , is 1/2 the Nyquist frequency in this example. Attenuation around the notch frequency is certainly not sharp for this filter. Tuning the notch is accomplished by adjusting the sampling period, T .

To obtain a sharp, high- Q notch, we can use a recursive notch filter design, adapted from Widrow and Stearns (1985):

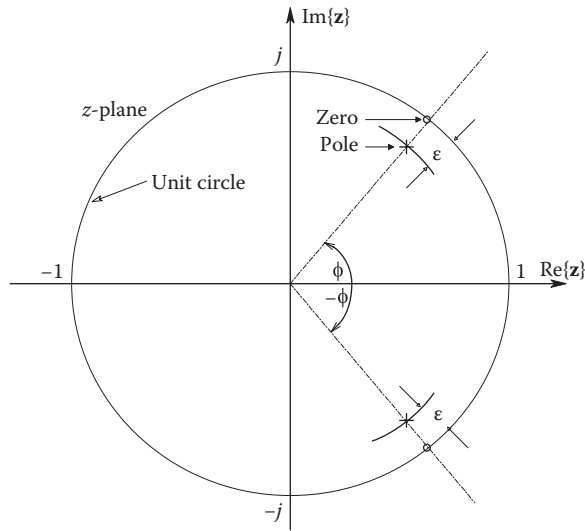
$$F_Z(z) = \frac{z^2 - z2\cos(\phi) + 1}{z^2 - z2\eta\cos(\phi) + \eta^2}, \quad (10.33)$$

where $\eta = 1 - \varepsilon$, $0 < \varepsilon \ll 1$, and the angle $\phi = (\omega_i/\omega_N)\pi$ is shown on the unit circle in Figure 10.4. ω_i , as in the previous example, is the frequency of the interfering sinusoid that we wish to eliminate.

Consideration of the numerator of $F(z)$ shows that the filter has complex-conjugate zeros on the unit circle at

$$z_{01} = \cos(\phi) + j\sin(\phi) \quad (10.34a)$$

$$z_{02} = \cos(\phi) - j\sin(\phi). \quad (10.34b)$$

**FIGURE 10.4**

A z -plane plot of the poles and zeros of the recursive quadratic notch filter of Equation 10.33 given by Widrow and Stearns (1985).

The poles of $F_Z(z)$ are easily shown to be just inside the unit circle, lying at $\pm\phi$ with respect to the real axis. Factoring the denominator yields

$$z_{P1} = \eta(z_{01}) \quad (10.35a)$$

$$z_{P2} = \eta(z_{02}). \quad (10.35b)$$

Following Widrow and Stearns, it can be shown that the recursive notch filter's Q is given by

$$Q = \frac{\text{Center frequency}}{-3 \text{ dB bandwidth}} = \frac{\omega_i T}{2(1 - \eta)}. \quad (10.36)$$

(The higher the Q , the sharper the notch.) The Widrow and Stearns adaptive filter can be tuned to any ω_i by calculating the numerical value of

$$\cos(\phi) \equiv \cos \left[\left(\frac{\omega_i}{\omega_N} \right) \pi \right], \quad (10.37)$$

and then using this value in the recursion formula for $y(nT)$:

$$y = \eta \left[x + (y_{-1} - x_{-1}) 2 \cos(\phi) + x_{-2} - \eta y_{-2} \right]. \quad (10.38)$$

Often, in signal processing applications, we are interested in estimating the points in time when a bandwidth-limited signal reaches its maxima or minima, that is, the points in

time where its first derivative goes to zero. The simplest form of *digital differentiator* is the FIR, two-point difference equation:

$$y = \left(\frac{1}{T}\right)(x - x_{-1}). \tag{10.39}$$

Figure 10.5A illustrates how Equation 10.39 estimates the slope of $x(t)$ by a straight-line approximation using the present and first past sample of $x(t)$. The digital transfer function for this simple routine is

$$D_z(z) = \frac{Y(z)}{X(z)} = \frac{z - 1}{Tz}. \tag{10.40}$$

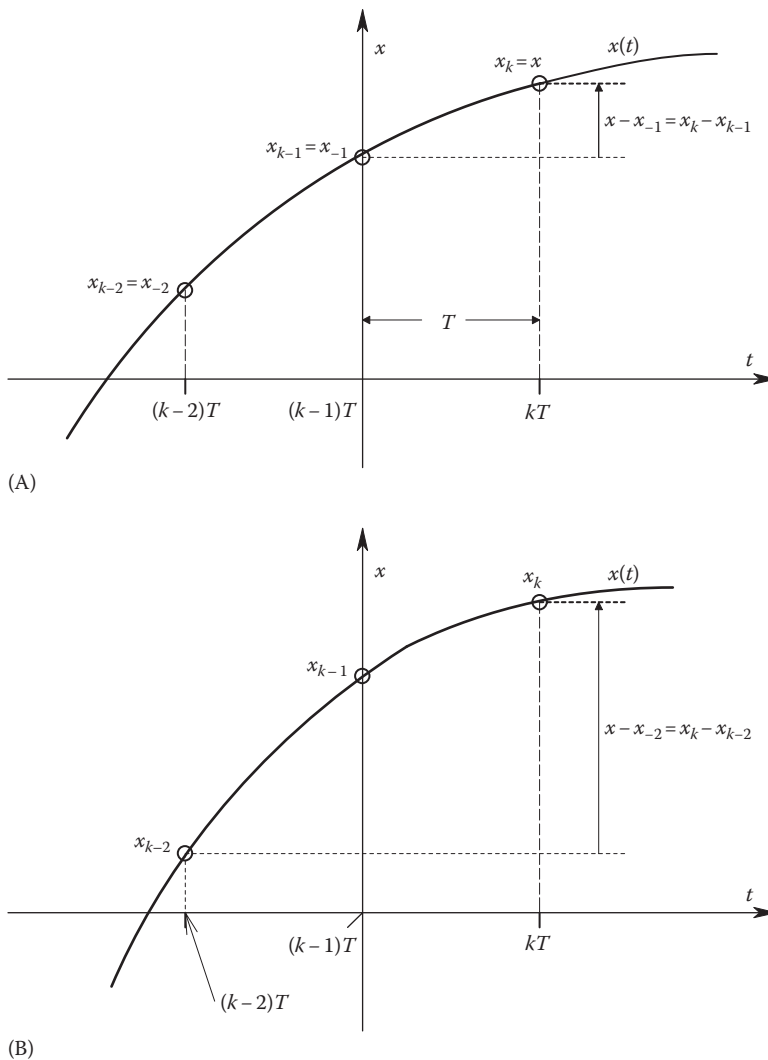


FIGURE 10.5 (A) Illustration of the simple, two-point, digital differentiator operating on x^* . (B) Illustration of the three-point, central difference digital differentiator given by Equation 10.47.

$D_Z(z)$ has a zero at $z = +1$ and a pole at the origin of the z -plane. By substituting $e^{j\omega T}$ for z in $D_Z(z)$, we can find the frequency response of the simple differentiator over $0 \leq \omega \leq \omega_N$:

$$\mathbf{D}_Z(j\omega T) = \left(\frac{2}{T}\right) \sin\left(\frac{\omega T}{2}\right) \exp\left[j\left(\frac{\pi}{2} - \frac{\omega T}{2}\right)\right]. \quad (10.41)$$

For low ω such that $\sin(\omega T/2) \approx \omega T/2$, in radians,

$$\mathbf{D}_Z(j\omega T) \cong \omega e^{j\pi/2}, \quad (10.42)$$

which is the frequency response function of an ideal (analog) differentiator. As ω increases, this $|\mathbf{D}_Z(j\omega T)|$ reaches a peak at the Nyquist frequency and then decreases.

Other digital differentiation routines are sometimes used. All digital differentiators depart from ideality as ω approaches the Nyquist frequency. In some cases, it is desirable to have the differentiator frequency response reach a peak and roll off at frequencies higher than $\omega_N/2$ to minimize the effects of noise on $x^*(t)$. These routines include the following (Tompkins and Webster 1981):

$$y = \left(\frac{1}{2T}\right)(x - x_{-2}) \quad (\text{three-point central difference}) \quad (10.43)$$

$$y = 0.1(2x + x_{-1} - x_{-3} - 2x_{-4}) \quad (\text{five-point, least squares, parabolic fit}) \quad (10.44)$$

$$y = \left(\frac{1}{28}\right)(3x + 2x_{-1} + x_{-2} - x_{-4} - 2x_{-5} - 3x_{-6}) \quad (\text{seven-point, LSPF}) \quad (10.45)$$

$$y = \left(\frac{1}{60}\right)(4x + 3x_{-1} + 2x_{-2} + x_{-3} - x_{-5} - 2x_{-6} - 3x_{-7} - 4x_{-8}) \quad (\text{nine-point}). \quad (10.46)$$

How the *three-point central difference differentiator* estimates the slope of $x(t)$ is illustrated in Figure 10.5B. The transfer function of this differentiator is easily seen to be

$$D_Z(z) = \frac{z^2 - 1}{2Tz^2}. \quad (10.47)$$

This transfer function has zeros at $z = 1$, $z = -1$, and two poles at the origin of the z -plane. The frequency response of $D_Z(z)$ is determined by substituting $e^{j\omega T}$ for z , and using the Euler relations,

$$\mathbf{D}_Z(j\omega T) = \left(\frac{1}{T}\right) \sin(\omega T) \exp\left[j\left(\frac{\pi}{2} - \omega T\right)\right]. \quad (10.48)$$

Like the two-point differentiator, this algorithm has a nearly ideal behavior at low frequencies. Its magnitude response peaks at $\omega = \omega_N/2$ and goes to zero at $\omega = \omega_N$. Note that the gains of the two- and three-point differentiators depend on a knowledge of the sampling period, but the five- and higher-point, LS, parabolic fit differentiator gains do not depend on T .

Digital integration routines allow us to estimate the area under a curve, using sampled data. In chemometric instrumentation systems, the outputs of various types of chromatographs

and spectrographs are often integrated in order to estimate the quantity of a chemical species in a sample eluent. Another obvious application of integrating a sampled variable is to compute its mean value or rms value.

The simplest form of digital integrator is the *rectangular integrator*, the action of which is illustrated in Figure 10.6A. All digital integration routines are recursive, that is, they add a present estimate of area to the cumulative old estimate of area. That is, the present integrator output $i(nT) = i$ is given by

$$i = i_{-1} + \Delta i. \quad (10.49)$$

In the case of the rectangular integrator,

$$i = i_{-1} + Tx_{-1}. \quad (10.50)$$

In terms of a z-transform transfer function, Equation 10.50 can be written as

$$\frac{I(z)}{X(z)} = H_{ZR}(z) = \frac{T}{z-1}. \quad (10.51)$$

Substituting $e^{j\omega T}$ for z gives us the frequency response of the rectangular integrator:

$$\mathbf{H}_{ZR}(j\omega T) = \frac{T/2}{\sin(\omega T/2)} \exp\left[-j\left(\frac{\pi}{2} + \frac{\omega T}{2}\right)\right]. \quad (10.52)$$

At low frequencies, $\mathbf{H}_{ZR}(j\omega T)$ behaves like

$$\mathbf{H}_{ZR}(j\omega T) \cong \left(\frac{1}{\omega}\right) \exp\left[\frac{-j\pi}{2}\right], \quad (10.53)$$

which is the behavior of an (ideal) analog integrator. At the Nyquist frequency,

$$\mathbf{H}_{ZR}(j\omega_N T) = \left(\frac{T}{2}\right) \exp[-j\pi]. \quad (10.54)$$

Many other integration routines exist, some of them quite complex. Here, we will review two other simple integration algorithms commonly used in DSP:

$$i = i_{-1} + \left[Tx_{-1} + \left(\frac{T}{2}\right)(x - x_{-1}) \right] \quad (\text{Trapezoidal integration cf. Figure 10.6B}) \quad (10.55)$$

$$i = i_{-2} + \left(\frac{T}{3}\right)[x + 4x_{-1} + x_{-2}] \quad (\text{Simpson's 1/3 rule cf. Figure 10.6C}). \quad (10.56)$$

From these difference equations, we can easily derive the *transfer function for trapezoidal integration* in z :

$$H_{ZT}(z) = \frac{T(z+1)}{2(z-1)}. \quad (10.57)$$

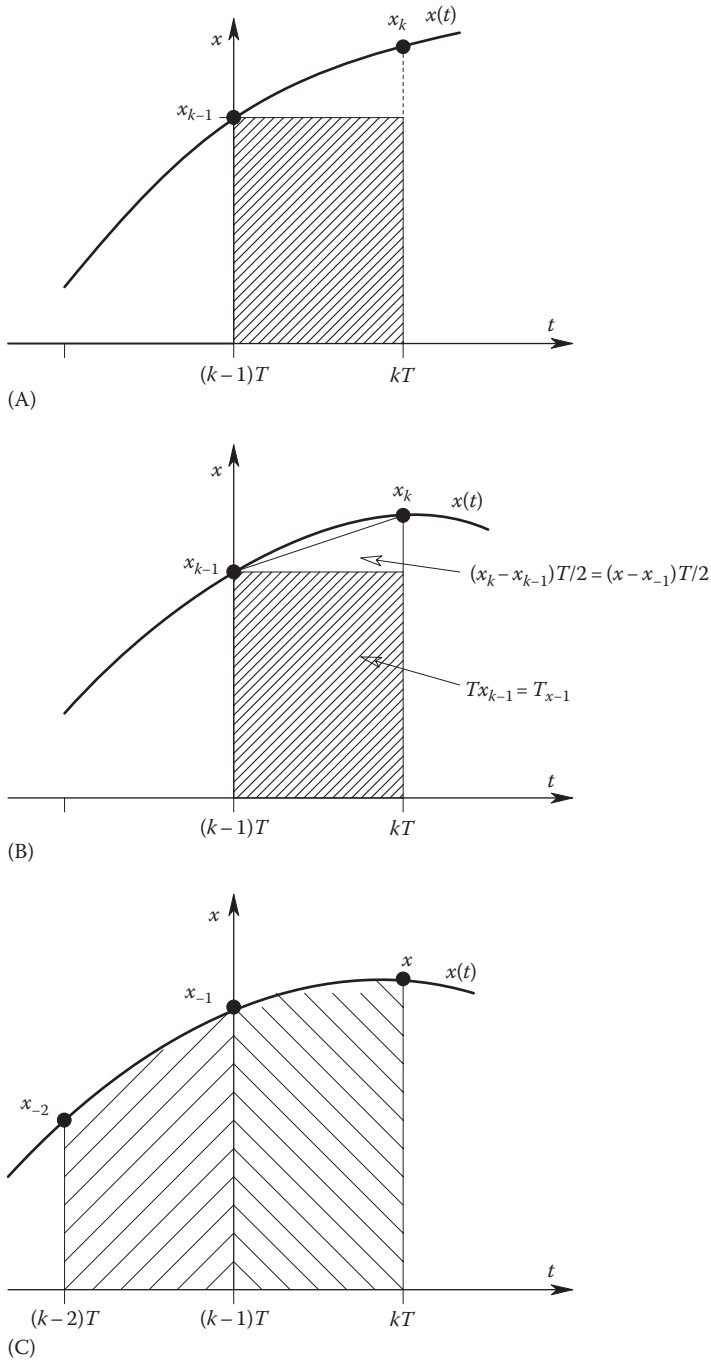
**FIGURE 10.6**

Illustration of three, common, digital integration routines in the time domain. (A) Simple rectangular integration; its transfer function is given by Equation 10.51. (B) Trapezoidal integration; see Equation 10.57. (C) Simpson's rule integration; see Equation 10.58.

For *Simpson's rule*,

$$H_{ZS}(z) = \frac{T(z^2 + 4z + 1)}{3(z^2 - 1)} \tag{10.58}$$

The frequency responses of these integrators are easily shown to be

$$\mathbf{H}_{ZT}(j\omega T) = \left(\frac{T}{2}\right) \cot\left(\frac{\omega T}{2}\right) \exp\left[\frac{-j\pi}{2}\right] \tag{10.59}$$

$$\mathbf{H}_{ZS}(j\omega T) = \frac{T[2 + \cos(\omega T)]}{3 \sin(\omega T)} \exp\left[\frac{-j\pi}{2}\right] \tag{10.60}$$

Note that the phase for both the trapezoidal and Simpson's rule integrators is that of an ideal integrator. However, their frequency response magnitudes depart from an ideal integrator at higher frequencies. Simpson's rule follows $1/\omega$ well out to $\omega = \omega_N/2$ and then reverses slope and increases sharply as ω increases above $2\omega_N/3$. Thus, Simpson's rule behaves poorly on any $x(t)$ containing noise power above $2/3$ the Nyquist frequency. The gain magnitude of the trapezoidal integrator follows the ideal frequency response curve closely up to about $\omega_N/3$ and then drops off to zero at ω_N . Thus, the trapezoidal integrator gives better accuracy when integrating noisy signals. If the PDS of $x(t)$ has little power above half the Nyquist frequency, the Simpson's integration rule may perform more accurately. The frequency response magnitude curves for the three integrators we have discussed are shown in Figure 10.7.

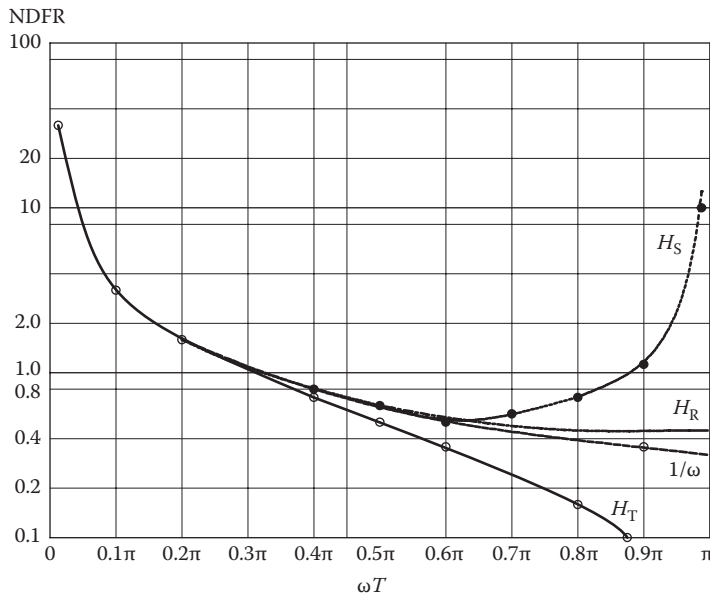


FIGURE 10.7 Normalized frequency response magnitudes of the three integrators shown in Figure 10.6. The frequency response magnitude of an ideal (analog) integrator, $1/\omega$, is shown for comparison. Note: H_R = rectangular integrator, H_T = trapezoidal integrator, and H_S = Simpson's rule integrator.

10.4 Discrete and Fast Fourier Transforms and Their Applications

Just as we can represent continuous, time-domain signals in the frequency domain using the CFT pair, shown in the following,

$$\mathbf{F}(j\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt \quad (10.61)$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{F}(j\omega)e^{j\omega t} d\omega, \quad (10.62)$$

we can also characterize discrete time signals of finite length in the frequency domain using the DFT and the inverse discrete Fourier transform (IDFT). In the CFT, $f(t)$ can exist over all time. In the DFT, we assume a finite number of samples, N , is taken of $x(t)$, each at an interval of T_s seconds. The total length of time window or epoch over which $x(t)$ is sampled is $T_E = (N - 1)T_s$. In computing the DFT, we assume $x(t) = x(t + NT_s)$, that is, $x(t)$ is periodic with a period of NT_s seconds.

The N elements of $\{x_n\}_N$ give rise to N elements of the DFT according to the relations

$$X_k = \sum_{m=0}^{N-1} x_m \exp\left(-j\frac{2\pi}{N}km\right), \quad k = 0, 1, \dots, N-1 \quad (10.63a)$$

or

$$X_k = \sum_{m=0}^{N-1} x_m \cos\left(\frac{2\pi km}{N}\right) - j \sum_{m=0}^{N-1} x_m \sin\left(\frac{2\pi km}{N}\right) \quad (10.63b)$$

The N terms of the DFT are complex numbers (vectors); they have real and imaginary parts or, equivalently, magnitudes and angles. The frequency spacing between adjacent terms of the spectrum, $\{X_n\}_N$, is given by the following equation:

$$\Delta\omega = \frac{2\pi}{T_E} = \frac{2\pi}{(N-1)T_s} \text{ r/s}. \quad (10.64)$$

The total frequency span of $\{X_k\}_N$ is 0 to $2\pi/T_s$ r/s.

The DFT has several interesting properties: It is a linear operator, that is, it obeys superposition, sic:

$$\text{DFT}_N\{Ap_k + Bq_k\} = A \text{DFT}_N\{p_k\} + B \text{DFT}_N\{q_k\}. \quad (10.65)$$

The DFT is periodic with period N . That is,

$$X_k = X_{k+N}. \quad (10.66)$$

A very important property of the DFT is that the DFT of real signals is *conjugate-symmetric* (Williams 1986). This can be written as

$$X_k = X_{N-1-k}^*, \quad 0 \leq k \leq N/2. \quad (10.67)$$

Another way of describing this property is to state that the real part of X_k is *even* about $k = N/2$ and the imaginary part of X_k is *odd* around $k = N/2$. Thus, the magnitude of X_k is *even* around $k = N/2$ and is equal to the magnitude of X_{N-1-k}^* . This means that one-half of a DFT spectrum of $\{x_n\}_N$ is redundant! Thus, N samples of $x(t)$ yield $N/2$ useful spectrum values ranging from 0 to $f_N = 1/2T_s$ Hz.

The inverse DFT can be written as

$$x_k = \frac{1}{N} \sum_{m=0}^{N-1} X_m \exp\left(+j \frac{2\pi}{N} mk\right) \quad (10.68)$$

The IDFT, also, is linear and periodic in N .

As an example of setting up a DFT routine, let us assume we need to resolve and examine in the frequency domain two, coherent (sinusoidal) signals to which broadband noise has been added. The signals are at 400 and 420 Hz. We wish to examine the magnitudes of $\{X_k\}_N$ to determine the relative levels of the signals and the noise. First, let us set the Nyquist frequency of the system to be 1000 Hz. The AAFs must have attenuated the signals plus noise to a negligible output power level at and above 1 kHz. The sampling rate required is thus 2000 samples/s. The signal spacing is 20 Hz. This implies that we should be able to resolve a $\Delta f = 2$ Hz. Thus $T_E = 1/\Delta f = 0.5$ s, and $N = T_E/T_s + 1 = 1001$ samples. (Typically, 1024 samples could be used in calculating $\{X_k\}_N$.)

10.4.1 Use of Data Windows to Improve Spectral Resolution

Direct computation of $\{X_k\}_N$ using Equation 10.63 is always done with a *windowing function* operating on the sampled data, $\{x_k\}_N$. That is, the product

$$x'_k = w_k x_k \quad (10.69)$$

is computed in the time domain, where w_k is the windowing function. The purpose of a windowing function is to reduce a phenomenon known as *spectral leakage*. If a pure, sinusoidal $x(t)$ is sampled and DFTed, we generally observe symmetrical, nonzero side lobes in $\{X_k\}_N$ around the main peak at the sinusoid's frequency. This phenomenon was called spectral leakage because spectral energy at the sinusoid's frequency effectively *leaks* into the adjacent frequency terms. What windows do in general is to reduce the magnitude of the leakage side lobes at the expense of the sharpness of the main spectral peak at the input sinusoid's frequency.

In general, in the frequency domain,

$${}^w X(j\omega) = W(j\omega) \otimes X(j\omega). \quad (10.70)$$

That is, the CFT of the windowed, sampled data sequence can be expressed as the *complex convolution* of the CFT of the windowing function with the CFT of $x(t)$.

To illustrate how windowing works, we first consider the case of the simplest, rectangular window consisting of N unit impulses, spaced T_s seconds apart. The rectangular window function is a truncated unit impulse train that multiplies (modulates) $x(t)$ in the time domain to generate the sampler output. The rectangular window function can be written as

$$w(t) = \{w_k\}_N = \sum_{k=0}^{N-1} \delta(t - kT_s). \quad (10.71)$$

A fundamental property of FTs is that the CFT of an even function in t is even in ω and has zero phase. It is therefore algebraically convenient to consider both the window and $x(t)$ to be even functions. This process is shown in Figure 10.8. We choose N to be an odd number so M will be an integer. Note that the epoch length T_E is

$$T_E = (N - 1)T_s = 2MT_s. \quad (10.72)$$

Hence,

$$M = \frac{N - 1}{2}. \quad (10.73)$$

The CFT of the rectangular window can be shown to be (Papoulis 1968)

$$W(\omega) = \frac{\sin[\omega(2M + 1)T_s/2]}{\sin(\omega T_s/2)} = \frac{\sin(\omega NT_s/2)}{\sin(\omega T_s/2)}. \quad (10.74)$$

For purposes of this example, let us assume that $x(t)$ is a cosine wave (even function): $x(t) = A \cos(\omega_0 t)$, where $\omega_0 < \omega_N$. The CFT of the cosine $x(t)$ is well known:

$$X(\omega) = A[\pi\delta(\omega + \omega_0) + \pi\delta(\omega - \omega_0)]. \quad (10.75)$$

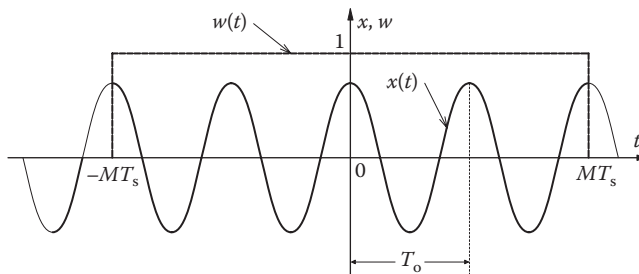


FIGURE 10.8

A unit rectangular (even) window function and a cosine wave. A finite number of samples of the cosine wave are generated by multiplying it by the 0, 1 rectangular window function.

Complex convolution now becomes real convolution, where the kernel only exists at $\omega = \pm\omega_0$. Thus, the CFT of the finite, sampled, rectangular-windowed cosine wave is

$${}^wX(\omega) = \frac{A}{2} \left\{ \frac{\sin\left[\frac{(\omega + \omega_0)NT_s}{2}\right]}{\sin\left[\frac{(\omega + \omega_0)T_s}{2}\right]} + \frac{\sin\left[\frac{(\omega - \omega_0)NT_s}{2}\right]}{\sin\left[\frac{(\omega - \omega_0)T_s}{2}\right]} \right\}. \tag{10.76}$$

The main spectral peaks in ${}^wX(\omega)$ are seen to occur at $\omega = \pm\omega_0$. The leakage is due to the side lobes of Equation 10.74. Figure 10.9 illustrates normalized plots of the DFTs of several window functions (see caption for details). Clearly, to reduce leakage and to get the best resolution, we must find a window function whose CFT has a narrow main lobe and low side lobes compared to the main lobe peak. Many such window functions have been devised and used. These include, but are not limited to (Papoulis 1977), *Bartlett*, *Tukey*, *Hamming*, *parabolic*, *maximum energy concentration*, *minimum energy moment*, *minimum amplitude moment*, *von Hann (Hanning)*, and *Kaiser*. (The latter two windows are described in Williams, 1986.) The Hamming window is one of the most widely used windows in DFT and FFT computations. Its CFT has low major side lobes. It consists of a cosine on a pedestal. For $N \geq 64$ samples, it can be shown that the pedestal height approaches 0.536 and the cosine term is multiplied by 0.464. Each sample of x_k is multiplied by a corresponding w_k :

$$w_k = 0.536 + 0.464 \cos\left[\frac{\pi(2k - N + 1)}{N}\right], \quad k = 0, 1, \dots, (N - 1). \tag{10.77}$$

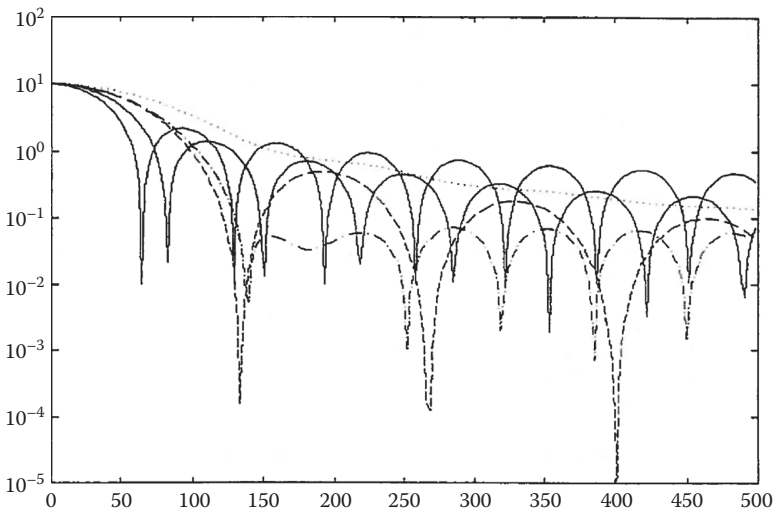


FIGURE 10.9 Normalized plots of the DFT magnitudes of several commonly used windowing functions. *Note:* The solid trace with the narrowest main lobe and the highest side lobes = rectangular (0, 1) window. Solid trace with wider main lobe and lower side lobes = Kaiser window. Dashed trace = Bartlett (triangular) window. Dash-dot trace = Hamming window. Dotted trace = triangle squared (Parzen) window. Computed with MATLAB® were 1024-point FFTs. A log magnitude vs. linear frequency scale is used.

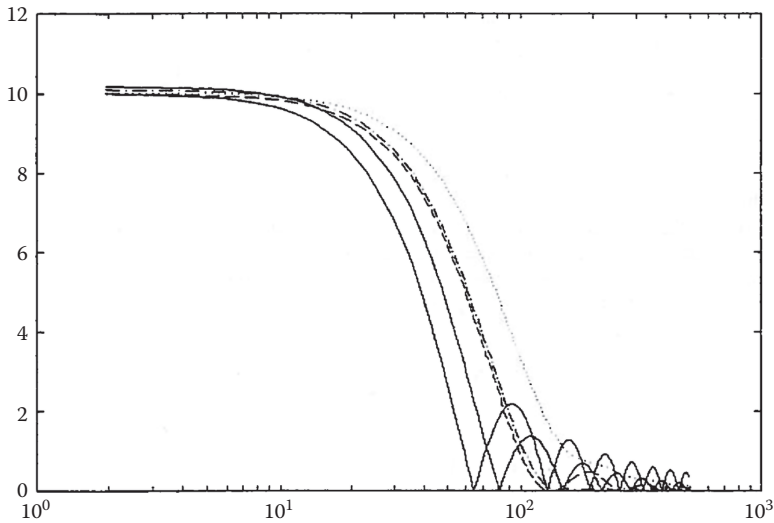


FIGURE 10.10

Normalized plots of the DFT magnitudes of the windows of Figure 10.9, except that a linear magnitude vs. log frequency scale is used. Note the trade-off between main lobe narrowness and side lobe height. The same trace identification used in Figure 10.9 is used.

In the symmetrical, even function case, we center the window function at $m = 0$, and we can write

$$w_m = 0.536 + 0.464 \cos \left[\frac{\pi 2m}{N-1} \right], \quad -N/2 \leq m \leq (N/2-1). \quad (10.78)$$

(integer m)

Figure 10.10 illustrates the major lobes of several common window functions. The DFTs have been normalized. Note that the base width, B , of the major spectral lobe of the Bartlett (triangular) and Hamming windows is about

$$B \approx \frac{8\pi}{(N-1)T_s} \text{ r/s}. \quad (10.79)$$

The B of the Parzen window is somewhat wider. The choice of the window algorithm to be used involves a trade-off between main lobe width and the amount of spectral leakage seen at a given distance from the frequency of the main peak. Generally, the narrower the main lobe, the greater the leakage. The selection of a window algorithm can be somewhat subjective, and the window selected can influence one's ability to interpret the DFT spectrograms. Most commercial DFT spectrum analyzers offer the user the choice of a plain, rectangular window, a cosine window such as the Tukey, Hanning or Hamming, or the Kaiser window.

10.4.2 Use of the DFT to Characterize Random Signals and Noise

A major use of the modern spectrum analyzer is to allow us to visualize the distribution of power with frequency in an analog waveform. This feature can be useful in vibration

analysis of mechanical structures, measurement of noise, determination of linear system transfer functions by the white noise method, etc.

In Chapter 3, we saw that the PDS is used to describe the power content at a particular frequency of an infinite-length, analog waveform. The units of the PDS are mean-squared volts (or units) per Hz. In modern engineering practice, we use the DFT with an appropriate window to estimate the square root of the PDS (RPDS) of an analog waveform. In this case, the RPDS units are rms volts (units)/ $\sqrt{\text{Hz}}$.

The discrete RPDS can be calculated in the following manner: First, the analog waveform $x(t)$ is passed through an analog antialiasing LPF to remove any significant spectral power at and above the Nyquist frequency, $\omega_N = \omega_s/2 = \pi/T_s$ r/s. Next, a finite length of the AAF's analog output, $x(t)$, is sampled at a rate of $1/T_s$ samples/s. N values of $x(nT_s)$ are stored in an array, $\{x_k\}_N$, in the computer. Each x_k is multiplied by the corresponding window weighting constant, w_k , giving a windowed data array, $\{w x_k\}_N$. N elements of the DFT of $\{w x_k\}$ are now calculated using symmetrical coefficients, giving $\{W X_m\}_N$. Because of the conjugate symmetry property of DFT output, only $N/2$ points in $\{W X_m\}_N$ are unique. We store a buffer array consisting of $\{W X_m\}_{N/2}$ for $m = 0$ to $m = (N/2 - 1)$. $m = 0$ corresponds to $f = 0$ and $k = (N/2 - 1)$ corresponds to $f = f_{\text{MAX}} = f_s/2 = f_N = 1/2T_s$ Hz in the spectrum.

The frequency spacing between Fourier coefficients is $\Delta f = 1/(N - 1)T_s$ Hz. The coefficients of a hypothetical, simple, $N = 8$ point DFT are shown in Figure 10.11. We note that the DFT coefficient magnitudes can be found from the Pythagorean theorem. Also, we observe that the important parameter, T_s , does not appear in the DFT operation, Equation 10.63. T_s is used to calculate the spectrum scale factors, Δf and f_{MAX} .

If $x(t)$ is a broadband (noisy) waveform, there will be a great deal of natural variation between corresponding spectrum coefficient's magnitudes in the computation of M similar spectra. Averaging the corresponding elements of M spectra permits a reduction of the standard deviation of a given element by a factor of about $1/\sqrt{M}$.

In the (now discontinued) HP/Agilent model 3582A DFT spectrum analyzer, noisy spectra were averaged in the following manner: M spectra are computed sequentially. Two arrays of length N are set up. One holds the average of the squares of the corresponding real parts of the DFT coefficients, the other the average of the squares of the corresponding imaginary parts of the DFT coefficients. After M spectra are calculated and the mean squares of the corresponding coefficients are stored, the corresponding mean squares are added, and the square root is taken of each corresponding sum, giving a true, rms, RPDS. These steps can be shown mathematically:

$$\overline{[\text{Re}\{X_k\}]^2} = \frac{1}{M} \sum_{i=1}^M [\text{Re}\{^i(X)_k\}]^2 \tag{10.80}$$

$$\overline{[\text{Im}\{X_k\}]^2} = \frac{1}{M} \sum_{i=1}^M [\text{Im}\{^i X_k\}]^2 \tag{10.81}$$

$$\text{rms } X_k = \sqrt{\overline{[\text{Re}\{X_k\}]^2} + \overline{[\text{Im}\{X_k\}]^2}}, \quad N/2 \leq k \leq (N - 1). \tag{10.82}$$

(integer k)

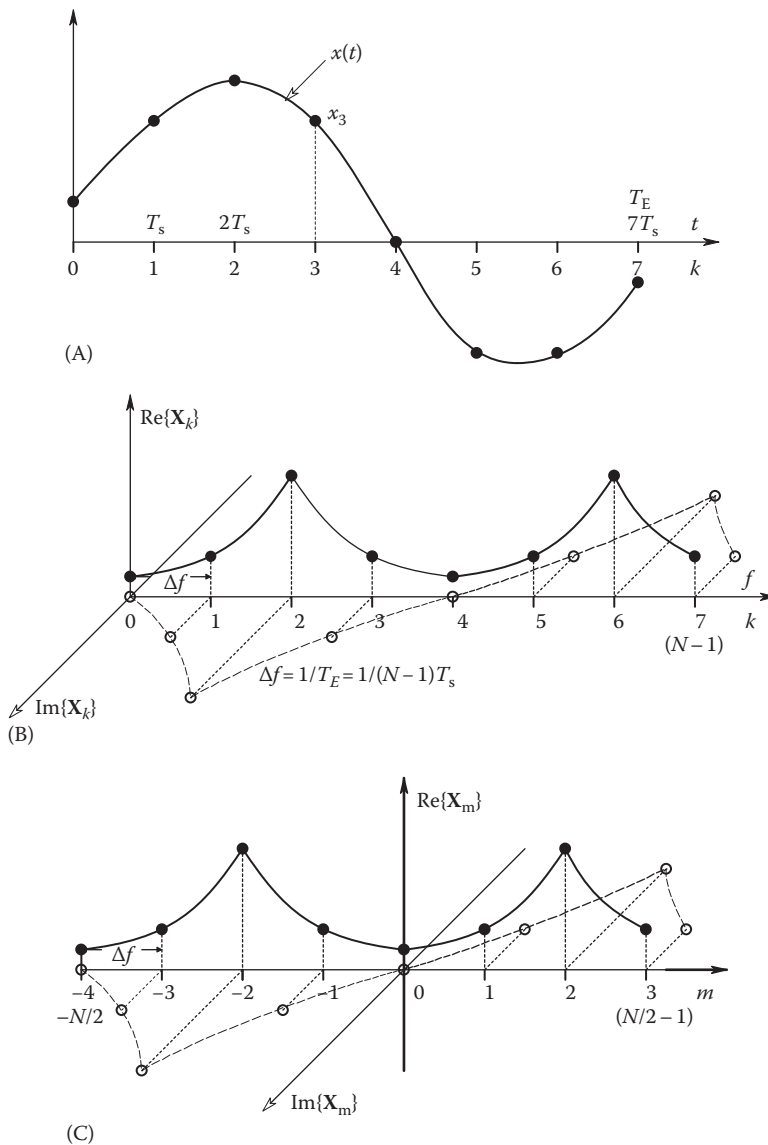


FIGURE 10.11

(A) Eight samples of a continuous time function, $x(t)$, are taken. An eight-sample, rectangular window is used. (B) The hypothetical DFT of the eight-point x_k is shown. Real and imaginary components of the complex DFT are shown for $0 \leq k \leq 7$ (eight samples). Note that the Pythagorean magnitude of the DFT is repeated for $0 \leq k \leq (N/2 - 1)$ and for $N/2 \leq k \leq (N - 1)$. This is a result of the property of *conjugate symmetry*. (C) By translating DFT coordinates, we see that we only need half of the N -point DFT to characterize x^* in the frequency domain.

The HP 3582A spectrum analyzer also offered the option of an exponential, moving average to reduce the uncertainty in calculating the spectra of nonstationary, noisy signals. *Nonstationarity* means the underlying physical processes producing the noise in $x(t)$ are changing in time, for example, a resistor may be heating up, and its increasing temperature produces an increasing resistance and increasing thermal noise. The exponential

averaging process weights the most recent samples more heavily than old samples. We illustrate it in the following for the real parts of the k th spectral element:

$$\begin{aligned} \overline{\exp[\operatorname{Re}\{X_k\}]^2} &= \left(\frac{1}{4}\right) \left\{ \left[\operatorname{Re}\{^1 X_k\} \right]^2 + \left(\frac{3}{4}\right)^1 \left[\operatorname{Re}\{^2 X_k\} \right]^2 + \left(\frac{3}{4}\right)^2 \left[\operatorname{Re}\{^3 X_k\} \right]^2 + \dots \right. \\ &\quad \left. + \left(\frac{3}{4}\right)^{M-2} \left[\operatorname{Re}\{^{M-1} X_k\} \right]^2 + \left(\frac{3}{4}\right)^{M-1} \left[\operatorname{Re}\{^M X_k\} \right]^2 \right\}. \end{aligned} \tag{10.83}$$

Here, the most recent DFT calculation is given the index 1 and the first, or oldest, the index M . Note that $(3/4)^8 = 0.100$ and $(3/4)^{16} = 0.0100$, so the *tail* of the moving average has negligible weight after computing 16 consecutive DFTs from $x(t)$. As in the case of calculating the rms X_k , the exponentially averaged rms X_k is found by the Pythagorean theorem:

$$\exp(\operatorname{rms} X_k) = \sqrt{\exp[\operatorname{Re}\{X_k\}]^2 + \exp[\operatorname{Im}\{X_k\}]^2}. \tag{10.84}$$

10.4.3 Fast Fourier Transform

In order to obtain a good approximation to the CFT of a bandwidth-limited waveform, we must calculate the DFT of the signal with a sampling rate above twice the highest frequency in the PDS of $x(t)$ and use enough samples to ensure a close spacing, Δf , between adjacent values of X_k . It is common to calculate spectra using N s ranging from $256 = 2^8$ to $4096 = 2^{12}$ and higher. The use of the DFT algorithm, Equation 10.63a, as written is seen to require 2^N calculations (multiplications and additions) to calculate each X_k coefficient's real and imaginary parts. However, there are N coefficients, so a total of $2N^2$ multiplications and additions are required to obtain the DFT, $\{X_k\}_N$. If the magnitudes of $\{X_k\}_N$ are to be displayed, then we must apply the Pythagorean theorem N times as well. It is clear that a large DFT calculated by the direct method will involve many high-precision computer operations and take a long time. For example, computation of an $N = 4096$ point DFT involves over 33.5 million multiplications and additions. This is time consuming.

To overcome the computational burden of direct computation of large DFTs, Cooley and Tukey (1965) devised an FFT means of calculating the DFT. Actually, there are now many variations on the original Cooley–Tukey FFT. It is not our purpose here to describe in detail the steps involved in calculating an FFT/DFT. Rather, we will outline the strategy used and explain why it is efficient.

Using the complex form of the DFT, direct calculation of the k th complex term can be written out as

$$X_k = x_0 e^{-j(2\pi k 0/N)} + x_1 e^{-j(2\pi k 1/N)} + \dots + x_i e^{-j(2\pi k i/N)} + \dots + x_{N-1} e^{-j(2\pi k [N-1]/N)}. \tag{10.85}$$

Each of the N exponential terms in Equation 10.85 defines an angle for the corresponding real x_k . Thus, N vector summations are required in order to find X_k in the conventional DFT, and a total of N^2 vector summations must be done to find the entire DFT, $\{X_k\}_N$.

In computing an FFT, use is made of the fact that the exponential angle operator is periodic in k_i/N . This periodicity leads to redundancy in the DFT calculations that is taken advantage of in an FFT algorithm by rearranging the order of the calculations so that there are $\log_2 N$ columns of vector summers, each of which contains N summers, for a total of $N \log_2(N)$ complex summers in an FFT array. Each summer sums two vectors. In FFT

operations, N is generally made a power of 2, for example, $N = 2^b$, where b is a positive integer, such as $b = 10$ for $N = 1024$. The angles of the vectors must still be calculated when the FFT array is set up and stored as constants. Their values depend only on k, i , and N . It is not practical to try to illustrate the signal flow graph of an FFT operation for N s over 8. Figure 10.12 shows an implementation of an eight-point FFT. Note the symmetrical rearrangement of the $x_{k,s}$ s at the FFT input. In some versions of the FFT, the output coefficients are rearranged.

The efficiency, η , of the FFT vs. the DFT calculated in the direct method can be calculated by assuming that computation speed is inversely proportional to the number of vector additions required in each case. So we can write

$$\eta = \frac{S_{\text{FFT}}}{S_{\text{DFT}}} = \frac{N}{\log_2(N)}. \tag{10.86}$$

For a 4096-point FFT, $\eta = 341.3$, for $N = 1024$, $\eta = 102.4$, etc.

There are many commercial FFT software routines available in various computer languages and programs, including FORTRAN, C, and MATLAB®. Also, companies

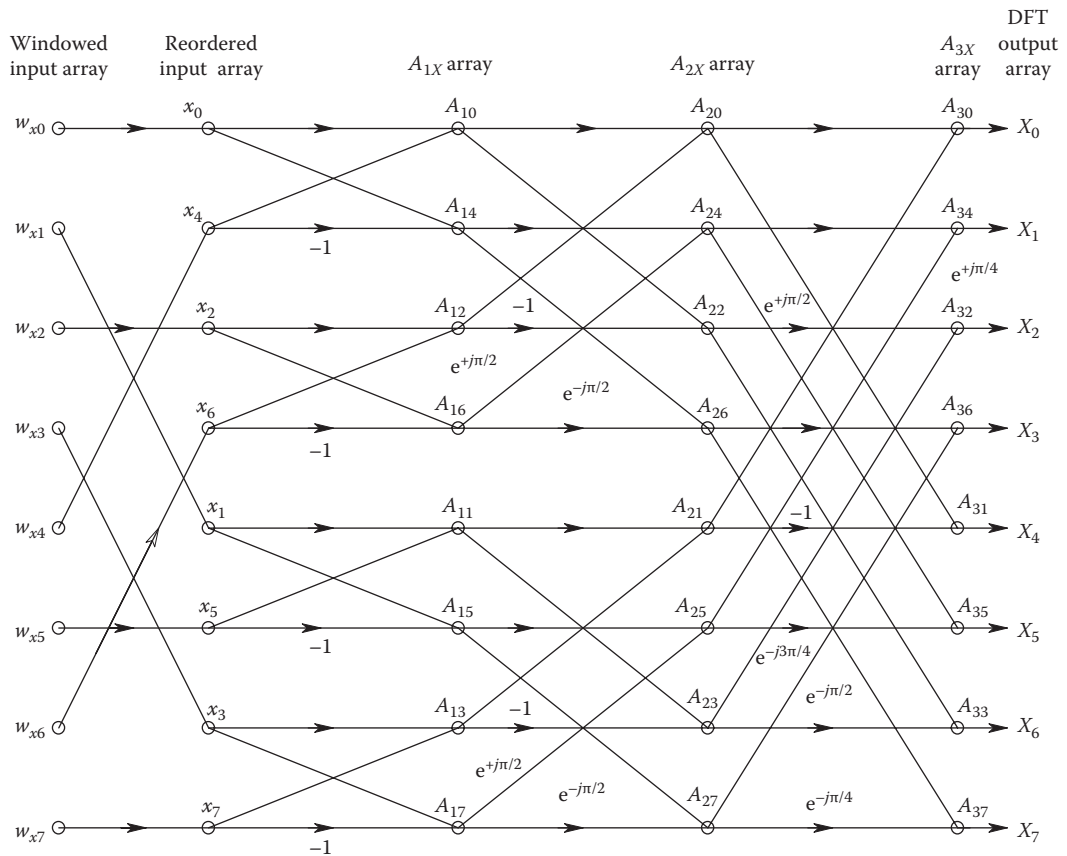


FIGURE 10.12 Flow graph showing an implementation of the Cooley–Tukey FFT algorithm for an eight-sample time function. Note that the FFT output values in the A_{3x} array are complex (i.e., they have real and imaginary parts).

offering data acquisition interface cards and DSP cards for PCs generally offer their software versions of FFT routines. Thus, there is a low probability that a person setting up a DSP capability in a measurement system will have to write their own FFT routine from scratch. For a tutorial on FFT operations, see Cerna and Harvey (2000).

DSP coprocessor boards: In the 1990s, there were a number of commercially available, plug-in, DSP coprocessor boards to compute FFTs in PCs and workstations. Designs for these boards constantly evolved, improving their speeds and capabilities. Many of them have now been discontinued as computer CPU speed and memory size has grown. Some of the commercially available, dedicated FFT cards included the discontinued *Array Microsystems* model a66540 frequency domain array processor (FDAP) card for VME bus computers that could do a 1024-point, real FFT in 13.3 μ s. The (now discontinued) *Ixthos* IXD7232 VME bus DSP card could run at a sustained 50 MFLOPS and do a 1024-point, complex FFT (radix 4) in 770 μ s. The *NI AFFA system* used two plug-in cards to make a powerful, audio-frequency FFT system for the Macintosh II PC with NuBus. An AFFA system does real-time audio spectrum analysis up to 20 kHz. It can do 64-, 126-, 256-, 512-, or 1024-point FFTs at sampling rates of 22.05, 24, 32, 44.1, or 48 k samples/s. It has rather elegant, menu-controlled operating software, which includes choice of rectangular, Hanning, Hamming, ExactBlackman, Blackman, or Blackman-Harris data windows. Many other companies, too numerous to list here, have made coprocessor, plug-in boards for PCs, and they offer operating software that enables flexible calculation of FFTs.

The current trend is not to rely on expensive, plug-in coprocessor boards for PCs and laptops. High-end Pentium™ systems with >3 GHz clocks have largely made them obsolete—let the CPU do the FFT. For example, *Datel*® phased out its line of board-level data acquisition and processing products as of February 28, 2004, including its powerful PCI-431 series of analog I/O, DSP coprocessor boards for the PCI bus. There is still a need for DSP coprocessors and DSP accelerator boards, however, for dedicated (non-PC), OEM DSP systems. *RTD Embedded Technologies*® offers a line of DSP interface and coprocessor and accelerator boards using Texas Instruments DSP chipsets.

10.5 Digital Routines for Interpolating Discrete Data

Often, in making and in interpreting measurements, we have the need to reconstruct a smooth analog curve between discrete data points. There are a variety of means of generating a smooth, continuous curve between a set of N discrete data points, $\{x_k, y_k\}_N$. Note that for a time-sampled $y(t)$, $x_k = kT_s$, $k = 0, 1, 2, \dots, N - 1$. One method of smooth curve generation is to compute a piecewise-linear approximation between the data points. In this approach, we generate a number of polynomial segments of relatively low order (0, 1, 2, 3) between the points. The data point pairs at the ends of each polynomial segment are called *knots*. Polynomials of order 2 or 3 are generally called *splines*. Polynomial approximations of orders 0 and 1 are shown in Figure 10.13. We will discuss *cubic splines* here because they are generally the most useful in terms of a trade-off between computational complexity and accuracy. The reader seeking more information on the field of splines should consult texts on numerical analysis such as those by Greenspan and Casulli (1993) and Scheid (1990).

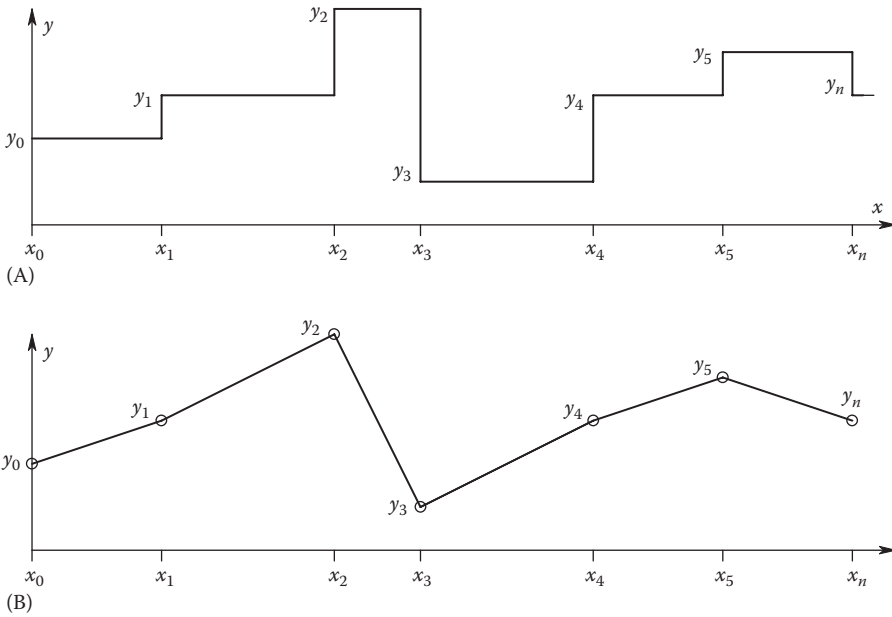


FIGURE 10.13

Illustration of zero-order (A) and first-order (B) polynomial approximations for interpolating between a set of discrete data points in order to regenerate the continuous time function, $x(t)$, from which the samples were taken.

A *cubic spline* is a cubic polynomial that is made to fit three consecutive data points, (x_k, y_k) , (x_{k+1}, y_{k+1}) , and (x_{k+2}, y_{k+2}) , or, if y is sampled data, (kT_s, y_k) , $([k + 1]T_s, y_{k+1})$, and $([k + 2]T_s, y_{k+2})$. The polynomial may be written as

$$y = a + bx + cx^2 + dx^3 \tag{10.87}$$

and

$$\dot{y} = b + 2cx + 3dx^2. \tag{10.88}$$

We assume that $x_0 = t = 0$ (local time origin) and that the intervals $x_i - x_{i-1} = T_s$ over the entire data set. Thus, $x_0 = 0, x_1 = T_s, x_k = kT_s$, etc. We also assume that the first derivative, \dot{y}_0 , is known or well estimated. To calculate the cubic spline for the first three data points $k = 0, 1, 2$, we can write

$$y_0 = a \tag{10.89a}$$

$$y_1 = a + bT_s + cT_s^2 + dT_s^3 \tag{10.89b}$$

$$y_2 = a + b2T_s + c4T_s^2 + d8T_s^3. \tag{10.89c}$$

Also, the first derivative at $x_0 = t = 0$ is from Equation 10.87:

$$\dot{y}_0 = b. \quad (10.89d)$$

We wish to find the coefficients a , b , c , and d for the $k = 0, 1, 2$ spline. This may be done by simultaneously solving Equations 10.89a through d. Doing so gives

$$a = y_0 \quad (10.90a)$$

$$b = \dot{y}_0 \quad (10.90b)$$

$$c = \frac{-7y_0 + 8y_1 - \dot{y}_2 - 6T_s y_0}{4T_s^2} \quad (10.90c)$$

$$d = \frac{3y_0 - 4y_1 + \dot{y}_2 + 2T_s y_0}{4T_s^3}. \quad (10.90d)$$

Substituting these values into the cubic spline formula yields

$$y(t) = y_0 + \dot{y}_0 t + \frac{-7y_0 + 8y_1 - y_2 - 6T_s \dot{y}_0}{4T_s^2} t^2 + \frac{3y_0 - 4y_1 + y_2 + 2T_s \dot{y}_0}{4T_s^3} t^3 \quad (10.91)$$

over the interval $0 \leq t \leq 2T_s$. By recursion, we can write a general cubic spline formula valid for any triplet of data points, each T_s seconds apart, (y_k, y_{k+1}, y_{k+2}) :

$$\begin{aligned} S_3(t) = & y_k + \dot{y}_k (t - kT_s) + \frac{-7y_k + 8y_{k+1} - y_{k+2} - 6T_s \dot{y}_k}{4T_s^2} (t - kT_s)^2 \\ & + \frac{3y_k - 4y_{k+1} + y_{k+2} + 2T_s \dot{y}_k}{4T_s^3} (t - kT_s)^3. \end{aligned} \quad (10.92)$$

The first derivative estimate needed for the cubic spline of the next, adjacent triad of data points is given by

$$\dot{y}_{k+2} = \dot{y}_k + \frac{2(y_{k+2} - 2y_{k+1} + y_k)}{T_s}. \quad (10.93)$$

Note that the first *knot* is at $k = 0$, the second knot is at $k = 2$, the third is at $k = 4$, etc.

As an example of calculating the cubic splines of a small data set using Equations 10.92 and 10.93, let $T_s = 1$ s, $\dot{y}_0 = 1$, $y_0 = 0$, $y_1 = 1$, $y_2 = 2$, $y_3 = 1$, $y_4 = 0$. Over the first interval, $0 \leq t \leq 2T_s = 2$ s, we have

$$S_{0,2}(t) = 0 + (1)t + (0)t^2 + (0)t^3 = t. \quad (10.94)$$

The derivative estimate at $k = 2$ is

$$\dot{y}_2 = 1 + \frac{2[2 - 2(1) + 0]}{1} = 1. \quad (10.95)$$

And over the interval, $2T_s \leq t \leq 4T_s$, the spline is

$$\begin{aligned} S_{2,4}(t) &= 2 + (1)(t-2) + \frac{[-7(2) + 8(1) - 6(1)(1)]}{4(1)}(t-2)^2 + \frac{[3(2) - 4(1) + (0) + 2(1)(1)]}{4(1)}(t-2)^3 \\ &= 2 + (t-2) - 3(t-2)^2 + (t-2)^3. \end{aligned} \quad (10.96)$$

Many other types of spline functions exist, as do variations on the cubic spline. For example, if both the first and second derivatives of the sampled variable are known at $t = 0$, then the cubic spline, $S(t)$, can be found between consecutive pairs of data points instead of between triplets of points. Using a development similar to that presented earlier for three-point, cubic splines, it is possible to show that for $kT_s \leq t \leq (k+1)T_s$,

$$Sk(t) = \dot{y}_k + y_k(t - kT_s) + \left(\frac{\ddot{y}_k}{2}\right)(t - kT_s)^2 + \frac{y_{k+1} - y_k - T_s\dot{y}_k - (T_s^2/2)\ddot{y}_k}{T_s^3}(t - kT_s)^3 \quad (10.97)$$

and

$$\dot{y}_k = \dot{y}_{k-1} + T_s\ddot{y}_{k-1} + 3\frac{y_k - y_{k-1} - T_s\dot{y}_{k-1} - (T_s^2/2)\ddot{y}_{k-1}}{T_s} \quad (10.98)$$

$$\ddot{y}_k = \ddot{y}_{k-1} + 6\frac{y_k - y_{k-1} - T_s\dot{y}_{k-1} - (T_s^2/2)\ddot{y}_{k-1}}{T_s^2}. \quad (10.99)$$

Although the scope of this discussion has been an introduction to cubic splines, it should be noted that the technique of splines is easily applied to data having nonuniform sampling periods, the resulting spline approximations can then be *sampled* at a uniform rate, and these M samples can be used to calculate a conventional DFT of $\{x_k y_k\}_N$ with $M > N$. Also, spline recursion equations can be put in tridiagonal matrix form for simultaneous solution of $Sk(t)$ from the entire data set at once (Scheid 1990). Some of the other interpolation methods that exist for analog data interpolation include the methods of LaGrange, Hermite, Gauss, Stirling, and Newton and the least squares method of fitting a continuous linear equation to a data set, $\{x_k, y_k\}_N$ (described in Chapter 1).

10.5.1 Estimating Missing Data at the Sampling Instants

We occasionally have a data sequence where one or more data points are missing, perhaps because the experimenter neglected to record them or because burst of noise completely

masked the true data value. If data point y_k is missing in a large data set, and the data set does not vary abruptly in time, then the missing data point can be found by linear interpolation:

$$y_k = \frac{y_{k+1} + y_{k-1}}{2}, \quad (10.100)$$

that is, the missing y_k can be approximated by the average of its nearest neighbors. In the event that the missing data point is part of a curve with nonzero second derivative, a more accurate estimate can be derived from the solution of four, simultaneous, cubic algebraic equations (Williams 1986). Two adjacent data points on either side of the missing data point are used:

$$y_k = \frac{-y_{k-2} + 4y_{k-1} + 4y_{k+1} - y_{k+2}}{6}. \quad (10.101)$$

Because cubic splines work with nonuniform data x -values, they also can be used to find a missing y_k . We assume y_{k-2} , y_{k-1} , and y_{k+1} are known and can be used in the spline calculation. \dot{y}_{k-2} also must be known or estimated.

A data point missing at the beginning or end of a finite data sequence, $\{x_k, y_k\}_N$, requires numerical extrapolation, which can also be carried out by a polynomial estimation routine using existing data (Williams 1988). The accuracy of extrapolation decreases as the interval (x_N, x_{N-1}) increases and as the bandwidth of $Y(j\omega)$ increases. If data are changing monotonically and are not noisy, then a reasonable value for y_N can be calculated.

We give an example of a simple extrapolator that uses the three-point central difference differentiator to estimate the slope of y_k at the end of the sequence, where $k = N - 1$. This slope is found from Equation 10.43:

$$m_{k-1} = \frac{y_{N-1} - y_{N-3}}{2T_s}. \quad (10.102)$$

Simple algebra tells us that the extrapolated value of y at $t = NT_s$ is

$$y_N = y_{N-1} + m_{k-1}T_s = 1.5y_{N-1} - 0.5y_{N-3}. \quad (10.103)$$

This same, basic procedure can be turned around to find an estimate of y_{-1} at $t = -T_s$.

The three-point, cubic spline approach can also be used to find an estimate for y_N at the end of the sequence $\{x_k, y_k\}_N$. The trick here is to redefine the local time origin so that the new $y_1 = y_N$ (value to be extrapolated), $y_0 = y_{N-1}$ (the last data sample of the sequence), $y_{-1} = y_{N-2}$, and $y_{-2} = y_{N-3}$ in the sequence. The first derivative at y_0 is estimated by the three-point central difference method:

$$\dot{y}_0 = \frac{y_0 - y_{-2}}{2T_s} = \frac{y_{N-1} - y_{N-3}}{2T_s}. \quad (10.104)$$

Equations 10.89a through c can now be used with appropriate changes of indices and the derivative estimate of Equation 10.104 to show that in terms of the original data set,

$$y_N = 3y_{N-2} - 2y_{N-3}. \quad (10.105)$$

Proof of Equation 10.105 is left as an exercise for the interested reader. Note that y_{N-1} curiously drops out of this cubic spline extrapolation formula.

10.6 Chapter Summary

DSP and its related mathematical field of numerical analysis are of considerable importance to all branches of engineering, particularly to electrical and systems engineers and to measurement and instrumentation engineers. Many textbooks have been written in these areas; we cite a representative group in the bibliography and references for this chapter. Our purpose in this chapter was to introduce some key concepts in DSP that are of particular use in instrumentation and measurement system design.

We first discussed the z -transform and digital filters as a linear means of operating on a sequence of sampled data in order to do a discrete filtering operation analogous to an analog filter, such as low-pass or band-pass filtering. We showed that the relevant range of input signal frequencies in digital filtering is from 0 to the Nyquist frequency, $f_s/2$. Examples of finite impulse response (FIR) low-pass and band-pass filters were given. We also examined the time-domain structure of the filtering process, showing how a computer program running in real time can implement a filtering algorithm by a series of synchronous delays, multiplications by constants, and summations. Recursive, IIR filters were introduced, and a second-order, high- Q , notch filter implementation was given as an example.

In measurement systems, the operations of digital differentiation and integration are also important. Several basic differentiation and integration routines were discussed and their algorithms described in the frequency domain. Integration routines were shown to be recursive, while differentiators generally use FIR algorithms.

A significant portion of this chapter was devoted to introducing the DFT. We examined how it can approximate the CFT and how window functions prevent side lobe *leakage*. The basics of the FFT algorithm to compute the DFT were described, an eight-point FFT example was illustrated, and an expression for FFT/DFT efficiency was derived. The availability of commercially available, dedicated DSP and FFT plug-in boards for PCs was cited as the state-of-the-art way to accelerate FFT and digital filter calculations in real time.

Finally, we considered the problem of reconstructing continuous, analog data between data samples. We introduced cubic splines as an efficient way to reconstruct analog data, as well as to interpolate missing data points and to extrapolate a data point at the end of a data sequence.

Problems

10.1 A certain continuous time function, $f(t)U(t)$, has the z -transform

$$F(z) = \frac{z}{z - 0.367879}$$

- Plot its pole and zero in the complex z -plane.
- Use long division to find $F(z)$ as an infinite series with terms of the form z^{-n} , $n = 0, 1, 2, \dots, \infty$.
- Write the sampled time function $f(nT) = f(n)$ by inspection in closed form. What is the original continuous $f(t)$?

10.2 Consider the discrete, linear, time-invariant system (DLTIS):

$$H(z) = \frac{z^2}{z^2 - 0.6z + 0.05}.$$

- A. Factor the denominator to find the system's two real poles in the z -plane.
 - B. Do a *partial fraction expansion* on $H(z)/z$ to find $H(z)$ as the sum of two real pole terms.
 - C. Write the weighting function, $h(n)u(n)$, as the sum of two inverse z -transforms.
- 10.3 A DLTIS, $H(z)$, has the weighting function $h(n) = u(n)e^{-naT}$, where T is the sampling period, and

$$u(n) = \delta(t) + \delta(t-T) + \delta(t-2T) + \dots + \delta(t-nT) + \dots = \sum_{n=0}^{\infty} \delta(t-nT).$$

This system has the discrete input $x(n) = u(n)e^{-nbT}$ for $n = 0, 1, 2, \dots, \infty$.

Find the system's discrete, time-domain output, $y(n)$. *Hint:* Use a partial fraction expansion on $Y(z) = X(z)H(z)$.

10.4 Use the relation

$$\sum_{n=0}^{N-1} \rho^n = \frac{1-\rho^N}{1-\rho}$$

to find the z -transform, $G(z)$, of the truncated exponential decay, $g(n) = e^{-naT}$, defined for $0 \leq n \leq N-1$ and zero for $n < 0$ and $n \geq N$. *Hint:* Note that there is an $(N-1)$ th order pole at the origin of the z -plane for $G(z)$.

- 10.5 Consider the DLTI system defined by the difference equation $y(n) = x(n) + (1/3)y(n-1)$. $y(n)$ is the present output, $x(n)$ the present input, and $x(n) = u(n)$, a unit step. There are zero initial conditions. Find an expression for and plot and dimension $y(n)$, $n = 0, 1, 2, 3, 4, \dots$
- 10.6 A causal FIR filter has the impulse response $y(n) = x(n) - x(n-4)$, where $x(t) = \delta(0)$ (a '1' at $t = nT = 0$):
- A. Write the filter's transfer function, $H(z)$, as a polynomial in z .
 - B. Plot the filter's poles and zeros in the z -plane.
 - C. Find the filter's frequency response, $\mathbf{H}(j\omega)$, by substituting $\mathbf{z} = e^{j\omega T}$ into $H(z)$. Use the Euler relation to find, plot, and dimension $|\mathbf{H}(j\omega)|$ and $\angle \mathbf{H}(j\omega)$ vs. ω .
 - D. Find and plot the filter's output, $y(n)$, in response to a unit step input to the filter.
- 10.7 The DE for a causal FIR notch filter is given as $y(n) = b_0x(n) + b_1x(n-1) + b_2x(n-2)$:
- A. Find the filter's transfer function, $H(z)$.
 - B. Find an expression for $\mathbf{H}(j\omega)$, magnitude, and phase. Let $b_0 = b_2$, and use the Euler relation for $\cos \theta$.
 - C. Find the b_0 and b_1 values required so that $\mathbf{H}(0) = 1 \angle 0$ and $\mathbf{H}(\pi/4T) = 0$.
- 10.8 Find the DFT, $\mathbf{X}(k)$, of the sequence, $x(n) = \sin(n\pi/N)$, $n = 0, 1, 2, \dots, N-1$. (*Hint:* Write $x(n)$ in exponential form.)

- 10.9** Find the DFT, $\mathbf{H}(k)$, of the Hamming window sequence, $h(n) = 0.54 + 0.42 \cos(n\pi/N)$, $n = 0, 1, 2, \dots, N - 1$.
- 10.10** Find the DFT, $\mathbf{X}(k)$, of the eight-point sequence $x(n) = 0, 0 \leq n \leq 4$; $x(n) = 1, 5 \leq n \leq 6$; $x(n) = 0, n = 7$.
- 10.11** $x(n)$ is the sampled input, $h(n)$ is the weighting function of a DTIF, and $y(n)$ is the output number sequence. Find $y(n)$ by linear, discrete convolution, following the procedure given in Figure 10.1. See Figure P10.11.

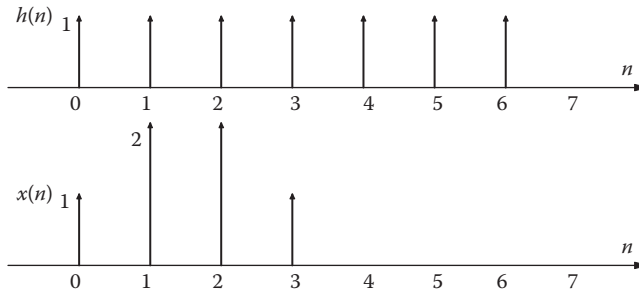


FIGURE P10.11

11

Solid-State Chemical Microsensors and Wireless Instrumentation

11.1 Introduction

In this chapter, we describe certain chemical microsensors, electronic *noses* (E-noses), and various radio ICs and their power sources and antennas used to wirelessly communicate data from remote chemical microsensors.

Silicon and MOX solid-state chemical sensors have been used to quantify numerous cations and anions in solution, for example, H^+ (pH), K^+ , Na^+ , Ca^{++} , Cl^- , I^- , HCO_3^- , CN^- , and SO_4^- , as well as concentrations (partial pressures) of gas molecules such as H_2 , O_2 , H_2S , NH_3 , CH_4 , SO_2 , CO_2 , CO , and many volatile organic compounds (VOCs) (Muller et al. 1991).

All solid-state chemical microsensors can be put into two major categories: *The first category* consists of devices that sense chemical analytes whose presence modulates the transport of electric charges (electrons, holes, ions) in the device. Examples of microsensors in this first category include chemFETs (chemical field-effect transistors), ion-selective field-effect transistors (ISFETs), ion-controlled diodes, Schottky diodes, thin-film tin oxide gas sensors, and chemiresistors. *The second category* includes all solid-state chemosensor devices not in the first category. These devices are based on a number of diverse technologies, including SAW oscillators, potentiometric gas sensors, and pyroelectric enthalpimetric sensors (Wohltjen 1984).

Electrochemical sensors can also be classified by their electroanalytical principles (Hierlemann and Baltes 2003, Janata 2003). These include, but are not limited to, the following:

1. *Voltammetric sensors*: These are based on the change in a device's I - V characteristic caused by the analyte. Generally, device current I is measured as a function of analyte concentration at varying voltages. *Amperometry* is a special case of voltammetry where the device voltage is held constant. ChemFETs, ISFETs, and Pd-Schottky diodes can be classified as voltammetric sensors.
2. *Potentiometric sensors*: These use the measurement of the open-circuit EMF of a sensing electrode pair (forming an electrochemical cell). Generally, the measured EMF is proportional to the logarithm of the concentration of the electroactive analyte and directly proportional to the Kelvin temperature of the cell (cf. the Nernst EMF equation).
3. *Conductometric sensors*: A small DC or AC potential is applied between two electrodes in a solution, and the conductance of the system is measured ($G = I/V$ siemens). The concentration of charge carriers determines the measured conductance.

An important application for gas-phase chemical sensors is in home and workplace smoke alarms. Here, gaseous combustion products (e.g., CO, CO₂, VOCs, and CH₄) above a threshold are detected. (For a comprehensive review of the gas detection technologies used in smoke alarms, see the tutorial paper by Warmack et al. 2012.)

11.2 Chemical Microsensor Designs

11.2.1 Silicon-Based Chemical Microsensors

11.2.1.1 Schottky Diode-Based Microsensors

A Schottky barrier diode (SBD) symbol and its static I - V curves (compared to a Si pn diode) are shown in Figure 11.1. Normally used in UHF electronic circuits because of its fast ON \rightarrow OFF times, the SBD architecture is also used for sensing low-molecular-weight gasses (Wohltjen 1984, Liu et al. 2004). Palladium is used for the anode metal in the SBD hydrogen gas sensor. The mechanism of operation depends on a change in the Pd work function caused by H₂ atoms adsorbing to the thin Pd layer and forming palladium hydride. Other gas-sensitive SBDs have also used Pd-TiO₂, Pd-CdS, Pd-ZnO, and PbS-Si metal layers to sense light-molecular-weight gasses such as H₂S, NH₃, SO₂, and CO, generally with poor selectivity. Typically, an SBD hydrogen sensor is packaged with a heater and an RTD temperature sensor so it can be run at a constant elevated temperature, enhancing its response time (Liu et al. 2004). A GaN metal semiconductor SBD can clearly detect ca. 100 ppm (H₂) in air at room temperature (Irokawa 2011).

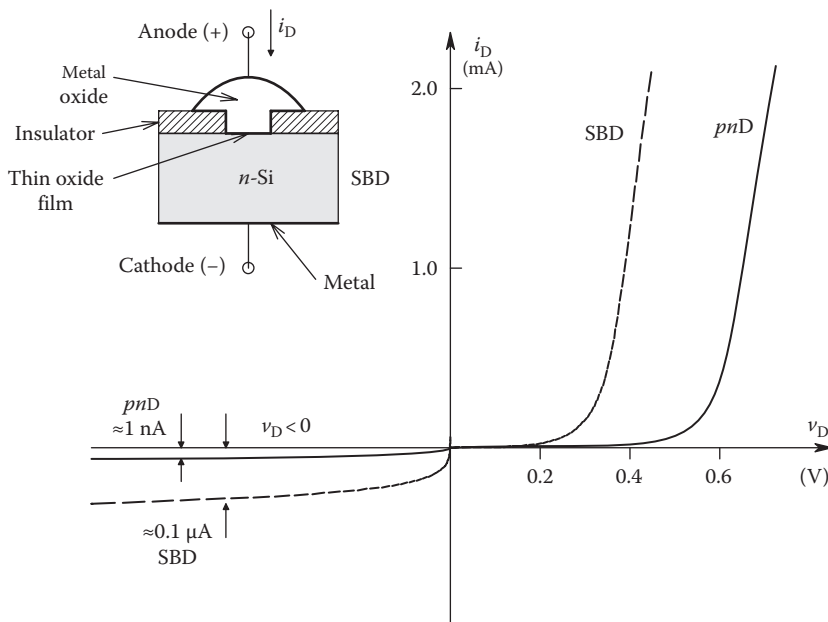


FIGURE 11.1

Diagram of an SBD. Its i_D vs. v_D curve is shown, along with the i_D vs. v_D curve for a typical, silicon, pn junction diode.

11.2.1.2 MOS ChemFET

The MOS chemFET (Wohltjen 1984, Senturia 1987, Wroblewski 2005) is also called the ISFET (Hierlemann and Baltes 2003). To understand how MOS chemFETs work, we will first describe the properties of a standard *p*-channel MOSFET used as an amplifier in electronic circuitry. The circuit symbol and defined voltages and currents for a *p*-MOSFET are shown in the upper left of Figure 11.2. In the MOSFET's saturation region, enhancement *p*-MOSFET drain current, i_D (also the source current), can be modeled by

$$i_D = -K[v_{GS} - V_T]^2 = KV_T^2 \left(1 - \frac{v_{GS}}{V_T}\right)^2, \text{ for } 0 < |v_{GS} + V_T| < |V_{DS}| \tag{11.1}$$

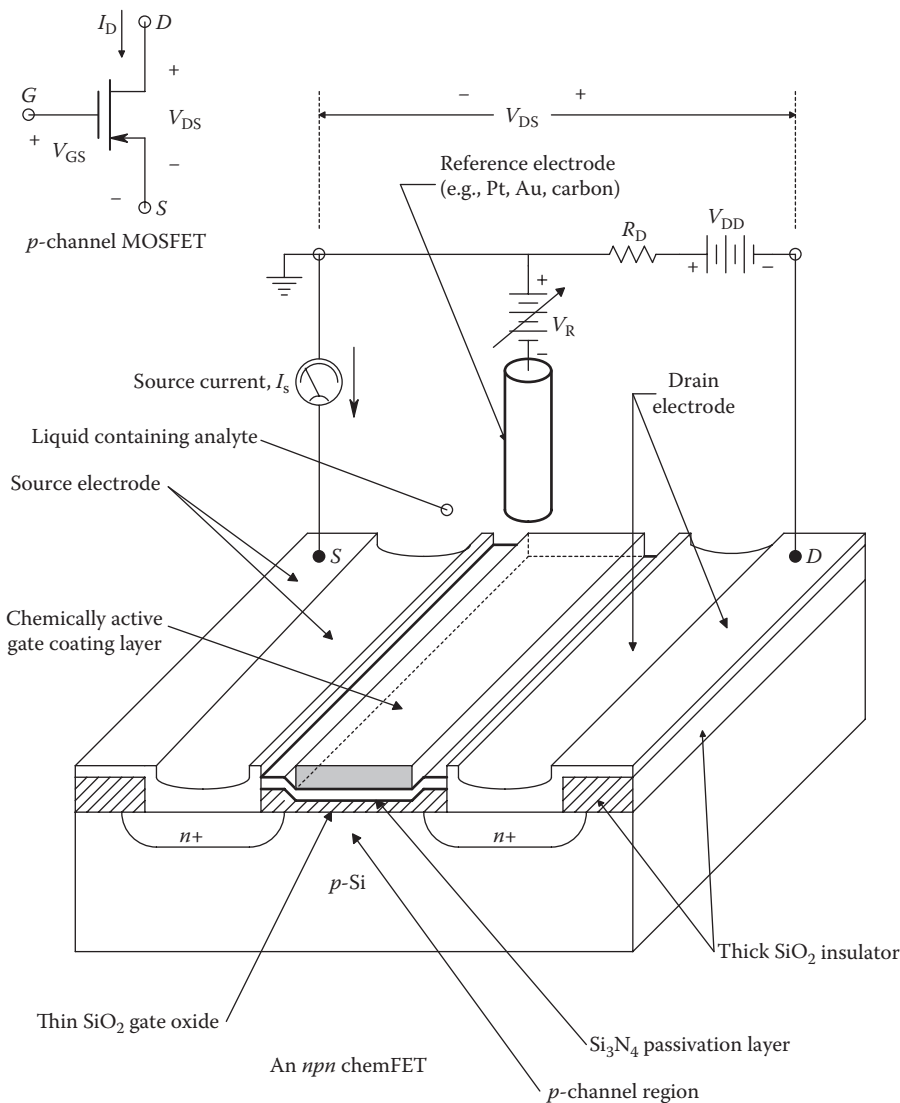


FIGURE 11.2 Diagram of an *n**p**n* chemFET with its basic biasing circuit. The inset shows the symbol for a *p*-channel MOSFET.

and in the MOSFET's ohmic region by

$$i_D = -K \left[2(v_{GS} - V_T)v_{DS} - v_{DS}^2 \right], \quad \text{for } 0 < |v_{DS}| < |v_{GS} - V_T| \quad (11.2)$$

where

v_{GS} is the gate to source voltage

K is the device's perveance in amps per volt²

By definition, $i_D \rightarrow 0$ at $V_T = v_{GS}$. V_T is the i_D cutoff v_{GS} value. The *saturation region* is where normal S-S MOSFET amplifier operation occurs, a region where i_D is relatively constant at fixed v_{GS} . The *ohmic region* is at low v_{DS} where i_D increases almost linearly with v_{DS} . The minus sign is used because i_D leaves the drain (enters the source) in a p -MOSFET. Other parameters (V_T , v_{GS} , and v_{DS}) are generally negative for a p -MOSFET. As an example, let $v_{GS} = -5$ V, $V_T = -3$ V, and the measured DC $i_D = -0.001$ A. The perveance K is thus found to be 0.00025 A/V². Note that i_D is defined as positive for n -MOSFETs. (See Northrop 2012, Ch. 2.)

A MOS chemFET hydrogen sensor generally has a Pd, Pd-alloy, Pt, or Ir gate electrode, similar to the metal electrode on an SBD. Again, H₂ gas molecules are adsorbed and diffuse rapidly through the thin metal (Pd) layer and absorb at the metal/SiO₂ interface, partly on the metal, partly on the oxide. Due to these absorbed layers, and the polarization phenomena at the interface, the drain current (I_D) is changed and the threshold voltage (V_T) is shifted. The V_T shift is proportional to the concentration of H₂ (the analyte) at the gate metal/oxide interface (Hierlemann and Baltes 2003). What is actually used to estimate the [H₂] is the change in chemFET drain (or source) current. As in the case of Schottky diode chemosensors, the Pd/alloy gate chemFET also is responsive to H₂S, NH₃, SO₂, and CO gasses, but there is interference between species, leading to poor selectivity. Detection of [H₂] at levels $\ll 1$ ppm is possible (Wohltjen 1984).

In the case of the ISFET used for liquid-phase ion sensing, the metal gate of the MOS chemFET is replaced by an aqueous electrolyte solution containing the analyte that is contacted by a reference electrode. (See Figure 11.2.) An electric field is set up between the reference electrode and an ion-selective material covering the ISFET's gate. The FET's I_D is changed by the potential at the gate/solution interface (v_{GS}) that in turn depends on the concentration of the specific ion sensed. Negative feedback can be used to keep $I_D([\text{ion}]) = I_D(\text{no ion})$ by electronically varying the reference electrode potential, V_R . Ions including H⁺ (pH), K⁺, Ca⁺⁺, Na⁺, Cl⁻, F⁻, and CN⁻ have been sensed using ISFETs (Wohltjen 1984).

ISFETs can be used to quantify the pH of a solution: a vapor-deposited gate layer of SiN, alumina, or oxynitride can be used to sense pH. ISFET gates can also be covered with organic ion-selective membranes such as polyurethane, silicone rubber, polystyrene, and polyacrylates containing ionophores to selectively sense K⁺, Na⁺, or Ag⁺ ions. Cell metabolism-induced pH changes have been quantified, and glucose concentration has been measured with ISFET sensors using a proton-producing electrochemical reaction (Hierlemann and Baltes 2003). It should be noted that several ISFET sensors for multiple analytes can be put on one MEM chip structure, along with A/D converters, RTD sensors, and a serial I/O interface.

ISFET analyte resolution (including the just noticeable difference [JND] in concentration and the detection threshold [DT]) is limited by noise and DC drift in I_D . There have been many papers on the solid-state physics governing noise in MOSFETs, but relatively few on noise in ISFETs (Jamasp et al. 1998, Jakobson and Nemirovsky 1999, Morgenshtein et al. 2002, Liu et al. 2011).

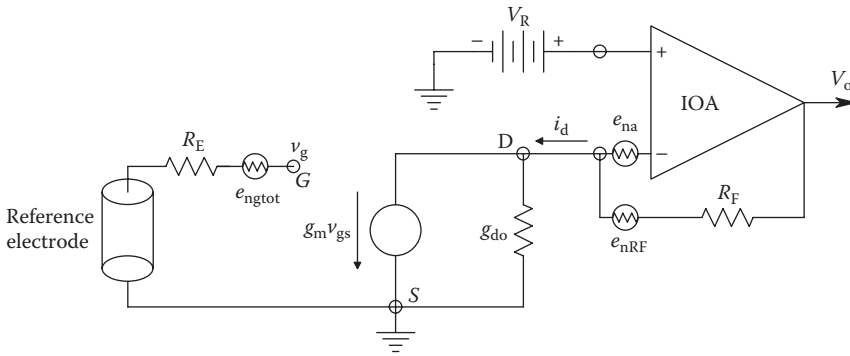


FIGURE 11.3
Schematic diagram of an ISFET and its noise sources.

Noise in ISFETs has several sources: There is noise external to the gate caused by the thermal agitation of charge carriers (ions) in solution; also, there is $1/f$ noise caused by ions interacting with the ISFET gate surface, the reference electrode surface, and test chamber surface. There also is an equivalent input noise voltage PDS from shot noise generated in the ISFET’s conducting channel.

It can be shown (BMF Wiki 2008) that there is an equivalent, referred-to-input (gate), two-source noise model for a saturated ISFET/MOSFET device; this model schematic is shown in Figure 11.3. The input (gate) noise voltage PDS has been shown to be

$$v_{ng}^2 = \frac{4kT\gamma g_{do}}{g_m^2} + \frac{FI_D^\alpha}{g_m^2 f} \text{ MSV/Hz}, \tag{11.3}$$

where

- γ is a bias-dependent factor of about 2/3
- I_D is the drain current magnitude at the ISFET’s operating (Q) point
- g_{do} is the SS, saturated channel conductance* at the Q point
- g_m is the ISFET’s SS transconductance at the Q point
- F and α are positive constants used to fit the $1/f$ noise data
- f is Hz frequency

The input current noise PDS has been shown to be given by

$$i_{ng}^2 = 4kT(2\pi f C_{gs})^2 \left[\frac{\delta}{(5g_{do})} + \frac{\gamma g_{do}}{g_m^2} + \frac{FI_D^\alpha}{(g_m^2 f)} \right] \text{ MSA/Hz}, \tag{11.4}$$

where C_{gs} is the gate–drain capacitance, $\delta \cong 1.33$, and the other parameters are defined for the input noise voltage PDS earlier. i_{ng}^2 is converted to an MS noise voltage/Hz in series with the gate by multiplying i_{ng}^2 by R_E^2 .

In addition to these device-generated noises referred to the input (gate), there is *external noise* arising in the electrolyte between the reference electrode and the gate surface. In general, this electrolyte will have an equivalent resistance, R_E , that is electrolyte concentration and

* $g_{do} = \partial i_D / \partial V_{DS}$ at Q. Q is at $V_{DS} = V_{Rv}$ and $V_{GS} = 0$.

temperature dependent. Thus, there will be an external thermal noise voltage in series with the gate, plus the electrochemical DC voltage generated by the analyte at the gate surface, Δv_{gs} . There will also be $1/f$ noise due to electrolyte ions interacting with surfaces. Thus, there will be a net source noise voltage PDS in series with the analyte's DC potential modeled by

$$e_{ns}^2 = 4kTR_E + \frac{D}{f} \text{ MSV/Hz}, \quad (11.5)$$

where D is a constant to fit the observed $1/f$ noise spectrum.

Thus, the DC electrochemical potential caused by analyte concentration in the seawater must be resolved above the total net gate input noise voltage, e_{ngtot}^2 referred to the ISFET's gate from ground, plus the OA's noise. The total MS noise PDS in series with the gate is thus

$$e_{ngtot}^2 = \left\{ \underset{\text{from solution}}{4kTR_E + \frac{D}{f}} \right\} + \left\{ \underset{\text{from gate noise voltage}}{\frac{[4kT\gamma g_{do} + FI_D^\alpha/f]}{g_m^2}} \right\} + R_E^2 \left\{ \underset{\text{from gate noise current}}{4kTG_E \times 4\pi^2 C_{gs}^2 f^2 \left[\frac{\delta}{(5g_{do})} + \frac{\gamma g_{do}}{g_m^2} + \frac{FI_D^\alpha}{(g_m^2 f)} \right]} \right\} \text{ MSV/Hz}. \quad (11.6)$$

Figure 11.3 illustrates an SS ISFET model for a grounded source, n -channel device connected to an OA transresistor; note the SS output voltage is $v_o = v_g g_m R_F$. In the SS, saturated ISFET model, g_m may be estimated from $g_m = \partial i_D / \partial v_{GS} \cong d [K(v_{GS} - V_T)^2] / dv_{GS}$ at the device's Q point. The FET's perveance, K , and its pinch-off voltage, V_T are needed to find g_m .

The OA is assumed to be ideal except for a white short-circuit input voltage noise, e_{na} . The OA's feedback resistor R_F is assumed to make white Johnson (thermal) noise. The battery V_R sets the summing junction's DC potential, hence the ISFET's drain, to $+V_R$. The DC $V_{GS} = 0$ V; hence, $I_D = I_{DSS}$. It is easily seen that the OA's output noise PDS, $S_{no}(f)$, equals

$$S_{no}(f) = \left[\underset{\substack{\text{MS noise from soln,} \\ \text{gate noise voltage} \\ \text{and current}}}{e_{ngtot}^2} \right] g_m^2 R_F^2 + \left\{ \underset{\substack{\text{white noise from} \\ \text{op amp}}}{e_{na}^2} (1 + R_F g_{do})^2 + \underset{\substack{\text{from feedback} \\ \text{resistor}}}{4kTR_F} \right\} \text{ MSV/Hz}. \quad (11.7)$$

To find the output MS SNR, we assume the analyte produces a small DC gate-source voltage, ΔV_{gs} . This is easily seen to produce a DC output voltage, $\Delta V_o = \Delta V_{gs} g_m R_F$. The mean-squared output SNR is thus

$$\text{SNR}_o = \frac{(\Delta V_{gs} g_m R_F)^2}{\int_{f_1}^{f_2} S_{no}(f) df}, \quad (11.8)$$

where $f_2 - f_1$ define a measurement noise bandwidth, B , which, for example, in the case of DC measurement with a noisy ISFET, might be $2.02 - 0.02 = 2.0$ Hz. $S_{no}(f)$ is defined in Equation 11.7.

Thus, it should be possible to predict the minimum detectable change in analyte concentration given a specified MS SNR_o .

Characterization of ISFET input noise and its minimization is an important research area (Hung et al. 1990, Liu et al. 2011).

11.2.2 Metal Oxide Gas Sensors

Modern laboratory analyses of gasses and VOCs have used the GC/MS. GC works because the transit times for different gasses and vapors through the chromatograph capillary tube are different, and although they are all injected at the same time in a bolus, they all have different affinities to the stationary-phase absorber material lining the tube, so that their velocities through the tube differ. The gas boluses are carried along inside the tube by an inert carrier gas, for example, argon, nitrogen, or helium. The different emerging analyte gas boluses are detected in the carrier gas by one of several types of gas molecule detector (see Northrop 2002, Sects. 8.2.7–8.2.8). The gas sample emerging from the GC detector is then passed to a mass spectrometer (quadrupole or magnetic) where the molecular mass of the specific eluent gas (or fragments thereof) sample is measured. Key parameters are the transit time for a sample in the GC tube at temperature T , the breadth of the sample peak detected, and the molecular mass read by the MS. Needless to say, a GC/MS is large, generally not field portable, and is expensive. The duration of a typical laboratory bench GC/MS run is between 10 and 100 min. Some short extra time is required for the system computer to arrange a data display graph and label and quantify the analytes. Thus, with the discovery and development of solid-state, MEM gas and vapor sensors, there has been a great emphasis to design small, field-portable, fast, accurate, less expensive analytical instruments to quantify gas analytes and vapors from VOCs. One approach, described in this section, is the use of MOXs, such as the n -type, including SnO_2 , Fe_2O_3 , ZnO_2 , and TiO_2 , which respond to oxidizing gasses, or the p -type, including nickel oxide or cobalt oxide, which respond to reducing compounds (Schaller et al. 1998, Schaller 2000). Also used are polymer-based sensors; all sensors generally exhibit a rise in conductance when the target gas analyte binds to its surface. As will be seen, MOX sensors will detect a number of gas analytes, but with poor selectivity. By manipulating sensor temperature and doping metals (e.g., Au, Pt, Pd), differential sensitivities can be improved.

11.2.2.1 Taguchi Gas Sensors

The first low-cost sensor for explosive gasses was invented in 1968 and patented by N. Taguchi in 1970 (US Patent No. 3,631,436) (Warmack et al. 2012). The basic Taguchi sensor consisted of two, nontouching, parallel coils of platinum–iridium wire, surrounded by a bead of SnO_2 n -semiconductor. The four leads from the two coils were used to suspend the bead (and coils) from pins protruding through a header. The sensor had a cylindrical package; it was surrounded by two layers of stainless steel gauze and measured 17 mm high by 16 mm in diameter (Watson and Tanner 1974). One coil was used as a heater and input electrode and the second as an output current (I_o) collector. Note that $I_o = V_B/R_S = V_B G_S$. A typical operating temperature for the Taguchi sensor bead was 300°C. In the presence of a reducing gas analyte (a gas that can be oxidized), cationic adsorption occurs that increases the apparent conductivity (G_S) of the bead. After removal of the analyte, adsorption of atmospheric O_2 again becomes the dominant process, and G_S decreases to the zero analyte value. Figure 11.4A illustrates schematically the original simple circuit for the Taguchi gas sensor threshold alarm. The presence of analyte above a threshold concentration lowered the resistance between the coil electrodes, allowing a buzzer alarm to sound. Figure 11.4B shows the simple DC circuit given by Watson and Tanner (1974) that gave quantitative output proportional to analyte concentration. The Watson and Tanner system was initially tested with petrol vapor (500 and 1000 ppm), carbon monoxide (1200 and 600 ppm), and ethanol

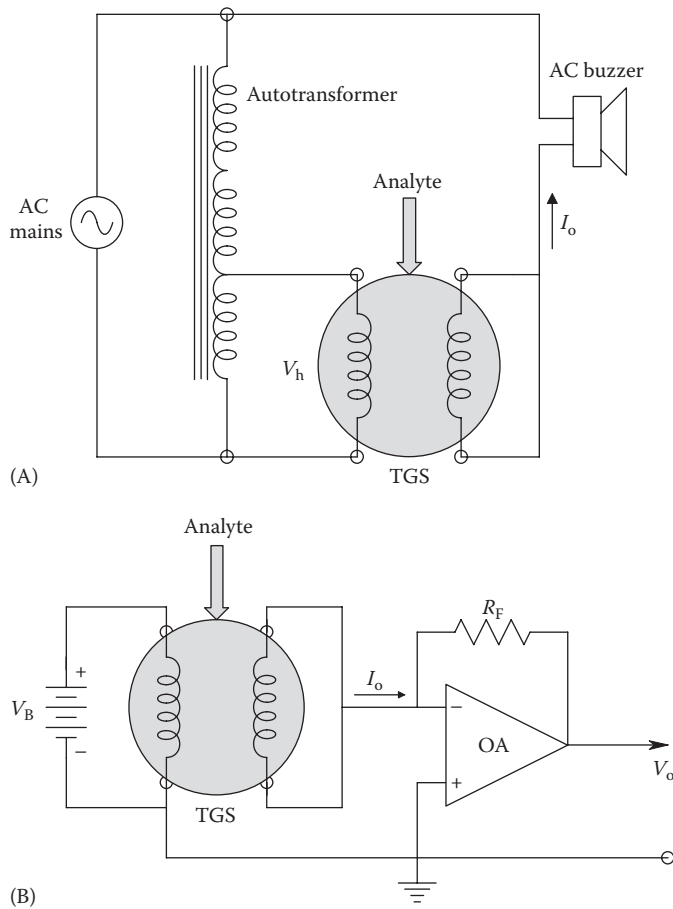


FIGURE 11.4

(A) Original Taguchi gas sensor circuit. It gave an audible alarm if the analyte gas concentration reached a threshold. (B) Taguchi sensor used to give a proportional, analog voltage output. A simple, OA transresistor circuit was used in this circuit designed by Watson and Tanner (1974).

vapor (200 and 400 ppm) to determine optimum operating parameters for the Taguchi sensor. Watson and Tanner favored reading out the DC output current as a function of analyte concentration. They observed that sensor I_o responses were almost linear for low analyte concentrations, but saturated for high concentrations. Note that $I_o = -V_o/R_F$ and $I_o = V_B G_S$, or $V_o = -V_B R_F G_S$. Watson and Tanner observed that for consistent readings, both the heater temperature (proportional to power dissipated) and the voltage across G_S must be regulated; they gave several simple circuits for these regulators. Warmack et al. (2012) observed that Taguchi tin oxide sensors have a general response to a reducing analyte given by an expression of the form $\Delta G/G_S = K[C]^\alpha$ in the low-concentration region of the analyte, $[C]$. At higher concentrations, the sensor saturates, and the exponent α drops rapidly.

11.2.2.2 Modern Metal Oxide Gas Sensors

A modern MEM tin oxide (SnO_2) gas sensor is relatively simple. It consists of a resistance heater on top of which is a thin film of SnO_2 to which electrical contact is made by two high-conductivity leads. Unlike the original Taguchi design, the heater is separate; it is

not used as an electrode in the sensor material. When the tin oxide film is exposed to the analyte gas (e.g., hydrogen), the gas molecules are adsorbed by the film and increase its conductivity, G_S . This conductivity increase can be sensed with a bridge or active circuit, or as ΔV_o , and related to the analyte gas' species and partial pressure. In a study by Guo et al. (2007), sensor G_S was driven by a constant current source (I_S), and the open-circuit output voltage, $V_o = I_S/G_S$, was read differentially. One group (Fort et al. 2003) reported on using an AC I_S and measuring the vector impedance Z_S of the SnO_2 sensor over a frequency range from 0 to 15 MHz. They concluded that impedance spectroscopy did not show a significant improvement in analyte discrimination over their chemical transient analysis.

Early studies with tin oxide gas sensors focused on the sensor's response to a single gas species, the ultimate goal being to realize an array of sensors to sense a number of simultaneously occurring gas analytes. SnO_2 -based gas sensors generally respond to a wide number of target gasses, but also exhibit a low level of selectivity to a given analyte. In communications terms, there is *cross talk* between analytes.

Figure 11.5A illustrates the construction of a typical thin-film tin oxide gas sensor. A resistance heater is used to raise the operating temperature of the sensor to ca. 350°C.

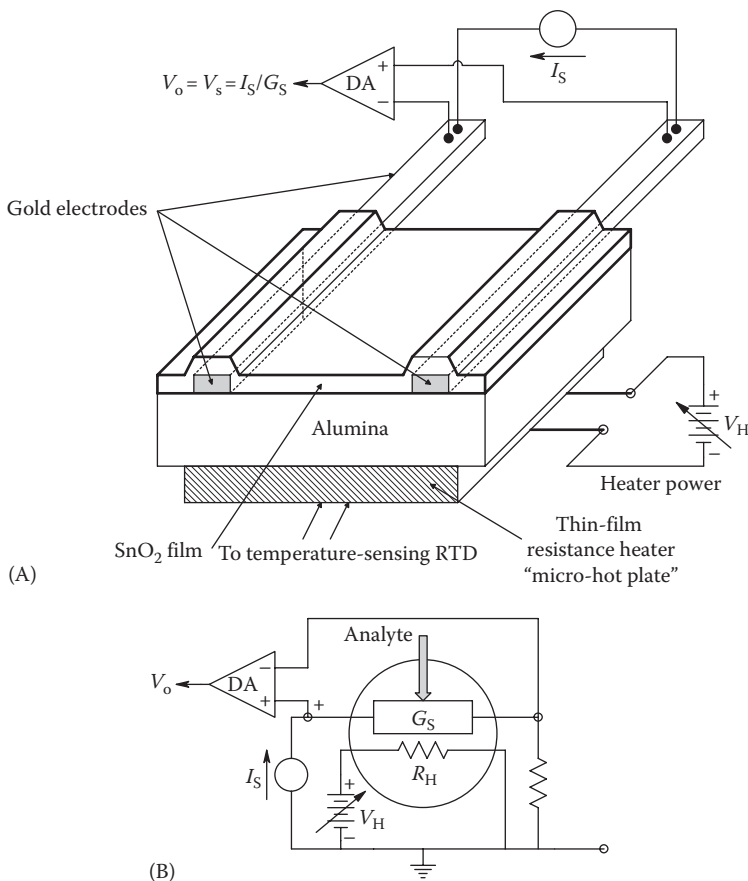


FIGURE 11.5

(A) Thin-film, tin oxide gas sensor. Bound analyte changes the sensor's conductance that is sensed by measuring the voltage across the sensor while passing a constant current, I_S , through it. (B) Schematic of the readout circuit. R_H is the heater resistance. Sensor temperature is proportional to V_H^2/R_H .

A 1993 paper by Suehle et al. showed that a SnO_2 film at 350°C increased its conductance from about 0.5×10^{-6} S to a peak of ca. 2×10^{-6} S when exposed to pure H_2 gas at 1.3 Pa pressure (a factor of 1:4 change). The sensor had a response time (time to reach SS, higher conductance) of about 200 s to pure H_2 (much too slow to be practical). When pure O_2 at 1.3 Pa was introduced, it oxidized the H_2 bound to the SnO_2 and reduced the conductivity to the 0 H_2 level of 0.5×10^{-6} S in about 200 s. (Note: $1 \text{ Pa} = 1 \text{ n/m}^2 = 1.45 \times 10^{-4} \text{ psi}$.)

Guo et al. (2007) reported on the performance of a 4×4 (16) element array of SnO_2 gas sensors in which the SnO_2 layer of each sensor was modified using metal additives and ion implantations. Minute amounts of metals such as Pt, Pd, and Au were used to modify the specific sensitivities of certain sensors in the array to a certain analyte. All 16 sensors were exposed to an analyte simultaneously and at the same temperature (300°C). Air was used to purge the array and reset the sensitivities of the 16 sensors. Sensor sensitivities could be calibrated by adjusting the constant sensor current, I_s , for each sensor. (See Figure 11.5B.) A DA was used to read out the output voltage, $V_o = V_s = I_s/G_s$. Guo et al. tested their array with H_2 (50, 100, 200 ppm), CH_4 (500, 3000, 5000 ppm), CO (25, 100, 250 ppm), and $\text{C}_2\text{H}_5\text{OH}$ vapor (50, 100, 200 ppm).

Of note was their use of *radar diagrams* that are polar plots in which a circle was divided into 16 equally spaced radii, numbered clockwise from 1 to 16. The distance from the center of the circle along each radius was proportional to that numbered sensor's SS response to a given gas and concentration. The tips of the 16 response vectors were connected to make a continuous contour that for each gas had a unique 2D shape. For example, the methane contours were *circular* (had nearly equal radii). Ethanol vapor produced a four-pointed *star*, with points (maximum responses) for sensors 2, 6, 10, and 14. Other sensors had much smaller responses to ethanol vapor. Thus, the shape of the radar diagram contours gave information as to the particular gas, and the radial distances were proportional to its concentration.

In a more recent publication, Far et al. (2008) investigated the performance of a 16-element tin oxide gas sensor array as the array temperature was modulated. They claimed: "Temperature modulation is shown to increase the number of our sensors from 16 physical sensors (integrated on-chip) up to 12,000 virtual sensors. This will enable the emulation of a very large number of sensors typically found in biological systems." This suggests $12,000/16 = 750$ response types/sensor for a presumed 750 different temperature steps.

Arrays of differentially doped tin oxide gas sensors can be used to make E-noses that can identify several VOCs at the same time. Fort et al. (2003) reported on their work with an eight-element tin oxide sensor array with the goal of developing analyte selectivity for gas mixture classifications. With wine characterization as a goal, they used aqueous test mixtures of 1%wt ethanol, 0.05%wt ethyl acetate, 0.08%wt glycerol, 0.05%wt isoamyl acetate, 0.05%wt linalool, and 0.05%wt hexanol, one at a time and in pairs. (Linalool is a commercial odorant/flavorant, aka *3,7-dimethyl-1,6-octadien-ol*, found naturally in cinnamon, bergamot, saffras, and orange blossom.) Three classification protocols were tested: (1) a transient analysis of $R_s(t)$ for various analytes at constant temperature, (2) SS AC vector impedance Z_s of the sensor for various analytes at constant temperature, and (3) $R_s(T)$ for sinusoidally programmed sensor temperature T with periods P of 40 and 180 s. These researchers found that the programmed temperature method, in

which $T \cong 220\{1 + 65 \sin [(2\pi/P)t]\}$ °C, gave better discrimination results than measurements using chemical transients or AC impedance. There was better separation of a given sensor's responses to the test analytes at the beginning and end of each sinusoidal cycle, presumably when the sensors' temperature returned to the baseline 220°C. There was a single peak in the responses to the test analytes over one heating cycle, but poor resolution at the peak between analytes.

In a 2012 paper, Chutia and Bhuyan examined the dynamics of two Figaro MOS/SnO₂ gas sensors (TGS 2610 and TGS 2620) to pulse inputs of various VOC analytes. The SS response of the sensors followed a power law relation, $R_s = K [C]^{-\alpha}$, where R_s is the SS sensor resistance, $[C]$ is the concentration of the gas, and α and K are constants that depend on the heater temperature and the type of gas. They tested the vapors of *ethanol*, *1-butanol*, *benzene*, *isopropyl alcohol*, and *isoamyl alcohol*, all at 500 ppm in air. They found that both the shape (rise and fall time) of the transient responses and the SS R_s value well characterized their sample gasses.

There is an extensive literature on tin oxide gas sensors. Their lack of analyte selectivity is being overcome by taking advantage of the fact that their selectivity (discrimination between gasses in a mixture) often changes markedly with temperature and the fact that rise times, fall times, and SS R_s are also good identifiers when a transient input is given to a sensor.

11.2.2.3 CO₂ Microsensor Based on Tin Oxide Doped with Copper Oxide

Xu et al. (2009) developed a novel CO₂ gas-sensing material, that is, nanocrystalline tin oxide (SnO₂) doped with cupric oxide (CuO). The sensor consisted of pairs of platinum interdigitating *fingers* ca. 1 mm long, 30 μm wide, spaced 30 μm, and microfabricated on a quartz substrate 250 μm thick. The fingers were chemically plated with SnO₂ and then doped with CuO. The interdigitated electrode area was 1.1 mm × 0.99 mm. The sensor had 11 pairs of interdigitating fingers. The presence of CO₂ gas caused a decrease in the electrical resistance between the two sets of interdigitated electrodes. For example, a prototype sensor had a doping level of 1:8 CuO/SnO₂ (molar ratio); its resistance was measured with various concentrations of CO₂ in air at 450°C. Fig. 2 (Xu et al. 2009) showed the sensor resistance as ca. $6.7 \times 10^8 \Omega$ (670 megohms) for pure air, ca. $5 \times 10^7 \Omega$ for pure N₂ gas, $5.2 \times 10^8 \Omega$ for 1% CO₂ (10,000 ppm) in air, $4.7 \times 10^8 \Omega$ for 2% CO₂ in air, $4.1 \times 10^8 \Omega$ for 3% CO₂ in air, and $3.8 \times 10^8 \Omega$ for 4% CO₂ in air.

One obvious way of reading out the CO₂ sensor of Xu et al. is with a Wheatstone bridge. Measurement of hundreds of megohms generally requires shielding and electrometer amplifiers to sense the null.

The current Earth's atmospheric CO₂ concentration is ca. 400 ppm, 1,000 ppm is the recommended upper level for indoor air quality (above 1,000 ppm causes fatigue and reduced concentration), 5,000 ppm is the limit of CO₂ in the workplace, and ca. 40,000 ppm (4%) is the [CO₂] in the human breath, assuming 20 liters of CO₂ are exhaled per hour.

Three to eight percent [CO₂] causes increased respiration rate and headaches; >10% results in nausea, vomiting, and unconsciousness; and over 20% produces rapid unconsciousness and death.

Additional discussion of CO₂ measurement by nondispersive IR spectroscopy may be found in Chapter 7.

11.3 Electronic Noses

11.3.1 Introduction

An E-nose is an electronic chemosensor system designed to detect and quantify analyte vapors, generally in air, often at extremely low concentrations. Early E-noses were bench-top systems; they rapidly evolved into field-portable and handheld systems such as the Cyranose® 320. The first practical application of an E-nose was the detection of land mines by the US Army during the Vietnam War. The concept of a chemical sensor array for odor classification was described for the first time by K.C. Persaud and G.H. Dodd in 1982 (Leake 2006). Applications of E-noses have rapidly evolved from the detection and measurement of toxic and hazardous VOCs to medical diagnostic applications and use in the food industry. They have been used to sense the VOCs in volcanic eruptions, leaks in gas mains, and groundwater purity. More recently, they have found application in medical diagnosis, specifically, the detection of cancers, various respiratory diseases, urinary tract infections, and bacterial vaginosis. Interested readers should consult the exhaustive review of the technology of E-noses and their applications by Wilson and Baietto (2009).

11.3.2 Rationale for Artificial Noses

The main subsystems of an E-nose are the following: (1) An array of chemical sensors has broad specificities to the presented analyte group (odors). Broad specificity is obtained, for example, by varying sensor operating temperatures and doping the sensor materials; (2) Sensor outputs are conditioned electronically; and (3) A computer discrimination system examines the sensor outputs (the odor's *smell print*) using various algorithms and, from an a priori knowledge of possible analyte components, predicts which are present and in what concentrations. For a given odor, there will be a specific pattern of sensor outputs from the array. An E-nose's analysis system can be *trained* to process a specific pattern response and identify and quantify the specific odor having a number of component analytes, such as decomposition or coffee. Haddad et al. (2010) trained an E-nose (the MOSES II) to recognize the general property of *odor pleasantness* using odors it never *smelled* before. Their system had over 90% accuracy at discriminating between categorically *pleasant* and *unpleasant* odors. The same results were obtained between two cultures (native Israeli and native Ethiopian) without retraining the apparatus. (The MOSES II E-nose uses an array of 16 sensors: 8 MOX and 8 quartz resonator [microbalance] sensors.)

The need for reliable, *field-portable* instruments that will rapidly and reliably sense analytes from contraband drugs; explosives of all sorts (e.g., semtex, C4, PETN, and nitroglycerine), including those in luggage and in land mines; as well as organic pollutants (VOCs) in soil, water, and air is a driving force for the development of measurement systems that will detect and measure the minute concentration of certain VOC analytes in the air when accompanied by other, irrelevant analytes. E-noses can be used to identify, as a group, the mixture of components that together form a specific odor, such as the bouquet of a wine. An E-nose can be used for the grim purpose of locating buried corpses after natural disasters or bombings and to evaluate the freshness and quality of meat, fish, fruit, dairy products, vegetables, and beverages (Schaller et al. 1998, Strike et al. 1999, Casalnuovo et al. 2006, Leake 2006).

Note that many less-portable or fixed technologies exist for VOC analysis, particularly in airports, to detect nitrates and other explosives. These include GC/MS systems and spectrophotometers. Even these instruments have been reduced in size so they can be brought into the field; they generally are not hand portable, however.

In the following sections, we will describe the state of the art in man-made VOC sensors and the overall design of the E-nose using artificial neural networks (ANNs).

In addition to E-noses that operate on gas-phase analytes, the less-known E-tongues detect liquid-phase, nonvolatile analytes using chemosensors.

11.3.3 Sensors for E-Nose Applications

In man's attempt to design artificial olfactory systems, it soon became evident that certain classes of sensors (e.g., conductive polymers, MOXs) had broad sensitivities to various VOCs (analytes). Thus, a sensor in an array would not only respond maximally to a specific analyte but also produce smaller outputs in response to other VOCs. In some respect, this behavior mimics a live olfactory receptor array, which, viewed at the receptor level, gives ambiguous responses. In mammals, the *in vivo* elimination of this *cross talk* is done neurally by the olfactory bulb and other parts of the brain (cf. Ch. 34 in Kandel et al. 1991). As you will see in Section 11.3.4, man has relied on computer algorithms and ANNs to sharpen the selectivity of his E-noses to the desired vapor-phase analytes.

In order to perfect an E-nose that will compete with the sensitivity of dogs, much effort has been directed to improving the sensitivity and molecular specificity of *nose* sensors to a desired vapor-phase, odorant molecule. In this section, we examine some of these sensors. There are currently several sensor technologies competing for the high [(specificity \times sensitivity)/cost] index for E-nose front ends. These sensors are summarized in Table 11.1.

TABLE 11.1

E-Nose Sensors: Some Sensors Useful in E-Noses

Sensor Type	Operating Mechanism	Threshold Sensitivity	Advantages	Disadvantages
CPF	Adsorbed analyte causes ΔG .	>10 ppt >5 ppm more common	Inexpensive, microfabricated, operates at room temperature	Very sensitive to humidity.
MOXs	Adsorbed analyte causes ΔG .	5–500 ppm	Inexpensive, microfabricated	Operate at high temperatures; big cross sensitivities.
MOSFETs	Gate charge altered by adsorbed analyte.	ppm	Can be integrated with interface ICs	Analyte bound to gate must be removed. Noise, poor sensitivity.
SAW sensor	Analyte absorption causes phase or frequency Δ .	1 pg mass change	High sensitivity	Complex interface signal conditioning.
Surface plasmon resonance	Photon absorption due to $\Delta\epsilon$ from analyte.	>0.01 ppb	High sensitivity	Expensive. Bound analyte must be removed to recycle.
Fluorescence	Analyte molecules tagged with fluorophore molecule.	ppb range	High sensitivity, high specificity	Expensive.

TABLE 11.2

CPF Detection Limits: Detection Limits for Various Gas-Phase CPF Materials + Dopant to Certain Analytes

Analyte	CPF Material + Dopant	Detection Limit	Sensor Type
NH ₃	PAni + SWNT	50 ppb	Chemiresistor
NO ₂	PPy + PET	<20 ppm	Chemiresistor
	PTh + CuPc	4.3 ppm	Chemiresistor
	PAni + In ₂ O ₃	<0.5 ppm	SAW device
HCl	PAni + FeAl	0.2 ppm	Chemiresistor
H ₂ S	PAni + heavy metal salts	<10 ppm	Chemiresistor
CO	PAni + FeAl	10 ppm	Chemiresistor
	PAni + In ₂ O ₃	<60 ppm	SAW
Water	PAni	<25 ppm	Chemiresistor
C ₂ H ₅ OH	PAni + Pd	<1 ppm	Chemiresistor
Methane	Poly(3-methylthiophene) +	Several ppm	Chemiresistor
Halide	MWNT		
Acetone	PTh copolymer	200–300 ppm	Chemiresistor
Toluene	PTh copolymer	20 ppm	Chemiresistor
Butylamine	Poly(anilineboronic acid)	10 ppb	Chemiresistor

Source: Bai, H. and Shi, G., *Sensors*, 7, 267, 2007.

The first class of VOC sensor we shall examine is the *conductive polymer film* (CPF). Typical CPFs include, but are not limited to, *polyacetylene* (PA), *polyaniline* (PAni), *polypyrrole* (PPy), *polythiophene* (PTh), *polyphenyl vinylene* (PPV), *poly(3-hexylthiophene)* (P3HT), and *poly(3,4-ethylene-dioxythiophene)* (PEDOT). CPFs have rapid response times, run at room temperature, and are less expensive to make than MOX MEM gas sensors. The data in Table 11.2 illustrate the detection limits for certain analytes and various CPF sensors.

In a review paper on CP analytical applications, Gupta et al. (2006) described a broad range of CPF sensors using bound enzymes and their principal analytes. These systems work in aqueous media. They are shown in Table 11.3. Some disadvantages of CPF sensors include their low selectivities, lack of long-term stability, and irreversibility. Their long-term loss of sensitivity may be due to the loss of doping in the CPF material, long-term oxidation, and gradual water adsorption from vapor in the air (Bai and Shi 2007). These problems generally do not affect the MOX sensors.

PEDOT-polystyrene sulfonate polymer composites doped with carbon particles have been used at CalTech to develop VOC sensor arrays (Sotzing et al. 2000a); PAni/carbon composite has also been examined for the same purpose (Sotzing et al. 2000b). Other CPF sensors have used films based on polymers of *indoles*, *aniline*, *thiophenes*, *p-phenylene vinylene*, and their derivatives (Schaller et al. 1998, Strike et al. 1999, Tan et al. 2001, Bai and Shi 2007, Yoon and Jang 2009, Pankaj et al. 2012).

A typical CPF sensor array contains up to 32 different polymer membranes having different, but overlapping, analyte sensitivities. The polymers are doped with conductive colloidal particles of carbon, gold, or silver. When a conductive particle-doped polymer film sensor is exposed to a VOC analyte, the analyte is adsorbed on the sensor's surface; it causes a physical swelling of the polymer membrane. The swelling causes the electrically conductive matrix to become less dense, in turn lowering the longitudinal

TABLE 11.3

Biosensors Using CPFs Plus Enzymes Reacting with Analytes in Aqueous Solution

Analyte	Enzyme	Polymer	Detection Method
Glucose	Glucose oxidase	PPy	A, P
		Poly(<i>N</i> -methyl pyrrole)	A
		PAni	A
		Polyindole	A
D-Alanine	D-Amino acid oxidase	PPy	A
Atrazine	Tyrosinase	PPy	A
Cholesterol	Cholesterol oxidase and cholesterol esterase	PPy	A
Choline	Choline oxidase	Substituted PPy	A
Glutamate	Glutamate dehydrogenase	PPy	A
Fructose	Fructose dehydrogenase	PPy	A
Hemoglobin	Pepsin	PAni	C
L-Lactate	Lactate oxidase	Polyphenylene diamine	A
		Lactate dehydrogenase	A
		PPy-polyvinylsulfonate	A
Lipids	Lipase	PAni	C
Phenols	Tyrosinase	Substituted PPy	A
Urea	Urease	PPy	A, P, C, Cap.
Uric acid	Uricase	PAni	A
Triglycerides	Lipase	PAni	C

Note: Detection methods: A = amperometry, P = potentiometry, C = conductometry, Cap. = measurement of capacitance between electrodes.

conductance of the membrane (the polymer is basically an insulator). Unfortunately, other physical factors such as temperature and humidity also affect the conductance, requiring that the sensor array be held at a constant temperature and humidity. The doped polymer films used at CalTech are on the order of 1 to 0.1 μm thick. The polymer gas sensors generally work at temperatures $<50^\circ\text{C}$, hence are sensitive to high humidity (Schaller et al. 1998). Thus, the humidity of air containing the analyte(s) should be controlled (held constant).

Response times for the CalTech CPF sensors to reach SS output, given a suddenly applied analyte, are reported to range from ca. 0.1 s to >100 s and depend on the specific VOC, the polymer, the temperature, and the sensor thickness (Lewis 2000). Lewis illustrated a sensor responding to 880 ppb of nitrotoluene vapor, the signal well above the noise, and a sensor sensitivity of 10 ppt for biogenic amines. Note that the analyte-caused ΔR of the CPF sensor can easily be sensed with a Wheatstone bridge or an OA transresistor (cf. Sections 4.2 and 6.3). An early study of human olfaction vs. CPF E-nose was described by Doleman and Lewis (2001). They found that their highest CPF E-nose sensitivity was ca. 10 times that of humans and could sense on the order of 0.1 ppm of the analyte tested.

A commercially available, handheld E-nose using the CalTech-developed CPF sensors is now marketed as the Cyranose 320®; this E-nose uses a 32 CPF sensor array (Li 2000). Each of the 32 CPF sensors in the Cyranose is made of a conductive carbon black material that is homogeneously blended throughout a specific, nonconducting polymer and

then deposited as a thin film on an alumina substrate. Contact is made with the film with two wires, forming a chemiresistor. Polymers with a wide range of properties are used to allow the array to distinguish between many types of vapors. The outputs of the 32-chemiresistor array are conditioned and input to an odor discrimination system. When a sensor responds to a specific analyte vapor, the film expands, raising the sensor resistance as a function of analyte concentration. This $\Delta R/R_0$ is measured (R_0 is the baseline resistance at temperature T_0 , and ΔR is the SS resistance increase caused by the analyte.) $\Delta R/R_0$ is seen to rise exponentially when exposed to a step of analyte concentration. The time constant is on the order of tens of seconds. The Cyranose[®] 320 processes the 32-element sensor array output using principal component analysis (PCA) for outlier diagnosis and four onboard supervised algorithms for building a model and predicting the unknowns:

- K-nearest neighbor (KNN)
- Soft independent modeling of class analogy (SIMCA)
- Fisher linear discrimination (FLD)
- Canonical discriminant analysis (CDA)

In order to run the statistical routines discussed earlier to identify the odor, first, PCA is run and then the operator must take 10–20 samples of the odor under measurement.

The second class of VOC analyte sensor is the MOS film (Moseley et al. 1991, Pitcher et al. 2003). Taguchi pioneered the work on tin oxide gas sensors in 1970. This class of sensor is operated at elevated temperatures (175°C–500°C). A wide range of MOS materials were investigated by Moseley et al. including the oxides of Ga, In, Mo, Nb, Sn, Ta, Ti, W, and Zn. Also examined for their analyte specificity were binary oxides such as CaSnO_3 , ternary oxides such as $\text{SrFe}_{0.8}\text{Mn}_{0.2}\text{O}_3$, various niobates such as BiNbO_4 , certain tantalates, perovskites, bronzes, and pyrochlores. Not all materials showed a resistance increase in response to a specific analyte; in fact, some showed a resistance *decrease* for certain VOC analytes. At 500°C, a $\text{BaSn}_{0.9}\text{Zr}_{0.1}\text{O}_3$ sensor prepared as a porous film 50 μm thick deposited on an alumina substrate showed responses (as $+\Delta R$) to Cl_2 , NO_2 , and SO_2 at 1% concentrations. It also detected H_2S at <2 ppm. Curiously, the $+\Delta R$ of this sensor increased monotonically to $[\text{H}_2\text{S}]$ up to 30 ppm, and then at ≥ 100 ppm, ΔR changed sign.

Moseley et al. (1991) concluded as follows:

The expectation is, and observation confirms, that most [MOX] materials show a response to a range of gases, and that spectacular selectivity is not commonly to be found. Even so, there is plenty of useful variability among the different materials [tested], and in some cases very great selectivity.

One can speculate that such selectivity tends to occur under conditions where the material is in a transition zone between n-type and p-type behaviour, where the effect of variations in the surface oxygen species has little effect on the conductivity and where the effect of specific surface coordination reactions might therefore be observed.

An experimental E-nose using nine various Figaro tin oxide sensors with a temperature and a humidity sensor was described by Keller et al. (1995). Sensor outputs were the nine inputs to an ANN that eliminated ambiguities and allowed the system to discriminate between the test substances: none (control), acetone, ammonia, isopropanol, lighter

fluid, and vinegar. Binary mixtures of the test substances were also presented: ammonia + isopropanol, ammonia + lighter fluid, ammonia + vinegar, and isopropanol + vinegar. The fuzzy ARTmap ANN algorithm was found to give the best detection performance, averaging 93.4% correct over all its trials.

The third class of VOC analyte sensor is the insulated gate MOSFET (aka IGFET). Figure 11.6 illustrates the basics of a simple n -channel MOSFET DC amplifier. When operating in the saturated channel mode, the DC drain current, I_D , is seen to increase as the gate–source (actually, the gate–channel) potential increases positively. When the MOSFET has depletion doping, $V_{GS} = 0$ produces a constant drain current, I_{DSS} . Thus, when the gate is shorted to the source, the MOSFET behaves like a DC current source. This behavior is illustrated in Figure 11.7. (MOSFETs can be made from doped silicon or gallium arsenide [GaAs].)

A chemical reaction at the gate that alters the gate charge can change V_{GS} and thus will change I_D . This ΔI_D is sensed; it is generally a monotonic function of the analyte concentration. For an effective MOSFET sensor, the induced ΔI_D must be resolved above noise and drift. Chemically sensitive MOSFETs are generally called chemFETs. They can be subdivided into ISFETs (ion-selective field-effect transistors), SURFETs (surface-modulated field-effect transistors), MEMFETs (membrane field-effect transistors), and POLFETs (polymer-covered

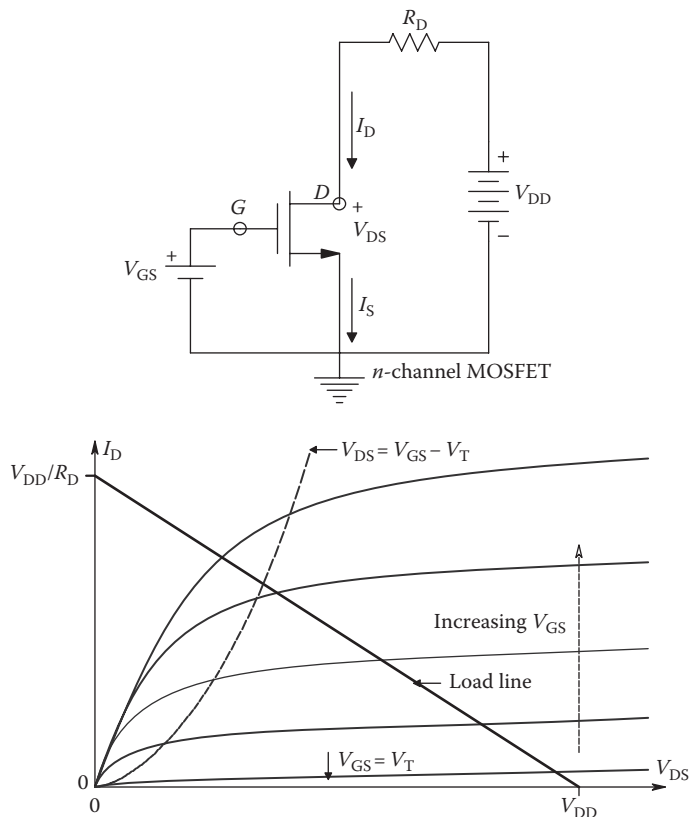
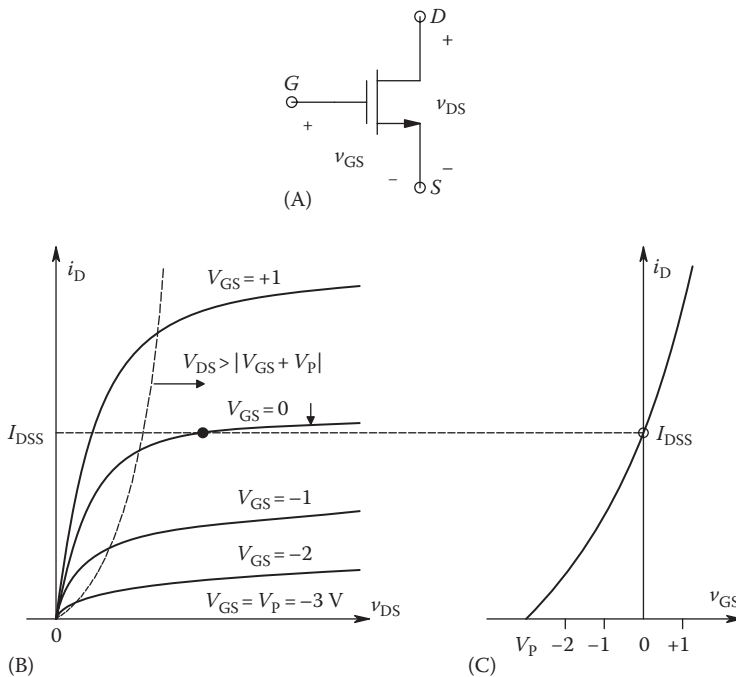


FIGURE 11.6 Simple, grounded source, n -channel MOSFET amplifier, showing its $i_D(V_{GS}, V_{DS})$ curves and load line.

**FIGURE 11.7**

(A) An *n*-channel MOSFET. (B). MOSFET operating curves. (C). MOSFET's i_D vs. v_{GS} curve for $V_{DS} > |v_{GS} + V_p|$. Note: $i_D = 0$ for $v_{GS} \leq V_p$.

gate field-effect transistors), depending on the details of their construction and application. (See the review papers by Wroblewski et al. [2004], Wroblewski [2005], and Khanna [2007].) Most chemFET designs have been used to sense analytes in aqueous solution.

In one embodiment of a chemFET gas sensor, the gate is covered with a thin polymer membrane; such a gas sensor is called a POLFET. The polymer can be electrodeposited or sprayed on the gate (Briand 2001a,b). POLFET sensors are noisy, they have relatively high $e_{na}s$ (cf. Section 3.1), and they also respond to temperature changes and stray ions in their gas input, all bad sources of uncertainty when making what are essentially DC measurements. In another embodiment of a gas-sensing chemFET, a porous (gas-permeable) metal gate is vapor deposited on a thin catalyst layer overlying the silicon oxide insulating the FET's channel. Gas penetrates the porous metal and undergoes some redox reaction in the catalyst, liberating electrons, protons, or ions that affect the gate-channel potential, hence I_D . Thus, MOSFET gas-phase sensors have comparatively poor sensitivity, generally in the ppm (Nagle et al. 1998). ChemFETs do offer the advantage of being able to be made by IC fabrication processes, however. To cancel DC drift in I_D , they can be used differentially, one device not being exposed to the gas analyte. It appears that gas-sensing chemFETs have not been the sensor of choice in commercial E-nose products (Nagle et al. 1998).

The fourth class of VOC analyte sensor we will examine is the SAW sensor. Figure 11.8 illustrates a simple SAW device. The SAW device is built on a thin piezoelectric material slab, usually quartz or LiNdO_3 (Pallàs-Areny and Webster 2001). A surface area between the electrodes is coated by a polymer film or other chemical having affinity for the VOC analyte. Electrodes at one end of the slab are excited by an oscillator, generally in the range of hundreds of MHz. The excitation induces ultrasonic mechanical vibrations in the piezomaterial that

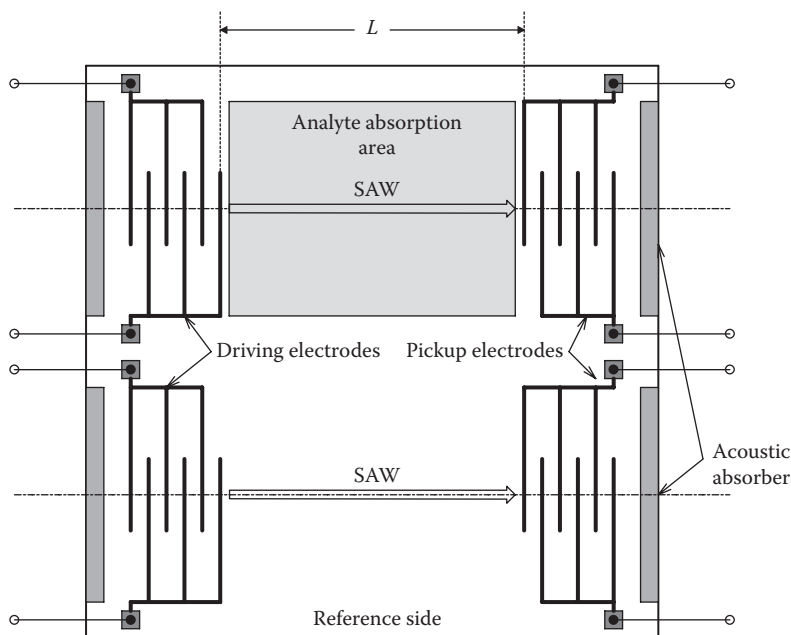


FIGURE 11.8
Differential SAW chemosensor. See text for description of operation.

propagate on its surface as Rayleigh (surface) waves, not unlike waves on water. To prevent wave reflections at the two ends of the piezoslub, an absorbing material is deposited, as shown in Figure 11.8. When the Rayleigh waves reach the electrodes at the far end of the coated piezoslub, a second set of electrodes picks up a vibration-induced voltage. The Rayleigh surface waves propagate with a finite velocity. When the analyte molecules are absorbed by the polymer, its mass is increased. The mass increase causes a fractional increase of the propagating Rayleigh wave's velocity. For low values of analyte concentration, this increase can be written as

$$v(M) = v_0 + \alpha \left(\frac{M}{A} \right) \tag{11.9}$$

or

$$v([C]) = v_0 \left(\frac{1 + \beta[C]}{v_0} \right), \tag{11.10}$$

where

v_0 is the surface wave propagation velocity in the absence of absorbed analyte

α and β are constants

(M/A) is the mass/area of the analyte absorbed on the surface

$[C]$ is SS analyte concentration over the membrane

Thus, the SAW device wave propagation velocity *increases* with absorbed analyte mass or analyte concentration around the SAW device. As a result, a vibration wave front launched from the excitation end will arrive at the pickup end $\delta t = L/v$ seconds later. Thus, for CW excitation, there will be a phase lag between output and input given by

$$\theta = \frac{2\pi f_o L}{v} \text{ radians,} \quad (11.11)$$

where

- f_o is the oscillator frequency
- L is the path length of the SAW
- v is the velocity of propagation

v is affected by the density of the absorbent film covering the slab surface and the temperature of the film and slab.

Assuming temperature of the SAW gas sensor is held constant, we can rewrite the phase shift to include analyte concentration:

$$\theta = \frac{2\pi f_o L}{v_o (1 + \beta[C]/v_o)} \cong \frac{2\pi f_o L}{v_o^2} (1 - \beta[C]). \quad (11.12)$$

Thus, the sensor's phase shift decreases with increasing analyte concentration, $[C]$:

$$\Delta\theta = -\frac{2\pi f_o L}{v_o^2} \beta[C]. \quad (11.13)$$

Note that by placing two identical SAW devices on the same chip, $\Delta\theta$ can be measured substantially independently of temperature. One sensor is exposed to the gas-phase analyte, the other is not. Both are at the same temperature. Note that to maximize phase-shift sensitivity, we need a long L , a high f_o , and a low v_o .

Another way of exploiting the increase of SAW velocity caused by the increased mass of the polymer layer is to use the SAW sensor's propagation lag $\delta_t = L/v$ to make a *feedback oscillator*. One then can detect small changes in the oscillation frequency as a function of $[C]$. δ_t can be viewed as a transport lag:

$$\delta_t = \frac{L}{v_o (1 + \beta[C]/v_o)} \cong \left(\frac{L}{v_o} \right) \left(\frac{1 - \beta[C]}{v_o} \right). \quad (11.14)$$

Figure 11.9 illustrates a simple negative feedback oscillator circuit. The amplifier has the transfer function, $A_v(s) = -K_{vo}/(s + a)$. K_{saw} is the volt/volt gain from input electrodes to output electrodes of the SAW device. The oscillator's loop gain transfer function is thus

$$A_L(s) = \frac{-K_{vo} K_{saw} e^{-s\delta t}}{s + a}. \quad (11.15)$$

The use of the root locus technique *angle criterion* for instability (Northrop 2000) gives us an expression for the frequency of oscillation. In general, the root locus plot is based on complex solutions of the vector equation:

$$1 - [A_L(s)] \cong 0. \quad (11.16)$$

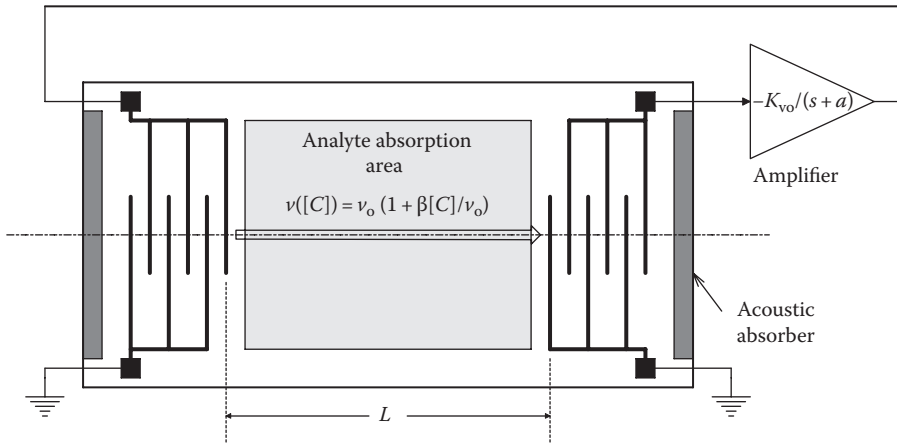


FIGURE 11.9 SAW oscillator. The resonant frequency drops in response to increased delay (phase shift) as analyte binds to the absorption area.

The system oscillates when the gain $K_{v_0}K_{saw}$ is high enough for two root locus branches to cross the $j\omega$ axis at $s = \pm j\omega_o$. This is where the complex variable $s = \sigma + j\omega = j\omega_o$. To find ω_o , we must solve the previous equation based on the *root locus angle criterion*. In general,

$$-A_L(j\omega_o) = -1. \tag{11.17}$$

The angle criterion is

$$\angle -A_L(j\omega_o) = -\pi. \tag{11.18}$$

Thus,

$$-\omega_o\delta_t - \tan^{-1}\left(\frac{\omega_o}{a}\right) = -\pi. \tag{11.19}$$

Equation 11.19 is a transcendental equation whose solution gives the SAW delay oscillator's oscillation frequency. Note that the oscillation frequency, ω_o , is a function of the SAW propagation delay, δ_t . There is no explicit solution to Equation 11.19, but we can linearize it by assuming $(\omega_o/a)^2 \ll 1$, so the arctangent function can be approximated by $\tan^{-1}(x) \cong x$ radians. Thus,

$$\omega_o\delta_t + \frac{\omega_o}{a} \cong \pi \tag{11.20}$$

↓

$$\omega_o \cong \frac{\pi}{\delta_t + 1/a} = \frac{a\pi}{a(L/v_o)(1 - \beta[C]/v_o) + 1} = \frac{a\pi}{[aL/v_o + 1] - aL\beta[C]/v_o^2}$$

↓

$$\cong \frac{a\pi}{[aL/v_o + 1]} \left[1 + \frac{aL\beta[C]/v_o^2}{[aL/v_o + 1]} \right] = \omega_c + \Delta\omega. \tag{11.21}$$

Thus, the incremental change in oscillator frequency due to the analyte absorption, $\Delta\omega$, can be written as

$$\Delta\omega = \frac{-\pi a^2 L \beta [C]}{[aL + v_o]^2} \text{ r/s.} \quad (11.22)$$

The fractional change in oscillator frequency due to analyte absorption is

$$\frac{\Delta\omega}{\omega_c} = \frac{-aL\beta[C]}{v_o(aL + v_o)}, \quad (11.23)$$

where ω_c is the oscillator frequency in the absence of analyte ($[C] = 0$).

Unfortunately, in polymer membrane-based SAW sensors, there is a trade-off between analyte sensitivity and response speed to transient changes in analyte concentration. A thin polymer film rapidly saturates with analyte molecules in the gas having a fixed VOC analyte concentration. Because it is thin, its total mass increase will be smaller than for a thick film. The thick film takes longer to saturate because of diffusion time; it also takes longer to lose the analyte into pure carrier gas (air). Maximum absorption sensitivity in thin films is in the range of femtograms ($\times 10^{-15}$ g) of analyte. Detectors for CO, HCl, H₂, H₂S, NH₃, NO₂, SO₂, various VOCs, and organophosphorous compounds (insecticides) have been developed using SAW sensor arrays (Staples 1999).

A fifth class of VOC analyte sensor is the quartz resonator. A sensor of this type is also called a quartz crystal microbalance (QCM). An analyte binding surface film is deposited on a quartz crystal that is part of an oscillator. When analyte molecules are absorbed on the film, or attach to it chemically, its mass increases and this lowers the frequency of oscillation of the oscillator. Mass increases on the order of 1 pg can be detected. If a liter of air has 1 pg of methane in it, this translates to a concentration of 1.4 ppb (1.4×10^{-9}) (Nagle et al. 1998). QCM research is being carried out at IBM's Zürich research lab.

A sixth class of VOC analyte sensor for E-nose applications is SPR. SPR is a quantum optic-based analytical method that makes use of surface binding of an analyte. SPR is a relatively new analytical tool that is just now being investigated for E-nose applications. SPR allows rapid, specific determination of the concentration of a variety of environmentally, medically, and biologically important analytes, either in liquid or gas phase. For example, specific bacteria, antigens, antibodies, theophylline, caffeine, gasses, pesticides, explosives, and controlled substances (opioid drugs) have been sensed with SPR in solution or as vapors. Threshold sensitivities for certain analytes have been reported as low as 0.01 ppb (10^{-11}). SPR technology is relatively simple and inexpensive to implement, compared with traditional analytical systems such as HPLC, mass spectrometry, IR spectrometry, and GC, and SPR instruments lend themselves well to field measurements. Since this section is about E-noses, the analytes considered are largely gas phase and are derived from a variety of sources including drugs, explosives, environmental contaminants, as well as from medical fluids such as blood, urine, feces, bacteria, and cancers.

SPR is a quantum optical phenomenon that occurs when a beam of monochromatic, LPL is reflected off a thin metal film, is vapor deposited on one side of a glass prism, or is incident at a critical angle on a gold-coated diffraction grating. SPR is described in detail in Section 7.7.3.

The reader should be aware that the design of SPR devices for chemical analysis is a rapidly growing field. One alternative configuration of the SPR sensor places the analyte solution over the surface of a plastic diffraction grating whose surface has been vapor

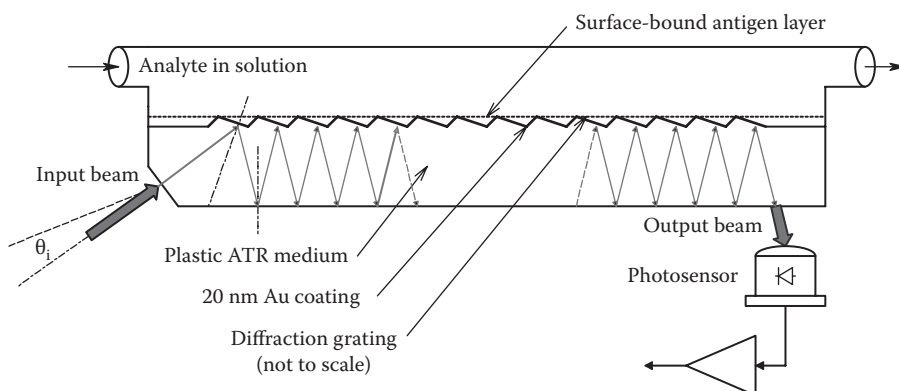


FIGURE 11.10

SPR grating/prism. There are multiple reflections from the grating allowing increased interaction of the input light with the analyte layer before the light exits.

deposited with gold, silver, or aluminum (gold is generally preferred because of its chemical stability). The plastic top of the grating acts as an ATR prism where light reflected from the grating where the beam strikes it initially is reflected back many times to the grating surface. Such a grating SPR design was proposed by Simon (1998). Figure 11.10 shows a side view of Simon's *long range*, SPR grating system. SPR occurs at a critical beam input angle, θ_i , giving a minimum of output light intensity when the coupling component of the monochromatic, TM-polarized, input beam wave vector satisfies the relation

$$\left(\frac{2\pi}{\lambda}\right)\sin(\theta_i) + \left(\frac{2\pi}{b}\right) = \left(\frac{2\pi}{\lambda}\right)\sqrt{\frac{(\kappa_a\kappa_m)}{(\kappa_a + \kappa_m)}}, \quad (11.24)$$

where

b is the grating constant

κ_a is the dielectric constant of the analyte

κ_m is the magnitude of the complex dielectric constant of the metal film

A very thin layer of polymer with high affinity to the analyte to be detected is chemically bound to the 20 nm gold film over the grating surface. The analyte binds to the polymer molecules, changing κ_a at the gold-polymer interface. This change of dielectric constant changes the interfacial refractive index and *retunes* the SPR to a new input angle, θ_i . Either by changing θ_i to retain a minimal light output or by measuring the *increase* in light output due to $\Delta\kappa_a$, the number or density of bound analyte molecules can be quantified. Simon (1998) claimed that a refractive index change in the analyte of less than 1% causes a 0.5° shift in θ_i to renull the system.

Other embodiments of the grating SPR system place the analyte over the gold film-covered grating; the monochromatic, TM-polarized light is directed through a thin film containing the analyte onto the gold-covered grating. Jory et al. (1995) reported on an exquisitely sensitive grating SPR system in which an AOTF element was used to control the wavelength of the incident beam to a precision of 0.0005 nm. They used their system to measure the concentration of NO_2 gas in N_2 (a gas-phase analyte). By depositing a thin layer of *phthalocyanine* over the gold coating, a wavelength shift of -0.004 nm renulled the

system from zero concentration when 0.01 ppm NO₂ in N₂ was applied. They claimed that the sensitivity of their system allowed detection of changes in the refractive index of the gas of 1×10^{-6} .

Capan et al. (2003) reported on a Kretschmann-type SPR system in which a thin film of *polymethyl methacrylate* (PMMA) over the gold film was found to respond most strongly to benzene vapors and less strongly to toluene, methyl benzene, and *m*-xylene vapors. Threshold sensitivity was about 100 ppm (10^{-4}).

In a 1999 review of SPR sensing, Homola et al. described the use of *polyethylene glycol* films in sensing the vapors of hydrocarbon, aldehydes, and alcohols. Chlorinated hydrocarbons could be sensed by SPR by their absorption in *polyfluoroalkylsiloxane* films, and *tetrachloroethane* could be resolved by absorption in a *polydimethylsiloxane* film. Teflon films have been used to sense aromatic hydrocarbon vapors. Still other coatings have been reported to be selective for other vapor analytes: for example, *PAni* was used to sense NO₂ and H₂S, and *bromocresol purple* was used to sense NH₃. When the SPR active metal film coating itself was palladium, which has affinity to hydrogen, it responded to molecular hydrogen, H₂.

SPR gas sensors of the Kretschmann design are apparently more expensive than the CPF and MOX sensors described earlier, and their sensitivities are generally on the order of ppm of analytes, rather than ppb, with some exceptions. Practically, any vapor analyte that can react with a chemical film bound to the gold film surface on a prism or grating with a strong affinity can be sensed by SPR. The reaction must cause a change in the refractive index or the permittivity at the gold surface in order to affect the SPR conditions (reflected intensity, λ and/or θ).

The problem of quick, efficient, SPR sensor regeneration remains to be solved, however. Once the bound surface reactant has combined with the analyte, the analyte or analyte complex must be totally removed from the bound reactant film before the next measurement can be made. That is, the surface reactant layer must be renewed or rejuvenated by some means that can be physical (e.g., heat) or chemical (a purging solvent that does not affect the ligand film over the gold).

The seventh class of gas-sensing systems is broad and includes traditional *chemical analytical systems* that are generally not hand carryable and do not appear suitable for E-noses. These include the following:

1. The *QPMS* that is lighter and more portable than a conventional MS that uses a magnetic field to separate ions. A QPMS produces a mass spectrogram with peaks representing ions broken off from the analyte. Its output peaks have a unique pattern for a given analyte (Northrop 2002).
2. All sorts of *spectrophotometers* that measure the absorption or transmission of light through a sample chamber containing the analyte gas. Light wavelengths from FIR through UV can be used to establish an analyte's unique signature spectrogram. Mixtures of analytes give less certain spectrophotometric results than do single analytes due to absorption peak overlaps.
3. GC exploits the physical property of vapor-phase analytes to propagate through a capillary adsorption column at different rates. (See Section 7.7.2.) The eluent peaks in a gas chromatogram are thus from the individual analytes in a mixture, separated in time. The temporal separation of the eluent peaks, given the column's temperature, is signature of the analytes; the areas of the peaks are proportional to the concentrations. Standard analyte samples are used to calibrate a gas

chromatograph. The type of detector used in a gas chromatograph determines its sensitivity and the class of analyte it is best suited for. For example, the MDA of chlorinated hydrocarbons using an *electron capture detector* is ca. 10 fg/mL (10^{-14} g/mL). A flame ionization detector's MDA is ca. 1 pg/mL (10^{-12} g/mL) for hydrocarbons (Northrop 2002). Because of its simplicity and sensitivity, efforts have been made to miniaturize GC columns and sensors. Ultrahigh-speed GC columns are used in Electronic Sensor Technology (EST)'s Z-Nose® model 4200 and in Alpha MOS' Heracles E-nose. The Z-Nose is about 10 times faster than E-nose systems such as the venerable Cyranose 320 (Leake 2006) and can create 2D Vapor Prints™ that aid in identifying the multiple analytes involved in a measurement.

There has been recent interest in the development of rapid, sensitive, field-portable, μ GC systems for use in E-nose applications. These systems often use air (instead of He or N₂ as a carrier gas in the column). The systems consist of a gas valve for sample introduction, a preconcentration trap, a μ GC column (generally etched in a silicon plate about 3.2 cm²), and an eluent detector. Two detectors in μ GC systems of note are plasma emission spectrum detection (μ PED) (Tienpont et al. 2008) and chemiresistor arrays (Zellers et al. 2011). The μ GC developed by Zellers et al., designed to sense VOCs from explosives, could sense 2,4- and 2,6-dinitrotoluene and the TNT taggant 2,3-dimethyl-2,3-dinitrobutane (DMNB) at low, sub-ppb ranges with a 1 L air sample in less than 3.5 min. The μ GC/ μ PED system of Tienpont et al. used He plus 1% H₂ as the carrier gas and was validated with a detection of *benzene*, *toluene*, *ethylbenzene*, and *o-xylene* in air, each at ≥ 0.13 μ g/L in less than 0.5 min.

At present, μ GC systems lie in a fuzzy area between E-noses and analytical lab instruments. Their greater complexity generally makes them portable, benchtop instruments, unlike the handheld Cyranose 320™.

4. Various *fluorescent techniques* for gas-sensing work in several ways. In one, a fluorescent probe compound that has an affinity for the vapor analyte is used. Fluorescence is induced by shining short-wavelength light (blue or UVA) on the probe compound. As the analyte reacts with the fluorescent probe substance, it can quench the probe's natural fluorescence (dims it proportional to the analyte concentration), or the bound analyte molecules can cause a shift in the fluorescent emission spectrum or increase the intensity of the fluorescence (Levitsky et al. 2001). A colorimetric approach described by Rakow and Suslick (2000) used an array of metalloporphyrin dyes whose colors change in response to analyte binding. These sensors permit the visual identification of ligating vapors (alcohols, amines, ethers, phosphines, phosphites, thioethers, thiols, arenes, halocarbons, and ketones). Sensitivity extends into the single ppm (10^{-6}) range. At the stage of development reported, the dye dots appear to be a qualitative means of analysis. However, intensity and wavelength variables should be able to yield quantitative data as well.

11.3.4 Feature Extraction in E-Noses

A central problem in using arrays of gas sensors having different sensitivities to a group (mixture) of analytes is finding an algorithm that will yield a reliable estimate of the analyte concentrations [C_i] in the mixture. In general, the fractional change in resistance of

the k th sensor in an array of N sensors can be written, assuming linearity of response (superposition) for low analyte concentrations:

$$\frac{\Delta R_k}{R_{ok}} = \sum_{i=1}^N \sigma_{ki} [C_i], \quad k = 1, 2, \dots, N, \quad (11.25)$$

where

σ_{ki} is the k th sensor's response sensitivity to the i th analyte in the mixture
 $[C_i]$ is the concentration of the i th analyte of unknown concentration in the mixture

The $\{\sigma_{ki}\}_N$ is generally known, and $\Delta R_k/R_{ok}$ is measured for each of the N sensors in the array.

From the N linear equations of the form of Equation 11.25, given $\{\sigma_{ki}\}$, the computer uses Cramer's rule to solve for $\{[C_1], [C_2], \dots, [C_N]\}$.

In their review paper on "Cross-reactive chemical sensor arrays," Albert et al. (2000) described and cited references for various linear and nonlinear algorithms that have been used to estimate $[C_i]$ s from MOS and CPF sensor arrays, as well as other gas sensors with cross sensitivities. These include *cluster analysis* (CA), *transformed cluster analysis* (TCA), PCA, *partial least squares* (PLS), and various architectures of ANNs. The ANN approach offers the advantage in that it can be *trained* to detect desired analytes. Keller et al. (1995) cited the early use of ANNs in E-nose design. Ouellette (1999) also discussed the use of ANNs with E-nose sensor arrays. The Cyranose® 320 E-nose that uses an array of 32 CP sensors uses PCA for detection of outliers and data reduction, along with three supervised algorithms for model building and predictions. These algorithms are KNN, *K-means*, and CDA.

It is beyond the scope of this text to treat the details of the data processing algorithms previously cited used to extract concentration data from sensor array outputs.

We have seen that E-noses are instruments designed to quantify the gas components in natural mixtures of gasses. Applications include the detection of food spoilage, disease diagnosis, bacterial identification, and quantifying the atmospheres in closed vehicles such as spacecraft, airliners, and submarines and the odors (and flavorant vapors) in wines (rancid wines contain *2,4,6-trichloroanisole* [TCA]), cheeses, beers, ales, and foods (meats, coffees, teas, mushrooms, fish, etc.). Grains (wheat, barley, and oats) were examined with a Nordic Sensor Technologies NST 3210 E-nose by Börjesson et al. in 1996. They correctly identified 75% of the samples using a four-class system: *moldy/musty*, *acid/sour*, *burnt*, and *normal*.

Growing bacteria also emit characteristic VOCs; hence, their unwanted presence in foods and wounds can be sensed by E-noses. Table 11.4 lists some medical diagnostic applications of E-noses.

For example, lung cancer was first reported by Di Natale et al. and Phillips et al. in 2003, using E-nose sensor arrays. In 2002, Dutta et al., using a Cyranose 320® E-nose, reported a 98% success in discriminating between six different bacterial species responsible for eye infections. These were *Escherichia coli*, *Haemophilus influenzae*, *Moraxella catarrhalis*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*, and *Streptococcus pneumoniae*. A Cyranose 320 E-nose was also used to diagnose gastroesophageal reflux disease (GERD) in patients with chronic obstructive lung disease (COLD). Timms et al. (2012) reported that they could discriminate between COLD patients with GERD and without GERD with an accuracy of 67.6%. The Cyranose 320 also could discriminate asthmatic patients with and without GERD with an accuracy of 85%, but failed to effectively separate COPD patients with and without GERD (64% accuracy).

TABLE 11.4

Medical VOCs: Summary of Important VOCs Analyzed by GC/MS

Material Sampled	Disorder/Infection	VOC
Urine	Urinary tract infection	Isovaleric acid, alkanes
Urine	Metabolic disorders	Isovaleric acid
Intraperitoneal fluid	Aerobic Gram-negative bacteria	Terpenes, ketones
Intraperitoneal fluid	Anaerobic bacterial infections	Acetic and butyric acids
Pus from infection	—	Isobutyric, isovaleric, isocaproic acids
Human breath	Breast cancer	Alkanes, monomethylated alkanes
Human breath	Lung cancer	Alkanes, monomethylated alkanes
Human breath	Acute asthma	Pentane
Alveolar air	Hepatic coma	Methyl mercaptan
Alveolar air	Rheumatoid arthritis	Pentane
Alveolar air	Schizophrenia	Pentane, carbon disulfide
Alveolar air	Diabetic ketosis	Acetone
Alveolar air	Cardiopulmonary disease	Acetone, ethanol
Blood plasma, cerebrospinal fluid	Hepatic encephalopathy	3-Methylbutanol

Source: Data from Turner, A.P.F. and Magan, N., *Nat. Rev. Microbiol.*, 2, 161, 2004.

E-noses use multisensor arrays based on how the mammalian olfactory system works. For a description of vertebrate olfaction, see Kandel et al. (1991) and Northrop (2001, 2005). E-noses using sensor arrays were first developed in the early 1980s (Casaliniuvo et al. 2006). The first E-nose sensor array used three MOS sensors (e.g., SnO₂) and could detect 20 odors (Persaud and Dodd 1982). As the technology of E-noses evolved, other sensors were used in arrays, including MOX silicon field-effect transistors, piezoelectric crystal resonators, SAW sensors, conducting organic polymers, and OF bundles in which the analyte affects absorbance, reflectance, fluorescence, or chemiluminescence—the analyte binds to a specific coating on the end of the fiber. Some E-noses use combinations of sensor types to enhance discrimination of analytes.

In spite of all the research on E-noses, the olfactory systems of trained dogs still appear to be the most reliable and most sensitive means for detecting drugs and explosives. Dogs have also been trained to detect the scent of lung cancer on a patient's breath. Certainly, we have a long way to go to invent an artificial nose with canine sensitivity. Dog threshold olfactory sensitivity to substances they have been trained to detect (e.g., cocaine, semtex, and C4) has been estimated to be on the order of tens of ppb (10⁻⁸), down to as great as ca. 500 ppt (5 × 10⁻¹⁰) (Waggoner et al. 1997).

It is well known that human olfactory sensitivity can vary more than a factor of 10³ between normal individuals. Some individuals are totally unable to smell certain odors, while others can. However, certain odors such as the *rotten egg* scent of H₂S can be sensed in concentrations of several parts per trillion. Methyl mercaptan (used as warning odorant in domestic natural gas) can be detected in air in concentrations as low as 7 × 10⁻¹³ M (700 nM). Olfaction in humans (and presumably in other mammals) is ca. 10⁴ times more sensitive than taste (gustation). It is well known that as persons age, there is a loss of threshold sensitivity for certain odors and the ability to discriminate between odors (Jacob 2003).

An E-nose is available 24/7; dogs require rest. An E-nose requires calibration, however, just as a dog requires careful training to respond to specific scents. Note that many less-portable or fixed technologies exist for VOC analysis, particularly in airports, to detect nitrates and explosives. These include GC/MSs and spectrophotometers. Even these instruments

have been reduced in size so they can be brought into the field; they generally are not hand portable, however. A swab of suspect analyte is brought to the GC/MS instrument.

11.3.5 Some Commercially Available E-Noses

In Table 11.5, we give a nonexhaustive list of E-noses, their manufacturers, and description of their operation.

The 18 sensors in the FOX 4000 E-nose have MOX architectures. The analyte for which they have maximum sensitivity is given in Table 11.6.

TABLE 11.5

E-Noses: A Nonexhaustive List of Some Popular E-Noses

Instrument Name	Manufacturer	Description
Cyranose 320	Intelligent Optical Systems, formerly Cyrano Sciences	Uses array of 32 conductive polymer sensors
i-PEN, i-PEN3	AIRSENSE	Uses 10 MOX sensors
GDA2	AIRSENSE	Portable ion mobility spectrometer
Reporter	AIRSENSE	Dispersive Raman spectrometer
FOX 2000, 3000, 4000	Alpha-MOS	Use MOX sensor arrays
GEMINI	Alpha-MOS	6-sensor array
HERACLES	Alpha-MOS	Uses ultrafast μ GC for VOCs
ASTREE	Alpha-MOS	Liquid-phase E-tongue—uses chemFETs
Prometheus E-nose	Alpha-MOS	18 MOX sensors, 6 ea. in 3 chambers, thence to a quadrupole MS
MOSES II	GSG Mess- und Analysengeräte	8 MOX sensors + 8 quartz microbalances
OMA (Osmetech Microbial Analyzer)	Osmetech Plc.	Microbial analyzer (VOCs)
OMA-UTI	Osmetech Plc.	Urinary tract infection detection
NST 3220E Lab Emission Analyzer	Nordic Sensor Technologies (NST)	Uses 10 MOSFET and 5 MOX sensors
Z-Nose 4200	EST	Uses fast μ GC
E-nose 1, model BH114	Bloodhound Sensors Ltd., Leeds, United Kingdom	
AromaScanner A32S	AromaScan, United Kingdom (now Osmetech, a unit of GenMark (2010))	Uses conductive polymer sensors

Note: The Cyranose[®] 320 was put on the market by Cyrano Sciences (CS) in 2000. At present, the Cyranose[®] 320 is marketed by Sensigent, Baldwin Park, CA. (www.sensigent.com/products/cyranose.htm). AIRSENSE Analytics, Schwerin, Germany, www.airsense.com/. Alpha-MOS (A French multinational corp. headquarters in Toulouse, FR and Hanover, MD.) <http://alpha-mo.ipower.com/>. Their Prometheus E-nose uses 18 MOX sensors, 6 ea. in 3 chambers, thence to a quadrupole MS. GSG Mess- und Analysengeräte of Bruchsal, Germany: see [www.gsg-analytical.com/english/Gasanalytik.htm/](http://www.gsg-analytical.com/english/Gasanalytik.htm). Osmetech plc, Crewe, United Kingdom, made Osmetech Microbial Analyzer (OMA) and OMA-UTI urinary tract infection detector. Nordic Sensor Technology, now AppliedSensor, www.appliedsensor.com/, made the NST 3210 E-nose; the NST 3220E was used in dairy product testing and evaluation by Schaller (2000). Electronic Sensor Technology, Newbury Park, CA, Z-Nose is based on fast gas chromatography. www.estcal.com/. Bloodhound Sensors Ltd. is now dissolved (2013).

TABLE 11.6

Principal Analytes Producing Large Responses in the Alpha MOS Fox 4000 E-Nose MOX Sensor Array

FOX 4000 Sensors	Principal Analytes
LY2/LG	Fluoride, chloride, oxynitride, sulfide
LY2/G	Ammonia, amines, carbon–oxygen compounds
LY2/AA	Alcohol, acetone, ammonia
LY2/GH	Ammonia, amine compounds
P40/2	Chlorine, H ₂ S, fluoride
P30/2	H ₂ S, ketone
T30/1	Polar compounds, HCl
P10/1	Nonpolar compounds, hydrocarbons, ammonia, Cl
P10/2	Nonpolar compounds, CH ₄ , C ₂ H ₆
P30/1	Hydrocarbons, ammonia, ethanol
T70/2	Toluene, xylene, CO
T40/1	Fluorine
P40/1	Fluorine, chlorine
LY2/gCTL	H ₂ S
LY2/gCT	Propane, butane
T40/2	Chlorine
PA/2	Ethanol, ammonia, amines
TA/2	Ethanol

Source: Data from Yu, Y.X. and Zhao, Y., Electronic nose integrated with chemometrics for rapid identification of foodborne pathogen, in: *Chemometrics in Practical Applications*, K. Varmuza, ed., InTech, Croatia, 2012, 15 pp. cdn.intechopen.com/pdfs/33613/InTech-Electronic-nose-integrated-with-chemometrics-for-rapid-identification-of-foodborne-pathogen.pdf (accessed November 28, 2012).

11.4 Radio ICs for Wireless Data Transmission

11.4.1 EM Spectrum

WDX is required for carried and implanted MEM sensors on moving animals and humans, vehicles, cargo, etc. Sensor implants and radio tags are used in areas such as hospital intensive care wards, in sports medicine, cardiology, physiological research, and ecological studies.

EM radiation has been classified by wavelength (and frequency); Table 11.7 denotes the commonly accepted frequency band designations.

The RF communication bands can also be described in terms of their wavelength ranges as ULF ($10^7 \leq \lambda \leq 10^8$ m), ELF (10^5 – 10^7 m), VLF (10^4 – 10^5 m), LF (10^3 – 10^4 m), MF (10^2 – 10^3 m), HF (10 – 10^2 m), VHF (1–10 m), UHF (10^{-1} –1 m), SHF (10^{-2} – 10^{-1} m), and EHF (10^{-3} – 10^{-2} m). RF bands can also be described by frequency ranges: for example, VLF (10–30 kHz), LF (30–300 kHz), MF (300 kHz–3 MHz), HF (3–30 MHz), VHF (30–300 MHz), UHF (300 MHz–3 GHz), and SHF (3–30 GHz).

The dimensions of the antennas used with Ws must be as small as possible, subject to the constraints of the RF wavelength used, the designed working range of the WPM device, and the radiation pattern required. (Radiation patterns of some multielement directive arrays

TABLE 11.7
EM Spectrum

Band	Wavelength λ	Frequency	Photon Energy
Cosmic rays	1 pm	300 EHz	1.24 MeV
Gamma rays	10 pm	30 EHz	124 keV
Gamma and x-rays	100 pm	3 EHz	12.4 keV
Far UV, x-rays	1 nm	300 PHz	1.24 keV
Near ultraviolet (NUV)	10 nm	30 PHz	124 eV
Visible light	100 nm	3 PHz	12.4 eV
Near infrared (NIR)	1 μm	300 THz	1.24 eV
MIR	10 μm	30 THz	124 meV
FIR	100 μm	3 THz	12.4 meV
EHF	1 mm	300 GHz	1.24 meV
SHF	1 cm	30 GHz	124 μeV
UHF	10 cm	3 GHz	12.4 μeV
VHF	1 m	300 MHz	1.24 μeV
HF	10 m	30 MHz	124 neV
MF	100 m	3 MHz	
LF	1 km	300 kHz	
Very low frequency (VLF)	10 km	30 kHz	
Ultralow frequency (ULF)	100 km	3 kHz	
Super low frequency (SLF)	1,000 km	300 Hz	
Extra low frequency (ELF)	10^4 m	30 Hz	
Extra extra low frequency (EELF)	10^5 m	3 Hz	

Notes: EHz = exaHertz = 10^{18} Hz, PHz = petaHertz = 10^{15} Hz, THz = teraHertz = 10^{12} Hz, GHz = gigaHertz = 10^9 Hz, MHz = megaHertz = 10^6 Hz, kHz = kiloHertz = 10^3 Hz. Photon energy is important at shorter wavelengths because of its role in breaking chemical bonds and initiating chemical reactions (Northrop 2012).

are described in the *ARRL Antenna Book* [1974].) Recall that the wavelength of a 2.4 GHz EM wave is $\lambda_{(\text{m})} = 0.3/f_{(\text{GHz})} = 0.3/2.4 = 0.125 \text{ m} = 12.5 \text{ cm}$. A conventional two-element dipole resonant at 2.4 GHz is $2 \times \lambda/4 = \lambda/2 = 6.25 \text{ cm}$ in overall length. Efforts to shrink the effective size of antennas in WPM and radio-frequency identification (RFID) applications have led to various planar antenna designs that can be etched onto PC boards. Such designs include the various meander and fractal antenna geometries described in Section 11.6.4.2.

When a WS modulator/transmitter (M/T) is implanted in an animal's or a person's body, or is swallowed, the RF wavelength in the body is shorter because the RF propagation velocity (c) in tissue is slower than in air. One implantable M/T described by Hashemi (2009) that used a center frequency of 915 MHz had a simple, optimum, $\lambda/4$, bent-single-wire antenna length of 8 cm in air, but the implanted device required a 5 cm, $\lambda/4$, bent-wire antenna, a 37.5% reduction in length. Of course, the water and conductive, ion-containing tissue surrounding the antenna attenuates the broadcast RF.

11.4.2 Some UHF Radio Chips

Many IC manufacturers now offer complete transceiver chipsets for use in the 2.4–2.5 GHz Industrial, Scientific, and Medical (ISM) band, suitable for WDX. We cite some examples (not an exhaustive list): National Semiconductor[®] designed a Bluetooth[™] “Serial Port Module”

IC, the LMX9820 (October 2006). This transceiver measures $10.1 \times 14.0 \times 1.9$ mm and contains, in addition to a digital smart radio, all circuitry required for two-way, Bluetooth (v.1.1 stack) communications protocol. Its internal memory supports up to three active Bluetooth links. The UART command/data port operates up to 921.6 kbaud. National also made the LMX3162 single-chip transceiver (March 2000). This chip radio has RF sensitivity to -93 dBm, with a typical system NF = 6.5 dB. This transceiver is used with external filters and an external microcontroller (MCU). An onboard, 1.3 GHz phase-locked loop is used for both modulation and demodulation functions.

Nordic VLSI ASA offers the nRF2401 single chip 2.4 GHz transceiver (2004). This radio has up to 125 channels; channel switching time is <200 μ s, with data rate up to 100 Mbps. It runs on 1.9–3.6 VDC and draws ≤ 10.5 mA pk. current at -5 dBm RF output power. Nordic Semiconductor also offers (April 20, 2009) a Bluetooth™ 2.4 GHz transceiver line, the μ Blue™. The first chip in this line is the nRF8001 single-mode slave chip, ideal for wearable sensors.

Micro Linear makes a 2.4 GHz Radio Chipset: the ML2712 2.4 GHz radio transceiver, the ML2713 Radio IF transceiver, and the ML7730 802.11 Baseband Processor. This 2000 system was intended for the IEEE 802.11 frequency-hopping spread spectrum standard.

Cypress Semiconductor Corp. offers the CYRF69103 low-power, 2.4 GHz radio on a chip (2013). This is an 8-bit flash-based MCU function and a 2.4 GHz transceiver in a single device. The radio has DSSS rates up to 250 kbps, a GFSK data rate of 1 Mbps, -97 dBm receiver sensitivity, programmable output power up to $+4$ dBm, automatic gain control (AGC), etc. The radio has a dual-conversion, low IF architecture optimized for power and range/robustness; it supports up to 98 discrete 1 MHz channels.

Qualcomm Atheros Communications offers a single-chip, AR6005 IEEE 802.11b/g/n WLAN transceiver (2011) for mobile and embedded applications, operating in the 2.3–2.5 GHz ISM band. This transceiver is compliant with the IEEE 802.11b, 802.11g, 802.11d, and 802.11n standards. In single-stream 802.11n mode, it provides up to 48 Mbps wireless throughput.

11.4.3 Antennas for Radio Chips

11.4.3.1 Introduction

The size of the antennas used in WDX and WPM applications generally must be as small as possible, subject to the constraints of the RF wavelength used, the designed working range of the WDX device, and the directionality desired. Recall that the in vacuo wavelength of a 2.4 GHz EM wave is $\lambda_{(m)} = 0.3/f_{(GHz)} = 0.3/2.4 = 0.125$ m = 12.5 cm. A conventional two-element, quarter-wave dipole resonant at 2.4 GHz is $2 \times \lambda/4 = \lambda/2 = 6.25$ cm in overall length. As you will see in Section 11.4.3, efforts to shrink the effective size of antennas in WDX, WPM, and RFID applications have led to various planar antenna designs that can be etched onto PC boards. Such designs include the various meander designs and fractal antenna geometries described in Section 11.4.3.2.

When a WDX M/T is implanted in an animal's or a person's body, or is swallowed, the RF wavelength in the body is shorter because the RF propagation velocity (c) in tissue is slower than in air. One implantable M/T described by Hashemi (2009) used a center frequency of 915 MHz and had a simple, optimum, $\lambda/4$, bent-single-wire antenna length of 8 cm in air, but the implanted device required a 5 cm, $\lambda/4$, bent-wire antenna, a 37.5% reduction in length. Of course, the conductive, water-containing tissue surrounding the antenna attenuated the broadcast RF.

11.4.3.2 Fractal Antennas

Obviously, all wireless sensors, radio tags, GPS receivers, and handheld communication systems require antennas, and there is a general need for antenna designs that are small (have dimensions $< \lambda/4$) and are planar (can fit on a PC board). Nathan Cohen (1997) was probably the first person to put a fractal antenna design to practice (in 1988), in an attempt to shrink his ham radio antenna. Since then, fractal space-filling antenna designs have proliferated in all areas of RF communications, including WPM, cell phones, RFID tag designs, and GPS systems operating in the UHF frequencies (Fractal Antenna 2013, Fractal FAQ 2013).

Fractals are iterative, self-replicating geometries found throughout nature and science. Mathematically, a deterministic fractal is a broken curve generated by repetitively applying a *generator function* (GF) with successive decreases in scale to an *initiator*. The initiator can be a straight line or a geometric area such as a triangle, circle, ring, cross, a Haar wavelet doublet, or a square. Deterministic fractals in two and three dimensions are also generated by mathematical formulas, for example, the Mandelbrot set equations, the Julia set equations, and Ferguson's *Sterling* fractal-generating computer program (Sterling 2013).

Random fractals are found in nature, for example, tree dendrition, a lightning bolt's branching, certain plants (e.g., ferns, romanesco broccoli), and frost crystals on glass. They can be generated on a computer by introducing random perturbations (noise) into a fractal-generating formula's constants.

There is growing use of various fractal element antenna (FEA) designs for miniature radio applications, such as RFID tags, mobile communications (e.g., cell phones, laptop computers), and GPS receivers. Fractal antennas are also referred to as multilevel, patch antennas and space-filling curves. The major feature of an FEA is the repetition of a geometrical motif over several scale sizes (usually no more than four). As receivers, they generally operate on the radiated EM field of the transmitting antenna in the far field (at 915 MHz, the far field exists greater than 5.2 cm from the transmitting antenna (Scharfeld 2001)).

Many FEA element designs are planar monopoles, that is, they exist on a 2D plane (*patch*). The FEA's ground plane is generally orthogonal to the planar monopole radiation element, giving the entire antenna a 3D domain. (See Werner and Ganguly (2003) for an in-depth review of FEA designs.) The evolution of a fifth-order, *space-filling*, Hilbert fractal pattern antenna element is shown in Figure 11.11 (Mitchell 2001, Wolfram 2010). In this case, the generator function is the same as the initiator. The gold-plated conductor pattern is laid down on a dielectric substrate such as glass, ceramic, or fiberglass.

One linear FEA design, illustrated in Figure 11.12, generates a 2D fractal in a plane by repetitively applying a triangular generator to a straight-line initiator with successive reductions in scale. This von Koch fractal process generates elements that can be used as dipoles or multielement FEAs. Generally, three to four iterations of the simple von Koch linear fractal algorithm are used. Zainud-Deen et al. (2009) gave a thorough analysis and evaluation of two, third iteration (K3) von Koch fractal elements connected in series as a monopole ($2L_e = 7.35$ cm, $\alpha = 70^\circ$), sandwiched between two dielectric layers. The antenna center frequency was ca. 910 MHz. Two K3 von Koch fractals were connected as a dipole antenna and the behavior studied by Ibrahim et al. (2006). They reported that their antenna had two resonances, one at 868 MHz and the other at 2.45 GHz. von Koch fractal monopoles and dipoles are, in general, more effective antennas than meander dipoles, shown in Figure 11.13.

A general property of FEAs is that they tend to be broadband, that is, have effective radiation properties over a relatively broad band of operating frequencies. Conventional

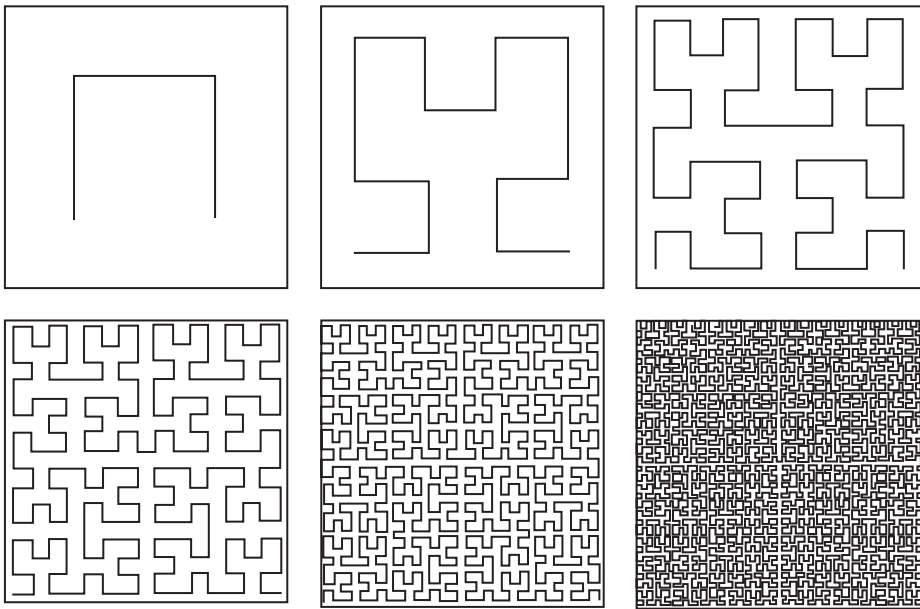


FIGURE 11.11
Evolution of a space-filling, Hilbert fractal mesh antenna design.

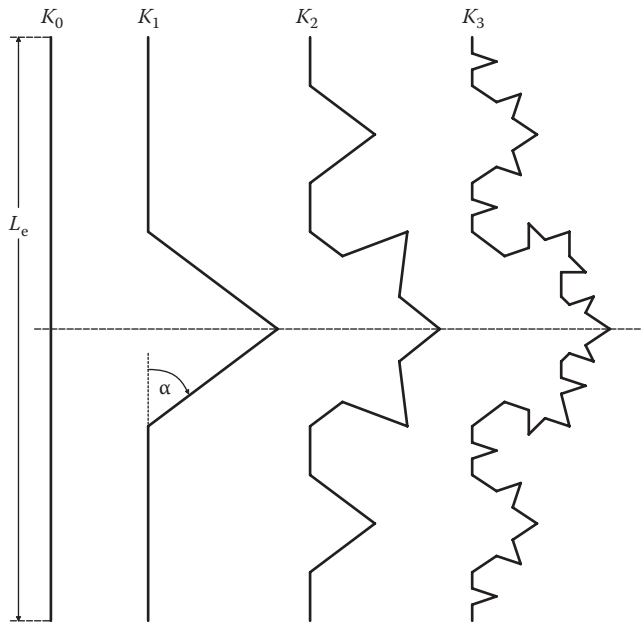


FIGURE 11.12
Evolution of a third-order Koch fractal linear antenna design.

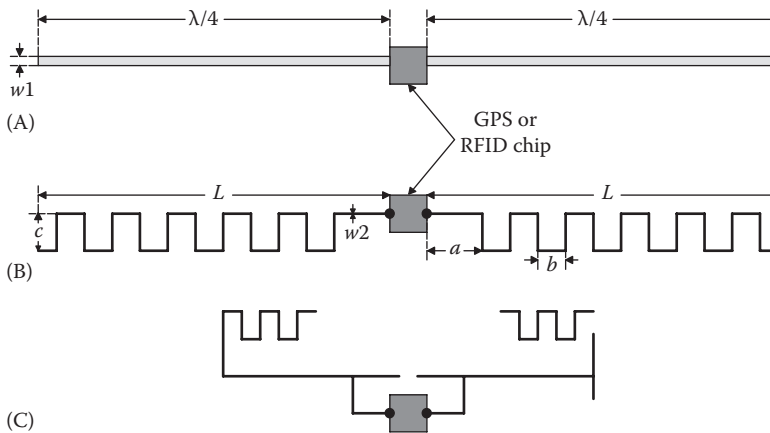


FIGURE 11.13

(A) Simple quarter-wave dipole. (B) Meander dipole. (C) Variation on the meander dipole.

linear dipoles and multielement arrays tend to have *tuned* or narrowband characteristics, although the log-periodic antenna array, such as used in broadcast TV reception, is relatively broadband. Log-periodic antennas have been called *fractal*, and log-periodic elements have been *fractalized* using von Koch triangular generators (see Karim et al. 2010) in order to reduce antenna size at a given frequency (by 26% for a series iteration having two triangle *bumps*/elements).

Another property of fractal antennas is that they are scalable; they are even useful at THz frequencies ($0.1\text{--}10 \times 10^{12}$ Hz) (Gaubert et al. 2004) where they find application in imaging ex vitro and in vitro tumors, including breast cancers (Berry et al. 2004, Fitzgerald et al. 2006, Townsend 2008). Note that the free-space wavelength at 1 THz = 0.3 mm, and a 1 THz *photon*, carries only 4 millielectron volts of energy (4 meV). X-rays, on the other hand, are ionizing radiations. Their photons carry $10^3\text{--}10^4$ eV. For example, triangular Sierpinski gasket FEAs operating around 1 THz were designed with a major edge size of 1600 μm , and the smallest, internal, equilateral triangles at iteration 4 are 100 μm on a side (Gaubert et al. 2004). See Figure 11.16 for the general shape of a triangular Sierpinski gasket FEA.

Fractal antenna design appears to be an art form as well as engineering science; many papers have appeared presenting and evaluating the performance of diverse FEA forms. For example, an unusual, planar, wheel-shaped FEA monopole radiator was described by Kumar and Malathi (2009). Their antenna had circular symmetry and used concentric iterations or repeats (wheels inside wheels) of the outer, basic *cog wheel* pattern shown in Figure 11.14. Their innovative *wheel* fractal used three, progressively smaller, concentric versions of the outer *wheel*. It had resonances at 1.34, 2.575, and 4.22 GHz, and the impedance bandwidths around the resonance frequencies were 40.77%, 19.02%, and 26.94%, respectively.

Another unusual planar fractal antenna with broadband characteristics was described by Liu et al. (2006). The basic geometry of their modified Sierpinski fractal monopole antenna is shown in Figure 11.15. The basic monopole consists of two, vertical, 60° metal triangles (88.9 mm high, 51.3 mm base) on a 200×200 mm ground plane. The authors show the resonance spectra when one large metal circle (the F1 generator) is added between the triangles, then three smaller circles plus the first circle, then 9 even smaller circles are added, and finally in the fourth iteration, 27 small circles are clustered in groups of 3 around each of the 9 even smaller circles. The fourth iteration Sierpinski fractal antenna is usable from 1 to 10 GHz.

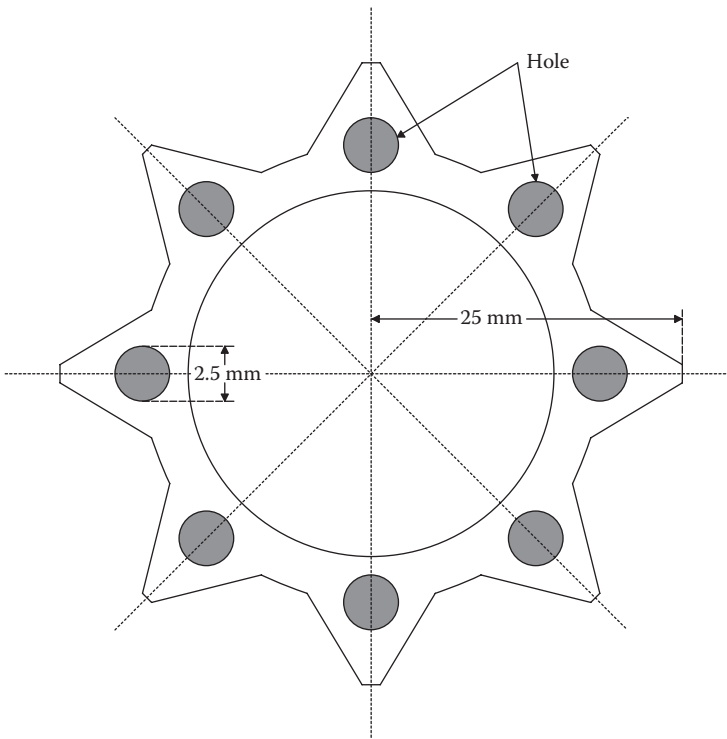


FIGURE 11.14
Cog wheel fractal seed. See text for description.

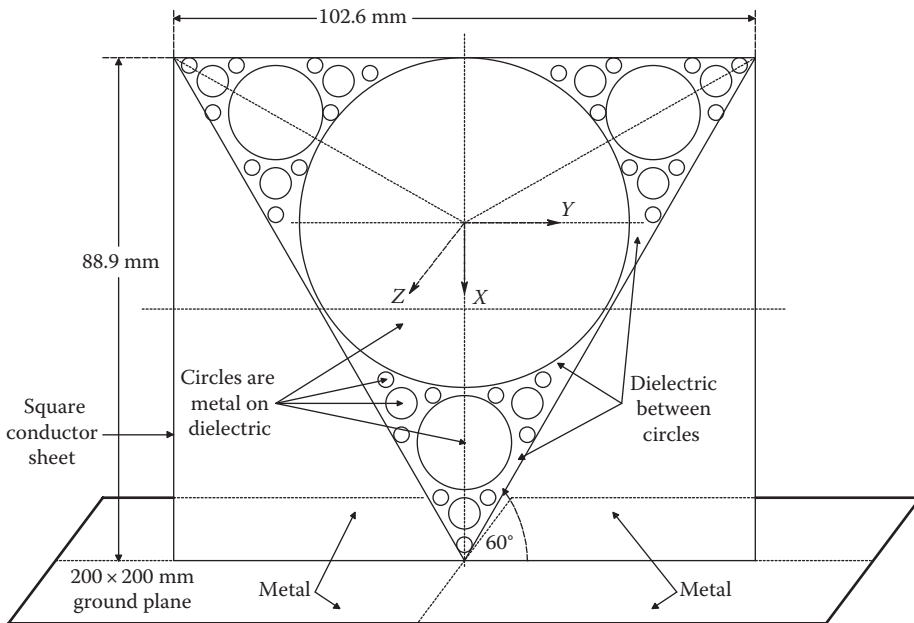


FIGURE 11.15
Geometry of a Sierpinski fractal monopole antenna. Metal circles are used on fiberglass.

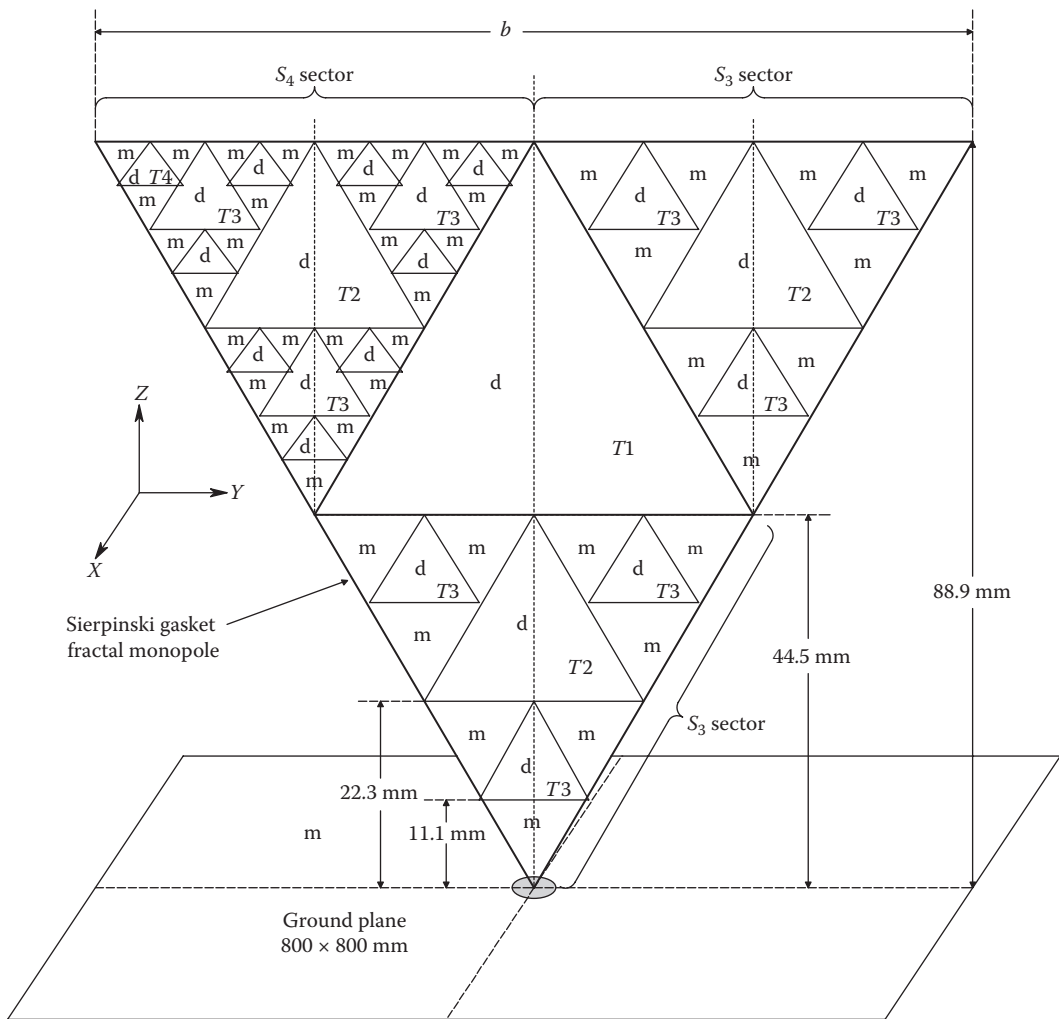
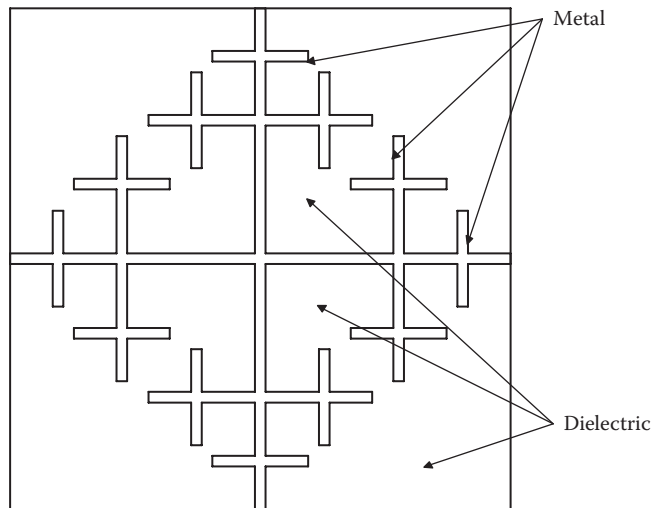


FIGURE 11.16

Another Sierpinski triangular, fractal monopole antenna. *Note:* m, metal; d, dielectric.

Puente-Baliarda et al. (1998) described yet another fractal monopole antenna. This fractal antenna is shown in Figure 11.16. Shown in the upper left is the fourth-order evolution (S_4) of the Sierpinski triangle gasket applied to a metal (m), vertical triangular monopole initiator (tip down) (the other 2/3 of the figure shows the (S_3) design). Each triangular area added to the metal initiator is dielectric (d in the figure). In each iteration, the triangles are one-half the size of those in the preceding iteration. Puente-Baliarda et al. tested a fifth-order (S_5) version of the triangular Sierpinski gasket monopole antenna, the smallest triangles of which were 1/32 the size of the original (S_0) triangular monopole. This monopole fractal antenna was useful over a broadband UHF range of ca. 0.5–14 GHz.

Figure 11.17 illustrates a second-order, space-filling fractal element based on crosses (Werner and Ganguly 2003). (The crosses are metal strips laid down on a dielectric substrate.) The F_0 initiator is the big cross, the F_1 iteration adds four cross pieces, and F_2

**FIGURE 11.17**

Werner and Ganguly (2003) space-filling fractal antenna, based on a cross generator or seed.

adds 12 smaller cross pieces. An F_3 (not shown) iteration can be obtained by crossing each of the 12 smaller cross pieces, giving 36 small crosses. The cross fractal tree element can be used as a monopole radiator with a ground plane or as a dipole.

A comprehensive comparison of UHF fractal antenna designs and their performances at various frequencies seems to be lacking in all of the extensive, available, current literature on the design and evaluation of FEAs.

In addition, there is a bewildering variety of FEA shapes and geometric evolutions. One principle does emerge, however; the higher the order of the fractals and the more space-filling that is used result in broader useful bandwidths and smaller antennas at a given frequency than do nonfractal designs (beams, dipoles, helices, etc.).

11.5 Wireless Patient Monitoring Systems

11.5.1 Introduction

Wireless patient monitoring (WPM), (or wireless medical telemetry services [WMTSs]), is a rapidly growing WDX technology applied to the monitoring of the vital signs of patients in all hospital departments (OR, RR, ICU, etc.), as well as ambulatory patients undergoing physical therapy and exercise. Analog sensors attached to patients measure parameters such as BP, ECG (various leads), EEG (various leads), blood O_2 saturation and heart rate, and body temperature and send their output signals to IC battery-powered multiplexors, digitizers, modulators, and RF transmitters, generally operating around 600 MHz, 1.4 GHz, or in the wireless local area networking (WLAN) bands at 2.4 and 5.1 GHz. These RF data transmissions are received at central stations where they are displayed in real time for nurses and clinicians and stored as part of a patient's permanent medical record. They are screened by a computer for unacceptable physiological values, which, if present, sound alarms.

WPM offers the flexibility of data interception by authorized nursing staff, residents, and attending physicians on their RF-equipped portable computers, as well as the ability to collect data from ambulatory patients as they exercise. Physicians can also access patient's electronic health records (EHRs) with their laptops and mobile PDAs, also by an RF link using protocols such as *Wi-Fi*TM or *Bluetooth*TM.

11.5.2 RF Communications Protocols Used in WDX and WPM

The signals from most physiological and environmental QUMs have relatively low bandwidths (see Sect. 1.1 in Northrop [2010]). In general, the bandwidths (equivalent frequency content) of endogenous physiological signals can range from nearly DC (ca. 1.2×10^{-5} Hz [or 12 μ Hz], a period of 24 h) to several kHz (Stefanovska 2007). Many endogenous hormones, neuromodulators, and behavioral parameters have circadian (ca. 24 h) cycles. Cardiac pumping spans 0.6–4 Hz; respiratory effort lies in a 0.145–0.6 Hz bandwidth (BW). Other neurogenic mechanical rhythms in the digestive system have even lower BWs. Bioelectric signals recorded from the skin surface have higher bandwidths: these include the EEG, EOG, ECoG, EMG, and ECG. Sounds emitted by organs recorded through the skin with microphones include heart valve and blood vessel turbulence sounds, sounds from the respiratory tract (e.g., rales), and sounds from arthritic joints. Such sounds generally require an audio signal-conditioning bandwidth, typically from 10 Hz to 3 kHz or so. Details on the bandwidths required to condition and reproduce physiological signals and their amplitudes can be found in Northrop (2002), and descriptions and analyses of various modulation and demodulation schemes for low-frequency WDX signals are also given in detail in Northrop (2012, Ch. 11).

Because the majority of physiological signals (x_m) recorded and transmitted in WPM have very low frequencies, the modulation scheme used must have frequency response to DC. FM and narrowband FM (NBFM) are ideal for this purpose. FM and NBFM signals can easily be demodulated using phase-locked loops (see Sect. 11.5.3 in Northrop 2012). Another modulation scheme that supports a DC or low-frequency signal is DSBSCM. Duty cycle or PWM of the RF carrier can also convey DC information. Detection is by rectification and low-pass filtering of the PWM RF and then subtracting out the DC level when the RF pulse duty cycle is 50% (for $x_m = 0$).

*Wi-Fi*TM is a trademark of the Wi-Fi Alliance that manufacturers may use to brand certified products that belong to a class of WLAN devices based on the IEEE 802.11 standards. The IEEE 802.11b standard defines 11 RF channels for *Wi-Fi*TM communications in the ISM 2.4 GHz band; the center frequencies are separated by 5 MHz and are at 2.412, 2.417, 2.422, 2.427, 2.432, 2.437, 2.442, 2.447, 2.452, 2.457, and 2.462 GHz. The absolute spectral width of each *Wi-Fi* channel is 22 MHz, so channels overlap (see *Wi-Fi* 2013, for a summary of *Wi-Fi* protocols). Because of channel overlap, *Wi-Fi* user density, RF emissions from Bluetooth devices, baby monitors, cordless telephones, microwave ovens, etc., interference issues have arisen for *Wi-Fi* users. These have been called *Wi-Fi pollution*.

There are more than 220,000 public *Wi-Fi hotspots* where *Wi-Fi* equipped laptops can access the Internet. *Wi-Fi* is also used for wireless Internet connections in tens of millions of homes, corporations, and university campuses worldwide. Embedded *Wi-Fi* systems have been used in portable (ambulatory) ECG monitors to permit home recording of a patient's ECG and its wireless transfer to the Internet, thence to a hospital. The current version of *Wi-Fi Protected Access* encryption (WPA2) as of 2010 is considered secure, provided that a strong passphrase is employed by users. The range of *Wi-Fi* communications using the 802.11b or 802.11g standards can be 32 m indoors and 95 m outdoors, using a standard antenna. In the new 802.11n standard, these ranges are effectively doubled.

ZigBee[™] has another established set of specifications governing wireless personal area networking (WPAN) systems using digital radio communications. ZigBee uses the 868–870 MHz and the 915 (902–928) MHz and 2.4 (2.4–2.5) GHz, ISM RF bands. Maximum data rates allowed on these bands are 250 kbps at 2.4 GHz, 40 kbps at 915 MHz, and 20 kbps at 868 MHz. ZigBee protocols have been standardized by the IEEE 802.14.4 Wireless Networking Standard. The ZigBee protocol's data bandwidth is only ca. one fourth that of Bluetooth (1 Mbps); this allows replacement of cables for many serial data applications, lower costs, and longer battery life (ZigBee 2013). ZigBee applications include, but are not limited to, personal, home, and hospital care; home automation; telecommunications; and commercial building automation. ZigBee devices can go from sleep to active mode in 15 ms or less, while Bluetooth wake-up delays are typically ca. 3 s. There are three different types of ZigBee devices:

1. The *ZigBee coordinator* (ZC)—forms the root of the network tree and can bridge to other networks. There is only one ZC in each network. It stores information about the network, including acting as the trust center and repository for security keys.
2. The *ZigBee router* (ZR)—as well as running an application function, the ZR can act as an intermediate router, passing on data from other devices.
3. The *ZigBee end device* (ZED)—can just talk to the parent node (either the ZC or ZR), but cannot relay data from other devices. This protocol allows the ZED to be asleep for a significant amount of time, giving longer battery life. A ZED requires the smallest memory and therefore can be less expensive to manufacture than a ZR or ZC.

Bluetooth[™] RF networking is rapidly becoming a preferred technology for WPM (Baisa 2005). It uses a radio transmission protocol called *frequency-hopping spread spectrum* that chops the data being sent into packets and transmits it on up to 79 bands of 1 MHz width in the range of 2.402–2.480 GHz (this is in the global ISM 2.4 GHz band that requires no licensing). In classic Bluetooth, the modulation is Gaussian frequency-shift keying (GFSK). It can achieve a gross data rate of up to 1 Mbit/s (1 Mbs). Other modulation schemes give extended data rates of 2 and 3 Mbit/s (Bluetooth 2013). Bluetooth provides a secure way to connect and exchange information between devices such as mobile phones and their accessories, faxes, telephones, laptops, PCs, printers, GPS receivers, digital cameras, and WPM devices. Bluetooth system specifications are developed and licensed by the Bluetooth Special Interest Group (BSIG). The BSIG is comprised of more than 13,000 companies. Bluetooth transmitters are classified by radiated power: Class 1 has a maximum permitted power of 100 mW and an effective range of ca. 100 m. Class 2's max. power = 2.5 mW; its range is ca. 10 m. Class 3 radiates 1 mW and has a range of ca. 1 m.

The Bluetooth protocol of frequency-hopping spread spectrum makes a Bluetooth signal difficult to eavesdrop on.

11.5.3 Commercial WPM Systems

The development of remote, WPM systems is a rapidly growing, competitive area of medical electronics (Kalorama 2010). Many corporations offer WPM product lines, most are for fixed hospital/clinical settings, and a few are portable systems intended for emergency medicine (field use). See Sect. 13.6.1 in Northrop (2012) for a nonexhaustive listing of commercial WPM systems.

11.6 Power Sources for Wireless Sensors, Effectors, and WDX

Clearly, wireless sensors, effectors, and WDX systems can be powered by compact, light, rechargeable batteries, such as lithium ion. A wireless device's receiver software can be programmed to detect loss of signal and give an alarm, should a battery fail or a sensor become disconnected from its power source or otherwise fail.

An emerging technology of *energy-harvesting* power sources that derive their power from ambient or *free energy* promises to create ultrareliable sources for low-power wireless sensors and short-range WDX. One important energy source for harvesting ICs is TEGs. TEGs are simply thermopiles consisting of many *n*- and *p*-type semiconductor (e.g., Bi_2Te_3) pellets arranged as series thermocouples, sandwiched between two, flat, ceramic, heat-conductive plates (cf. Section 6.4.1) (see Figure 11.18). Many *pn* junctions in series are packed between a flat upper heat source and a lower heat sink (TEG Power 2013, Marlow 2013). The Seebeck effect produces open-circuit, TEG EMFs on the order of $10\text{--}50\text{ mV}/\Delta T^\circ\text{C}$, depending on the TEG size and the temperature difference, ΔT , between the two plates in contact with the *hot* and *cold* junctions. Note that in WPM sensor applications, $\Delta T \approx 38 - 21 = 17^\circ\text{C}$, using the temperature difference between the human body and the hospital room as the energy source. Yang and Liu (2010) reported on their study of the feasibility of using TEG generators attached to human skin as a heat source. They concluded that only low-power applications below $1\text{--}2\text{ mW}$ can run efficiently on body heat. These include some WPM applications such as a TEG-powered pulse oximeter. (The TEG was worn on the wrist like a watch.)

Figure 11.19 illustrates a simplified schematic of the Linear Technology Corp. LTC3108-1 energy harvester chip, showing its use to power a wireless patient sensor. Note that in the LTC3108-1, the low DC voltage from the external TEG is converted to a higher AC voltage by chopping and by using a 1:100 step-up transformer, then the secondary voltage AC is rectified, regulated, and filtered. A battery can be connected to the V_{STORE} input terminal to provide backup, as shown in the figure. The maximum power output of the LTC3108 is in the range of $200\text{ }\mu\text{W}$ to 1 mW , depending on the specific TEG used, the ΔT ,

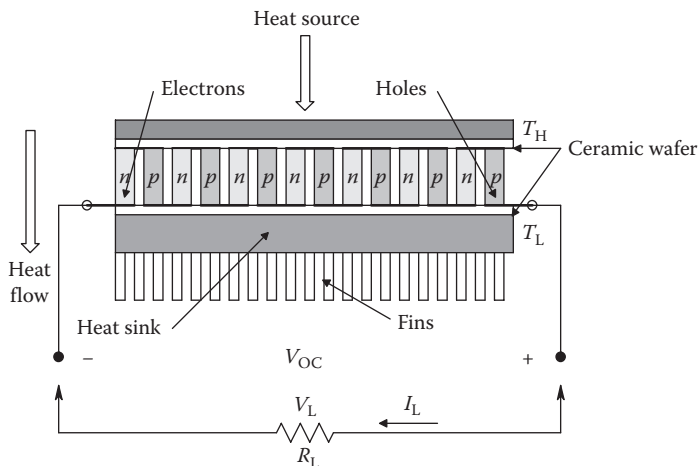


FIGURE 11.18

Cross section through a TEG. In general, the TEG OCV $\propto T_H - T_L$. Also, as T_H increases, so does the Thevenin resistance of the TEG. T_H is the high temperature of the heat absorber; T_L is the low temperature of the heat sink.

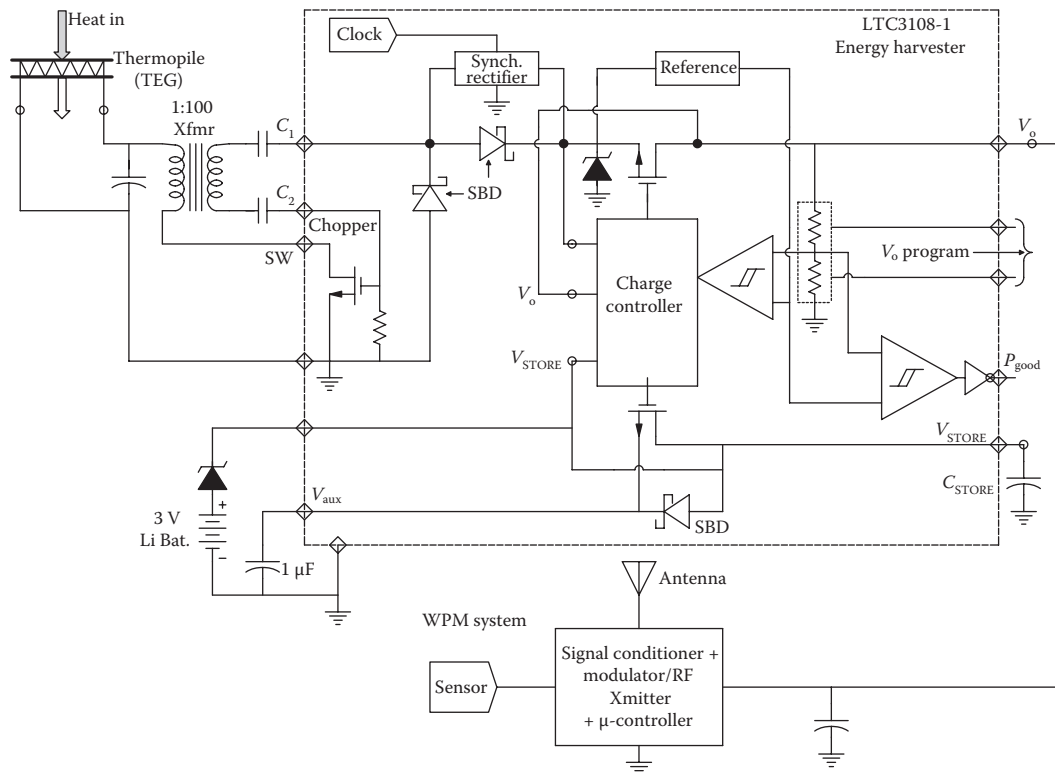


FIGURE 11.19

Author's simplified schematic of the Linear Technologies' LTC3108-1 energy harvester microchip. In this case, the power source is low DC voltage from a thermopile (TEG). The chip is shown powering a WPM system. See text for description.

the TEG's $V_{\max} \times I_{\max}$ product, and the chopper transformer turns ratio. The LTC3108-1 can be programmed to output 2.5 V, 3.0 V, 3.7 V, or 4.5 VDC to power the WPM sensor, MCU, and RF transmitter.

Details of energy harvester ICs that can be powered by TEGs, piezoelectric generators, and solar sources may be found in the Linear Technology *Journal of Analog Innovation*, LT (2010). Also see EDN (2012) for an article on energy harvester chips and applications.

M.A. Karami and D.J. Inman (2011, 2012) have investigated the feasibility of powering implantable cardiac pacemakers with the mechanical vibrational energy harvested from the human heartbeat. Their energy harvester power source is intended to trickle-charge pacemaker batteries, which generally supply average power on the order of $1 \mu\text{W}$ to the pacemaker. The first piezoelectric generator they designed was a thin, zigzag (crank or meander design) piezoelectric cantilever, mass loaded at the far end. Tests indicated that it could produce ca. $10 \mu\text{W}$ when located *near* a beating heart (*not attached* to the heart pericardium), when the heart is beating at its normal rate. One, experimental, 11-member meander structure resonated close to 50 Hz. A problem that they solved was maintaining the piezoelectric cantilever's power sensitivity off resonant frequency.

To improve their harvester performance, the piezoelectric cantilever was modified to make it a nonlinear mechanical oscillator by adding permanent magnets to the system: one magnet was attached to the tip of the cantilever and two adjacent to the tip on the

fixed frame. The repelling force between the tip magnet and the frame magnets opposes the elastic restoring force of the beam. When the repulsive magnetic force between the tip and frame magnets exceeds the restoring elastic force of the beam, the zero deflection position becomes unstable. There are now two equilibrium points for the deflection of the cantilever on either side of center. Karami and Inman reported that their nonlinear beam was 27×27 mm, and the tip and frame magnets measure $25.4 \times 3.18 \times 3.18$ mm. Using the prototype nonlinear piezoelectric cantilever, the system delivered more than $3 \mu\text{W}$ of power over a simulated heart rate of 7–700 beats/min (bpm). At 280 bpm, the power output peaked at $17 \mu\text{W}$. Above 420 bpm, power output dropped to ca. $2 \mu\text{W}$. Typically, the prototype nonlinear harvester's power output was between 5 and $10 \mu\text{W}$ over most of this wide heart rate range. (No circuits or details on experimental test protocols were given in the Karami and Inman papers.)

Many physical phenomena powering low-voltage, energy-harvesting power sources have intermittent (quasiperiodic or aperiodic) outputs. These voltage sources include, but are not limited to, solar cells, wind turbines, wave and tidal power sources, and, as described earlier, piezoelectric sources driven by mechanical energy from quasiperiodic physical efforts such as breathing (chest expansion, diaphragm movement) and tissue vibrations from heart contractions, artery pulsations, etc. These intermittent electrical energy outputs must be rectified, smoothed, and stored to provide steady, conditioned, DC power sources for stand-alone MEM electronic systems.

The *electric double-layer capacitor* (EDLC) (also known as a supercapacitor or ultracapacitor) can be used for these storage applications. The intermittent power source output is made unipolar (rectified), and the charge is stored in an EDLC. An EDLC has a much higher power density than batteries. Also, the energy density of EDLCs is typically on the order of hundreds of times greater than a conventional electrolytic capacitor. For example, a typical D-cell (battery)-sized, low-voltage electrolytic capacitor may have a capacitance of several thousand microfarads; an EDLC of the same size and voltage rating may have a capacitance of *several farads*, an increase of two orders of magnitude. Larger EDLCs can have capacities of up to 5000 F. Certain EDLCs have reached specific energies (aka energy densities) of ca. 30 Wh/kg (1.8 kJ/kg) (Zheng 2002, Jeol.com. 2007). As is well known, a discharged capacitor's voltage rises exponentially when charged from a constant voltage DC source voltage in series with an equivalent resistance, R_s . The charging time constant is simply $R_s C$ seconds. Likewise, a capacitor discharges its voltage exponentially into a resistive load, R_L , with time constant $R_L C$ seconds. A supercapacitor can be charged more quickly (linearly) by an electronic current source. The total electric energy stored in a capacitor is given by the well-known relations:

$$W = \frac{1}{2} C V_c^2 = \frac{1}{2} \frac{Q^2}{C} = \frac{1}{2} Q V_c \text{ joules, or Watt-seconds,} \quad (11.26)$$

where

C is the capacitance in Farads

V_c is the capacitor's terminal voltage in volts

Q is the charge stored in the capacitor in coulombs: $Q \equiv C V_c$

For example, a 5000 F EDLC charged to 3 V holds 22.5 kilojoules (kJ) of energy. It is enough to run a power tool or emergency lantern or buffer a single solar panel!

As we have seen, EDLCs are low-voltage devices, a property that makes them useful for energy-harvesting applications required to power implantable, medical MEMS.

Other *macro* applications of EDLCs include (1) the power source for diesel engine starting in tanks, locomotives, submarines, and trucks; (2) their use as energy accumulators in dynamic braking systems in electric vehicles; (3) their use in powering electric buses (capabus) (fast recharge makes this possible); (4) their use in portable power tools; (5) their use in portable communications devices; and (6) their use as backup power for emergencies (radios, flashlights).

The advantages of using EDLCs as energy storage buffers include (1) long life, with little degradation over hundreds of thousands of charge/discharge (C/D) cycles; (2) low cost per C/D cycle; (3) very high rates of C/D (no chemical reactions are involved); (4) very low internal resistance, hence low heating levels during deep C/D; (5) high specific output power that can exceed 6000 (J/s)/kg at 95% efficiency; (6) no corrosive electrolytes and low material toxicity (e.g., activated carbon from coconut shells); and (7) simple charging—no full charge detection is needed as with batteries. There is no danger of overcharging. *Disadvantages of EDLCs include* the following: (1) The maximum voltage that individual EDLCs can operate at is from 2 to 3 V; (2) Complicated electronic interfaces are required to step discharge voltage up and maintain a constant output voltage (or current) as the EDLC voltage falls; (3) EDLCs can be operated in series to increase maximum stored voltage, but individual capacitors require voltage-balancing electronics to prevent overvoltages; (4) EDLCs have the highest dielectric absorption of any type of capacitor; (5) Charged EDLCs can lose their charge (self-discharge) faster than most electrolytic capacitors and faster than a rechargeable battery (secondary cell); and (6) The amount of energy stored per unit weight is generally less than that of an electrochemical battery (ca. 500 J/kg for a standard EDLC vs. 9.6 kJ/kg for a typical Li-ion battery).

11.7 Absorbable Electronic Circuit Implants

Recently, a new class of IC fabricated from silicon, magnesium, and silk protein has been designed for use in implantable biosensors and environmental sensors. These circuits, called *transient electronics*, have been described by Hwang et al. (2012). The outstanding feature of this class of circuits is that they are water soluble; hence, after a predetermined time in vivo or in situ, they self-destruct, literally dissolving. In the case of bioimplants, they literally biodegrade to harmless compounds and ions, which can be metabolized and/or excreted. For example, a piece of Si nanomembrane (NM) of dimensions $3 \times 3 \text{ nm} \times 70 \text{ nm}$ thick reacts with water forming silicon hydroxide, thus:



or



In phosphate-buffered saline (pH = 7.4) at 37°C, the Si NM's dissolution rate was 4.5 nm/day (Hwang et al. 2012). Si is used for diodes, photodiodes, BJTs, and MOSFETs in the transient circuit designs.

Magnesium metal was used as conductors (or resistors) in the *transient* electronic circuits: a 150 nm thick Mg *wire* deposited on top of a 10 nm thick MgO insulator on a processed

silk pad was immersed in deionized water. The Mg and MgO were observed to dissolve completely in ca. 3 h in the Hwang et al. paper, forming magnesium hydroxide. The reactions are as follows:



or



These authors also stated that the processed silk protein circuit substrates were also biodegradable (similar to silk sutures).

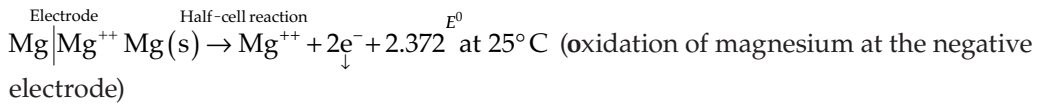
The lifetime of an implanted transient electronic circuit is determined by the time required for its silk protein overcoat to dissolve and let body fluids reach the Si and Mg circuit elements, and the silk protein substrate, dissolving them.

A simple, planar, demonstration circuit was shown consisting of (1) a helical Mg/MgO inductor, ca. 4 mm in diameter; (2) Mg interconnecting wires; (3) a Mg resistor (thin Mg wire); (4) a Si diode; (5) a Mg/MgO capacitor (MgO is the dielectric); and (6) a Si/MgO/Mg transistor. A histological section of the implant site in a mouse was done 3 weeks postimplant. It showed nearly complete absorption of the transient circuit's metals and partial absorption of the silk substrate film (Hwang et al. 2012).

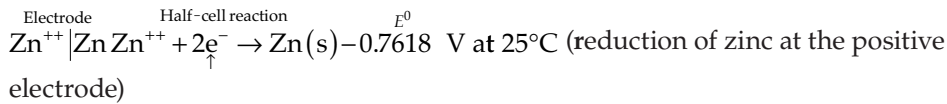
Ongoing work by Hwang et al. is being done to design more complicated transient circuits for in vivo and in situ measurement applications. Some prototype absorbable circuit modules include strain sensors based on Si NM resistors, arrays of addressable Si NM photodetectors with blocking diodes, arrays of *p*-channel MOSFETs that use Mg source (S), gate (G), and drain (D) electrodes and MgO gate dielectric, simple MOS logic gates, etc. Other circuit modules under consideration are an RLC oscillator. This could be made voltage tunable by using a varactor diode capacitor. *pn* diodes can also be used to measure temperature or ionizing radiation, and Schottky diodes can be used to quantify certain chemical species.

A major problem with the design of completely absorbable implanted circuits is the power source. Batteries, such as Li cells, are not suitable. However, there are other possible sources: (1) RF power can be inductively coupled through the skin and rectified and filtered to a DC voltage (currently used); (2) Thermopiles (the thermocouple materials would have to be absorbable and biocompatible. The working ΔT would be between an animal's deep core temperature and cooler, cutaneous temperature); (3) Power sources for transient electronic sensors used in environmental measurements could also use silk membrane-covered Si photoelectric arrays; and (4) Electrochemical cells using the Cl⁻ ions in extracellular fluid (ECF) are another possibility. Continuously exposed to ECF, the battery electrodes would have to last the design duration of the transient implanted circuit, hence be thicker than the Mg and Si circuit wires and components and also be biocompatible. Clearly, there is a need for a nontoxic, biodegradable primary electrochemical cell (nonrechargeable battery) for use with transient circuit implants. We note that Au, Fe, and Zn are also biocompatible metals. Gold metal does not biodegrade (oxidize) and is harmless in small quantities. Iron (Fe⁺⁺) is found in hemoglobin and myoglobin, and zinc is an important ion in many important protein enzymes, in particular *carbonic anhydrase*, *lactic dehydrogenase*, and certain *peptidases* used in the digestion of proteins. Zinc is also found in the intestinal mucosa, kidney tubules, and in the epithelial cells of many glands in the

body. If a primary cell were made from Zn and Mg electrodes, it would have a standard potential of ca. $E_{Cell}^0 = 1.61$ V at 25°C, according to the following redox half-cell reactions:

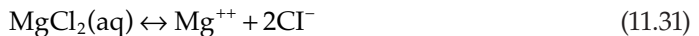


Anode (– bat. terminal: e[–]s lost from Mg anode)

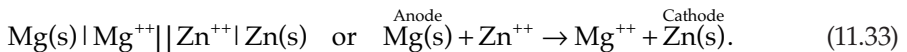


↑ Cathode (+ bat. terminal: e[–]s enter Zn cathode)

Note that ionized MgCl₂, ZnCl₂, and chloride ions are involved, that is, Cl[–], Mg⁺⁺, and Zn⁺⁺. Thus,



The Mg/Zn cell can be written as



The standard cell potential at 25°C is $E_{Cell}^0 = E^0$ (oxidation) + E^0 (reduction) = +2.372 V + (–0.7618 V) = 1.610 V.

If an iron cathode were used, $\text{Fe}^{++} + 2\text{e}^- \rightarrow \text{Fe(s)}$ $E^0 = -0.44$ V, and $E_{Cell}^0 = 1.93$ V.

Besides the problem of supplying *no-touch* power for internal WPM sensors, another challenge that will limit the development of complicated transient electronic circuits is the design of complex Si LSI circuits to function on the soluble silk protein substrate.

11.8 Chapter Summary

In this chapter, we have described various chemical microsensor designs, including Schottky diode–based gas sensors, MOX (tin oxide) gas sensors, and chemical sensors based on the ubiquitous MOSFET (e.g., the chemFET, the ISFET). The evolution of arrays of gas sensors into E-noses and E-nose applications was described.

The EM spectrum was described. Radio ICs for VHF and UHF WDX were listed, and antennas for these chips, including various space-saving fractal designs, were covered.

WPM systems were reviewed, including RF communications protocols used.

Power sources for wireless sensors, effectors, and WDX systems were covered, including energy-harvesting technologies. Lastly, we described the emerging technology of absorbable ICs for sensor implants used in animal ecology.

Problems

- 11.1** The attenuation constant of an EM plane wave in a lossy medium is given by the parameter α . The *skin depth* δ of an EM wave in a lossy medium is defined as the distance $\delta = \alpha^{-1}$ it must travel in order to reduce its amplitude by $e^{-1} = 0.368 = 36.8\%$ (see Balanis 1989, Ch. 4). (This problem has relevance to implantable wireless sensors.)
- Calculate and plot δ in m vs. frequency in Hz for adipose tissue (log-log plot) over 10^3 – 10^7 Hz. Given: $\sigma = 0.3$ S/m over 10^3 – 10^7 Hz; $\epsilon_r = 10^4$ at 1 kHz, 400 at 10 kHz, 40 at 100 kHz, 20 at 1 MHz, and 10 at 10 MHz; and $\mu_r \cong 1$ over 10^3 – 10^7 Hz.
 - Calculate and plot δ in m vs. frequency in Hz for muscle (log-log plot) over 10^3 – 10^7 Hz. Given: $\sigma = 0.4$ S/m over 10^3 – 10^7 Hz; $\epsilon_r = 4 \times 10^5$ at 1 kHz, 2.5×10^4 at 10 kHz, 8×10^3 at 100 kHz, 1.05×10^3 at 1 MHz, and 1.05×10^2 at 10 MHz; and $\mu_r \cong 1$ over 10^3 – 10^7 Hz. (Parameters from Gabriel et al. [1996] and Collins et al. [2002].)
- 11.2** Describe as many intermittent, low-level, energy sources (AC and DC) as you can that might be used to power an *energy-harvesting* IC power source for wireless sensors and WDX. (Don't forget photonic and electrochemical sources.) Describe the power levels involved.
- 11.3** An n -channel depletion ISFET is used to sense the level of an organic pollutant in a seawater sample.

The sensor circuit is zeroed by placing the ISFET in analyte-free seawater of the same salinity and temperature as the sample. The ISFET is operated in its saturated drain mode. The measurement is difficult because of the fact that the QUM must be detected as a DC level in the presence of white noise, $1/f$ noise, and f^n noise. (Refer to Equation 11.6.) There is also noise from the signal-conditioning OA (see Figure P11.3):

- Find the mean-squared output noise voltage/Hz, v_{on}^2 for (1) the noise from the solution (cf. Equation 11.5), (2) the noise from the gate noise voltage, (3) the noise from the gate equivalent input current noise, and (4) the noise from the OA and R_F . Given: $4kT = 1.656 \times 10^{-20}$, the solution's $R_E = 10^4 \Omega$, ISFET $I_{DSS} = 1.0$ mA, $V_T = -3$ V, $v_{GSQ} = 0$ V, $C_{gs} = 1$ pF, $g_{d0} = 10^{-4}$ S, $D = 10^{-17}$, $F = 10^{-20}$, $\gamma = 0.67$, $\alpha = 1.0$, $\delta = 1.33$. OA $e_{na} = 3$ nVRMS/ $\sqrt{\text{Hz}}$, $i_{na} \approx 0$, $R_F = 10^4 \Omega$. (You will need to find numerical values for the ISFET's perveance, K , and g_m at $i_D = I_{DSS} = 1.0$ mA.)
- Find an expression for the system's MS output noise voltage. The Hz noise bandwidth is $B = 1.1 - 0.1 = 1.0$

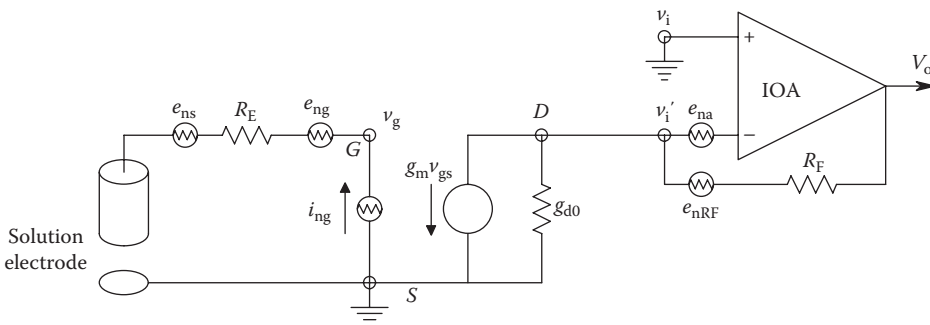


FIGURE P11.3

- C. When the analyte is added to the seawater, there is a DC shift in v_o , $\Delta v_o = \Delta v_{gs} g_m R_F$, which must be resolved above the total output RMS noise. Find the Δv_{gs} that will give an RMS output $\text{SNR}_o = 10$ in the 1.0 Hz noise bandwidth, B .
- 11.4** A 1000 F supercapacitor is charged to $V_c = 3.0$ V:
- Find the energy stored in joules.
 - The capacitor is discharged at a constant current so its voltage follows $V_c(t) = 3.0 - \rho t$. The discharging is stopped automatically at $t = T_s = 60$ s when $V_c(T_s) = 1.50$ V. Find a numerical value for ρ for these conditions.
 - Find a numerical value for the discharge current, I_D , for these conditions.
 - How long in hours will it take to discharge the supercapacitor to $V_c = 1.50$ V with a constant discharge current $I_D = 1.0$ mA?
 - How much energy in joules is discharged from the capacitor at $t = 60$ s?
- 11.5** A SnO_2/CuO MEM CO_2 sensor was developed by Xu et al. (2006) (see Section 11.2.2.3). In this problem, it is used in a Wheatstone bridge to sense ppm increases in ambient CO_2 concentration above that in room air (400 ppm). See Figure P11.5; consider the electrometer differential amplifier (EDA) disconnected for this problem. Two sensors are used in the bridge. One (R_R) is kept in CO_2 -free air as a reference; its resistance is $6.7 \times 10^8 \Omega$. From Fig. 2 in the Xu et al. paper, we can approximate the measurement sensor's resistance by $R_M = 6.7 \times 10^8 \Omega - 1.5 \times 10^4 \Omega/\text{ppm CO}_2$. Both sensors are run at 450°C . Let $R_1 = R_2 = 10^3 \Omega$, $V_B = 10\text{Vdc}$:
- Derive an expression for the DC bridge output V_o in volts/ppm CO_2 .
 - Now, consider the thermal (Johnson) noise arising in the bridge resistors that interferes with the resolution of the CO_2 measurements. Consider R_R and R_M to be $6 \times 10^8 \Omega$ at 450°C . Assume the noise bandwidth is 2 Hz. Also, let $R_1 = R_2 = 10^3 \Omega$ and be at 27°C . Find the total MS noise voltage at the V_o nodes.
 - What must the increment in $[\text{CO}_2]$ be in ppm for the DC V_o to be 10 times the total RMS noise at the V_o terminals?
- 11.6** Consider again Figure P11.5 with the amplifier attached. The EDA has an equivalent, single-ended, short-circuited input, white voltage noise root power spectrum, $e_{na} = 16.0$ nVrms/ $\sqrt{\text{Hz}}$. The EDA also has a white current noise root power spectrum, $i_{na} = 10.0$ fArms/ $\sqrt{\text{Hz}}$. We wish to again find the minimum $[\text{CO}_2]$ in ppm to give an

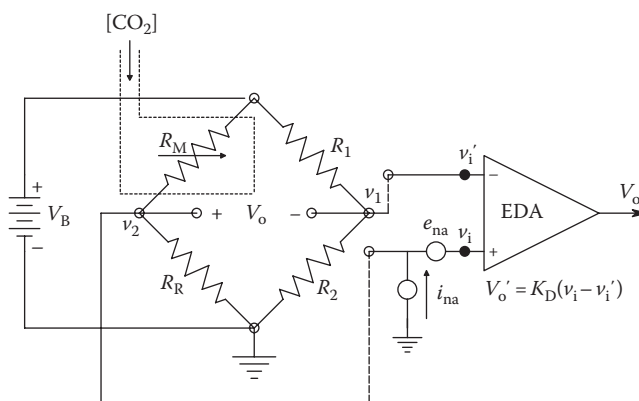


FIGURE P11.5

RMS SNR ratio of 10 at V_o' . Consider the thermal noise from R_R and R_M , as well as e_{na} and i_{na} . Let the EDA have a differential DC gain of $K_D = 10^3$.

- 11.7** Figure 11.18 illustrates a sectional side view of a TEG using many pn junctions. Note that a thermal EMF (V_{OC}) is generated because electrons are concentrated at the cold ends (at T_L) of the n materials, while holes are dominant in the cold ends of the p materials (see Marlow 2013). The TEG can be simply modeled by a Thevenin open-circuit voltage (V_{OC}) in series with a Thevenin equivalent resistance, R_G . Note that V_{OC} is a function of $T_H - T_L$:
- A. Find an expression for the R_L value that will extract maximum power from the TEG, and also that maximum power, given V_{OC} and R_G of the TEG.
 - B. Under maximum power delivery conditions, find an expression for the internal power dissipation of the TEG.

12

Introduction to Mechanical Microsensors

12.1 Introduction to Microelectromechanical Systems

Microelectromechanical systems (MEMSs) have a wide range of sensor applications, from measurement of angular rotation (rate gyro), linear acceleration, fluid pressure, and fluid viscosity to chemical and biological analysis. Chemical and biological analysis is generally done by measuring the change in cantilever or sprung mass resonant frequency as analyte molecules or cells bind to specific ABs or other specific chemical binding sites attached to the vibrating cantilever's surface. The development of MEM sensors goes back over 25 years; it is a rapidly growing field with many novel applications. See the review papers on MEMS by Schmidt and Howe (1987), Raiteri et al. (2001), Baltes et al. (2002), Huang et al. (2005), Vashist (2007), Finot et al. (2008), and Cox et al. (2012).

12.1.1 Nanoelectromechanical Systems

Ultraminiaturized MEMSs are called nanoelectromechanical systems (NEMSs). (See the review articles on NEMS by Gebeshuber 2012, Ekinici and Roukes 2005, Craighead 2000, and Roukes 2000.) The scale of NEMS is generally on the order of single micrometers (μm) to hundreds of nm, depending on application. One widely studied NEM structure is the mechanically resonant, double-clamped, silicon beam. Beams with lengths on the order of $10\ \mu\text{m}$ and thicknesses about $100\ \text{nm}$ have been fabricated with resonant frequencies from 2 to $134\ \text{MHz}$. These NEM beams have high mechanical Q_s ranging from 10^3 to 10^5 (Ekinici and Roukes 2005). Roukes (2000) reported that a doubly clamped Si beam of length $L = 0.1\ \mu\text{m}$, width $w = 0.01\ \mu\text{m}$, and thickness $t = 0.01\ \mu\text{m}$ had a mechanical resonant frequency of $7.7\ \text{GHz}$.

One means of mechanically exciting a NEM beam is to pass an alternating current through the beam at right angles to a DC magnetic field, \mathbf{B} . This generates an AC mechanical force perpendicular to the beam, vibrating it. In vector notation, $d\mathbf{F} = i d\mathbf{l} \times \mathbf{B}$, where $d\mathbf{l}$ is in the direction of the current, $i = I_0 \cos(2\pi ft)$, and the differential vector force element $d\mathbf{F}$ is the vector cross product of $i d\mathbf{l}$ and \mathbf{B} . In the desired case of orthogonal vectors and a beam of length L , $F/L = iB\ \text{N/m}$. The deflection amplitude of a vibrating beam can be measured electronically (electrostatically or electromagnetically) or even optically (see Ekinici and Rouke 2005). Note that as in the case of MEM cantilevers, an analyte bound to the NEM beam lowers its resonant frequency or reduces its vibration amplitude if the beam is excited at its fixed, no-load resonant frequency.

12.2 MEM Accelerometer and Pressure Sensor ICs

12.2.1 MEM Linear Accelerometers

MEM accelerometers are widely available from many manufacturers: see the comprehensive listing by Sensorportal (2013). A nonexhaustive list of MEMS accelerometer manufacturers includes Analog Devices; Honeywell; Bosch; Delco; DigiSens; Freescale Semiconductor; Honeywell; Kionix; Memsic; mindsensors.com; Murata; Panasonic; Rieker, Inc.; Sentera Technology; Silicon Designs, Inc.; Silicon Sensing Systems/Japan; Spectrum Sensors and Controls, Inc.; STMicroelectronics; and IC Sensors–Milpitas, CA. These accelerometers are offered with a wide range of sensitivities and ranges. Applications range from airbag deployment triggers (vehicle impact detection), antitheft devices, vibration monitoring, vehicle stability control, shock detection in shipping delicate systems, washing machine spin out-of-balance sensing, and autonomous aircraft stability control (along with XYZ rate gyro chips) to video games.

Figure 12.1A and B illustrates a simplified schematic (top view) of one of the early monolithic, micromachined, linear accelerometers, the now discontinued Analog Devices ADXL50. The proof mass was constrained to only move in the X -(input) direction. Micromachined *finger* electrodes protruded from the two sides of the mass bar in the Y -direction. Between each pair of protruding fingers on each side was a pair of fixed electrodes. Under zero acceleration conditions (Figure 12.1A), the proof mass bar is centered, and the total capacitance between both the A and B fixed electrodes and the finger electrodes is C_0 (in the ADXL50 accelerometer, there were 42 side fingers, 21 on a side). When the case (and mass) is given a constant acceleration to the right ($+x$ -direction), the Newtonian force, $F = M\ddot{x}$, acts on the mass, moving it a distance Δx so that the spring restoring force of the tethers equals the acceleration force. That is, when $F = M\ddot{x} = K_s \delta x$, the steady-state mass deflection is

$$\delta x = \frac{M\ddot{x}}{K_s} \text{ m.} \quad (12.1)$$

The displacement, δx , linearly *increases* the capacitance between the B electrodes and the fingers and *decreases* the capacitance between the A electrodes and the fingers. The A electrodes are driven by a 1 MHz square wave; the B electrodes are driven by the square wave's complement (180° phase shift). The voltage on the A electrodes has a 3.4 V DC mean, and the B electrode voltage has a 0.2 V DC mean. The total C_0 between each set of fixed electrodes and the mass *fingers* is 0.1 pF. At a full-scale acceleration of 50 g, δx is such that $\Delta C_{\max}/C_0 = 0.1$. The mass of the movable bar and fingers is given as $0.1 \mu\text{g} = 10^{-10} \text{ kg}$, and the minimum resolvable δx is said to be $0.2 \text{ \AA} = 20 \text{ pm}$. The total area of the ADXL50 MEM accelerometer is $500 \times 625 \mu\text{m}$. The output sensitivity is 19 mV/g ($1 \text{ g} = 9.8 \text{ m/s}^2$). The ADXL50's frequency response is low pass and goes from DC to -3 dB at 1 kHz and is down 20 dB at 10 kHz. Beam-tether mechanical resonance is at ca. 24 kHz. The equivalent input acceleration noise of the ADXL50 has a 6.5 RMS $\text{mg}/\sqrt{\text{Hz}}$ noise bandwidth. Another discontinued Analog Devices' micromachined, linear accelerometer model, the ADXL105, has a full-scale input range of $\pm 5 \text{ g}$ and an equivalent input g -noise of 225 rms $\mu\text{g}/\sqrt{\text{Hz}}$. The ADXL105 also has a signal bandwidth of 0–10 kHz (-3 dB). The ADXL78 iMEMS accelerometer is a fourth-generation surface micromachined accelerometer. It has three full-scale ranges: 37, 55, and 70 g. It has a -3 dB frequency response at 400 Hz and an equivalent input noise of 1.1 $\text{mg}/\sqrt{\text{Hz}}$ over a 10–400 Hz bandwidth. Its major application is automotive airbag triggering. Another current (2013) Analog Devices' linear accelerometer is the digital output, MEMS ADXL362.

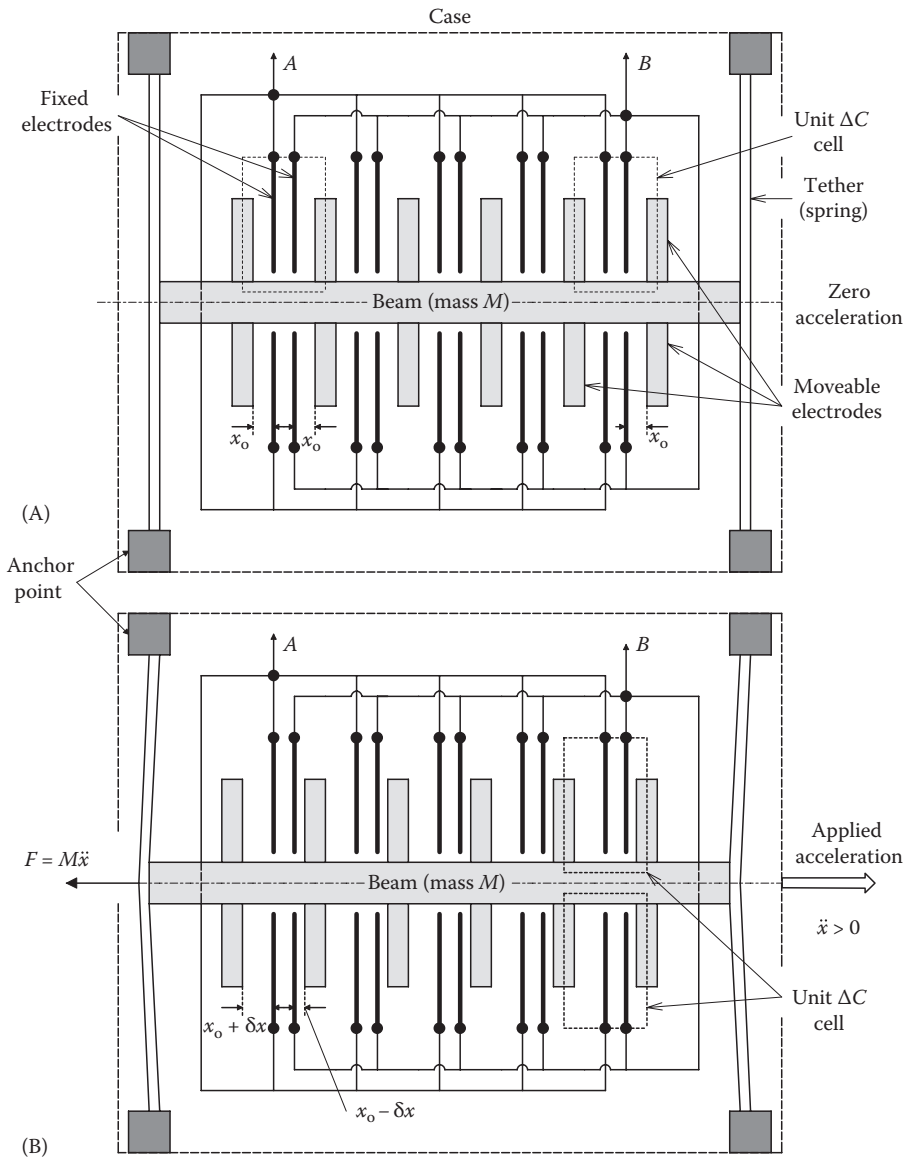


FIGURE 12.1 (A) Top view of an ADXL50-type micromachined linear accelerometer with $\ddot{x} = 0$. (B) Top view of the MEM accelerometer accelerating to the right. Note that the steady-state displacement of the beam mass, δx , is determined by an equilibrium between acceleration (Newtonian) force and spring (Hooke) force, so $\delta x = M\ddot{x}/K_s$, where K_s is the net spring constant.

Figure 12.2 shows a simplified schematic diagram of the electronic signal processing done onboard the ADXL50 accelerometer chip. The 42-beam *fingers* are reduced to one equivalent finger and the fixed side electrodes to a pair. The differential capacitor produces a double-sideband, suppressed-carrier (DSBSC) modulated signal at the summing junction of the buffer amplifier. The amplitude of this signal is proportional to ΔC , hence $|\ddot{x}|$. The phase of the signal depends on the sign of ΔC , hence \dot{x} . The ADXL50 clock frequency is 1 MHz. Note $\Delta C \propto \delta x$.

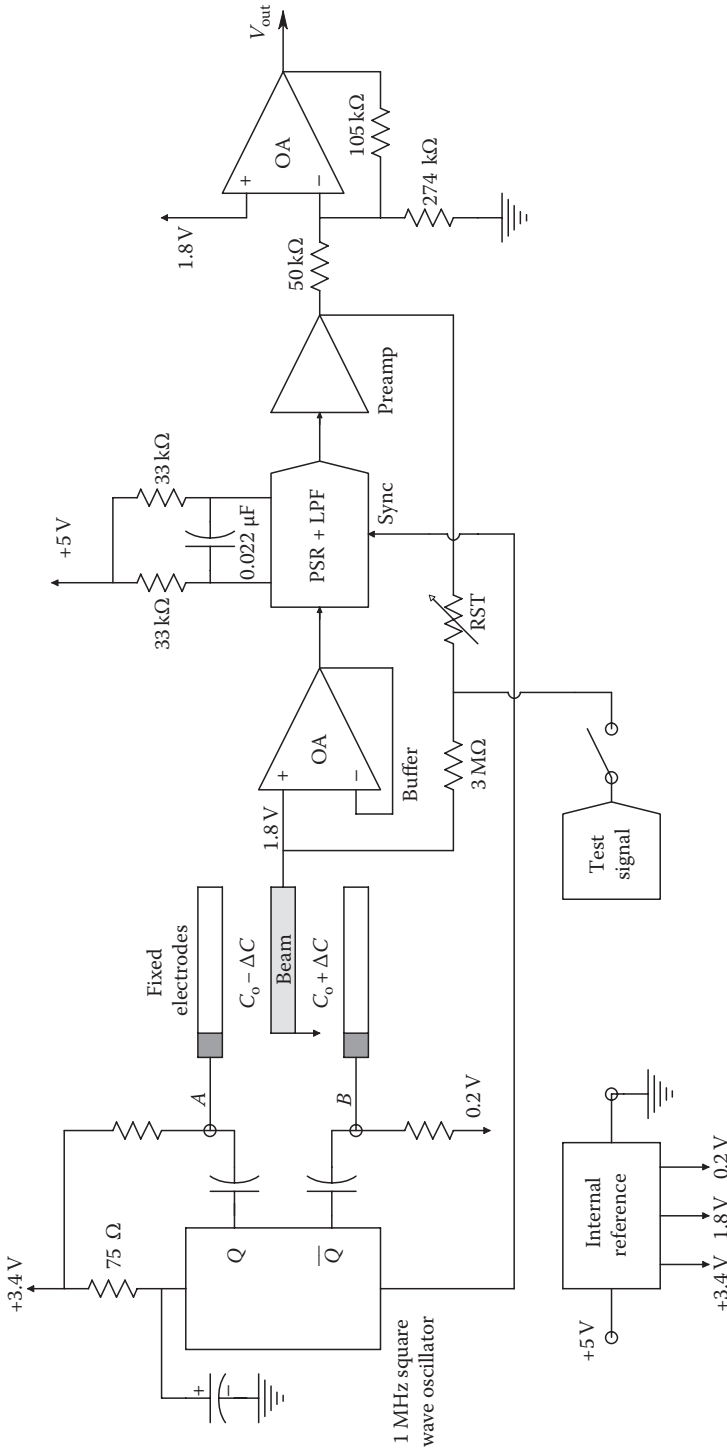


FIGURE 12.2 Simplified circuit of the mass deflection readout system in the ADXL50 accelerometer. Capacitance change ΔC is proportional to beam deflection, δx. ΔC is detected by a De Sauty bridge-phase-sensitive rectifier-LPF circuit.

Honeywell also offers a line of micromachined linear accelerometers in their ASA line of sensors.

Acceleration is sensed simply by a three-plate, differential capacitor. Fixed upper and lower silicon plates sandwich a spring-suspended middle plate whose mass responds to the applied acceleration by deflecting and causing the desired $C_o + \Delta C$ and $C_o - \Delta C$, which is detected by a bridge and phase-sensitive rectifier. The capacitive g sensor of the Honeywell ASA series is much simpler mechanically than the beam and finger architecture used by Analog Devices. Honeywell (2013) also makes general purpose piezoelectric and piezoresistive accelerometers.

12.2.2 MEM Pressure Sensors

MEM pressure sensors have many applications ranging from automotive (tire low air pressure warning, fuel pressure, manifold pressure, brake fluid pressure [at brake cylinders], etc.); physiological research (airway pressure, blood pressure, cerebrospinal fluid pressure, etc.); and heating, ventilation, and air conditioning (HVAC) industrial control to aerospace applications. Their advantage? They are small, relatively accurate, and inexpensive.

The simplest MEM pressure sensor is probably a micromachined diaphragm, which deflects proportional to the pressure difference across it. The diaphragm deflection can be sensed by several means: (1) piezoresistors bonded to the deflecting diaphragm acting through a Wheatstone bridge circuit, (2) a capacitance bridge, and (3) optically, by light beam deflection or interferometry. Signal conditioning is generally on the chip.

Welham et al. (1996) described a laterally driven, micromachined resonant pressure sensor. Their system employed a laterally driven, micromachined, resonant proof mass mounted centrally on a single silicon crystal diaphragm. The polysilicon resonator mass was supported by four beams, each 150 μm long, forming a hammock-type flexure. The resonator was excited electrostatically, and its motion is sensed by two 25-plate comb capacitors. The resonator's fundamental frequency was 58 kHz (at 0 pressure). It oscillated in the plane of the diaphragm in one dimension. Applied pressure deflected the diaphragm attached to the resonator and *raised* the resonant frequency linearly with pressure: 58–85 kHz for P ranging from 0 to 3.5 bar (1 bar = 14.5 psi). Their initial results showed that the vibrating mass pressure sensor has a sensitivity of 8.8 kHz/bar, and a $Q = 50$ in air.

Stampfer et al. (2006) described NEM pressure sensors they fabricated using single-wall, carbon nanotubes as piezoresistive strain sensors. Pressure-induced strain caused a nanotube resistance change from ca. 79 k Ω to ca. 88 k Ω as the Δp was raised from 0 to 120 kPa (1 kPa = 0.145 psi = 10^3 N/m²).

An interesting, proof-of-concept, 10-sensor, MEM array was developed by Kottapalli et al. (2012) that could measure fluid pressure and velocity. The array measured 60 mm (L) \times 25 mm (W) \times 0.4 mm (H) and was fabricated using liquid crystal polymer (LCP) backing and diaphragms to which are bonded metal piezoresistors. Wheatstone bridges are used to convert Δp -caused- ΔR to output voltages. The intended application of the designed pressure/velocity sensing array is in the guidance and control of autonomous underwater vehicles (AUVs). The MEM pressure array had a pressure voltage sensitivity of 14.3 $\mu\text{V}/\text{Pa}$ and a fluid velocity resolution of 25 mm/s.

General Electric offers a wide range of MEM pressure sensors for many applications. Configurations include surface-mount chips, DIP packages, mechanical mounts, disposable medical pressure sensors, and dies for incorporation into sensor packages (GE 2013). Other manufacturers of MEM pressure sensors include Melexis (2013), STMicroelectronics (2013), and Motorola (2013). Motorola makes a wide variety of MEM pressure sensors that

support three types of pressure measurement: differential, gauge, and absolute. These sensors are all based on a patented design that uses a monolithic silicon piezoresistor, which generates a changing output voltage with variations in applied pressure. The piezoresistor is ion implanted on a thin Si diaphragm. The strain gauge is an integral part of the diaphragm; thus, there are no temperature effects due to differences in thermal expansion of the strain gauge and the diaphragm.

12.3 MEM Rate Gyros

12.3.1 Introduction

There are several MEM rate gyro designs. They may be broadly divided into those using the Coriolis forces acting on a flat proof mass given oscillating, *linear* displacement(s) along one or two axes (x, y) in the XY plane (Fujita et al. 1997, Seshia et al. 2002, Apostolyuk 2006, Dhillon and Grutter 2007, Sun et al. 2008) and those that use Coriolis forces associated with a disk or ring given angular oscillation around a *central, perpendicular axis* (Juneau et al. 1997, Ayazi and Najafi 2001, Martynenko et al. 2010). Other MEM gyro designs have used TFs (Bernstein et al. 1993, QRS14 2011) and vibrating beams (Maenaka and Shiozawa 1994).

Piazza and Stephanou (2002) have described the design of a prototype, unique, MEM thermofluidic rate gyro (see Section 7.2.2.4). Their MEM thermofluidic gyro makes use of the Coriolis force acting on a fluid flowing at constant velocity through a linear channel. The Coriolis force is the result of a rotational velocity Ω around a z -axis perpendicular to the fluid channel, which lies in the XY plane. In general, the differential force on a mass element δm with velocity \mathbf{v} is $d\mathbf{f}_c = 2\delta m (\mathbf{v} \times \Omega)$ N, where $\delta m = \Delta v \rho$, ρ is the fluid density in kg/m^3 , and Δv is a differential fluid volume. The vector cross product is taken.

Because fluid laminar flow generates a parabolic velocity distribution in the flow channel, \mathbf{v} varies in the y -direction symmetrically about the channel center (x -axis), with \mathbf{v}_{max} occurring at the center ($y = 0$) and $\mathbf{v} = 0$ at the channel walls. The net result is that \mathbf{f}_c due to Ω causes an asymmetrical skewing of fluid velocity along the y -axis. This asymmetrical flow causes a differential cooling of the two thermal sensors located on either side of the fluid channel. Thus, ΔT between the thermosensors is thus proportional to Ω . Piazza and Stephanou investigated the use of water as the fluid in a simulation of their gyro's performance. (No doubt a denser, freeze-resistant fluid could also be used.) Their simulation showed that their gyro responded linearly over $\Omega = \pm 150^\circ/\text{s}$, with a resolution of ca. $0.75^\circ/\text{s}$. (Not discussed in their paper was the means of circulating the fluid at constant velocity, either onboard the gyro chip or externally.)

12.3.2 Oscillating Disk Rate Gyro

A micromachined, oscillating disk, rate gyro was developed at UC Berkeley by T. Juneau for his PhD research (Juneau and Pisano 1996, Juneau et al. 1997). Figure 12.3 illustrates the basic geometry of this innovative angular rate sensor. In this first example, consider a thin disk is suspended on two torsion springs along the z -axis. The ends of the springs are attached to the gyro case (not shown). The disk is caused to oscillate sinusoidally at its resonant frequency, ω_m , around its spin (z) axis with a low amplitude, sinusoidal motion with peak value, Ψ . One of 12 electrostatic *motors* arranged around the circumference of

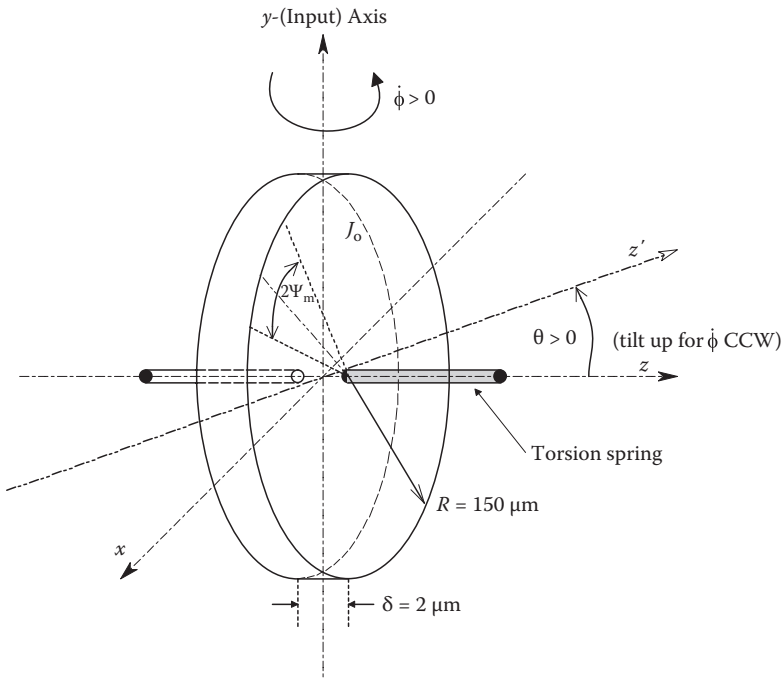


FIGURE 12.3

Micromachined, oscillating flywheel used in the MEM oscillating disk rate gyro. The oscillation around the z -axis is $\Psi(t) = \Psi_m \cos(\omega_m t)$, where $\omega_m = \sqrt{K_T/J_o}$ is the natural, mechanical resonant frequency of the disk around its z -axis. Angular velocity $\dot{\phi}$ input around the y -(input) axis causes a moment that tries to tilt the z -axis up in the ZY plane.

the disk is shown at 2 o'clock in Figure 12.4A. Electrostatic attraction/repulsion between each electrode fixed to the disk and the stationary drive electrodes is used to generate a periodic torque, which maintains a constant-amplitude disk oscillation around the z -axis. The undamped mechanical resonant frequency of the disk can be shown to be

$$f_m = \frac{1}{2\pi} \sqrt{\frac{K_\Psi}{J_o}} \text{ Hz} \tag{12.2}$$

where

K_Ψ is the net torsional spring constant of the disk
 J_o is its moment of inertia around its central (z) axis

The induced oscillation around the z -axis can be written as

$$\Psi(t) = \Psi_m \sin(\omega_m t). \tag{12.3}$$

Hence,

$$\dot{\Psi}(t) = \omega_m \Psi_m \cos(\omega_m t) \tag{12.4}$$

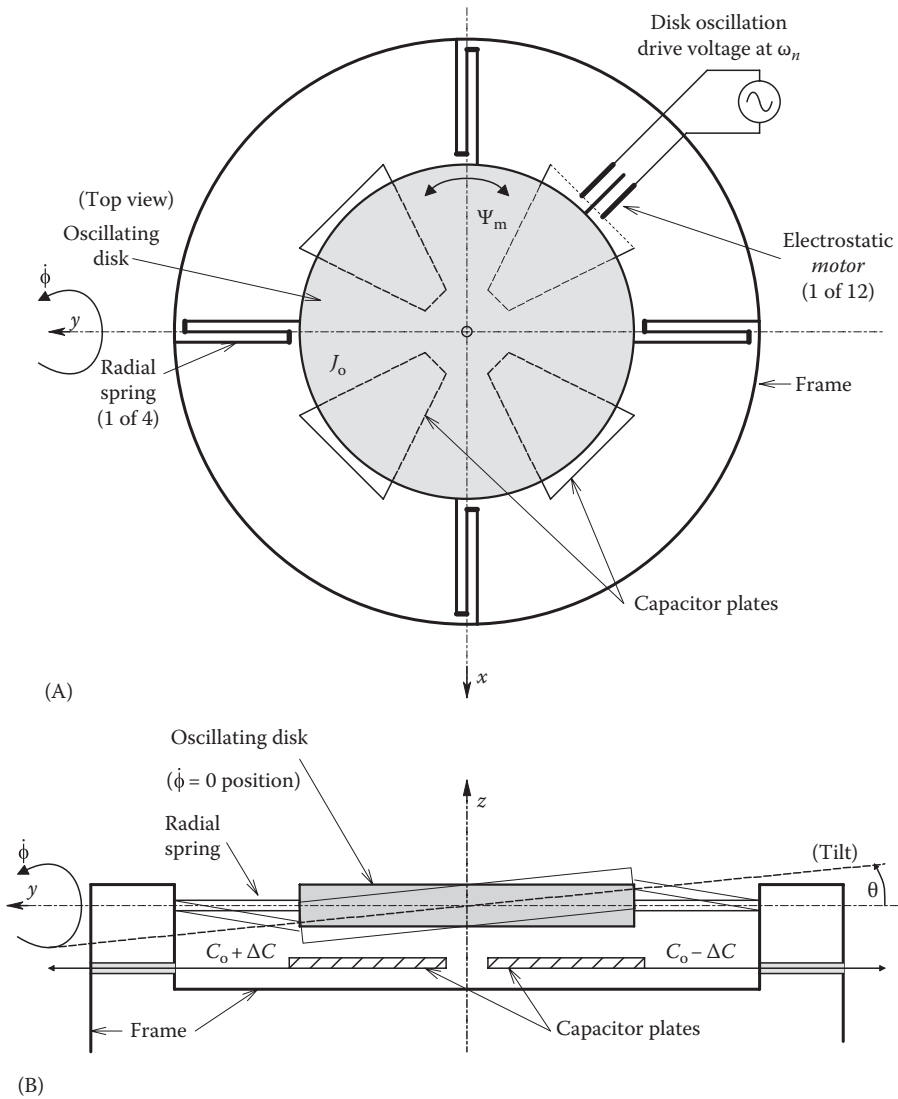


FIGURE 12.4

(A) Top view of an oscillating disk MEM rate gyro. Twelve electrostatic, capacitor motors drive the disk oscillations at ω_n r/s. (B) Side view of the oscillating disk MEM rate gyro. As a result of case input velocity, $\dot{\phi}$, the disk is forced to tilt as shown. The tilt is periodic and is sensed by a $\pm\Delta C$ to the fixed capacitor plates.

and

$$\ddot{\Psi}(t) = -\omega_m^2 \Psi_m \sin(\omega_m t). \tag{12.5}$$

The disk oscillation frequency in the Juneau disk gyro was on the order of 400 Hz.

It is observed that if the gyro case (and disk) disk is rotated CCW at a constant velocity $\Omega = \dot{\phi}$ about its input y -axis, a torque will be developed that tries to tilt the disk spin (z) axis up in the ZY plane. This tilt, θ , due to the torque can be measured to sense Ω .

The following mathematical development using the Lagrange equations (Cannon 1967) illustrates how the oscillating disk rate gyro works: The total *kinetic energy*, T , of the oscillating disk is

$$T = \frac{1}{2} J_o [\dot{\Psi} + \dot{\phi} \sin(\theta)]^2 + \frac{1}{2} J_1 [\dot{\theta}^2 + \dot{\phi}^2 \cos^2(\theta)] \quad (12.6)$$

where

J_o is the moment of inertia of the disk around the z (spin) axis

J_1 is its moment of inertia around the x - and y -axes

The *potential energy* of the disk, U , is stored in the torsion springs on the z -axis and the spring that restrains the tilt θ in the YZ plane:

$$U = \frac{1}{2} K_\Psi \Psi^2 + \frac{1}{2} K_\theta \theta^2 \quad (12.7)$$

Now the Lagrangian function for the system is $L \equiv T - U$, sic:

$$L = \frac{1}{2} J_o [\dot{\Psi}^2 + 2\dot{\Psi}\dot{\phi}\sin(\theta) + \dot{\phi}^2 \sin^2(\theta)] + \frac{1}{2} J_1 [\dot{\theta}^2 + \dot{\phi}^2 \cos^2(\theta)] - \frac{1}{2} [K_\Psi \Psi^2 + K_\theta \theta^2]. \quad (12.8)$$

There are three Lagrangian coordinates (q_k) for the disk gyro:

$q_1 = \Psi$: Rotor \angle on the spin (y) axis

$q_2 = \theta$: Rotor tilt \angle (θ in YZ plane)

$q_3 = \phi$: Input \angle around y -axis

1. For $q_1 = \Psi$,

$$\frac{\partial L}{\partial \Psi} = -K_\Psi \Psi \quad (12.9a)$$

$$\frac{\partial L}{\partial \dot{\Psi}} = J_o [\dot{\Psi} + \dot{\phi} \sin(\theta)] \quad (12.9b)$$

$$\frac{\partial}{\partial t} \left(\frac{\partial L}{\partial \dot{\Psi}} \right) = J_o [\ddot{\Psi} + \ddot{\phi} \sin(\theta) + \dot{\phi} \dot{\theta} \cos(\theta)]. \quad (12.9c)$$

Now the torque around the spin z -axis is given by

$$\frac{\partial}{\partial t} \left(\frac{\partial L}{\partial \dot{\Psi}} \right) - \frac{\partial L}{\partial \Psi} = M_\Psi. \quad (12.10)$$

We assume that the disk tilt is restrained by a very stiff spring so $\theta \rightarrow 0$ and $\dot{\theta} \rightarrow 0$. Thus, the torque required to oscillate the disk at constant amplitude is

$$M_\Psi \cong J_o \ddot{\Psi} - K_\Psi \Psi. \quad (12.11)$$

2. For $q_2 = \theta$,

$$\begin{aligned}\frac{\partial L}{\partial \theta} &= \frac{1}{2} J_o \left[2\dot{\Psi} + \dot{\phi} \cos(\theta) + -\dot{\phi}^2 (-2 \sin(2\theta)) \frac{1}{2} + \frac{1}{2} J_1 \dot{\phi}^2 \frac{1}{2} (-2 \sin(2\theta)) \right] - K_\theta \theta \\ &= J_o \dot{\Psi} \dot{\phi} \cos(\theta) + \frac{1}{2} J_o \dot{\phi}^2 \sin(2\theta) - \frac{1}{2} J_1 \dot{\phi}^2 \sin(2\theta) - K_\theta \theta\end{aligned}\quad (12.12a)$$

$$\frac{\partial L}{\partial \theta} = 0 \quad (12.12b)$$

$$\frac{\partial}{\partial t} \left(\frac{\partial L}{\partial \dot{\theta}} \right) = 0. \quad (12.12c)$$

Thus,

$$M_\theta = \frac{\partial}{\partial t} \left(\frac{\partial L}{\partial \dot{\theta}} \right) - \frac{\partial L}{\partial \theta} = -J_o \ddot{\Psi} \dot{\phi} + K_\theta \theta = -\dot{\phi} J_o \left[\omega_m \Psi_m \cos(\omega_m t) \right] + K_\theta \theta. \quad (12.13)$$

Equation 12.13 is an expression for the torque on the tilt axis, that is, the *output torque*. M_θ lies in the YZ plane.

Finally, we examine the torque around the input or y -axis:

3. $q_3 = \phi$:

$$\frac{\partial L}{\partial \phi} = 0 \quad (12.14a)$$

$$\frac{\partial L}{\partial \dot{\phi}} = J_o \dot{\Psi} \sin(\theta) + J_o \dot{\phi} \frac{1}{2} [1 - \cos(2\theta)] + J_1 \dot{\phi} \frac{1}{2} [1 + \cos(2\theta)] \quad (12.14b)$$

$$\begin{aligned}\frac{\partial}{\partial t} \left(\frac{\partial L}{\partial \dot{\phi}} \right) &= J_o \ddot{\Psi} \sin(\theta) + \frac{1}{2} J_o \ddot{\phi} - \frac{1}{2} J_o \ddot{\phi} \cos(2\theta) + \frac{1}{2} J_o \dot{\phi} \dot{\theta} 2 \sin(2\theta) + \frac{1}{2} J_1 \ddot{\phi} \\ &\quad + \frac{1}{2} J_1 \dot{\phi} \dot{\theta} \cos(2\theta) - \frac{1}{2} J_1 \dot{\phi} \dot{\theta} 2 \sin(2\theta).\end{aligned}\quad (12.14c)$$

Now, assuming $\theta \rightarrow 0$ and $\dot{\theta} \rightarrow 0$, the expression for the input torque reduces to

$$M_\phi = \frac{\partial}{\partial t} \left(\frac{\partial L}{\partial \dot{\phi}} \right) - \frac{\partial L}{\partial \phi} \equiv J_1 \ddot{\phi} = J_1 \dot{\Omega}. \quad (12.15)$$

Note that rotation rate of Juneau's oscillating disk gyro can be sensed around the x -axis as well as the y -axis. When the x -axis is the input axis, because of symmetry, the disk experiences a torque in the XZ plane. Figure 12.4B illustrates

schematically how the Juneau MEM gyro is suspended and how the small disk tilt is sensed by incremental capacitance changes. From Equation 7.81, we find

$$\theta = \dot{\phi} \frac{J_o \omega_m \Psi_m \cos(\omega_m t)}{K_\theta} = \Omega \frac{J_o \omega_m \Psi_m \cos(\omega_m t)}{K_\theta} \quad (12.16)$$

where

K_θ is the stiffness (torsional spring constant) of the four radial suspension springs in the YZ plane in Newton-meters/radian
 Ω is the input angle rate of change (velocity) that is sensed

The actual small angular tilt, θ , was sensed by measuring the change in capacitance using two pairs of capacitor plates located under the disk; the plates of a pair are opposite each other. Each pair of plates can be connected to a De Sauty bridge-type circuit. (The De Sauty bridge was described in Section 5.5.1.4.)

Note from Equation 7.84 that $\theta \propto \dot{\phi}$ and $\Delta C \propto \theta$, so after phase-sensitive demodulation of the bridge output, $V_o \propto \Omega$. Two orthogonal pairs of capacitor plates are required to resolve both disk tilt θ due to rotation $\dot{\phi}$ around the y -axis and also orthogonal disk tilt due to rotation rate $\dot{\alpha}$ around the x -axis (not shown). The De Sauty bridge drive voltage must have a frequency, $\omega_b \gg \omega_m$. The AC $[\cos(\omega_m t)]$ term in the bridge output is again phase sensitive rectified to yield a DC $V'_o \propto \Omega_y$. As we commented earlier, the Juneau oscillating disk gyro can be used to sense *two* angle rates simultaneously, for example, roll *and* pitch (Juneau and Pisano 1996). Thus, the outputs of the two capacitance bridges must be functionally combined to resolve the exact disk tilt vector, which is the result of simultaneous Ω_y and $\dot{\alpha}$ inputs. (It is left as a chapter exercise to derive an expression for the tilt due to a rotation rate, $\dot{\alpha}$, around the x -axis and to design at the block diagram level an electronic system to resolve the net tilt due to $\dot{\phi}$ and $\dot{\alpha}$.) Juneau reported that his micromachined (MEM) disk gyro prototype could resolve $\dot{\phi} \geq 0.1667^\circ/\text{s}$ in a 0–25 Hz bandwidth.

In the 47 years following Juneau's MEM gyro prototype, there has been much activity to develop sensitive, micromachined rotation rate sensors. The interested reader should consult the early review paper by Yazdi et al. (1998) and the paper describing more recent results in this R&D area by Acar and Shkel (2004) and Apostolyuk (2006). More recently, Sun et al. (2008) described a novel, vibrating mass, MEM rate gyro that used optical interferometry to read out mass tilt caused by $\Omega = \dot{\phi}$ (input angular velocity). An oscillating, rectangular proof mass was also used in a MEM rate gyro described by Seshia et al. (2002). As an example of MEM gyros, Silicon Sensing Systems in Japan (2013) markets a family of ultraminiature, MEM silicon ring gyros. Their CRM100 gyro package measures $4.8 \times 5.7 \times 1.2$ mm and has both analog and digital outputs. The gyro's range can be configured to span $\pm 75^\circ/\text{s}$, $\pm 150^\circ/\text{s}$, $\pm 300^\circ/\text{s}$, and $\pm 900^\circ/\text{s}$. Gyro bandwidth can be adjusted up to 150 Hz.

An ultrasensitive MEM rate gyro application was described by Arnaudov and Angelov (2005). Their system was used to measure the Earth's rotation rate. The Earth's nominal rotation rate is $\Omega_E = 360^\circ/24 \text{ h} = 15^\circ/\text{h} = 4.178074 \times 10^{-3}^\circ/\text{s}$, so direct use of MEM gyros is difficult; DSP must be used. Arnaudov and Angelov used a BEI Technologies, Inc., Systron Donner Inertial Division, type HZI-90-100A MEM gyro. This micromachined, vibrating quartz, TF sensor has a standard range of $\pm 90^\circ/\text{s}$, an output scale factor of $22.2 \text{ mV}/(^\circ/\text{s})$, a resolution and threshold of $0.004^\circ/\text{s}$, an x -axis g sensitivity $< 0.006 (^\circ/\text{s})/g$, and a bias short-term stability (over 100 s, at constant temp.) of $< 0.05^\circ/\text{s}$. Their approach to resolve the very noisy and biased Ω_E signal was to periodically reorient the gyro from being aligned with

the Ω_E axis to being \perp to it (zero Ω_E input). Each measurement consisted of averaging data from $M = 3,000\text{--}10,000$ cycles at each of the two sensor positions. (At each individual sensor position, 256 samples at a 2 ms rate were taken.) They found that $\Omega_E = 0.00454^\circ/\text{s}$, a total error of 8.7%.

12.3.3 Vibrating Mass MEM Gyros

A vibrating mass MEM gyroscope is shown schematically in plan view in Figure 12.5A and B. It consists of an *outer case* (generally fastened to the vehicle), a *decoupling frame* having mass M_1 , and an inner, oscillating proof mass, M_2 , which is attached to the decoupling frame by springs. The decoupling frame in turn is attached to the outer case by springs.

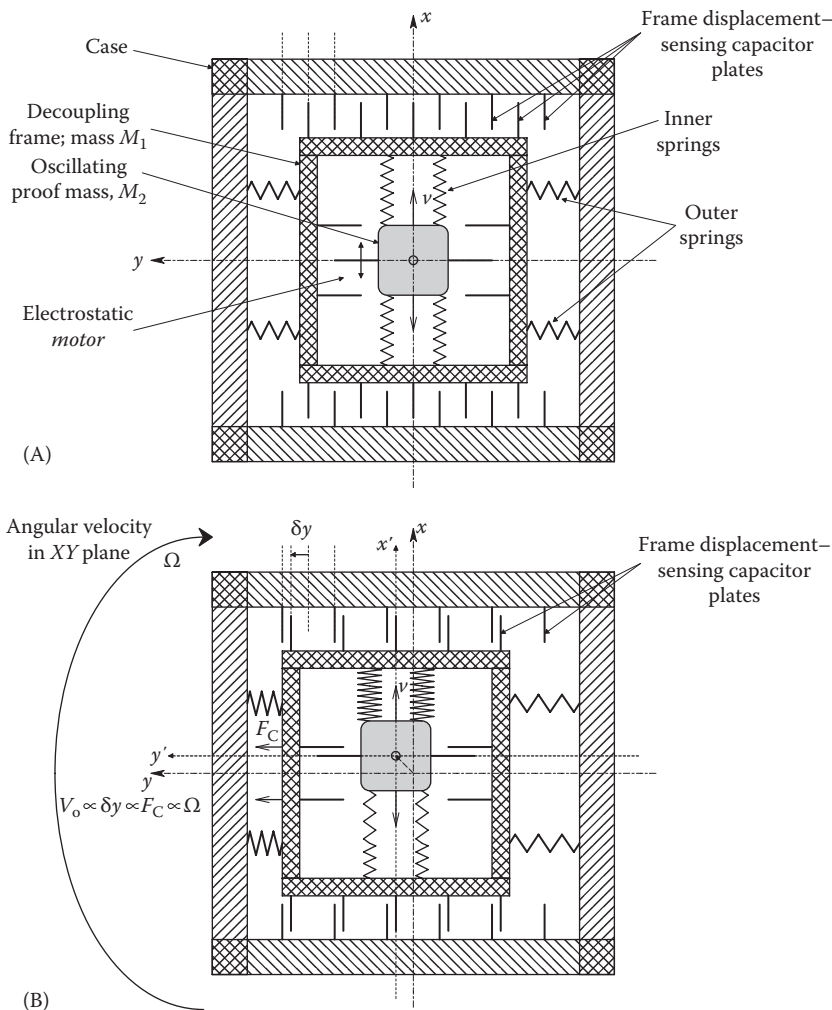


FIGURE 12.5 (A) Top view of a linearly vibrating mass, MEM rate gyro when the case input velocity, Ω , around the z - (input) axis is zero. Note the inner decoupling frame suspending the mass and in turn suspended from the case. (B) When the case input angular velocity to be sensed, Ω , is clockwise, as shown, the decoupling frame is forced to the left by Coriolis force as the oscillating mass moves in the $+x$ -direction, and vice versa.

The input or sensing axis (Ω) of this MEM gyro is the z -axis (out of page). Shown at the top and bottom of the case and decoupling frame are a series of normally equally spaced capacitor plates used to sense frame (and proof mass) y -displacement due to Coriolis force acting on the mass. For purposes of analysis, the proof mass M_2 is made to oscillate in the $\pm x$ -direction in the XY plane at frequency ω_m , using electrostatic *motors*. That is, $\delta_x = \delta_{x0} \sin(\omega_m t)$. Since the input (z) axis is orthogonal to the XY plane of rotation, the mass M_2 (hence the decoupling frame) experiences a net Coriolis force of $F_C = 2M_2 v_x \Omega = 2M_2 \delta_{x0} \omega_m \cos(\omega_m t) \Omega$ N. This AC force has the same frequency as the M_2 drive displacement and is directed along the $\pm y$ -axis. Thus, the AC Coriolis force causes the decoupling frame to oscillate laterally with a magnitude proportional to Ω . The $F_C(t)$ acting on the frame's y -springs causes a small oscillating displacement of the frame (including M_2), δy , which in turn causes a small $\Delta C(t)$ in the capacitor plates. The $\Delta C(t)$ is demodulated and any DC offset is removed to obtain a $V_o \propto \Omega$.

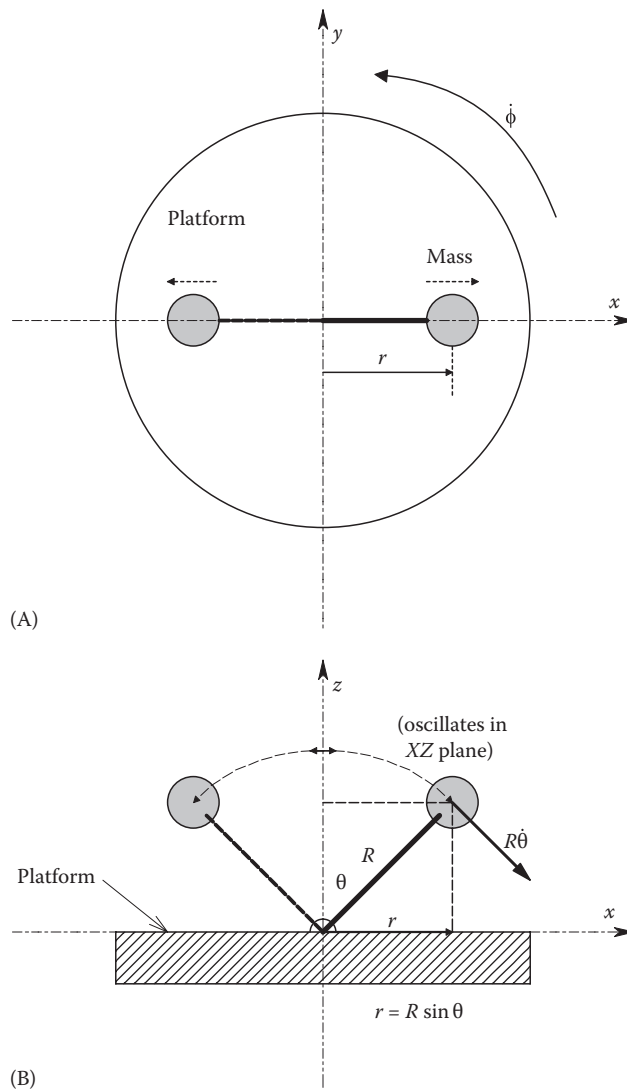
In the now discontinued Analog Devices' $\pm 150^\circ/\text{s}$, ADXRS150 MEM yaw rate gyro chip, the on-chip demodulator electronics could resolve capacitance changes as small as 12×10^{-21} F (12 zF) from frame deflections as small as 16 fm (16×10^{-15} m) (Geen 2003). (1 Å unit = 10^{-10} m.) The ADXRS150 has two matched proof mass resonators/decoupling frame, spring, and capacitor assemblies on the chip. Their outputs have reversed polarity, and the two demodulator outputs are connected differentially in order to reject common-mode artifacts caused by linear acceleration shocks. Specifications of the ADXRS150 MEM angular velocity gyro include $0.05^\circ/\text{s}/\sqrt{\text{Hz}}$ noise, a sensitivity of ca. 12.5 mV/($^\circ/\text{s}$) over $\pm 150^\circ/\text{s}$, nonlinearity = 0.1% of full-scale, frequency response, and DC to 40 Hz (user adjustable). Thus, in a 16 Hz output bandwidth, there would be $0.2^\circ/\text{s}$ RMS equivalent noise. The ADXRS612, $\pm 250^\circ/\text{s}$ MEM yaw rate gyro is offered by Analog Devices.

12.3.4 Quartz Tuning Fork MEM Rate Gyro

When a pair of masses is driven to oscillate with equal amplitude, but in opposite directions (180° out of phase), in the same plane, rotation about the input (Z) axis generates Coriolis moments that act orthogonally at the vibrating mass attachment point (see Figure 12.6). Equations for these Coriolis moments were developed in Section 7.2.2.3. The same principle of vibrating masses has been used by Systron Donner Inertial Division, Concord, CA, to make a family of angular rate gyros using MEM technology. The vibrating masses are the vibrating arms of a dual, quartz MEM TF. For example, Systron Donner Inertial Division's GyroChip™ II, QRS14 MEM angular rate sensor has a $\pm 100^\circ/\text{s}$ full-scale range, a ≤ 50 Hz signal bandwidth, a threshold of $< 0.004^\circ/\text{s}$, and an output noise of $< 0.05^\circ/\text{s}/\sqrt{\text{Hz}}$. The sensor linearity is given at $< 0.05\%$ of full range (QRS 2012). The readout system for the Systron Donner Inertial's QRS14 rate gyro is shown in Figure 12.7, adapted from Figure 4 in QRS (2012). The excitation system has been omitted from Figure 12.7 for simplicity. Note that the Coriolis moments induced in the top quartz TF MEM structure by $\dot{\phi}$ are coupled to and excite oscillations in the bottom (readout) TF. A PSD, followed by an LPF, is used to extract an output (DC) signal, V_o , proportional to the angular input rate, $\dot{\phi}$. Systron Donner Inertial also offers the QRS116 Single-Axis Tactical Grade Analog Gyroscope ($\pm 100^\circ/\text{s}$) and the QRS28 Dual-Axis Gyroscope.

12.3.5 Applications of MEM Gyros

MEM gyros have a wide variety of applications in both military and civilian areas. For example, they are used in automotive ride stabilization and rollover detection and also in

**FIGURE 12.6**

(A) (Top view): A single vibrating mass rate gyro. The mass oscillates in the XZ plane. (B) Side view of vibrating mass rate gyro. Platform rotational velocity, $\dot{\phi}$, around the z -axis causes torque to be developed around the mass pivot. See text for description.

video camera image stabilization (especially in photography from helicopters and boats). They are also used for roll, pitch, and yaw stabilization in autonomous flying vehicles (drones), in input devices in virtual reality video games, and for stabilizing two-wheeled vehicles and *walking* robots. They have been used to study human balance dynamics and implement postural control prosthesis systems (Wall and Weinberg 2003). In civil engineering, MEM gyros can be used to measure bridge torsional sway due to wind loading and traffic and also angular sway at the top of skyscrapers due to high wind and earthquakes.

The precision of MEM gyros is generally not good enough for long-range inertial navigation (e.g., in submarines or space vehicles). Fiber-optic (Sagnac) and/or traditional spinning rotor gyros are generally used in these applications.

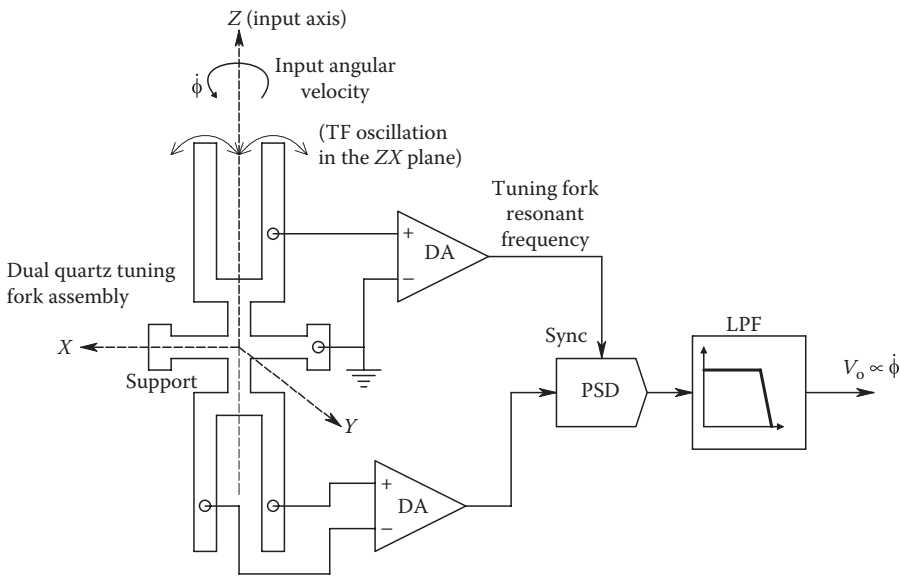


FIGURE 12.7

Diagram of Systron Donner's QRS14 quartz, dual TF, rate gyro. See text for description.

One advantage of MEM gyros is that their packages are small; for example, the ADXRS150 gyro measures $7 \times 7 \times 3.65$ mm, and the SSSJ MEM ring gyro measures $4.8 \times 5.7 \times 1.2$ mm. Their small size makes them ideal for applications in drone aircraft and helicopters.

12.4 Cantilever-Like MEMS and NEMS and Their Applications

12.4.1 Introduction

The development and application of cantilever MEMS is an active and growing research area. One factor inspiring this R&D effort is the ability to mass-produce this class of sensor at relatively low cost. Another factor is the extreme sensitivity to the QUM (the analyte). Applications, described in more detail in the succeeding text, include acting as a mass spectrometer, the detection of specific bacteria or cells, quantization of certain biomolecules, sensing vapors from explosives and pollutants, and in drug discovery. The development of micro- and nanocantilever sensing began in the mid-1990s. An excellent, broad, technical summary of cantilever-like MEM sensors and their applications may be found in Boisen et al. (2011).

In general, sensing relies on molecules binding to a monolayer deposited on the cantilever, increasing its mass. This binding and resulting bound mass increase can be sensed by (1) measuring the bound mass-induced static strain in the cantilever and, more sensitively, (2) by measuring the shift in the self-resonant frequency of the cantilever carrying the bound mass.

12.4.2 Applications of Cantilever-Based MEMS Sensors

Resonant MEM and NEM cantilever sensors are primarily used to measure chemical and biochemical analytes in air and water at ultralow concentrations. Most MEMS use an

ultrathin ligand coating that binds to a specific analyte. When the cantilever is given a specific AB coating, it will bind to a specific biomolecule, such as a prostate-specific antigen (PSA), used to indicate the possible presence of prostate cancer (Wu et al. 2001). The bound mass of the analyte will cause a shift in the cantilever's resonant frequency, or if driven at its unbound resonant frequency, it will exhibit a reduced amplitude of oscillation due to mass attachment. Whole bacteria or bacterial fragments can also be ligated by specific ABs bound to the cantilever for pollution assays and to sense pathogenic *Escherichia coli* and *Salmonella enterica* in lesions (Ilic et al. 2001, Weeks et al. 2003). Nucleic acids have also been successfully quantified. For example, messenger RNA (mRNA) biomarker candidates in a pool of cellular RNA have been sensed at the picomolar level *without target amplification* (Zhang et al. 2006). DNA oligos have been quantified by binding to complimentary *target* DNA on the cantilever. The authors claim that the method can detect target DNA at a concentration of 0.05 nM or lower (Su et al. 2003). Braun et al. (2009) used resonant cantilever sensors to study the interaction between T5 bacteriophage and specific membrane proteins bonded to the beam. A clear shift of resonance frequency was observed for a phage concentration of 3 pM. The binding properties of the vancomycin-type antibiotics were investigated for vancomycin-sensitive and vancomycin-resistant peptides associated with MRSA and vancomycin-resistant *Enterococcus* sp. (Ndieyira et al. 2008).

Chemical vapors from explosives have also been sensed using vibrating MEM cantilevers. Yang et al. (2010) reported on a MEM cantilever system that can sense TNT vapor at the 100 ppt level. Pinnaduwege et al. (2003) reported that vapors from the explosives TETN and RDX could be detected in a 10–30 ppt range, using a gold-coated silicon cantilever functionalized with a bound, self-assembled monolayer (SAM) of *4-mercaptobenzoic acid*. A SAM of *6-mercaptopnicotinic acid* on a silicon microcantilever has also been found effective in sensing explosive vapors (Zuo et al. 2001). (See the review papers on SAMs by Prashar [2012] and Schreiber [2000].)

NEMS mass spectrometry has also been investigated using an oscillating microbridge (a microcantilever fixed at both ends) (Naik et al. 2009). Gold nanoparticles of known mass, 2.5 nm in diameter, were used to examine the resonant frequency shifts as a function of the distribution and density of the particles on the bridge. Boisen et al. (2011) stated: "By comparing the distribution of measured frequency changes with the distribution expected for particles from a monodisperse source, it was possible to derive an average mass for the gold nanoparticles." Next, Naik et al. used a solution of bovine albumen protein. From 578 binding events (resonant frequency jumps), it was possible to generate a mass histogram showing that the albumen sample consisted of monomers as expected, but also dimers, trimers, and pentamers, and to calculate the ratios between them. The work of Naik et al. demonstrated that it is possible to register the binding events of individual molecules (Boisen et al. 2011).

A summary of cantilever MEMS and NEMS applications should include the following (Vashist 2007):

- Humidity sensors
- Herbicide sensors
- Metal-ion sensors (in solution)
- Temperature sensors
- Viscosity sensors
- Calorimetry sensors

- Magnetic bead detection sensors
- IR radiation detection sensors
- Explosive detection sensors
- VOC detection
- Cancer detection microchips
- Myoglobin detection sensors
- Biosensors for coronary heart disease
- Sensors to detect single nucleotide polymorphisms (SNPs)
- Sensors to detect specific ABs or antigens
- Sensors to detect specific bacteria (e.g., *E. coli*)

12.4.3 Design of Cantilever MEMS and NEMS

12.4.3.1 Cantilever Materials and Fabrication

Silicon-based MEM microcantilevers are typically ca. 1 μm thick and 450–950 μm long. Widths (w) and thicknesses (t) are varied. Ultrathin cantilevers with $t < 200$ nm have been constructed (Ramos et al. 2009). Silicon, silicon nitride, and silicon oxide have been used (Vashist 2007). MEMS cantilevers are generally more than just pure silicon. They can have two or more, thin, overlying layers deposited on them, for example, a 3 nm layer of Cr, on top of which is a 40 nm layer of Au, in turn covered with a thin layer of selective organic ligand molecules to bind to a desired analyte (e.g., a specific AB). A piezoresistive material such as doped Si to sense cantilever strain can also be deposited on the bottom of the beam. In most resonance sensitivity analysis, the thin beam coatings are generally ignored compared to the 1000 nm Si beam thickness.

Polymer-based cantilevers also have utility. Not only are they micromachineable, but their Young's moduli are ca. two orders of magnitude lower than Si-based materials. Thus, the stiffness of polymer-based cantilevers is reduced, and their mass sensitivity is larger. Polymer cantilevers must have thicknesses (t) comparable to those of silicon to realize this increased mass sensitivity. Young's modulus Y of pure silicon ranges from 130 to 188 GPa, depending on the crystal structure orientation relative to the applied stress (Si is anisotropic). Y for silicon nitride (Si_3N_4) is 180–290 GPa ($1.8\text{--}2.9 \times 10^{10}$ dynes/cm²) compared to ca. 4.4 GPa (4.4×10^8 dynes/cm²) for SU-8 epoxy photoresist. Other polymers such as *polyimide*, *polystyrene*, *polypropylene*, *polyethylene terephthalate*, *fluoropolymers*, *parylene*, and *TOPAS*[®] have been used. Details of silicon and polymer cantilever micromachining techniques are beyond the scope of this chapter and are given in the extensive review paper by Boisen et al. (2011).

12.4.3.2 Cantilever Readout Methods

There are two major means for sensing MEM cantilever analyte mass loading. One is to sense the static cantilever deflection in the gravitational field due to the mass of the analyte bound to the cantilever. The other is to mechanically vibrate the cantilever at its resonant frequency, f_r , and observe the resonant frequency shift, Δf_r , due to mass loading by the bound analyte or the decrease in vibration amplitude when the beam is driven at its unloaded resonant frequency, f_r .

Static cantilever deflection, proportional to bound analyte mass, and frequency shift in cantilever resonance can be quantified by several means. The major ones are

- Optical deflection of a collimated light beam directed to a mirror on the free end of the cantilever. The reflected beam is directed to an integrated photodiode. Optical system geometry is such that V_o from the diode is proportional to δx , the cantilever deflection. Optical deflection can be used on static (DC) and vibrating (AC) cantilevers.
- Another means that is extremely sensitive is to measure the differential capacitance change between two fixed plates and the cantilever. This means is useful for nanometer-sized cantilevers as well as micron-sized devices. It works for static (DC) and vibrating (AC) cantilever deflections.
- Piezoresistive sensing of cantilever deflection requires that a piezoresistive material (basically a strain gauge), generally a metal such as gold, is bound to the surface of the cantilever. When the cantilever deflects, a strain is induced in the piezoresistor that incrementally increases its resistance. This ΔR is generally sensed with a Wheatstone bridge circuit. Piezoresistance sensing of cantilever beam deflection is used for both static (DC) and vibrating cantilever (AC) applications.
- Piezoelectric sensing of bound mass is generally done using a piezoelectric cantilever material. The cantilever is a piezoelectric material such as aluminum nitride (Cleland et al. 2001) or is coated with one. The piezoelectric cantilever is caused to vibrate at its resonant frequency, f_r , and the Δf_r due to bound analyte mass is sensed. f_r s greater than 80 MHz have been reported. Recently, it has been reported that the aluminum nitride piezoelectric cantilevers have been sandwiched between two electrodes, enabling the cantilever to be driven electrically at its f_r . By adding a DC bias voltage to the drive voltage, it has been possible to tune the f_r with a sensitivity of 34 kHz/V, allowing the Δf_r caused by added analyte mass to be nulled to f_r ($m = 0$). The ΔV (DC) is thus a function of m (Boisen et al. 2011). Piezoelectric cantilevers cannot work at DC to measure static deflection (cf. Section 6.4.3).

12.4.3.3 Static MEM Cantilever Deflection Sensing with Piezoresistors

An example of a MEM cantilever sensor that uses piezoresistive strain gauges to sense ultrasmall masses bound to the cantilever is shown in Figure 12.8B. One (active) cantilever is coated with a thin layer of a ligand that binds a fluid-phase molecular analyte (gray spheres). The analyte mass bound to the ligand causes the active cantilever to deflect some δx , straining the piezoresistive traces on the top of the cantilever. This causes an incremental increase in the resistance of the piezoresistor. Also, an increase in temperature of both cantilevers above a reference temperature will increase their resistances. This sensitivity to bound mass and temperature can be written as follows: $R_M = R_o(1 + \alpha\Delta T) + km$, where R_o is the piezoresistance sensing element's unstrained resistance at T_o , k is a sensitivity factor, α is the piezoresistor's temperature coefficient of resistance, m is the bound mass in kg, and $\Delta T \equiv (T - T_o) \geq 0$. The reference cantilever cannot bind to the analyte; hence, its resistance is given by $R_R = R_o(1 + \alpha\Delta T)$. Figure 12.8A illustrates the circuit of a simple Wheatstone bridge used to sense a $V_o \propto m$. The resistors R_o are held at temperature T_o . The bridge excitation voltage V_s is audio-frequency AC, to minimize $1/f$ noise associated with V_o amplification. V_o is amplified differentially to minimize common-mode noise from

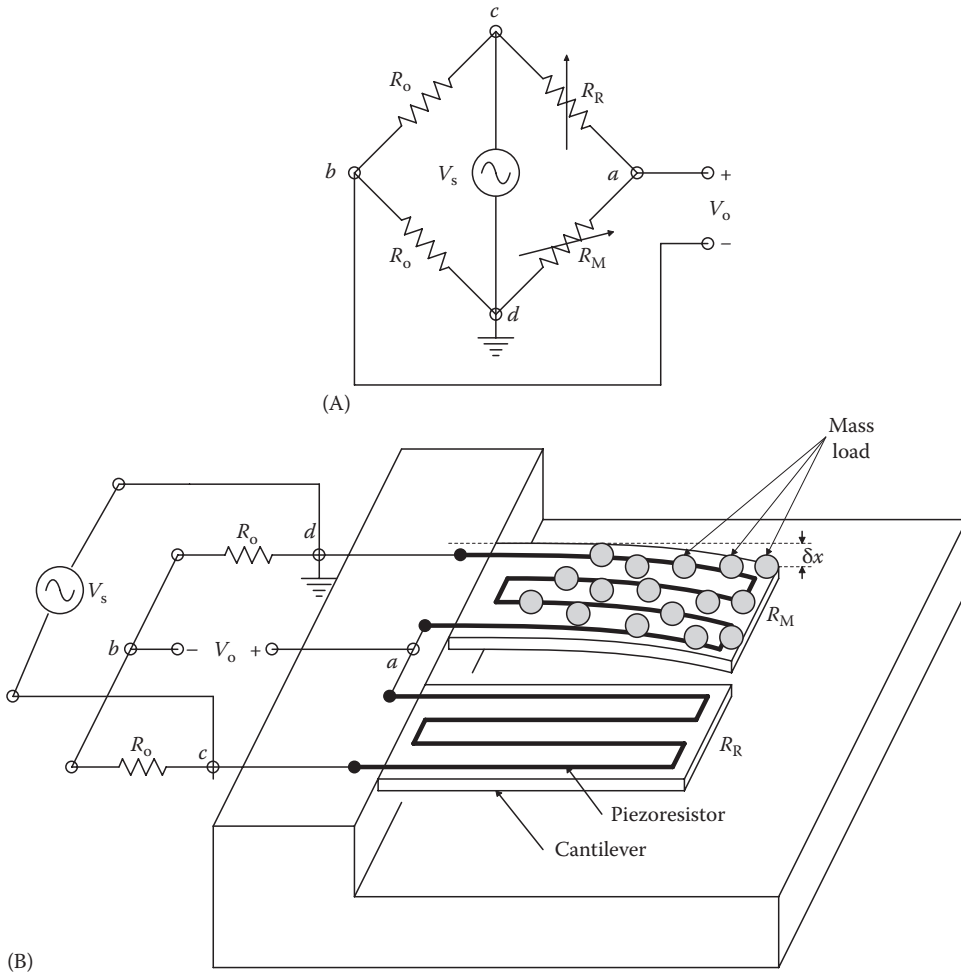


FIGURE 12.8 (A) Wheatstone bridge circuit of the MEM, dual-cantilever bound mass sensor. (B) Diagram of the dual-cantilever chemisensor. See text for description.

vibration and external AC field pickup in the circuit. Using the bridge circuit, it is easy to show that V_o is given by

$$V_o \cong \frac{km}{4R_o} \left[1 - \left(\frac{\alpha\Delta T + km}{2R_o} \right) \right] \cong \frac{km}{4R_o} V_s \quad (12.17)$$

It is clear that from the circuit geometry, the first-order temperature dependence term of the piezoresistors, $R_o\Delta\delta T$, cancels out in the numerator, and the second-order (denominator) terms are negligible. Boisen et al. (2011) reported that when gold is used for the piezoresistor, a mass resolution of as small as 1 attogram was reached ($1 \text{ ag} = 1 \times 10^{-18} \text{ g}$). Wheatstone bridge resolution is about $1 \mu\Omega$ (cf. Section 3.8.4). So k in Ω/kg of analyte in Equation 12.17 needs to be around 1015.

Various polymers that are micromachineable, such as polystyrene, parylene, and TOPAS®, have been used instead of silicon compounds for MEM cantilevers. Polymer cantilever MEMs are less expensive to fabricate and offer greater sensitivity because they have lower Young's moduli; hence, k in the km term can be larger.

12.4.3.4 Cantilever Beam Resonance Frequency

An end-clamped cantilever beam can be shown to have a number of bending (oscillatory) modes; in each mode, the free end of the beam has the maximum deflection, Δx . In the *first mode*, there is simple bending of the beam from its clamped end. In the *second mode*, there is one zero deflection node at ca. 0.68 of the beam length, L . The third deflection mode exhibits two zero deflection nodes, and the fourth mode, three zero deflection nodes, etc. The details of beam bending and oscillation modes are well covered in the review paper by Boisen et al. (2011), and also in Ch. 13 of Cannon (1967). Raiteri et al. (2001) and Arecco (2003) gave the fundamental oscillation (resonance) frequency of an end-clamped, rectangular cantilever as

$$f_{r1} \cong \frac{0.162}{L^2} \sqrt{\frac{Y}{\rho}} \text{ Hz} \quad (12.18)$$

where

L is the beam length in cm

ρ is the beam mass density in g/cm³ (2.33 for Si)

Y is Young's modulus of beam material in dynes/cm²

Note that the constant 0.1639 can vary among authors due to model assumptions; Finot et al. (2008) used 0.1624, Fukuma et al. (2012) used 0.162, and Teva et al. (2006) and Olson (1943) used 0.1615.

A simpler expression for f_{r1} was given by Vashist (2007):

$$f_{r1} = \frac{0.32\sqrt{k}}{\sqrt{m}} \text{ Hz} \quad (12.19)$$

where

k is the cantilever's Hookean spring constant ($k = \delta x/F$)

m is the cantilever's mass ($m = \rho tLw$ g)

ρ is the beam density in g/cm³

t is the beam thickness in cm (in x -direction)

L is the beam length in cm

w is the beam width in cm

Olson (1940) gave the following CGS f_{r1} expression (Equation 12.20) for the fundamental resonance frequency of an end-clamped rectangular cantilever beam:

$$f_{r1} = \frac{0.5596}{L^2} \sqrt{\frac{YK^2}{\rho}} \text{ Hz} \quad (12.20)$$

where

Y is Young's modulus of the beam in dynes/cm²

K is the radius of gyration of the beam in cm ($K = t/\sqrt{12}$ for a rectangular beam)

L is the beam's length in cm

ρ is the density of the beam in g/cm³

Bound analyte mass generally lowers f_{r1} as it increases the effective ρ .

12.4.3.5 Excitation of Mechanical Oscillation of MEM Cantilevers

The mechanical (end deflection) frequency response of a cantilever, $\Delta z/F_z$, is constant from $f = 0$ to just below f_r , then it rises to the resonant peak at $f = f_r$, then falls off rapidly to 0 as $f \gg f_r$. It is the shift in this resonant peak that is used to quantify bound analyte.

An ingenious means of mechanically exciting (vibrating) MEM cantilevers using the magnetic Lorenz force was described by Li et al. (2009). Figure 12.9 illustrates this method. The cantilever strain is read out with a bonded piezoresistive element at the base of the beam. The beam has a thin aluminum conductor bonded to its outer edges. The conductor carries an AC electric current at the beam's unloaded mechanical resonant frequency, f_r . A small, strong, permanent, NdFeB magnet is located so its \mathbf{B} field points along the long axis of the cantilever, perpendicular to the current-carrying conductor across the free end of the cantilever. A force \mathbf{F} is generated from the interaction of the current with the perpendicular \mathbf{B} field. The differential force vector can be shown to be given by the vector cross product: $d\mathbf{F} = I(d\mathbf{i} \times \mathbf{B})$, where I is the current and $d\mathbf{i}$ is a unit vector in the direction of the current. From the circuit geometry, the total force acting perpendicular to the end of the cantilever is $F_z = i(t)wB$ N (see *vector cross product* in the Glossary), where $i(t)$ is the AC current in the aluminum wire, w is the width of the end of the cantilever, and B is the magnitude of the permanent magnet's B field (50 mT in the paper by Li et al.), assumed to be constant over w . F_z is directed upward for the current and \mathbf{B} field directions shown. Since $i(t) = I_0 \sin(2\pi f_r t)$, the force will also be AC and will excite the beam to vibrate in the vertical (z) direction at the cantilever's mechanical resonance frequency, f_r ; thus, the Δz excursion at the end of the beam will be maximum. This maximum Δz at f_r can be sensed with the piezoresistor bonded to the base of the beam. As the beam is mass loaded by the analyte, the mechanical resonant frequency will shift downward by some small Δf (i.e., a new $f'_r = f_r - \Delta f$ will exist). Δf is proportional to the total bound mass, m .

Mechanical oscillation of a microcantilever at its resonant frequency can also be accomplished using electrostatic attraction between the beam and fixed capacitor plates. From basic

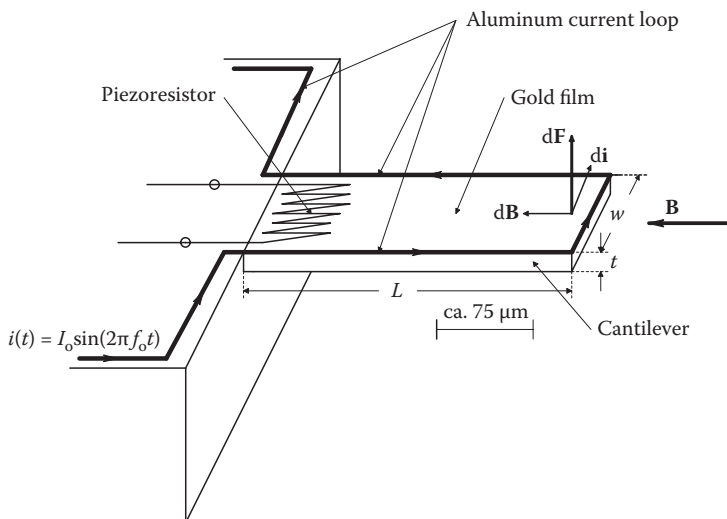


FIGURE 12.9

Diagram of a vibrating cantilever MEM chemisensor. A magnetic force motor drives the cantilever at its mechanical resonance frequency. As analyte mass binds to the cantilever surface, the resonance frequency is lowered. This Δf is sensed. The cantilever must be purged of bound analyte before it can be reused.

physics, the attractive force between two oppositely charged flat capacitor plates of area A m², in vacuo, can be written as follows: $F = Ae_0E^2/2$ N, where E is the electric field between the plates in V/m and ϵ_0 is the permittivity of vacuum in (F/m) (Sears 1953). Note that the electric field is alternately switched between the top surface of the cantilever and upper capacitor plate and between the bottom surface of the cantilever and lower capacitor plate. If the spacing between the upper plate and the beam is d , then the zero deflection electric field is $E = V_0/d$ V/m, where V_0 is the voltage between the upper beam surface and the upper capacitor plate. Thus, an upward electrostatic force of $F = Ae_0(V_0/d)^2$ N acts uniformly on the cantilever. When the beam is put between two capacitor plates and they are alternately switched on and off at the beam's resonant frequency (the top plate is at V_0 , lower plate at 0 V for $1/(2f_r)$ s, then the lower plate is at V_0 , while the upper plate is at 0 V for $1/2f_r$ s, etc.), the resulting alternating, up/down electrostatic forces can be made to drive the cantilever at its mechanical resonance frequency, ω_r . This same alternating capacitor force can be used to drive an oscillating disk gyro about its pivot axis (see Section 12.3.2) or an oscillating gyro proof mass (cf. Section 12.3.3).

12.4.3.6 Vibrations in a Beam Clamped at Both Ends

A change in the resonance frequency in a doubly clamped beam is also a sensitive means of detecting analyte molecules bound to it. Olson (1943) gave the fundamental resonance frequency of a doubly clamped beam as

$$f_{r1} = \frac{1.028t}{L^2} \sqrt{\frac{Y}{\rho}} \text{ Hz} \quad (12.21)$$

where

L is the beam length in cm

t is the beam thickness in cm

Y is the beam material's Young's modulus in dynes/cm²

ρ is the beam material's density in gm/cm³

Nanoscale beams have been made of monocrystalline silicon carbide (SiC) and have been given resonance frequencies ranging from VHF and UHF to low GHz frequencies (Huang et al. 2005). Measured mechanical resonance Q generally decreases with increasing f_r .

12.4.3.7 Vibrations in a Circularly Clamped Thin Plate

Olson (1943) gave the fundamental frequency of oscillation of a thin circular plate (diaphragm) under no radial tension as

$$f_{r1} = \frac{0.467t}{R^2} \sqrt{\frac{Y}{\rho(1-\sigma^2)}} \text{ Hz} \quad (12.22)$$

where

R is the plate radius in cm

ρ is the plate mass density in g/cm³

Y is Young's modulus in dynes/cm²

t is the plate's thickness in cm

σ is the material's Poisson's ratio (see Glossary)

CGS units were used.

Again, the binding of an analyte to the plate's surface will change f_{r1} . In a MEMS sensor using a vibrating plate (diaphragm), this Δf_{r1} can be sensed optically, electrostatically, or by piezoresistors bonded to the diaphragm's surface. Vibrations in the plate can be excited electrostatically or by passing alternating current through a conducting coil bound to the plate's surface subject to a permanent magnet's constant (DC) B field.

12.4.4 Quartz Tuning Fork Biosensor

We have seen how vibrating masses, disks, and piezoelectric crystals and cantilevers can be used for sensing bound mass from a variety of analytes by measurement of the resonant frequency (f_r) shift. Adding to this list of MEMS is a prototype quartz TF biosensor, developed and tested by Su et al. (2002). A TF is basically two cantilevers bound to the same base. This geometry acts to damp out overtones (harmonics) and produces a nearly pure fundamental frequency at f_r (the first harmonic) (Olson 1943). Two means of exciting TF mechanical oscillations were described by Su et al.: (1) *In the open-loop mode*, excitation of the quartz TF biosensor is accomplished by fastening the TF to a piezoelectric plate and exciting plate oscillations at a desired frequency; these are mechanically coupled to the TF, causing it to vibrate. The amplitudes and frequencies of the TF oscillations are sensed with electrodes attached to the quartz TF; and (2) *In the closed-loop, self-excitation mode*, the quartz TF's vibrations are sensed and input to a tuned circuit whose center frequency is the f_r of the TF. The output of the tuned circuit is amplified and fed to the excitation piezoplate.

Su et al. observed that the reproducibility in the resonant frequency shift due to mass loading by bound analyte was ca. 2 ppm, yielding a detection mass limit for human IgG AB protein of ca. 5 $\mu\text{g}/\text{mL}$. The authors observed that the reproducibility could be improved by using higher-frequency TFs. They also described other technical problems with their TF microsensor.

12.4.5 Liquid Viscosity Measurement Using Piezoceramic Resonators

The resonant frequency of a piezoelectric resonator depends largely on its mechanical properties such as shape, thickness, and Young's modulus. Any external medium that adds to the resonator's mass or increases the energy coupled from the resonator to the medium will generally lower the resonant frequency and the Q of the resonator.

The electrical equivalent circuit of a piezoceramic resonator disk has been generally described by a capacitor, C_0 , in parallel with a series RLC circuit (R_1 , C_1 , L_1) describing the resonator's electrical behavior around its electromechanical resonance frequency at its terminals. For example, an Abracon Corp. Application Note gave numerical values for these model parameters for a 4 MHz disk resonator: $C_0 = 36.3$ pF, $C_1 = 4.4$ pF, $R_1 = 8 \Omega$, $L_1 = 385 \mu\text{H}$. Saluja and Kalonia (2004, 2005) described a fluid viscosity measurement system that used sample volumes from 8 to 10 μL . They used quartz disk resonators with 5 and 10 MHz fundamental (unloaded) frequency and having 1.37 cm diameters. Their system was characterized by the parallel C_0 -RLC circuit model, described earlier. When loaded with the viscous fluid, the resonator's resonant frequency dropped and its Q was reduced. This was accounted for by adding a series R_2 and L_2 to the series

R_1, C_1, L_1 branch in the unloaded resonator model. In general, the resonator's complex admittance magnitude was shown to be

$$|Y(j\omega)| = \left\{ \left[\frac{\text{Real}}{R^2 + (\omega L - 1/\omega C)^2} \right]^2 + \left[\omega C_0 - \frac{\text{Imaginary}}{R^2 + (\omega L - 1/\omega C)^2} \right]^2 \right\}^{1/2} \quad (12.23)$$

where

$$R = R_1 \text{ (unloaded) or } R_1 + R_2 \text{ (loaded)}$$

$$L = L_1 \text{ (unloaded) or } L_1 + L_2 \text{ (loaded)}$$

$$C = C_1 \text{ (unloaded)}$$

The loaded C was assumed to be C_1 (unchanged), as was C_0 .

Note that the resonant frequency of the resonator, ω_r , can be solved for by setting the imaginary term in Equation 12.23 equal to zero. Measurement of the test solution's viscosity could be characterized by the frequency shift, $\Delta\omega_r = (\omega_{rU} - \omega_{rL})$, where ω_{rU} is the unloaded resonant frequency and ω_{rL} is the resonant frequency with the QUM on the resonator disk. Note that the resonance frequency is defined here as the frequency ω_r , where $Y(j\omega_r) = \text{real}$.

Saluja and Kalonia (2004) used a Hewlett Packard/Agilent HP4194A impedance/gain-phase analyzer to characterize the RLC parameters of the unloaded and loaded resonators. By examining the frequency shift and the R_2 and L_2 values, they were able to obtain consistent measurement of their test solutions' $\sqrt{\rho\eta}$ in $\text{gm/cm}^2/\text{s}^{1/2}$, where η is the viscosity in centipoise and ρ the solution's density in gm/cm^3 . The instrument developed by Saluja and Kalonia (2005) was shown to have application in the rheological characterization of proteins. An advantage of the resonant viscosity measurement system of Saluja and Kalonia was that the QUM could easily be washed off the resonator, then the resonator dried making it ready for another measurement after recalibration of ω_{rU} .

Another approach to fluid viscosity measurement was described by Papi et al. (2006). In this study, they developed an expression for the viscosity (η) of a test liquid using the resonant frequency of a cantilever in vacuo (ω_0), its resonant frequency in the test fluid (ω_s), and also the cantilever's dimensions (L, W, T), its density (ρ_0), and the density of the test liquid (ρ_s). They showed that the measured sample viscosity (η_s) could be modeled by the following equation:

$$\eta_s = \omega_s \alpha \left[\frac{\omega_0^2}{\omega_s^2} - (1 + \beta) \right]^2 \quad (12.24)$$

where

$$\alpha \equiv 8\rho_0^2 T / \pi^2 \rho_s a_2^2$$

$$\beta \equiv \pi \rho_s W / 4\rho_0 T a_1$$

$$a_1 \equiv 1.0553$$

$$a_2 \equiv 3.7997$$

The authors observed that their results were strictly valid for those samples with viscosity larger than pure water used as a reference. When water was used as a reference liquid, they derived the model for the sample viscosity:

$$\eta_s = \frac{Y_s \omega_s \eta_w}{Y_w \omega_w} \quad (12.25)$$

where

ω_w is the cantilever resonant frequency in pure water

η_w is the viscosity of water

$Y_s \equiv \left[\omega_0^2 / \omega_s^2 - (1 + \beta) \right]^2$ for the sample

$Y_w \equiv \left[\omega_0^2 / \omega_w^2 - (1 + \beta) \right]^2$ for water

Note that the resonance frequency of the cantilever in vacuo still appears in Y_s and Y_w .

12.5 Chapter Summary

The rapidly growing field of MEM sensors has been introduced in this chapter. MEM linear accelerometers, their design, properties, and applications have been described, as well as MEM pressure sensors. Next, we considered some designs for MEM rate gyros based on periodically moving masses, namely, the oscillating disk gyro, vibrating mass MEM gyros, and the quartz *TF* rate gyro.

Many sensor designs have been based on cantilever-like MEMs. In some, the QUM binds to the microcantilever beam and its mass is detected by beam deflection. In other cantilever MEMs, the mass of the bound QUM causes a change in the mechanical resonant frequency of the cantilever, which is sensed. Similarly, QUM bound mass also can change the oscillation frequency of a quartz TF or resonator plate.

Problems

- 12.1** Use Equation 12.21 to *calculate* the mechanical fundamental resonant frequency in Hz of a silicon beam *clamped at both ends*, having dimensions $L = 40 \mu\text{m}$, $W = 2.3 \mu\text{m}$, and $t = 5 \mu\text{m}$. Given for silicon: $Y = 120 \text{ GPa}$, $\rho = 2.33 \text{ g/cm}^3 = 2.33 \times 10^3 \text{ kg/m}^3$. Use SI units.
- 12.2** Use Olson's CGS formula (Equation 12.20) to calculate the fundamental resonance frequency of a silicon carbide NEM, end-supported cantilever having the following specifications: $L = 1110 \text{ nm}$, $w = 120 \text{ nm}$, $t = 75 \text{ nm}$, device mass $m = 44.2 \text{ fg}$, $Y = 450 \text{ GPa}$, and $\rho = 2.33 \times 10^3 \text{ kg/m}^3$.

12.3 Cheung (2010) gave a formula for the static, small center deflection of a thin, edge-clamped circular diaphragm:

$$\delta_{\text{center}} = \frac{3r^4P(1-\nu^2)}{16Yt^3} \text{ m}$$

where

- P is the applied differential pressure in Pa
- r is the diaphragm radius in m
- ν is Poisson's ratio
- Y is Young's modulus in Pa
- t is the diaphragm thickness in m

Note: 1 Pa = 1 N/m². Let $Y = 185 \text{ MPa}$, $r = 1.0 \text{ }\mu\text{m}$, $\nu = 0.23$, and $t = 20 \text{ nm}$.

A. Find the deflection sensitivity, δ_{center} in $\mu\text{m}/\text{P}$ in kPa.

B. How would you measure δ_{center} ?

12.4 Consider the simple, 1D mass-spring-viscous friction mechanical system of Figure P12.4. It is caused to vibrate sinusoidally by an applied sinusoidal force, $f(t) = F_0 \sin(\omega t)$. The system can be described by the second-order ODE: $M\ddot{x} + B\dot{x} + Kx = f(t)$, where M is the mass of the vibrating system in kg, B is the damping constant in N/(m/s), and K is the Hookean spring constant in N/m. The force $f(t)$ is in newtons.

- A. Find expressions for the mechanical system's undamped resonant frequency, ω_n , and its Q in terms of M , B , and K .
- B. Assuming a steady-state sinusoidal force, find the complex value of \mathbf{X}/\mathbf{F} for $\omega = \omega_n$ and for $\omega = 0$ (a DC force).
- C. Derive an expression for the frequency ω_p , where $|\mathbf{X}/\mathbf{F}(j\omega)|$ is maximum. *Hint:* It is generally lower than ω_n .
- D. Make a dimensioned Bode plot (magnitude and phase) for $\mathbf{X}/\mathbf{F}(j\omega)$.
- E. Now, let the mass be increased by a small increment, δ_m : show how this affects the system's undamped natural frequency, ω_n . Derive an expression for the change in ω_n , $\Delta\omega = \omega_n - \omega'_n$, caused by the mass increment.

12.5 The frequency response function of a second-order, low-pass, underdamped, mechanical vibrating system is given by the following general expression:

$$\frac{\mathbf{Y}}{\mathbf{F}}(j\omega) = \frac{K}{\left[(j\omega)^2/(\omega_n^2) + (j\omega)/(\omega_n Q) + 1 \right]} = \mathbf{T}(j\omega)$$

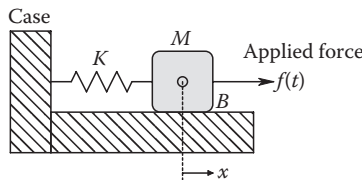


FIGURE P12.4

where Q is the system's *quality factor*, K is a constant, Y is the amplitude of oscillation, F is the sinusoidal driving force, and ω_n is the system's undamped natural frequency.

- A. Give an expression for $|T(j\omega_n)|$.
- B. Find an expression for the value of $\omega = \omega_p$ that will maximize $|T(j\omega)|$.
- C. Find an expression for the peak value of $|T(j\omega_p)|$.
- D. Note that in a second-order system, $Q \equiv 1/2\xi$. ξ is the damping factor. What is the largest ξ such that $|T(j\omega)|$ has no peak for any $\omega > 0$.

12.6 A linear MEM accelerometer consists of a proof mass M kg attached to a case by a spring with spring constant K N/m. Viscous damping B N/(m/s) exists between the mass and the case. When the case is accelerated in the $x_c > 0$ direction, the mass moves relative to the housing in the direction x_m shown in Figure P12.6. x_m is the mass displacement in *inertial space*. x_c is the case displacement in *inertial space*.

- A. Write the Newtonian force balance ODE for the mass, given case acceleration \ddot{x}_c . Find the transfer function:

$$\frac{(X_e - X_m)}{\ddot{X}_c}(s).$$

- B. Given $M = 0.1 \mu\text{g} = 10^{-10}$ kg. Find the spring K required so that a constant case acceleration of 0.1 g will produce a relative mass displacement of $(x_c - x_m) = 50$ pm.
- C. Find the accelerometer's natural frequency $f_n = \omega_n/2\pi$.
- D. What value of B in N/(m/s) is required to give the accelerometer critical damping, $\xi = 0.7071$.

12.7 Figure P12.7 illustrates the top and side view of a MEM vibrating mass angular rate gyro. The mass is forced to vibrate in a slot micromachined in the disk with amplitude $r(t) = r_o \sin(\omega_o t)$. When the disk (and mass) is rotated at angular velocity ϕ , the mass generates a Coriolis force $F(t)$ on the side of the slot with a moment arm, $r(t)$, generating an instantaneous torque $T(t)$ on the disk holding the mass. The disk is supported by a torque post allowing no significant motion of the disk. The torque post has a piezoresistor strain gauge to sense the oscillating torque generated by the Coriolis force from the vibrating mass.

- A. Use the Lagrange equations to find expressions for $L = T - U$, $T = \text{mass KE}$, $U = \text{mass PE}$.
- B. Give expressions for $\partial L/\partial\phi$, $\partial L/\partial\dot{\phi}$, and $d/dt(\partial L/\partial\dot{\phi})$. Assume $\ddot{\phi} = 0$.
- C. Find an expression for the instantaneous torque on the torque post, $T(t, \phi)$.
- D. What signal processing would you use to extract $V_o = K\dot{\phi}$?

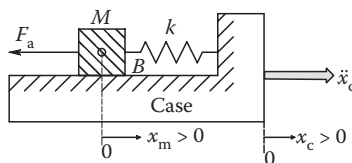


FIGURE P12.6

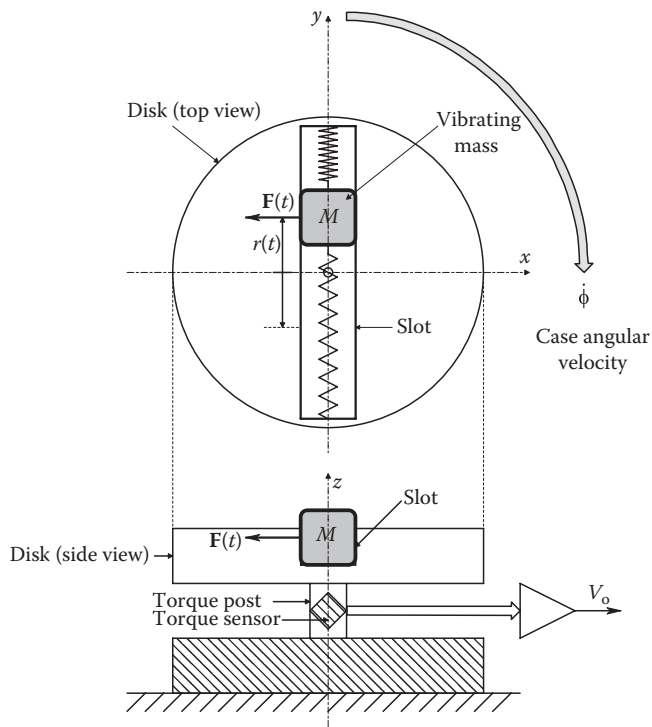


FIGURE P12.7

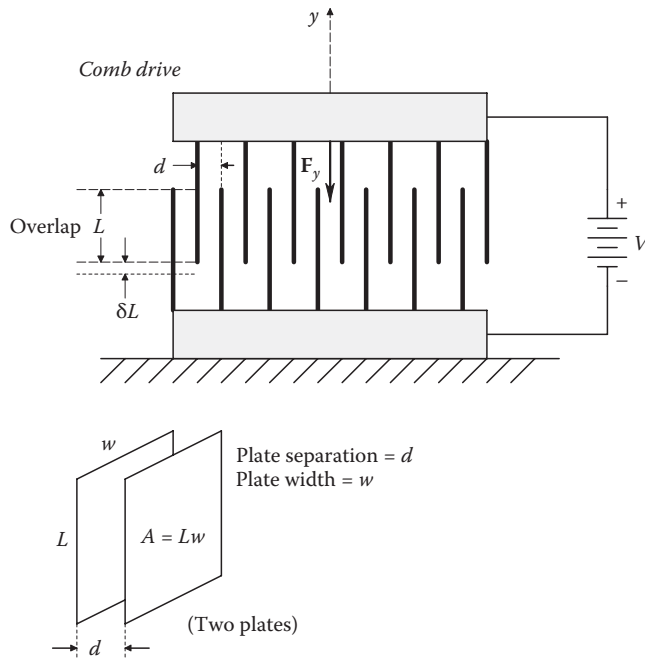


FIGURE P12.8

- 12.8** The schematic of a MEM capacitor *comb drive* is shown in Figure P12.8. It is used to generate a constant force, F_y : The seven bottom plates are fixed, and the upper seven plates can move together.
- Assume N interleaving plates on the top and bottom ($N = 7$ in Figure P12.8). Using the simple parallel plate capacitor model, find an expression for the total capacitance of the *comb capacitor*. The plate separation d is fixed, and the overlap L can vary.
 - When a voltage V is applied across the capacitor, energy is stored in the electric field between the plates. This is given by the well-known relation, $W = 1/2 CV^2$ J. Find an expression for the downward force, F_y . (Note: $dW = F_y dL$.)
- 12.9** Consider the Juneau oscillating disk rate gyro shown in Figures 12.3 and 12.4.
- Derive an expression for the disk tilt moment, M_β , due to $\dot{\alpha}$, the CCW input rotation rate around the x -axis.
 - Show how you can sense the dynamic disk tilt angle, β , using the four fixed capacitor plates shown in Figure 12.4. The disk is at ground potential, and the plates can be connected in pairs.

13

Examples of the Design of Measurement Systems

13.1 Introduction

In this final chapter, we describe the needs for, the systems design philosophies of, and the means of realizing systems used in the measurement of (13.2) *a self-nulling, microdegree polarimeter to measure glucose in bioreactors*; (13.3) *a system to locate and measure the intensity of partial discharges in the insulation of coaxial, HV power cables*; (13.4) *the design of a closed-loop, constant-phase, pulsed laser velocimeter and rangefinder for moving targets*; and (13.5) *the design of capacitive sensors for the detection of hidden objects*. These four systems have been chosen to illustrate certain design principles as well as to describe diverse measurement systems.

The author and his graduate students have designed and developed these unique measurement systems; therefore, their designs reflect their personal bias toward certain circuit architectures and manufacturers' components. It is well appreciated that systems design in electrical engineering is an art as well as a science. The designer must balance criteria on system performance with cost and ease of manufacturing, safety, system reliability, ergonomics, environmental concerns, and even aesthetics. There are often several designs that can lead to the same end product, and no clear set of factors appears to favor one design over the others. It is in these fuzzy cases that a designer is more free to show personal preferences, that is, exhibit his or her *design style*.

In the first design example described later, we describe the need for, the design approach, the specifications, and evaluation of the performance of a self-nulling, microdegree polarimeter.

13.2 Self-Nulling, Microdegree Resolution Polarimeter to Measure Glucose in Bioreactors

A *polarimeter* is an instrument used to measure the angle of rotation of LPL when it is passed through an *optically active material*. (Polarimeters are introduced in Section 6.7.2.) There are many types of polarimeters, and there are a number of optically active substances found in living systems; probably the most important is the sugar, D-glucose-glucose, dissolved in water. Clear biological liquids such as urine, extracellular fluid, saliva, and the aqueous humor of the eyes contain dissolved D-glucose whose molar concentration is an increasing function of the blood glucose concentration (Northrop 2002). D-glucose-glucose

concentration is also an important parameter in the fermentation industry and also in biotechnology applications where genetically engineered cells are grown in bioreactors.

Using polarimetry, D-glucose-glucose concentration can be measured from the simple relation:

$$\phi = [\alpha]_{\lambda}^T CL \text{ degrees,} \quad (13.1)$$

where ϕ is the measured optical rotation of LPL of wavelength λ passed through a sample chamber of length L containing the optically active analyte (D-glucose-glucose) at concentration C and temperature T .

The constant $[\alpha]_{\lambda}^T$ is called the *specific optical rotation* of the analyte; its units are generally in degrees per (optical path length unit \times concentration unit). $[\alpha]_{\lambda}^T$ for D-glucose-glucose in 25°C water and at 512 nm is 0.0695 millidegrees/(cm \times g/dL). Thus, the optical rotation of a 1 g/L solution of D-glucose-glucose in a 10 cm cell is 69.5 millidegrees (m°).

Note that $[\alpha]_{\lambda}^T$ can have either sign, depending on the analyte, and if N optically active substances are present in the sample chamber in solution, the *net optical rotation* is given by superposition, assuming no analyte interactions:

$$\phi = L \sum_{k=1}^N [\alpha]_{\lambda}^{T_k} C_k. \quad (13.2)$$

To describe how a polarimeter works, we must first describe what is meant by LPL. When light from an incoherent source such as a tungsten lamp is treated as an electromagnetic wave phenomenon (rather than photons), we observe that the propagating radiation is composed of a broad spectrum of wavelengths and polarization states. To facilitate the description of LPL, let us examine a monochromatic ray propagating in the z -direction at the speed of light in the medium in which it is traveling. This velocity in the z -direction is $v = c/n$ m/s, where c is the velocity of light in vacuo and n is the *refractive index* of the propagation medium (generally >1.0). A beam of monochromatic light can have a variety of polarization states, including *random* (unpolarized), *circular* (CW or CCW), *elliptical*, and *linear* (Balanis 1989). Figure 13.1 illustrates an LPL ray in which the \mathbf{E} vector propagates entirely in the XZ plane

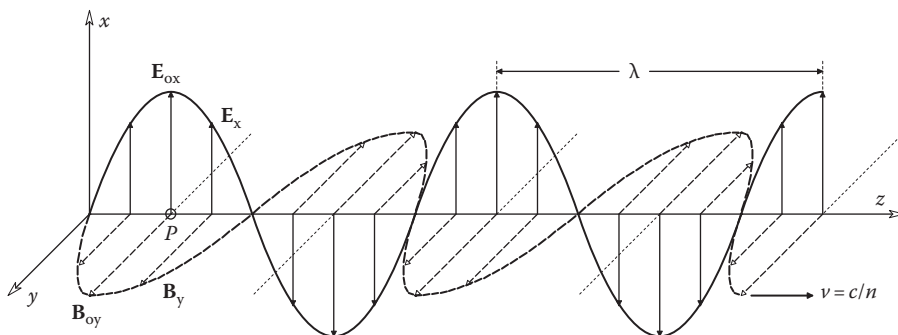


FIGURE 13.1

Diagram of a linearly polarized electromagnetic wave propagating in space. The \mathbf{E} vector propagates entirely in the XZ plane, and the orthogonal \mathbf{B}_y vector lies in the YZ plane; \mathbf{E} and \mathbf{B} are in-phase. Both vectors are functions of time and distance.

and its orthogonal \mathbf{B}_y vector lies in the YZ plane. \mathbf{E} and \mathbf{B} are in-phase. Both vectors are functions of time and distance. For the $\mathbf{E}_x(t, z)$ vector,

$$\mathbf{E}_x(t, z) = \mathbf{E}_{ox} \cos \left[\frac{2\pi v}{\lambda} t - \frac{2\pi}{\lambda} z \right] = \mathbf{E}_{ox} \cos[kvt - kz], \tag{13.3}$$

where

- $k \equiv 2\pi/\lambda$
- v is the speed of light in the solvent
- $v = c/n$, where c is the speed of light in vacuo and n is the solvent's refractive index

Note that the \mathbf{E} vector need not lie in the XZ plane; its plane of propagation can be tilted some angle θ with respect to the x -axis and still propagate in the z -direction (the orthogonal \mathbf{B} vector is then rotated θ with respect to the y -axis). When describing LPL, we examine the positive maximums of the \mathbf{E} vector and make a vector with this maximum and at an angle θ with the x -axis in the XY plane.

Figure 13.2 illustrates what happens to the LPL of wavelength λ passing through a sample chamber containing an optically active analyte. Polarizer P_1 converts the input light to a linearly polarized beam, which is passed through the sample chamber. As the result of the interaction of the light with the optically active solute, the emergent LPL ray is rotated by an angle, θ . In this figure, the analyte is *dextrorotatory*, that is, the emergent \mathbf{E} vector, \mathbf{E}_2 , is rotated clockwise when viewed in the direction of propagation (along the $+z$ -axis). (D-glucose-glucose is dextrorotatory.) To measure the optical rotation of the sample, a second polarizer, P_2 , is rotated until a null in the light intensity striking the photosensor is noted (in its simplest form, the photosensor can be a human's eye). This null is when P_2 's pass axis is orthogonal to $\mathbf{E}_2 \perp \theta$. Since polarizer P_1 is set with its pass axis aligned with the x -axis (i.e., at 0°), it is easy to find θ , thence the concentration of the optically active analyte, $[G]$. In the simple polarimeter of Figure 13.2, and in other polarimeters, the polarizers

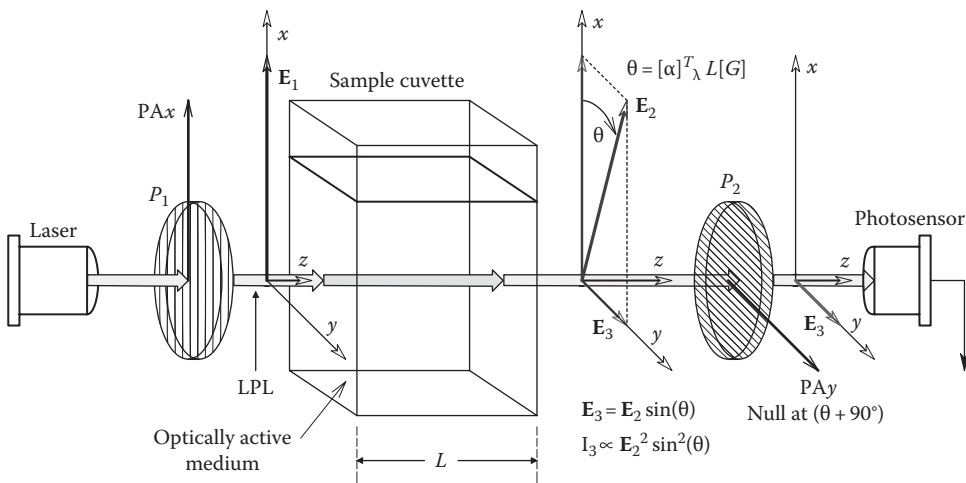


FIGURE 13.2

Diagram of a simple, manually nulled polarimeter. Two linear polarizers are used. The LPL is passed through a sample cuvette of length L containing an optically active analyte, such as D-glucose-glucose. The analyte causes the emergent LPL \mathbf{E} vector to be rotated θ degrees. Polarized P_2 is rotated ($90^\circ + \theta$) to achieve a null at the output, detected by eye or photosensor.

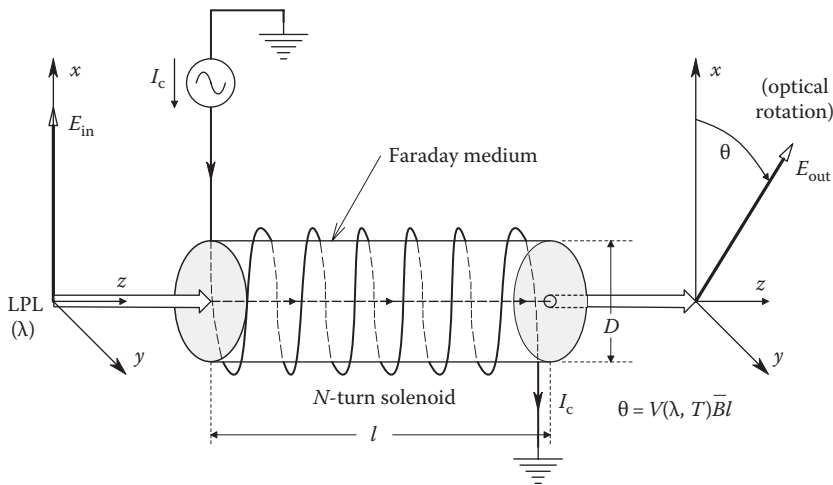


FIGURE 13.3

Diagram of an FR. The entering LPL E vector is rotated θ° by interaction with the axial, B_z vector and the Faraday medium.

are generally made from calcite crystals (Glan calcite) and have *extinction ratios* of ca. 10^4 . (Extinction ratio is the ratio of LPL intensity passing through the polarizer when the LPL’s polarization axis is aligned with the polarizer’s pass axis to the intensity of the emergent beam when the polarization axis is made 90° [is orthogonal] to the polarizer’s axis.)

To make an electronic polarimeter, we need an electrical means of nulling the polarimeter output, as opposed to physically rotating P_2 around the z -axis. One means of generating an electrically controlled optical rotation is to use the *Faraday magneto-optic effect*. Figure 13.3 illustrates a FR. A FR has two major components: (1) a core of a transparent, magneto-optically active medium (many transparent gasses, liquids, and solids exhibit magneto-optic activity; for example, a lead glass rod with optically flat ends or a glass test chamber with optically flat ends containing a gas or liquid can be used); and (2) a solenoid coil wound around the rod or test chamber (core) that generates an axial magnetic field, B , inside the rod or chamber collinear with the entering LPL beam. The exiting LPL beam undergoes optical rotation according to the simple Faraday relation (Hecht 1987):

$$\theta_m = V(\lambda, T)\bar{B}_z l, \tag{13.4}$$

where

θ_m is the Faraday optical rotation in degrees

$V(\lambda, T)$ is the magneto-optically active material’s *Verdet constant*, having units of degrees/($T \times m$)

Verdet constants are given in a variety of different units, such as 10^{-3} min of arc/(gauss \times cm). B_z is the average axial flux density collinear with the optical path, and l is the length of the path where the LPL is exposed to the axial B field.

A material’s Verdet constant *increases* with *decreasing* wavelength and with *increasing* temperature. When a Verdet constant is specified, the wavelength and temperature at which it was measured must be given. For example, the Verdet constant of distilled water at $20^\circ C$ and 578 nm is $218.3^\circ/(T \times m)$ (Hecht 1987, Ch. 8). This value may appear large, but the length

of the test chamber on which the solenoid is wound is $10\text{ cm} = 10^{-1}\text{ m}$, and $1\text{ T} = 10^4\text{ gauss}$, so actual rotations tend to be $< \pm 5^\circ$. The Verdet constant for lead glass under the same conditions is about six times larger. Not all Verdet constants are positive; for example, an aqueous solution of ferric chloride or solid amber has negative Verdet constants.

Two well-known formulas for the axial \mathbf{B} inside a solenoid are given as follows (Kraus 1953). At the center of the solenoid, on its axis,

$$B_z = \frac{\mu NI}{\sqrt{4R^2 + l^2}} \text{ T.} \tag{13.5}$$

At either end of the solenoid, on its axis,

$$B_z = \frac{\mu NI}{2\sqrt{R^2 + l^2}} \text{ T.} \tag{13.6}$$

It can be shown by averaging over l that B_z over the length, l , of the solenoid is approximately

$$\bar{B}_z \cong \infty \frac{3\mu NI}{4\sqrt{4R^2 + l^2}} \rightarrow \frac{3\mu NI}{4l} \text{ T, when } l \gg 2R. \tag{13.7}$$

It should be clear that an FR can be used to electrically null a simple polarimeter. See Figure 13.4 for an illustration of a simple, open-loop, DC, electrically nulled polarimeter.

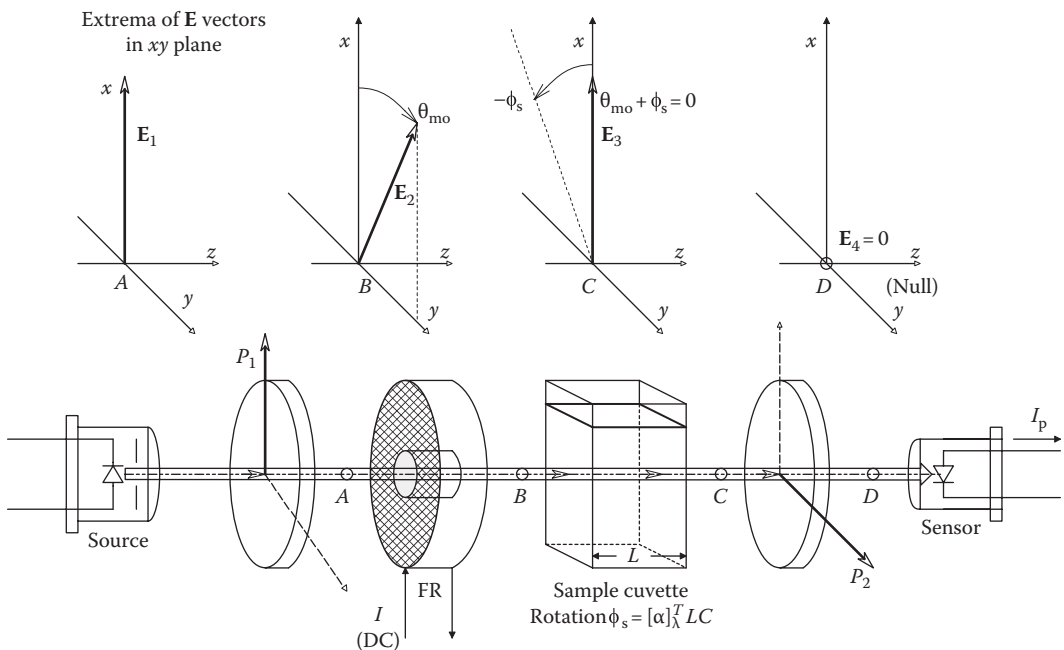


FIGURE 13.4

In this polarimeter, polarizer P_2 is fixed at 90° with respect to the pass axis of P_1 . DC, I , is passed through the FR coil, generating a magnetic field that rotates \mathbf{E}_1 θ_{mo} degrees. The analyte in the cuvette rotates the output \mathbf{E} vector, \mathbf{E}_3 , $-\phi_s$ degrees. I is adjusted so $\theta_{mo} + \phi_s = 0$. This produces a null at point D . Thus, I is proportional to ϕ_s , hence C .

The DC FR current is varied until the output photosensor registers a null ($E_4 = 0$). P_2 's axis remains orthogonal to P_1 's axis. At null, the optical rotation from the FR is equal and opposite to the optical rotation from the optically active analyte. Thus,

$$V(\lambda, T) \frac{3\mu NI}{4l} l = [\alpha]_{\lambda}^T CL. \tag{13.8}$$

And the concentration of the OA analyte is given by

$$C = \frac{3V\mu NI}{4[\alpha]_{\lambda}^T L}. \tag{13.9}$$

Note that C is dependent on three well-known parameters (N , l , and L) and the two are less well known and are temperature and wavelength dependent (V and $[\alpha]$). This type of operation yields accuracy in the tenths of a degree.

Figure 13.5 illustrates an improvement on the accuracy of the static polarimeter shown in Figure 13.4. An audio-frequency AC is passed through the FR coil, causing the emergent polarization vector at point B to rock back and forth sinusoidally in the XY plane at audio frequency, $f_m = \omega_m/2\pi$ Hz. The maximum rocking angle is θ_{mo} . This rocking or

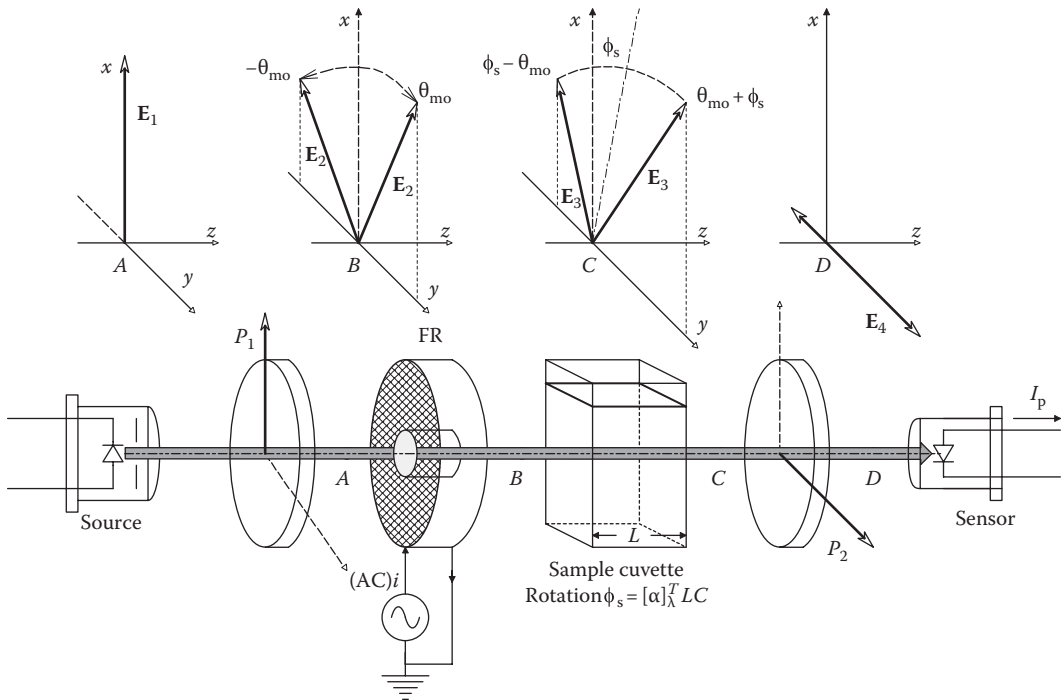


FIGURE 13.5

In this polarimeter, the E_1 vector at point A has its angle sinusoidally modulated in the XY plane by passing an AC through the FR coil; $\theta_m(t) = \theta_{mo} \sin(\omega_m t)$. This AC-modulated E_2 vector at point B is then given a DC optical rotation of ϕ_s by the optically active sample in the cuvette. The second polarizer's pass axis is orthogonal to that of P_1 . Hence, the E_4 vector that is transmitted is time varying and lies along the y -axis. See the text for analysis.

angle-modulated E_2 vector is next passed through the optically active sample where at any instant, the angle of E_2 has the rotation ϕ_s added to it, as shown at point C. Next, the asymmetrically rocking E_3 vector is passed through polarizer P_2 , which only passes the y -components of E_3 , as E_{4y} at point D. Thus, $E_{4y}(t)$ can be written as

$$E_{4y}(t) = E_{4y0} \sin[\phi_s + \theta_{mo} \sin(\omega_m t)]. \quad (13.10)$$

The sinusoidal component in the $\sin[*]$ argument is from the audio-frequency polarization angle modulation. The instantaneous intensity of the light at D, $i_4(t)$, is proportional to $E_{4y}^2(t)$:

$$i_4(t) = E_{4y0}^2 \left[\frac{c\epsilon_0}{2} \right] \sin^2[\phi_s + \theta_{mo} \sin(\omega_m t)] \text{ W/m}^2. \quad (13.11)$$

Now by trigonometric identity, $\sin^2(x) = 1/2 [1 - \cos(2x)]$, and we have

$$i_4(t) = E_{4y0}^2 \left[\frac{c\epsilon_0}{2} \right] \frac{1}{2} \{1 - \cos[2\phi_s + 2\theta_{mo} \sin(\omega_m t)]\}. \quad (13.12)$$

Since the angle argument of the cosine is small, that is, $|2(\phi_s + \theta_{mo})| < 3^\circ$, we may use the approximation $\cos(x) \cong (1 - x^2/2)$. Assume the instantaneous photosensor output voltage, $v_p(t)$, is proportional to $i_4(t)$, so we can finally write

$$v_p(t) = K_P i_4(t) \cong K_P \left\{ E_{4y0}^2 \left(\frac{c\epsilon_0}{2} \right) \right\} \frac{1}{2} \left[1 - \left[1 - \frac{4\phi_s^2 + 4\theta_{mo}^2 \sin^2(\omega_m t) + 8\phi_s \theta_{mo} \sin(\omega_m t)}{2} \right] \right\} \quad (13.13)$$

$$v_p(t) = K_V \left[\phi_s^2 + \theta_{mo}^2 \frac{1}{2} (1 - \cos(2\omega_m t)) + 2\phi_s \theta_{mo} \sin(\omega_m t) \right], \quad (13.14)$$

where $K_V = K_P E_{4y0}^2 (c\epsilon_0/2)$.

Thus, the photosensor output voltage contains three components: (1) a DC component, which is not of interest; (2) a double-frequency component, which has no information on ϕ_s ; and (3) a fundamental frequency sinusoidal term whose peak amplitude is proportional to ϕ_s . By using an HPF, we can block the DC components, and by using a PSR synchronized to the FR modulating current sinusoid at ω_m , we can recover a DC signal proportional to ϕ_s while rejecting the $2\omega_m$ sinusoidal term.

By feeding back a DC current proportional to ϕ_s into the FR, the system can be made closed loop and self-nulling. The first such feedback polarimeter was invented by Gilham in 1957; it has been improved and modified by various workers since then (Rabinovitch et al. 1982, Coté et al. 1990, 1992, 1993, Cameron and Coté 1997, Northrop 2002). Instruments of this sort can resolve ϕ_s to better than ± 20 microdegrees (μ°).

A conventional, closed-loop, polarization-angle-modulated, Gilham polarimeter is shown in Figure 13.6. The DC light source can be a laser or LAD. (We used a 512 nm [green]

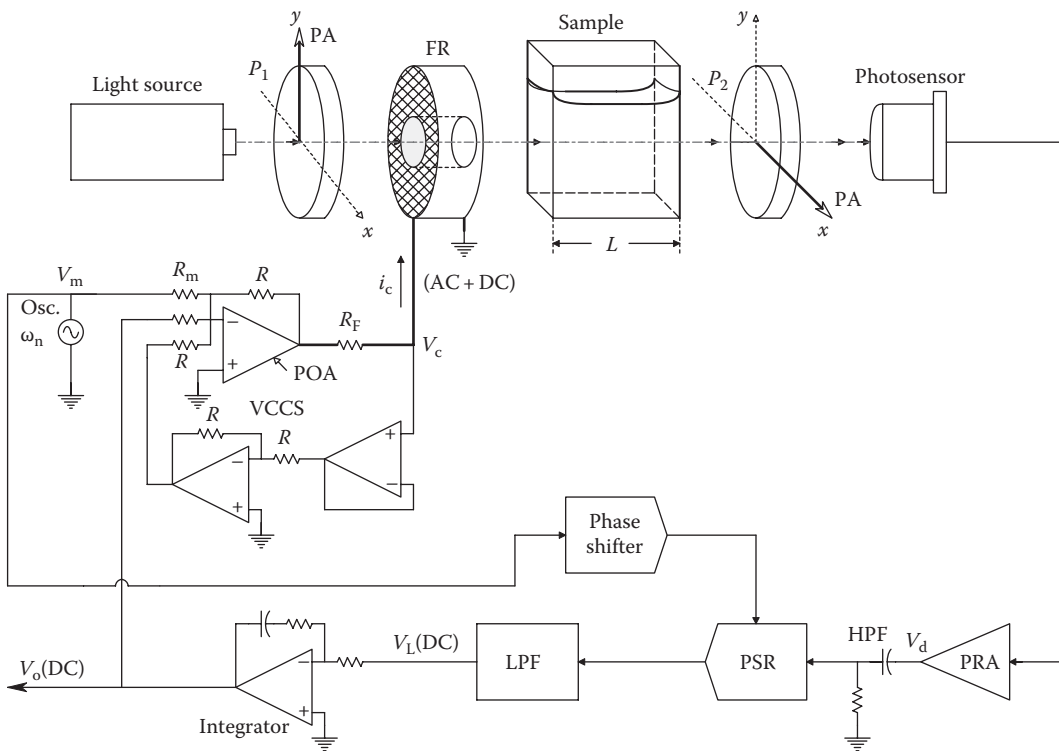


FIGURE 13.6

Diagram of a self-nulling, feedback polarimeter of the Gilham type developed in the author's laboratory. The FR has two functions: (1) It angle modulates the LPL, so the angle of the LPL exiting the sample cuvette is $\theta_m(t) = \theta_{m0} \sin(\omega_m t)$; and (2) A DC is also passed through its coil to counterbalance the DC rotation caused by the sample in the cuvette. See text for a description of the system operation. A VCCS is used to drive the FR coil rather than a VCVS to eliminate heating effects on the coil's resistance. See the text for analysis.

diode laser.) The light from the laser is passed through a calcite Glan-laser polarizer, P_1 , to improve its degree of polarization and then through the FR. In the FR, the input LPL is polarization angle modulated at frequency f_m . θ_{m0} is made about 2° . After passing through the sample, the angle modulation is no longer symmetrical around the y -axis because the optically active analyte's rotation has been added to the modulation angle. As shown previously, this causes a fundamental frequency sinusoidal voltage to appear at the output of the HPF whose peak amplitude is proportional to the desired ϕ_s . The PSR and LPF output a DC voltage, V_L , proportional to ϕ_s .

V_L is integrated and the DC integrator output, V_o , is used to add a DC nulling current component to the FR AC input. This nulling current rotates the angle-modulated output of the FR so that when it passes through the sample, there is no fundamental frequency term in the photosensor output, V_d . Thus, the system is at null, and the integrator output, V_o , is proportional to $-\phi_s$. Integration is required to obtain a type 1 control system that has zero SS error to a constant DC input (Ogata 1990). (In this design, the integrator is given a zero to make it have a proportional plus integral [PI] transfer function to make the closed-loop system stable.)

Browne et al. (1997) devised a feedback polarimetry system that has microdegree resolution, designed to measure glucose in bioreactors. Their system is illustrated in

Figure 13.7. The system of Browne et al. works the same as the polarimeter system of Figure 13.6, with the exception that the FR and the sample cuvette are replaced with a single cylindrical chamber around which is wound the solenoid coil. The 10 cm sample chamber contains D-glucose-glucose dissolved in water and other dissolved salts, some of which may be optically active. (Cells and proteins are filtered out, leaving clear liquid in the sample chamber.)

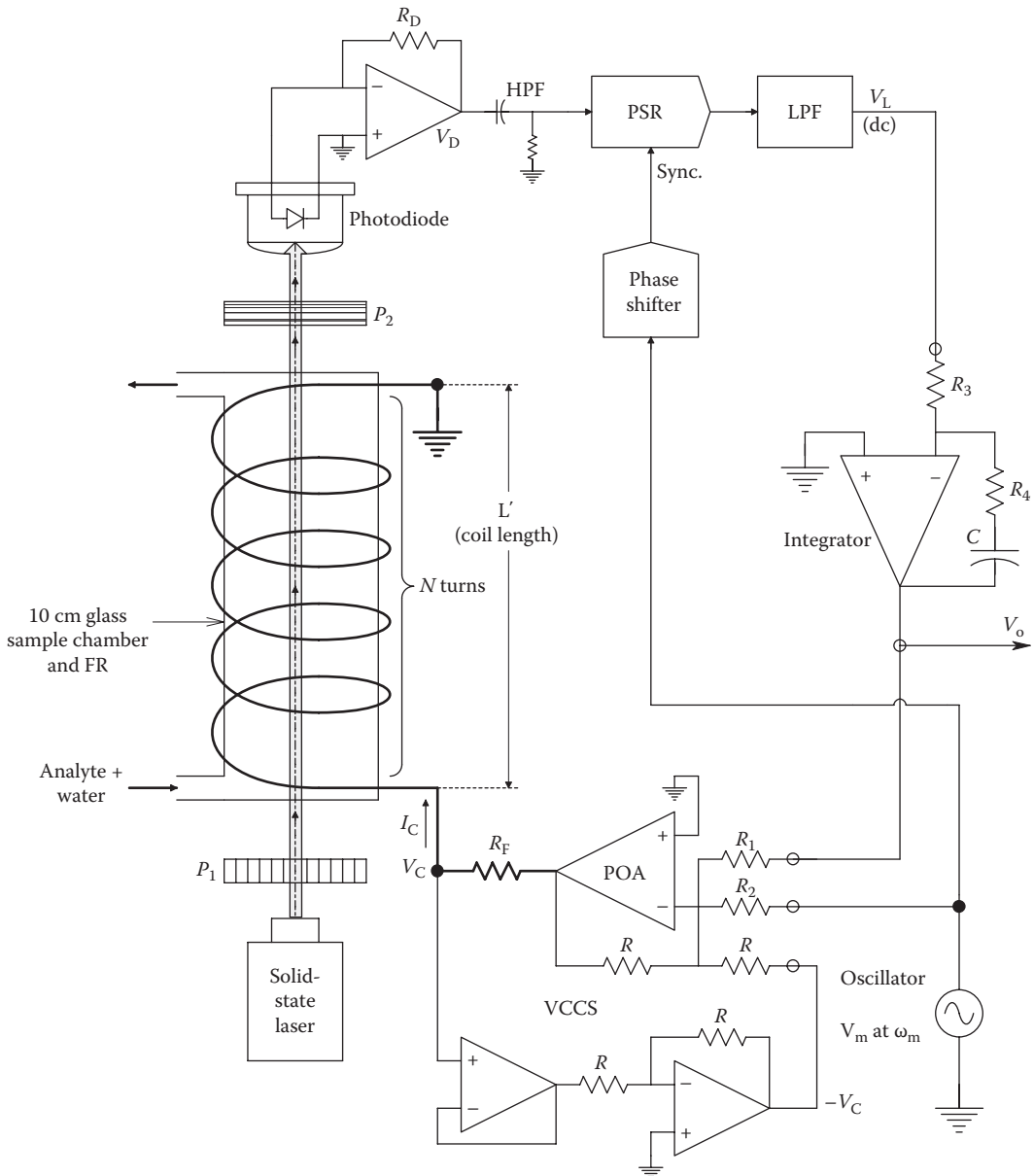


FIGURE 13.7

A closed-loop, servo polarimeter of Browne et al. (1997) that uses the water solvent of the analyte as an FR medium. It was designed to measure glucose in bioreactors. Again, a VCCS is used to drive the coil to avoid heating effects on the coil resistance. See the text for analysis.

The solenoid current is supplied from a three-OA precision VCCS (see Section 4.4). The reason for using a VCCS to drive the coil is because the Faraday magneto-rotary effect is proportional to the axial **B** field, which in turn depends on the current through the coil. Because the coil is lossy, it gradually heats from the power dissipated by the AC modulation excitation and the DC nulling current. As the coil wire temperature increases, its resistance goes up. Thus, if a voltage source (OA) was used to drive the coil, **B** would be a function of coil temperature as well as the input voltage, V_C . This **B** temperature dependence is eliminated by using the VCCS.

The FR can be eliminated from this polarimeter because the water solvent in the chamber has a Verdet constant, and when it is subject to AC and DC axial magnetic fields, they cause optical rotation of the transmitted LPL. The glucose solute is optically active, so we find that in the same path length, optical rotation occurs because of the dissolved D-glucose-glucose and Faraday rotation occurs because of the action of **B** on the water.

One might ask whether the dissolved glucose itself has a Verdet constant and is also acted on by the axial **B** field. It probably does, but the Faraday magneto-rotary effect on the dissolved glucose is much smaller than that on the water solvent. In any case, the Faraday rotation of all solutes present adds (or subtracts) a small amount to the polarization angle modulation produced by the AC **B** component acting on the water. The (feedback) nulling DC also may act on the solutes, affecting system gain, K_{FR} , slightly but uniformly.

It is possible to summarize the operation of this interesting system by a block diagram, shown in Figure 13.8. The input to the measurement system is the physical angle of rotation of the polarization vector, ϕ_s , caused by passing the LPL through the optically active analyte. The negative feedback of the system output voltage, V_o , leads to a Faraday counterrotation of the LPL, ϕ_F , forming a net (error) rotation, ϕ_e . The DC counterrotation, ϕ_F , is simply

$$\phi_F = V\bar{B}L' = V\left(\frac{3}{4}\right)\frac{\mu N}{L'}L'I_c = \left[V\left(\frac{3}{4}\right)\mu N\right]\left[\frac{-V_o R}{R_1 R_F}\right] = -K_{FR}V_o. \tag{13.15}$$

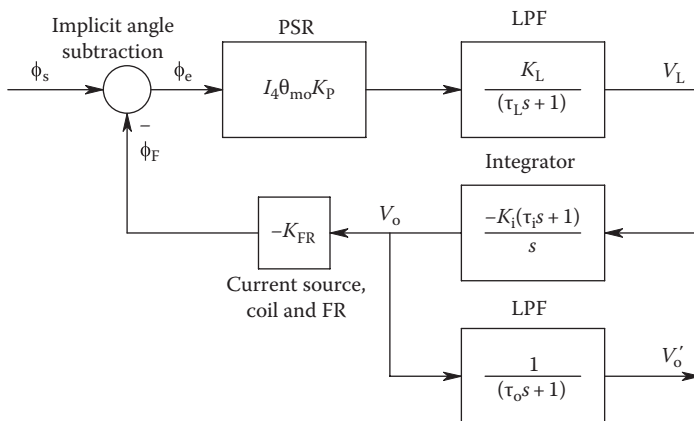


FIGURE 13.8 Systems block diagram describing the dynamics of the feedback polarimeters of Figures 13.6 and 13.7.

Clearly, $K_{FR} = V(3/4)\mu NR/(R_1 R_F)$ degrees/volt. The closed-loop transfer function of the system is found from the block diagram:

$$\frac{V_o}{\phi_s}(s) = \frac{-I_4 \theta_{mo} K_P K_L K_i (\tau_i s + 1)}{s^2 \tau_L + s [1 + \tau_i I_4 \theta_{mo} K_P K_L K_i K_{FR}] + I_4 \theta_{mo} K_P K_L K_i K_{FR}}. \quad (13.16)$$

It is of interest to consider the DC or SS response of this closed-loop polarimeter to a constant ϕ_s .

This can be shown to be simply

$$\frac{V_o}{\phi_s} = \frac{-1}{K_{FR}}. \quad (13.17)$$

Equation 13.17 is a rather serendipitous result. It tells us that the system's calibration depends only on K_{FR} , which depends on the transconductance of the VCCS, the dimensions and number of turns of the coil, the Verdet constant, and the permeability (μ) of water. The DC gain is seen to be independent of the light intensity (I_4), the depth of polarization angle modulation (θ_{mo}), the frequency of modulation (ω_m), and the system's gain parameters (K_P , K_L , and K_i) and time constants (τ_L and τ_i).

The prototype system built and tested by Browne et al. (1997) was evaluated for D-glucose-glucose concentrations (x) from 60 to 140 mg/dL. By least-mean-square error curve fitting, the output voltage (y) was found to follow the linear model

$$y = 0.0544x + 0.0846 \text{ V}, \quad (13.18)$$

with $R^2 = 0.9986$. System resolution was $10.1 \mu^\circ/\text{mV}$ out, at 27°C with a 10 cm cell, and the glucose sensitivity was $55.4 \text{ mV}/\text{mg}\%$. Noise in V_o was reduced by LPF (not shown in Figure 13.7).

Note that in this instrument, OAs were used exclusively for all active components, including the photodiode current-to-voltage converter (transresistor) (cf. Section 2.6.4), the PSR (cf. Section 2.8), the LPF (cf. Section 7.2), the PI integrator (cf. Section 6.6), the oscillator (cf. Section 5.5), and the VCCS (cf. Section 4.4).

13.3 Design of a System to Detect, Measure, and Locate Partial Discharges in High-Voltage Coaxial Power Cables

In the process of aging, the insulation of HV, underground, coaxial power cables may develop physical defects such as water trees and other cavities. Under certain conditions, these defects can serve as the foci for massive insulation failures (arc-overs). Long before catastrophic insulation failure occurs, the water trees and other defect sites, subject to periodic (60 Hz AC) electric field stress in the dielectric, begin to arc over internally, producing PDs. Individual PDs are not fatal events for the cable. However, once begun, they can cumulatively lead to one in time. It is our intention here to describe the signal conditioning circuitry involved in measuring PDs, not the sophisticated digital signal processing techniques used to pinpoint their location in a buried cable.

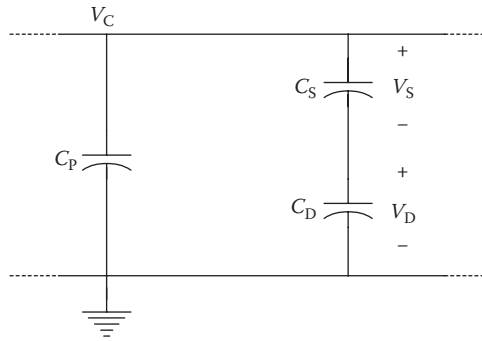


FIGURE 13.9 Low-frequency equivalent circuit of a 60 Hz, underground, coaxial power cable. C_D is the tiny capacitance of the PD cavity. See text for description.

A simple equivalent, low-frequency, lumped capacitive circuit for a power cable with a PD site is shown in Figure 13.9. C_p is the total capacitance between the inner conductor and the outer braided shield. C_D is the tiny capacitance of the water tree or void across which the PD occurs, and C_s is the total series capacitance between the void and the inner conductor and the braid. Generally, $C_p \gg C_s, C_D$.

The PD occurs when the gas or vapor dielectric in the defect cavity breaks down under electrostatic stress, permitting the charge stored in the defect walls to dissipate, temporarily reducing the electric field in the defect. A simplified electrical equivalent of the arc process is shown in Figure 13.10.

For each following AC cycle the PD is repeated, etc. The instantaneous applied voltage on the cable at which a given PD occurs is called the *inception voltage*. At 60 Hz, a coaxial, HV power cable behaves like a large capacitor, C_p . A typical 15 kV, cross-linked polyethylene (XLPE), coaxial power cable has a capacitance of about 60 pF/ft. In the frequency band from 20 kHz to about 50 MHz, the cable behaves like a lossy, RLC transmission line (see Section 9.8), with a characteristic impedance, Z_o , of around 45 Ω .

When the water tree or other insulation defect arcs over at the AC PD inception voltage, the resulting PD injects what is effectively an impulse of charge into the transmission line equivalent circuit of the cable. The PD discharge transient is over in picoseconds; however,

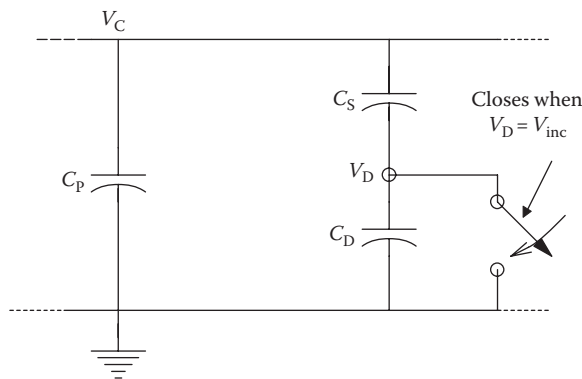


FIGURE 13.10 A simplified schematic illustrating the PD process.

a voltage pulse propagates away from the arc site in both directions on the cable. The direct PD voltage transient recorded at the excitation/measurement (E/M) end of the cable ranges between 1 and 10 mV peak amplitude and several hundred ns in duration. It is superimposed on the kilovolt range, 60 Hz cable excitation voltage, which presents a problem in its separation and measurement.

The direct PD transient first arrives at the E/M end of the power cable at time T_0 . The transient that travels to the far end of the cable is reflected at its open-circuit end and then travels back past the PD site to the E/M end (see Section 9.8). The Thevenin impedance of the HV source driving the E/M end of the cable is higher than the cable's Z_o , so that all PD transients arriving at the E/M end are reflected to the open-circuited, far end of the cable, where they are again reflected, and so on. By examining the *reflection series* of a PD pulse at the E/M end of the cable, and knowing the physical length of the cable, it is possible to estimate the location of the PD site. A typical reflection series from a single PD, recorded from a 15 kV, underground residential distribution (URD), XLPE dielectric, coaxial power cable in the author's laboratory is shown in Figure 13.11. Note how successive pulses are attenuated and broadened by the cable's LPF characteristics. In the field, if a PD is judged to be severe, the section of cable containing the PD site is excised, and a new cable section is spliced in, or the entire cable may be replaced with a new one.

The initial voltage transient at the PD site can be modeled, following the lumped-parameter development of Mason (1965). Three capacitors are used in Mason's model: C_D is the equivalent capacitance of the insulation defect giving rise to the PD. C_S is the equivalent series capacitance between C_D and the cable's outer shield and the cable's center conductor. C_p is the total cable capacitance, less C_S and C_D in series. A PD occurs when the 60 Hz voltage on the cable reaches the defect's inception voltage, V_{inc} . At this time, the dielectric of the insulation defect breaks down, and the voltage across C_D quickly goes to zero as its charge is neutralized. This situation is illustrated in Figure 13.10. At $t = 0^-$, just before the PD occurs, the charge stored in C_p is $Q_p(0^-) = C_p V_{inc}$. It is easy to see that the charges stored

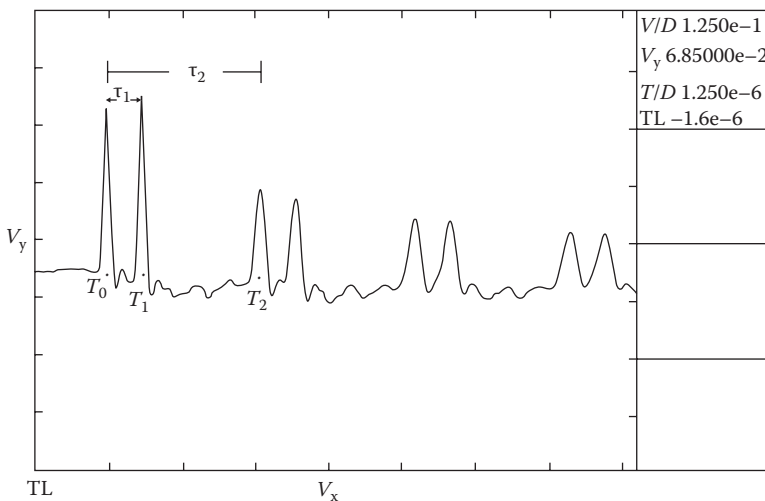


FIGURE 13.11

PD waveforms generated by an insulation defect near the far end of a 730 ft coaxial power cable. Amplitude of the first recorded PD pulse in the series is 375 mV. $\tau_1 = 600$ ns and $\tau_2 = 2.640$ μ s. Pulses have been conditioned by a voltage gain of 100 as well as HPF.

on C_S and C_D are equal and equal to $Q_{SD} = V_{inc} C_S C_D / (C_S + C_D)$. Thus, the voltage across the defect cavity at breakdown is

$$V_D = \frac{Q_{SD}}{C_D} = V_{inc} \frac{C_S}{C_S + C_D}. \quad (13.19)$$

The voltage across C_S is

$$V_S = \frac{Q_{SD}}{C_S} = V_{inc} \frac{C_D}{C_S + C_D}. \quad (13.20)$$

Obviously, $V_{inc} = V_D + V_S$.

The PD occurs at $t = 0$. At $t = 0+$, $V_D \rightarrow 0$, $Q_D \rightarrow 0$, and $V_S \rightarrow V_{inc}$. Now the total charge on C_P and C_S is

$$Q_{PS} = C_P V_{inc} + Q_{SD} \text{ (on } C_S \text{)}. \quad (13.21)$$

Assuming the transmission line's series inductance prevents instantaneous longitudinal charge flow, the voltage across the PD site at $t = 0+$ is now

$$V' = \frac{Q_{PS}}{C_P + C_S} = \frac{V_{inc} [C_P + C_S C_D / (C_S + C_D)]}{C_P + C_S}. \quad (13.22)$$

The initial voltage change is nearly instantaneous, with a rise time estimated to be in picoseconds. The voltage change at the site is approximately

$$\Delta V = V' - V_{inc} = V_{inc} \left[\frac{C_P + C_S C_D / (C_S + C_D)}{C_P + C_S} - 1 \right] \cong \frac{-V_{inc} C_S^2}{C_P (C_S + C_D)}. \quad (13.23)$$

Let us calculate a typical PD initial amplitude at the PD site: Let $C_P = 10$ nF, $V_{inc} = 2000$ V, $C_S = 1$ pF, and $C_D = 3$ pF. From Equation 13.23, $\Delta V = -50$ mV. The 60 Hz excitation on the power cable is over 92 dB larger than the initial PD amplitude! As shown in Figure 13.11, as the PD *spike* propagates along the cable, it is attenuated and rounded due to the attenuation of high frequencies by the cable's low-pass transfer function. Measurements on the mV range PD reflection series require the separation of the PD transients from the high 60 Hz excitation voltage. Figure 13.12 illustrates a simplified block diagram of the signal acquisition portion of the PD detection, location, and characterization system described in the following US patents: Mashikian et al. (1989, 1993, 1995).

The cable is excited from a clean, variable-voltage, 60 Hz HV source. There must be no corona or PDs in the HV power source, which can be confused with or mask PD signals from the cable. A Tettex AG (now Haefely Test AG) type 3380/100/70, 70 kV, 10 nF, compressed gas, standard HV capacitor was used to couple voltages on the cable under test to the first HPF (HPF1) used to condition the PD signal. A low-noise, OA, signal conditioning amplifier (A_1) with a gain of 10 is used to boost the PD signal at the output of HPF1. A_1 is followed by a second HPF (HPF2) to further reduce 60 Hz voltage. The output of HPF2 is

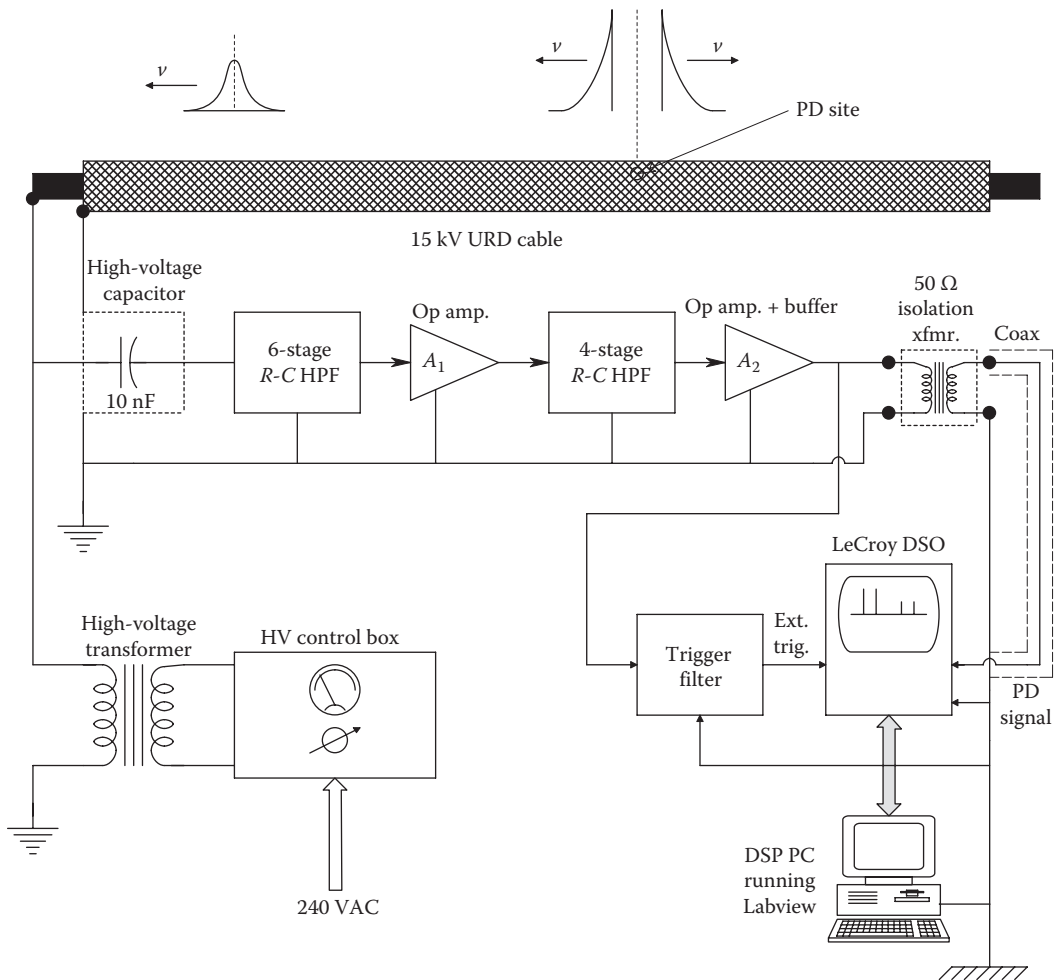


FIGURE 13.12

Block diagram of the PD measurement system of Mashikian et al. (1989). The OA amplifiers each have a voltage gain of 10 and -3 dB bandwidths at 5 MHz. The HPFs are shown in detail in Figure 13.13.

conditioned by a second broadband amplifier (A_2) with a gain of 10. AD840 OAs were used because they have low-voltage noise, high f_T and slew rate, and low cost. A_2 drives a 50 Ω , high-frequency, isolation transformer (North Hills Signal Processing Corp., model 0016PA, 20 Hz to 20 MHz, 50 Ω Z_{in} and Z_{out}), which is used to prevent ground loops and protect the DSO in the event of a massive cable arc-over. A 50 Ω coaxial cable carried the PD signals to a LeCroy DSO, which served to digitize and record the PD transients. Digital records from the DSO's memory or floppy disks are passed on to a PC used for further DSP operations on the signals. The schematics of the HPFs are shown in Figure 13.13. Figure 13.14 shows the calculated frequency response of the two, cascaded HPFs and amplifiers. The gain of the combined filters and amplifiers is -262 dB at 60 Hz. This means that the 60 Hz HV cable is attenuated by a factor of 7.94×10^{-14} . Thus, a 10 kV pk, 60 Hz input to the filter produces a 0.794 nV pk, 60 Hz interference at its output. The combined HPFs have a -3 dB frequency of 25.2 kHz, and the passband gain is $+40$ dB, as expected. The SNR of weak, noisy PDs is

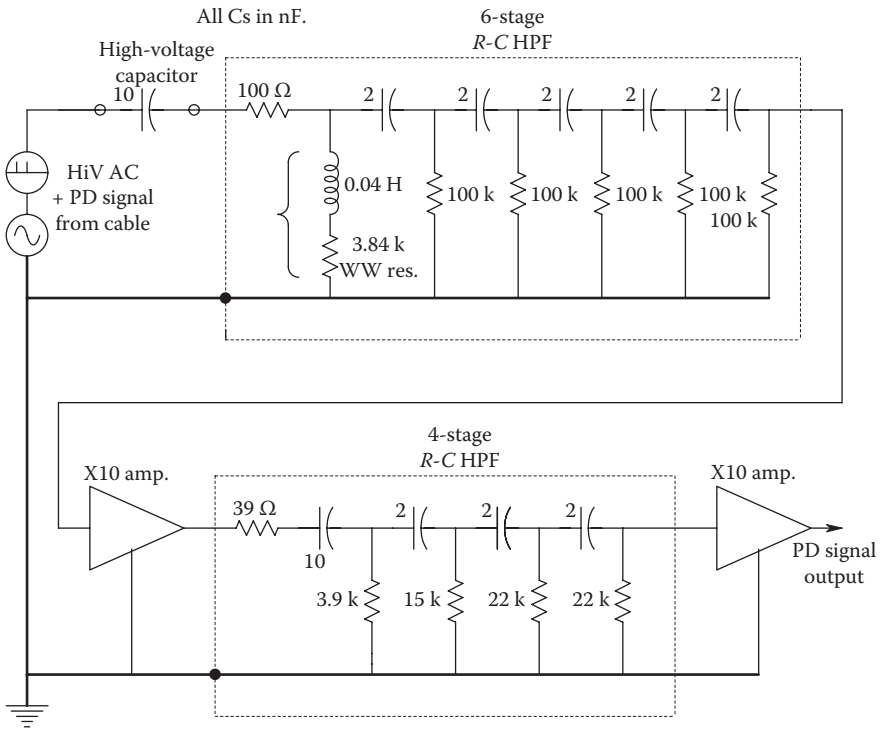


FIGURE 13.13 Schematic diagram of the signal conditioning HPFs. The OA circuits are represented simply by gain of 10 VCVSs.

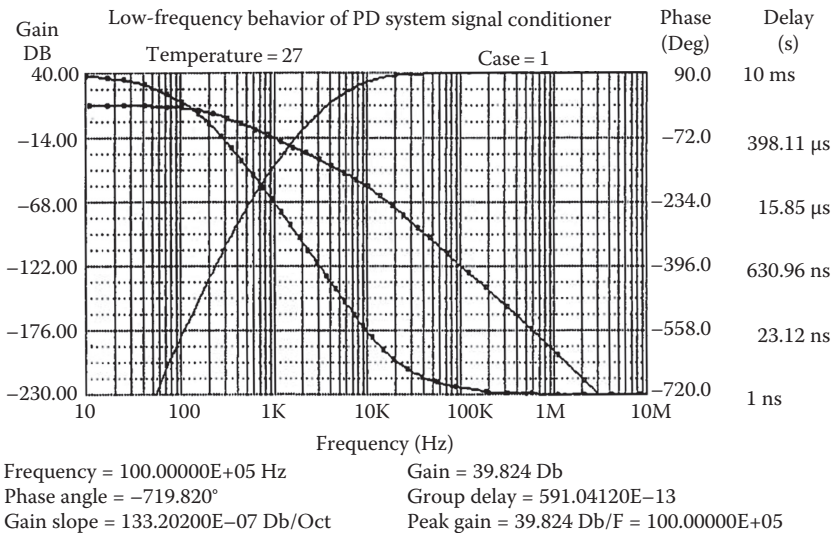


FIGURE 13.14 Calculated frequency response of the HPFs shown in Figure 13.13. Note that the total filter attenuation at 60 Hz is -262 dB. See text for description.

improved by signal averaging on the LeCroy DSO. Averaging is only possible given reliable, consistent triggering of the DSO sweep by the peak of the first (largest) PD transient recorded at the E/M end of the energized power cable. A special NBPF is used on the PD waveforms to more reliably extract the triggering pulses for signal averaging on the DSO. When the PD SNR is too low for reliable averaging on the DSO, we use a cross-correlation technique on the PC to establish a reference time $T = 0$ at the start of each noisy PD pulse group, so averaging can take place reliably. The DSO sampling period is normally set at 20 ns/sample; 16,000, 8-bit samples are taken on a delayed input signal, so the full extent of the first (trigger) PD pulse can be seen.

After a suitable PD reflection series or averaged reflection series is acquired, it can be transferred to the PC on the IEEE-488 bus for further DSP operations, which include proprietary PD site location algorithms. One location algorithm deconvolves the broad, recorded pulse series with the cable transfer function in order to better estimate the true times of occurrence of the pulses in the reflection series. Another location algorithm uses a maximum likelihood (ML) method with cross correlation (Knapp et al. 1990). The ML location algorithm has been found to be robust in the presence of noise. PD site location is based on the a priori knowledge of the cable's length. From the PD reflection series shown in Figure 13.11, we can find the PD propagation velocity on the cable, c_c , from the fact that it takes τ_2 seconds for a pulse to travel from the E/M end of the cable to the far end and then return a distance of $2L$ meters. Thus,

$$c_c = \frac{2L}{\tau_2} \text{ m/s.} \quad (13.24)$$

c_c on most XLPE HV cables is about 200 m/ μ s or 0.667 times the speed of light.

The true time of origin of the PD is initially unknown, so we work with the time differences between the first three pulses in the reflection series. Thus,

$$\tau_1 = T_1 - T_0 \quad (13.25)$$

and

$$\tau_2 = T_2 - T_0. \quad (13.26)$$

The time differences are typically measured using the maxima of the pulses corrected for broadening and attenuation by the cable's transfer function. Equation 13.24 can be used to write

$$\tau_1 = (2L - x_d)/c_c - x_d/c_c. \quad (13.27)$$

Equation 13.27 is solved for x_d , the distance from the E/M end of the cable to the site of PD origin. c_c is eliminated by substituting Equation 13.24 into Equation 13.27. The final result is

$$x_d = L(1 - \tau_1/\tau_2). \quad (13.28)$$

Thus, we have only to know the total cable length, L , and estimate the time differences τ_1 and τ_2 in order to find an estimate for x_d .

In the embodiment described, the PD measurement and locating system is able to locate artificial PD sites at known locations on a cable with an accuracy of 99.5% (Knapp et al. 1990). When the PD site is very near on either end of the cable (<30 m from the ends), the PD pulse peaks merge, and estimation of the pulse origin times is not ordinarily possible. By splicing an extra 30 m of defect-free, identical cable to the far end of the cable under test, we can obtain distinct separation of the PD pulses and improved location resolution for PD sites closer than about 30 m from the cable ends.

Another problem that affects the SNR of the system and hence the resolution of x_d is the pickup of coherent interference in the system's 20 kHz to 10 MHz signal bandwidth. Such interference is from RF sources such as local AM broadcast stations. In certain cases, when the cable is above ground, such RFI can be of the same order of magnitude as the PD reflection series, confounding measurements. Our approach to this problem is to do a spectral analysis of the system noise to locate the frequencies of the coherent interference; we then implement one or more adaptive, digital, FIR notch (band-reject) filters on the system computer to remove the interference, leaving the PD waveforms largely intact. Of course, good grounding, shielding, and guarding practices were followed. As mentioned earlier, a broadband isolation transformer was used to reduce common-mode noise. However, RFI can be picked up by the cable under test in spite of these practices. System broadband random noise arises in the head-stage amplifier and in the HPFs. AD840 OAs were used for broadband amplifiers. These OAs have $f_T = 400$ MHz, $\eta = 400$ V/ μ s, and $e_{na} = 4$ nV/ $\sqrt{\text{Hz}}$. Their broadband noise is ca. 10 μ V rms in a 10 Hz to 10 MHz bandwidth.

Our instrumentation allowed the detection and reliable site estimation (x_d) of PD pulses as low as 80 μ V peak at the E/M end of the cable using a 5 MHz signal bandpass. Although the estimated PD site location is found in terms of % of the total known cable length, the actual physical location of the PD site is complicated by the fact that the cable plunges into its trench and then snakes a certain amount along its path. In practice, a hole is dug to the cable at the expected site, and then a transient test pulse is injected into the unenergized cable by magnetic induction. A fast current pulse is passed through a coil with about 25 turns around the outside of the cable. This induces a PD-like transient and a corresponding reflection series, which can be compared to the stored PD series. A coincidence of pulse peaks in the two series means the test coil is very near the PD inception site. Accuracy with the current pulse *vernier* method is limited by the DSO sampling rate. For repetitive signals, such as the current test pulses, the effective sampling period of the DSO is 2 ns/samples. This interval corresponds to a distance on the cable of 0.4 m, giving a comparison location error of 0.2% on a 200 m cable. The results of ongoing development of DSP algorithms for pulse time difference estimation suggest that a routine location accuracy of 99.8% will be possible. This accuracy translates into a ± 0.5 m error in finding a PD site on a 250 m cable. The use of sophisticated DSP filtering techniques on noisy data is what has made the prototype PD measurement instrument evolve into a successful, field-portable system.

13.4 Design of a Closed-Loop, Constant-Phase, Pulsed Laser Ranging System and Velocimeter

Laser velocimeters are used by police to measure vehicle velocities on highways. They are also used to measure the relative velocity and range between two space vehicles during docking maneuvers. In the police application, the pulsed laser is generally directed from a

stationary transmitter/receiver in line with vehicle travel, as from a bridge overpass or on a curve to a nonspecular surface on the moving vehicle. A telescope collimated with the laser is used to collect the partially reflected signal and direct it to a fast, sensitive photodetector, such as an avalanche photodiode. There is a Doppler shift on the frequency of the return signal, but this is not what is measured. The time delay between transmission and the detection of the return pulse is what is quantified. In 3.33 ns, light travels 1 m. Unlike Doppler radar velocimeters, the LAVERA systems are highly directional, subject only to the transmitted laser beam's divergence. It is this specificity that reduces target ambiguity when one or more other vehicles are in close proximity to the desired target. In order to image the laser spot on the target vehicle, the vehicle surface should be a nonspecular (unlike a mirror), diffuse reflector with a high albedo (brightness). A mirrorlike finish at an angle to the transmitted beam will cause the reflected beam to be directed away from the receiver telescope, and not be sensed, generating a *stealth* condition. Thus, slanted windshields and chrome bumpers on automobiles make poor target surfaces. A black vinyl bumper is also a poor target because of its low albedo. Such poorly responding targets are called *uncooperative*. A white painted surface is more cooperative, and a retroreflecting surface makes an ideal, cooperative target, as nearly all of the laser's energy is directed back parallel to its beam.

Early LAVERA systems used by police generally operated in an open-loop mode, the laser being pulsed at a constant rate (e.g., 381 pps in an early Laser Atlanta Optics, Inc., unit). The time between transmitting any one pulse and the detection of its return image from the moving vehicle is measured and stored in a digital memory. One way of measuring the return time, which is proportional to the target range, is to have the k th transmitted pulse start a fast Miller integrator circuit (similar to the horizontal sweep circuit in an analog oscilloscope) generating a ramp of voltage. A fast, T&H circuit follows the ramp until the k th return pulse triggers the T&H circuit to hold the ramp voltage at that instant. The held ramp voltage is then converted to digital form by an ADC, whose output is stored in digital memory. The previous $(k - 1)$ th digital word, proportional to the range at time t_{k-1} , is subtracted from the present k th digital word, which is proportional to the range at time t_k . The difference, appropriately scaled, is proportional to vehicle velocity. In practice, an FIR smoothing routine operates on the sequence of differences to reduce noise. This approach has two serious drawbacks: (1) It does not provide range information, and (2) its bandwidth, hence its ability to respond to sudden changes in vehicle velocity, is limited by the low, constant pulse rate of the laser source.

Another open-loop LAVERA system, described by Koskinen et al. (1992), used a higher, fixed, pulse rate (between 1 and 50 MHz) and an ECL time-to-amplitude converter (TAC), which is, in effect, a precise phase detector.

The components of a typical LAVERA system consist of a pulsed, LAD source, usually about 1.5 μm wavelength, collimating optics, a target (moving or stationary, cooperative or uncooperative), a receiving telescope, a high-speed photosensor, an electronic signal conditioning amplifier, a time-to-voltage converter to estimate target range, signal processing circuitry (digital and/or analog) to give speed, and a readout display. In the following description of the closed-loop, constant-phase, LAVERA system, we describe the systems architecture of the range and velocity estimating circuits and do not consider the details of the pulse circuits, the laser, optics, the sensor, or preamplifier.

Figure 13.15 shows a functional block diagram of the closed-loop LAVERA system developed by the author. Note that this system will work either with a pulsed or sinusoidal light intensity output. The loop integrator has a compensatory zero associated with it, giving the closed-loop system a circle root locus. The zero is at $s = -1/\tau_c$ in the s -plane.

There are several possible DPD circuits that can be used. One DPD circuit that we used was a DPD based on the Motorola 4044 IC; it compares the phase between the k th transmitted

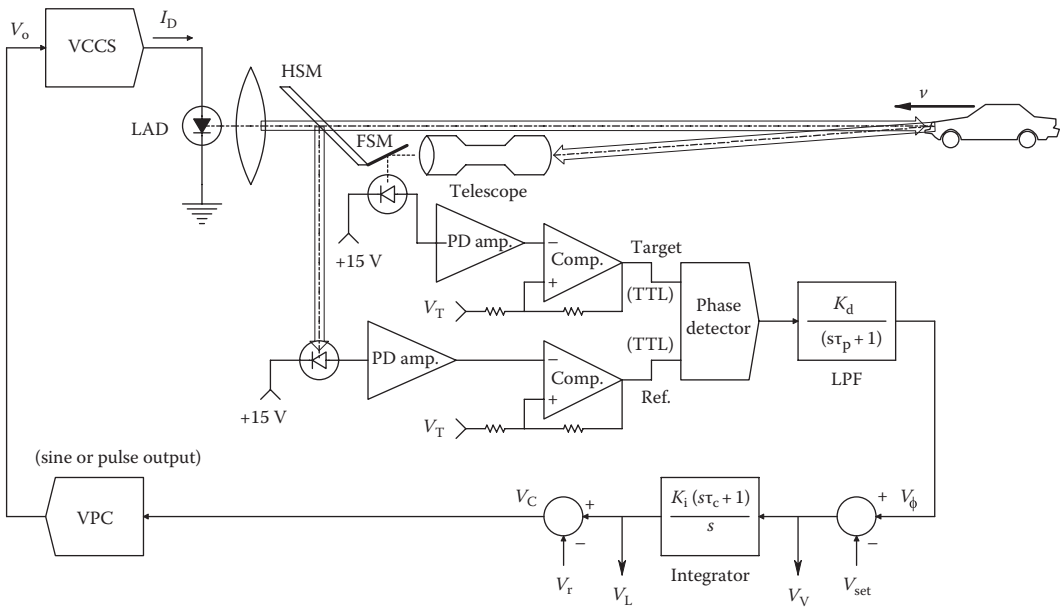


FIGURE 13.15

Block diagram of the closed-loop, constant-phase, pulsed LAVERA system devised by the author. The heart of the system is the VPC-controlled oscillator, V_V is the system output voltage proportional to target velocity, and V_L is the output voltage proportional to the target range, L .

pulse and its *echo* received $\tau_k = 2L_k/c$ s later, where L_k is the distance to the target (assumed constant over τ_k) and c is the velocity of light in air. If the frequency of the pulses, f_c is constant, the phase lag between received and transmitted pulses can be expressed as

$$\phi = 360 \frac{2L}{cT_c} = (2.40 \times 10^{-6}) L f_c \text{ degrees } (L \text{ in m, } f_c \text{ in Hz}). \tag{13.29}$$

The *average* voltage at the output of the LPF conditioning the output of the DPD is $V_\phi = K_d\phi$. V_ϕ is next conditioned by an integrator with phase lead compensation, $K_i(s\tau_c + 1)/s$. The output of the integrator, V_L , has a DC level, V_r subtracted from it to form V_C , the input to the VPC. Do not confuse the VPC with the well-known VCO or VCO IC. The frequency output of a VCO behaves according to the rule $f_c = K_V V_C + c$.

In the VPC, it is the *period* of the oscillator output that is directly proportional to V_C . Thus, for the VPC, we can write

$$T_c = \frac{1}{f_c} = K_P V_C + d. \tag{13.30}$$

The VPC output, V_o , can be sinusoidal or pulsed, depending on the application. V_o drives a VCCS, which provides current to drive the LAD. The LAD optical output provides a reference phase input to the DPD IC. The entire system operates as a closed-loop, constant-phase, type I control system. The *period* of the VPC output is automatically adjusted as L changes to maintain a constant ϕ and V_ϕ . Because the closed-loop system has a pole at the origin of the s -plane, it has zero SS phase error for constant L (stationary targets) and a small SS phase error

for targets moving at a constant velocity ($dL/dt = v_o$). We can use this general property of a type I control system to examine the SS relationships governing the output voltage proportional to velocity, V_v , and the output voltage proportional to range, V_L . If the target is at rest at range L_o , then V_v will be zero, and thus $V_\phi = K_d\phi = V_{SET}$. After some algebra, the SS V_{L_o} is found to be

$$V_{L_o} = \frac{K_d 2.40 \times 10^{-6} L_o}{K_V V_{SET}} + (V_r - d/K_V) = S_L L_o + \lambda, \quad (13.31)$$

where λ is a constant. Let the desired, SS phase difference in the system be ϕ_o . If we substitute $V_{SET}/K_d = \phi_o$ into Equation 13.31, we obtain

$$V_L = \frac{2.40 \times 10^{-6} L_o}{K_P \phi_o} + (V_r - d/K_V). \quad (13.32)$$

Now let the target be moving at a constant velocity, $dL/dt = v_o$. Now V_v is nonzero and may be shown to be

$$V_v \cong \infty \frac{dV_L/dt}{K_i} = \frac{K_d 2.40 \times 10^{-6} v_o}{K_i K_V V_{SET}} = S_V v_o. \quad (13.33)$$

We now consider the system's dynamics. Clearly, the system is nonlinear because the input, L , appears as a multiplicative input. Figure 13.16A illustrates the nonlinear system's block diagram of the LAVERA system. In order to *linearize* the system to examine its loop gain, we consider the nonlinear transfer function of the VPC as a linear gain with a DC component, of the form, $f_c = S_V V_C + e$, where S_V and e are determined at the system's operating point. By differentiation,

$$S_V = \frac{df_c}{dV_C} = \frac{-K_P}{(K_P V_C + d)^2} = -K_P (f_c)^2 = -K_P \left[\frac{\phi_o}{(2.40 \times 10^{-6} L_o)} \right]^2. \quad (13.34)$$

The linearized system is shown in Figure 13.16B. Its loop gain is second order and is given by

$$A_L(s) = -\frac{K_V \phi_o^2 K_d K_i (s\tau_c + 1)}{2.40 \times 10^{-6} L_o s (s\tau_p + 1)}. \quad (13.35)$$

It can easily be shown that the closed-loop, nonlinear system's natural frequency squared is

$$\omega_n^2 = \frac{K_V \phi_o^2 K_d K_i}{\tau_p L_o 2.40 \times 10^{-6}} (r/s)^2. \quad (13.36)$$

The system's damping factor for complex-conjugate poles is also nonlinear and is given by

$$\zeta = \left(\frac{1}{2} \right) \frac{2.40 \times 10^{-6} L_o + K_V \phi_o^2 K_d K_i \tau_c}{\sqrt{K_V \phi_o^2 K_d K_i 2.4 \times 10^{-6} L_o \tau_p}}. \quad (13.37)$$

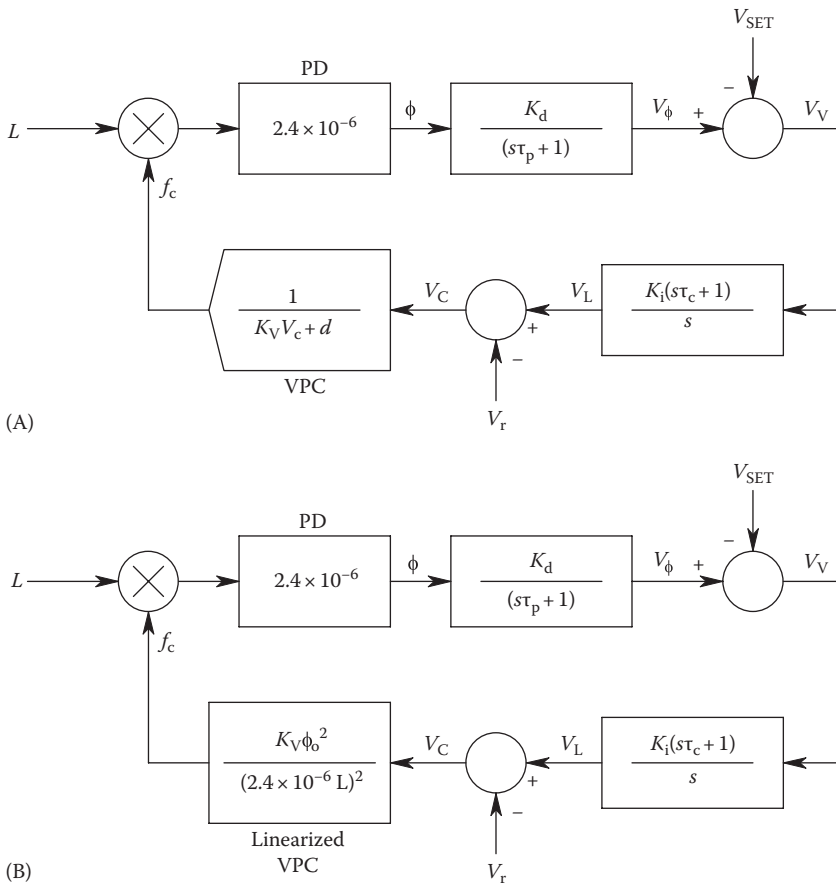


FIGURE 13.16 (A) Block diagram of the non-linear LAVERA system illustrating the VPC and the $L \times f_c$ multiplier operation. (B) LAVERA block diagram in which the VPC has been linearized. Its loop gain is given by Equation 13.35.

As an example, let the following parameters be used: $L_o = 100$ m, $\phi_o = 25.0^\circ$, $K_i = 500$, $\tau_p = 0.01$ s, $\tau_c = 0.0001$ s, $K_d = 0.01$, $K_V = 4.75E-6$. We calculate $S_V = 2.4E-6 / (K_i K_p \phi_o) = 4.042E-5$ V/m/s. The range sensitivity is $S_L = 2.4E-6 / (K_p \phi_o) = 2.0211E-2$ V/m. The undamped natural frequency is

$$\omega_n = \sqrt{\frac{K_V \phi_o^2 K_d K_i}{2.4E-6 \tau_p L_o}} = 78.64 \text{ r/s, or } f_n = 12.5 \text{ Hz.} \tag{13.38}$$

ω_n is seen to increase as range, L , decreases. The natural frequency calculated earlier can easily be increased at 100 m by increasing K_i , K_d , etc., and by decreasing τ_p . The closed-loop system damping factor at a range of 100 m is $\zeta = 0.64$ (underdamped). If L varies from 10 to 300 m, the SS VPC frequency varies from 1.042 MHz to 34.7 kHz.

An experimental investigation of the LAVERA system described previously was undertaken in the author’s laboratory by Nelson (1999b). A simplified circuit schematic diagram for a simple VPC is shown in Figure 13.17. A CW, sinusoidal amplitude modulated light beam was used because the amplifiers used to condition the photonic signals required

much lower bandwidths than had a pulsed output been used; hence cost was lower. The VPC oscillator was made from an analog reciprocal circuit (Northrop 2004) that drives a Maxim MAX038 linear VFC VCO. The VPC sinusoidal output drives a three-OA VCCS circuit, which in turn provides current to the two NIR (675 nm), 5 mW, LADs. One is the transmitter and the other is used as a phase reference source with an ND filter to prevent saturating the reference photodiode. The sinusoidal signals from the LADs are conditioned and then changed to TTL square waves by a pair of fast comparators. The phase detector in this case is a simple EOR gate followed by a DC shifter and LPF. After integration and rectification to prevent V_L from going <0 , and zener clipping to limit the positive input, V_L drives the VPC. Nelson's system was developed as a prototype navigational aid for the blind. A two-channel acoustic output was attached to V_V and V_L . In one embodiment, the frequency of a CW audio tone was made inversely proportional to range, L , and object velocity toward the subject elicited a second signal of clicks whose rate was proportional to v . Object velocity away from the subject produced no clicks; the decreasing frequency of the CW tone signaled negative v .

Note that the system architecture of the LAVERA system can be extended to ultrasound to simultaneously measure velocity and distance of objects moving in air or water or even in the body (such as aneurisms or heart motion). Unlike Doppler systems, a LAVERA system gives object displacement as well as velocity. At long ranges, pulsed ultrasound bursts can be used; for short-range applications, CW ultrasound is indicated.

13.5 Design of Capacitive Sensors for the Detection of Hidden Objects

13.5.1 Introduction

Historically, there have been many approaches to sensing small changes in capacitance (ΔC). Some of these techniques were developed to measure small changes in dielectric constant, small changes in mechanical displacement (of a diaphragm or cantilever) associated with the measurement of pressure or force, small changes in capacitance caused by the presence of buried objects having dielectric constants different from soil (this might include land mines or ceramic pipes buried in soil or concrete), the presence of humans in places they should not be (including burglars, workers near dangerous machinery such as robot arms, and kids near school busses), liquid level changes, water content in foods, etc. A human body is mostly water; thus, it has a high dielectric constant, which makes it easy to sense by ΔC methods.

One of the earliest means devised for sensing ΔC is the De Sauty bridge (see Section 5.5.1.4 and Figure 5.7). The bridge is generally operated around null. The output of a De Sauty bridge given a small capacitance unbalance can be written as

$$\begin{aligned} V_o &= V_1 - V_2 = V_b \left\{ \frac{1/j\omega C_o}{[R + 1/j\omega C_o]} - \frac{1/j\omega(C_o + \Delta C)}{[1/j\omega(C_o + \Delta C) + R]} \right\} \\ &\downarrow \\ V_o &\cong V_b \left\{ \frac{j\omega R C_o (\Delta C / C_o)}{[j\omega R C_o + 1]^2} \right\}. \end{aligned} \quad (13.39)$$

For $\omega = 1/RC_o$ r/s, the bridge has maximum sensitivity. The output can then be shown to be

$$V_o = V_b \left[\frac{1}{2} \left(\frac{\Delta C}{C_o} \right) \right]. \tag{13.40}$$

As we stated in Chapter 5, stray wiring capacitances from the bridge corners to ground and the input capacitance of the differential amplifier make the output conditions more complex than the simple output (Equation 13.40). Once a De Sauty bridge is taken off the lab bench, its symmetry is hard to preserve, and it requires compensation for stray capacitances to make it sensitive to very small $\Delta C/C_o$.

Other means of sensing small changes in capacitance have relied on various servo nulling techniques for capacitance bridges. Two such OA feedback circuits are shown in Figures 13.18 and 13.20.

In the simple feedback (driven guard) circuit of Figure 13.18, the AC voltage at the OA output is examined:

$$V_o = I_C R + V_s = V_s [j\omega CR + 1]. \tag{13.41}$$

It is instructive to examine both the *phase* and *magnitude sensitivity* of V_o . The phase of V_o is

$$\phi = \tan^{-1}(\omega RC). \tag{13.42}$$

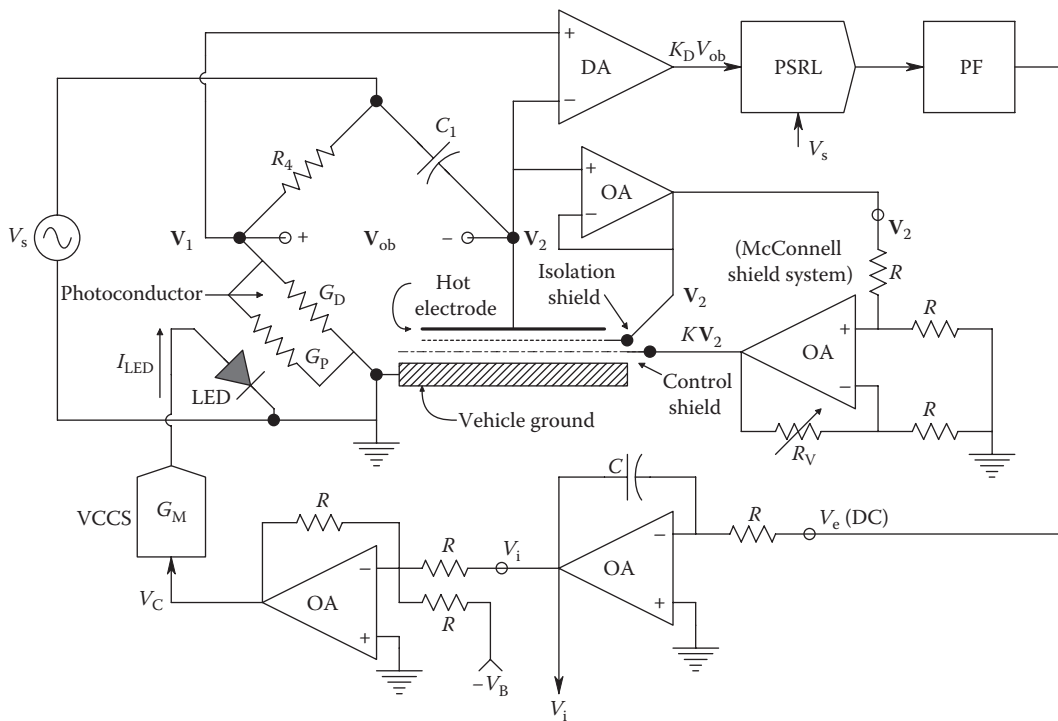


FIGURE 13.18

An active circuit that can measure changes in capacitance as changes of voltage or output phase. A driven guard shield is used to enhance sensitivity.

The phase sensitivity of V_o with respect to C is

$$S_C^\phi = \frac{d\phi}{dC} = \frac{\omega R}{1+(\omega RC)^2} = \frac{\omega RC}{C[1+(\omega RC)^2]} = 0.5/C \quad \text{for } \omega = \frac{1}{RC} \text{ r/s.} \quad (13.43)$$

The magnitude of V_o is

$$M = \left| \frac{V_o}{V_s} \right| = \sqrt{1+(\omega RC)^2}. \quad (13.44)$$

The magnitude sensitivity with respect to C is

$$S_C^M = \frac{dM}{dC} = \frac{\omega^2 C^2 R^2}{C\sqrt{1+(\omega RC)^2}} = \frac{1}{C\sqrt{2}} = 0.071/C, \quad \text{for } \omega = \frac{1}{RC}. \quad (13.45)$$

Phase can be measured under controlled, low-noise conditions to millidegrees. Thus,

$$\Delta\phi = S_C^\phi \Delta C = 0.001^\circ = \left(\frac{0.5}{C_o} \right) \Delta C. \quad (13.46)$$

Assume $C_o = 100$ pF, then a ΔC of ca. 0.1 pF could be resolved by direct phase measurement; this is poor resolution. However, it is easier to measure voltage with accuracy than it is phase. For example, if we can resolve 1 μ V and $V_s = 10$ V,

$$V_o = \Delta M |V_s| = S_C^M |V_s| \Delta C = 10^{-6} V = \frac{0.7071}{C_o} \times 10 \Delta C. \quad (13.47)$$

Thus, the resolvable $\Delta C \cong 1.4 \times 10^{-17}$ F or 14 attofarads (aF). This is a useful range of resolution, which in fact can be increased by several orders of magnitude by the use of lock-in amplification. In practice, this resolution will be several orders of magnitude less because of system noise and drift.

13.5.2 Self-Balancing Circuits Used to Measure $\Delta C/C_o$

Figure 13.19 illustrates a self-nulling De Sauty bridge in which a small change in C_2 unbalances the bridge. We can write $C_2 = C_o + \Delta C$. The unbalance is sensed by a PSR. The DC error voltage, V_e , is integrated and added to a DC level, $-V_B$, which maintains static bridge balance. $V_C = V_B - V_i$ is the input to a VCCS with transconductance G_M S. The VCCS output current, I_{LED} , causes an LED to emit light on a photoconductor, causing its light-controlled conductance, G_P , to increase. In general, $G_{PC} = G_P + G_D$. G_D is the photoconductor's fixed dark conductance. Thus, the photoconductor's conductance is changed by V_i , so the bridge rebalances to compensate for the ΔC . The integrator in the feedback loop insures that $V_e \rightarrow 0$ in the SS, following a step change in ΔC . This means that $V_o = 0$:

$$V_o = \frac{1/(G_P + G_D)}{1/(G_P + G_D) + R_4} - \frac{C_1}{C_1 + C_2} = 0. \quad (13.48)$$

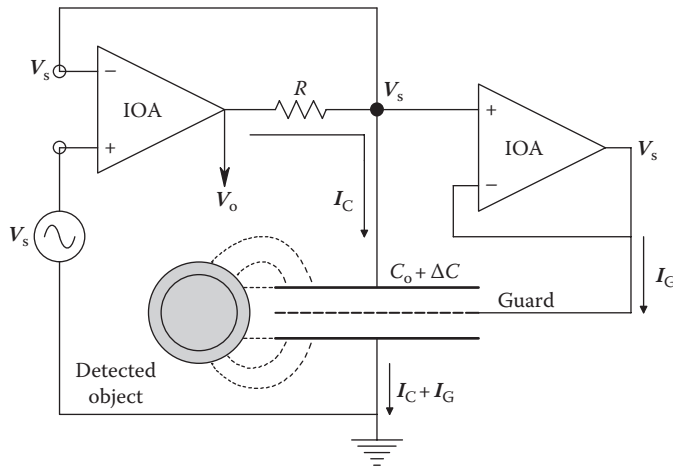


FIGURE 13.19 A self-nulling De Sauty bridge, from a design by McConnell (1993). V_i can be shown to be proportional to ΔC .

From Equation 13.48, we see that null occurs for

$$C_2 = C_1 R_4 G_{PC} \tag{13.49}$$

or when

$$C_o + \Delta C = C_1 R_4 (G_P + G_D). \tag{13.50}$$

The relation between the photoconductance G_P and V_C can be described by the following:

$$I_{LED} = G_M V_C. \tag{13.51}$$

$$P_i = K_{LED} G_M V_C \quad (\text{optical power, watts}). \tag{13.52}$$

$$G_P = K_{PC} P_i = K_{PC} K_{LED} G_M V_C = K_P V_C = K_P (V_B - V_i). \tag{13.53}$$

Substituting Equation 13.53 into Equation 13.50, we find

$$C_o + \Delta C = C_1 R_4 [K_P (V_B - V_i) + G_D]. \tag{13.54}$$

Initial balance is accomplished by setting V_B :

$$V_B = \frac{C_o - C_1 R_4 G_D}{C_1 R_4 K_P}. \tag{13.55}$$

(Note that V_B can be zero if $C_o = C_1 R_4 G_D$.) Incremental changes in C_2 are nulled by the integrated error output, V_i :

$$V_i = \frac{-\Delta C}{C_1 R_4 K_P}. \tag{13.56}$$

Note that V_i can have either sign, depending on ΔC . Self-balancing bridges can be quite sensitive, but as we remarked previously, unsymmetrical stray capacitances to ground from the bridge elements can prevent a perfect null of V_o in phase with V_s . Stray capacitances will give V_o a quadrature component.

Another approach to self-nulling detection of ΔC is shown in Figure 13.20. This system, devised by the author, is a parametric feedback system (Northrop 2000); it achieves an automatic null by keeping the RMS AC through the capacitor constant by varying the frequency of the AC voltage source driving the capacitor. A VPC variable-frequency oscillator is used instead of the customary VCO in which output frequency is proportional to the DC driving voltage, V_C . In a VPC, the output period is proportional to V_C . That is, $\tau = (K_P V_C + d) = 1/f$. The VPC is used to linearize the system's output, V_i . The integrator in the feedback loop forces $V_e \rightarrow 0$ in the SS. Thus, the DC outputs of the matched true-RMS converter chips, $\underline{V_C}$ and $\underline{V_o}$, are equal. From inspection of Figure 13.20,

$$I_c = V_o j\omega(C_o + \Delta C), \tag{13.57}$$

$$V_c = V_o j\omega(C_o + \Delta C) R_F, \tag{13.58}$$

$$\underline{V_C} = \frac{V_o}{\sqrt{2}} \omega(C_o + \Delta C) R_F, \tag{13.59}$$

$$\underline{V_C} = \frac{V_o}{\sqrt{2}}. \tag{13.60}$$

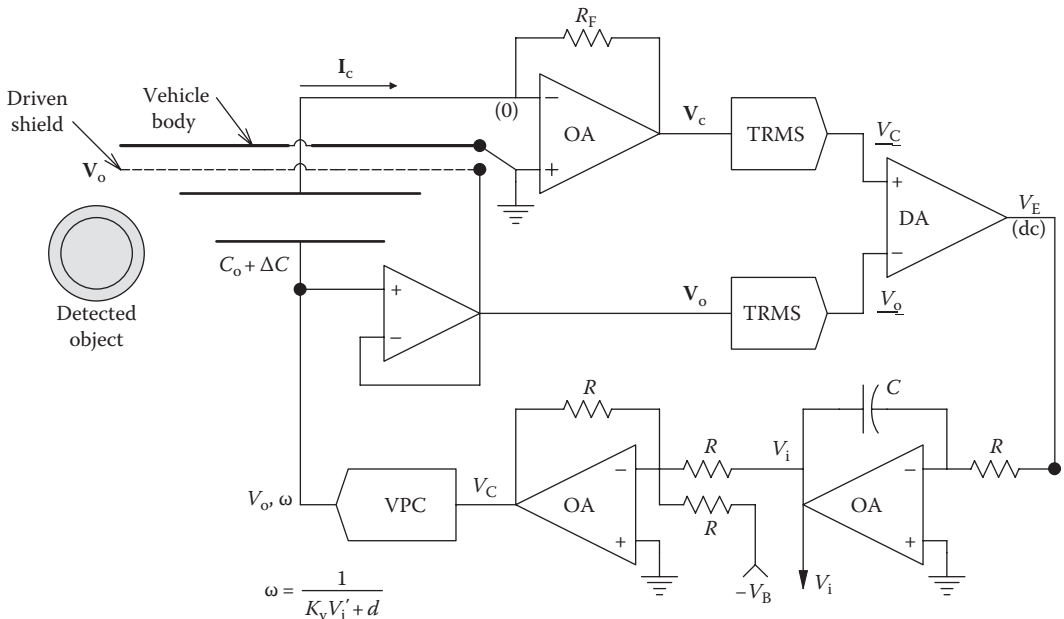


FIGURE 13.20 A self-nulling, ΔC -sensing system devised by the author. V_i is proportional to ΔC .

Now at null, in the SS,

$$\underline{V}_c = \frac{V_o}{\sqrt{2}} \omega R_F C_o (1 + \Delta C / C_o) = \underline{V}_o = \frac{V_o}{\sqrt{2}} \quad (13.61)$$

$$\downarrow$$

$$\omega R_F C_o (1 + \Delta C / C_o) = 1 \quad (13.62)$$

$$\downarrow$$

$$R_F C_o (1 + \Delta C / C_o) = \omega^{-1} = K_P (V_B - V_i) + d. \quad (13.63)$$

For SS null with $\Delta C \rightarrow 0$, V_B is set so

$$V_B = C_o R_F / K_P - d / K_P. \quad (13.64)$$

Again, the integrator output, V_i , nulls the ΔC component. Hence,

$$V_i = -\frac{R_F C_o}{K_P} \frac{\Delta C}{C_o}. \quad (13.65)$$

Using *typical* numbers, we can find the VPC system's sensitivity. Assume a $V_i = 1 \mu\text{V}$ can be detected reliably, also $C_o = 100 \text{ pF}$, $R_F = 10^7 \Omega$, and $K_P = 2 \times 10^{-5} \text{ s/V}$. Thus,

$$10^{-6} = \frac{10^7 \times 10^{-10}}{2 \times 10^{-5}} \frac{\Delta C}{C_o} \quad (13.66)$$

$$\downarrow$$

$$\frac{\Delta C}{C_o} = 2 \times 10^{-8}. \quad (13.67)$$

Because the frequency changes with ΔC , it is instructive to examine the *frequency sensitivity*, $S_{\Delta C}^\omega$, with capacitance. Note that

$$\omega = \frac{1}{K_P V_C + d}. \quad (13.68)$$

At null,

$$\omega = \frac{1}{K_P [(C_o R_F / K_P - d / K_P) - R_F \Delta C / K_P] + d} = \frac{1}{R_F (C_o - \Delta C)} \text{ r/s}. \quad (13.69)$$

Now

$$S_{\Delta C}^\omega = \frac{d\omega}{d\Delta C} = \frac{-1}{R_F (C_o - \Delta C)^2}. \quad (13.70)$$

Now the fractional change in frequency, $\Delta\omega$, is simply

$$\Delta\omega = S_{\Delta C}^{\omega} \Delta C. \quad (13.71)$$

Let the detectable $\Delta\omega = 10^{-2}$ r/s out of $\omega_o = 10^4$ r/s. Again, let $R_F = 10^7 \Omega$ and $C_o = 10^{-10}$ F. From Equation 13.71, we find the detectable $\Delta C = 10^{-15}$ F or 1 fF. This amounts to 10 ppm of C_o . Thus, the frequency sensitivity is not as great as the V_i sensitivity (0.02 ppm).

13.5.3 Summary

As we observed in the introduction to this section, capacitance sensors have found diverse applications for a number of years. In 1981, Stahovec was issued US Patent # 4,300,116 for *Safety Method and Apparatus for Sensing the Presence of Individuals Adjacent [to] a Vehicle*. Stahovec's system was designed for school bus safety. It was based on the detection of the unbalance voltage of a variant of the De Sauty bridge. His sensing capacitors employed a *driven guard shield* to minimize C_o while maximizing ΔC . Before stopping the bus, the bridge was servo balanced; it was then left open loop to detect a child near its capacitor electrodes while the bus was halted. Balance was effected using a photoconductor and an LED, similar to the system shown in Figure 13.19.

NASA has described several capacitive proximity sensors for use in space and industrial applications. One such system was described by Vranish (1992a,b, 1994) in a Technical Support Package (TSP) from *NASA Tech Briefs*. Vranish's system uses a *driven capacitor shield* and senses ΔC by the changes the object causes in the frequency of an oscillator connected to the capacitor sensor about a ca. 20 kHz carrier. Apparently, no servo- or autobalancing strategy was used in Vranish's system. McConnell (1993) also described a capacitive proximity sensor electrode system with improved object sensitivity that used *two* driven shields, one with unity gain and the other with a gain that can be set over a range of 0.5–2.0. An embodiment of McConnell's shielding system is shown in Figure 13.19. In McConnell's TSP, he shows an improved sensitivity for a gain of 1.3. The optimum gain will, of course, depend on the actual electrode and shield geometry used.

Use of a capacitive sensor system to sense land mines turns out to have many problems. The work in the author's laboratory has shown that the air gap between the sensing capacitor and the ground must be constant as the sensor is scanned over the ground surface (i.e., the ground must be perfectly flat), the soil must be homogeneous (no large rocks), and the mine must have a significantly different dielectric constant than the soil around it. Soil moisture will also affect measurements; pure water has a dielectric constant of $\epsilon = 78.5$ at 25°C.

13.6 Chapter Summary

In this chapter, we have described the designs of four novel measurement systems: The first is a sensitive microdegree polarimeter used to measure the concentration of optically active D-glucose-glucose used in bioreactors to grow cells. This system uses a self-nulling design in which the water solvent in the bioreactor's growth medium is used as a Faraday medium to linearly modulate the LPL used in the instrument. The use of the water solvent for modulation and feedback is novel, because normally an expensive glass-core FR would be used.

The second system was designed to measure and locate PDs caused by insulation faults on coaxial, HV power cables. These measurements are important to power utilities companies because they indicate the sites where catastrophic insulation failures can and will occur and where new cable sections should be spliced in to avoid cable failures.

The third example of instrument system design is a closed-loop, constant-phase, pulsed LAVERA system. It has novel design in its use of a VPC-controlled oscillator. The same system architecture can be extended to ultrasonic ranging systems, using either pulsed or CW ultrasound. The LAVERA system provides simultaneous target range and velocity output information.

The fourth example illustrates the use of feedback to automatically null systems used to measure very small changes in object capacitance ($\Delta C/C_0$). In one example, a self-balancing bridge run at constant frequency is compared to a novel feedback system in which the current through the capacitive sensor was kept constant by varying the operating frequency by a VPC.

The accuracy of all four systems is seen to be limited by noise. In all cases, low-noise amplification must be used. Further improvement of system performance in noise was achieved in the PD measurement system by implementing digital notch filters to remove coherent interference, using cross correlation to align noisy PD waveforms for ensemble averaging and other DSP algorithms for PD site location. The effects of noise in the LAVERA system are minimized by using a high-power laser, low-noise preamplifiers and OAs, and fast logic.

Glossary

Acronyms and Abbreviations

AC	Alternating current (<i>adj. and n.</i>)
ADC	Analog-to-digital converter
AF	Active filter
ANN	Artificial neural network
APD	Avalanche photodiode
ATR	Attenuated total reflection
B	Magnetic flux density in tesla (1 tesla = 10^4 gauss = 10^4 maxwells/cm ² = 1 W/m ²)
BJT	Bipolar junction transistor
BPF	Band-pass filter
<i>ca.</i>	(Latin: <i>circa</i>) About
CCD	Charge-coupled device (or detector)
CCO	Current-controlled oscillator
CCVS	Current-controlled voltage source
CCW	Counterclockwise
CFOA	Current feedback op amp
CGS	Centimeter–gram–second units of measurement
CM	Common mode
CMRR	Common-mode rejection ratio
COLD	Chronic obstructive lung disease
CPF	Conductive polymer film
CRT	Cathode ray tube
CSOA	Chopper-stabilized op amp
CUT	Circuit under test
CVD	Chemical vapor deposition
CW	Continuous wave (as opposed to pulsed)
<i>D</i>	Capacitor dissipation factor
DA	Differential (or difference) amplifier
DC	Direct current (<i>adj. and n.</i>), constant
d-c	Direct coupled (amplifier)
DCE	Data communications equipment (in an RS-232C system)
DDS	Dual-differential subtractor
DIP	Dual, in-line package
DM	Difference mode
DRFP	Dynamic range, floating point (ADC)
DSBSCM	Double-sideband, suppressed-carrier modulation
DTE	Data terminal equipment
DUT	Device under test
ECL	Emitter-coupled logic

E/M	Excitation/measurement end (of a cable)
EM	Electromagnetic (<i>adj.</i>)
EMF	Electromotive force (in volts)
EMI	Electromagnetic interference (noise)
EMR	Electromagnetic radiation
EOA	Electrometer op amp
EPROM	Electronically-programmed read-only memory
eV	Electron volt
FDNR	Frequency-dependent negative resistance
FEA	Fractal element antenna
FET	Field-effect transistor
FO	Fiber optic
FOC	Fiber-optic cable
FOM	Fiber-optic mechanosensor
F-P	Fabry–Pérot (interferometer)
<i>g</i>	The mean acceleration of Earth's gravity: $1\ g = 9.8\ \text{m/s}^2$. Also, <i>g</i> is the abbreviation for gram mass
GBWP	Gain \times bandwidth product (of an amplifier). Also, f_T
GERD	Gastroesophageal reflux disease
GIC	Generalized impedance converter (circuit)
GMT	Geiger–Mueller tube
GNSS	Global navigation satellite system
GPa	Gigapascal (SI pressure unit) = $10^9\ \text{Pa} = 10^9\ \text{N/m}^2$. ($1\ \text{GPa} = 10^8\ \text{dyn/cm}^2$)
HV	High voltage
IA	Instrumentation amplifier
IC	Integrated circuit
ICU	Intensive care unit
IGFET	Insulated gate field-effect transistor
INT(*)	Integral value of the number, (*)
IOA	Ideal op amp
ISFET	Ion-sensitive field-effect transistor
JFET	Junction field-effect transistor
JJ	Josephson junction
KCL	Kirchoff's current law
KVL	Kirchoff's voltage law
LAD	Laser diode
LAVERA	Laser velocity and ranging (system)
LDV	Laser Doppler velocimetry
LE	Limiting error
LIA	Lock-in amplifier
LMS	Least mean square
LN ₂	Liquid nitrogen
LPF	Low-pass filter
LPL	Linearly polarized light
LVDT	Linear variable differential transformer
m	Maxwell (unit of magnetic flux)
MDA	Minimum detectable amount
MEM	(<i>adj.</i>) Microelectromechanical
MEMS	(<i>n.</i>) Microelectromechanical system(s)

MKS	Meter–kilogram–second measurement system
MOCS	Magneto-optic current sensor
MOS	Metal oxide semiconductor
MOSFET	Metal oxide semiconductor field-effect transistor
MOX	Metal oxide
MS, ms	Mean squared
M/T	Modulator/transmitter
N_A	Avogadro's number: $6.022\,137 \pm 0.000\,007 \times 10^{23} \text{ mol}^{-1}$
NBPF	Narrow band-pass filter
ND	(<i>adj.</i>) Neutral density
NDS	Nondispersive spectrophotometry
NDW	Neutral density wedge
NEMS	Nanoelectromechanical system
NTC	Negative temperature coefficient (resistance)
OA	Optically active
ODE	Ordinary differential equation
OF	Optical fiber
OR	Operating room
Pa	Pascal (N/m^2), SI pressure unit (1 kPa = 0.14504 psi)
PC	Photoconductor; personal computer; printed circuit
PDF	Probability density function
PDS	Power density spectrum
PLL	Phase-lock loop
PMT	Photomultiplier tube
POLFET	Polymer-covered gate field-effect transistor
PPOF	Polarization-preserving optical fiber
PSR	Phase-sensitive rectifier
PVDF	Polyvinylidene fluoride
PyM	Pyroelectric material
QUM	Quantity under measurement
R-C	Resistance coupled (amplifier)
RFI	Radio-frequency interference
RFID	Radio-frequency identification (transponder tags)
RH	Relative humidity
RIG	Rate-integrating gyro
RLC	(<i>adj.</i>) Resistance, inductance, and capacitance
RMS (or rms)	Root mean square
RR	Recovery room
RTD	Resistance temperature detector
SAM	Self-assembling monolayer
SBSL	Single-bubble sonoluminescence
SCR	Silicon-controlled rectifier
SI	(Abbreviated from French: <i>Système International D'Unites.</i>) Modern form of the metric system adapted by international treaty in 1960
SPDT	Single pole, double throw (switch)
SPR	Surface plasmon resonance
SRAM	Static read-only memory
SS	Steady state
SV	Satellite vehicle

TCXO	Temperature-controlled crystal oscillator
TEG	Thermoelectric generator (thermopile)
TM	Transverse magnetic (EM wave propagation)
TNAE	Total noise equivalent acceleration (of an MEM accelerometer)
TRMS	True root mean square
TSP	Technical support package (NASA)
TTL	Transistor-transistor logic
UHF	Ultrahigh frequencies
USD	United States' dollars
UTI	Urinary tract infection
VCCS	Voltage-controlled current source
VCO	Voltage-controlled oscillator
VCVS	Voltage-controlled voltage source
VHF	Very high frequencies
VI	Virtual instrument
VMRG	Vibrating mass rate gyro
VOC	Volatile organic compound
VPC	Voltage-to-period converter
Wb	Weber, SI unit for magnetic flux (1 Wb = 10^8 maxwells)
WDX	Wireless data transmission
WPM	Wireless patient monitoring
WS	Wireless sensor
[X]	Concentration of a substance, X
XLPE	Cross-linked polyethylene (insulation)
Y	Young's (stretch) modulus (SI units in Pa)

Glossary of Terms

Amperometry: (In the electrochemical detection of substances.) A constant potential voltage (V) is applied between two electrodes, and the change in current in time due to the analyte is monitored.

Aprotic solvents (liquids): Solvents that will dissolve many salts but lack an acidic hydrogen (proton). Characteristically, they can accept hydrogen bonds, do not have acidic hydrogen centers, and dissolve organic salts such as tetraethylammonium iodide. Examples of polar aprotic solvents include ethyl acetate, acetone, acetonitrile, dioxane, and dimethyl sulfoxide.

ARGOS satellite telemetry system: Used principally for wildlife tracking. Data from mobile transmitter platforms (tags) are uplinked in the 401.618–401.680 MHz band to nongeostationary, low polar-orbiting, environmental satellites (POES). These satellites are put in orbit and maintained by NOAA and MetOp of the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). Orbital altitude is 850 km; orbital period is *ca.* 100 min. ARGOS received messages are (1) stored on the onboard recorder and then retransmitted to the ground each time the SV passes over the main receiving stations located at Wallops Island, VA; Fairbanks, AK; and Svalbard, Norway. (2) The messages can also be retransmitted to other regional reception stations in range. At any given moment, each satellite

in the system sees all ARGOS transmitters within a circle on the Earth's surface with *ca.* 5000 km diameter. As a satellite orbits, the visibility circle sweeps a 5000 km wide swath around the Earth, covering both poles. Due to the Earth's rotation, the swath shifts 25° west (2800 km at the equator) around the polar axis at each revolution. This causes an overlap between successive swaths. Since overlap increases with latitude, the number of daily passes over a given transmitter or receiver also increases with their latitude. At the poles, each SV sees each polar transmitter on every pass, approximately 14 times per day.

The ARGOS system is widely used to track marine animals tagged with pop-up archival tags (PAT), platform terminal transmitters (PTT), or GPS data-logging tags.

Avogadro's number (N_A): N_A is the number of particles (atoms or molecules) in one mole (gram molecular weight) of an isotopically pure substance. Its value accepted by the US NIST is $6.022\,141\,5 \pm 0.000\,001\,0 \times 10^{23} \text{ mol}^{-1}$ (Fox and Hill 2007). Fox (2012) calculated N_A to be $6.022\,141\,089\,796\,636\,994\,702\,80 \times 10^{23} \text{ mol}^{-1}$ for carbon 12.

Conductometry: (In electrochemical and gas-phase sensor analysis.) The conductance between the electrodes of an electrochemical cell or an MOX gas sensor is defined as $G = I/V$ siemens. That is, G is calculated from the measured I , given an applied V . V can be varied.

Coriolis acceleration: In the simplest case, a mass M with vector velocity v lies in the XY plane, which is rotating with angular velocity Ω r/s. The moving mass will be acted on by a *Coriolis acceleration* a_c due to the rotation of the mass at ω r/s around a rotation vector Ω having an axis perpendicular to the XY plane of rotation and a magnitude equal to ω , the angular velocity of the rotating system. The vectors v and Ω are separated in general by an angle θ . In this case, $\theta = 90^\circ$. Thus, in general, in an x, y, z coordinate system, we can write the vector cross product for the Coriolis vector acceleration:

$$\mathbf{a}_c = 2v \times \Omega = -2\Omega \times v.$$

In more general terms, the vector v is rotated through θ into Ω , and the direction of the a_c vector is \perp to the plane containing v and Ω and lies along the y -axis. Its magnitude is given by $a_c = 2v\Omega \sin \theta$. θ is the angle between Ω and v .

Coriolis force: A mass M moving in a linear path with some vector velocity v on a platform rotating with angular velocity Ω is acted on by a Coriolis force, which can be shown to be given by the vector cross product:

$$\mathbf{F}_c = 2Mv \times \Omega = -2M\Omega \times v$$

where M is the mass of the moving element and Ω is the angular velocity vector, which has a magnitude equal to the rotation rate ω of the platform and is directed perpendicular to the plane of rotation by the right-hand screw rule. (It is colinear with the axis of rotation of the rotating platform.) The \times symbol represents the *vector cross product* operator. The vector cross product between two vectors, \mathbf{A} and \mathbf{B} , separated by an angle θ is a vector $\mathbf{C} = \mathbf{A} \times \mathbf{B}$, whose magnitude is $C = AB \sin \theta$ and whose direction is perpendicular to the plane containing \mathbf{A} and \mathbf{B} and is directed by a right-hand screw rotated in the direction of \mathbf{A} rotated into fixed \mathbf{B} .

On the surface of a sphere rotating at ω r/s, the object (mass M) velocity v can be resolved into three vector components, v_e , v_n , and v_u , where v_e is the east vector component, v_n is the north component, and v_u is the vertical (upward) component. The velocity vectors v_n and v_e are tangential to the Earth's surface; v_u is

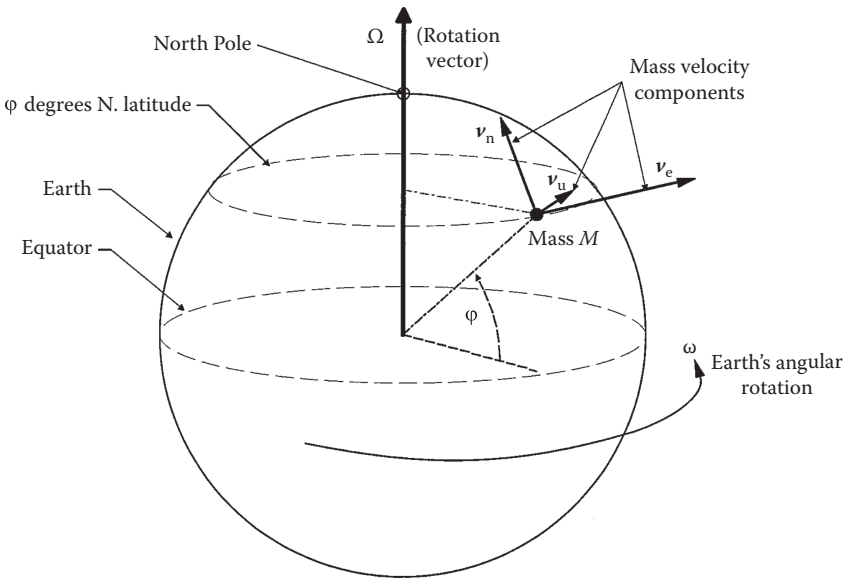


FIGURE GL.1
Diagram of the rotating Earth, showing mass velocity vectors.

perpendicular to it. All angles and vectors in Figure GL.1 are shown as positive. ϕ is the latitude angle (measured from the equator). On the Earth, Ω lies along the North Pole (spin axis). Thus,

$$\Omega \equiv \omega \begin{pmatrix} 0 \\ \cos \phi \\ \sin \phi \end{pmatrix} \mathbf{k} \text{ and } \mathbf{v} = \begin{pmatrix} v_e \\ v_n \\ v_u \end{pmatrix}. \text{ Thus } |\mathbf{F}_c| = |-2M\Omega \times \mathbf{v}| = 2\omega M \begin{pmatrix} v_n \sin \phi - v_u \cos \phi \\ -v_e \sin \phi \\ v_e \cos \phi \end{pmatrix}.$$

When considering the Coriolis effect on winds and water, v_u is small, and the vertical component of the Coriolis force is negligible compared to gravity. Thus, only the horizontal velocity components matter, and we can write

$$\mathbf{v} = \begin{pmatrix} v_e \\ v_n \end{pmatrix} \text{ and } \mathbf{F}_c = M \begin{pmatrix} v_n \\ -v_e \end{pmatrix} (2\omega \sin \phi).$$

From the preceding vector matrix equations, we can see that if $v_n = 0$ and ω and $\phi > 0$, a mass point moving due east experiences a Coriolis acceleration forcing it due south. Similarly, setting $v_e = 0$, a mass element moving due north is acted on by a Coriolis force moving it due east.

Cramer's rule: (For solving simultaneous, linear algebraic equations.) For example, assume an $N = 3$ system. The linear equations are

$$\begin{aligned} \delta R_1/R_{01} &= \sigma_{11}C_1 + \sigma_{12}C_2 + \sigma_{13}C_3 \\ \delta R_2/R_{02} &= \sigma_{21}C_1 + \sigma_{22}C_2 + \sigma_{23}C_3 \\ \delta R_3/R_{03} &= \sigma_{31}C_1 + \sigma_{32}C_2 + \sigma_{33}C_3. \end{aligned}$$

In matrix notation,

$$\delta\mathbf{R}/\mathbf{R}_o = \mathbf{C}[\boldsymbol{\sigma}],$$

where $[\boldsymbol{\sigma}]$ is the 3×3 \mathbf{A} matrix, \mathbf{C} is a 1×3 column matrix, etc.

To solve for C_1 , C_2 , and C_3 , given $\delta\mathbf{R}/\mathbf{R}_o$ and $[\boldsymbol{\sigma}]$,

$$C_1 = \frac{\begin{bmatrix} \delta R_1/R_{o1} & \sigma_{12} & \sigma_{13} \\ \delta R_2/R_{o2} & \sigma_{22} & \sigma_{23} \\ \delta R_3/R_{o3} & \sigma_{32} & \sigma_{33} \end{bmatrix}}{\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}}, C_2 = \frac{\begin{bmatrix} \sigma_{11} & \delta R_1/R_{o1} & \sigma_{13} \\ \sigma_{21} & \delta R_2/R_{o2} & \sigma_{23} \\ \sigma_{31} & \delta R_3/R_{o3} & \sigma_{33} \end{bmatrix}}{\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}}, \text{ etc.}$$

Detectability: The extent to which something is detectable. Detectability depends, among other things, on sensor system output signal-to-noise ratio. The sound, image, waveform, etc., must be larger than the background or surround spatiotemporal noise. An image may be rendered undetectable even if its amplitude is above the background noise if it is defined by patterns with specific optical wavelengths: think camouflage.

Etalon: The etalon is an essential component in a Fabry–Pérot interferometer. An etalon is an optical interferometer in which many impinging rays of light undergo multiple reflections between two partially reflecting surfaces and whose resulting optical transmission is a diffraction pattern with circular symmetry that is periodic in wavelength and appears as a sequence of high-contrast, intensity rings. There are solid and air-spaced etalons. Etalons can be used as precise wavelength references in telecommunications applications where the periodicity of the signal provides an array of reference frequencies for a telecommunications grid. An etalon can be used in a feedback loop to stabilize the wavelength of a laser.

Free spectral range: It is the spacing in optical frequency or wavelength ($\Delta\lambda$) between two successive reflected or transmitted optical intensity maxima of an interferometer of diffractive optical system.

Peak frequency response of a second-order, low-pass system: A second-order, quadratic, low-pass system has the basic frequency response function:

$$\frac{Y}{X}(j\omega) = \frac{1}{(j\omega)^2/\omega_n^2 + j\omega/(\omega_n Q) + 1},$$

where

ω is the applied radian frequency

ω_n is the system's undamped natural frequency in r/s

Q is the system's resonance quality factor = $1/(2\xi)$

ξ is the quadratic system's damping factor (see Figures 1.3 and 1.4 in Chapter 1)

Y is the output

X is the SS sinusoidal input of frequency ω

The magnitude of the frequency response can be written as

$$\left| \frac{Y}{X}(j\omega) \right| = \frac{1}{\sqrt{(1 - \omega^2/\omega_n^2)^2 + \omega^2/(\omega_n^2 Q)}}$$

By differentiating the denominator and setting the result = 0, we can show that $|Y/X|$ has a peak at $\omega_p < \omega_n$:

$$\omega_p = \omega_n \sqrt{[1 - 1/(2Q^2)]}$$

At $Q > 10$, $\omega_p \rightarrow \omega_n$.

Half-cell: (In electrochemistry.) A battery cell is made up from two half-cells. Each half-cell electrode has a standard potential: $E^0 = -\Delta F^0/n\mathfrak{F}$ volts, where \mathfrak{F} is the Faraday number, ΔF^0 is the change in free energy for the half-cell's chemical reaction when the activities of all the reactants and products are unity, and n is the number of equivalents reacting. For any half-cell reaction, ΔF^0 is constant at any given temperature and independent of pressure. For example, the standard half-cell electrode potential for the *oxidation* of zinc is +0.7618 V. The reaction is $\text{Zn(s)} = \text{Zn}^{++} + 2e^-$. *By definition*, the oxidation of hydrogen (hydrogen half-cell reference electrode) has a standard potential of $E^0 \equiv 0.0000$ V. The reaction is $\text{H}_2(\text{g}, 1 \text{ atm}) = 2\text{H}^+ + 2e^-$. The standard half-cell potential for the *reduction* of silver ions is $E^0 = +0.7991$ V. The reaction is $\text{Ag}^+ + e^- = \text{Ag(s)}$. (This is silver plating.)

Magnetization: The process of making a material temporarily or permanently magnetic. That is, an applied *magnetic intensity*, H (amps/m), causes a *magnetic flux density*, B (w/m²), to exist in the material. In general, $B = \mu H$, where μ is the material's *permeability* in webers/(amp-m). μ is a function of the material magnetized, the mechanical strain on the material (*Villari effect*), temperature, H , and dH/dt , among other variables.

Nernst chemical EMF: (Half-cell.) It can be shown (Maron and Prutton 1958) that the electrical potential of a typical chemical half-cell is

$$E = E_0 - \frac{RT}{nF} \ln \left(\frac{1}{a_B^n} \right) = E_0 - \frac{RT}{F} \ln \left(\frac{1}{a_B} \right) \text{V,}$$

where a typical half-cell reaction is $2M(\text{s}) + nB^- \longleftrightarrow nC\uparrow + kD + E(\text{s})$; T is the Kelvin temperature; F is the Faraday number, 96,500 abs. coulombs; the gas constant $R = 8.315$ joules/(K g-mol); n is the integer number of equivalents of B reacting; E_0 is the *standard potential* of the half-cell; and a_B is the activity of the B ion. k is an integer constant; M is a metal reactant; C, D, and E are reaction products; and (s) denotes an insoluble substance.

Optically active material: A liquid, liquid solution, or solid that has the property of rotating the plane of the E-vector of a transmitted beam of LPL. In a solution, optical rotation generally follows the law:

$\phi = CL[\alpha]_\lambda^T$, where C is the concentration of the optically active solute, L is the optical path length, and $[\alpha]_\lambda^T$ is the *specific optical rotation* of the optically active substance in degrees/(m path length \times molar concentration at Kelvin temperature T

and wavelength λ in nm). (Note: Other units of rotation ϕ , path length L , concentration unit C , temperature T , and wavelength λ are often used.) For dilute solutions of N optically active solutes, the net optical rotation may be given by

$$\phi = \sum_{k=1}^N [\alpha_k]_{\lambda}^T C_k L.$$

For example, using the sodium D line as a light source ($\lambda = 5893$ nm), an aqueous solution of D-glucose has an $[\alpha] = 52.5 + 0.025d$ deg/dm at 20°C, where $d = \text{conc. in g/100 g solution}$ and the path length L is in decimeters (dm).

In general, the magnitude of $[\alpha]_{\lambda}^T$ increases as λ decreases.

Perveance: A high-perveance transistor or vacuum tube is a power device; that is, its collector, drain, or plate current has a substantially higher range than for a low-power, small-signal, voltage amplifier device. For example, for an FET, $I_D = P(V_{GS}^2, V_{DS})$. P is the perveance factor; its units are A/V^2 .

Poisson's ratio: In mechanics, when a rod or bar is subjected to tensile stress, it not only elongates in the (axial) direction of stress, but also its transverse dimensions decrease. For a rectangular bar of rest length l_o , and an original transverse dimension, w_o , Poisson's ratio is defined as

$$\sigma \equiv -\frac{\Delta w/w_o}{\Delta l/l_o} \text{ (dimensionless).}$$

The ratio, $\Delta w/w_o$, is the same for all transverse dimensions, except for nonisotropic materials such as certain crystals whose properties are different along different axes (such as silicon). The minus sign is introduced to keep $\sigma > 0$, since an increase in length always results in a decrease in transverse dimensions, and vice versa. $\sigma \approx 0.3$ for most metals. Thus, for a rectangular bar of length l_o , width b_o , and height c_o , the strains are

$$\frac{\Delta b}{b_o} = -\frac{\sigma \Delta l}{l_o} = \frac{\sigma F}{AY},$$

$$\frac{\Delta c}{c_o} = -\frac{\sigma \Delta l}{l_o} = \frac{\sigma F}{AY},$$

where

Y is Young's modulus for the material

F is the axial force on the bar

$$A = b_o c_o$$

It can be shown that the bar undergoes a fractional volume change due to the applied F :

$$\frac{\Delta V}{V_o} = (1 - 2\sigma) \frac{\Delta l}{l_o},$$

where $V_o = l_o b_o c_o$, $(1 - 2\sigma)$ is always > 0 , and $\Delta V > 0$ for a tensile force.

Note that for certain MEMS materials, σ has the following values: Si, 0.23; polysilicon, 0.23; polyimide, 0.42 (Cheung 2010).

Potentiometry: (In electrochemical analysis of substances.) The open-circuit voltage (EMF) is measured across a pair of electrodes (a cell) as a function of time, given a step input of analyte. Redox reactions generally occur in the cell.

Q: Quality factor of a tuned or resonant electronic or mechanical system having a frequency response function $\frac{V_o}{V_i}(j\omega)$.

The Q is defined as $Q = f_{pk}/\Delta f$, where f_{pk} is the Hz frequency, $|V_o/V_i|$ is maximum, and Δf is the Hz frequency difference between the two frequencies, f_h and f_l , where $|V_o/V_i| = 0.707 |V_o/V_i|_{max}$. It is a measure of the *sharpness* of tuning or resonance. $\Delta f = f_h - f_l$ is called the half-power bandwidth. In general, in a quadratic low-pass system,

$$\frac{X_o}{X_i}(j\omega) = \frac{K\omega_n^2}{[(j\omega)^2 + (j\omega)(2\xi\omega_n) + \omega_n^2]} = \frac{K\omega_n^2}{[(j\omega)^2 + (j\omega)(\omega_n/Q) + \omega_n^2]}$$

$$\text{where } |X_o/X_i(0)| = K, \quad |X_o/X_i(j\omega_n)| = KQ,$$

where ξ is the damping factor, ω_n is the undamped resonant frequency (r/s) where the phase of $X_o/X_i = -90^\circ$, and $Q \equiv 1/(2\xi)$. In a second-order, tuned, resonant system, the frequency response magnitude has a marked peak around ω_n for $0 \leq \xi \leq 0.3$ (see Sect. 5.5.2 in Northrop [2012]).

Q can also be defined for an oscillatory system as equal to 2π (stored potential energy per cycle/energy dissipated per cycle) [Blom et al. 1992, Cox 2011]. The Q of a vibrating cantilever sensor is reduced by coating the cantilever surface with thin, nonelastic coatings that adsorb analytes.

Q is also used as a measure of inductor purity. In the simple series R - L model for an inductor at a low frequency, Q is defined in this context as $Q \equiv X_L/R_L = \omega L/R_L$. Ideally, $R_L \rightarrow 0$ (a lossless inductor).

Relative humidity: The RH of an air/water system is defined as the ratio of the *partial pressure of water vapor* in the system (P_w) to the *saturated vapor pressure* of water (P_{ws}) over a pure air/water system at a specified temperature. That is, $RH(T) = 100 \times P_w/P_{ws}$. For example, $P_{ws}(25^\circ\text{C})$ is known to be 23.756 mmHg. At 0°C , $P_{ws} = 4.579$ mmHg.

Thermal (Johnson) noise from resistors: Broadband, Gaussian white noise from resistors. Thermal noise has a theoretical, one-sided power density spectrum of $e_N(f) = 4kTR$ MS V/Hz, where T is the Kelvin temperature of the resistor R and k is Boltzmann constant = 1.38062×10^{-23} J/K.

Total noise equivalent acceleration (TNEA): A conventional mass-spring-dashpot accelerometer is subject to noise from thermal motion of the proof mass, M . Note that a small mass, as found in MEM accelerometers, will have a higher RMS acceleration noise than from a larger macro system. The RMS TNEA can be shown to be given by (Gabrielson 1993)

$$\text{TNAE}_{\text{rms}} = \sqrt{\frac{a_n^2}{\Delta f}} = \frac{\sqrt{4kTD}}{M} = \sqrt{\frac{4kT\omega_r}{QM}} \sqrt{\text{m/s}^2/\text{Hz}}$$

(noise equivalent acceleration spectral density),

where

Q is the mechanical resonance quality factor

ω_r is the resonance frequency of the proof mass-spring-dashpot inertial accelerometer system

M is the mass

D is dashpot damping constant in newtons/(m/s)

k is Boltzmann constant

T is the Kelvin temperature

Δf is the Hz noise bandwidth.

Type 1 feedback system: A simple, general, single-loop, linear, negative feedback system has a *loop gain* given by

$$A_L(s) = \frac{-H(s)G(s)}{s^p},$$

where $X(s)$ is the system input and $Y(s)$ is the feedback system output. $E(s)$ is the system error = $X(s) - Z(s)$. $H(s)$ is the forward gain = $Y(s)/E(s)$, $G(s)$ is the feedback gain = $Z(s)/Y(s)$, s^p is a p th-order pole (integrator) in the feedback loop, generally in series with $G(s)$. $p = 0, 1, 2, \dots$ If the type exponent $p = 0$, the system is said to be *type zero*; when $p = 1$, the system is *type 1*; and $p = 2$ gives a *type 2 system*.

Using the Laplace final value theorem, a *type 1 system* can be shown to have zero SS DC error, given a step input, $x(t) = X_0 U(t)$, at $t = 0$, but has a finite S-S DC error, $R/[G(0)H(0)]$, when the input $x(t)$ is a ramp, $x(t) = Rt$, $t \geq 0$. A *type 0 system* has an SS DC error of $e_{ss} = X_0/[1 + G(0)H(0)]$ for a step input of height X_0 and an infinite SS error for a ramp input.

Vector cross product: A vector cross product is used to describe *Coriolis acceleration*, \mathbf{a}_c , and certain electromagnetic phenomena:

$$\mathbf{a}_c = 2(\mathbf{v} \times \boldsymbol{\Omega}) = -2(\boldsymbol{\Omega} \times \mathbf{v}).$$

The magnitude of the acceleration is given by $a_c = 2v \Omega \sin \theta$. Figure GL.2 illustrates two cases of finding \mathbf{a}_c . In **A**, the vectors are orthogonal, and the platform is the XY plane, which is rotating CCW with velocity ω r/s. The angular velocity vector, $\boldsymbol{\Omega}$, is found by multiplying a unit vector \mathbf{k} , which lies along the z -axis (perpendicular to the XY plane), by ω . The direction of \mathbf{k} is given by the RH screw rule, that is, it is the direction a normal wood screw would advance is rotated if turned CCW, *viewed from above the XY plane*. (Note that this is CW rotation if viewed from the bottom [slot end]). Now, if we rotate the mass velocity vector \mathbf{v} into $\boldsymbol{\Omega}$ through angle θ , an RH screw would advance along the positive y -axis, giving the direction of \mathbf{a}_c shown. In Figure GL.2B, the vector \mathbf{v} does not lie in the XY plane. When it is rotated into $\boldsymbol{\Omega}$, the RH screw rule points \mathbf{a}_c in a direction perpendicular to the plane formed by \mathbf{v} and $\boldsymbol{\Omega}$. In both cases, the magnitude of \mathbf{a}_c is given by $a_c = 2 v \Omega \sin \theta$.

Some general vector cross product identities are

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A}$$

$$\mathbf{A} \times (\mathbf{B} + \mathbf{C}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C}$$

$$(k\mathbf{A}) \times \mathbf{B} = k(\mathbf{A} \times \mathbf{B})$$

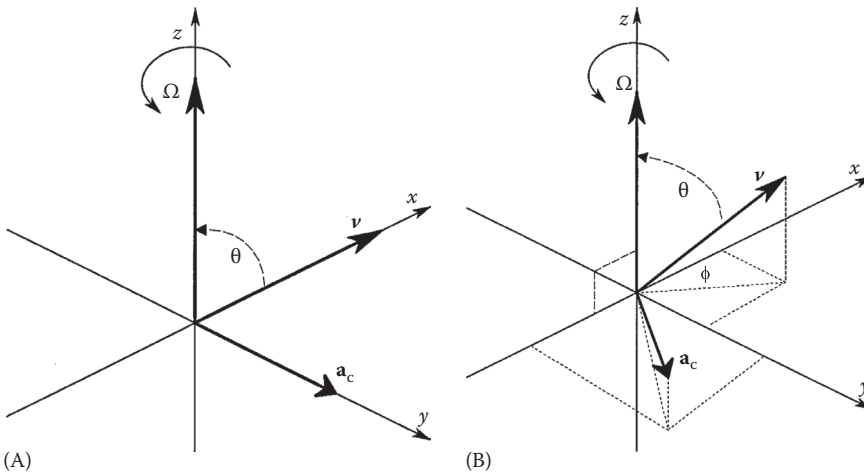


FIGURE GL.2

(A) Orthogonal vectors illustrating vector cross product. Ω is the angular velocity vector. (B) Nonorthogonal vectors' ν and \mathbf{a}_c cross product.

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \det(\mathbf{ABC})$$

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B})$$

$$(\mathbf{A} \times \mathbf{B}) \times (\mathbf{C} \times \mathbf{D}) = \det(\mathbf{ABD})\mathbf{C} - \det(\mathbf{ABC})\mathbf{D}$$

Consider three orthogonal unit vectors \mathbf{i}_x , \mathbf{j}_y , and \mathbf{k}_z :

$$\mathbf{i} \times \mathbf{j} = \mathbf{k} = -\mathbf{j} \times \mathbf{i}$$

$$\mathbf{j} \times \mathbf{k} = \mathbf{i} = -\mathbf{k} \times \mathbf{j}$$

$$\mathbf{k} \times \mathbf{i} = \mathbf{j} = -\mathbf{i} \times \mathbf{k}.$$

Also

$$\mathbf{i} \times \mathbf{i} = \mathbf{j} \times \mathbf{j} = \mathbf{k} \times \mathbf{k} = \mathbf{0}$$

because $\sin 0 = 0$.

Three-dimensional cross products can also be written in matrix notation. Representing the two vectors as sums of their orthogonal components,

$$\mathbf{A} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$$

$$\mathbf{B} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}.$$

We can write

$$\mathbf{A} \times \mathbf{B} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \mathbf{C}.$$

Villari effect: (Also known as the *inverse magnetostrictive effect*.) Magnetostriction characterizes the shape change of a ferromagnetic material (e.g., iron, cobalt) during *magnetization*. The inverse magnetostrictive (Villari) effect characterizes the

change of domain magnetization when a mechanical stress is applied to the material. This change can be *positive* (magnetization is *increased* by tension) like in pure iron or *negative* (magnetization is *decreased* by tension) like in nickel.

Wavenumber: The wavenumber ν of an optical (light) frequency f is $\nu \text{ cm}^{-1} = k \times f$, where f is in Hz and $k = 3.33 \times 10^{-12} \text{ cm}^{-1}/\text{Hz}$. Also, $\nu \text{ cm}^{-1} = 10^{-2}/\lambda$, where λ is the light wavelength in meters.

Work function: The minimum thermodynamic work (energy) required to remove an electron from a solid material to a point immediately outside the solid surface. The work function Φ is measured in electron-volts (eV). If the material has a crystal-line structure, orientation of the crystal will affect the work function; for example, the work function for Si ranges from 4.60 to 4.85 eV, and Φ for Au ranges from 5.1 to 5.47 eV. Li has a low Φ of 2.3 eV. Tungsten (W), used for vacuum tube filaments, has a Φ ranging from 4.32 to 5.22 eV. Palladium (Pd) used in Schottky barrier diode chemosensors for H_2 has Φ from 5.22 to 5.6 eV.

Young's modulus: A mechanical property of materials. The linear stretch or Young's modulus is given by

$$Y = \frac{\text{tensile or compressive stress}}{\text{tensile or compressive strain}} = \frac{F/A}{\Delta L/L_0}$$

Units : dyn/cm^2 or gigapascals ($1 \text{ GPa} = 10^8 \text{ dyn}/\text{cm}^2$).

Y for silicon can range from 130 to 188 GPa, depending on the crystal structure orientation relative to the applied stress (Si is anisotropic) (Hopcroft et al. 2010, Kaajakari 2013). Y for silicon carbide (SiC) is *ca.* 450 GPa, Y for polyimide plastic was measured at 3.91 GPa (Sasaki et al. 2011), and SU-8 photoresist polymer has $Y = 4.4 \text{ GPa}$ (ME189 2013).

References

- Acar, C. and A.M. Shkel. 2004. Structural design and experimental characterization of torsional micromachined gyroscopes with non-resonant drive mode. *J. Micromech. Microeng.* 14: 15–25.
- Agilent. 2013. Agilent Technologies GC systems homepage. Accessed May 13, 2013 at: www.chem.agilent.com/en-US/Products-Services/Instruments-Systems/Gas-Chromatography/Pages/default.aspx.
- Albert, K.J. et al. 2000. Cross-reactive chemical sensor arrays. *Chem. Rev.* 100: 2595–2626.
- AN211. 2002. Applications of magnetic position sensing, 8pp. Accessed April 12, 2013 at: <http://www51.honeywell.com/aero/common/documents/Applications-of-Magnetic-Position-Sensors.pdf>.
- Analog Devices. 2003. AD7709 Data Sheet Rev A, 03/2003. AD7709: 16-Bit Sigma Delta ADC with Current Sources, Switchable Reference Inputs and I/O Port. Accessed 22 March, 2014 at: www.analog.com/en/analog-to-digital-converters/ad-converters/ad7709/products/products.html.
- Analog Devices. 2013. AD538 Analog computational circuit. Accessed April 10, 2013 at: www.analog.com/en/special-linear-functions/analog-multipliersdividers/ad538/products/product.html.
- Anderson, K.F. 1998 (March). NASA's Anderson loop. *IEEE Instrum. Meas. Mag.* 1(1): 5–15.
- Angelo, E.J., Jr. 1969. *Electronics: BJTs, FETs and Microcircuits*. McGraw-Hill Book Co., New York.
- Apex. 2012. 276pp brochure on power op amps and their applications. Accessed April 10, 2013 at: www.apexanalog.com/wp-content/uploads/2012/08/TechSemV12.pdf.
- Apostolyuk, V. 2006. Theory and design of micromechanical vibratory gyroscopes. Astrise Corp Technical Paper. Accessed October 19, 2012 at: www.astrise.com/research/library/memsgyro.pdf.
- Arecco, D. 2003. Analysis and preliminary characterization of a MEMS cantilever-type chemical sensor. MS dissertation, Worcester Polytechnic Institute, Worcester, MA, 220pp.
- Arnaudov, R. and Y. Angelov. 2005. Earth rotation measurement with micromechanical yaw-rate gyro. *Meas. Sci. Technol.* 16: 2300–2306.
- ARRL. 1974. *The ARRL Antenna Book*, 13th edn. American Radio Relay League, Inc., Newington, CT.
- Aseltine, J.A. 1958. *Transform Method in Linear System Analysis*. McGraw-Hill Book Co., New York.
- Ayazi, F. and K. Najafi. 2001. A HARPSS polysilicon vibrating ring gyroscope. *JMEMS* 10(2): 169–179.
- Azaro, R., E. Zeni, P. Rocca, and A. Massa. 2007. Innovative design of a planar fractal-shaped GPS/GSM/Wi-Fi antenna. *Microw. Opt. Technol. Lett.* 50(3): 825–839.
- Bai, H. and G. Shi. 2007. Gas sensors based on conducting polymers. *Sensors* 7: 267–307.
- Baisa, N. 2005. Designing wireless interfaces for patient monitoring equipment. RF Design, April 1, 2005. Accessed November 3, 2011 at: <http://rfdesign.com/mag/504rfd4b.pdf>.
- Balanis, C.A. 1989. *Advanced Engineering Electromagnetics*. John Wiley & Sons, New York.
- Baltes, H., O. Brand, A. Hierlemann, D. Lange, and C. Hagleitner. 2002. CMOS MEMS—Present and future. In *Proceedings of the 15th IEEE International Conference on Micro Electro Mechanical Systems*, Las Vegas, NV. pp. 459–466.
- Barnes, J.R. 1987. *Electronic System Design: Interference and Noise Control Techniques*. Prentice Hall, Englewood Cliffs, NJ.
- Barnes Engineering. 1983. *Handbook of Infrared Radiation Measurements*. Barnes Engineering Co., Stamford, CT.
- Barney, G.C. 1985. *Intelligent Instrumentation: Microprocessor Applications in Measurement and Control*. Prentice Hall, Englewood Cliffs, NJ.
- Bartz, M. 1993 (December). What is dithering? *Hewlett-Packard J.*
- Barwicz, A. and W.J. Bock. 1990. An electronic high-pressure measuring system using polarimetric fiber optic sensor. *IEEE Trans. Instrum. Meas.* 39(6): 976–981.
- Beckwith, T.G. and N.L. Buck. 1961. *Mechanical Measurements*. Addison-Wesley, Reading, MA.
- Berger, A.J., T.-W. Koo, I. Itzkan, G. Horowitz, and M.S. Feld. 1999. Multicomponent blood analysis by near-infrared Raman spectroscopy. *Appl. Opt.* 38(13): 2916–2926.

- Bernstein, J. et al. 1993. A micromachined comb-drive tuning fork rate gyroscope. *MEMS* 93: 143–148.
- Berry, E. et al. 2004. Multispectral classification techniques for terahertz pulsed imaging: An example in histopathology. *Med. Eng. Phys.* 26: 423–430.
- Betzler, K. 2002. Fabry-Perot interferometer. Online tutorial paper from the University of Osnabrück. Accessed January 11, 2013 at: www.repairfaq.org/sam/laser/fabryperot.pdf.
- BioNavis. 2012. How does surface plasmon resonance work? 5pp web tutorial. Accessed December 12, 2012 at: www.bionavis.com/technology/spr/.
- Biosensing. 2012. Technical notes: 101—Principle of SPR detection: Intensity profile and shift of the SPR angle. 102—SPR sensitivity and detection limit. 103—Surface plasmon resonance v. Quartz crystal microbalance. 11pp. Accessed December 26, 2012 at: www.biosensingusa.com/Application101.html.
- Bluetooth. 2013. Bluetooth technology tutorial. 27 pp online document. Accessed May 15, 2013 at: www.radio-electronics.com/info/wireless/bluetooth/bluetooth_overview.php.
- BMF Wiki. 2008. Noise in MOSFETs and HBTs. Accessed May 6, 2013 at: http://bmf.ece.queensu.ca/mediawiki/index.php/Noise_in_MOSFETs_and_HBTs.htm.
- Bock, W.J., T.R. Wolinski, and A. Barwicz. 1990. Development of a polarimetric optical fiber sensor for electronic measurement of high pressure. *IEEE Trans. Instrum. Meas.* 39(5): 548–555.
- Boisen, A. et al. 2011. Cantilever-like micromechanical sensors. *Rep. Prog. Phys.* 74: 036101. 30pp.
- Bradford, M. 1975. Helicopter electrostatic discharge control system. MS dissertation in EE, University of Connecticut, Storrs, CT. (R.B. Northrop, adviser.)
- Braun, T. et al. 2009. Quantitative time-resolved measurement of membrane protein-ligand interactions using microcantilever array sensors. *Nat. Nanotechnol.* 4: 179.
- Breer, H. 1997. Sense of smell: Signal recognition and transduction in olfactory neurons. In *Handbook of Biosensors and Electronic Noses*, E. Kress-Rogers, ed. CRC Press, Boca Raton, FL, Chapter 22.
- Brenner, M.P., S. Hilgenfeldt, and D. Lohse. 2002. Single-bubble sonoluminescence. *Rev. Mod. Phys.* 74: 425–484.
- Briand, D. 2001a. Thermally isolated microelectronic devices for gas sensing applications. ScD dissertation, Institute of Microtechnology, University of Neuchâtel, Neuchâtel, Switzerland. 280pp. Accessed April 28, 2013 at: infoscience.epfl.ch/record/137625/files/These_BriandD.pdf.
- Briand, D. et al. 2001b. Thermally isolated MOSFET for gas sensing application. *IEEE Electron Device Lett.* 22(1): 11–13. Accessed April 28, 2013 at: ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=892428.
- Browne, A.F., T.R. Nelson, and R.B. Northrop. 1997. Microdegree polarimetric measurement of glucose concentrations for biotechnology applications. In *Proceedings of the 23rd Annual Northeast Bioengineering Conference*, J.R. Lacourse, ed. University of New Hampshire, Durham, NH, 5/21, 22/97. pp. 9–10.
- Bullock, T.H. and G.A. Horridge. 1965. Structure and function in the nervous systems of invertebrates. In *Arthropoda: Receptors Other Than Eyes*, Vol. II. W.H. Freeman and Co., San Francisco, CA, pp. 1150–1151, Chapter 18.
- Cameron, B.D. and G.L. Coté. 1997. Noninvasive glucose sensing utilizing a digital closed-loop polarimetric approach. *IEEE Trans. Biomed. Eng.* 44(12): 1221–1227.
- Cannon, R.H., Jr. 1967. Dynamics of physical systems. Sect. 5.7 “Lagrange’s Method”. In *Coupled Modes of Natural Motion: Many Degrees of Freedom*. McGraw-Hill Book Co., New York, Chapter 13.
- Capan, R., A.K. Ray, A.K. Hassan, and T. Tanriserver. 2003. Poly(methyl methacrylate) films for organic vapour sensing. *J. Phys. D Appl. Phys.* 36: 1115–1119.
- Caruso, M.J., T. Bratland, C.H. Smith, and R. Schneider. 1998. A new perspective on magnetic field sensing. 19pp. *Honeywell Corp. Report*. Accessed February 4, 2014 at: www51.honeywell.com/aero/common/documents/myaerospacecatalog-documents/Defense_Brochures-documents/Magnetic_Literature_Technical_Article-documents/A-New-Perspective-on-Magnetic-Field-Sensing.pdf.
- Caruso, M.J. and L.S. Withanawasam. 1999. Vehicle detection and compass applications using AMR magnetic sensors. In *Sensors Expo. Proceedings*, pp. 477–489. Accessed February 3, 2014 at: masters.donntu.edu.ua/2007/kita/gerus/library/amr.pdf.

- Casalinuovo, I.A., D. Di Pierro, M. Coletta, and P. Di Francesco. 2006. Application of electronic noses for disease diagnosis and food Spoilage detection. *Sensors* 6: 1428–1439.
- CDC. 2011. Measuring radiation. 3pp. Accessed May 03, 2013 at: <http://emergency.cdc.gov/radiation/measurement.asp>.
- Cease, T.W. and P. Johnston. 1990. A magneto-optical current transducer. *IEEE Trans. Power Deliv.* 5(2): 548–555.
- Cerna, M. and A.F. Harvey. 2000. The fundamentals of FFT-based signal analysis and measurements. Web tutorial on FFT. Accessed April 25, 2013 at: www.lumerlink.com/courses/ece697/docs/Papers/The Fundamentals of FFT-Based Signal Analysis and Measurements.pdf.
- Cheo, P.K. 1990. *Fiber Optics and Optoelectronics*. Prentice Hall, Englewood Cliffs, NJ.
- Cheung, N. 2010. Micro-electromechanical systems (MEMS) fabrication. On-line lecture notes, EE143, lecture 24, 42pp. Accessed May 29, 2013 at: www-inst.berkeley.edu/~ee143/fa10/lectures/Lec_24.pdf.
- Clarke, J. and R.H. Koch. 1988 (14 October). The impact of high temperature superconductivity on SQUID magnetometers. *Science* 242: 217–223.
- Cleland, A.N., M. Pophristic, and I. Ferguson. 2001. Single crystal aluminum nitride nanomechanical resonators. *Appl. Phys. Lett.* 69: 2070.
- Cline, J.D. and D.S. Grant. 1974. Electrostatic charge sensing probe. US Patent No. 3,857,066. Issued 24 December 1974. Dayton Aircraft Products, Assignee.
- Cohen, N. 1997. Fractal antenna applications in wireless telecommunications. In *Proceedings of the IEEE Electronic Industries Forum of New England*, Boston, MA. pp. 43–49.
- Collins, C.M. et al. 2002. Numerical calculations of the static magnetic field in three-dimensional multi-tissue models of the human head. *Magn. Reson. Imaging* 20: 413–424.
- Cooley, J.W. and J.W. Tukey. 1965. An algorithm for the machine computation of complex Fourier series. *Math. Comput.* 19: 297–301.
- Cooper, W.D. 1978. *Electronic Instrumentation and Measurement Techniques*, 2nd edn. Prentice Hall, Englewood Cliffs, NJ.
- Copeland, J., R. Yates, and L. Ruggiero. 2004. Wolverine population assessment in Glacier National Park. Progress Report, US Forest Service. 18pp. Accessed May 03, 2013at: www.wolverinefoundation.org/wp-content/uploads/2011/02/glacier04.pdf.
- Costa, D. et al. 2010. Accuracy of ARGOS locations of pinnipeds at-sea estimated using Fastloc GPS. *PLoS ONE* 5(1): e8677, 9pp.
- Coté, G.L., M.D. Fox, and R.B. Northrop. 1990. Optical polarimetric sensor for blood glucose measurement. In *Proceedings of the 16th Annual Northeast Bioengineering Conference*, R. Gaumond, ed. IEEE Press, New York, pp. 102–103.
- Coté, G.L., M.D. Fox, and R.B. Northrop. 1992. Noninvasive optical polarimetric glucose sensing using a true phase measurement technique. *IEEE Trans. Biomed. Eng.* 39(7): 752–756.
- Coté, G.L., M.D. Fox, and R.B. Northrop. 1993. Optical glucose sensor apparatus and method. US Patent #5,209,231.
- Craighead, H.G. 2000. Nanoelectromechanical systems. *Science* 290: 1532–1535.
- Crum, L.A. 1994 (September). Sonoluminescence. *Phys. Today* 22–29.
- Crum, L.A. and T.J. Matula. 1997 (30 May). Shocking revelations. *Science* 276: 1348–1349.
- Culshaw, B. 1984. *Optical Fiber Sensing and Signal Processing*. Peter Perigrinus, London, U.K.
- Cutting, C.L., A.C. Jason, and J.L. Wood. 1956. A capacitance-resistance hygrometer. *J. Sci. Instrum.* 33(1): 36.
- Dahake, S.L. et al. 1983. Progress in the realization of the units of capacitance, resistance and inductance at the National Physical Laboratory, India. *IEEE Trans. Instrum. Meas.* 32(1): 5–8.
- Davis, C.M. 1986. *Fiber Optic Sensor Technology Handbook*. Optical Technologies, Inc., Herndon, VA.
- Delahaye, F., A. Fau, D. Dominguez, and M. Bellon. 1987. Absolute determination of the farad and ohm, and measurement of the quantized Hall resistance $R_H(2)$ at LCIE. *IEEE Trans. Instrum. Meas.* 36(2): 205–207.
- DePriest, D. 2006. Differential GPS. Web paper: Accessed May 03, 2013 at: www.gpsinformation.org/dale/dgps.htm.

- Dhillon, N. and K. Grutter. 2007 Design, layout and simulation of a Silicon DRIE based tactical grade gyroscope. Report for the design project for ME218/EECS245. University of California, Berkeley, CA. Accessed April 28, 2013 at: www.me.berkeley.edu/~ndhillon/Publications/PDF/reports/7_gyroscope_design.pdf.
- Didenko, Y.T., W.B. McNamara III, and K.S. Suslick. 2000 (19 October). Molecular emission from single-bubble sonoluminescence. *Nature* 407: 877–879.
- Di Natale, C. et al. 2003. Lung cancer identification by analysis of breath by means of an array of non-selective gas sensors. *Biosen. Bioelectron.* 18: 1209–1218.
- Doleman, B.J. and N.S. Lewis. 2001. Comparison of odor detection thresholds and odor discriminabilities (sic) of a conducting polymer composite electronic nose versus mammalian olfaction. *Sens. Actuat.* 72: 41–50.
- Drake, A.D. and D.C. Leiner. 1984. A fiber Fizeau interferometer for measuring minute biological displacements. *IEEE Trans. Biomed. Eng.* 31(7): 507–511.
- Drysdale, C.V., A.C. Jolley and G.F. Tagg. 1952. *Electrical Measuring Instruments: Part 1*, 2nd edn. John Wiley & Sons, New York.
- Du, Z. 1993. A frequency-independent, high resolution phase meter. MS dissertation in Biomedical Engineering at the University of Connecticut, Storrs, CT. (R.B. Northrop, major adviser.)
- Dwight, H.B. 1969. *Tables of Integrals and Other Mathematical Data*. MacMillan Co., Toronto, Ontario, Canada.
- EDN. 2012. Energy harvesting. Accessed April 26, 2013 at: cds.linear.com/docs/en/article/ElecDesign/energy_harvesting_0912.pdf.
- Egan, W.F. 1981. *Frequency Synthesis by Phaselock*. John Wiley & Sons, New York.
- Ekinici, K.L. and M.L. Roukes. 2005. Nanoelectromechanical systems. *Rev. Sci. Instrum.* 76: 061101. DOI: 10.1063/1.1927327. 12pp.
- Endo, T., M. Koyangi, and A. Nakamura. 1983. High-accuracy Josephson potentiometer. *IEEE Trans. Instrum. Meas.* 32(1): 267–271.
- Exar Integrated Systems, Inc. 1979. *Phase-Locked Loop Data Book*. Exar Integrated Systems, Sunnyvale, CA.
- FAA. 2005. Wide-area augmentation system performance analysis report. Report #14. 100pp. FAA/William J. Hughes Technical Center. Atlantic City International Airport, NJ 08405. Accessed February 5, 2013 at: www.nts.gov/REPORTS/waaspan14.pdf.
- Far, A., B. Guo, F. Flitti, and A. Bermak. 2008. Temperature modulation for tin-oxide gas sensors. In *Proceedings of the Fourth IEEE Symposium on Electronic Design, Test & Applications*, Hong Kong, January 23–25, 2008. DOI: 10.1109/DELTA.2008.106, pp. 378–381.
- Farrand. 1996. Precision Inductosyn® Position Transducers. 15pp web article. Accessed February 1, 2013 at: www.ruhle.com/PDFFiles/FarrandControlsBrochure.pdf.
- Fellows, K.R. 1997. The design of a non-dispersive spectrophotometer to measure oxyhemoglobin. MS dissertation in Biomedical Engineering, The University of Connecticut, Storrs, CT. (R.B. Northrop major advisor.)
- Finden, H.J. and B.A. Horlock. 1957 (July). The Inductosyn and its application. *J. Brit. IRE* 369–383.
- Finot, E., A. Passian, and T. Thundat. 2008. Measurement of mechanical properties of cantilever shaped materials. *Sensors* 8: 3497–3541. DOI: 10.3390/s8053497.
- Fitzgerald, A.J. et al. 2006. Terahertz pulsed imaging of human breast tumors. *Radiology* 239(2): 533–540.
- Foreman, J.W. Jr., E.W. George, and R.D. Lewis. 1993a. Measurement of localized flow velocities in gasses with a laser Doppler flow meter. In *Selected Papers on Laser Doppler Velocimetry*, R.J. Adrian, ed. SPIE Press, Bellingham, WA, pp. 8–9.
- Foreman, J.W. Jr., R.D. Lewis, J.R. Thornton, and H.J. Watson. 1993b. Laser Doppler velocimeter for measurement of localized flow velocities in liquids. In *Selected Papers on Laser Doppler Velocimetry*, R.J. Adrian, ed. SPIE Press, Bellingham, WA, p. 10.
- Fort, A. et al. 2003. Tin oxide gas sensing: Comparison among different measurement techniques for gas mixture classification. *IEEE Trans. Instrum. Meas.* 52(3): 921–926.

- Foster, M.W., D.J. Ferrel, and R.A. Lieberman. 1994. Surface plasmon resonance biosensor miniaturization. *Proc. SPIE* 2293: 122–131.
- Fox, M.D. 1978. Multiple crossed-beam ultrasound Doppler velocimetry. *IEEE Trans. Sonics Ultrason.* 25(5): 281–286.
- Fox, M.D. 1987. True volume flow measurement with multiple beam ultrasound Doppler. In *Proceedings of the 13th Annual New England Bioengineering Conference*. IEEE Press, Philadelphia, PA, 2: 357–360.
- Fox, M.D. and J.F. Donnelly. 1978. Simplified method for determining piezoelectric constants for thickness mode transducers. *J. Acoust. Soc. Am.* 64(5): 1261–1265.
- Fox, M.D. and W.M. Gardiner. 1988. Three-dimensional Doppler velocimetry of flow jets. *IEEE Trans. Biomed. Eng.* 35(10): 834–841.
- Fox, M.D. and L.G. Puffer. 1978. Model for short-term movements in *Stapelia variegata* L. *Plant Physiol.* 61: 209–212.
- Fox, R. and T. Hill. 2007 (March/April). An exact value for Avogadro's number. *Am. Sci.* 3pp. Accessed April 8, 2013 at: www.americanscientist.org/issues/id.368,y.0,no.,content.true_page.1.css.print/issue.aspx.
- Fox, R.F. 2012. An exact value for Avogadro's number redux redux. Web paper. Accessed May 03, 2013 at: www.fefox.com/ARTICLES/AEVARR.pdf.
- Fractal Antenna. 2013. Fractal Antenna Systems, Inc. Homepage. (Company founded in 1995 by Dr. Nathan Cohen.) Accessed May 15, 2013 at: www.fractenna.com/about/about.html.
- Fractal FAQ. 2013. Fractal antenna FAQ. 6pp web article. Accessed May 15, 2013 at: www.fractenna.com/faq/faq.html.
- Franco, S. 1988. *Design with Operational Amplifiers and Integrated Circuits*. McGraw-Hill Book Co., New York.
- Frank, C.J. 1998 (September). Raman analysis in pharmaceuticals. Kaiser Optical Systems' *Raman Review*. pp. 1–3.
- Frog. 2012. On-line information on Defiant Technologies' Frog-4000™ hand-held gas chromatograph. Accessed May 13, 2013 at: www.defiant-tech.com/frog-4000.php.
- Fujita, T. et al. 1997. Two-dimensional micromachined gyroscope. In *Proceedings of the 1997 International Conference on Solid-State Sensors and Actuators*, Chicago, IL, June 16–19, 1997. IEEE, pp. 887–890.
- Fukuma, T. et al. 2012. Atomic-resolution imaging in liquid by frequency modulation atomic force microscopy using small cantilevers with megahertz-order resonance frequencies. *Nanotechnology* 23: 135706, 12pp.
- Gabriel, S., R.W. Lau, and C. Gabriel. 1996. The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz to 20 GHz. *Phys. Med. Biol.* 41: 2251–2269.
- Gamma-Sci. 2012a. Introduction to [UDT] Position Sensors. 4pp tutorial. Accessed December 17, 2012 at: www.gamma-sci.com/position-sensor-introduction/.
- Gamma-Sci. 2012b. Continuous Position Sensors. (List of UDT position sensor models.) Accessed December 17, 2012 at: www.gamma-sci.com/products/position-sensor/.
- Gaubert, C. et al. 2004. THz fractal antennas for electrical and optical semiconductor emitters and receptors. *Phys. Status Solidi C* 1(6): 1439–1444.
- GE. 2013. Pressure MEMS. Product listing. Accessed January 29, 2013 at: www.ge-mcs.com/pressure-mems.html.
- Gebeshuber, I.C. 2012. Nanoelectromechanical systems (NEMS)—Introduction, application and challenges of nanoelectromechanical systems, 4pp web review. Accessed May 22, 2013 at: www.azonanano.com/article.aspx?ArticleID=2465.
- Geddes, L.A. and L.E. Baker. 1968. *Principles of Applied Biomedical Instrumentation*. John Wiley & Sons, New York.
- Geddes, N.J. and C.R. Lawrence. 1997. Monitoring Immunoreactions with SPR. In *Handbook of Biosensors and Electronic Noses*, E. Kress-Rogers, ed. CRC Press, Boca Raton, FL. 349–368, Chapter 16.
- Geen, J. 2003. New iMEMS angular rate-sensing gyroscope. *Analog Dialog*. 37: 1–4.

- George, W.K. and J.L. Lumley. 1973. The laser Doppler velocimeter and its application to the measurement of turbulence. *J. Fluid Mech.* 60(2): 321–362.
- Gilham, E.J. 1957. A high-precision photoelectric polarimeter. *J. Sci. Instrum.* 34: 435–439.
- Goldstein, R.J. and W.F. Hagen. 1993. Turbulent flow measurements utilizing the Doppler shift of scattered laser radiation. In *Selected Papers on Laser Doppler Velocimetry*, R.J. Adrian, ed. SPIE Press, Bellingham, WA, p. 475.
- Goldstein, R.J. and D.K. Kreid. 1967. Measurement of laminar flow development in a square duct using a laser Doppler flowmeter. *J. Appl. Mechanics.* 34: 813–818.
- Gowar, J. 1984. *Optical Fiber Communications*. Prentice Hall, Englewood Cliffs, NJ.
- Graeme, J.G. 1974. *Applications of Operational Amplifiers*. McGraw-Hill, New York.
- Greenspan, D. and V. Casulli. 1993. *Numerical Analysis for Applied Mathematics, Science and Engineering*. Addison-Wesley Pub. Co., Reading, MA.
- Groselj, D. 2010. Guidance on instrumentation for calibration laboratories including RICs. Instruments and Observing Methods Report No. 101. World Meteorological Organization. WMO/TD-No. 1543. Accessed December 10, 2012 at: www.wmo.int/pages/prog/www/IMOP/publications/IOM-1012_RIC-Instrumentation.pdf.
- GSAT. 2013. Data sheet for USB Dongle GPS receiver with built-in SiRFStar GPS chip. Accessed April 16, 2013 at: <http://glsat.manufacturer.globalsources.com/si/6008802441322/pdtl/Portable-GPS/1031397653/GPS-USB-Dongle.htm>.
- Grubb, W.T. and L.H. King. 1980. Palladium-palladium oxide pH electrodes. *Anal. Chem.* 52: 270–273.
- Guo, B., A. Bermak, P.C.H. Chan, and G-Z. Yan. 2007. A monolithic integrated 4 × 4 tin oxide gas sensor array with on-chip multiplexing and differential readout circuits. *Solid-State Electron.* 51: 69–76.
- Gupta, N., S. Sharma, I. A. Mir, and D. Kumat. 2006. Advances in sensors based on conducting polymers. *J. Sci. Ind. Res.* 65: 549–557.
- Haddad, R., A. Medhanie, Y. Roth, D. Harel, and N. Sobel. 2010. Predicting odor pleasantness with an electronic nose. *PLoS Comput. Biol.* 6(4): 11.
- Hall, H.T. 1958. Some high-pressure, high-temperature apparatus design considerations: Equipment for used at 1000,000 atmospheres and 3000°C. *Rev. Sci. Instrum.* 29: 267–275.
- Hamamatsu. 2013. PMT catalog. Accessed April 15, 2013 at: www.hamamatsu.com/us/en/3001.html.
- Hamilton, C.A. et al. 1991. A 24 GHz Josephson array voltage standard. *IEEE Trans. Instrum. Meas.* 40(2): 301–304.
- Hanlon, E.B. et al. 2000. Prospects for *in vivo* Raman spectroscopy. *Phys. Med. Biol.* 45(2): R1–R59.
- Hansen, A.T. 1983. Fiber-optic pressure transducers for medical applications. *Sens. Actuat.* 4: 545–554.
- Harnden, J.D., Jr., F.D. Martzloff, W.G. Morris, and F.B. Golden. 1972 (9 October). Metal-oxide varistor: A new way to suppress transients. *Electronics*. Accessed April 11, 2013 at: www.eeel.nist.gov/817/pubs/spd-anthology/files/MOV_announce_BW.pdf.
- Harrington, D. 2010. Tracking the wolverine, 9 pp web post. Accessed May, 03, 2013 at: <http://danielharrington.wordpress.com/2010/01/31/tracking-the-wolverine/>.
- Hashemi, K. 2009. Subcutaneous transmitter. Open Source Instruments, Inc., Web paper. Accessed May 13, 2013 at: www.opensourceinstruments.com/Electronics/A3013/M3013.html.
- Hecht, E. 1987. *Optics*, 2nd edn. Addison-Wesley, Reading, MA.
- Helfrick, A.D. and W.D. Cooper. 1990. *Modern Electronic Instrumentation and Measurement Techniques*. Prentice Hall, Englewood Cliffs, NJ.
- Hey, J.C. and W.P. Kram, eds. 1978. *Transient Voltage Suppression Manual*, 2nd edn. General Electric Co., Auburn, NY.
- Hierlemann, A. and H. Balthes. 2003. CMOS-based chemical microsensors. *Analyst* 128: 15–28.
- Hiller, R.A. and B.P. Barber. 1995 (February). Producing light from a bubble of air. *Sci. Am.* 96–98.
- Hiller, R.A., K. Weninger, S.J. Putterman, and B.P. Barber. 1994 (14 October). Effect of noble gas doping in single-bubble sonoluminescence. *Science* 266: 248–250.
- Hochberg, R.C. 1986. Fiber-optic sensors. *IEEE Trans. Instrum. Meas.* 35(4): 447–450.

- Honeywell. 2013. General purpose accelerometers, Web announcement. Accessed April 26, 2013 at: <https://measurementsensors.honeywell.com/pages/Category.aspx?cat=Honeywell&category=PRODUCTTYPES-ACCELEROMETERS-GENERALPURPOSE/>.
- Hong, H.-D. 1994. Optical interferometric measurement of skin vibration for the diagnosis of cardiovascular disease. PhD dissertation in Biomedical Engineering, University of Connecticut, Storrs, CT. (M.D. Fox, adviser.)
- Hong, H.-D. and M.D. Fox. 1993. Detection of skin displacement and capillary flow using an optical stethoscope. In *Proceedings of the 19th Annual Northeast Bioengineering Conference*, Newark, NJ. March 18–19, 1993. IEEE Press, New York, pp. 189–190.
- Huang, X.M.H. et al. 2005. VHF, UHF and microwave frequency nanomechanical resonators. *New J. Phys.* 7: 15.
- Hung, K.K. et al. 1990. A unified model for the flicker noise in metal-oxide-semiconductor field effect transistors. *IEEE Trans. Electron. Dev.* 37(3): 654–665.
- Hwang, S.-W. et al. 2012 (28 September). A physically transient form of silicon electronics. *Science* 337: 1640–1644.
- Ibrahhiem, A. et al. 2006. Bi-band fractal antenna design for RFID applications at UHF. In *Proceedings of EuCAP 2006*, Nice, France, November 6–10, 2006, 3pp.
- Ibuka, M., S. Naito, and T. Furuya. 1983. Point-contact Josephson voltage standard. *IEEE Trans. Instrum. Meas.* 32(1): 276–279.
- Igarishi, T., Y. Koizumi, and M. Kanno. 1968. Determination of an absolute capacitance by a horizontal cross-capacitor. *IEEE Trans. Instrum. Meas.* 17(4): 226–231.
- Ilic, B. et al. 2001. Single cell detection with micromechanical oscillators. *J. Vac. Sci. Technol. B* 19: 2825.
- Irokawa, Y. 2011. Hydrogen sensors using nitride-based semiconductor diodes: The role of metal/semiconductor interfaces. *Sensors* 11: 674–695. DOI: 10.3390/s110100664.
- Jacob, T. 2003. Olfaction: A tutorial on the sense of smell, Web tutorial. Accessed April 28, 2013 at: www.cf.ac.uk/bio/staffinfo/jacob/teaching/sensory/olfact1.html.
- Jaeger, K.B., P.D. Levine, and C.A. Zack. 1991. Industrial experience with a quantified Hall effect system. *IEEE Trans. Instrum. Meas.* 40(2): 256–261.
- Jakobson, C.G. and Y. Nemirovsky. 1999. 1/f noise in ion sensitive field effect transistors from sub-threshold to saturation. *IEEE Trans. Electron Dev.* 48(1): 259–260.
- Jamasb, S. et al. 1998. Accurate continuous monitoring using ISFET-based biosensors based on characterization and modeling of drift and low frequency noise. In *Proceedings of the 20th Annual Conference of the IEEE EMB Society*. IEEE Service Center, Piscataway, NJ, 20(6): 2864–2867.
- James, H.J., N.B. Nichols, and R.S. Philips. 1947. *Theory of Servomechanisms*. McGraw-Hill, New York.
- Janata, J. 2003. Electrochemical microsensors. *Proc. IEEE* 91(6): 864–869.
- Jeol.com. 2007 (3 October). A 30 Wh/kg supercapacitor for solar energy and a new battery. Jeol.com. Accessed September 13, 2011 at: www.jeol.com/NEWSEVENTS/PressReleases/tabid/521/articleType/ArticleView/articleId/112/A-30-Whkg-Supercapacitor-for-solar-Energy-and-a-New-Battery.aspx.
- Johnston, J.M. 1999. Canine detection capabilities: Operational implications of recent R&D findings, 7pp web paper. Accessed April 19, 2013 at: http://barksar.org/K-9_Detection_Capabilities.pdf.
- Jory, M.J. et al. 1995. A surface-plasmon-based optical sensor using acousto-optics. *Meas. Sci. Technol.* 6: 1193–1200.
- Juang, J.-N. and R. Radharamanan. 2009. Evaluation of ring laser and fiber optic gyroscope technology, 9pp web paper. Accessed May 03, 2013 at: www.asee.org/documents/sections/middle-atlantic/fall-2009/01-Evaluation-Of-Ring-Laser-And-Fiber-Optic-Gyroscope-Technology.pdf.
- Juneau, T. and A.P. Pisano. 1996 (June). Micromachined dual axis angular rate sensor. In *Proceedings of the IEEE Solid State Sensor and Actuator Workshop*, Hilton Head Island, SC.
- Juneau, T., A.P. Pisano, and J.H. Smith. 1997 (June). Dual-axis operation of a micromachined rate gyroscope. In *Technical Digest of the Ninth International Conference Solid-State Sensors and Actuators (Transducers'97)*, Chicago, IL, pp. 883–886.

- Kalorama. 2010. Remote and wireless patient monitoring markets. Kalorama Information, R566-530, 8pp. Accessed June 17, 2010 at: www.mindbranch.com/catalog/print_product_page.jsp?code=R566-530.
- Kaiser, S., T. Maier, A. Grossmann, and C. Zimmerman. 2001. Fizeau interferometer for phase-shifting interferometry in ultrahigh vacuum. *Rev. Sci. Instrum.* 72(9): 3726–3727.
- Kanai, H., K. Noma, and J. Hong. 2001 (December). Advanced spin-valve GMR-head. *Fujitsu Sci. Tech. J.* 37(2): 174–182.
- Kandel, E.R., J.H. Schwartz, and T.M. Jessel. 1991. *Principles of Neural Science*, 3rd edn. Appleton and Lange, Norwalk, CT.
- Karami, M.A. and D.J. Inman. 2011. Electrochemical modeling of the low-frequency zigzag micro-energy harvester. *J. Int. Mater. Syst. Struct.* 22: 271–281.
- Karami, M.A. and D.J. Inman. 2012. Powering pacemakers from heartbeat vibrations using linear and nonlinear energy harvesters. *Appl. Phys. Lett.* 100: 042901, 4pp.
- Karim, M.N.A. et al. 2010. Log periodic fractal Koch antenna for UHF band applications. *Prog. Electromagn. Res.* 100: 201–218.
- Keiser, G. 1983. *Optical Fiber Communications*. McGraw-Hill, New York.
- Keithley Instruments, Inc. 2004. Low Level Measurement, Precision DC Current, Voltage, and Resistance Measurements, 6th edn. Accessed March 21, 2014 at: www.Keithley.com/knowledgecenter/knowledgecenter_pdf/LowLevMsHandbk.pdf.
- Keller, P.E. et al. 1995 (12 October). Electronic noses and their applications. In *Proceedings of the IEEE Nothcon/Technical Applications Conference (TAC'95)*, Portland, OR.
- Khanna, V.K. 2007. Advances in chemical sensors, biosensors and microsystems based on ion-sensitive field-effect transistor. *Indian J. Pure Appl. Phys.* 45: 345–353.
- Kibble, B.P. 1983. Realizing the ampere by levitating a superconducting mass—A suggested procedure. *IEEE Trans. Instrum. Meas.* 32(1): 144.
- Kinley, H. 1980. *The PLL Synthesizer Cookbook*. Tab Books, Inc., Blue Ridge Summit, PA.
- Knapp, C.H., R. Bansal, M.S. Mashikian, and R.B. Northrop. 1990. Signal processing techniques for partial discharge site location in shielded cables. *IEEE Trans. Power Deliv.* 5(2): 859–865.
- Koo, T.-W. et al. 1999. Reagentless blood analysis by near-infrared Raman spectroscopy. *Diab. Technol. Ther.* 1(2): 153–157.
- Koskinen, M., J. Typpö, and J. Kostamovaara. 1992. A fast time-to-amplitude converter for pulsed time-of-flight laser rangefinding. *SPIE* 1633: 128–136.
- Kottapalli, A.G.P., M. Asadnia, J.M. Miao, G. Barbastathis, and M.S. Triantafyllou. 2012. A flexible liquid crystal polymer MEMS pressure sensor array for fish-like underwater sensing. *Smart Mater. Struct.* 21: 13.
- Koujili, M. 2008. Inductosyn, 13pp on-line paper. Accessed February 1, 2013 at: http://project-wire-scanner.web.cern.ch/project-wire-scanner/Notes/rapport_inductosyn.pdf.
- Kraus, J.D. 1953. *Electromagnetics*. McGraw-Hill Book Co. Inc., New York.
- Kumar, R. and P. Malathi. 2009. On the design of wheel shape fractal antenna. *Int. J. Recent Trends Eng.* 2(6): 98–100.
- Lambert, J.L., J.M. Morookian, S.J. Sirk, and M.S. Borchert. 2002. Measurement of aqueous glucose in a model anterior chamber using Raman spectroscopy. *J. Raman Spectrosc.* 33(7): 524–529.
- Langley, R.B. 2000 (April). Smaller and smaller: The evolution of the GPS receiver. *GPS World.* 11: 54–58.
- Lathi, B.P. 1965. *Signals, Systems and Communication*. John Wiley & Sons, New York.
- Leake, L.L. 2006. Electronic noses and tongues. *Food Technol.* 6(6): 96–102.
- Levitsky, I., S.G. Krivoshylykov, and J.W. Grate. 2001. Rational design of a Nile red/polymer composite film for fluorescence sensing of organophosphonate vapors using hydrogen bond acidic polymers. *Anal. Chem.* 73: 3441–3448.
- Lewis, N.S. 2000. Use of conducting polymer composite chemically sensitive resistors as an “electronic nose”. Accessed May 30, 2013 at: www.its.caltech.edu/~mmrc/nsl/nose.htm.
- Lewis, S.B., R. Flynn, and N. Barten. 2008. Wolverine population ecology in Berners Bay, Alaska. Wildlife research progress report, Alaska Department of Fish and Game, Division of Wildlife Conservation, St. Juneau, AK, 23 pp. Accessed May 03, 2013 at: www.wolverinefoundation.org/wp-content/uploads/2911/02/BBAKProgress2008.pdf.

- Li, J. 2000 (August). The Cyranose chemical vapor analyzer. *Sensors* 17: 56–60, 9pp. Accessed November 26, 2012 at: <http://archives.sensorsmag.com/articles/0800/56/main>.
- Li, X. et al. 2009. Integrated MEMS/NEMS resonant cantilevers for ultrasensitive biological detection. *J. Sens.* 2009: 10, Article ID 637874.
- Lia, R., K. Gibblea, and K. Szymaniec. 2011. Improved accuracy of the NPL-CsF2 primary frequency standard: evaluation of distributed cavity phase and microwave lensing frequency shifts, 7pp. Web paper. Accessed April 10, 2013 at: <http://arxiv.org/ftp/arxiv/papers/1107/1107.2412.pdf>.
- Lion, K.S. 1959. *Instrumentation in Scientific Research: Electrical Input Transducers*. McGraw-Hill Book Co., Inc, New York.
- Liu, C.C., P.J. Hesketh, and G.W. Hunter. 2004. Chemical microsensors. *The Electrochemical Society Interface*. Summer 2004, pp. 22–27.
- Liu, J.-C. et al. 2006. Modified Sierpinski fractal monopole antenna with Descartes circle theorem. *Microw Opt. Technol. Lett.* 48(5): 909–911.
- Liu, Y. et al. 2011. An extended CMOS ISFET model incorporating the physical design geometry and the effects of performance and offset variation. *IEEE Trans. Electron Devices*. 58(12): 4414–4422.
- Lohse, D. 2002 (July 25). Sonoluminescence: Inside a micro-reactor. *Nature* 418: 381–383.
- LT. 2010 (October). Ultralow voltage energy harvester uses thermoelectric generator for battery-free wireless sensors. *Linear Technol. J. Analog Innov.* 20(3): 39.
- Macovski, A. 1983. *Medical Imaging Systems*. Prentice Hall, Englewood Cliffs, NJ.
- Maenaka, K. and T. Shiozawa. 1994. A study of silicon angular rate sensors using anisotropic etching technology. *Sens. Actuat. A*. 43: 72–77.
- Marcuse, D. 1981. *Principles of Optical Fiber Measurements*. Academic Press, New York.
- Marlow. 2013. 7. How do thermoelectric generators (TEGs) work? 3 pp web essay. Accessed June 21, 2013 at: www.marlow.com/resources/general-faq/7-how-do-thermoelectric-generators-work.html.
- Maron, S.H. and C.F. Prutton. 1958. *Principles of Physical Chemistry*. MacMillan Co., New York.
- Martynenko, Y.G., I.V. Merkuriev, and V.V. Podalkov. 2010. Dynamics of a ring micromechanical gyroscope in the forced oscillation mode. *Gyrosc. Navig.* 1(1): 43–51.
- Mashikian, M.S., R.B. Northrop, R. Bansal, and C.L. Nikias. 1989. Method and instrumentation for the detection, location, characterization of partial discharges and faults in electric power cables. US Patent No. 4,887,041, December 12, 1989.
- Mashikian, M.S., R.B. Northrop, R. Bansal, and F. Palmieri. 1993. Method and apparatus for the detection and location of faults and partial discharges in shielded cables. US Patent No. 5,272,439, December 21, 1993.
- Mashikian, M.S., R.B. Northrop, and D. Sui. 1995. Cable fault detection using a high-voltage alternating polarity DC signal superposed with a system frequency AC signal. US Patent No. 5,448,176, September 5, 1995.
- Mason, J.H. 1965. Discharge detection and measurements. *Proc. IEE*. 112(7): 1407–1422.
- Mason, W.P. 1957. Properties of transducer materials. In *American Institute of Physics Handbook*, D.E. Gray, ed. McGraw-Hill, New York, pp. 3–119, Chapter 3.
- McConnell, R.L. 1993. Capacitive proximity sensors with additional driven shields. *NASA Tech Briefs Technical Support Package GSC-13475*. Goddard Space Flight Center, Greenbelt, MD 20771.
- Melexis. 2013. Microelectronic integrated systems. Accessed January 29, 2013 at: www.melexis.com.
- Mikron. 2013. Catalog on blackbody radiation calibration sources, 8pp. Accessed April 12, 2013 at: www.transcat.com/PDF/blackbody.pdf.
- Millman, J. 1979. *Microelectronics*. McGraw-Hill Book Co., New York.
- Mitchell, K. 2001. Tutorial: Hilbert curve coloring. Accessed July 28, 2010 at: www.fractalus.com/kerry/tutorials/hilbert/hilbert-tutorial.html.
- Möller, K.D. 1988. *Optics*. University Science Books, Mill Valley, CA.
- Mopsik, F.I. 1984. Precision time-domain dielectric spectrometer. *Rev. Sci. Instrum.* 55(1): 79–87.
- Morgan, D. 1978. Revised data-interface standards. Motorola AN781A. Accessed April 23, 2013 at: www.viste.com/LON/NeuronChip/documents/APPSLIT.pdf.

- Morgenshtein, A., U. Dinnar, and Y. Nemirovsky. 2002. ISFET operation in pass-transistor mode without readout circuits, 4pp web paper. Accessed May 06, 2013 at: [webee.technion.ac.il/people/arkadiy/papers/ISFET Operation in Pass Transistor Mode without Readout Circuits.pdf](http://webee.technion.ac.il/people/arkadiy/papers/ISFET%20Operation%20in%20Pass%20Transistor%20Mode%20without%20Readout%20Circuits.pdf).
- Moseley, P.T., J. Norris, and D. Williams. 1991. *Techniques and Mechanisms in Gas Sensing*. Adam Hilger, New York.
- Moss, W.C., D.B. Clarke, and D.A. Young. 1997a (May 30). Calculated pulse widths and spectra of a single sonoluminescing bubble. *Science* 376: 1398–1401.
- Moss, W.C., D.B. Clarke, and D.A. Young. 1997b. Star in a jar: A new model for single-bubble sonoluminescence. In *Proceedings of the 133rd ASA Meeting*, State College, PA, 4pp.
- Motorola. 2013. Pressure sensors: Motorola master selection guide, 17pp. Accessed February 5, 2013 at: http://noel.feld.cvut.cz/hw/motorola/books/sg73/pdf/5_9sensor_msg.pdf.
- MTS. 2013. Magnetostrictive linear position sensors, 4pp web tutorial. Accessed May 06, 2013 at: www.sensorland.com/HowPage024.html (Also see: www.mtssensors.com/).
- Muller, R.S., R.T. Howe, S.D. Senturia, R.L. Smith, and R.M. White. 1991. *Microsensors*. IEEE Press, New York.
- Nagle, H.T., S.S. Schiffman, and R. Gutierrez-Osuna. 1998. The how and why of electronic noses. *IEEE Spectrum*. 35(9): 22–34.
- Naik, A.K. et al. 2009. Towards single-molecule nanomechanical mass spectrometry. *Nat. Nanotechnol.* 4: 445–450.
- Nanavati, R.P. 1975. *Semiconductor Devices: BJTs, JFETs, MOSFETs, and Integrated Circuits*. Intext Educational Publishers, New York.
- Napolitano, F. 2010. Fiber-optic gyroscopes key technological advantages, 8pp web article. Accessed May 03, 2013 at: www.ixsea.com/pdf/fog-key-advantages.pdf.
- NASA. 2013. Johannes Kepler: His life, his laws and times, 6pp. Accessed April 30, 2013 at: kepler.nasa.gov/Mission/JohannesKepler/.
- National Instruments. 2003a. How IVI-C instrument driver technology enables system longevity and platform portability, 3pp. Accessed April 29, 2013 at: www.ni.com/ivi/white-paper/3433/en.
- National Instruments. 2003b. IVI architecture. Accessed April 29, 2013 at: www.ni.com/ivi/white-paper/3433/en.
- National Instruments. 2003c. *IVI: Comparison of IVI-C and IVI-COM Drivers*. Accessed 4/29/13 at: www.ni.co/ivi/c_com.htm.
- National Instruments. 2013. What is PXI? 9pp web tutorial. Accessed April 29, 2013 at: www.ni.com/white-paper/4811/en.
- Natus. 2013. Grass instruments company history, 2pp. Accessed April 17, 2013 at: www.grasstechnologies.com/company/history.html.
- Navon, D.H. 1975. *Electronic Materials and Devices*. Houghton-Mifflin, Boston, MA.
- Ndieyira, J.W. et al. 2008. Nanomechanical detection of antibiotic mucopeptide binding in a model for superbug drug resistance. *Nat. Nanotechnol.* 3(11): 691–696.
- Nelson, R.A. 1999a. The global positioning system: A national resource. Accessed April 28, 2013 at: www.atcourses.com/global_positioning_system.htm.
- Nelson, T.R. 1999b. Development of a type 1 nonlinear feedback system for laser velocimetry and ranging. MS dissertation in Biomedical Engineering at the University of Connecticut, Storrs, CT. (R.B. Northrop, major adviser.)
- Nemat, A. 1990. A digital frequency independent phase meter. *IEEE Trans. Instrum. Meas.* 39(4): 665–666.
- Neuman, M.R. 1988. Neonatal monitoring. In *Encyclopedia of Medical Devices and Instrumentation*. J.G. Webster, ed. John Wiley & Sons, New York.
- Newport Corporation. 2012. *Diffraction Grating Handbook*, Richardson Grating Laboratory, New York. Accessed December 11, 2012 at: <http://gratings.newport.com/library/handbook/toc.asp>.
- Nicati, P.A. and P. Robert. 1988. Stabilised current sensor using Sagnac interferometer. *J. Phys. E: Sci. Instrum.* 21: 791–796.

- NIST. 2013. International aspects of the SI. Accessed June 13, 2013 at: <http://physics.nist.gov/cuu/Units/international.html>.
- NOAA. 2010. Draft environmental assessment. Gulf of the Farallones National Marine Sanctuary White Shark research permit application for project entitled: Fine scale, long-term tracking of adult white sharks, Web document. Accessed May 03, 2013 at: http://farallones.noaa.gov/eco/sharks/pdf/mcsi_draft_ea_sept_24_2010.pdf.
- Northrop, R.B. 1980. Ultrasonic respiration/convulsion monitoring apparatus and method for its use. US Patent No. 4,197,856, Issued April 15, 1980.
- Northrop, R.B. 1990. *Analog Electronic Circuits: Analysis and Applications*. Addison-Wesley, Reading, MA.
- Northrop, R.B. 2000. *Endogenous and Exogenous Regulation and Control of Physiological Systems*. CRC Press, Boca Raton, FL.
- Northrop, R.B. 2001. *Introduction to Dynamic Modeling of Neuro-Sensory Systems*. CRC Press, Boca Raton, FL.
- Northrop, R.B. 2002. *Non-Invasive Instrumentation and Measurements in Medical Diagnosis*. CRC Press, Boca Raton, FL.
- Northrop, R.B. 2003. *Signals and Systems Analysis in Biomedical Engineering*. CRC Press, Boca Raton, FL.
- Northrop, R.B. 2004. *Analysis and Application of Analog Electronic Circuits to Biomedical Instrumentation*. CRC Press, Boca Raton, FL.
- Northrop, R.B. 2005. *Introduction to Instrumentation and Measurements*, 2nd edn. Taylor & Francis, Boca Raton, FL.
- Northrop, R.B. 2010. *Signals and Systems Analysis in Biomedical Engineering*, 2nd edn. CRC Press, Boca Raton, FL.
- Northrop, R.B. 2012. *Analysis and Application of Analog Electronic Circuits to Biomedical Instrumentation*, 2nd edn. CRC Press, Boca Raton, FL.
- Northrop, R.B. and B.M. Decker. 1977. Design of a no-touch infant apnea monitor. In *Proceedings of the Fifth Annual North-East Bioengineering Conference*. Pergamon Press, New York, pp. 245–248.
- Northrop, R.B. and H.M. Horowitz. 1966. An instantaneous pulse frequency demodulator for neurophysiological applications. *Proceedings of the Symposium on Biomedical Engineering*, Milwaukee, WI, Vol. 1, pp. 5–8.
- Northrop, R.B., J.-M. Wu, and H.M. Horowitz. 1967. An instantaneous frequency cardiometer. In *Digest of Seventh International Conference on Engineering in Medicine and Biology*. IEEE press. Stockholm, Sweden, p. 417.
- Olfaction. 2010. Sensory abilities, 12pp. web paper. Accessed April 19, 2012 at: <http://cynologist.com/index.php/anatomy-of-dog/sensory-abilities-olfaction/>.
- Oliver, B.M. and J.M. Cage. 1971. *Electronic Measurements and Instrumentation*. McGraw-Hill, New York.
- Olson, H.F. 1940. *Elements of Acoustical Engineering*. D. Van Nostrand Co. Inc., New York.
- Omega Corporation. 2013. *Relative Humidity Measurement Instruments*. Accessed April 30, 2013 at: www.omega.ca/shop/sectionSC.asp?section=HU&book=temperature.
- Oriel. 2013a. Oriel photomultiplier tubes. Accessed April 15, 2013 at: <http://assets.newport.com/pdfs/e5582.pdf>.
- Orion. 2013b. Orion ion selective electrodes. Accessed April 15, 2013 at: [ww.pollardwater.com/pages_product/L1302008electrodes.asp](http://www.pollardwater.com/pages_product/L1302008electrodes.asp).
- Ortec. 2000. Pulse-processing electronics for single-photon counting. Application Note AN51, Revelation 1, 8pp. Accessed April 15, 2013 at: www.ortec-online.com/library/index.aspx.
- Ott, H.W. 1976. *Noise Reduction Techniques in Electronic Systems*. John Wiley & Sons, New York.
- Ouellette, J. 1999 (February). Electronic noses sniff out new markets. *Indust. Phys.* 5(1): 26–29.
- Pallás-Areny, R. and J.G. Webster. 1991. *Sensors and Signal Conditioning*. John Wiley & Sons, Inc., New York.
- Pallás-Areny, R. and J.G. Webster. 2001. *Sensors and Signal Conditioning*, 2nd edn. John Wiley & Sons, Inc., New York.
- Pankaj, K., S.H. Kumar, and K. Sukhjeet. 2012. Conducting polymer based potentiometric sensors. *Res. J. Chem. Environ.* 16(3): 125–133.

- Pankratz, C.K. 1998. *Microchannel Plates and MCP Detectors and Imaging Systems*. Del Mar Photonics, San Diego, CA, A MCP tutorial, 12pp. Accessed April 29, 2013 at: www.dmphotronics.com/MCP_MCPIImageIntensifiers/mcp_references.htm.
- Papi, M, G. Arcovito, and M. De Spirito. 2006. Fluid viscosity determination by means of uncalibrated atomic force microscopy cantilevers. *Appl. Phys. Lett.* 88: 194102 (2006), 3pp.
- Papoulis, A. 1968. *Systems and Transforms with Applications to Optics*. McGraw-Hill, New York.
- Papoulis, A. 1977. *Signal Analysis*. McGraw-Hill Book Co., New York.
- P-E Optoelectronics. 2003. CPM-channel photomultiplier: Operating instruction, 10pp. Accessed May 03, 2013 at: www.hofoo.com.cn/uploadfiles/cpmoperatinginstruction.pdf.
- Persaud, K. and G. Dodd. 1982. Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose. *Nature* 299: 325–355.
- Peslin, R. et al. 1975. Frequency response of the chest: Modeling and parameter estimation. *J. Appl. Physiol.* 39(4): 523–534.
- Peterson, J.I. and S.R. Goldstein. 1982. A miniature fiber optic pH sensor potentially suitable for glucose measurements. *Diabetes Care* 5(3): 272–274.
- Peura, R.A. and Y. Mendelson. 1984. Blood glucose sensors: An overview. *Proceedings of the IEEE/NSF Symposium on Biosensors*, IEEE Press, New York, pp. 63–68.
- Phillips, M. et al. 2003. Detection of lung cancer with volatile markers in the breath. *Chest* 123: 1788–1792.
- Piazza, G. and P. Stephanou. 2002. Micromechanical thermo-fluidic single-axis yaw rate sensor. Berkeley EECS Student paper. Accessed October 22, 2012 at: www-bsac.eecs.berkeley.edu/projects/ee245/Reference_Papers/Piazza_Stephanou_EECS245_Fall02.pdf.
- Pimmel, R.L. et al. 1977. Instrumentation for measuring respiratory impedance by forced oscillations. *IEEE Trans. Biomed. Eng.* 24(2): 89–93.
- Pinnaduwa, L.A., V. Boiadjev, J.E. Hawk, and T. Thundat. 2003. Sensitive detection of plastic explosives with self-assembled monolayer-coated microcantilevers. *Appl. Phys. Lett.* 83: 1741.
- Piquemal, F. et al. 1991. Direct comparison of quantized Hall resistances. *IEEE Trans. Instrum. Meas.* 40(2): 234–236.
- Pitcher, S., J.A. Thiele, H. Ren, and J.F. Vetelino. 2003. Current/voltage characteristics of a semiconductor metal oxide gas sensor. *Sensors Actuat. B.* 93: 454–462.
- Pöpel, R. et al. 1991. Nb/Al₂O₃/Nb Josephson voltage standards at 1 V and 10 V. *IEEE Trans. Instrum. Meas.* 40(2): 298–300.
- Prashar, D. 2012. Self assembled monolayers—A review. *Int. J. Chem. Tech. Res.* 4(1): 258–265.
- Prazak, P. and W.B. Scott. 1975 (March 20). Radiation monitor has linear output. *Electronics* 117.
- Princeton Instruments. 2013. Raman spectroscopy systems (homepage). Accessed May 6, 2013 at: www.princetoninstruments.com/products/specsys/default.aspx?
- Puente-Baliarda, C. et al. 1998. On the behavior of the Sierpinski multiband antenna. *IEEE Trans. Antennas Propag.* 46(4): 517–524.
- PXI Systems Alliance. 2005. PXI-1™ hardware specifications. Revelation 2.2, June 13, 2005, 57pp. Accessed April 25, 2013 at: www.pxisa.org/userfiles/files/Specifications/PXIHWSPEC22.pdf.
- QRS. 2012. Quartz MEMS technology. Systron Donner inertial online information. Accessed December 12, 2012 at: www.systron.com/support/technology.
- QRS14. 2011. QRS14 user's guide: MEMS angular rate sensor Model QRS14, 8pp. Accessed April 29, 2013 at: www.systron.com/sites/default/files/964014-d_qrs14_1_.pdf.
- Quayle, S. 2013. SI radiation measurement units: Conversion factors, 1p. Accessed May 01, 2013 at: www.stevequayle.com/index.php?s=239.
- Quinn, E.J. and T. Posipanko. 1970. Pressure transducer system for continuous measurement of minute pressure changes. *Rev. Sci. Instrum.* 41: 475–476. doi:10.1063/1.1684555.
- Rabinovitch, B., W.F. March, and R.L. Adams. 1982. Noninvasive glucose monitoring of the aqueous humor of the eye. Part I. Measurements of very small optical rotations. *Diabetes Care* 5(3): 254–258.
- Raiteri, R. et al. 2001. Micromechanical cantilever-based biosensors. *Sens. Actuat. B.* 79: 115–126.
- Rakow, N.A. and K.S. Suslick. 2000 (August 17). A colorimetric sensor array for odor visualization. *Nature* 406: 710–713.

- Ramos, D. et al. 2009. Arrays of dual nanomechanical resonators for selective biological detection. *Anal. Chem.* 81: 2274.
- Reymann, D. 1991. A practical device for 1 nV accuracy measurements with Josephson arrays. *IEEE Trans. Instrum. Meas.* 40(2): 309–311.
- Ribeiro, P.C., S.J. Williamson, and L. Kaufman. 1988. SQUID arrays for simultaneous magnetic measurements: Calibration and source localization performance. *IEEE Trans. Biomed. Eng.* 35(7): 551–559.
- Ripka, P., M. Tondra, J. Stokes, and R. Beech. 1998. AC-driven AMR and GMR magnetosensors. *Proceeding Eurosensors XII*, Southampton, U.K., pp. 967–969.
- Rogers, A.J. 1973. Optical technique for measurement of current at high voltage. *Proc. IEEE.* 120(2): 261–267.
- Rogers, A.J. 1976. Method for simultaneous measurement of voltage and current on high voltage lines using optical techniques. *Proc. IEEE.* 123(10): 957–960.
- Rogers, A.J. 1977. The electrogyration effect in crystalline quartz. *Proc. Royal. Soc. Lond. A.* 353: 177–192.
- Roukes, M.L. 2000. Nanoelectromechanical systems. Technical digest of the 2000 solid-state sensor and actuator workshop, Hilton Head Island, SC, 6/4-2/2000, 10pp.
- RTK GPS. 2010. High-end DGPS and RTK systems, 4pp Periodic Report: Geospatial Extension Program, Utah State University, Logan, UT. Accessed February 1, 2013 at: <http://extension.usu.edu/nasa/hm/on-target/gps-tutorial>.
- Ruhle. 2009. Inductosyn® applications. Accessed February 1, 2013 at: www.ruhle.com/applicat.htm.
- Saluja, A. and D.S. Kalonia. 2004. Measurement of fluid viscosity at microliter volumes using quartz impedance analysis. *AAPS Pharm. Sci. Tech.* 5(3): 14, Article 47. www.aapspharmacitech.org.
- Saluja, A. and D.S. Kalonia. 2005. Application of ultrasonic shear rheometer to characterize rheological properties of high protein concentration solutions at microliter volume. *J. Pharm. Sci.* 94(6): 1161–1168.
- Schaller, E. 2000. Applications and limits of “electronic noses” in the evaluation of dairy products. Dr. of Technical Sciences Thesis at the ETH, Zurich, Switzerland. Accessed December 10, 2012 at: e-collection.library.ethz.ch/eserv/eth/23664/eth-23664-02.pdf.
- Schaller, E., J.O. Bosset, and F. Escher. 1998. “Electronic noses” and their application to food. *Lebensm. Wiss. u. Technol.* 31: 305–316.
- Scharfeld, T.A. 2001. An analysis of the fundamental constraints on low cost passive radio-frequency identification system design. MS dissertation in mechanical engineering, MIT, Cambridge, MA, 115pp.
- Scheid, F.S. 1990. *2000 Solved Problems in Numerical Analysis*. McGraw-Hill, New York.
- Schmidt, M.A. and R.T. Howe. 1987. Silicon resonant microsensors. *Ceramic Engineering and Science Proceedings*, 8 September 9–October 10, 1987, pp. 1019–1034. Accessed June 11, 2013 at: www-mtl.mit.edu/researchgroups/mems/people/schmidt/conferences/3.SchmidtCESSeptOct87.pdf.
- Schreiber, F. 2000. Structure and growth of self-assembling monolayers. *Prog. Surf. Sci.* 65: 151–256. www.elsevier.com/locate/progsurf.
- Sears, F.W. 1949. *Optics*. Addison-Wesley Press Inc., Cambridge, MA.
- Sears, F.W. 1953. *Electricity and Magnetism*. Addison-Wesley Publishing Co., Cambridge, MA.
- Seaturtle. 2010. *Wildlife Tracking*, Web paper. Accessed May 03, 2013 at: www.wildlifetracking.org/faq.shtml (also see: www.seaturtle.org).
- SensiQ. 2012. Overview of Surface Plasmon Resonance, Revelation A.00. 5pp web tutorial. Accessed December 26, 2012 at: www.discoverensiq.com/uploads/file/support/spr/Overview_of_SPR.pdf.
- Sensorportal. 2013. List of MEMS accelerometers, 3pp. Hotlinks to accelerometer manufacturers. Accessed April 28, 2013 at: www.sensorportal.com/HTML/SENSORS/Accelerometers_Manuf.htm.
- Senturia, S.D. 1987. The role of the MOS structure in integrated sensors. *Sens. Actuat.* 4(4): 507–526.
- Seshia, A.A., R.T. Howe, and S. Montague. 2002. An integrated microelectromechanical resonant output gyroscope. Paper presented at the *15th IEEE International Conference on Micro Electro Mechanical Systems*, January 22–24, pp. 722–726. Accessed April 29, 2013 at: www-eng.lbl.gov/~dw/if/ideas/mems/mems_2002_papers/177.pdf.

- Shafer-Peltier, K.E., C.L. Haynes, M.R. Glucksberg, and R.P. van Duyne. 2002. Toward a glucose biosensor based on surface-enhanced Raman scattering. *J. Am. Chem. Soc.* 125(2): 588–593.
- Shida, L. et al. 1989. SI value of quantized Hall resistance based on ETL's calculable capacitor. *IEEE Trans. Instrum. Meas.* 38(2): 252–255.
- Shields, J.Q., R.F. Dziuba, and H.P. Layer. 1989. New realization of the ohm and farad using the NBS calculable capacitor. *IEEE Trans. Instrum. Meas.* 38(2): 249–251.
- SI. 2006. *The International System of Units*, 8th edn. Institute of Electrical and Electronics Engineers, New York. Accessed June 16, 13 at: www.bipm.org/utis/common/pdf/si_brochure_8_en.pdf.
- Siakavara, K. 2007. A compact fractal microstrip antenna for GPS and terrestrial radio services. *Int. J. Electron* 94(3): 277–283.
- Silicon Sensing Systems, Japan. 2013. Main web page: Accessed May 01, 2013 at: www.sssj.co.jp/en/index.html.
- Simon, H.J. 1998 (December 8). Sensor using long range surface plasmon resonance with diffraction double grating. US Patent No. 5,846, 843.
- Simpson, D. 2003. Universal Serial Bus (USB) information, 11pp. Accessed April 29, 2013 at: www.sss-mag.com/usb.html#overview1.
- Sims, D.W., N. Queiroz et al. 2009. Long-term GPS tracking of ocean sunfish *Mola mola* offers a new direction in fish monitoring. *PLoS One* 4(10): 7. Accessed May 03, 2013 at: www.plosone.org/article/info:doi/10.1371/journal.pone.0007351.
- Sirohi, R.S. and H.P. Kothiyal. 1991. *Optical Components, Systems and Measurement Techniques*. Marcel Dekker, New York.
- Smith, C.H. and R.W. Schneider. 1999. Low-field magnetic sensing with GMR sensors, 13pp. Presented at Sensors EXPO-Baltimore, May, 1999. Accessed April 12, 2012 at: www.nve.com/Downloads/lowfield.pdf.
- Soltani, P.K., D. Wysniewski, and K. Swartz. 1999. Amorphous selenium direct radiography for industrial imaging. *Paper 22. Proceedings Conference Computerized Tomography for Industrial Applications and Image Processing in Radiology*, March 15–17, Berlin, Germany, pp. 123–132.
- Sotzing, G.A. et al. 2000a. Preparation and properties of vapor detector arrays formed from poly(3,4-ethelenedioxy)thiophene-poly(styrene-sulfonate) insulating polymer composites. *Anal. Chem.* 72: 3181–3190.
- Sotzing, G.A. et al. 2000b. Highly sensitive detection and discrimination of biogenic amines utilizing arrays of polyaniline/carbon black composite vapor detectors. *Chem. Mat.* 12: 593–595.
- Sprague Electric Co. 1987. *Hall Effect and Optoelectronic Sensors*. Sprague Semiconductor Group, Concord, NH. Data Book SN-500.
- Stahovec, J.L. 1981. Safety method and apparatus for sensing the presence of individuals adjacent to a vehicle. US Patent No. 4,300,116. November 10, 1981.
- Stampfer, C., T. Helbling, D. Obergfell, B. Schöberle, M.K. Tripp, A. Jungen, S. Roth, V.M. Bright, and C. Hierold. 2006. Fabrication of single-walled carbon-nanotube-based pressure sensors. *Nano Lett.* 6(2): 233–237.
- Staples, E.J. 1999. A new electronic nose. *Sensors* 16: 33–40.
- Stark, H., F.B. Tuteur, and J.B. Anderson. 1988. *Modern Electrical Communications*, 2nd edn. Prentice Hall, Englewood Cliffs, NJ.
- Stefanovska, A. 2007. Coupled oscillations. *IEEE Eng. Med. Biol. Mag.* 26(6): 25–29.
- Sterling. 2013. Homepage for free Sterling software download. Accessed May 15, 2013 at: <http://sterling3.software.informer.com/>.
- STMicroelectronics. 2013. Piezoresistive pressure sensor LPS331AP. Accessed January 29, 2013 at: www.st.com/internet/com/TECHNICAL_RESOURCES/TECHNICAL_LITERATURE/DATASHEET/DM00036916.pdf.
- Stone, H.S. 1982. *Microcomputer Interfacing*. Addison-Wesley Publishing Co., Reading, MA.
- Stout, M.B. 1950. *Basic Electrical Measurements*. Prentice Hall Inc., Englewood Cliffs, NJ.
- Stout, M.B. 1960. *Basic Electrical Measurements*, 2nd edn. Prentice Hall Inc., Englewood Cliffs, NJ.

- Strike, D.J., M.G.H. Meijerink, and M. Koudelka-Hep. 1999. Electronic noses—A mini-review. *Fresenius J. Anal. Chem.* 364: 499–505.
- Su, M., S. Li, and V.P. Dravid. 2003. Microcantilever resonance-based DNA detection with nanoparticle probes. *Appl. Phys. Lett.* 82(20): 3562–3564.
- Su, X., C. Dai, J. Zhang, and S.J. O’Shea. 2002. Quartz tuning fork biosensor. *Biosens. Bioelectron.* 17: 111–117.
- Sun, B., B. Zhang, and M.T. Khan. 2008. Modeling and formulation of a novel microoptomechanical gyroscope. *J. Nanomater.* 2008: 9, Article ID 429168, doi:10.1155/2008/429168.
- Suslick, K.S., W.B. McNamara III, and Y. Didenko. 1999. Hot spot conditions during multi-bubble cavitation. In: *Sonochemistry and Sonoluminescence*, Crum, L.A. et al. eds. Kluwer Publishers, Dordrecht, the Netherlands, pp. 191–204.
- Tan, T.T., V.O. Schmitt, Q. Lucas, and S. Isz. 2001. Electronic noses and electronic tongues. *LabPLus Int.* September/October 16–19.
- Tarr, R.V. and P.G. Steffes. 1993 (September 14). Non-invasive blood glucose measurement system and method using stimulated Raman spectroscopy. US Patent No. 5,243,983.
- Tarr, R.V. and P.G. Steffes. 1998. The noninvasive measure of D-glucose in the ocular aqueous humor using stimulated Raman spectroscopy. *IEEE/LEOS Newslett.* 12(2): 22–27.
- Taub, H. and D.L. Schilling. 1977. *Digital Integrated Electronics*. McGraw-Hill, New York.
- Taylor, B.N. 1990. New international representations of the volt and ohm effective January 1, 1990. *IEEE Trans. Instrum. Meas.* 39(1): 2–5.
- TEG Power. 2013. Expanding your energy options, 3pp web tutorial on TEGs. Accessed June 21, 2013 at: <http://tegpowers.com>.
- Tektronix 1990. *Product Catalog*. Tektronix Inc., Wilsonville, OR.
- Teva, J. et al. 2006. A femtogram resolution mass sensor platform, based on SOI electrostatically driven resonant cantilever. Part I: Electromechanical model and parameter extraction. *Ultramicroscopy* 106: 800–807.
- Thompson, A.M. and D.G. Lampard. 1956. A new theorem in electrostatics and its application to calculable standards of capacitance. *Nature* 177: 188.
- Thompson, H.D. and W.H. Stevenson. 1979. *Laser Velocimetry and Particle Sizing*. Hemisphere Publishing Co., Washington, DC.
- TI. 2005. 38pp datasheet on TLC2642 series chopper-stabilized op amps. Accessed April 10, 2013 at: www.ti.com/lit/ds/symlink/tlc2652.pdf (Rev. 2/05).
- Tienpont, B., F. David, W. Witdouck, D. Vermeersch, H. Stoeri, and P. Sandra. 2008. Features of a micro-gas chromatograph equipped with enrichment device and microchip plasma emission detection (μ PED) for air monitoring. *Lab Chip.* 8: 1819–1828.
- Timms, C., P.S. Thomas, and D.H. Yates. 2012. Detection of gastro-oesophageal reflux disease (GORD) in patients with obstructive lung disease using exhaled breath profiling. (Abstract). *J. Breath Res.* 6(1): 1.
- Tompkins, W.J. and J.G. Webster, eds. 1981. *Design of Microcomputer-Based Medical Instrumentation*. Prentice Hall, Englewood Cliffs, NJ.
- Tompkins, W.J. and J.G. Webster, eds. 1988. *Interfacing Sensors to the IBM® PC*. Prentice Hall, Englewood Cliffs, NJ.
- Townsend, P. 2008. Terahertz radiation. Web paper. Accessed November 20, 2010 at: www.paultownsend.co.uk/research/fundamentals/terahertz-radiation/.
- Tremblay, Y., P.W. Robinson, and D.P. Costa. 2009. A parsimonious approach to modeling animal movement data. *PLoS One* 4(3): 11, e4711.
- TSI, Inc. 2013. *TSI Thermal Anemometry Probes*. Accessed April 16, 2013 at: www.tsi.com/uploadedFiles/_Site_Root/Products/Literature/Catalogs/Hotwire_Catalog_2980465.pdf.
- Turner, A.P.F. and N. Magan. 2004 (February). Electronic noses and disease diagnostics. *Nat. Rev. Microbiol.* 2: 161–166.
- u-blox. 2009a. GPS: Essentials of satellite navigation, 174pp. Document No. GPS-X-02007-D. Accessed February 5, 2013 at: [202.114.120.10/gnss/docs/GPS Compendium \(GPS-X-02007\).pdf](http://202.114.120.10/gnss/docs/GPS%20Compendium%20(GPS-X-02007).pdf).

- u-blox. 2009b. GPS Antennas. Application Note GPS-X-08014-A1. 40pp. Accessed February 5, 2013 at: [www.u-blox.ch/images/downloads/Product_Docs/GPS-Antennas_ApplicationNote\(GPS-X-08014\).pdf](http://www.u-blox.ch/images/downloads/Product_Docs/GPS-Antennas_ApplicationNote(GPS-X-08014).pdf).
- u-blox. 2013. Product catalog 13. Accessed February 1, 2013 at: [www.fontekcorp.com/fontekcorp/Portals/0/U-blox GPS Modules Product List.pdf](http://www.fontekcorp.com/fontekcorp/Portals/0/U-blox%20GPS%20Modules%20Product%20List.pdf).
- UDT. 2012. Non-contact position sensing using optical detectors, 6pp tutorial paper by UDT Instruments. Accessed December 17, 2012 at: www.gamma-sci.com/wp-content/uploads/2012/06/Non-Contact-Position-Sensing-Using-Using-Optical-Detectors.pdf.
- U.S. Coast Guard. 2013. Navigation center: NDGPS general information. Main page. Accessed April 30, 2013 at: www.navcen.uscg.gov/dgps/?pageName=dgpsMain.
- Vashist, S.K. 2007 (May 22). A review of microcantilevers for sensing applications. *Azojono. (J. Nanotechnol. Online)*. 3: 1–15. 10pp. Accessed 5/30/13 at: www.azonano.com/article.aspx?ArticleID=1927.
- van der Ziel, A. 1974. *Introductory Electronics*. Prentice Hall, Englewood Cliffs, NJ.
- Vaughan, M. and J.M. Vaughan. 1989. *The Fabry Perot Interferometer: History, Theory, Practice, and Applications*. Adam Hilger, New York.
- von Klitzing, K. 1986. The quantum Hall effect. In *The Physics of the Two-Dimensional Electron Gas*. J.T. Devreese and F.M. Perters, eds. Plenum Press, New York.
- von Klitzing, K., G. Dorda, and M. Pepper. 1980. New method for high accuracy determination of the fine structure constant based in quantized Hall resistance. *Phys. Rev. Lett.* 45: 494–496.
- Vranish, J.M. 1992a. Capacitive proximity sensor has longer range. NASA Tech Briefs Technical Support Package, GSC-13377. Goddard Space Flight Center, Greenbelt, MD.
- Vranish, J.M. 1992b (August). Capacitive proximity sensor has longer range. NASA Tech Briefs, pp. 22–23.
- Vranish, J.M. 1994 (August). Capacitive sensor with driven shields and bridge circuit. NASA Tech Briefs, pp. 34–35.
- Waggoner, P. et al. 1997. Canine olfactory sensitivity to cocaine hydrochloride and methyl benzoate. In *First Annual Symposium on Enabling Technologies for Law Enforcement*. International Society for Optical Engineering, Boston, MA.
- Wall, C., III and M.S. Weinberg. 2003 (March/April). Balance prosthesis for postural control. *IEEE Eng. Med. Biol. Mag.* 22: 84–90.
- Warmack, R.J., M. Wise, and D. Wolf. 2012 (April). Home smoke alarms: A technology roadmap, 32pp online technical report. Sponsored by: US Fire Administration, FEMA, Oak Ridge National Laboratory, and US Consumer Product Safety Commission. Accessed May 8, 2013 at: www.cpsc.gov/PageFiles/93425/homesmokealarm.pdf.
- Watriasiewicz, B.M. and M.J. Rudd. 1976. *Laser Doppler Measurements*. Butterworths, London, U.K.
- Watson, J. and D. Tanner. 1974. Applications of the Taguchi gas sensor to alarms for inflammable gases. *Radio Electron. Eng.* 44(2): 85–91.
- Webster, J.G. 1978. *Medical Instrumentation: Application and Design*. Houghton-Mifflin, Boston, MA.
- Webster, J.G., ed. 1992. *Medical Instrumentation*, 2nd edn. Houghton-Mifflin, Boston, MA.
- Weeks, B.L. et al. 2003. A microcantilever-based pathogen detector. *Scanning* 25: 297.
- Welham, C.J., J.W. Gardner, and J. Greenwood. 1996. A laterally-driven micromachined resonant pressure sensor. *Sens. Actuat. A*. 52: 86–91.
- Werner, D.H. and S. Ganguly. 2003. An overview of fractal antenna engineering research. *IEEE Antennas Propag. Mag.* 45(1): 38–57.
- Widrow, B. and S.D. Stearns. 1985. *Adaptive Signal Processing*. Prentice Hall, Englewood Cliffs, NJ.
- Wiener, P. and B. Elektronik. 2013. Introduction to VME/VXI/VXS standards, 11pp web paper. Accessed April 25, 2013 at: file.wiener-d.com/documentation/General/WIENER_VME_VXI_VXS_introduction_1.0.pdf.
- Wi-Fi. 2013. Wi-Fi/WLAN channels, frequencies, bands and bandwidths. 8pp web tutorial. Accessed May 15, 2013 at: www.radio-electronics.com/info/wireless/wi-fi/80211-channels-number-frequencies-bandwidth.php.
- Williams, C.S. 1986. *Designing Digital Filters*. Prentice Hall, Englewood Cliffs, NJ.

- Wilson, A.D. and M. Baietto. 2009. Applications and advances in electronic-nose technologies. *Sensors* 9: 5099–5148. doi:10.3390/s90705099.
- Wilson, W.L., Jr. 1996. A sensitive magnetoresistive MEMS acoustic sensor. Final Report. Contract No. 961004. Rice University, Houston, TX, 28pp.
- Wingard, L.B. et al. 1982. Potentiometric measurement of glucose concentration with an immobilized glucose oxidase/catalase electrode. *Diabetes Care* 5(3): 199–202.
- Wise, D.L., ed. 1989. *Applied Biosensors*. Butterworths, Boston, MA.
- Wiza, J.L. 1979. Microchannel plate detectors. *Nucl. Instrum. Methods* 162: 587–601.
- Wohltjen, H. 1984. Chemical microsensors and microinstrumentation. *Anal. Chem.* 56(1): 87A–103A.
- Wolf, S. 1983. Interference signals and their elimination or reduction. *Guide to Electronic Measurements and Laboratory Practice*, 2nd edn. Prentice Hall, Englewood Cliffs, NJ. Chapter 16.
- Wolfbeis, O.S. 1991. *Fiber Optic Sensors and Biosensors*, Vol. 1, CRC Press, Boca Raton, FL.
- Wolfram. 2010. Hilbert curve. Accessed July 28, 2010 at: <http://mathworld.wolfram.com/HilbertCurve.html>.
- Wroblewski, W. 2005. Field effect transistors (FETs) as transducers in electrochemical sensors, 7pp web paper tutorial. Accessed April 30, 2013 at: csrg.ch.pw.edu.pl/tutorials/isfet/.
- Wroblewski, W., A. Dybko, E. Malinowska, and Z. Brzozka. 2004. Towards advanced chemical microsensors. *Talanta* 63: 33–39.
- Wu, G. et al. 2001. Bioassay of prostate-specific antigen (PSA) using microcantilevers. *Nat. Biotechnol.* 19: 856.
- Xu, J.C., G.W. Hunter, D. Lukco, C-C. Liu, and B.J. Ward. 2009. Novel carbon dioxide microsensor based on tin oxide nanomaterial doped with copper oxide. *IEEE Sensors J.* 9(3): 235–236.
- Xu, L., S. Lindfors, and K. Stadius. 2009. A digitally controlled 2.4-GHz oscillator in 65-nm CMOS. *Analog Integr. Circ. Signal Process.* 58(1): 35–42.
- Yang, E.S. 1988. *Microelectronic Devices*. McGraw-Hill Book Co., New York.
- Yang, Y. et al. 2010. Quad-cantilever microsensors with a low-cost single-sided micro-machining technique for trace chemical vapor detection. *Microelectronic Eng.* 87: 2317–2322.
- Yang, Y. and J. Liu. 2010. Evaluation of the power-generation capacity of wearable thermoelectric power generator. *Front. Energy Power Eng. China* 4(3): 346–357.
- Yarov, A. 1991. *Optical Electronics*. Saunders, Philadelphia, PA.
- Yasui, K. 1998. Effect of surfactants on single-bubble sonoluminescence. *Phys. Rev. E.* 58(4): 4560–4567.
- Yazdi, Y., F. Ayozi, and K. Najafi. 1998. Micromachined inertial sensors. *Proc. IEEE.* 86(8): 1640–1658.
- Yeh, Y. and H.Z. Cummins. 1964. Localized fluid flow measurements with a HeNe laser spectrometer. *Appl. Phys. Lett.* 4(10): 176.
- Yoon, H. and J. Jang. 2009. Conducting-polymer nanomaterials for high-performance sensor applications: Issues and challenges. *Adv. Funct. Mater.* 19: 1567–1576.
- Young, P.R. 1996. Basic mass spectroscopy. Organic chemistry online. Accessed May 09, 2013 at: people.stfx.ca/tsmithpa/Chem361/labs/spec/MS.htm.
- Yu, Y.X. and Y. Zhao. 2012. Electronic nose integrated with chemometrics for rapid identification of foodborne pathogen. In *Chemometrics in Practical Applications*. K. Varmuza, ed. InTechOpen, Rijeka, Croatia, Chapter 9. Accessed February 9, 2014 at: cdn.intechopen.com/pdfs/33613/InTech-Electronic-nose-integrated-with-chemometrics-for-rapid-identification-of-foodborne-pathogen.pdf 15 pp.
- Zainud-Deen, S.H., H.A. Malhat, and K.W. Adwalla. 2009. Fractal antennas for passive UHF RFID applications. *Prog. Electromagn. Res. B* 16: 209–228.
- Zareian-Jahromi, M.A. 2009 (May 22). MEMS-based micro gas chromatography: Design, fabrication and characterization. Master of Science Dissertation in Electrical and Computer Engineering at the Virginia Polytechnic University & State University, Blacksburg, VA, 71pp. Accessed May 6, 2013 at: scholar.lib.vt.edu/theses/available/etd-06052009-023407/unrestricted/Master_Thesis_Final_Submitted_7_20_2009.pdf.

- Zellers, E.T., G. Serrano, H. Chang, and L.K. Amos. 2011. A micro gas chromatograph for high speed determination of explosive markers. *Proceeding of the Transducers Conference*, Beijing, China, June 5–9, 2011, pp. 2082–2085. Accessed December 07, 2012 at: wims2.org/publications/papers/1308056241-zellers_amos_transducers_11.pdf.
- Zhang, J. et al. 2006. Rapid and label-free nanomechanical detection of biomarker transcripts in human RNA. *Nat. Nanotechnol.* 1: 214–220. doi: 10.1038/nnano.2006.134.
- Zheng, J.P. 2002. The limitation of energy density for battery /double layer capacitor hybrid cell. In *Proc. Symposium on Advances in Electrochemical Capacitors & Hybrid Power Systems. Electrochemical Society Proceedings*, R.J. Brodd, ed. 1p. Accessed December 14, 2011 at: www.electrochem.org/dl/ma/201/pdfs/0228.pdf.
- ZigBee. 2013. *Zigbee Tutorial*. Accessed May 13, 2013 at: www.tutorial-reports.com/wireless/zigbee/tutorial.php.
- Zijlstra, P., P.M.R. Paulo, and M. Orrit. 2012 (June). Optical detection of single non-absorbing molecules using the surface plasmon resonance of a gold nanorod. *Nat. Nanotechnol.* 7: 379–382.
- Zuch, E.L. 1981. *Data Acquisition and Conversion Handbook*, 4th edn. Datel-Intersil, Inc., Mansfield, MA.
- Zuo, G.M. et al. 2001. Dual-SAM functionalization on integrated cantilevers for specific trace-explosive sensing and non-specific adsorption suppression. *Nanotechnology* 18: 255501.

Weighing in on the growth of innovative technologies, the adoption of new standards, and the lack of educational development as it relates to current and emerging applications, the third edition of **Introduction to Instrumentation and Measurements** uses the authors' 40 years of teaching experience to expound on the theory, science, and art of modern instrumentation and measurements (I&M).

What's New in This Edition

This edition includes material on modern integrated circuit (IC) and photonic sensors, micro-electro-mechanical (MEM) and nano-electro-mechanical (NEM) sensors, chemical and radiation sensors, signal conditioning, noise, data interfaces, and basic digital signal processing (DSP), and upgrades every chapter with the latest advancements. It contains new material on the designs of micro-electro-mechanical (MEMS) sensors, adds two new chapters on wireless instrumentation and microsensors, and incorporates extensive biomedical examples and problems.

Containing 13 chapters, this third edition

- Describes sensor dynamics, signal conditioning, and data display and storage
- Focuses on means of conditioning the analog outputs of various sensors
- Considers noise and coherent interference in measurements in depth
- Covers the traditional topics of DC null methods of measurement and AC null measurements
- Examines Wheatstone and Kelvin bridges and potentiometers
- Explores the major AC bridges used to measure inductance, Q , capacitance, and D
- Presents a survey of sensor mechanisms
- Includes a description and analysis of sensors based on the giant magnetoresistive effect (GMR) and the anisotropic magnetoresistive (AMR) effect
- Provides a detailed analysis of mechanical gyroscopes, clinometers, and accelerometers
- Contains the classic means of measuring electrical quantities
- Examines digital interfaces in measurement systems
- Defines digital signal conditioning in instrumentation
- Addresses solid-state chemical microsensors and wireless instrumentation
- Introduces mechanical microsensors (MEMS and NEMS)
- Details examples of the design of measurement systems

Introduction to Instrumentation and Measurements is written with practicing engineers and scientists in mind, and is intended to be used in a classroom course or as a reference. It is assumed that the reader has taken core EE curriculum courses or their equivalents.

K20793

 **CRC Press**
Taylor & Francis Group
an informa business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK



