


FOUNDATIONS OF ELECTROMAGNETIC THEORY

FOURTH EDITION

REITZ • MILFORD • CHRISTY



Digitized by the Internet Archive
in 2022 with funding from
Kahle/Austin Foundation

Foundations of Electromagnetic Theory

FOURTH EDITION

JOHN R. REITZ

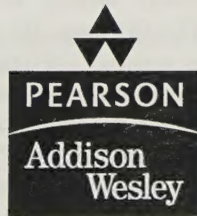
Research Laboratory, Ford Motor Company

FREDERICK J. MILFORD

Battelle Memorial Institute

ROBERT W. CHRISTY

Dartmouth College



San Francisco Boston New York
Cape Town Hong Kong London Madrid Mexico City
Montreal Munich Paris Singapore Sydney Tokyo Toronto

Publisher Jim Smith
Sponsoring Editor Stuart Johnson
Editorial Assistant Jennifer Duggan
Managing Editor Kazia Navas
Production Supervisor Jennifer Brownlow Bagdigian
Production Coordinator Patricia A. Oduor
Text Design and Editorial Production Services Sylvia Dovner/Technical Texts
Cover Designer Eileen Hoff
Executive Marketing Manager Scott Dustan
Manufacturing Buyer Jeffrey Sargent
Prepress Buyer Sarah McCracken
Technical Art Supervisor Joe Vetere
Compositor Universities Press
Printer Bradford & Bigelow
Cover Photo Roman Kruchuk/Shutterstock

ISBN-10 0-201-52624-7 (casebound edition)
ISBN-10 0-321-58174-1 (paperback edition)
ISBN-13 978-0-321-58174-7 (paperback edition)

Library of Congress Cataloging-in-Publication Data

Reitz, John R.

Foundations of electromagnetic theory/John R. Reitz, Frederick J. Milford,
Robert W. Christy.—4th ed.

p. cm.

Includes bibliographical references and index.

ISBN 0-201-52624-7

I. Electromagnetic theory. I. Milford, Frederick J. II. Christy, Robert W. III. Title

QC670.R4 1992

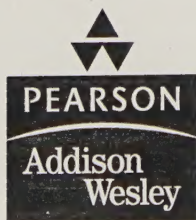
530.1'41—dc20

91-36491

CIP

Copyright ©1993, 2009 Pearson Education, Inc., publishing as Pearson Addison-Wesley, 1301 Sansome St., San Francisco, CA 94111. All rights reserved. Manufactured in the United States of America. This publication is protected by Copyright and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. To obtain permission(s) to use material from this work, please submit a written request to Pearson Education, Inc., Permissions Department, 1900 E. Lake Ave., Glenview, IL 60025. For information regarding permissions, call (847) 486-2635.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and the publisher was aware of a trademark claim, the designations have been printed in initial caps or all caps.



Preface

Although Maxwell's equations were formulated over one hundred years ago, the subject of electromagnetism has not remained static. Advanced undergraduate students in science, to whom we are directing our attention, today approach the subject with a qualitative understanding of atomic phenomena. At the same time, they have acquired a good background in mathematics and for the first time are in a position to solve some of the important problems of classical physics. In addition, many of these students are computer literate. The present volume evolved from the teaching of courses in electricity and magnetism to physics majors at Case Institute of Technology and at Dartmouth College. A course in electromagnetism is ideally suited to a development of the concepts of vector analysis, partial differential equations, boundary-value problems, and numerical computations via the desktop computer. The sections involving these techniques are written in such a way that little previous knowledge of the subject is required.

We believe that building up electricity and magnetism from the basic experimental laws is the correct approach at the intermediate level, and we have followed this approach through a rigorous exposition of the fundamentals. We have also been careful to include a number of appropriate examples to bridge the gap between the formal development of the subject and the problems. A full understanding of the electric and magnetic fields inside matter can be obtained only after the atomic nature of materials is appreciated. Hence we have used elementary atomic concepts freely in the development of macroscopic theory. The macroscopic electromagnetic behavior of isotropic dielectrics, conductors, magnetic materials, plasmas, and superconductors is treated in separate chapters (Chapters 4, 7, 9, 14, and 15, respectively), and a simple discussion of the microscopic theory of these classes of matter (except superconductors) is also provided (in Chapters 5, 7, 10, and 14). This coverage is extended to time-dependent fields in Chapter 19.

Much of modern physics (and engineering) involves time-dependent electromagnetic fields in which Maxwell's displacement current plays a crucial role. Chapters 16 through 20 develop the application to electromagnetic waves and wave propagation—including the connection with optical phenomena, which is the frequency range that has displaced microwaves in technological interest. The relativistic formulation of electromagnetic theory has been put into the last chapter (Chapter 22), although it can be taken up at any point after Chapter 16.

This fourth edition of the book has been changed in a number of ways. The most prominent change is the inclusion of a self-contained discussion of the solution of numerical problems in electromagnetism by means of the personal computer (Appendix I) and the addition of sections devoted to the numerical solution of Laplace's equation, other differential equations, and electric circuit problems (Sections 3-12, 13-11, and 15-4). Several computer routines written in BASIC language are provided; for example, for the solution of a system of linear equations and the finite-difference method for the solution of a differential equation. (If the reader chooses not to include this computer-related material, it is easily bypassed, since this material is localized in Appendix I and in the sections indicated; it is not essential to other parts of the book.) The sections on propagation of electromagnetic waves have been partially rewritten and include new sections (in Chapter 19) on linearity and causality and on propagation in dispersive media.

Other changes from earlier editions include more worked examples, about thirty additional figures, and an update of the materials chapters—particularly Chapters 14 and 15. All tables of data and references to other books have been updated; a new table on the properties of superconductors has been added. About fifty new problems have been added to the problem sets at the ends of chapters.

As an aid to the reader, the more difficult problems are labeled with an asterisk. Sections and chapters in the text that are starred are not essential to its further development and may be omitted in an abbreviated study. A summary section at the end of each chapter identifies key ideas and formulas.

Contents

* Starred sections and chapters may be omitted without loss of continuity

1 Vector Analysis	1	
1-1 Definitions	1	
1-2 Vector Algebra	2	
1-3 Gradient	5	
1-4 Vector Integration	9	
1-5 Divergence	12	
1-6 Curl	14	
1-7 The Vector Differential Operator ∇	17	
1-8 Further Developments	19	
1-9 Summary	22	
2 Electrostatics	26	
2-1 Electric Charge	26	
2-2 Coulomb's Law	27	
2-3 The Electric Field	31	
2-4 The Electrostatic Potential	33	
2-5 Conductors and Insulators	36	
2-6 Gauss's Law	37	
2-7 Application of Gauss's Law	40	
2-8 The Electric Dipole	43	
2-9 Multipole Expansion of Electric Fields	46	
2-10 The Dirac Delta Function	48	
2-11 Summary	50	
3 Solution of Electrostatic Problems	56	
3-1 Poisson's Equation	56	
3-2 Laplace's Equation	58	
3-3 Laplace's Equation in One Independent Variable	60	
3-4 Solutions to Laplace's Equation in Spherical Coordinates: Zonal Harmonics	60	
3-5 Conducting Sphere in a Uniform Electric Field	63	
3-6 Solutions to Laplace's Equation in Cylindrical Coordinates: Cylindrical Harmonics	65	
*3-7 Laplace's Equation in Rectangular Coordinates	66	
*3-8 Laplace's Equation in Two Dimensions: General Solution	67	
3-9 Electrostatic Images	69	
3-10 Line Charges and Line Images	74	
3-11 System of Conductors: Coefficients of Potential	75	
3-12 Numerical Solution of Electrostatic Problems	77	
3-13 Solutions of Poisson's Equation	91	
3-14 Summary	92	
4 The Electrostatic Field in Dielectric Media	97	
4-1 Polarization	98	
4-2 Field Outside of a Dielectric Medium	99	
4-3 The Electric Field Inside a Dielectric	103	
4-4 Gauss's Law in a Dielectric: The Electric Displacement	106	
4-5 Electric Susceptibility and Dielectric Constant	109	
4-6 Point Charge in a Dielectric Fluid	110	
4-7 Boundary Conditions on the Field Vectors	112	
4-8 Boundary-Value Problems Involving Dielectrics	115	
*4-9 Force on a Point Charge Embedded in a Dielectric	118	
4-10 Method of Images for Problems Involving Dielectrics	120	
4-11 Summary	122	

5 Microscopic Theory of Dielectrics	127	8-2 Forces on Current-Carrying Conductors	193
5-1 Molecular Field in a Dielectric	127	8-3 The Law of Biot and Savart	197
5-2 Induced Dipoles: A Simple Model	131	8-4 Elementary Applications of the Biot and Savart Law	199
5-3 Polar Molecules: The Langevin-Debye Formula	132	8-5 Ampere's Circuital Law	204
*5-4 Permanent Polarization: Ferroelectricity	136	8-6 The Magnetic Vector Potential	207
5-5 Summary	138	8-7 The Magnetic Field of a Distant Circuit	208
6 Electrostatic Energy	141	8-8 The Magnetic Scalar Potential	210
6-1 Potential Energy of a Group of Point Charges	142	8-9 Magnetic Flux	212
6-2 Electrostatic Energy of a Charge Distribution	143	8-10 Summary	212
6-3 Energy Density of an Electrostatic Field	146	9 Magnetic Properties of Matter	218
6-4 Energy of a System of Charged Conductors: Coefficients of Potential	148	9-1 Magnetization	218
6-5 Coefficients of Capacitance and Induction	150	9-2 The Magnetic Field Produced by Magnetized Material	222
6-6 Capacitors	151	9-3 Magnetic Scalar Potential and Magnetic Pole Density	225
6-7 Forces and Torques	154	9-4 Sources of the Magnetic Field: Magnetic Intensity	226
6-8 Summary	157	9-5 The Field Equations	227
7 Electric Current	162	9-6 Magnetic Susceptibility, Permeability, and Hysteresis	229
7-1 Nature of the Current	162	9-7 Boundary Conditions on the Field Vectors	235
7-2 Current Density: Equation of Continuity	165	9-8 Boundary-Value Problems Involving Magnetic Materials	237
7-3 Ohm's Law: Conductivity	167	9-9 Current Circuits Containing Magnetic Media	242
7-4 Steady Currents in Continuous Media	170	*9-10 Magnetic Circuits	245
7-5 Approach to Electrostatic Equilibrium	173	*9-11 Magnetic Circuits Containing Permanent Magnets	248
7-6 Resistance Networks and Kirchhoff's Laws	175	9-12 Summary	251
7-7 Microscopic Theory of Conduction	181	*10 Microscopic Theory of Magnetism	256
7-8 Summary	185	10-1 Molecular Field Inside Matter	257
8 The Magnetic Field of Steady Currents	190	10-2 Origin of Diamagnetism	259
8-1 The Definition of Magnetic Induction	190	10-3 Origin of Paramagnetism	261
		10-4 Theory of Ferromagnetism	263
		10-5 Ferromagnetic Domains	265
		10-6 Ferrites	268
		10-7 Summary	268

- 11 Electromagnetic Induction 271**
- 11-1 Electromagnetic Induction 271
 - 11-2 Self-Inductance 277
 - 11-3 Mutual Inductance 279
 - 11-4 The Neumann Formula 280
 - 11-5 Inductances in Series and in Parallel 281
 - 11-6 Summary 284
- 12 Magnetic Energy 289**
- 12-1 Magnetic Energy of Coupled Circuits 290
 - 12-2 Energy Density in the Magnetic Field 292
 - 12-3 Forces and Torques on Rigid Circuits 294
 - *12-4 Hysteresis Loss 298
 - 12-5 Summary 301
- 13 Slowly Varying Currents 305**
- 13-1 Transient and Steady-State Behavior 306
 - 13-2 Kirchhoff's Laws 307
 - 13-3 Elementary Transient Behavior 308
 - 13-4 Steady-State Behavior of a Simple Series Circuit 312
 - 13-5 Series and Parallel Connection of Impedances 314
 - 13-6 Power and Power Factors 316
 - 13-7 Resonance 317
 - *13-8 Mutual Inductances in ac Circuits 320
 - *13-9 Mesh and Nodal Equations 324
 - *13-10 Driving Point Impedance and Transfer Functions 328
 - *13-11 Solving Network Equations with the Computer 329
 - 13-12 Summary 336
- *14 Physics of Plasmas 341**
- 14-1 Theoretical Approaches to Plasma Dynamics 342
 - 14-2 Electrical Neutrality in a Plasma 343
 - 14-3 Particle Orbits and Drift Motion in a Plasma 344
 - 14-4 Magnetic Mirrors 349
 - 14-5 The Hydromagnetic Equations 350
 - 14-6 The Pinch Effect 353
 - 14-7 Magnetic Confinement Systems for Controlled Thermonuclear Fusion 355
 - 14-8 Plasma Oscillations and Wave Motion 358
 - 14-9 Summary 362
- 15 Electromagnetic Properties of Superconductors 365**
- 15-1 The History of Superconductivity 365
 - 15-2 Perfect Conductivity and Perfect Diamagnetism of Superconductors 370
 - 15-3 Examples Involving Perfect Flux Exclusion 371
 - 15-4 The London Equations 374
 - 15-5 Summary 384
- 16 Maxwell's Equations 386**
- 16-1 The Generalization of Ampere's Law: Displacement Current 386
 - 16-2 Maxwell's Equations and Their Empirical Basis 389
 - 16-3 Electromagnetic Energy 390
 - 16-4 The Wave Equation 394
 - 16-5 Monochromatic Waves 395
 - 16-6 Boundary Conditions 398
 - 16-7 The Wave Equation with Sources 402
 - 16-8 Summary 408
- 17 Propagation of Monochromatic Electromagnetic Waves 412**
- 17-1 Plane Monochromatic Waves in Nonconducting Media 412
 - 17-2 Polarization 418
 - 17-3 Energy Density and Flux 421
 - 17-4 Plane Monochromatic Waves in Conducting Media 423
 - *17-5 Spherical Waves 430
 - 17-6 Summary 437

18 Monochromatic Waves in Bounded Regions 441

- 18-1 Reflection and Refraction at the Boundary of Two Nonconducting Media: Normal Incidence 441
- 18-2 Reflection and Refraction at the Boundary of Two Nonconducting Media: Oblique Incidence 445
- 18-3 Brewster's Angle: Critical Angle 451
- 18-4 Complex Fresnel Coefficients: Reflection from a Conducting Plane 454
- 18-5 Reflection and Transmission by a Thin Layer: Interference 462
- 18-6 Propagation Between Parallel Conducting Plates 469
- 18-7 Waveguides 474
- 18-8 Cavity Resonators 477
- 18-9 Summary 479

19 Dispersion and Oscillating Fields in Dispersive Media 484

- 19-1 Linearity and Causality 484
- *19-2 Frequency Response and Dispersion Relations 491
- 19-3 Drude-Lorentz Harmonic Oscillator Model 494
- 19-4 Resonance Absorption by Bound Charges 499
- 19-5 The Drude Free-Electron Theory 507
- *19-6 Dielectric Relaxation 513
- *19-7 Oscillating Fields in Dispersive Media 517
- 19-8 Summary 521

20 The Emission of Radiation 525

- 20-1 Radiation from an Oscillating Dipole 525
- 20-2 Radiation from a Half-Wave Antenna 529
- 20-3 Radiation from a Group of Moving Charges 531
- *20-4 Near and Intermediate Zone Fields 535

- 20-5 Radiation Damping: Thomson Cross Section 537
- 20-6 Summary 541

21 Electrodynamics 545

- 21-1 The Lienard-Wiechert Potentials 545
- 21-2 The Field of a Uniformly Moving Point Charge 548
- 21-3 The Field of an Accelerated Point Charge 551
- 21-4 Radiation Fields for Small Velocities 554
- 21-5 Summary 556

22 The Special Theory of Relativity 558

- 22-1 Physics Before 1900 558
- 22-2 The Lorentz Transformation and Einstein's Postulates of Special Relativity 561
- 22-3 Geometry of Space-Time 566
- 22-4 The Lorentz Transformation as an Orthogonal Transformation 567
- 22-5 Covariant Form of the Electromagnetic Equations 569
- 22-6 Transformation Law for the Electromagnetic Field 573
- 22-7 The Field of a Uniformly Moving Point Charge 574
- 22-8 Summary 576

Appendix I

- Computers and Electromagnetic Theory 579

Appendix II

- Coordinate Transformations, Vectors, and Tensors 594

Appendix III

- Systems of Units 599

Appendix IV

- Vector Differential Operators 603

Appendix V

- The Dirac Delta Function 605

Appendix VI

Fourier Integrals (Fourier
Transforms) 607

Appendix VII

Functions of a Complex Variable:
Dispersion Relations 610

**Answers to Odd-Numbered
Problems 619**

Index 624

Vector Analysis

In the study of electricity and magnetism, a great saving in complexity of notation may be accomplished by using the notation of vector analysis. In providing this valuable shorthand, vector analysis also brings to the forefront the physical ideas involved in equations. It is the purpose of this chapter to give a brief but self-contained exposition of basic vector analysis and to provide a rather utilitarian knowledge of the field that is required for a treatment of electricity and magnetism. Readers already familiar with vector analysis will find this chapter a useful review and an introduction to the notation of the text.

1-1 DEFINITIONS

In the study of elementary physics, several kinds of quantities are encountered; in particular, we distinguish between vectors and scalars. For our purposes, it is sufficient to define a scalar as follows.

A *scalar* is a quantity that is completely characterized by its magnitude.

Examples of scalars are numerous: mass, time, volume, etc. A simple extension of the idea of a scalar is a *scalar field*—a function of position that is completely specified by its magnitude at all points in space.

A vector may be defined as follows.

A *vector* is a quantity that is completely characterized by its magnitude and direction.

As examples of vectors, we cite position from a fixed origin, velocity, acceleration, force, etc. The generalization to a *vector field* gives a function of position that is completely specified by its magnitude and direction at all points in space.

These definitions may be refined and extended; in fact, in Appendix II they are replaced by more subtle definitions in terms of transformation properties. In addition, more complicated kinds of quantities, such as tensors, may be defined. Scalars and vectors will, however, largely suffice for our purposes until Chapter 22.

1-2

VECTOR ALGEBRA

Since the algebra of scalars is familiar to the reader, this algebra will be used to develop vector algebra. In order to proceed with this development it is convenient to have a representation of vectors, for which purpose we introduce a three-dimensional Cartesian coordinate system. This three-dimensional system will be denoted by the three variables x, y, z or, when it is more convenient, x_1, x_2, x_3 . With respect to this coordinate system, a vector is specified by its x -, y -, and z -components. Thus a vector* \mathbf{V} is specified by its components V_x, V_y, V_z , where $V_x = |\mathbf{V}| \cos \alpha_1$, $V_y = |\mathbf{V}| \cos \alpha_2$, $V_z = |\mathbf{V}| \cos \alpha_3$, the α 's being the angles between \mathbf{V} and appropriate coordinate axes. The scalar $|\mathbf{V}| = \sqrt{V_x^2 + V_y^2 + V_z^2}$ is the *magnitude* of the vector \mathbf{V} , or its *length*. In the case of vector fields, each of the components is to be regarded as a function of x, y , and z . It should be emphasized at this point that we introduce a representation of the vectors with respect to a Cartesian coordinate system only for simplicity and ease of understanding; all of the definitions and operations are, in fact, independent of any special choice of coordinates.

The sum of two vectors is defined as the vector whose components are the sums of the corresponding components of the original vectors. Thus if \mathbf{C} is the sum of \mathbf{A} and \mathbf{B} , we write

$$\mathbf{C} = \mathbf{A} + \mathbf{B} \quad (1-1)$$

and

$$C_x = A_x + B_x, \quad C_y = A_y + B_y, \quad C_z = A_z + B_z \quad (1-2)$$

This definition of the vector sum is completely equivalent to the familiar parallelogram rule for vector addition.

Vector subtraction is defined in terms of the negative of a vector, which is the vector whose components are the negatives of the corresponding

* Vector quantities are denoted by boldface symbols.

components of the original vector. Thus if \mathbf{A} is a vector, $-\mathbf{A}$ is defined by

$$(-\mathbf{A})_x = -A_x, \quad (-\mathbf{A})_y = -A_y, \quad (-\mathbf{A})_z = -A_z \quad (1-3)$$

The operation of subtraction is then defined as the addition of the negative and is written

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B}) \quad (1-4)$$

Since the addition of real numbers is associative and commutative, it follows that vector addition (and subtraction) is also associative and commutative. In vector notation this appears as

$$\begin{aligned} \mathbf{A} + (\mathbf{B} + \mathbf{C}) &= (\mathbf{A} + \mathbf{B}) + \mathbf{C} \\ &= (\mathbf{A} + \mathbf{C}) + \mathbf{B} = \mathbf{A} + \mathbf{B} + \mathbf{C} \end{aligned} \quad (1-5)$$

In other words, the parentheses are not needed, as indicated by the last form.

Proceeding now to the process of multiplication, we note that the simplest product is a scalar times a vector. This operation results in a vector, each component of which is the scalar times the corresponding component of the original vector. If c is a scalar and \mathbf{A} is a vector, the product $c\mathbf{A}$ is a vector, $\mathbf{B} = c\mathbf{A}$, defined by

$$B_x = cA_x, \quad B_y = cA_y, \quad B_z = cA_z \quad (1-6)$$

It is clear that if \mathbf{A} is a *vector field* and c a *scalar field*, then \mathbf{B} is a new vector field that is *not* necessarily a constant multiple of the original field.

If two vectors are to be multiplied, there are two possibilities, known as the vector and scalar products. Considering first the scalar product, we note that this name derives from the scalar nature of the product, although the alternative names, *inner product* and *dot product*, are sometimes used. The definition of the scalar product, written $\mathbf{A} \cdot \mathbf{B}$, is

$$\mathbf{A} \cdot \mathbf{B} = A_x B_x + A_y B_y + A_z B_z \quad (1-7)$$

This definition is equivalent to another, and perhaps more familiar, definition—that is, as the product of the magnitudes of the original vectors times the cosine of the angle between these vectors. If \mathbf{A} and \mathbf{B} are perpendicular to each other,

$$\mathbf{A} \cdot \mathbf{B} = 0$$

The scalar product is commutative. The length of \mathbf{A} is

$$|\mathbf{A}| = \sqrt{\mathbf{A} \cdot \mathbf{A}}$$

The vector product of two vectors is a vector, which accounts for the name. Alternative names are *outer product* and *cross product*. The vector product is written $\mathbf{A} \times \mathbf{B}$. If \mathbf{C} is the vector product of \mathbf{A} and \mathbf{B} , then

$\mathbf{C} = \mathbf{A} \times \mathbf{B}$, or

$$\begin{aligned} C_x &= A_y B_z - A_z B_y, & C_y &= A_z B_x - A_x B_z, \\ C_z &= A_x B_y - A_y B_x \end{aligned} \quad (1-8)$$

It is important to note that the cross product depends on the order of the factors; interchanging the order introduces a minus sign:

$$\mathbf{B} \times \mathbf{A} = -\mathbf{A} \times \mathbf{B}$$

Consequently,

$$\mathbf{A} \times \mathbf{A} = 0$$

This definition is equivalent to the following: The vector product is the product of the magnitudes times the sine of the angle between the original vectors, with the direction given by a right-hand screw rule.*

The vector product may be easily remembered in terms of a determinant. If \mathbf{i} , \mathbf{j} , and \mathbf{k} are unit vectors—that is, vectors of unit magnitude, in the x -, y -, and z -directions, respectively—then

$$\mathbf{A} \times \mathbf{B} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix} \quad (1-9)$$

If this determinant is evaluated by the usual rules, the result is precisely our definition of the cross product.

The preceding algebraic operations may be combined in many ways. Most of the results so obtained are obvious; however, there are two triple products of sufficient importance to merit explicit mention. The triple scalar product $D = \mathbf{A} \cdot \mathbf{B} \times \mathbf{C}$ is easily found to be given by the determinant

$$D = \mathbf{A} \cdot \mathbf{B} \times \mathbf{C} = \begin{vmatrix} A_x & A_y & A_z \\ B_x & B_y & B_z \\ C_x & C_y & C_z \end{vmatrix} = -\mathbf{B} \cdot \mathbf{A} \times \mathbf{C} \quad (1-10)$$

This product is unchanged by an exchange of dot and cross or by a cyclic permutation of the three vectors. Note that parentheses are not needed, since the cross product of a scalar and a vector is undefined.

The other interesting triple product is the triple vector product $\mathbf{D} = \mathbf{A} \times (\mathbf{B} \times \mathbf{C})$. By a repeated application of the definition of the cross product, Eq. (1-8), we find

$$\mathbf{D} = \mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B}) \quad (1-11)$$

* Let \mathbf{A} be rotated into \mathbf{B} through the smallest possible angle. A right-hand screw rotated in this manner will advance in a direction perpendicular to both \mathbf{A} and \mathbf{B} ; this direction is the direction of $\mathbf{A} \times \mathbf{B}$.

which is frequently known as the *back cab rule*. It should be noted that in the cross product the parentheses are vital; without them the product is not well defined.

At this point one might well inquire as to the possibility of vector division. Division of a vector by a scalar can, of course, be defined as multiplication by the reciprocal of the scalar. Division of a vector by another vector, however, is possible only if the two vectors are parallel. On the other hand, it is possible to write general solutions to vector equations and so accomplish something closely akin to division. Consider the equation

$$c = \mathbf{A} \cdot \mathbf{X} \quad (1-12)$$

where c is a known scalar, \mathbf{A} is a known vector, and \mathbf{X} is an unknown vector. A general solution to this equation is

$$\mathbf{X} = \frac{c\mathbf{A}}{\mathbf{A} \cdot \mathbf{A}} + \mathbf{B} \quad (1-13)$$

where \mathbf{B} is an arbitrary vector that is perpendicular to \mathbf{A} —that is, $\mathbf{A} \cdot \mathbf{B} = 0$. What we have done is very nearly to divide c by \mathbf{A} ; more correctly, we have found the general form of the vector \mathbf{X} that satisfies Eq. (1-12). There is no unique solution, and this fact accounts for the vector \mathbf{B} . In the same fashion, we may consider the vector equation

$$\mathbf{C} = \mathbf{A} \times \mathbf{X} \quad (1-14)$$

where \mathbf{A} and \mathbf{C} are known vectors and \mathbf{X} is an unknown vector. The general solution of this equation is

$$\mathbf{X} = \frac{\mathbf{C} \times \mathbf{A}}{\mathbf{A} \cdot \mathbf{A}} + k\mathbf{A} \quad (1-15)$$

where k is an arbitrary scalar. Thus, \mathbf{X} as defined by Eq. (1-15) is very nearly the quotient of \mathbf{C} by \mathbf{A} ; the scalar k takes account of the nonuniqueness of the process. If \mathbf{X} is required to satisfy both Eqs. (1-12) and (1-14), then the result is unique (if it exists) and is given by

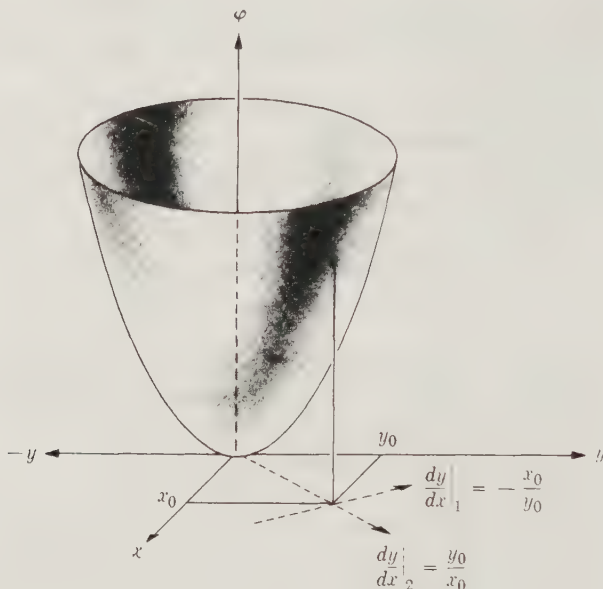
$$\mathbf{X} = \frac{\mathbf{C} \times \mathbf{A}}{\mathbf{A} \cdot \mathbf{A}} + \frac{c\mathbf{A}}{\mathbf{A} \cdot \mathbf{A}} \quad (1-16)$$

1-3 GRADIENT

The extensions of the ideas introduced in the previous section to differentiation and integration, i.e., vector calculus, will now be considered. The simplest of these is the relation of a particular vector field to the derivatives of a scalar field. It is convenient first to introduce the idea of the *directional derivative* of a function of several variables, which is just the rate of change of the function in a specified direction. The directional derivative

FIGURE 1-1

The function $\varphi(x, y) = x^2 + y^2$ plotted against x and y in a three-dimensional graph



of a scalar function φ is usually denoted by $d\varphi/ds$; it must be understood that ds represents an infinitesimal displacement in the direction being considered, and that ds is the scalar magnitude of ds . If ds has the components dx , dy , dz , then

$$\begin{aligned} \frac{d\varphi}{ds} &= \lim_{\Delta s \rightarrow 0} \frac{\varphi(x + \Delta x, y + \Delta y, z + \Delta z) - \varphi(x, y, z)}{\Delta s} \\ &= \frac{\partial \varphi}{\partial x} \frac{dx}{ds} + \frac{\partial \varphi}{\partial y} \frac{dy}{ds} + \frac{\partial \varphi}{\partial z} \frac{dz}{ds} \end{aligned}$$

In order to clarify the idea of a directional derivative, consider a scalar function of two variables. Thus, $\varphi(x, y)$ represents a two-dimensional scalar field. We may plot φ as a function of x and y as is done in Fig. 1-1 for the function $\varphi(x, y) = x^2 + y^2$. The directional derivative at the point x_0, y_0 depends on the direction. If we choose the direction corresponding to $dy/dx = -x_0/y_0$, then we find

$$\left. \frac{d\varphi}{ds} \right|_{x_0, y_0} = \frac{\partial \varphi}{\partial x} \frac{dx}{ds} + \frac{\partial \varphi}{\partial y} \frac{dy}{ds} = \left[2x_0 - 2y_0 \frac{x_0}{y_0} \right] \frac{dx}{ds} = 0 \quad (1-17a)$$

Alternatively, if we choose $dy/dx = y_0/x_0$, we find

$$\left. \frac{d\varphi}{ds} \right|_{x_0, y_0} = \left(2x_0 + 2 \frac{y_0^2}{x_0} \right) \sqrt{\frac{x_0^2}{x_0^2 + y_0^2}} = 2\sqrt{x_0^2 + y_0^2} \quad (1-17b)$$

since $ds = \sqrt{(dx)^2 + (dy)^2}$. As a third possibility, choose $dy/dx = \alpha$; then

$$\left. \frac{d\varphi}{ds} \right|_{x_0, y_0} = (2x_0 + 2\alpha y_0)(1 + \alpha^2)^{-1/2} \quad (1-17c)$$

If this result is differentiated with respect to α and the derivative is set equal to zero, then the value of α , for which the derivative is a maximum or minimum, is found. When we perform these operations, we obtain $\alpha = y_0/x_0$, which simply means that the direction of maximum rate of change of the function $\varphi = x^2 + y^2$ is the radial direction. If the direction is radially outward, then the maximum is the maximum rate of increase; if it is radially inward, it is a maximum rate of decrease or minimum rate of increase. In the direction specified by $dy/dx = -x_0/y_0$, the rate of change of $x^2 + y^2$ is zero. This direction is tangent to the circle $x^2 + y^2 = x_0^2 + y_0^2$. Clearly, on this curve, $\varphi = x^2 + y^2$ does not change. The direction in which $d\varphi/ds$ vanishes gives the direction of the curve $\varphi = \text{constant}$ through the point being considered. These lines, which are circles for the function $x^2 + y^2$, are completely analogous to the familiar contour lines or lines of constant altitude that appear on topographic maps. Figure 1-2 shows the function $\varphi = x^2 + y^2$ replotted as a contour map.

The idea of contour lines may be generalized to a function of three variables, in which case the surfaces, $\varphi(x, y, z) = \text{constant}$, are called *level surfaces* or *equipotential surfaces*. The three-dimensional analog to Fig. 1-2 is the only practical way of graphing a scalar field for a three-dimensional space. The gradient of a scalar function may now be defined as follows.

The *gradient* of a scalar function φ is a vector whose magnitude is the maximum directional derivative at the point being considered and whose direction is the direction of the maximum directional derivative at the point.

FIGURE 1-2
The function $\varphi(x, y)$ of Fig. 1-1 expressed as a contour map in two dimensions

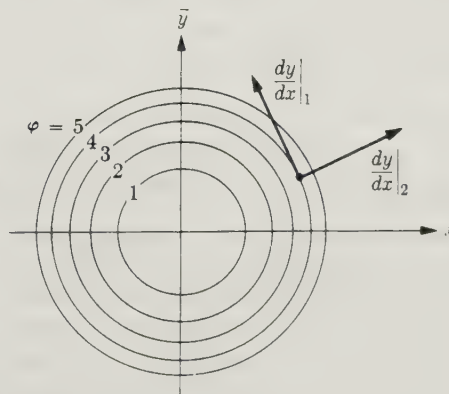
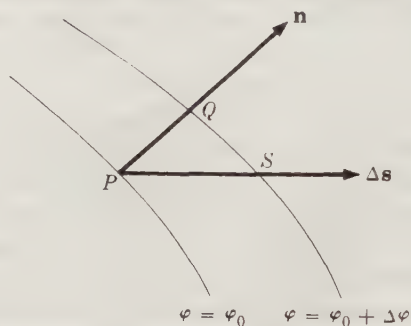


FIGURE 1-3

Parts of two level surfaces of the function $\varphi(x, y, z)$. $|\mathbf{grad} \varphi|$ at P equals the limit as $PQ \rightarrow 0$ of $\Delta\varphi/PQ$, and $d\varphi/ds$ is the corresponding limit of $\Delta\varphi/PS$.



It is evident that the gradient has the direction normal to the level surface of φ through the point being considered. The most common symbols for the gradient are ∇ and \mathbf{grad} .^{*} In terms of the gradient, the directional derivative is given by

$$\frac{d\varphi}{ds} = |\mathbf{grad} \varphi| \cos \theta \quad (1-18)$$

where θ is the angle between the direction of ds and the direction of the gradient. This result is immediately evident from the geometry of Fig. 1-3. If we write ds for the vector displacement of magnitude ds , then (1-18) can be written

$$\frac{d\varphi}{ds} = \mathbf{grad} \varphi \cdot \frac{ds}{ds} \quad (1-19)$$

This equation enables us to find the explicit form of the gradient in any coordinate system in which we know the form of ds . In rectangular coordinates, we know that $ds = \mathbf{i} dx + \mathbf{j} dy + \mathbf{k} dz$. We also know that

$$d\varphi = \frac{\partial \varphi}{\partial x} dx + \frac{\partial \varphi}{\partial y} dy + \frac{\partial \varphi}{\partial z} dz$$

From this equation and (1-19), it follows that

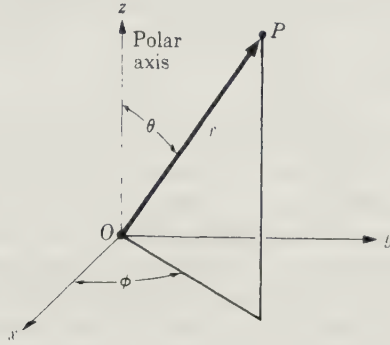
$$\begin{aligned} \frac{\partial \varphi}{\partial x} dx + \frac{\partial \varphi}{\partial y} dy + \frac{\partial \varphi}{\partial z} dz &= (\mathbf{grad} \varphi)_x dx + (\mathbf{grad} \varphi)_y dy \\ &\quad + (\mathbf{grad} \varphi)_z dz \end{aligned}$$

Equating coefficients of differentials of independent variables on both sides

^{*} We will use the latter notation at first.

FIGURE 1-4

Definition of the polar coordinates r , θ , and ϕ



of the equation gives

$$\mathbf{grad} \varphi = \mathbf{i} \frac{\partial \varphi}{\partial x} + \mathbf{j} \frac{\partial \varphi}{\partial y} + \mathbf{k} \frac{\partial \varphi}{\partial z} \quad (1-20)$$

in rectangular coordinates. In a more complicated case, the procedure is the same. In spherical polar coordinates, with r , θ , ϕ as defined in Fig. 1-4, we have

$$d\varphi = \frac{\partial \varphi}{\partial r} dr + \frac{\partial \varphi}{\partial \theta} d\theta + \frac{\partial \varphi}{\partial \phi} d\phi \quad (1-21)$$

and

$$ds = \mathbf{a}_r dr + \mathbf{a}_\theta r d\theta + \mathbf{a}_\phi r \sin \theta d\phi \quad (1-22)$$

where \mathbf{a}_r , \mathbf{a}_θ , and \mathbf{a}_ϕ are unit vectors in the r , θ , and ϕ directions, respectively. Applying (1-19) and equating coefficients of independent variables yields

$$\mathbf{grad} \varphi = \mathbf{a}_r \frac{\partial \varphi}{\partial r} + \mathbf{a}_\theta \frac{1}{r} \frac{\partial \varphi}{\partial \theta} + \mathbf{a}_\phi \frac{1}{r \sin \theta} \frac{\partial \varphi}{\partial \phi} \quad (1-23)$$

in spherical coordinates.

1-4 VECTOR INTEGRATION

While there are other aspects to differentiation involving vectors, it is convenient to discuss vector integration first. For our purposes, we may consider three kinds of integrals: line, surface, and volume, according to the nature of the differential appearing in the integral. The integrand may be either a vector field or a scalar field; however, certain combinations of integrands and differentials give rise to uninteresting integrals. Those of most interest here are the scalar line integral of a vector, the scalar surface integral of a vector, and the volume integrals of both vectors and scalars.

If \mathbf{F} is a vector field, a line integral of \mathbf{F} is written

$$\int_{a(C)}^b \mathbf{F}(\mathbf{r}) \cdot d\mathbf{l} \quad (1-24)$$

where C is the curve along which the integration is performed, a and b the initial and final points on the curve, and $d\mathbf{l}$ an infinitesimal vector displacement along the curve C . Since $\mathbf{F} \cdot d\mathbf{l}$ is a scalar it is clear that the line integral is a scalar. The definition of the line integral follows closely the Riemann definition of the definite integral. The segment of C between a and b is divided into a large number of small increments $\Delta\mathbf{l}_i$; for each increment an interior point is chosen and the value of \mathbf{F} at that point found. The scalar product of each increment with the corresponding value of \mathbf{F} is found and the sum of these computed. The line integral is then defined as the limit of this sum as the number of increments becomes infinite in such a way that each increment goes to zero. This definition may be compactly written as

$$\int_{a(C)}^b \mathbf{F} \cdot d\mathbf{l} = \lim_{N \rightarrow \infty} \sum_{i=1}^N \mathbf{F}_i \cdot \Delta\mathbf{l}_i$$

It is important to note that the line integral usually depends not only on the endpoints a and b but also on the curve C along which the integration is to be done, since the magnitude and direction of $\mathbf{F}(\mathbf{r})$ and the direction of $d\mathbf{l}$ depend on C and its tangent, respectively. The line integral around a closed curve is of sufficient importance that a special notation is used for it, namely,

$$\oint_C \mathbf{F} \cdot d\mathbf{l} \quad (1-25)$$

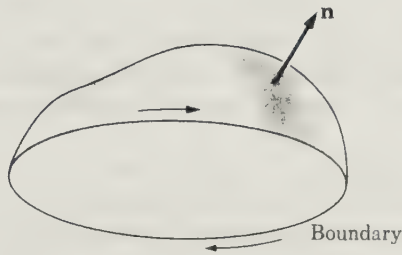
The integral around a closed curve is usually not zero; the class of vectors for which the line integral around *any* closed curve is zero is of considerable importance. For this reason, one often encounters line integrals around undesignated closed paths. For example,

$$\oint \mathbf{F} \cdot d\mathbf{l} \quad (1-26)$$

This notation is useful only in those cases where the integral is independent of the contour C within rather wide limits. If any ambiguity is possible, it is wise to specify the contour. The basic approach to the evaluation of line integrals is to obtain a one-parameter description of the curve and then use this description to express the line integral as the sum of three ordinary one-dimensional integrals. In all but the simplest cases, this procedure is long and tedious. Fortunately, however, it is seldom necessary to evaluate the integrals in this fashion. As will be seen later, it is often possible to show that the line integral does not depend on the path between the endpoints. In the latter case, a simple path may be chosen to simplify the integration.

FIGURE 1-5

Relation of normal \mathbf{n} to a surface and the direction of traversal of the boundary



If \mathbf{F} is again a vector, a surface integral of \mathbf{F} is written

$$\int_S \mathbf{F} \cdot \mathbf{n} \, da \quad (1-27)$$

where S is the surface over which the integration is to be performed, da is an infinitesimal area on S , and \mathbf{n} is a unit normal to da . There is a twofold ambiguity in the choice of \mathbf{n} , which is resolved by taking \mathbf{n} to be the outward drawn normal if S is a closed surface. If S is not closed and is finite, then it has a boundary, and the sense of the normal is important only with respect to the arbitrary positive sense of traversing the boundary. The positive sense of the normal is the direction in which a right-hand screw would advance if rotated in the direction of the positive sense on the bounding curve, as illustrated in Fig. 1-5. The surface integral of \mathbf{F} over a closed surface S is sometimes denoted by

$$\oint_S \mathbf{F} \cdot \mathbf{n} \, da$$

Comments exactly parallel to those made for the line integral can be made for the surface integral. This surface integral is clearly a scalar; it usually depends on the surface S , and cases where it does not are particularly important. The definition of the surface integral is made in a way exactly comparable to that of the line integral. The detailed formulation is left for the problem section.

If \mathbf{F} is a vector and φ a scalar, then the two volume integrals in which we are interested are

$$J = \int_V \varphi \, dv \quad \mathbf{K} = \int_V \mathbf{F} \, dv \quad (1-28)$$

Clearly J is a scalar and \mathbf{K} a vector. The definitions of these integrals reduce quickly to just the Riemann integral in three dimensions except that in \mathbf{K} one must note that there is one integral for each component of \mathbf{F} . These integrals are sufficiently familiar to require no further comment.

1-5 DIVERGENCE

Another important operator, which is essentially a derivative, is the divergence operator. The divergence of vector \mathbf{F} , written $\text{div } \mathbf{F}$, is defined as follows.

The *divergence* of a vector is the limit of its surface integral per unit volume as the volume enclosed by the surface goes to zero. That is,

$$\text{div } \mathbf{F} = \lim_{V \rightarrow 0} \frac{1}{V} \oint_S \mathbf{F} \cdot \mathbf{n} \, da$$

The divergence is clearly a scalar point function (scalar field), and it is defined at the limit point of the surface of integration. This definition has several virtues: It is independent of any special choice of coordinate system, and it may be used to find the explicit form of the divergence operator in any particular coordinate system.

In rectangular coordinates, the volume element $\Delta x \Delta y \Delta z$ provides a convenient basis for finding the explicit form of the divergence. If one corner of the rectangular parallelepiped is at the point x_0, y_0, z_0 , then

$$\begin{aligned} F_x(x_0 + \Delta x, y, z) &= F_x(x_0, y, z) + \Delta x \left. \frac{\partial F_x}{\partial x} \right|_{x_0, y, z} \\ F_y(x, y_0 + \Delta y, z) &= F_y(x, y_0, z) + \Delta y \left. \frac{\partial F_y}{\partial y} \right|_{x, y_0, z} \\ F_z(x, y, z_0 + \Delta z) &= F_z(x, y, z_0) + \Delta z \left. \frac{\partial F_z}{\partial z} \right|_{x, y, z_0} \end{aligned} \quad (1-29)$$

where higher-order terms in Δx , Δy , and Δz have been omitted. Since the area element $\Delta y \Delta z$ is perpendicular to the x -axis, $\Delta z \Delta x$ is perpendicular to the y -axis, and $\Delta x \Delta y$ is perpendicular to the z -axis, the definition of the divergence becomes

$$\begin{aligned} \text{div } \mathbf{F} &= \lim_{V \rightarrow 0} \frac{1}{\Delta x \Delta y \Delta z} \left\{ \iint F_x(x_0, y, z) \, dy \, dz \right. \\ &\quad + \Delta x \Delta y \Delta z \frac{\partial F_x}{\partial x} + \int F_y(x, y_0, z) \, dx \, dz \\ &\quad + \Delta x \Delta y \Delta z \frac{\partial F_y}{\partial y} + \int F_z(x, y, z_0) \, dx \, dy \\ &\quad + \Delta x \Delta y \Delta z \frac{\partial F_z}{\partial z} - \int F_x(x_0, y, z) \, dy \, dz \\ &\quad \left. - \int F_y(x, y_0, z) \, dx \, dz - \int F_z(x, y, z_0) \, dx \, dy \right\} \end{aligned} \quad (1-30)$$

The minus signs associated with the last three terms account for the fact that the outward drawn normal is in the direction of the negative axes in these cases. The limit is easily taken, and the divergence in rectangular coordinates is found to be

$$\operatorname{div} \mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \quad (1-31)$$

In spherical coordinates, the procedure is similar. The volume enclosed by the coordinate intervals Δr , $\Delta \theta$, $\Delta \phi$ is chosen as the volume of integration. This volume is $r^2 \sin \theta \Delta r \Delta \theta \Delta \phi$. Because the area enclosed by the coordinate intervals depends on the values of the coordinates (which is not the case with rectangular coordinates), it is best to write $\mathbf{F} \cdot \mathbf{n} \Delta a$ in its explicit form:

$$\begin{aligned} \mathbf{F} \cdot \mathbf{n} \Delta a &= F_r r^2 \sin \theta \Delta \theta \Delta \phi \\ &+ F_\theta r \sin \theta \Delta \phi \Delta r + F_\phi r \Delta r \Delta \theta \end{aligned} \quad (1-32)$$

It is clear from this expression that $r^2 F_r \sin \theta$, rather than just F_r , must be expanded in Taylor series. Similarly, it is the coefficient of the products of coordinate intervals that must be expanded in the other terms. Making these expansions and using them to evaluate the surface integral in the definition of the divergence gives

$$\begin{aligned} \operatorname{div} \mathbf{F} &= \lim_{v \rightarrow 0} \frac{1}{r^2 \sin \theta \Delta r \Delta \theta \Delta \phi} \left\{ \frac{\partial}{\partial r} (F_r r^2 \sin \theta) \Delta r \Delta \theta \Delta \phi \right. \\ &\quad \left. + \frac{\partial}{\partial \theta} (F_\theta r \sin \theta) \Delta \theta \Delta r \Delta \phi + \frac{\partial}{\partial \phi} (F_\phi r) \Delta \phi \Delta r \Delta \theta \right\} \quad (1-33) \end{aligned}$$

Taking the limit, the explicit form of the divergence in spherical coordinates is found to be

$$\operatorname{div} \mathbf{F} = \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 F_r) + \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta F_\theta) + \frac{1}{r \sin \theta} \frac{\partial F_\phi}{\partial \phi} \quad (1-34)$$

This method of finding the explicit form of the divergence is applicable to any coordinate system, provided that the forms of the volume and surface elements or, alternatively, the elements of length are known.

The physical significance of the divergence is readily seen in terms of an example taken from fluid mechanics. If \mathbf{V} is the velocity of a fluid, given as a function of position, and ρ is its density, then $\oint_S \rho \mathbf{V} \cdot \mathbf{n} da$ is clearly the net amount of fluid per unit time that leaves the volume enclosed by S . If the fluid is incompressible, the surface integral measures the total source of fluid enclosed by the surface. The preceding definition of the divergence then indicates that it may be interpreted as the limit of the source strength per unit volume, or the source density of an incompressible fluid.

An extremely important theorem involving the divergence may now be stated and proved.

Divergence theorem. The integral of the divergence of a vector over a volume V is equal to the surface integral of the normal component of the vector over the surface bounding V . That is,

$$\int_V \operatorname{div} \mathbf{F} \, dv = \oint_S \mathbf{F} \cdot \mathbf{n} \, da$$

Consider the volume to be subdivided into a large number of small cells. Let the i th cell have volume ΔV_i and be bounded by the surface S_i . It is clear that

$$\sum_i \oint_{S_i} \mathbf{F} \cdot \mathbf{n} \, da = \oint_S \mathbf{F} \cdot \mathbf{n} \, da \quad (1-35)$$

where in each integral on the left, the normal is directed outward from the volume being considered. Since outward to one cell is inward to the appropriate adjacent cell, all contributions to the left side of Eq. (1-35) cancel except those arising from the surface of S , and Eq. (1-35) is essentially proved. The divergence theorem is now obtained by letting the number of cells go to infinity in such a way that the volume of each cell goes to zero.

$$\oint_S \mathbf{F} \cdot \mathbf{n} \, da = \lim_{\Delta V_i \rightarrow 0} \sum_i \left\{ \frac{1}{\Delta V_i} \oint_{S_i} \mathbf{F} \cdot \mathbf{n} \, da \right\} \Delta V_i \quad (1-36)$$

In the limit, the sum on i becomes an integral over V , and the ratio of the integral over S_i to ΔV_i becomes the divergence of \mathbf{F} . Thus,

$$\oint_S \mathbf{F} \cdot \mathbf{n} \, da = \int_V \operatorname{div} \mathbf{F} \, dv \quad (1-37)$$

which is the divergence theorem. We shall have frequent occasion to exploit this theorem, both in the development of the theoretical aspects of electricity and magnetism and for the very practical purpose of evaluating integrals.

1-6 **CURL**

The third interesting vector differential operator is the curl. The curl of a vector, written **curl** \mathbf{F} , is defined as follows.

The *curl* of a vector is the limit of the ratio of the integral of its cross product with the outward drawn normal, over a closed surface, to the volume enclosed by the surface as the volume goes

to zero. That is,

$$\mathbf{curl} \mathbf{F} = \lim_{V \rightarrow 0} \frac{1}{V} \oint_S \mathbf{n} \times \mathbf{F} da \quad (1-38)$$

The parallelism between this definition and the definition of the divergence is quite apparent; instead of the scalar product of the vector with the outward drawn normal, one has the vector product. Otherwise the definitions are the same. A different but equivalent definition is more useful. This alternative definition is as follows.

The component of $\mathbf{curl} \mathbf{F}$ in the direction of the unit vector \mathbf{a} is the limit of a line integral per unit area, as the enclosed area goes to zero, this area being perpendicular to \mathbf{a} . That is,

$$\mathbf{a} \cdot \mathbf{curl} \mathbf{F} = \lim_{S \rightarrow 0} \frac{1}{S} \oint_C \mathbf{F} \cdot d\mathbf{l} \quad (1-39)$$

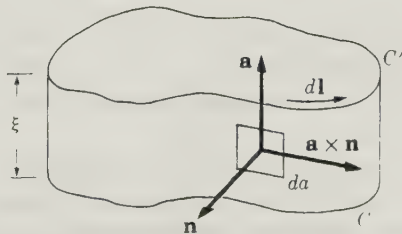
where the curve C , which bounds the surface S , is in a plane normal to \mathbf{a} .

It is easy to see the equivalence of the two definitions by considering a plane curve C and the volume swept out by this curve when it is displaced a distance ξ in the direction of the normal to its plane, as shown in Fig. 1-6. If \mathbf{a} is a normal to this plane, then taking the dot product of \mathbf{a} with the first definition of the curl, Eq. (1-38), gives

$$\mathbf{a} \cdot \mathbf{curl} \mathbf{F} = \lim_{V \rightarrow 0} \frac{1}{V} \oint_S \mathbf{a} \cdot \mathbf{n} \times \mathbf{F} da \quad (1-40)$$

Since \mathbf{a} is parallel to the normal for all of the bounding surface except the narrow strip bounded by C and C' , only the integral over this surface need be considered. For this surface we note that $\mathbf{a} \times \mathbf{n} da$ is just $\xi d\mathbf{l}$, where $d\mathbf{l}$ is

FIGURE 1-6
Volume swept out by displacing the plane curve C in the direction of its normal, \mathbf{a}



an infinitesimal displacement along C . Since, in addition, $V = \xi S$, the limit of the volume integral is just

$$\mathbf{a} \cdot \mathbf{curl} \mathbf{F} = \lim_{V \rightarrow 0} \frac{1}{\xi S} \oint \xi \mathbf{F} \cdot d\mathbf{l}$$

which reduces to the second form of our definition upon canceling the ξ 's. This equivalence can be shown without the use of the special volume used here; however, so doing sacrifices much of the simplicity of the proof given above.

The form of the curl in various coordinate systems can be calculated in much the same way as was done with the divergence. In rectangular coordinates, the volume $\Delta x \Delta y \Delta z$ is convenient. For the x -component of the curl only the faces perpendicular to the y - and z -axes contribute. Recalling that $\mathbf{j} \times \mathbf{k} = -\mathbf{k} \times \mathbf{j} = \mathbf{i}$, the nonvanishing contributions from the faces of the parallelepiped to the x -component of the curl give

$$\begin{aligned} (\mathbf{curl} \mathbf{F})_x &= \lim_{V \rightarrow 0} \frac{1}{V} \{ [-F_y(x, y, z + \Delta z) + F_y(x, y, z)] \Delta x \Delta y \\ &\quad + [F_z(x, y + \Delta y, z) - F_z(x, y, z)] \Delta x \Delta z \} \end{aligned} \quad (1-41)$$

Making a Taylor series expansion and taking the limit gives

$$(\mathbf{curl} \mathbf{F})_x = \frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \quad (1-42)$$

for the x -component of the curl. The y - and z -components may be found in exactly the same way. They are

$$(\mathbf{curl} \mathbf{F})_y = \frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}, \quad (\mathbf{curl} \mathbf{F})_z = \frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \quad (1-43)$$

The form of the curl in rectangular coordinates can be easily remembered if it is noted that it is just the expansion of a three-by-three determinant, namely,

$$\mathbf{curl} \mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_x & F_y & F_z \end{vmatrix} \quad (1-44)$$

Finding the form of the curl in other coordinate systems is only slightly more complicated and is left to the problem section.

As with the divergence, we encounter an important and useful theorem involving the curl, known as Stokes's theorem.

Stokes's theorem. The line integral of a vector around a closed curve is equal to the integral of the normal component of its curl over any surface bounded by the curve. That is,

$$\oint_C \mathbf{F} \cdot d\mathbf{l} = \int_S \mathbf{curl} \mathbf{F} \cdot \mathbf{n} \, da \quad (1-45)$$

where C is a closed curve that bounds the surface S .

The proof of this theorem is quite analogous to the proof of the divergence theorem. The surface S is divided into a large number of cells. The surface of the i th cell is called ΔS_i and the curve bounding it is C_i . Since each of these cells must be traversed in the same sense, it is clear that the sum of the line integrals around the C_i 's is just the line integral around the bounding curve; all of the other contributions cancel. Thus

$$\oint_C \mathbf{F} \cdot d\mathbf{l} = \sum_i \oint_{C_i} \mathbf{F} \cdot d\mathbf{l}$$

It remains only to take the limit as the number of cells becomes infinite in such a way that the area of each goes to zero. The result of this limiting process is

$$\begin{aligned} \oint_C \mathbf{F} \cdot d\mathbf{l} &= \lim_{\Delta S_i \rightarrow 0} \sum_i \frac{1}{\Delta S_i} \oint_{C_i} \mathbf{F} \cdot d\mathbf{l} \Delta S_i \\ &= \int_S \mathbf{curl} \mathbf{F} \cdot \mathbf{n} \, da \end{aligned}$$

which is Stokes's theorem. This theorem, like the divergence theorem, is useful both in the development of electromagnetic theory and in the evaluation of integrals. It is perhaps worth noting that both the divergence theorem and Stokes's theorem are essentially partial integrations.

THE VECTOR DIFFERENTIAL OPERATOR ∇

We now introduce an alternative notation for the three types of vector differentiation that have been discussed—namely, gradient, divergence, and curl. This notation uses the vector differential operator *del*, defined in Cartesian coordinates as

$$\nabla = \mathbf{i} \frac{\partial}{\partial x} + \mathbf{j} \frac{\partial}{\partial y} + \mathbf{k} \frac{\partial}{\partial z} \quad (1-46)$$

Del is a differential operator in that it is used only in front of a function of (x, y, z) , which it differentiates; it is a vector in that it obeys the laws of

vector algebra.* In terms of del, Eqs. (1-20), (1-31), and (1-44) are expressed as follows:

$$\mathbf{grad} = \nabla,$$

$$\nabla\varphi = \mathbf{i} \frac{\partial\varphi}{\partial x} + \mathbf{j} \frac{\partial\varphi}{\partial y} + \mathbf{k} \frac{\partial\varphi}{\partial z} \quad (1-20)$$

$$\text{div} = \nabla \cdot,$$

$$\nabla \cdot \mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \quad (1-31)$$

$$\text{curl} = \nabla \times,$$

$$\nabla \times \mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_x & F_y & F_z \end{vmatrix} \quad (1-44)$$

The operations expressed with del are themselves independent of any special choice of coordinate system. Any identities that can be proved using the Cartesian representation hold independently of the coordinate system. Del can be expressed in a non-Cartesian (curvilinear) orthonormal coordinate system in a form analogous to Eq. (1-46) with the appropriate distance elements, but it must be remembered in applying it that the unit vectors in such coordinate systems are themselves functions of position and have to be differentiated.† The important integral theorems, according to Eqs. (1-19), (1-45), and (1-37), are

$$\int_{a(C)}^b \nabla\varphi \cdot d\mathbf{l} = \int_a^b d\varphi = \varphi \Big|_a^b = \varphi_b - \varphi_a \quad (1-47)$$

$$\int_S \nabla \times \mathbf{F} \cdot \mathbf{n} \, da = \oint_C \mathbf{F} \cdot d\mathbf{l} \quad (1-45)$$

$$\int_V \nabla \cdot \mathbf{F} \, dv = \oint_S \mathbf{F} \cdot \mathbf{n} \, da \quad (1-37)$$

These give the integral of a derivative of a function over a region of n dimensions, in terms of the values of the function itself on the $(n - 1)$ -dimensional boundary of the region, for $n = 1, 2, 3$. Because the del

* It is also a vector in terms of its transformation properties, as shown in Appendix II.

† Results for cylindrical and spherical coordinates are found in Appendix IV. An elementary discussion is given by H. T. Yang, *American Journal of Physics*, vol. 40, p. 109 (1972).

operator obeys the rules of vector algebra, it is convenient to use it in calculations involving vector analysis, and henceforth we shall express the gradient, divergence, and curl in terms of ∇ . It should be noted that ∇ is a *linear* operator:

$$\nabla(a\varphi + b\psi) = a\nabla\varphi + b\nabla\psi$$

$$\nabla \cdot (a\mathbf{F} + b\mathbf{G}) = a\nabla \cdot \mathbf{F} + b\nabla \cdot \mathbf{G}$$

$$\nabla \times (a\mathbf{F} + b\mathbf{G}) = a\nabla \times \mathbf{F} + b\nabla \times \mathbf{G}$$

if a and b are constant scalars.

1-8 FURTHER DEVELOPMENTS

The gradient, divergence, and curl operations may be repeated on appropriate kinds of fields. For example, it makes sense to take the divergence of the gradient of a scalar field. Indeed, this combined operation is sufficiently important that it has a special name, the *Laplacian*. It does not, however, make sense to take the curl of the divergence of a vector field because doing so involves taking the curl of a scalar, which is undefined. There are altogether five second-order operators that make sense, and two of these yield zero. All five, however, are very important in the study of electromagnetic fields.

The first of these is the *Laplacian operator*, which is defined as the divergence of the gradient of a scalar field, and which is usually written ∇^2 ,

$$\nabla \cdot \nabla = \nabla^2$$

In rectangular coordinates,

$$\nabla^2\varphi = \frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2} \quad (1-48)$$

This operator is of great importance in electrostatics and will be considered at length in Chapter 3.

The curl of the gradient of any scalar field is zero. This statement is most easily verified by writing it out in rectangular coordinates. If the scalar field is φ , then

$$\nabla \times (\nabla\varphi) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{\partial\varphi}{\partial x} & \frac{\partial\varphi}{\partial y} & \frac{\partial\varphi}{\partial z} \end{vmatrix} = \mathbf{i} \left(\frac{\partial^2\varphi}{\partial y \partial z} - \frac{\partial^2\varphi}{\partial z \partial y} \right) + \dots = 0$$

(1-49)

which verifies the original statement. In operator notation,

$$\nabla \times \nabla = 0$$

The divergence of any curl is also zero. This result is verified directly in rectangular coordinates by writing

$$\begin{aligned} \nabla \cdot (\nabla \times \mathbf{F}) &= \frac{\partial}{\partial x} \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right) \\ &+ \frac{\partial}{\partial y} \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} \right) + \dots = 0 \end{aligned} \quad (1-50)$$

The two other possible second-order operations are taking the curl of the curl or the gradient of the divergence of a vector field. It is left as an exercise to show that in rectangular coordinates

$$\nabla \times (\nabla \times \mathbf{F}) = \nabla(\nabla \cdot \mathbf{F}) - \nabla^2 \mathbf{F} \quad (1-51)$$

where the Laplacian of a vector is the vector whose rectangular components are the Laplacians of the rectangular components of the original vector. In any coordinate system other than rectangular the Laplacian of a vector is defined by Eq. (1-51).

Another way in which the application of the vector differential operators may be extended is to apply them to various products of two vectors and scalars. The six possible combinations of differential operators and products are tabulated in Table 1-1. These identities may be readily verified in rectangular coordinates, which is sufficient to assure their validity in any coordinate system. A derivative of a product of more than two functions, or a higher than second-order derivative of a function, can be calculated by repeated applications of the identities in Table 1-1, which is therefore exhaustive. The formulas can be easily remembered from the

TABLE 1-1
Differential Vector
Identities

$\nabla \cdot \nabla \varphi = \nabla^2 \varphi$	(1-1-1)
$\nabla \cdot \nabla \times \mathbf{F} = 0$	(1-1-2)
$\nabla \times \nabla \varphi = 0$	(1-1-3)
$\nabla \times (\nabla \times \mathbf{F}) = \nabla(\nabla \cdot \mathbf{F}) - \nabla^2 \mathbf{F}$	(1-1-4)
$\nabla(\varphi \psi) = (\nabla \varphi) \psi + \varphi \nabla \psi$	(1-1-5)
$\nabla(\mathbf{F} \cdot \mathbf{G}) = (\mathbf{F} \cdot \nabla) \mathbf{G} + \mathbf{F} \times (\nabla \times \mathbf{G}) + (\mathbf{G} \cdot \nabla) \mathbf{F} + \mathbf{G} \times (\nabla \times \mathbf{F})$	(1-1-6)
$\nabla \cdot (\varphi \mathbf{F}) = (\nabla \varphi) \cdot \mathbf{F} + \varphi \nabla \cdot \mathbf{F}$	(1-1-7)
$\nabla \cdot (\mathbf{F} \times \mathbf{G}) = (\nabla \times \mathbf{F}) \cdot \mathbf{G} - (\nabla \times \mathbf{G}) \cdot \mathbf{F}$	(1-1-8)
$\nabla \times (\varphi \mathbf{F}) = (\nabla \varphi) \times \mathbf{F} + \varphi \nabla \times \mathbf{F}$	(1-1-9)
$\nabla \times (\mathbf{F} \times \mathbf{G}) = (\nabla \cdot \mathbf{G}) \mathbf{F} - (\nabla \cdot \mathbf{F}) \mathbf{G} + (\mathbf{G} \cdot \nabla) \mathbf{F} - (\mathbf{F} \cdot \nabla) \mathbf{G}$	(1-1-10)

rules of vector algebra and ordinary differentiation. The only ambiguity could be in (1-1-6) where $\mathbf{F} \cdot \nabla$ occurs (not $\nabla \cdot \mathbf{F}$).

Some particular types of functions come up often enough in electromagnetic theory that it is worth noting their various derivatives now. For the function $\mathbf{F} = \mathbf{r}$,

$$\begin{aligned}\nabla \cdot \mathbf{r} &= 3 \\ \nabla \times \mathbf{r} &= 0 \\ (\mathbf{G} \cdot \nabla)\mathbf{r} &= \mathbf{G} \\ \nabla^2 \mathbf{r} &= 0\end{aligned}\tag{1-52}$$

For a function that depends only on the distance $r = |\mathbf{r}| = \sqrt{x^2 + y^2 + z^2}$,

$$\varphi(r) \text{ or } \mathbf{F}(r): \quad \nabla = \frac{\mathbf{r}}{r} \frac{d}{dr}\tag{1-53}$$

For a function that depends on the scalar argument $\mathbf{A} \cdot \mathbf{r}$, where \mathbf{A} is a constant vector,

$$\varphi(\mathbf{A} \cdot \mathbf{r}) \text{ or } \mathbf{F}(\mathbf{A} \cdot \mathbf{r}): \quad \nabla = \mathbf{A} \frac{d}{d(\mathbf{A} \cdot \mathbf{r})}\tag{1-54}$$

For a function that depends on the argument $\mathbf{R} = \mathbf{r} - \mathbf{r}'$, where \mathbf{r}' is treated as a constant,

$$\nabla = \nabla_R\tag{1-55}$$

$$\nabla_R = \mathbf{i} \frac{\partial}{\partial X} + \mathbf{j} \frac{\partial}{\partial Y} + \mathbf{k} \frac{\partial}{\partial Z}$$

where $\mathbf{R} = X\mathbf{i} + Y\mathbf{j} + Z\mathbf{k}$. If \mathbf{r} is treated as constant instead,

$$\nabla = -\nabla'\tag{1-56}$$

where

$$\nabla' = \mathbf{i} \frac{\partial}{\partial x'} + \mathbf{j} \frac{\partial}{\partial y'} + \mathbf{k} \frac{\partial}{\partial z'}$$

There are several possibilities for the extension of the divergence theorem and of Stokes's theorem. The most interesting of these is Green's theorem, which is

$$\int_V (\psi \nabla^2 \varphi - \varphi \nabla^2 \psi) dv = \oint_S (\psi \nabla \varphi - \varphi \nabla \psi) \cdot \mathbf{n} da\tag{1-57}$$

This theorem follows from the application of the divergence theorem to the vector

$$\mathbf{F} = \psi \nabla \varphi - \varphi \nabla \psi$$

TABLE 1-2
Vector Integral
Theorems

$$\int_S \mathbf{n} \times \nabla \varphi \, da = \oint_C \varphi \, d\mathbf{l} \quad (1-2-1)$$

$$\int_V \nabla \varphi \, dv = \oint_S \varphi \mathbf{n} \, da \quad (1-2-2)$$

$$\int_V \nabla \times \mathbf{F} \, dv = \oint_S \mathbf{n} \times \mathbf{F} \, da \quad (1-2-3)$$

$$\int_V (\nabla \cdot \mathbf{G} + \mathbf{G} \cdot \nabla) \mathbf{F} \, dv = \oint_S \mathbf{F}(\mathbf{G} \cdot \mathbf{n}) \, da \quad (1-2-4)$$

Using this \mathbf{F} in the divergence theorem, we obtain

$$\int_V \nabla \cdot [\psi \nabla \varphi - \varphi \nabla \psi] \, dv = \oint_S (\psi \nabla \varphi - \varphi \nabla \psi) \cdot \mathbf{n} \, da \quad (1-58)$$

Using the identity (Table 1-1) for the divergence of a scalar times a vector gives

$$\nabla \cdot (\psi \nabla \varphi) - \nabla \cdot (\varphi \nabla \psi) = \psi \nabla^2 \varphi - \varphi \nabla^2 \psi \quad (1-59)$$

Combining Eqs. (1-58) and (1-59) yields Green's theorem. Some other integral theorems are listed in Table 1-2.

This concludes our brief discussion of vector analysis. In the interests of brevity, the proofs of many results have been relegated to the problems. No attempt has been made to achieve a high degree of rigor. The approach has been utilitarian: What we will need, we have developed; everything else has been omitted.

1-9

SUMMARY

Three different types of vector differentiation can be expressed by the **vector differential operator del**, ∇ , namely, gradient, divergence, and curl:

$$\nabla \varphi = \mathbf{i} \frac{\partial \varphi}{\partial x} + \mathbf{j} \frac{\partial \varphi}{\partial y} + \mathbf{k} \frac{\partial \varphi}{\partial z}$$

$$\nabla \cdot \mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}$$

$$\nabla \times \mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_x & F_y & F_z \end{vmatrix}$$

Del is a linear operator. Repeated applications of it, or its application to

products of functions, produce formulas that can be derived in rectangular coordinates, but are independent of the coordinate system. They can be remembered by the rules of vector algebra and ordinary differentiation. The derivatives of a few special functions are worth committing to memory. The most important **integral theorems** about the derivatives are

$$\int_a^b \nabla \varphi \cdot d\mathbf{l} = \varphi \Big|_a^b$$

$$\int_S \nabla \times \mathbf{F} \cdot \mathbf{n} \, da = \oint_C \mathbf{F} \cdot d\mathbf{l} \quad (\text{Stokes's theorem})$$

$$\int_V \nabla \cdot \mathbf{F} \, dv = \oint_S \mathbf{F} \cdot \mathbf{n} \, da \quad (\text{Divergence theorem})$$

These are generalizations of the fundamental theorem of calculus.

PROBLEMS

1-1 The vectors from the origin to the points A, B, C, D are

$$\mathbf{A} = \mathbf{i} + \mathbf{j} + \mathbf{k}$$

$$\mathbf{B} = 2\mathbf{i} + 3\mathbf{j}$$

$$\mathbf{C} = 3\mathbf{i} + 5\mathbf{j} - 2\mathbf{k}$$

$$\mathbf{D} = \mathbf{k} - \mathbf{j}$$

Show that the lines \overline{AB} and \overline{CD} are parallel and find the ratio of their lengths.

1-2 Show that the following vectors are perpendicular:

$$\mathbf{A} = \mathbf{i} + 4\mathbf{j} + 3\mathbf{k} \quad \mathbf{B} = 4\mathbf{i} + 2\mathbf{j} - 4\mathbf{k}$$

1-3 Show that the following vectors form the sides of a right triangle:

$$\mathbf{A} = 2\mathbf{i} - \mathbf{j} + \mathbf{k} \quad \mathbf{B} = \mathbf{i} - 3\mathbf{j} - 5\mathbf{k} \quad \mathbf{C} = 3\mathbf{i} - 4\mathbf{j} - 4\mathbf{k}$$

1-4 By squaring both sides of the equation $\mathbf{A} = \mathbf{B} - \mathbf{C}$ and interpreting the result geometrically, prove the "law of cosines."

1-5 (a) Show that the following are unit vectors in the xy -plane making angles α, β with the x -axis:

$$\mathbf{A} = \mathbf{i} \cos \alpha + \mathbf{j} \sin \alpha \quad \mathbf{B} = \mathbf{i} \cos \beta + \mathbf{j} \sin \beta$$

(b) By means of a scalar product, obtain the formula for $\cos(\alpha - \beta)$.

1-6 If \mathbf{A} is a constant vector and \mathbf{r} is the vector from the origin to the point (x, y, z) , show that the following is the equation of a plane:

$$(\mathbf{r} - \mathbf{A}) \cdot \mathbf{A} = 0$$

1-7 With \mathbf{A} and \mathbf{r} defined as in Problem 1-6, show that the following is the equation

of a sphere:

$$(\mathbf{r} - \mathbf{A}) \cdot \mathbf{r} = 0$$

1-8 Using the dot product, find the cosine of the angle between the body diagonal of a cube and one of the cube edges.

1-9 Prove the law of sines for a triangle by using the vector cross product with $\mathbf{A} + \mathbf{C} = \mathbf{B}$.

1-10 If \mathbf{A} , \mathbf{B} , \mathbf{C} , are vectors from the origin to the points A , B , C , show that the following is perpendicular to the plane ABC :

$$(\mathbf{A} \times \mathbf{B}) + (\mathbf{B} \times \mathbf{C}) + (\mathbf{C} \times \mathbf{A})$$

1-11 Verify that Eq. (1-15) is a solution to Eq. (1-14) by direct substitution. [Note that Eq. (1-14) implies that \mathbf{C} is perpendicular to \mathbf{A} .]

1-12 (a) Show that \mathbf{A} , \mathbf{B} , and \mathbf{C} are not linearly independent if

$$\mathbf{A} \cdot \mathbf{B} \times \mathbf{C} = 0$$

(b) Are the following vectors linearly independent?

$$\mathbf{A} = \mathbf{j} + 3\mathbf{k}, \quad \mathbf{B} = \mathbf{i} - 2\mathbf{k}, \quad \mathbf{C} = \mathbf{i} + \mathbf{j} + \mathbf{k}$$

1-13 (a) Show that the unit vector normal to the surface $\varphi(\mathbf{r}) = \text{constant}$ is

$$\mathbf{n} = \nabla\varphi/|\nabla\varphi|$$

(b) Find \mathbf{n} for the ellipsoid

$$\varphi = ax^2 + by^2 + cz^2$$

1-14 Find the gradient of φ in cylindrical coordinates, given that $d\mathbf{s} = dr\mathbf{a}_r + r d\theta\mathbf{a}_\theta + dz\mathbf{k}$. It should be noted that r and θ have different meanings here than in Eqs. (1-21) and (1-22). In spherical coordinates, r is the magnitude of the radius vector from the origin, and θ is the polar angle. In cylindrical coordinates, r is the perpendicular distance from the cylinder axis, and θ is the azimuthal angle about this axis.

1-15 From the definition of the divergence, obtain an expression for $\nabla \cdot \mathbf{F}$ in cylindrical coordinates.

1-16 (a) Find the divergence of the vector

$$\mathbf{i}(x^2 + yz) + \mathbf{j}(y^2 + zx) + \mathbf{k}(z^2 + xy)$$

(b) Also find the curl.

1-17 Is $\nabla \times \mathbf{F}$ necessarily perpendicular to \mathbf{F} for every vector function \mathbf{F} ? Justify your answer.

1-18 For any two scalar functions φ and ψ , prove that

$$\nabla^2(\varphi\psi) = \varphi\nabla^2\psi + \psi\nabla^2\varphi + 2\nabla\varphi \cdot \nabla\psi$$

1-19 If \mathbf{r} is the vector from the origin to the point (x, y, z) , prove the formulas

$$\nabla \cdot \mathbf{r} = 3 \quad \nabla \times \mathbf{r} = 0 \quad (\mathbf{u} \cdot \nabla)\mathbf{r} = \mathbf{u}$$

[Note: \mathbf{u} is any vector.]

1-20 If \mathbf{A} is a constant vector, show that

$$\nabla(\mathbf{A} \cdot \mathbf{r}) = \mathbf{A}$$

1-21 Prove identities (1-1-7) and (1-1-9) in Table 1-1.

1-22 If r is the magnitude of the vector from the origin to the point (x, y, z) , and $f(r)$ is an arbitrary function of r , prove that

$$\nabla f(r) = \frac{\mathbf{r}}{r} \frac{df}{dr}, \quad \nabla \times [f(r)\mathbf{r}] = 0$$

1-23 Prove that

$$\nabla \cdot \mathbf{F}(r) = \frac{\mathbf{r}}{r} \cdot \frac{d\mathbf{F}}{dr}$$

1-24 If $\xi = \mathbf{A} \cdot \mathbf{r}$, prove that

$$\nabla \varphi(\xi) = \mathbf{A} \frac{d\varphi}{d\xi}$$

1-25 Verify Eq. (1-51) in rectangular coordinates, where $\nabla^2 \mathbf{F}$ in these coordinates is as defined in the text.

1-26 Prove identities (1-2-2) and (1-2-4) in Table 1-2. [*Hint:* Use the divergence theorem and one or more identities from Table 1-1.]

Electrostatics

The subject of electricity is introduced in this chapter with the empirical concepts of charge and the force law between charges (Coulomb's law). But we shall use the mathematical tools of Chapter 1 to express Coulomb's law in other, more powerful formulations: the electric potential formulation and Gauss's law, both of which will be very important to the subsequent development of the subject.

2-1

ELECTRIC CHARGE

While the first observation of the electrification of objects by rubbing is lost in antiquity, it is commonly known that rubbing a hard rubber comb on a piece of wool endows the rubber with the ability to pick up small pieces of paper. As a result of rubbing the two objects together (strictly speaking, as a result of bringing them into close contact), both the rubber and the wool acquire a new property: They are *charged*. This experiment serves to introduce the concept of *charge*. But charge, itself, is not created during this process; the total charge, or the sum of the charges on the two bodies, is still the same as before electrification. In the light of modern physics, we know that microscopic charged particles—specifically, electrons—are transferred from the wool to the rubber, leaving the wool positively charged and the rubber comb negatively charged.

Charge is a fundamental and characteristic property of the microscopic particles that make up matter. In fact, all atoms are composed of protons, neutrons, and electrons, and two of these particles bear charges. But even though on a microscopic scale matter is composed of a large number of charged particles, the powerful electrical forces associated with these particles are fairly well hidden in a macroscopic observation. The reason is that there are two kinds of charge, positive and negative, and an ordinary

piece of matter contains approximately equal amounts of each kind. From the macroscopic viewpoint, then, charge refers to net charge, or excess charge. When we say that an object is charged, we mean that it has an excess charge, either an excess of electrons (negative) or an excess of protons (positive). In this and the following chapters, charge will usually be denoted by the symbol q .

It is an experimental observation that charge can be neither created nor destroyed. The total charge of a closed system cannot change. From the macroscopic point of view, charges may be regrouped and combined in different ways; nevertheless, we may state that *net charge is conserved in a closed system*.

2-2 COULOMB'S LAW

Toward the end of the eighteenth century, techniques in experimental science achieved sufficient sophistication to make possible refined observations of the forces between electric charges. The results of these observations, which were extremely controversial at the time, can be summarized in three statements.

1. There are two and only two kinds of electric charge, now known as positive and negative.
2. Two point charges exert on each other forces that act along the line joining them and are inversely proportional to the square of the distance between them.
3. These forces are also proportional to the product of the charges, are repulsive for like charges, and attractive for unlike charges.

The last two statements, with the first as preamble, are known as *Coulomb's law* in honor of Charles Augustin de Coulomb (1736–1806), who was one of the leading eighteenth century students of electricity.

Coulomb's law for point charges may be concisely formulated in the vector notation of Chapter 1 as

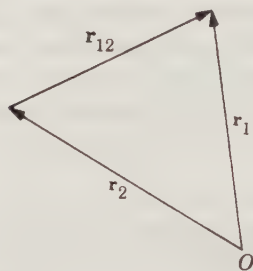
$$\mathbf{F}_1 = C_u \frac{q_1 q_2}{r_{12}^2} \frac{\mathbf{r}_{12}}{r_{12}} \quad (2-1)$$

$$\mathbf{r}_{12} = \mathbf{r}_1 - \mathbf{r}_2$$

where \mathbf{F}_1 is the force on charge q_1 , \mathbf{r}_{12} is the vector to q_1 from q_2 , r_{12} is the magnitude of \mathbf{r}_{12} , and C_u is a constant of proportionality about which more will be said later. In Eq. (2-1) a unit vector in the direction of \mathbf{r}_{12} has been formed by dividing \mathbf{r}_{12} by its magnitude, a device that will be used frequently. If the force on q_2 is to be found, it is only necessary to change every subscript 1 to 2 and every 2 to 1. Understanding this notation is important, since in future work it will provide a technique for keeping track

FIGURE 2-1

Vector \mathbf{r}_{12} , extending from the point at the tip of vector \mathbf{r}_2 to the point at the tip of the vector \mathbf{r}_1 . Clearly, $\mathbf{r}_{12} = -\mathbf{r}_{21}$.



of field and source variables. Figure 2-1 shows the vector \mathbf{r}_{12} with respect to an arbitrary origin O .

Coulomb's law applies to point charges. In the macroscopic sense, a "point charge" is one whose spatial dimensions are very small compared with any other length pertinent to the problem under consideration, and we shall use the term "point charge" in this sense. To the best of our knowledge, Coulomb's law also applies to the interactions of atomic particles such as protons and electrons. Equation (2-1) is found to hold for the electrostatic repulsion between nuclei at distances greater than about 10^{-14} meter; at smaller distances, the powerful, but short-ranged, nuclear forces dominate the picture.

Equation (2-1) is an experimental law. Nevertheless, there is both theoretical and experimental evidence to indicate that the inverse square law is exact—that is, the exponent of r_{12} is exactly 2. By an indirect experiment,* it has been shown that the exponent of r_{12} can differ from 2 by no more than one part in 10^{15} .

The constant C_u in Eq. (2-1) requires some comment, since it determines the system of units. The units of force and distance are presumably those belonging to one of the systems used in mechanics: The most direct procedure here would be to set $C_u = 1$, and choose the unit of charge such that Eq. (2-1) agrees with experiment. This procedure is adopted in the gaussian system of units. Other procedures are also possible and may have certain advantages; for example, the unit of charge may be specified in advance. It was shown by Giorgi in 1901 that all of the common electrical units, such as the ampere, volt, ohm, henry, and so forth can be combined with one of the mechanical systems—namely, the mks (meter-kilogram-second system)—to form a system of units for all electric and

* E. R. Williams, J. E. Faller, and H. A. Hill, *Physical Review Letters*, vol. 26, p. 721 (1971). Similar experiments were performed earlier. Maxwell established the exponent of 2 to within one part in 20,000.

magnetic problems. An advantage of this system is that the results of electric circuit calculations come out in the electrical units that are used in the laboratory. We shall use the rationalized mks or *Giorgi system* of units in the form known as SI (*Système International*). Since in this system q is measured in coulombs (C), \mathbf{r} in meters, and \mathbf{F} in newtons (N), C_u must have the dimensions of newton meters² per coulomb².

The size of the SI unit of charge, the coulomb, is established from magnetic experiments. This choice of units makes $C_u = 8.9874 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$.^{*} We make the apparently complicated substitution, $C_u = 1/4\pi\epsilon_0$, in the interest of simplifying other equations. The constant ϵ_0 , which will occur repeatedly, is known as the *permittivity of free space* and is numerically equal to $8.854 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2$.

In the mks system, *Coulomb's law* for the force between two point charges can thus be written

$$\mathbf{F}_1 = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r_{12}^2} \frac{\mathbf{r}_{12}}{r_{12}} \quad (2-1a)$$

In Appendix III the definitions of the coulomb, the ampere, the permeability, and permittivity of free space are related to one another and to the velocity of light in a logical way; since a logical formulation of these definitions requires a knowledge of magnetic phenomena and of electromagnetic wave propagation, it is not appropriate to pursue them now. In Appendix III, the gaussian system of units is also discussed. Until Chapter 4, every formula can be changed to gaussian units simply by replacing ϵ_0 with $1/4\pi$.

If more than two point charges are present, the mutual forces are determined by the repeated application of Eq. (2-1). In particular, if a system of N charges is considered, the force on the i th charge is given by

$$\mathbf{F}_i = q_i \sum_{j \neq i}^N \frac{q_j}{4\pi\epsilon_0 r_{ij}^3} \mathbf{r}_{ij} \quad (2-2)$$

$$\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$$

where the summation on the right is extended over all of the charges except the i th. Equation (2-2) is the superposition principle for forces, which says that the total force acting on a body is the vector sum of the individual forces that act on it.

^{*} The approximate value of $C_u \approx 9 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$ is perhaps more easily remembered.

A simple extension of the idea of N interacting point charges is the interaction of a point charge with a continuous charge distribution. We deliberately choose this configuration to avoid certain subtleties that may be encountered when the interaction of two continuous charge distributions is considered.

Before proceeding further, the meaning of a continuous distribution of charge should be examined. It is well known that electric charge is found in multiples of a basic charge—that of the electron. In other words, if any charge were examined in great detail, its magnitude would be found to be an integral multiple of the magnitude of the electronic charge. For the purposes of macroscopic physics, this discreteness of charge causes no difficulties simply because the electronic charge has a magnitude of 1.6019×10^{-19} C, which is extremely small. The smallness of the basic unit means that macroscopic charges are invariably composed of a very large number of electronic charges; thus, in a macroscopic charge distribution, any small element of volume contains a large number of electrons.

We may describe a charge distribution in terms of a *charge density function*, defined as the limit of the charge per unit volume as the volume becomes infinitesimal. Care must be used, however, in applying this kind of description to atomic problems, since in such cases only a small number of electrons are involved, and the process of taking the limit is meaningless. Leaving aside atomic cases, we may proceed as though a segment of charge might be subdivided indefinitely. We thus describe the charge distribution by means of point functions.

A *volume charge density* is defined by

$$\rho = \lim_{\Delta V \rightarrow 0} \frac{\Delta q}{\Delta V} \quad (2-3)$$

and a *surface charge density* is defined by

$$\sigma = \lim_{\Delta S \rightarrow 0} \frac{\Delta q}{\Delta S} \quad (2-4)$$

From what has been said about q , it is evident that ρ and σ are net charge, or excess charge, densities. It is worthwhile mentioning that in typical solid materials even a very large charge density ρ will involve a change in the local electron density of only about one part in 10^9 .

If charge is distributed through a volume V with a density ρ , and on the surface S that bounds V with a density σ , then the force exerted by this charge distribution on a point charge q located at \mathbf{r} is obtained from Eq.

(2-2) by replacing q_j with $\rho_j dv'_j$ (or with $\sigma_j da'_j$) and proceeding to the limit:

$$\begin{aligned} \mathbf{F}_q &= \frac{q}{4\pi\epsilon_0} \int_V \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} \rho(\mathbf{r}') dv' \\ &+ \frac{q}{4\pi\epsilon_0} \int_S \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} \sigma(\mathbf{r}') da' \end{aligned} \quad (2-5)$$

The variable \mathbf{r}' is used to locate a point within the charge distribution—that is, it plays the role of the source point \mathbf{r}_j in Eq. (2-2). It may appear at first sight that if point \mathbf{r} falls inside the charge distribution, the first integral of Eq. (2-5) should diverge, but that is not the case. The region of integration in the vicinity of \mathbf{r} contributes a negligible amount, and the integral is well behaved (see Problem 2-5).

It is clear that the force on q as given by Eq. (2-5) is proportional to q , and the same is true in Eq. (2-2). This observation leads us to introduce a vector field that is independent of q , namely, the force per unit charge. This vector field, known as the electric field, is considered in detail in the following section.

2-3 THE ELECTRIC FIELD

The *electric field* at a point is defined operationally as the limit of the force on a test charge placed at the point to the charge of the test charge—the limit being taken as the magnitude of the test charge goes to zero. The customary symbol for the electric field is \mathbf{E} .

In vector notation, the definition of \mathbf{E} becomes

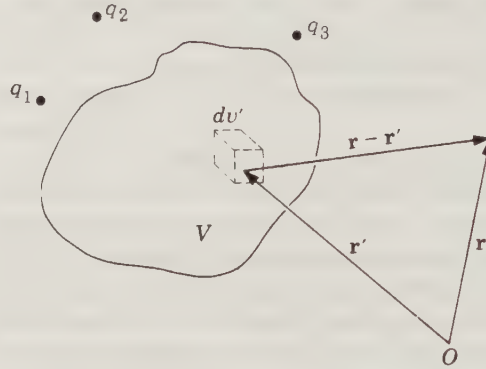
$$\mathbf{E} = \lim_{q \rightarrow 0} \frac{\mathbf{F}_q}{q} \quad (2-6)$$

The limiting process is included in the definition of \mathbf{E} to ensure that the test charge does not affect the charge distribution that produces \mathbf{E} . If, for example, positive charge is distributed on the surface of a conductor (a conductor is a material in which charge is free to move), then bringing a test charge into the vicinity of the conductor will cause the charge on the conductor to redistribute itself. If the electric field were calculated using the ratio of force to charge for a finite test charge, the field obtained would be that due to the redistributed charge rather than that due to the original charge distribution.

In the special case where one of the charges of the charge distribution can be used as a test charge, the limiting process is unnecessary. In this case the electric field at the location of the test charge will be that produced by

FIGURE 2-2

Geometry of \mathbf{r} , \mathbf{r}' and $\mathbf{r} - \mathbf{r}'$. The vector \mathbf{r} defines the observation point (field point), and \mathbf{r}' ranges over the entire charge distribution, including point charges.



all of the rest of the charge distribution. There will be no redistribution of charge, since the proper charge distribution obtains under the influence of the entire charge distribution, including the charge being used as test charge. In certain other cases, notably those in which the charge distribution is *specified*, the force will be proportional to the size of the test charge. In these cases, too, the limit is unnecessary; however, if any doubt exists, it is always safe to use the limiting process.

Equations (2-2) and (2-5) provide a ready means for obtaining an expression for the electric field due to a given distribution of charge (refer to Fig. 2-2). Let the charge distribution consist of N point charges q_1, q_2, \dots, q_N located at the points $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$, respectively, and a volume distribution of charge specified by the charge density $\rho(\mathbf{r}')$ in the volume V and a surface distribution characterized by the surface charge density $\sigma(\mathbf{r}')$ on the surface S . If a test charge q is located at the point \mathbf{r} , it experiences a force \mathbf{F} given by

$$\begin{aligned} \mathbf{F} = & \frac{q}{4\pi\epsilon_0} \sum_{i=1}^N q_i \frac{\mathbf{r} - \mathbf{r}_i}{|\mathbf{r} - \mathbf{r}_i|^3} + \frac{q}{4\pi\epsilon_0} \int_V \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} \rho(\mathbf{r}') dv' \\ & + \frac{q}{4\pi\epsilon_0} \int_S \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} \sigma(\mathbf{r}') da' \end{aligned} \quad (2-7)$$

due to the given charge distribution. The electric field at \mathbf{r} is the limit of the ratio of this force to the test charge q . Since the ratio is independent of q , the electric field at \mathbf{r} is just

$$\begin{aligned} \mathbf{E}(\mathbf{r}) = & \frac{1}{4\pi\epsilon_0} \sum_{i=1}^N q_i \frac{\mathbf{r} - \mathbf{r}_i}{|\mathbf{r} - \mathbf{r}_i|^3} + \frac{1}{4\pi\epsilon_0} \int_V \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} \rho(\mathbf{r}') dv' \\ & + \frac{1}{4\pi\epsilon_0} \int_S \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} \sigma(\mathbf{r}') da' \end{aligned} \quad (2-8)$$

Equation (2-8) is very general; in most cases one or more of the terms will not be needed.

The quantity we have just defined, the electric field, may be calculated at each point in space in the vicinity of a system of charges or of a charge distribution. Thus $\mathbf{E} = \mathbf{E}(\mathbf{r})$ is a vector point function, or a vector field. This field has a number of interesting mathematical properties, which we shall proceed to develop in the following sections and in the next chapter. As an aid to visualizing the electric field structure associated with a particular distribution of charge, Michael Faraday (1791–1867) introduced the concept of *lines of force*. A line of force is an imaginary line (or curve) drawn in such a way that its direction (or tangent) at any point is the direction of the electric field at that point.

Consider, for example, the electric field structure associated with a single positive point charge q_1 . The lines of force are radial lines radiating outward from q_1 . Similarly, the lines of force associated with an isolated negative point charge are also radial lines, but this time the direction is inward (i.e., toward the negative charge). These two examples are extremely simple, but they nevertheless illustrate an important property of the field lines: The lines of force terminate on the sources of the electric field—that is, on the charges producing the electric field.

Figure 2-3 shows two simple electric fields that have been mapped with the aid of lines of force.

2-4 THE ELECTROSTATIC POTENTIAL

As noted in Chapter 1, if the curl of a vector vanishes, then the vector may be expressed as the gradient of a scalar. The electric field given by Eq. (2-8) satisfies this criterion. We note that taking the curl of Eq. (2-8) involves differentiating with respect to r . This variable appears in the equation only in functions of the form $(\mathbf{r} - \mathbf{r}')/|\mathbf{r} - \mathbf{r}'|^3$, and hence it will suffice to show that functions of this form have zero curl. Using the formula from Table 1-1 for the curl of the product (vector times scalar) gives

$$\nabla \times \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} = \frac{1}{|\mathbf{r} - \mathbf{r}'|^3} \nabla \times (\mathbf{r} - \mathbf{r}') + \left[\nabla \frac{1}{|\mathbf{r} - \mathbf{r}'|^3} \right] \times [\mathbf{r} - \mathbf{r}'] \quad (2-9)$$

A direct calculation (see Problem 1-19) shows that

$$\nabla \times (\mathbf{r} - \mathbf{r}') = 0 \quad (2-10)$$

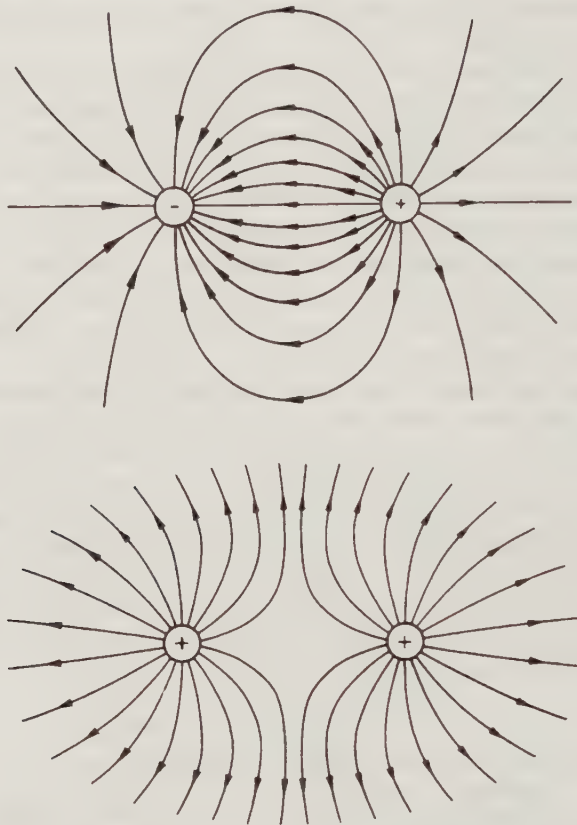
and (see Problem 1-22) that

$$\nabla \frac{1}{|\mathbf{r} - \mathbf{r}'|^3} = -3 \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^5} \quad (2-11)$$

These results, together with the observation that the vector product of a

FIGURE 2-3

The mapping of an electric field with the aid of lines of force



vector with a parallel vector is zero, suffice to prove that

$$\nabla \times \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} = 0 \quad (2-12)$$

Since each contribution of Eq. (2-8) to the electric field is of this form, we have demonstrated that the curl of the electric field is zero. Equation (2-12) indicates that a scalar function exists whose gradient is the electric field. It remains to find this function. That is, we now know that a function exists that satisfies

$$\mathbf{E}(\mathbf{r}) = -\nabla\varphi(\mathbf{r}) \quad (2-13)$$

but we have yet to find the form of the function φ . It should be noted that it is conventional to include the minus sign in Eq. (2-13) and to call φ the electrostatic potential.

The *electrostatic potential* due to a point charge q_1 is

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{q_1}{|\mathbf{r} - \mathbf{r}_1|} \quad (2-14)$$

which is readily verified by direct differentiation. Thus, the potential that corresponds to the electric field of Eq. (2-8) is

$$\begin{aligned} \varphi(\mathbf{r}) = & \frac{1}{4\pi\epsilon_0} \sum_{i=1}^N \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|} + \frac{1}{4\pi\epsilon_0} \int_V \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dv' \\ & + \frac{1}{4\pi\epsilon_0} \int_S \frac{\sigma(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} da' \end{aligned} \quad (2-15)$$

which is also easily verified by direct differentiation.

It may seem that Eqs. (2-14) and (2-15) were obtained in a rather arbitrary fashion. However, since all that is required of φ is that it satisfy Eq. (2-13), which has been verified directly, the means by which φ was obtained is immaterial.

The electrostatic potential φ can be obtained directly as soon as its existence is established. Since φ is known to exist, we may write

$$\int_{\text{ref}}^{\mathbf{r}} \mathbf{E}(\mathbf{r}') \cdot d\mathbf{r}' = - \int_{\text{ref}}^{\mathbf{r}} \nabla\varphi \cdot d\mathbf{r}' \quad (2-16)$$

where ref stands for a reference point at which φ is zero. From the definition of the gradient,

$$\nabla\varphi \cdot d\mathbf{r}' = d\varphi \quad (2-17)$$

Using Eq. (2-17) in Eq. (2-16) converts it into the integral of a perfect differential, which is easily done. The result is

$$- \int_{\text{ref}}^{\mathbf{r}} \nabla\varphi \cdot d\mathbf{r}' = -\varphi(\mathbf{r}) = \int_{\text{ref}}^{\mathbf{r}} \mathbf{E}(\mathbf{r}') \cdot d\mathbf{r}' \quad (2-18)$$

which is really the inverse of Eq. (2-13). If the electric field due to a point charge is used in Eq. (2-18), and the reference point or lower limit in the integral is taken at infinity, with the potential there zero, the result is

$$\varphi(\mathbf{r}) = \frac{q}{4\pi\epsilon_0 r} \quad (2-19)$$

This result is just a special case of Eq. (2-14), namely, the case where \mathbf{r}_1 is zero. While this derivation can be extended to obtain Eq. (2-15), the procedure is too cumbersome to include here.

Another interesting and useful aspect of the electrostatic potential is its close relation to the potential energy associated with the conservative

electrostatic force. The potential energy associated with an arbitrary conservative force is

$$U(\mathbf{r}) = - \int_{\text{ref}}^{\mathbf{r}} \mathbf{F}(\mathbf{r}') \cdot d\mathbf{r}' \quad (2-20)$$

where $U(\mathbf{r})$ is the potential energy at \mathbf{r} relative to the reference point at which the potential energy is arbitrarily taken to be zero. Since in the electrostatic case $\mathbf{F} = q\mathbf{E}$, it follows that if the same reference point is chosen for the electrostatic potential and for the potential energy, then the electrostatic potential is just the potential energy per unit charge. This idea is sometimes used to introduce the electrostatic potential, but we feel that the introduction by means of Eq. (2-13) emphasizes the importance of the electrostatic potential in determining the electrostatic field. There is, of course, no question about the equivalence of the two approaches.

The utility of the electrostatic potential in calculating electric fields can be seen by contrasting Eqs. (2-8) and (2-15). Equation (2-8) is a vector equation; to obtain the electric field from it, it is necessary to evaluate three sums or three integrals for each term. At best, this procedure is tedious, and in some cases the integrals are almost impossible to do. Equation (2-15), on the other hand, is a scalar equation and involves only one sum or integral per term. Furthermore, the denominators appearing in this equation are all of the form $|\mathbf{r} - \mathbf{r}'|$, which simplifies the integrals compared with those of Eq. (2-8). This simplification is sometimes sufficient to make the difference between doing the integrals and not doing them. The objection that after doing the integrals of Eq. (2-15) it is still necessary to differentiate the result is readily answered by observing that differentiation can always be accomplished if the derivatives exist, and is in fact usually much easier than integration. In Chapter 3, it will be seen that the electrostatic potential is even more important in those problems where the charge distribution is not specified, but must rather be determined in the process of solving the problem.

In the mks system the unit of energy is the newton-meter or joule. The unit of potential is joule/coulomb, but this unit occurs so frequently that it is given a special name, the volt (V). The unit of the electric field is the newton/coulomb or the volt/meter.

CONDUCTORS AND INSULATORS

So far as their electrostatic behavior is concerned, materials may be divided into two categories: conductors of electricity and insulators (dielectrics). *Conductors* are substances, like the metals, which contain large numbers of essentially free charge carriers. These charge carriers (electrons in most cases) are free to wander throughout the conducting material; they respond to almost infinitesimal electric fields; and they continue to move as long as

they experience a field. These free carriers carry the electric current when a steady electric field is maintained in the conductor by an external source of energy.

Dielectrics are substances in which all charged particles are bound rather strongly to constituent molecules. The charged particles may shift their positions slightly in response to an electric field, but they do not leave the vicinity of their molecules. Strictly speaking, this definition applies to an ideal dielectric—that is, one showing no conductivity in the presence of an externally maintained electric field. Real physical dielectrics show a feeble conductivity, but in a typical dielectric, the conductivity is 10^{20} times smaller than that of a good conductor. Since 10^{20} is a tremendous factor, it is usually sufficient to say that dielectrics are nonconductors.

Certain materials (semiconductors, electrolytes) have electrical properties intermediate between conductors and dielectrics. So far as their behavior in a static electric field is concerned, these materials behave like conductors. However, their transient response is somewhat slower (i.e., it takes longer for these materials to reach equilibrium in a static field).

In this and the following four chapters, we shall be concerned with materials in *electrostatic* fields. Dielectric polarization, although a basically simple phenomenon, produces some rather complicated effects; hence, we shall delay its study until Chapter 4. Conductors, on the other hand, may be treated quite easily in terms of concepts that have already been developed.

Since charge is free to move in a conductor, even under the influence of very small electric fields, the charge carriers (electrons or ions) move until they find positions in which they experience no net force. When they come to rest, the interior of the conductor must be a region devoid of an electric field, because the charge carrier population in the interior is by no means depleted, and if a field persisted, the carriers would continue to move. *Thus, under static conditions, the electric field in a conductor vanishes.* Furthermore, since $\mathbf{E} = 0$ in a conductor, the potential is the same at all points in the conducting material. In other words, *under static conditions, each conductor forms an equipotential region of space.*

2-6 GAUSS'S LAW

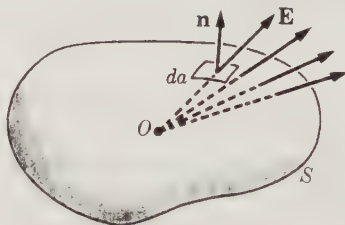
An important relationship exists between the integral of the normal component of the electric field over a closed surface and the total charge enclosed by the surface. This relationship, known as Gauss's law, will now be investigated in more detail. The electric field at point \mathbf{r} due to a point charge q located at the origin is

$$\mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\epsilon_0 r^3} \mathbf{r} \quad (2-21)$$

Consider the surface integral of the normal component of this electric field

FIGURE 2-4

An imaginary closed surface S that encloses a point charge at the origin



over a closed surface (such as that shown in Fig. 2-4) that encloses the origin and, consequently, the charge q :

$$\oint_S \mathbf{E} \cdot \mathbf{n} \, da = \frac{q}{4\pi\epsilon_0} \oint_S \frac{\mathbf{r} \cdot \mathbf{n}}{r^3} \, da \quad (2-22)$$

The quantity $(\mathbf{r}/r) \cdot \mathbf{n} \, da$ is the projection of da on a plane perpendicular to \mathbf{r} . This projected area divided by r^2 is the solid angle subtended by da , which is written $d\Omega$. It is clear from Fig. 2-5 that the solid angle subtended by da is the same as the solid angle subtended by da' , an element of the surface area of the sphere S' whose center is at the origin and whose radius is r' . It is then possible to write

$$\oint_S \frac{\mathbf{r} \cdot \mathbf{n}}{r^3} \, da = \oint_{S'} \frac{\mathbf{r}' \cdot \mathbf{n}}{r'^3} \, da' = 4\pi$$

which shows that

$$\oint_S \mathbf{E} \cdot \mathbf{n} \, da = \frac{q}{4\pi\epsilon_0} 4\pi = \frac{q}{\epsilon_0} \quad (2-23)$$

in the special case described above. If q lies outside of S , it is clear from Fig. 2-6 that S can be divided into two areas, S_1 and S_2 , each of which subtends

FIGURE 2-5

Construction of the spherical surface S' as an aid to evaluation of the solid angle subtended by da

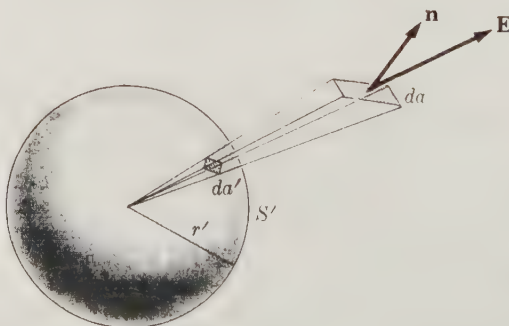


FIGURE 2-6

Closed surface S divided into two surfaces, S_1 and S_2 , each of which subtended the same solid angle at q



the same solid angle at the charge q . For S_2 , however, the direction of the normal is toward q , while for S_1 it is away from q . Therefore, the contributions of S_1 and S_2 to the surface integral are equal and opposite, and the total integral vanishes. Thus, if the surface surrounds a point charge q , the surface integral of the normal component of the electric field is q/ϵ_0 , while if q lies outside the surface, the surface integral is zero. The preceding statement applies to any closed surface, even to so-called re-entrant ones. A study of Fig. 2-7 is sufficient to verify that this is indeed the case.

If several point charges q_1, q_2, \dots, q_N are enclosed by the surface S , then the total electric field is given by the first term of Eq. (2-8). Each charge subtends a full solid angle (4π); hence Eq. (2-23) becomes

$$\oint_S \mathbf{E} \cdot \mathbf{n} \, da = \frac{1}{\epsilon_0} \sum_{i=1}^N q_i \quad (2-24)$$

This result can be readily generalized to the case of a continuous distribution of charge characterized by a charge density. If each element of charge $\rho \, dv$ is considered as a point charge, it contributes $\rho \, dv/\epsilon_0$ to the surface integral of the normal component of the electric field provided it is inside the surface over which we integrate. The total surface integral is then the sum of all contributions of this form due to the charge inside the surface.

FIGURE 2-7

An element of solid angle cutting the surface S more than once



If S is a closed surface that bounds the volume V ,

$$\oint_S \mathbf{E} \cdot \mathbf{n} \, da = \frac{1}{\epsilon_0} \int_V \rho \, dv \quad (2-25)$$

which is known as Gauss's law.

The term on the left, the integral of the normal component of the electric field over the surface S , is sometimes called the *flux* of the electric field through S .

Gauss's law may be expressed in yet another form by using the divergence theorem. The divergence theorem (1-37) states that

$$\oint_S \mathbf{F} \cdot \mathbf{n} \, da = \int_V \nabla \cdot \mathbf{F} \, dv$$

If this theorem is applied to the surface integral of the normal component of \mathbf{E} , it yields

$$\oint_S \mathbf{E} \cdot \mathbf{n} \, da = \int_V \nabla \cdot \mathbf{E} \, dv \quad (2-26)$$

which, when substituted into Eq. (2-25), gives

$$\int_V \nabla \cdot \mathbf{E} \, dv = \frac{1}{\epsilon_0} \int_V \rho \, dv \quad (2-27)$$

Equation (2-27) must be valid for all volumes—that is, for any choice of the volume V . This requirement can be satisfied only if the integrands appearing on the left and on the right in the equation are equal. Thus, the validity of Eq. (2-27) for any choice of V implies that

$$\nabla \cdot \mathbf{E} = \frac{1}{\epsilon_0} \rho \quad (2-28)$$

This result may be thought of as a differential form of Gauss's law.

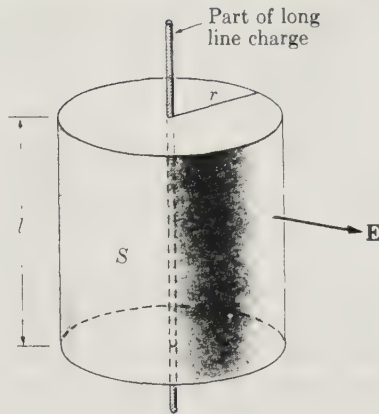
2-7

APPLICATION OF GAUSS'S LAW

Equation (2-28) or, more properly, a modified form of this equation, which will be derived in Chapter 4, is one of the basic differential equations of electricity and magnetism. While Eq. (2-28) is important in this role, Gauss's law also has practical utility. This practicality of the law lies largely in providing a very easy way to calculate electric fields in sufficiently symmetric situations. In other words, in certain highly symmetric situations of considerable physical interest, the electric field may be calculated by using Gauss's law instead of by the integrals given above or by the

FIGURE 2-8

A cylindrical surface to be used with Gauss's law to find the electric field produced by a long line charge



procedures of Chapter 3. This use of Gauss's law accomplishes a major saving in effort.

In order that Gauss's law be useful in calculating the electric field, it must be possible to choose a closed surface such that the electric field has a normal component that is either zero or a single fixed value at every point on the surface. As an example, consider a very long line charge of charge density λ per unit length, as shown in Fig. 2-8. The symmetry of the situation clearly indicates that the electric field is radial and independent of both position along the wire and angular position around the wire. These observations lead us to choose the surface shown in Fig. 2-8. For this surface, it is easy to evaluate the integral of the normal component of the electric field. The circular ends contribute nothing, since the electric field is parallel to them. The cylindrical surface contributes $2\pi r l E_r$, since \mathbf{E} is radial and independent of the position of the cylindrical surface. Gauss's law then takes the form

$$2\pi r l E_r = \frac{\lambda l}{\epsilon_0} \quad (2-29)$$

Equation (2-29) can be solved for E_r to give

$$E_r = \frac{\lambda}{2\pi\epsilon_0 r} \quad (2-30)$$

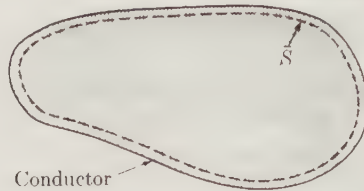
From this result, the potential can be obtained by integration:

$$\varphi = \frac{\lambda}{2\pi\epsilon_0} \ln r + C$$

The saving of effort accomplished by the use of Gauss's law will be more

FIGURE 2-9

A gaussian surface S constructed inside a charged conductor



fully appreciated by solving Problem 2-4, which involves direct application of Eq. (2-8).

Another important result of Gauss's law is that the charge (net charge) of a charged conductor resides on its outer surface. We saw in Section 2-5 that the electric field inside a conductor vanishes. We may construct a gaussian surface anywhere inside the conductor. By Gauss's law, the net charge enclosed by each of these surfaces is zero. Finally, we construct the gaussian surface S of Fig. 2-9, and again see that the net charge enclosed is zero. The only place left for the charge that is not in contradiction with Gauss's law is on the surface of the conductor. Since there is no charge in the interior, part of the material could be removed without changing anything. Thus, the charge of a conducting *shell* must reside entirely on the *outer* surface.

The electric field just outside a charged conductor must be normal to the surface of the conductor. This fact follows because the surface is an equipotential, and $\mathbf{E} = -\nabla\phi$. Let us assume that the charge on a conductor is given by the surface density function σ . If Gauss's law is applied to the small pillbox-shaped surface S of Fig. 2-10, then

$$E \Delta S = \left(\frac{\sigma}{\epsilon_0} \right) \Delta S$$

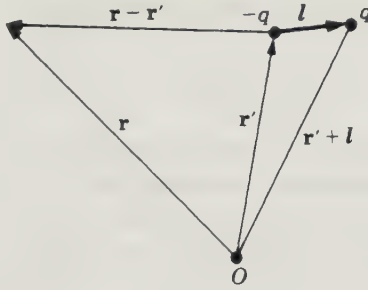
where ΔS is the area of one of the pillbox bases. Hence, for the electric field

FIGURE 2-10

Application of Gauss's law to the closed, pillbox-shaped surface S that intersects the surface of a charged conductor



FIGURE 2-11
Geometry involved in calculating the electric field $\mathbf{E}(\mathbf{r})$ due to two point charges



just outside a conductor,

$$E = \frac{\sigma}{\epsilon_0} \quad (2-31)$$

2-8 THE ELECTRIC DIPOLE

Two equal and opposite charges separated by a small distance form an electric dipole. The electric field and potential distribution produced by such a charge configuration can be investigated with the aid of the formulas of Sections 2-3 and 2-4. Suppose that a charge $-q$ is located at the point \mathbf{r}' and a charge q is located at $\mathbf{r}' + \mathbf{l}$, as shown in Fig. 2-11, then the electric field at an arbitrary point \mathbf{r} may be found by direct application of Eq. (2-8). The electric field at \mathbf{r} is found to be

$$\mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\epsilon_0} \left\{ \frac{\mathbf{r} - \mathbf{r}' - \mathbf{l}}{|\mathbf{r} - \mathbf{r}' - \mathbf{l}|^3} - \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} \right\} \quad (2-32)$$

The correct electric field for any value of q and any value of the separation l is given by Eq. (2-32); however, it is not easy to interpret. What we want is the dipole field, and in the dipole the separation l is small compared with $\mathbf{r} - \mathbf{r}'$; hence, we may expand Eq. (2-32), keeping only the first nonvanishing term. Since this procedure is of general utility, we will consider it in detail. The primary difficulty in making this expansion is caused by the denominator of the first term of Eq. (2-32). The reciprocal of this denominator can be rewritten as

$$\begin{aligned} |\mathbf{r} - \mathbf{r}' - \mathbf{l}|^{-3} &= [(\mathbf{r} - \mathbf{r}')^2 - 2(\mathbf{r} - \mathbf{r}') \cdot \mathbf{l} + l^2]^{-3/2} \\ &= |\mathbf{r} - \mathbf{r}'|^{-3} \left[1 - \frac{2(\mathbf{r} - \mathbf{r}') \cdot \mathbf{l}}{|\mathbf{r} - \mathbf{r}'|^2} + \frac{l^2}{|\mathbf{r} - \mathbf{r}'|^2} \right]^{-3/2} \end{aligned}$$

In the last form it is easy to expand by the binomial theorem, keeping only

terms linear in l . The result of this expansion is

$$|\mathbf{r} - \mathbf{r}' - l|^{-3} = |\mathbf{r} - \mathbf{r}'|^{-3} \left\{ 1 + \frac{3(\mathbf{r} - \mathbf{r}') \cdot l}{|\mathbf{r} - \mathbf{r}'|^2} + \dots \right\} \quad (2-33)$$

where terms involving l^2 have been dropped. Using Eq. (2-33) in Eq. (2-32) and again keeping only terms linear in l gives

$$\mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\epsilon_0} \left\{ \frac{3(\mathbf{r} - \mathbf{r}') \cdot l}{|\mathbf{r} - \mathbf{r}'|^5} (\mathbf{r} - \mathbf{r}') - \frac{l}{|\mathbf{r} - \mathbf{r}'|^3} + \dots \right\} \quad (2-34)$$

Equation (2-34) gives that part of the electric field, due to a finite electric dipole, which is proportional to the separation of the charges. There are other contributions proportional to the square, the cube, and higher powers of the separation. If, however, the separation is small, these higher powers contribute very little. In the limit as l goes to zero, all of the terms vanish unless the charge becomes infinite. In this limit, with the product ql remaining constant, all terms except the term linear in l vanish. In this limit a *point dipole* is formed. A point dipole has no net charge, no extent in space, and is completely characterized by its dipole moment, which is the limit of ql as l goes to zero. We use the symbol \mathbf{p} to represent the electric dipole moment and write

$$\mathbf{p} = ql \quad (2-35)$$

In terms of the dipole moment, Eq. (2-34) may be written

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \left\{ \frac{3(\mathbf{r} - \mathbf{r}') \cdot \mathbf{p}}{|\mathbf{r} - \mathbf{r}'|^5} (\mathbf{r} - \mathbf{r}') - \frac{\mathbf{p}}{|\mathbf{r} - \mathbf{r}'|^3} \right\} \quad (2-36)$$

The potential distribution produced by a point dipole is also important and could be found by looking for a function with gradient equal to the right side of Eq. (2-36). It is, however, easier to apply Eq. (2-15) to the charge distribution consisting of two point charges separated by a small distance. Using the notation of Eq. (2-32), the potential distribution is given by

$$\varphi(\mathbf{r}) = \frac{q}{4\pi\epsilon_0} \left[\frac{1}{|\mathbf{r} - \mathbf{r}' - l|} - \frac{1}{|\mathbf{r} - \mathbf{r}'|} \right] \quad (2-37)$$

By expanding the first term in exactly the same way as was done for the first term of Eq. (2-32) and retaining only the linear term in l , Eq. (2-37) can be put in the form

$$\varphi(\mathbf{r}) = \frac{q}{4\pi\epsilon_0} \frac{(\mathbf{r} - \mathbf{r}') \cdot l}{|\mathbf{r} - \mathbf{r}'|^3} \quad (2-38)$$

This equation is valid to the same approximation as Eq. (2-34); namely, terms proportional to l^2 and to higher powers of l are neglected. For a point

dipole, \mathbf{p} , Eq. (2-38) is exact; however, it is better written as

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{\mathbf{p} \cdot (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} \quad (2-39)$$

Equation (2-39) gives the potential $\varphi(\mathbf{r})$ produced by an electric dipole, and from this potential, the electric field (Eq. 2-36) can be determined.

It is also interesting to inquire about the potential energy of an electric dipole placed in an *external electric field*. In the case of two charges, $-q$ at \mathbf{r} and q at $\mathbf{r} + \mathbf{l}$, in an electric field described by the potential function $\varphi_{\text{ext}}(\mathbf{r})$, the potential energy is

$$U = -q\varphi_{\text{ext}}(\mathbf{r}) + q\varphi_{\text{ext}}(\mathbf{r} + \mathbf{l}) \quad (2-40)$$

If \mathbf{l} is small compared with \mathbf{r} , $\varphi_{\text{ext}}(\mathbf{r} + \mathbf{l})$ may be expanded in a power series in \mathbf{l} and only the first two terms kept. The expansion gives

$$\varphi_{\text{ext}}(\mathbf{r} + \mathbf{l}) = \varphi_{\text{ext}}(\mathbf{r}) + \mathbf{l} \cdot \nabla \varphi_{\text{ext}} \quad (2-41)$$

where the value of the gradient at point \mathbf{r} is to be used. If this expansion is used in Eq. (2-40), the result is

$$U = q\mathbf{l} \cdot \nabla \varphi_{\text{ext}} \quad (2-42)$$

Going to the limit of a point dipole gives simply

$$U(\mathbf{r}) = \mathbf{p} \cdot \nabla \varphi_{\text{ext}} \quad (2-43)$$

which is, of course, exact. Since the electric field is the negative gradient of the electrostatic potential, an alternative form of Eq. (2-43) is

$$U(\mathbf{r}) = -\mathbf{p} \cdot \mathbf{E}_{\text{ext}}(\mathbf{r}) \quad (2-44)$$

$U(\mathbf{r})$ is the potential energy of a dipole \mathbf{p} in an external electric field \mathbf{E}_{ext} , where $\mathbf{E}_{\text{ext}}(\mathbf{r})$ is evaluated at the location of the dipole.

It is important to note that two potentials have been discussed in this section. In Eqs. (2-37), (2-38), and (2-39), the electrostatic potential *produced* by an electric dipole is considered. In Eqs. (2-40) through (2-43), the dipole is considered to be in an *existing electric field* described by a potential function $\varphi_{\text{ext}}(\mathbf{r})$. This electric field is due to charges other than those comprising the dipole; in fact, the dipole field must be excluded to avoid an infinite result. This statement could lead us to rather complicated questions concerning self-forces and self-energies that we cannot discuss here; however, it may be noted that the potential energy resulting from the interaction of an electric dipole with its own field arises from forces exerted on the dipole by itself. Such forces, known in dynamics as internal forces, do not affect the motion of the dipole as a whole. For our purposes, further consideration of this question will be unnecessary.

2-9 MULTIPOLE EXPANSION OF ELECTRIC FIELDS

It is apparent from the definition of dipole moments given previously that certain aspects of the potential distribution produced by a specified distribution of charge might well be expressed in terms of its electric dipole moment. In order to do so, it is necessary to define the electric dipole moment of an arbitrary charge distribution. Rather than make an unmotivated definition, we shall consider a certain expansion of the electrostatic potential due to an arbitrary charge distribution. To reduce the number of position coordinates, a charge distribution in the neighborhood of the origin of coordinates will be considered. The further restriction will be made that the charge distribution can be entirely enclosed by a sphere of radius a , which is small compared with the distance to the point of observation. An arbitrary point within the charge distribution will be designated by \mathbf{r}' , the charge density at that point by $\rho(\mathbf{r}')$, and the observation point by \mathbf{r} (see Fig. 2-12). The potential at \mathbf{r} is given by

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int_V \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dv' \quad (2-45)$$

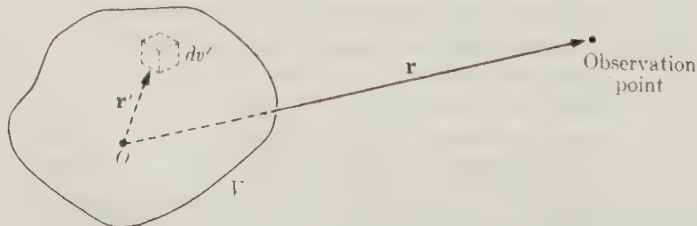
where dv' is used to designate an element of volume in the charge distribution and V denotes the entire volume occupied by the charge distribution. In view of the restriction made above to points of observation that are remote from the origin, the quantity $|\mathbf{r} - \mathbf{r}'|^{-1}$ can be expanded in a series of ascending powers of \mathbf{r}'/r . The result of such an expansion is

$$\begin{aligned} |\mathbf{r} - \mathbf{r}'|^{-1} &= (r^2 - 2\mathbf{r} \cdot \mathbf{r}' + r'^2)^{-1/2} \\ &= \frac{1}{r} \left\{ 1 - \frac{1}{2} \left[-\frac{2\mathbf{r} \cdot \mathbf{r}'}{r^2} + \frac{r'^2}{r^2} \right] + \frac{113}{222} \left[\quad \right]^2 + \dots \right\} \end{aligned} \quad (2-46)$$

where only the first three terms are explicitly indicated. It should be noted that while $(r'/r)^2$ is negligible compared with $2\mathbf{r}' \cdot \mathbf{r}/r^2$, it may not be dropped in the first set of brackets because it is of the same order as the dominant term in the second set of brackets. Using Eq. (2-46) in Eq. (2-45)

FIGURE 2-12

Localized charge in the volume V with charge density $\rho(\mathbf{r}')$. The electric field is to be calculated at point \mathbf{r} .



and omitting terms involving the cube and higher powers of \mathbf{r}' yields

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int_V \left\{ \frac{1}{r} + \frac{\mathbf{r} \cdot \mathbf{r}'}{r^3} + \frac{1}{2} \left[\frac{3(\mathbf{r} \cdot \mathbf{r}')^2}{r^5} - \frac{r'^2}{r^3} \right] + \dots \right\} \rho(\mathbf{r}') dv' \quad (2-47)$$

Since \mathbf{r} does not involve the variable of integration \mathbf{r}' , all of the \mathbf{r} dependence may be taken from under the integral sign to obtain

$$\begin{aligned} \varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \left\{ \frac{1}{r} \int_V \rho(\mathbf{r}') dv' + \frac{\mathbf{r}}{r^3} \cdot \int_V \mathbf{r}' \rho(\mathbf{r}') dv' \right. \\ \left. + \sum_{i=1}^3 \sum_{j=1}^3 \frac{1}{2} \frac{x_i x_j}{r^5} \int_V (3x'_i x'_j - \delta_{ij} r'^2) \rho(\mathbf{r}') dv' \right\} \end{aligned} \quad (2-48)$$

where x_i, x_j are Cartesian components of \mathbf{r} ; x'_i, x'_j are the Cartesian components of \mathbf{r}' ; and δ_{ij} is defined as follows:

$$\delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

It is easy to interpret Eq. (2-48). The first integral in the equation is clearly the total charge, and the first term is the potential that would result if this total charge were concentrated at the origin.

The second integral is a straightforward generalization of the dipole moment concept to a continuous charge distribution and is called the *dipole moment of the charge distribution*. The second term in Eq. (2-48) is the potential that would result if a point dipole equal to the dipole moment of the charge distribution were located at the origin of coordinates. It is interesting to note that the dipole moment of a charge distribution is independent of the origin of coordinates if the total charge is zero. To verify this statement, consider a new coordinate system with *origin* at \mathbf{R} in the old system. Denoting a point with respect to the old system by \mathbf{r}' and the same point with respect to the new system by \mathbf{r}'' , we have

$$\mathbf{r}' = \mathbf{r}'' + \mathbf{R} \quad (2-49)$$

The dipole moment with respect to the old system is

$$\mathbf{p} = \int_V \mathbf{r}' \rho(\mathbf{r}') dv' = \int_V (\mathbf{r}'' + \mathbf{R}) \rho(\mathbf{r}') dv' = \int_V \mathbf{r}'' \rho dv' + \mathbf{R} Q \quad (2-50)$$

Thus, \mathbf{p} is independent of the origin if Q is zero.

The third term of Eq. (2-48) can be written

$$\sum_{i=1}^3 \sum_{j=1}^3 \frac{1}{2} \frac{x_i x_j}{r^5} Q_{ij} \quad (2-51)$$

where

$$Q_{ij} = \int_V (3x'_i x'_j - \delta_{ij} r'^2) \rho(\mathbf{r}') dv' \quad (2-52)$$

There are nine components of Q_{ij} corresponding to i, j equal to 1, 2, 3. Of these nine components, six are equal in pairs, leaving six distinct components. This set of quantities forms the quadrupole moment tensor* and represents an extension of the dipole moment concept. There are higher-order moments generated by keeping higher-order terms in the expansion of Eq. (2-48). These higher-order multipoles are important in nuclear physics, but will not be considered further in this book.

The electric multipoles are used, as Eq. (2-48) indicates, to approximate the potential (and electric field) of a charge distribution. There are, however, many other uses—all in the framework of approximating a real extended charge distribution by point charges, point dipoles, and so forth. These approximations often make it possible to solve extremely difficult problems.

2-10
THE DIRAC DELTA FUNCTION†

In the general expressions for electric field and potential in Eqs. (2-8) and (2-15), we have distinguished between point charges and continuous charge distributions. For economy of notation, if for no other reason, it would be useful to be able to express point charges as a special case of a general charge-density function $\rho(\mathbf{r})$. The *Dirac delta function* $\delta(\mathbf{r})$ can serve this purpose, and in addition it is a valuable mathematical tool in many calculations. We write

$$\rho(\mathbf{r}) = q\delta(\mathbf{r}) \quad (\text{point charge}) \quad (2-53)$$

where

$$\delta(\mathbf{r}) = 0 \quad \text{for } \mathbf{r} \neq 0$$

and

$$\int \delta(\mathbf{r}') dv' = 1 \quad (2-54)$$

The delta function, thus, gives a mathematical expression to the physical idea of a point charge at $\mathbf{r} = 0$: The integrated charge density is q , but all of the charge is located exactly at the origin. The delta function is obviously a very highly singular mathematical function if it can be zero everywhere

* Tensors are a generalization of vectors and an elementary discussion is given in Appendix II.

† Other properties of the delta function are discussed in Appendix V.

except at a single point and yet have a nonzero integral. Nevertheless it is a legitimate mathematical object, which leads to no difficulties if one is cautious, for example, by not trying to differentiate it like a continuous function. A variation of it can be made to represent a surface charge density $\sigma(\mathbf{r})$ —that is, a charge distribution that vanishes everywhere except on a certain surface. With these extensions, the single integral over $\rho(\mathbf{r})$ is enough in Eqs. (2-8) and (2-15). For this application, we note that

$$\int F(\mathbf{r}')\delta(\mathbf{r}') dv' = F(0) \quad (2-55)$$

where F is any scalar or vector function, since the integrand vanishes except at $\mathbf{r}' = 0$. Furthermore,

$$\int F(\mathbf{r}')\delta(\mathbf{r}' - \mathbf{r}_0) dv' = F(\mathbf{r}_0) \quad (2-56)$$

Thus, if $\rho(\mathbf{r}') = q_i\delta(\mathbf{r}' - \mathbf{r}_i)$,

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int \frac{q_i\delta(\mathbf{r}' - \mathbf{r}_i)}{|\mathbf{r} - \mathbf{r}'|} dv' = \frac{1}{4\pi\epsilon_0} \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|}$$

and

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int \frac{q_i\delta(\mathbf{r}' - \mathbf{r}_i)}{|\mathbf{r} - \mathbf{r}'|^3} (\mathbf{r} - \mathbf{r}') dv' = \frac{1}{4\pi\epsilon_0} \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|^3} (\mathbf{r} - \mathbf{r}_i)$$

for a point charge q_i at \mathbf{r}_i . With this characterization of point charges, it is not necessary to include a specific summation over q_i in Eqs. (2-8) and (2-15).

Some other properties of the delta function can be obtained as consequences of Gauss's law in differential form:

$$\nabla \cdot \mathbf{E} = \frac{1}{\epsilon_0} \rho \quad (2-28)$$

For a point charge q at $\mathbf{r} = 0$, with Eq. (2-21)

$$\nabla \cdot \frac{q}{4\pi\epsilon_0} \frac{\mathbf{r}}{r^3} = \frac{1}{\epsilon_0} q\delta(\mathbf{r})$$

or

$$\nabla \cdot \frac{\mathbf{r}}{r^3} = 4\pi\delta(\mathbf{r}) \quad (2-57)$$

Also, since

$$\nabla \frac{1}{r} = \frac{\mathbf{r}}{r} \frac{d}{dr} \left(\frac{1}{r} \right) = -\frac{\mathbf{r}}{r^3}$$

then,

$$\nabla^2\left(\frac{1}{r}\right) = -4\pi\delta(\mathbf{r}) \quad (2-58)$$

Equations (2-57) and (2-58) are important enough that a more direct derivation is worthwhile. A straightforward differentiation shows that the divergence is zero everywhere but at the origin:

$$\nabla \cdot \frac{\mathbf{r}}{r^3} = \nabla\left(\frac{1}{r^3}\right) \cdot \mathbf{r} + \frac{1}{r^3} \nabla \cdot \mathbf{r} = -\frac{3}{r^4} \frac{\mathbf{r}}{r} \cdot \mathbf{r} + \frac{3}{r^3} = 0, \quad r \neq 0$$

At $r = 0$ the divergence is $-\infty + \infty$, which is indeterminate. The divergence theorem, however, applied to a small sphere of radius R around the origin, gives

$$\int_V \nabla \cdot \frac{\mathbf{r}}{r^3} dv = \oint_S \frac{\mathbf{r} \cdot \mathbf{n}}{r^3} da = \frac{1}{R^2} \oint_S da = 4\pi$$

Since the volume integral is 4π no matter how small the radius R , the integrand can be represented as $4\pi\delta(\mathbf{r})$, in agreement with Eq. (2-57). In other words, the delta function permits the divergence theorem to be applied to \mathbf{r}/r^3 , even in a region that contains the singularity at the origin. The delta function is extremely useful whenever an integral over the divergence of \mathbf{r}/r^3 , or over the Laplacian of $1/r$, is encountered.

2-11
SUMMARY

Electrostatics is based on **Coulomb's law**, which for a point charge q_1 at the origin and a point charge q at \mathbf{r} gives the electrostatic force on q as

$$\mathbf{F}_e = \frac{1}{4\pi\epsilon_0} \frac{qq_1 \mathbf{r}}{r^2 r}$$

where $\epsilon_0 = 8.854 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2$ in mks units and $\epsilon_0 = 1/4\pi$ in gaussian units. It is convenient to treat q as a test charge and abstract from it by defining the electric field \mathbf{E} corresponding to the electric force \mathbf{F}_e :

$$\mathbf{F}_e = q\mathbf{E}$$

The **electrostatic field** at \mathbf{r} due to the source charge q_1 located at $\mathbf{r}_1 = 0$ is

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{q_1 \mathbf{r}}{r^2 r}$$

The curl and divergence of \mathbf{E} are both of fundamental importance:

$$\nabla \times \frac{\mathbf{r}}{r^3} = 0$$

$$\nabla \cdot \frac{\mathbf{r}}{r^3} = 4\pi\delta(\mathbf{r})$$

where the **Dirac delta function** is defined by

$$\delta(\mathbf{r}) = 0, \quad \mathbf{r} \neq 0$$

$$\int \delta(\mathbf{r}) dv = 1$$

The delta function has the property that for any function F

$$\int F(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}_0) dv = F(\mathbf{r}_0)$$

• Coulomb's law can be generalized to systems of many source charges, or to a **continuous distribution of charge density** $\rho(\mathbf{r})$ defined so that the element of charge in a volume element dv is

$$dq = \rho(\mathbf{r}) dv$$

For a point charge q_i at \mathbf{r}_i ,

$$\rho(\mathbf{r}) = q_i \delta(\mathbf{r} - \mathbf{r}_i)$$

Since the forces and fields are additive,

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} \rho(\mathbf{r}') dv'$$

Since ∇ is a linear operator,

$$\nabla \times \mathbf{E} = 0$$

$$\nabla \cdot \mathbf{E} = \frac{1}{\epsilon_0} \rho(\mathbf{r})$$

These are the fundamental differential equations that must be satisfied locally at every point by all electrostatic fields. (In fact, the divergence equation is satisfied even by time-dependent fields, and it is *one of the four fundamental Maxwell equations.*)

• **Gauss's law** follows from the divergence equation by integrating both sides over an arbitrary volume V and applying the divergence theorem:

$$\oint_S \mathbf{E} \cdot \mathbf{n} da = \frac{1}{\epsilon_0} Q$$

where

$$Q = \int_V \rho(\mathbf{r}) dv$$

is the total charge inside V bounded by S . Gauss's law has practical usefulness for calculating E in a few special situations where it can be argued on general grounds that \mathbf{E} must be constant in magnitude and

direction relative to some chosen surface S . It also shows that the *charge on a conductor* must lie *on its outer surface* and that just outside the surface, $E = \sigma/\epsilon_0$.

- The existence of an **electrostatic potential** function $\varphi(\mathbf{r})$ follows from the curl equation, such that

$$\mathbf{E} = -\nabla\varphi$$

For a given \mathbf{E} -field,

$$\varphi(\mathbf{r}) = -\int_{\text{ref}}^{\mathbf{r}} \mathbf{E} \cdot d\mathbf{l}$$

For a specified charge distribution,

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dv'$$

This equation is easier to evaluate than the integral for \mathbf{E} . The scalar potential φ is related to the **potential energy** U of the conservative electrostatic force by

$$U = q\varphi$$

- At some distance away from the region where the source charges ρ are located, the **multipole expansion** of φ is useful:

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \left[\frac{Q}{r} + \frac{\mathbf{p} \cdot \mathbf{r}}{r^3} + \dots \right]$$

where

$$\mathbf{p} = \int_V \mathbf{r}' \rho(\mathbf{r}') dv'$$

is the dipole moment of the charge distribution. Usually the first nonvanishing term in the expansion is most important; we consider only the first two terms.

PROBLEMS

2-1 Two particles, each of mass m and having charges q and $2q$, respectively, are suspended by strings of length l from a common point. Find the angle θ that each string makes with the vertical.

2-2 Two small identical conducting spheres have charges of $2.0 \times 10^{-9} \text{ C}$ and $-0.4 \times 10^{-9} \text{ C}$, respectively. (a) When they are placed 4 cm apart, what is the force between them? (b) If they are brought into contact and then separated by 4 cm, what is the force between them?

2-3 Point charges of $4 \times 10^{-9} \text{ C}$ are situated at each of three corners of a square whose side is 15 cm. Find the magnitude and direction of the electric field at the vacant corner point of the square.

- 2-4** An infinitely long line charge has uniform charge density λ per unit length. Using direct integration, find the electric field at a distance r from the line.
- 2-5** (a) A circular disk of radius R has a uniform surface charge density σ . Find the electric field at a point on the axis of the disk at a distance z from the plane of the disk. (b) A right circular cylinder of radius R and height L is oriented along the z -axis. It has a nonuniform volume density of charge given by $\rho(z) = \rho_0 + \beta z$ with reference to an origin at the center of the cylinder. Find the force on a point charge q placed at the center of the cylinder.
- 2-6** A thin, conducting, spherical shell of radius R is charged uniformly with total charge Q . By direct integration, find the potential at an arbitrary point (a) inside the shell, (b) outside the shell.
- 2-7** (a) Two point charges, $-q$ and $+\frac{1}{3}q$, are situated at the origin and at the point $(a, 0, 0)$ respectively. At what point along the x -axis does the electric field vanish? (b) In the x, y -plane make a plot of the equipotential surface which goes through the point just referred to. Is this point a true minimum in the potential?
- 2-8** Show that the $\varphi = 0$ equipotential surface of the preceding problem is spherical in shape. What are the coordinates of the center of this sphere?
- 2-9** Given a right circular cylinder of radius R and length L containing a uniform charge density ρ , calculate the electrostatic potential at a point on the cylinder axis but external to the distribution.
- 2-10** Given a region of space in which the electric field is everywhere directed parallel to the x -axis, prove that the electric field is independent of the y - and z -coordinates in this region. If there is no charge in this region, prove that the field is also independent of x .
- 2-11** The dielectric strength of air (i.e., the electric field above which the air becomes conducting) is 3×10^6 V/m. (a) What is the highest possible potential of an isolated spherical conductor of radius 8 cm? (b) What is the radius of a spherical conductor which could hold 1 coulomb of charge?
- 2-12** A conducting object has an arbitrarily shaped hollow cavity in its interior. If a point charge q is introduced into the cavity, prove that the charge $-q$ is induced on the surface of the cavity. [Hint: Use Gauss's law.]
- 2-13** The electric field in the atmosphere at the earth's surface is approximately 200 V/m, directed downward. At 1400 m above the earth's surface, the electric field in the atmosphere is only 20 V/m, again directed downward. (a) What is the average charge density in the atmosphere below 1400 m? (b) Does this consist predominantly of *positive* or *negative* ions?
- 2-14** Two infinite parallel conducting plates are separated by the distance d . The plates have uniform charge densities σ and $-\sigma$, respectively, on their inside surfaces. (a) Obtain an expression for the electric field between the plates. (b) Prove that the electric field in the regions external to the plates is zero. [Note: Two charged parallel conducting plates of finite area produce essentially the same electric field in the region between them as was found above provided the dimensions of the plates are large compared with the separation d ; such an arrangement is called a *capacitor* (see Chapter 6).]
- 2-15** A spherical charge distribution has a volume charge density that is a function only of r , the distance from the center of the distribution. In other words, $\rho = \rho(r)$.

If $\rho(r)$ is as given below, determine the electric field as a function of r . Integrate the result to obtain an expression for the electrostatic potential $\varphi(r)$, subject to the restriction that $\varphi(\infty) = 0$. (a) $\rho = A/r$ with A a constant for $0 \leq r \leq R$; $\rho = 0$ for $r > R$. (b) $\rho = \rho_0$ (i.e., constant) for $0 \leq r \leq R$; $\rho = 0$ for $r > R$.

2-16 An infinitely long circular rod of radius R contains a uniform charge density ρ . Use Gauss's law to find the electric field for $r > R$ and $r < R$.

2-17 (a) Calculate the curl and divergence of \mathbf{r}/r^a . (b) What charge density $\rho(r)$ would produce the following field:

$$\mathbf{E} = \frac{q}{4\pi\epsilon_0 r^a} \mathbf{r}$$

(c) What is the potential of this field?

2-18 Suppose the exponent in the Coulomb field were not exactly 3, but $a = 3 - \delta$, where $\delta \ll 1$. Calculate the integral of $\nabla \cdot \mathbf{E}$ over a spherical volume of radius R centered on the charge q .

2-19 The *screened Coulomb potential*

$$\varphi = \frac{q}{4\pi\epsilon_0} \frac{e^{-r/\lambda}}{r}$$

is appropriate for a charge q in a semiconducting medium. Calculate the corresponding electric field and charge density.

2-20 A spherical conductor of radius R and charge Q is surrounded by a concentric spherical shell at radius $2R$. The shell is uniformly charged with charge $-2Q$. (a) Use Gauss's law to calculate the electric field at all points in space. (b) Obtain the potential at all points.

2-21 A long straight line charge of charge λ per unit length extends along the z -axis from a to ∞ . Another line charge with $-\lambda$ per unit length extends along the $-z$ -axis from $-a$ to $-\infty$. (a) By direct integration, determine the electric potential in cylindrical coordinates. (b) Find the r component of the electric field.

2-22 Using Eq. (2-39) for the potential produced by a dipole \mathbf{p} , make a plot of the traces of equipotential surfaces in a plane containing the dipole. For convenience, the dipole may be located at the origin. Use the results obtained to sketch in some of the lines of force. Compare the result with Fig. 2-3.

2-23 (a) Show that the force acting on a dipole \mathbf{p} placed in an external electric field \mathbf{E}_{ext} is $\mathbf{p} \cdot \nabla \mathbf{E}_{\text{ext}}$. (b) Show that the torque acting on the dipole in this field is

$$\boldsymbol{\tau} = \mathbf{r} \times [(\mathbf{p} \cdot \nabla) \mathbf{E}_{\text{ext}}] + \mathbf{p} \times \mathbf{E}_{\text{ext}}$$

where \mathbf{r} is the vector distance to the dipole from the point about which the torque is to be measured. The quantity $\mathbf{p} \times \mathbf{E}_{\text{ext}}$, which is independent of the point about which the torque is computed, is called the turning couple acting on the dipole.

2-24 Three charges are arranged in a linear array. The charge $-2q$ is placed at the origin, and two charges, each of $+q$, are placed at $(0, 0, l)$ and $(0, 0, -l)$, respectively. (a) Find a relatively simple expression for the potential $\varphi(\mathbf{r})$ which is valid for distances $|\mathbf{r}| \gg l$. (b) Make a plot of the equipotential surfaces in the x, z -plane.

2-25 What is the quadrupole moment tensor of the charge distribution discussed in Problem 2-24?

2-26 By using delta functions for the charge distribution of point charges, show that the dipole moment of a pair of point charges $\mathbf{p} = ql$ follows from the general definition

$$\mathbf{p} = \int \mathbf{r}' \rho(\mathbf{r}') dv'$$

2-27 Suppose a molecule is represented by a charge $-2q$ at the origin and charges $+q$ at l_1 and l_2 , with $|l_1| = |l_2| = l$. (a) Find the dipole moment of the molecule. (b) For H_2O , $l = 0.958 \times 10^{-10}$ m and the angle between l_1 and l_2 is $\theta = 105^\circ$. If $p = 6.14 \times 10^{-30}$ C · m, find the effective charge q .

2-28 Obtain the electric field of a point dipole by calculating the gradient of

$$\varphi = \frac{1}{4\pi\epsilon_0} \frac{\mathbf{p} \cdot \mathbf{r}}{r^3}$$

Solution of Electrostatic Problems

The solution to an electrostatic problem is straightforward for the case in which the charge distribution is everywhere specified, for then, as we have seen, the potential and electric field are given directly as integrals over this charge distribution:

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int \frac{dq'}{|\mathbf{r} - \mathbf{r}'|} \quad (3-1)$$

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int \frac{(\mathbf{r} - \mathbf{r}') dq'}{|\mathbf{r} - \mathbf{r}'|^3} \quad (3-2)$$

However, many of the problems encountered in practice are not of this type. If the charge distribution is not specified in advance, it may be necessary to determine the electric field *first*, before the charge distribution can be calculated. For example, an electrostatic problem may involve several conductors, with either the potential or total charge of each conductor given, but the distribution of surface charge will not be known in general, and must be obtained as part of the solution to the problem.

Our aim in this chapter is to develop an alternative approach to electrostatic problems, and to accomplish this goal, we first derive the fundamental differential equation that must be satisfied by the potential φ . For the present we shall disregard problems involving dielectric bodies; problems of this type will be solved in Chapter 4.

3-1**POISSON'S EQUATION**

All of the basic relationships that we need here were developed in the preceding chapter. First, we have the differential form of Gauss's law:

$$\nabla \cdot \mathbf{E} = \frac{1}{\epsilon_0} \rho \quad (3-3)$$

Furthermore, in a purely electrostatic field, \mathbf{E} may be expressed as minus the gradient of the potential φ :

$$\mathbf{E} = -\nabla\varphi \quad (3-4)$$

Combining Eqs. (3-3) and (3-4), we obtain

$$\nabla \cdot \nabla\varphi = -\frac{\rho}{\epsilon_0} \quad (3-5a)$$

As noted in Chapter 1, it is convenient to think of the divergence of the gradient as a single differential operator, $\nabla \cdot \nabla$ or ∇^2 , which is called the *Laplacian*.

The Laplacian is a scalar differential operator and Eq. (3-5a) is a differential equation, known as *Poisson's equation*:

$$\nabla^2\varphi = -\frac{\rho}{\epsilon_0} \quad (3-5b)$$

The operator ∇^2 involves differentiation with respect to more than one variable. Hence, Poisson's equation is a *partial differential equation* that may be solved once we know the functional dependence of $\rho(x, y, z)$ and the appropriate boundary conditions.

The operator ∇^2 , just like the ∇ , $\nabla \cdot$, and $\nabla \times$, makes no reference to any particular coordinate system. In order to solve a specific problem, we must write ∇^2 in terms of x, y, z or r, θ, ϕ , or some other set of coordinates. The choice of the particular set of coordinates is arbitrary, but substantial simplification of the problem is usually achieved by choosing a set compatible with the symmetry of the electrostatic problem. The form taken by $\nabla^2\varphi$ in various coordinate systems is easily found by first taking the gradient of φ and then operating with $\nabla \cdot$. Using specific expressions from Chapter 1 (or from Appendix IV), we have the following forms.*

Rectangular coordinates:

$$\nabla^2\varphi \equiv \frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2} \quad (3-6)$$

Spherical coordinates:

$$\nabla^2\varphi \equiv \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial\varphi}{\partial r} \right) + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial\theta} \left(\sin\theta \frac{\partial\varphi}{\partial\theta} \right) + \frac{1}{r^2 \sin^2\theta} \frac{\partial^2\varphi}{\partial\phi^2} \quad (3-7)$$

* For the form of the Laplacian in other, more complicated coordinate systems, the reader is referred to the references at the end of this chapter.

Cylindrical coordinates:

$$\nabla^2\varphi \equiv \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial \varphi}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 \varphi}{\partial \theta^2} + \frac{\partial^2 \varphi}{\partial z^2} \quad (3-8)$$

It should be noted that r and θ have different meanings in Eqs. (3-7) and (3-8). In spherical coordinates, r is the magnitude of the radius vector from the origin and θ is the polar angle. In cylindrical coordinates, r is the perpendicular distance from the cylinder axis and θ is the azimuthal angle about this axis.

3-2

LAPLACE'S EQUATION

In a certain class of electrostatic problems involving conductors, all of the charge is found either on the surface of the conductors or in the form of fixed point charges. In these cases, ρ is zero at most points in space.

Where the charge density vanishes, the Poisson equation reduces to the simpler form

$$\nabla^2\varphi = 0 \quad (3-9)$$

which is *Laplace's equation*.

Suppose we have a set of N conductors maintained at the potentials $\varphi_1, \varphi_2, \dots, \varphi_N$. Our problem is to find the potential at all points in space outside of the conductors. We can do so by finding a solution to Laplace's equation that reduces to $\varphi_1, \varphi_2, \dots, \varphi_N$ on the surfaces of the appropriate conductors. Such a solution to Laplace's equation may be shown to be unique—that is, there is no other solution to Laplace's equation that satisfies the same boundary conditions. A proof of this statement will be given below. The solution to Laplace's equation that we find in this way is not applicable to the interior of the conductors, because the conductors have surface charge, which implies a discontinuity in the gradient of φ across the surface (see Section 2-7). But we have already seen that the interior of each conductor is a region of constant potential, so the solution to our problem is complete.

We shall describe in some detail two methods for solution of Laplace's equation: The first is a method for compounding a general solution to Eq. (3-9) from particular solutions (basis functions) in a coordinate system dictated by the symmetry of the problem; the second is the method of images. In addition, a completely general solution to the problem in two dimensions will be found. Before taking up these specific procedures, however, we prove some important properties of the solution to Laplace's equation.

Theorem I. If $\varphi_1, \varphi_2, \dots, \varphi_n$ are all solutions of Laplace's equation, then

$$\varphi = C_1\varphi_1 + C_2\varphi_2 + \dots + C_n\varphi_n \quad (3-10)$$

where the C 's are arbitrary constants, is also a solution.

The proof of this theorem follows immediately from the fact that

$$\begin{aligned} \nabla^2\varphi &= \nabla^2C_1\varphi_1 + \nabla^2C_2\varphi_2 + \dots + \nabla^2C_n\varphi_n \\ &= C_1\nabla^2\varphi_1 + C_2\nabla^2\varphi_2 + \dots + C_n\nabla^2\varphi_n \\ &= 0 \end{aligned}$$

Through the use of Theorem I, we may superimpose two or more solutions of Laplace's equation in such a way that the resulting solution satisfies a given set of boundary conditions. Examples will be given in the following sections.

Theorem II: Uniqueness theorem. Two solutions of Laplace's equation that satisfy the same boundary conditions differ at most by an additive constant.

To prove this theorem, we consider the closed region V_0 exterior to the surfaces S_I, S_{II}, \dots, S_N of the various conductors in the problem and bounded on the outside by a surface S , the latter being either a surface at infinity or a real physical surface that encloses V_0 . Let us assume that φ_1 and φ_2 are two solutions of Laplace's equation in V_0 that also have the *same* boundary conditions on $S, S_I, S_{II}, \dots, S_N$. These boundary conditions may be specified by assigning values of either φ or $\partial\varphi/\partial n$ on the bounding surfaces.

We define a new function $\Phi = \varphi_1 - \varphi_2$. Obviously, $\nabla^2\Phi = \nabla^2\varphi_1 - \nabla^2\varphi_2 = 0$ in V_0 . Furthermore, either Φ or $\mathbf{n} \cdot \nabla\Phi$ vanishes on the boundaries. Let us apply the divergence theorem to the vector $\Phi\nabla\Phi$:

$$\int_{V_0} \nabla \cdot (\Phi\nabla\Phi) dv = \int_{S+S_I+\dots+S_N} \Phi\nabla\Phi \cdot \mathbf{n} da = 0$$

since the second integral vanishes. The divergence may be expanded according to Eq. (1-1-7) of Table 1-1 to give

$$\nabla \cdot (\Phi\nabla\Phi) = \Phi\nabla^2\Phi + (\nabla\Phi)^2$$

But $\nabla^2\Phi$ vanishes at all points in V_0 , so that the divergence theorem reduces

in this case to

$$\int_{V_0} (\nabla\Phi)^2 dv = 0$$

Now $(\nabla\Phi)^2$ must be either positive or zero at each point in V_0 , and since its integral is zero, it is evident that $(\nabla\Phi)^2 = 0$ is the only possibility.

The theorem is essentially proved. A function whose gradient is zero at all points cannot change: Hence, at all points in V_0 , Φ has the same value that it has on the bounding surfaces. If the boundary conditions have been given by specifying φ_1 and φ_2 on the surfaces S, S_1, \dots, S_N , then since $\Phi = 0$ on these surfaces, it vanishes throughout V_0 . If the boundary conditions are given in terms of $\partial\varphi_1/\partial n$ and $\partial\varphi_2/\partial n$, then $\nabla\Phi$ equals zero at all points in V_0 and $\nabla\Phi \cdot \mathbf{n} = 0$ on the boundaries. The only solution compatible with the last statement is Φ equal to a constant.

3-3 LAPLACE'S EQUATION IN ONE INDEPENDENT VARIABLE

If φ is a function of one variable only, Laplace's equation reduces to an ordinary differential equation. Consider the case where φ is $\varphi(x)$, a function of the single rectangular coordinate x . Then,

$$\frac{d^2\varphi}{dx^2} = 0 \quad \text{and} \quad \varphi(x) = ax + b \quad (3-11)$$

is the general solution, where a and b are constants chosen to fit the boundary conditions. This result was found in the preceding chapter for the potential between two charged conducting plates oriented normal to the x -axis.

The situation is no more complicated in other coordinate systems where φ is a function of a single variable. In spherical coordinates where φ equals $\varphi(r)$, Laplace's equation and its general solution become

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d\varphi}{dr} \right) = 0, \quad \varphi(r) = -\frac{a}{r} + b \quad (3-12)$$

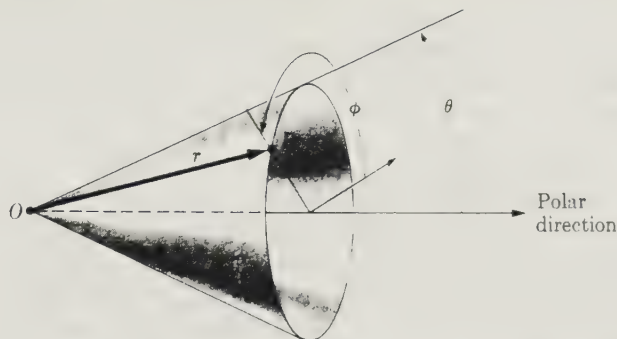
The general solution to Laplace's equation in cylindrical coordinates for a function that is independent of θ and z —that is, for $\varphi(r)$ —is left as an exercise in the problem section.

3-4 SOLUTIONS TO LAPLACE'S EQUATION IN SPHERICAL COORDINATES: ZONAL HARMONICS

We next turn our attention to solutions of Laplace's equation where φ is a function of more than one variable. Many of the problems of interest to us deal with conductors in the shape of spheres or cylinders, and thus solutions

FIGURE 3-1

Location of the point P in terms of the spherical coordinates r , θ , ϕ



of Laplace's equation in either spherical or cylindrical coordinates are needed. We first take up the spherical problem, but we shall find it expedient to *limit the discussion* to cases in which φ is independent of the azimuthal angle ϕ . This limitation restricts the class of problems that we shall be able to solve. Nevertheless, many interesting physical problems fall into this restricted category, and more complicated problems are really beyond the scope of this book.

For the spherical case, φ is $\varphi(r, \theta)$, where r is the radius vector from a fixed origin O and θ is the polar angle (Fig. 3-1). Using Eq. (3-7), Laplace's equation becomes in this case

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \varphi}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \varphi}{\partial \theta} \right) = 0 \quad (3-13)$$

This partial differential equation will be solved by a technique known as *separation of variables*. A solution of the form $\varphi(r, \theta) = Z(r)P(\theta)$ is substituted into Eq. (3-13), yielding

$$\frac{1}{r^2} P(\theta) \frac{d}{dr} \left(r^2 \frac{dZ}{dr} \right) + \frac{Z(r)}{r^2 \sin \theta} \frac{d}{d\theta} \left(\sin \theta \frac{dP}{d\theta} \right) = 0 \quad (3-14)$$

Note that the partial derivatives have been replaced by total derivatives, since Z and P are each functions of one variable only. Dividing through by $\varphi(r, \theta)$ and multiplying through by r^2 , we transform Eq. (3-14) into

$$\frac{1}{Z} \frac{d}{dr} \left(r^2 \frac{dZ}{dr} \right) = - \frac{1}{P \sin \theta} \frac{d}{d\theta} \left(\sin \theta \frac{dP}{d\theta} \right) \quad (3-15)$$

The left side of this equation is a function of r only and the right side is a function of θ . The only way in which a function of r can equal a function of θ for all values of r and θ is for both functions to be constant. Hence, let each side of Eq. (3-15) equal k , where k is the "separation constant."

Not all values of k necessarily yield solutions that are acceptable on physical grounds. Consider the θ equation, known as *Legendre's equation*,

TABLE 3-1Legendre Polynomials
for $n = 0, 1, 2,$ and 3

n	$P_n(\theta)$
0	1
1	$\cos \theta$
2	$\frac{1}{2}(3 \cos^2 \theta - 1)$
3	$\frac{1}{2}(5 \cos^3 \theta - 3 \cos \theta)$

first:

$$\frac{1}{\sin \theta} \frac{d}{d\theta} \left(\sin \theta \frac{dP}{d\theta} \right) + kP = 0 \quad (3-16)$$

The only physically acceptable solutions that are defined over the full range of θ (from 0 to π) correspond to $k = n(n + 1)$, where n is a positive integer. The solution for a particular n will be denoted by $P_n(\theta)$. Solutions of Eq. (3-16) for other values of k are ill-behaved in the vicinity of $\theta = 0$ or $\theta = \pi$ radians, becoming infinite or even undefined at these values of θ .* These solutions cannot be made to fit physical boundary conditions and hence must be discarded.†

The acceptable solutions, the $P_n(\theta)$, are polynomials in $\cos \theta$ and are usually referred to as *Legendre polynomials*. The first four Legendre functions are given in Table 3-1. It is evident from Eq. (3-16) that the P_n may be multiplied by any arbitrary constant.

We now return to the radial equation

$$\frac{d}{dr} \left(r^2 \frac{dZ}{dr} \right) = n(n + 1)Z \quad (3-17)$$

where we have used the explicit form of k , which gave acceptable θ solutions. Inspection of Eq. (3-17) shows that two independent solutions are

$$Z_n = r^n \quad \text{and} \quad Z_n = r^{-(n+1)}$$

Solutions of Laplace's equation are obtained as the product $\varphi_n(r, \theta) = Z_n(r)P_n(\theta)$, where particular care must be exercised to have Z and P

* The discussion here has been all too brief. The interested reader is referred to more mathematical texts for a detailed treatment of Legendre's equation. See, for example, the books listed at the end of this chapter. Legendre's equation is usually written in a different form by substituting $x = \cos \theta$, and its solutions are then denoted by $P_n(x)$ or $P_n(\cos \theta)$.

† This statement requires some qualification. In some electrostatic problems, the regions around $\theta = 0$ and $\theta = \pi$ may be naturally excluded, for example, by conducting conical surfaces. Under these conditions, solutions of Eq. (3-16) with other values of k could be used. Problems of this type will not be considered here.

correspond to the same value of n , which is a mandatory requirement since both sides of Eq. (3-15) are equal to the same constant, namely, $n(n + 1)$.

We have now solved Laplace's equation in spherical coordinates and have obtained the solutions that are known as *zonal harmonics*:

$$\varphi_n = r^n P_n(\theta) \quad \text{or} \quad \varphi_n = r^{-(n+1)} P_n(\theta) \quad (3-18)$$

where $P_n(\theta)$ is one of the polynomials listed in Table 3-1, and n is a positive integer or zero. The zonal harmonics form a set of functions appropriate for solution of Laplace's equation with spherical boundaries and azimuthal symmetry. These solutions may be combined according to Theorem I. Several of the zonal harmonics are already well known to us: One of the $n = 0$ solutions, namely, $\varphi = \text{constant}$, is a trivial solution of Laplace's equation and is valid in any coordinate system; the zonal harmonic r^{-1} is the potential of a point charge; and $r^{-2} \cos \theta$ is the potential of a dipole.

3-5

CONDUCTING SPHERE IN A UNIFORM ELECTRIC FIELD

We shall illustrate the usefulness of zonal harmonics for electrostatic problems having spherical symmetry by solving the problem of an uncharged conducting sphere placed in an *initially* uniform electric field \mathbf{E}_0 . The lines of a uniform electric field are parallel, but the presence of the conductor alters the field in such a way that the field lines strike the surface of the conductor, which is an equipotential surface, normally. If we take the direction of the initially uniform electric field as the polar direction (z -direction), and if we make the origin of our coordinate system coincide with the center of the sphere, then from the symmetry of the problem it is clear that the potential will be independent of azimuthal angle ϕ and may be expressed as a sum of zonal harmonics.

The spherical conductor of radius a is an equipotential surface. Let us denote its potential by φ_0 . Our problem is to find a solution to Laplace's equation in the region outside the sphere, which reduces to φ_0 on the sphere itself and which has the correct limiting form at large distances away. The solution may be formally written as

$$\begin{aligned} \varphi(r, \theta) = & A_1 + C_1 r^{-1} + A_2 r \cos \theta + C_2 r^{-2} \cos \theta \\ & + \frac{1}{2} A_3 r^2 (3 \cos^2 \theta - 1) + \frac{1}{2} C_3 r^{-3} (3 \cos^2 \theta - 1) + \dots \end{aligned} \quad (3-19)$$

where the A 's and C 's are arbitrary constants. At large r , the electric field will be only slightly distorted from its initial form, and the potential will be

that appropriate to a uniform electric field:

$$\begin{aligned} [\mathbf{E}(r, \theta)]_{r \rightarrow \infty} &= \mathbf{E}_0 = E_0 \mathbf{k} \\ [\varphi(r, \theta)]_{r \rightarrow \infty} &= -E_0 z + \text{constant} \\ &= -E_0 r \cos \theta + \text{constant} \end{aligned} \quad (3-20)$$

Hence, in order to make Eqs. (3-19) and (3-20) agree at large r , $A_2 = -E_0$. Furthermore, all the A 's from A_3 up must be set equal to zero.

The term $C_1 r^{-1}$ produces a radial field that, as we might expect, is compatible only with a spherical conductor bearing net total charge. Since our problem deals with an uncharged conductor, the constant C_1 must be set equal to zero. At the surface of the sphere, $\varphi = \varphi_0$, and the potential must become independent of angle θ . The two terms involving $\cos \theta$ may be made to cancel each other, but the terms with higher inverse powers of r cannot be canceled one against the other because they contain *different* Legendre functions. The only possibility is to set all the C_i 's with $i \geq 3$ equal to zero. Equation (3-19) now becomes

$$\begin{aligned} \varphi(r, \theta) &= A_1 - E_0 r \cos \theta + C_2 r^{-2} \cos \theta, \quad \text{for } r \geq a \\ \varphi(a, \theta) &= \varphi_0 \end{aligned} \quad (3-21)$$

Since the two expressions must be equal at $r = a$, $A_1 = \varphi_0$ and $C_2 = E_0 a^3$.

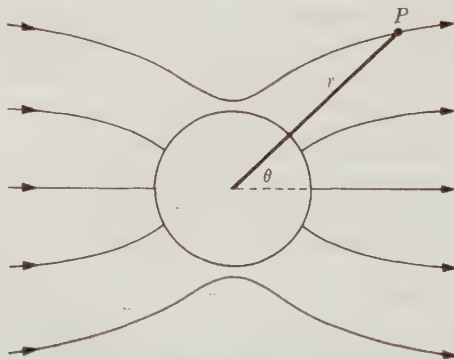
From the final expression for the potential, we may calculate not only the electric field at all points in space (Fig. 3-2) but also the surface density of charge on the conducting sphere:

$$\left. \begin{aligned} E_r &= -\frac{\partial \varphi}{\partial r} = E_0 \left(1 + 2 \frac{a^3}{r^3} \right) \cos \theta \\ E_\theta &= -\frac{1}{r} \frac{\partial \varphi}{\partial \theta} = -E_0 \left(1 - \frac{a^3}{r^3} \right) \sin \theta \end{aligned} \right\} \quad \text{for } r \geq a \quad (3-22)$$

$$\sigma(\theta) = \epsilon_0 E_r|_{r=a} = 3\epsilon_0 E_0 \cos \theta \quad (3-23)$$

FIGURE 3-2

Lines of electric flux for the case of a conducting sphere placed in a uniform electric field



The total charge on the sphere,

$$Q = a^2 \int_0^\pi \sigma(\theta) 2\pi \sin \theta d\theta$$

is obviously zero, which agrees with our initial assumption.

SOLUTIONS TO LAPLACE'S EQUATION IN CYLINDRICAL COORDINATES: CYLINDRICAL HARMONICS

Laplace's equation in cylindrical coordinates may also be solved by the method of separation of variables. Here again it will be expedient to work out solutions for only a restricted class of problems; namely, those in which the potential is independent of the coordinate z . These solutions are appropriate for certain problems involving a long straight cylindrical conductor or wire, but are *not* appropriate for those dealing with a short cylindrical segment.

If the potential is independent of z , Laplace's equation in cylindrical coordinates becomes

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial \varphi}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 \varphi}{\partial \theta^2} = 0 \quad (3-24)$$

Substitution of $\varphi = Y(r)S(\theta)$ reduces the equation to

$$\frac{r}{Y} \frac{d}{dr} \left(r \frac{dY}{dr} \right) = - \frac{1}{S} \frac{d^2 S}{d\theta^2} = k \quad (3-25)$$

where k again plays the role of a separation constant. The θ -equation is particularly simple; it has the solutions $\cos k^{1/2}\theta$ and $\sin k^{1/2}\theta$. But if these solutions are to make sense physically, each must be a single-valued function of θ . Thus,

$$\cos k^{1/2}(\theta + 2\pi) = \cos k^{1/2}\theta$$

$$\sin k^{1/2}(\theta + 2\pi) = \sin k^{1/2}\theta$$

Or, to put it differently, after θ has gone through its full range from 0 to 2π , the function must join *smoothly* to its value at $\theta = 0$. Such is the case only if $k = n^2$, n being an integer. We may further require n to be positive (or zero) without losing any of these solutions.

Returning now to the r -equation, we are able to verify easily that $Y(r)$ is r^n or r^{-n} ; unless $n = 0$ when $Y(r) = \ln r$ or $Y(r) = \text{constant}$. Hence, the required solutions to Laplace's equation, the so-called *cylindrical harmonics*, are

$$\begin{array}{ll} 1 & \ln r \\ r^n \cos n\theta & r^{-n} \cos n\theta \\ r^n \sin n\theta & r^{-n} \sin n\theta \end{array}$$

These functions form a complete set for the variables r, θ in cylindrical coordinates, and the potential $\varphi(r, \theta)$ may be developed as a superposition of cylindrical harmonics in accordance with Theorem I.

*3-7 LAPLACE'S EQUATION IN RECTANGULAR COORDINATES

In rectangular coordinates, the variables may be separated by making the substitution

$$\varphi(x, y, z) = f_1(x)f_2(y)f_3(z)$$

whereby Laplace's equation reduces to

$$\frac{1}{f_1(x)} \frac{d^2 f_1}{dx^2} + \frac{1}{f_2(y)} \frac{d^2 f_2}{dy^2} = - \frac{1}{f_3(z)} \frac{d^2 f_3}{dz^2} \quad (3-26a)$$

The left side of this equation is a function of x and y , and the right side is a function of z only; hence both sides must be equal to the same constant, k . This constant is the first separation constant. The two equations obtained from Eq. (3-26a) are

$$\frac{d^2 f_3}{dz^2} + k f_3 = 0 \quad (3-26b)$$

$$\frac{1}{f_2} \frac{d^2 f_2}{dy^2} = k - \frac{1}{f_1} \frac{d^2 f_1}{dx^2}$$

The latter equation has been written such that the variables x and y are separated. Each side of this equation is now set equal to $-m$ (the second separation constant). Thus,

$$\frac{d^2 f_2}{dy^2} + m f_2 = 0 \quad (3-26c)$$

$$\frac{d^2 f_1}{dx^2} - (k + m) f_1 = 0 \quad (3-26d)$$

Equations (3-26b), (3-26c), and (3-26d) are easily solved. One of the typical solutions for $\varphi(x, y, z)$ is

$$\varphi(x, y, z) = A e^{-(k+m)^{1/2} x} \cos m^{1/2} y \cos k^{1/2} z \quad (3-27)$$

The other seven independent solutions for a pair of separation constants (k, m) are obtained by making one or more of the following substitutions: $+(k+m)^{1/2} x$ for $-(k+m)^{1/2} x$, $\sin m^{1/2} y$ for $\cos m^{1/2} y$, and $\sin k^{1/2} z$ for $\cos k^{1/2} z$.

* Starred sections may be omitted without loss of continuity.

Thus far there are no restrictions on k or m , but boundary conditions on the problem usually restrict k (or m) to a discrete set of positive or negative values. It is worthwhile making the point that it is the boundary conditions that really pick out the *pertinent solutions* to a partial differential equation. For fixed x and y the following function is just the *Fourier series expansion* for an *arbitrary* even function of z :

$$\varphi(x, y, z) = \sum_p \sum_q A_{pq} e^{-(p^2+q^2)^{1/2}x} \cos py \cos qz$$

The individual solutions in Eq. (3-27) do not represent potentials that correspond to simple physical situations. The case where both separation constants are zero does, however, correspond to an interesting physical situation; hence, we turn our attention to this case. From Eq. (3-26d), it is evident that $f_1(x) = a_1x$, or $f_1(x) = \text{constant}$, is a solution; from Eq. (3-26c), we obtain $f_2(y)$, and so on. Thus,

$$\begin{aligned} \varphi(x, y, z) = & A_1xyz + A_2xy + A_3yz + A_4xz \\ & + A_5x + A_6y + A_7z + A_8 \end{aligned} \quad (3-28a)$$

where the A 's are arbitrary constants. This solution may be applied to the case where three conducting planes intersect at right angles. If these planes are the coordinate planes xy , yz , and zx , and all are at the same potential, then

$$\varphi(x, y, z) = A_1xyz + A_8 \quad (3-28b)$$

It is left as an exercise for the reader to determine the surface charge density on the coordinate planes that is compatible with Eq. (3-28b).

3-8

LAPLACE'S EQUATION IN TWO DIMENSIONS: GENERAL SOLUTION

If the potential is a function of only two rectangular coordinates, Laplace's equation is written

$$\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} = 0 \quad (3-29a)$$

It is possible to obtain the general solution to this equation by means of a transformation to a new set of independent variables. Nevertheless, it should be emphasized that such a transformation leads to a simplification of the original equation only in the two-dimensional case. Let

$$\xi = x + iy, \quad \eta = x - iy$$

where $i = \sqrt{-1}$ is the unit imaginary number. In terms of these relationships,

$$\frac{\partial^2}{\partial x^2} = \frac{\partial^2}{\partial \xi^2} + 2 \frac{\partial^2}{\partial \xi \partial \eta} + \frac{\partial^2}{\partial \eta^2}$$

$$\frac{\partial^2}{\partial y^2} = -\frac{\partial^2}{\partial \xi^2} + 2 \frac{\partial^2}{\partial \xi \partial \eta} - \frac{\partial^2}{\partial \eta^2}$$

and

$$\nabla^2 \varphi = 4 \frac{\partial^2 \varphi}{\partial \xi \partial \eta} = 0 \quad (3-29b)$$

It is evident that a very general solution to (3-29b) is

$$\varphi = F_1(\xi) + F_2(\eta) = F_1(x + iy) + F_2(x - iy) \quad (3-30)$$

where F_1 and F_2 are *arbitrary* functions that are suitably* continuous and differentiable. The functions F_1 and F_2 are complex quantities in general, but two real functions may be constructed in the following way. First let $F_2(x - iy) = F_1(x - iy)$ —that is, let the two functions F_1 and F_2 have the same dependence on their arguments. Then,

$$\varphi_1 = F_1(x + iy) + F_1(x - iy) = 2 \operatorname{Re} [F_1(x + iy)]$$

where Re stands for *real part of*. Furthermore, the second real potential function is

$$\varphi_2 = -i[F_1(x + iy) - F_1(x - iy)] = 2 \operatorname{Im} [F_1(x + iy)]$$

where Im stands for *imaginary part of*. Thus, the real and imaginary parts of any complex function $F(x + iy)$ are both solutions of Laplace's equation.

The solutions found in this way are not restricted to any particular coordinate system. For example, the cylindrical harmonics of Section 3-7 are obtained from the complex functions† $(x + iy)^n = r^n e^{in\theta}$, and $\ln(x + iy) = \ln r + i\theta$. On the other hand, when it comes to solving a particular two-dimensional problem, there is no standard procedure for finding the appropriate complex function. This method generates so many solutions that it is not possible to enumerate them all and cast out those that do not agree with boundary conditions on the problem. In simple cases, the required functions may be found by trial and error; in other cases, the method of *conformal mapping* (which is beyond the scope of this book) may be useful.

* The function F must be "analytic" in the (x, y) domain of interest. See Appendix VII, Eq. (VII-3).

† The cylindrical and rectangular coordinates are related in the usual way: $x = r \cos \theta$, $y = r \sin \theta$.

ELECTROSTATIC IMAGES

For a given set of boundary conditions the solution to Laplace's equation is unique, so that if one obtains a solution $\varphi(x, y, z)$ by any means whatever, and if this φ satisfies all boundary conditions, then a complete solution to the problem has been effected. The *method of images* is a procedure for accomplishing this result without specifically solving a differential equation. It is not universally applicable to all types of electrostatic problems, but enough interesting problems fall into this category to make it worthwhile discussing the method here.

Suppose the potential may be written in the following way:

$$\varphi(\mathbf{r}) = \varphi_1(\mathbf{r}) + \frac{1}{4\pi\epsilon_0} \int_S \frac{\sigma(\mathbf{r}') da'}{|\mathbf{r} - \mathbf{r}'|} \quad (3-31)$$

where φ_1 is either a specified function or easily calculable, and the integral represents the contribution to the potential from surface charge on all conductors appearing in the problem. The function σ is not known. It may happen, and this is the essence of the image-charge method, that the last term in Eq. (3-31) can be replaced by a potential φ_2 that is due to a *specified* charge distribution, usually point or line charges. This replacement is possible so long as the surfaces of all conductors coincide with equipotential surfaces of the combined $\varphi_1 + \varphi_2$. The specified charges producing φ_2 are called *image charges*. They do not really exist, of course. Their apparent location is *inside* the various conductors, and the potential $\varphi = \varphi_1 + \varphi_2$ is a valid solution to the problem only in the exterior region.

EXAMPLE 3-1 Point Charge Near a Conducting Plane

We shall now take up some specific examples that demonstrate the advantages of the image-charge method. It should be pointed out that the solution of these example problems would be very difficult if we had to use other analytic methods. Consider first the problem of a point charge q placed near a conducting plane of infinite extent. To formulate the problem mathematically, let the conducting plane coincide with the yz -plane, and let the point charge lie on the x -axis at $x = d$ (see Fig. 3-3a). The potential fits the prescription Eq. (3-31), with

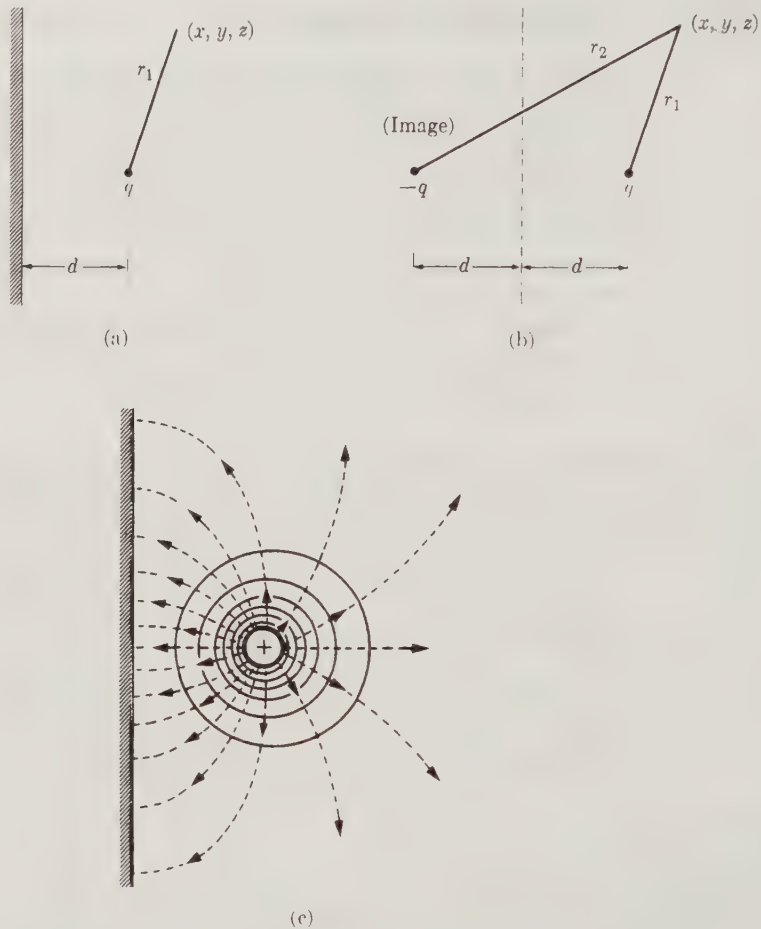
$$\varphi_1(x, y, z) = \frac{1}{4\pi\epsilon_0 r_1} = \frac{q}{4\pi\epsilon_0 \sqrt{(x-d)^2 + y^2 + z^2}} \quad (3-32)$$

Solution: Consider a different problem, that of two point charges (q and $-q$) a distance $2d$ apart, as shown in Fig. 3-3(b). The potential of these two charges,

$$\varphi(x, y, z) = \frac{q}{4\pi\epsilon_0 r_1} - \frac{q}{4\pi\epsilon_0 r_2} \quad (3-33)$$

FIGURE 3-3

Problem of a point charge and conducting plane solved by means of the image-charge method: (a) original problem; (b) location of image charge; (c) lines of force (dashed) and equipotential surfaces (solid)



not only satisfies Laplace's equation at all points exterior to the charges, but also reduces to a constant (namely, zero) on the plane that perpendicularly bisects the segment joining the two charges. Thus, Eq. (3-33) satisfies the boundary conditions of the original problem. Because solutions to Laplace's equation are unique, Eq. (3-33) is the correct potential in the entire half-space to the right of the conducting plane. The charge $-q$ gives rise to the potential:

$$\varphi_2(x, y, z) = -\frac{q}{4\pi\epsilon_0 r_2} = -\frac{q}{4\pi\epsilon_0 \sqrt{(x+d)^2 + y^2 + z^2}} \quad (3-34)$$

This charge is called the *image* of the point charge q . Naturally, the image does not really exist, and Eq. (3-32) does *not* give correctly the potential to the left of the conducting plane in Fig. 3-3(a).

The electric field \mathbf{E} in the region to the right of the conducting plane may be obtained as the negative gradient of Eq. (3-33). Since the surface of the conducting plane represents an interface joining two solutions of Laplace's equation, namely, $\varphi = 0$ and Eq. (3-33), the discontinuity in the electric field is accommodated by a surface charge density σ on the plane:

$$\sigma(y, z) = \epsilon_0 E_x|_{x=0} = -\frac{qd}{2\pi(d^2 + y^2 + z^2)^{3/2}} \quad (3-35)$$

The lines of force and equipotential surfaces appropriate to the original problem are shown in Fig. 3-3(c). These are the same lines of force and equipotential surfaces appropriate to the two point-charge problem in Fig. 3-3(b) except that in the latter case the flux lines would continue into the left half-plane. It is evident from the figure that *all* of the electric flux lines that would normally converge on the image charge are intercepted by the plane in Fig. 3-3(c). Hence, the total charge on the plane is equal to that of the image charge $-q$. This same result may be obtained mathematically by integrating Eq. (3-35) over the entire surface (see Problem 3-14).

It is evident that the point charge q exerts an attractive force on the plane, because the induced surface charge is of the opposite sign. By Newton's law of action and reaction, this force is equal in magnitude to the force exerted on q by the plane. Since the point charge experiences no force due to its own field,

$$\mathbf{F} = -q \nabla \varphi_2 \quad (3-36)$$

which is just the force exerted on it by the image charge.

EXAMPLE 3-2
Point Charge Near a
Right-Angle Corner

Another problem that may be solved simply in terms of images is that of determining the electric field of a point charge q in the vicinity of a right-angle intersection of two conducting planes (Fig. 3-4a).

Solution: The positions of the necessary image charges are shown in Fig. 3-4(b). It is readily seen that the two planes shown dotted in the

FIGURE 3-4
 Point charge in a right-angle corner

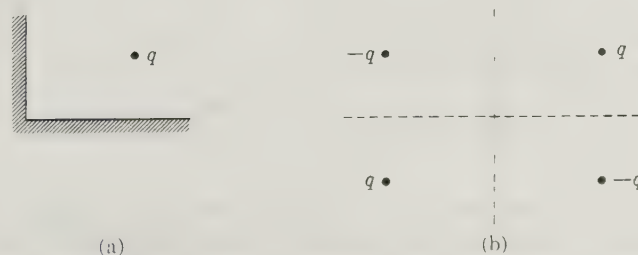


FIGURE 3-5

Point charge q in the vicinity of a conducting sphere. q' is the image charge.

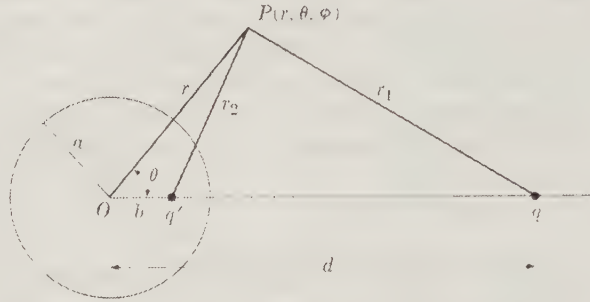


figure are surfaces of zero potential due to the combined potentials of q and the three image charges.

The principal difficulty in solving a problem by the image technique is that of finding a group of image charges that, together with the originally specified charges, produce equipotential surfaces at the conductors. The problem is straightforward only in cases where the geometry is simple. Such is the case, however, in the next example.

EXAMPLE 3-3
Point Charge and
Conducting Sphere

Consider the problem posed by a point charge q in the vicinity of a conducting sphere; it requires a single image charge to make the sphere a surface of *zero potential*. An additional image charge is needed to change the potential of the sphere to some other constant value.

Solution: We first determine the magnitude and location of the image q' which together with the point charge q produces zero potential at all points on the sphere. The geometry of the situation is shown in Fig. 3-5. The point charge q is a distance d from the center of the sphere, and the radius of the sphere is a . It is apparent from the symmetry of the problem that the image charge q' will lie on the line passing through q and the center of the sphere.

The desired results are most easily obtained by using spherical coordinates, with the origin of coordinates at the center of the sphere. Let the polar axis be taken as the line joining q to the origin. The distance b and the magnitude of q' are to be determined in terms of the specified quantities: q , d , a . The potential at an arbitrary point P due to q and q' is given by

$$\begin{aligned}\varphi(r, \theta, \phi) &= \frac{q}{4\pi\epsilon_0 r_1} + \frac{q'}{4\pi\epsilon_0 r_2} \\ &= \frac{1}{4\pi\epsilon_0} \left[\frac{q}{\sqrt{r^2 + d^2 - 2rd \cos \theta}} + \frac{q'}{\sqrt{r^2 + b^2 - 2rb \cos \theta}} \right]\end{aligned}\quad (3-37)$$

On the surface of the sphere, $r = a$ and $\varphi(a, \theta, \phi) = 0$ for all θ and ϕ . But from expression (3-37), $\varphi(a, \theta, \phi)$ can equal zero for all θ only if the two square roots are proportional to each other, which requires that $b = a^2/d$, for then

$$\sqrt{a^2 + b^2 - 2ab \cos \theta} = \frac{a}{d} \sqrt{d^2 + a^2 - 2ad \cos \theta}$$

Hence,

$$b = \frac{a^2}{d} \quad (3-38)$$

and furthermore,

$$q' = -\frac{a}{d}q \quad (3-39)$$

These equations serve to specify the location and magnitude of the first image charge.

A second image charge q'' may be placed at the center of the sphere without destroying the equipotential nature of the spherical surface. The magnitude of q'' is arbitrary; it may be adjusted to fit the boundary conditions on the problem. Thus, a complete solution to the point charge-conducting sphere problem has been effected; the potential at all points exterior to the sphere is

$$\varphi(r, \theta, \phi) = \frac{1}{4\pi\epsilon_0} \left[\frac{q}{r_1} + \frac{q'}{r_2} + \frac{q''}{r} \right] \quad (3-40)$$

The potential of the spherical conductor itself is

$$\varphi(a, \theta, \phi) = \frac{q''}{4\pi\epsilon_0 a} \quad (3-41)$$

and the surface density of charge on the sphere is

$$\sigma(\theta, \phi) = -\epsilon_0 \left. \frac{\partial \varphi}{\partial r} \right|_{r=a} \quad (3-42)$$

All the lines of force that would normally converge on the image charges are intercepted by the sphere. Hence, the total charge on the sphere, Q , is equal to the sum of the image charges:

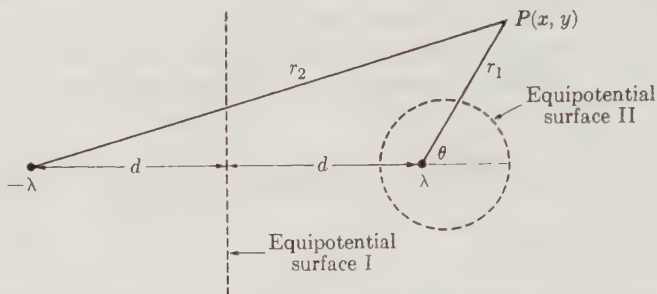
$$Q = q' + q'' \quad (3-43)$$

This result may be verified by direct integration of Eq. (3-42).

Special cases of interest are the *grounded* sphere: $\varphi(a) = 0$, $q'' = 0$, and the *uncharged spherical conductor*: $q'' = -q'$.

FIGURE 3-6

Two infinitely long, parallel line charges (of charge λ and $-\lambda$ per unit length), shown cutting the plane of the paper



3-10 LINE CHARGES AND LINE IMAGES

Thus far, our image technique has been limited to problems involving point charges and hence point images. In this section we shall take up several problems that may be solved by means of line image charges. Consider two infinitely long, parallel, line charges, with charges λ and $-\lambda$ per unit length, respectively, as shown in Fig. 3-6. The potential at any point is given by

$$\varphi = -\frac{\lambda}{2\pi\epsilon_0} [\ln r_1 - \ln r_2] = -\frac{\lambda}{2\pi\epsilon_0} \ln \frac{r_1}{r_2} \quad (3-44)$$

where r_1 and r_2 are the perpendicular distances from the point to the two line charges. The equipotentials are obtained by setting Eq. (3-44) equal to a constant, a procedure which is equivalent to requiring that

$$\frac{r_1}{r_2} = M \quad (3-45)$$

where M is constant. Hence, the equipotentials may be specified by Eq. (3-45).

The equipotential corresponding to $M = 1$ is the plane located halfway between the two line charges, shown as equipotential surface I in Fig. 3-6. The potential of the plane is zero. Hence, the problem of a long line charge oriented parallel to a conducting plane has been effectively solved. The potential in the half-space is given correctly by Eq. (3-44). Let us assume that the line charge shown on the right side of the figure is the specified charge, which is at a distance d from the conducting plane. Then the line charge on the left side of the figure plays the role of an image. Again, the total charge on the plane is equal to that of the image charge.

Let us next consider equipotential surfaces corresponding to other values of M . The general form of the surface may be found by expressing r_1 and r_2 in rectangular coordinates. For convenience, we choose the origin of the coordinate system on the positive line charge and make this charge coincide with the z -axis. We let the second line charge be located at

$x = -2d, y = 0$. Now

$$r_1^2 = x^2 + y^2$$

and

$$r_2^2 = (x + 2d)^2 + y^2$$

so that Eq. (3-45) becomes, after a little algebraic manipulation,

$$x^2 + y^2 - \frac{4M^2xd}{1 - M^2} = \frac{4M^2d^2}{1 - M^2} \quad (3-46)$$

which is the equation of a circular cylinder extending parallel to the z -axis. If M is less than one, the cylinder surrounds the positive line charge, as does equipotential surface II of the figure. The axis of the cylinder goes through the point

$$x = \frac{2M^2d}{1 - M^2}, \quad y = 0 \quad (3-47)$$

and the radius of the cylinder is

$$R_c = \frac{2Md}{1 - M^2} \quad (3-48)$$

EXAMPLE 3-4

Charged Conducting Cylinder Oriented Parallel to a Conducting Plane

We are now in a position to solve a number of interesting problems involving cylindrical conductors, but only one of this type will be discussed. Consider the problem of a long cylindrical conductor that is in the vicinity of a conducting plane and oriented parallel to it. The cylinder bears the charge λ per unit length. Figure 3-6 may serve to illustrate the problem, the two conductors coinciding with the dotted surfaces.

Solution: Both line charges are images in this case, and the potential in the region surrounding the cylinder and to the right of the plane is given by Eq. (3-44). It is evident that the charge induced on the plane is equal to $-\lambda$ per unit distance in the z -direction.

SYSTEM OF CONDUCTORS AND COEFFICIENTS OF POTENTIAL

In the preceding sections, several important methods for obtaining solutions to Laplace's equation have been discussed. Although general in scope, these methods are limited by practical considerations to problems in which the conductors have rather simple shapes. When their shapes are complicated, a complete mathematical solution to the problem in analytic form is out of the

question. Nevertheless, certain conclusions can be drawn about the system just because the potential satisfies Laplace's equation. In fact, we shall prove here that a linear relationship exists between the potential of one of the conductors and the charges on the various conductors in the system. The coefficients in this relationship, the so-called *coefficients of potential*, are functions only of the geometry (specifically, they do not depend on either charges or potentials) and, although not always calculable analytically, they may be determined numerically or directly from experiment.

Suppose there are N conductors in fixed geometry. Let all of the conductors be uncharged except conductor j , which bears the charge Q_{j0} . The appropriate solution to Laplace's equation in the space exterior to the conductors will be given the symbol $\varphi^{(j)}(x, y, z)$, and the potential of each of the conductors will be indicated by $\varphi_1^{(j)}, \varphi_2^{(j)}, \dots, \varphi_j^{(j)}, \dots, \varphi_N^{(j)}$. Now let us change the charge of the j th conductor to λQ_{j0} . The function $\lambda\varphi^{(j)}(x, y, z)$ satisfies Laplace's equation, since λ is a constant; that the new boundary conditions are satisfied by this function may be seen from the following argument. The potential at all points in space is multiplied by λ . Thus, all derivatives (and in particular the gradient) of the potential are multiplied by λ . Because $\sigma = \epsilon_0 E_n$, it follows that all charge densities are multiplied by λ . Thus, the charge of the j th conductor is λQ_{j0} , and all other conductors remain uncharged.

A solution of Laplace's equation that fits a particular set of boundary conditions is unique; therefore we have found *the correct solution*, $\lambda\varphi^{(j)}(x, y, z)$, to our modified problem. The interesting conclusion we draw from this discussion is that the potential of each conductor is proportional to the charge Q_j of conductor j . That is,

$$\varphi_i^{(j)} = p_{ij}Q_j, \quad (i = 1, 2, \dots, N) \quad (3-49)$$

where p_{ij} is a constant that depends only on the geometry.

The same argument may be applied to the case where conductor k is charged: $Q_k = \nu Q_{k0}$, with all other conductors being uncharged. Here the appropriate solution to Laplace's equation is $\nu\varphi^{(k)}(x, y, z)$, where $\varphi^{(k)}$ is the solution for $\nu = 1$. It is apparent, then, that

$$\lambda\varphi^{(j)}(x, y, z) + \nu\varphi^{(k)}(x, y, z) \quad (3-50)$$

is a solution appropriate to the case where both conductors are charged. Again we appeal to the uniqueness of a solution for a given set of boundary conditions. Thus, Eq. (3-50) is *the solution* for this case, and the potential of each conductor may be written as

$$\varphi_i = p_{ij}Q_j + p_{ik}Q_k, \quad (i = 1, 2, \dots, N) \quad (3-51)$$

This result may be generalized immediately to the case where all N conductors are charged:

$$\varphi_i = \sum_{j=1}^N p_{ij}Q_j \quad (3-52)$$

This equation is the linear relationship between potential and charge that we have been seeking. The coefficients p_{ij} are called the *coefficients of potential*. In Chapter 6, it will be shown that the array of these coefficients is symmetrical—that is, $p_{ij} = p_{ji}$.

NUMERICAL SOLUTION OF ELECTROSTATIC PROBLEMS

Our previous discussions of solutions to Laplace's equation have always used coordinate systems that were well adapted to the boundary conditions—that is, rectangular coordinates for rectangular boundary conditions, cylindrical coordinates for cylindrical boundaries, and so on. Not all boundaries match convenient coordinate systems. Furthermore, even if they do, it may be very difficult to satisfy the boundary conditions analytically. These are some of the reasons why numerical approximations to solutions of Laplace's equation are important. While the subject has a long history, it has been the recent, ready availability of large-scale digital computers that has brought renewed interest in numerical techniques.

Of the many approaches that have been used, two general types dominate: finite difference methods and finite element methods. *Finite difference methods* are characterized by a regular mesh superimposed on the region in which a solution is to be found. At each point on the mesh Laplace's equation is approximated by approximating the second derivatives. The resulting algebraic equations are then solved, usually by an iterative technique, for the values of the potential at the mesh points. The *finite element method*, on the other hand, divides the region (or in some methods the boundary of the region) in which the solution is to be found into subelements adapted to the boundaries and covering the region without overlap. In each of these finite elements, a simple analytical approximation to the solution is made. The solutions are required to satisfy the boundary conditions and to be continuous across boundaries between elements. Some adjustable parameters are left in the approximation, and these parameters are determined by minimizing the energy of the system. It is a little difficult to follow this abstract description, so we shall discuss each method in terms of the solution of Laplace's equation in one dimension. The equation and boundary conditions are

$$d^2\varphi/dx^2 = 0, \quad \varphi(0) = 0, \quad \varphi(1) = 1 \quad (3-53)$$

This equation has been considered in Section 3-3. Its exact solution is easily found to be $\varphi(x) = x$. We shall first use the finite difference method to find an approximate solution.

The first step is to divide the interval from $x = 0$ to $x = 1$ into N equal segments of length $1/N$. At the i th dividing point, Laplace's equation is

approximated by

$$[(\varphi_{i+1} - \varphi_i) - (\varphi_i - \varphi_{i-1})]/(1/N)^2 = 0 \quad (3-54)$$

At the end points $\varphi(0) = 0$ and $\varphi(1) = 1$. The $N - 1$ equations of Eq. (3-54) can be solved exactly (the solution is $\varphi_i = i/N$), but it is more usual to use an iterative technique to find a series of approximate solutions that converge to the exact solution. One basis for the iteration is the rearrangement of Eq. (3-54) into the form

$$\varphi_i = \frac{\varphi_{i+1} + \varphi_{i-1}}{2} \quad (3-55)$$

The solution is started by assuming a set of values for the φ_i 's. Frequently the choice $\varphi_i = 0$ is made for all interior points and of course the specified boundary values, $\varphi_0 = 0$ and $\varphi_1 = 1$, on the boundary. A new set of φ_i 's is calculated by using Eq. (3-55), starting either at $i = 1$ or $i = N - 1$ and always using the latest available value of the φ_i 's on the righthand side. We define the "difference" between two iterations as the maximum of the absolute value of the difference between the values of the potential calculated by the two iterations for all points on the mesh. The iteration is continued until the difference between the last two iterations is reduced to a specified value. This process has two shortcomings: It converges quite slowly, and the difference between the last two iterations is a questionable measure of the error—that is, the difference between the last iteration and the exact solution.

The rate of convergence can be improved by "overrelaxation." To overrelax, Eq. (3-55) is rewritten as

$$\varphi_i = w \frac{\varphi_{i+1} + \varphi_{i-1}}{2} + (1 - w)\varphi_i \quad (3-56)$$

where w is an arbitrary nonzero constant. Considered as difference equations, Eq. (3-55) and Eq. (3-56) are equivalent. Consequently their exact solutions are identical. However, when Eq. (3-56) is used as a basis for an iterative method of approximating the solution—that is, when the latest available values for the quantities on the righthand side are used to calculate a new value for the lefthand side—one finds substantial variations in the rate of convergence and the accuracy of the approximation.

Using superscripts to keep track of the iterations, we write Eq. (3-56) for the $(n + 1)$ th iteration for the value of the potential at point i as

$$\varphi_i^{n+1} = w \frac{\varphi_{i+1}^n + \varphi_{i-1}^n}{2} + (1 - w)\varphi_i^n \quad (3-57)$$

assuming that we are starting with $i = 1$. The value of w determines the rate of convergence, and there is usually a distinct optimum. This optimum depends on the equation being solved, the boundary, the size of the mesh,

TABLE 3-2

	$D_m = 0.1$	0.01	0.001	0.0001	0.00001	
Number of Iterations to Satisfy Iteration Criterion	$w = 1.0$					
D_m in Finite Difference Approximation to One-Dimensional Laplace Equation. ($\varphi(0) = 0$, $\varphi(1) = 1.0$; number of segments $N = 10$.)	Iterations	3	18	41	64	87
	Max. error	0.42	0.087	0.0086	0.00086	0.000085
	$w = 1.5$					
	Iterations	2	9	14	19	25
	Max. error	0.20	0.012	0.0015	0.00017	1.3×10^{-5}

and the value of the iteration parameter, D_m . We require that the difference (as defined above) between the last two iterations be less than D_m . The optimum value of w can be found a priori in only a few cases, but in general a value of about 1.5 greatly speeds the convergence. Although there is no guarantee, the value of w that produces the fastest convergence often produces the most accurate solution for a given value of the iteration parameter.

This iteration procedure is very easy to program, and doing so is left to the problem section. Without, however, considering the details of the program, it is instructive to look at some results. For 10 segments, Table 3-2 shows the number of iterations required to satisfy various iteration criteria, $\max \{|\varphi_i^{n+1} - \varphi_i^n|\} < D_m$. If $w = 1$ it takes a large number of iterations to satisfy the criterion even with $D_m = 0.01$ (a relatively large value). Also with $w = 1$ the accuracy (the upper bound of the magnitude of the difference between the exact value and the calculated value) is roughly an order of magnitude worse than D_m . For $w = 1.5$, which is not the optimum value, it takes only about one third as many iterations to satisfy a given iteration criterion, and what is more interesting, the accuracy is an order of magnitude better than the $w = 1$ case for the same value of D_m . If N is larger, the improvement in convergence obtained by choosing the optimum value of w is striking. For example, if $N = 50$ and $D_m = 0.001$, 233 iterations are required with $w = 1$, while only 52 are required for $w = 1.9$. In practical problems, the exact solution is, of course, not known. One must rely on some test such as our iteration criterion to determine when the iteration has been carried far enough to give the required accuracy. It is usually worth spending some time looking at convergence numerically before trying to find a final solution to a new problem. Experience helps and some of the problems at the end of the chapter provide for experimentation for that purpose.

We next consider a finite element approach to the same problem. Again we want to solve the boundary value problem of Eq. (3-53), but this time, by using a simple finite element method. To implement the process, the interval $x = 0$ to $x = 1$ is divided into a number of elements by a finite set of points $0 < x_1 \cdots < 1$. The number of points is arbitrary, as are the

locations of the points. For our problem, we divide the interval into three elements by the points x_1 and x_2 . In each element, the potential φ is approximated by some convenient function. The approximating functions do not need to satisfy Laplace's equation, but those we choose for this problem will. In fact, we choose linear approximations (which we denote by u) to the potential φ :

$$\begin{aligned} u &= b_1 x, & 0 \leq x \leq x_1 \\ u &= a_2 + b_2 x, & x_1 \leq x \leq x_2 \\ u &= a_3 + b_3 x, & x_2 \leq x \leq 1 \end{aligned} \quad (3-58)$$

The first of these equations includes the boundary condition $u(0) = 0$. Requiring that $u(1) = 1$ and matching the values of the potentials at x_1 and x_2 gives

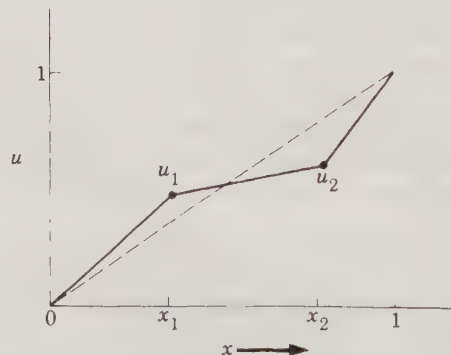
$$\begin{aligned} b_1 &= \frac{u_1}{x_1} \\ a_2 &= u_1 - \frac{u_2 - u_1}{x_2 - x_1} x_1, & b_2 &= \frac{u_2 - u_1}{x_2 - x_1} \\ a_3 &= u_2 - \frac{u_3 - u_2}{x_3 - x_2} x_2, & b_3 &= \frac{u_3 - u_2}{x_3 - x_2} \end{aligned} \quad (3-59)$$

Here $x_3 = 1$, $u_3 = 1$, and u_1 and u_2 are as yet unknown values. This trial potential is shown in Fig. 3-7.

Now let $\varphi(x)$ be the true solution of Laplace's equation and define $e(x) = u(x) - \varphi(x)$; that is, $e(x)$ is the error of the approximate solution. Since $u(x)$ has been chosen to satisfy the same boundary conditions as $\varphi(x)$, $e(0) = e(1) = 0$. We now integrate $(du/dx)^2$ (which equals $[d(\varphi + e)/dx]^2$)

FIGURE 3-7

Trial function for use in finite element method for determining the electrostatic potential



over the interval 0 to 1:

$$F = \int_0^1 \left(\frac{du}{dx}\right)^2 dx = \int_0^1 \left(\frac{d\varphi}{dx}\right)^2 dx + 2 \int_0^1 \frac{d\varphi}{dx} \frac{de}{dx} dx + \int_0^1 \left(\frac{de}{dx}\right)^2 dx \quad (3-60)$$

The second integral will be shown to be zero by writing $(d\varphi/dx)(de/dx) = d/dx[e(d\varphi/dx)] - e d^2\varphi/dx^2$. Since φ satisfies Laplace's equation, $d^2\varphi/dx^2 = 0$. The other term is an exact differential. Consequently,

$$\int_0^1 \frac{d}{dx} \left(e \frac{d\varphi}{dx} \right) dx = e \frac{d\varphi}{dx} \Big|_0^1 = 0 \quad (3-61)$$

since $e = 0$ on the boundaries. Therefore, Eq. (3-60) becomes

$$F = \int_0^1 \left(\frac{du}{dx}\right)^2 dx = \int_0^1 \left(\frac{d\varphi}{dx}\right)^2 dx + \int_0^1 \left(\frac{du}{dx} - \frac{d\varphi}{dx}\right)^2 dx \quad (3-62)$$

Now du/dx is a function of the b 's, or equivalently of u_1 and u_2 , by Eqs. (3-58) and (3-59). If the integral on the left in Eq. (3-62) is minimized by varying the parameters in the trial function, then the mean square difference between the electric field ($-du/dx$) corresponding to the trial function and the true electric field will also be minimized. In this sense we obtain the best approximation to the true solution $\varphi(x)$.*

In the simple case being considered the integral F can be computed easily. In terms of the b 's it is

$$\begin{aligned} F &= \int_0^{x_1} b_1^2 dx + \int_{x_1}^{x_2} b_2^2 dx + \int_{x_2}^1 b_3^2 dx \\ &= b_1^2 x_1 + b_2^2 (x_2 - x_1) + b_3^2 (1 - x_2) \\ &= \frac{u_1^2}{x_1} + \frac{(u_2 - u_1)^2}{x_2 - x_1} + \frac{(1 - u_2)^2}{1 - x_2} \end{aligned} \quad (3-63)$$

Differentiating with respect to u_1 and u_2 and setting the two derivatives equal to zero gives two equations from which u_1 and u_2 can be determined. The result is $u_1 = x_1$ and $u_2 = x_2$. In other words, this particular finite element approximation gives the exact solution. This good fortune is

* This analysis can be carried through in exactly the same way in three dimensions (using the divergence theorem); in this case the quantity to be minimized is the integral of $\nabla u \cdot \nabla u$ over the volume of the system. The analysis can also be done if the value of the potential is specified on part of the boundary and the value of the electric field (normal derivative) is specified on the rest of the boundary.

primarily due to the simplicity of the problem. If we had used approximations of the form $u = a + bx^2$, for example, we would still have found an approximate solution, but it would not have been the exact solution. If the interval had been divided into a larger number of elements (or if the approximation functions had more parameters), there would have been more equations to solve for the u_i 's (and other parameters), and consequently we would almost certainly have had to resort to a numerical program for solving them. The complexities in the finite element problem are setting up the finite element decomposition, finding the linear equations, and then solving them.

Several things are apparent from the preceding considerations of a very simple problem. The finite difference problem is easier to understand and easier to set up, but it requires a uniform mesh. Because there are no restrictions on the sizes of the elements, the finite element method is easier to adapt to irregular geometries, and the mesh can easily be made finer where the potential is varying rapidly and coarser where it is varying more slowly. The result is better accuracy for a given amount of computation. The finite element method also gives an analytic approximation to the potential in each element. The finite difference method gives the potential only at discrete points and thus requires interpolation or other approximations to compute the field. Remember, however, that the finite element method is significantly more difficult to set up. There is much more to be said about each of these methods, and there is an extensive literature about each.

We next consider some more realistic, although still relatively simple, examples of numerical techniques. The computer programs implemented in the following examples use the programming language known as BASIC. A brief introduction to BASIC is given in Appendix I.

EXAMPLE 3-5
Potential in a Rectangular
Region

The problem is to find an approximation to the potential in a two-dimensional rectangular region with the boundaries held at specified potentials as shown in Fig. 3-8 with $V_{\text{TOP}} = 1.0 \text{ V}$, $V_{\text{LEFT}} = 0.3 \text{ V}$, $V_{\text{RIGHT}} = 0.7 \text{ V}$, and $V_{\text{BOTTOM}} = 0.0 \text{ V}$.

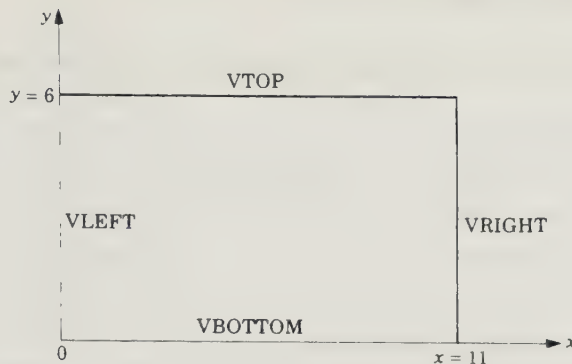
Solution: The finite difference method will be used. The iteration equation for this problem is

$$\varphi_{i,j}^{n+1} = \frac{w}{4} [\varphi_{i,j+1}^n + \varphi_{i+1,j}^n + \varphi_{i,j-1}^{n+1} + \varphi_{i-1,j}^{n+1}] + (1-w)\varphi_{i,j}^n \quad (3-64)$$

where the superscripts indicate the iteration number, and w is again the relaxation parameter. The iteration process is easy to program and a listing (in BASIC) for a straightforward program is given in Fig. 3-9. Lines 100 to 250 comprise the header and the data input lines. Lines 270

FIGURE 3-8

Rectangular region subject to boundary conditions to be solved numerically for electrostatic potential

**FIGURE 3-9**

BASIC program for two-dimensional finite difference approximation to Laplace's equation

```

100 '*****
110 '*
120 '*   LAPLAC4A: NUMERICAL SOLUTION OF LAPLACE'S EQUATION IN 2D RECTANGLE *
130 '*           NO ESSENTIAL LINE NUMBERS *
140 '* *
150 '*****
160 'The rectangle is 7x12 including the boundaries. The mesh is 1x1 squares.
170 DIM V(10, 20)
180 CLS : SCREEN 0
190 '
200 'ENTER DATA
210 PRINT "Enter values between 0 and 1 for the potentials on the four sides."
220 INPUT "VTOP = ", VT: INPUT "VBOT = ", VB
230 INPUT "VLEFT = ", VL: INPUT "VRIGHT = ", VR
240 PRINT "Enter value of iteration parameter, DM.": INPUT "DM = ", DM
250 PRINT "Enter the value of the relaxation parameter, w.": INPUT "w = ", W
260 '
270 LOCATE 23, 1: PRINT "Press any key to continue"
280 C$ = INPUT$(1): CLS : IT = 0
290 '
300 'Enter starting values for the potential at the grid points and on the bdy.
310 FOR J = 2 TO 11
320   JP = 6 * J - 6
330   V(1, J) = VT: LOCATE 1, JP: PRINT VT
340   V(7, J) = VB: LOCATE 19, JP: PRINT VB
350 NEXT J
360 FOR I = 2 TO 6
370   V(I, 1) = VL: V(I, 12) = VR: IP = 3 * I - 2
380   FOR J = 2 TO 11: V(I, J) = 0: NEXT
390   LOCATE IP, 1: PRINT V(I, 1)
400   FOR J = 2 TO 12: JP = 6 * J - 6: LOCATE IP, JP: PRINT V(I, J): NEXT
410 NEXT I
420 '
430 'Iteration procedure
440 IT = 0
450 DW = 1
460 WHILE DW > DM
470   IT = IT + 1
480   DW = 0
490   LOCATE 22, 5: PRINT "Iteration no. = "; IT
500   FOR I = 2 TO 6
510     FOR J = 2 TO 11
520       IP = 3 * I - 2: JP = 6 * J - 6: IA = IP + 1
530       VIJ = W * ((V(I - 1, J) + V(I + 1, J) + V(I, J - 1) + V(I, J + 1))
540         / 4) + (1 - W) * V(I, J)
550       DC = ABS(VIJ - V(I, J)): 'Residual at point I,J.
560       IF DC > DW THEN DW = DC: 'Largest residual in a given iteration.
570       V(I, J) = VIJ
580       LOCATE IP, JP: PRINT USING "####"; V(I, J)
590       LOCATE IA, JP: PRINT USING "####"; DC
600     NEXT J
610   NEXT I
620   WEND
630 LOCATE 22, 25: PRINT "- Value of DW achieved <="; DW
640 LOCATE 21, 1: PRINT "DM = "; DM, "w = "; W
650 END

```

FIGURE 3-10

Potential in a rectangular region calculated by the finite difference method (Example 3-5)

	1	1	1	1	1	1	1	1	1	1
.3	.6269	.7445	.7724	.8146	.8265	.8344	.8405	.8447	.8419	.8140
.3	.4630	.5588	.6105	.6393	.6572	.6706	.6831	.6962	.7088	.7141
.3	.3665	.4170	.4516	.4750	.4923	.5077	.5250	.5484	.5830	.6335
.3	.2859	.2912	.3038	.3169	.3294	.3428	.3607	.3895	.4412	.5370
.3	.1860	.1581	.1553	.1595	.1656	.1735	.1854	.2076	.2555	.3731
	0	0	0	0	0	0	0	0	0	0

to 410 set the initial values and print them in an array on the screen (the seven by twelve array was chosen because it is the largest array that will show four significant figures for the potential and the differences between the last two iterations on a standard screen). Lines 430 to 610 are the iteration routine. Within the outermost loop, line 460 checks to see if the iteration criterion is satisfied. If not, another iteration is made. Lines 570 and 580 print the value of the potential and the residual for each iteration as it is completed. If the value of the largest residual—that is, the difference between the last two iterations—is less than the specified iteration parameter, line 630 prints that value. The results obtained for an iteration parameter 10^{-6} with $w = 1.5$ are shown in Fig. 3-10. It required only 24 iterations to achieve these results and they almost certainly satisfy the difference equations to a high order of accuracy. Unfortunately, there is no guarantee that the result is an equally good approximation to the solution to Laplace's equation. We defer further consideration of this important issue until later in this section.

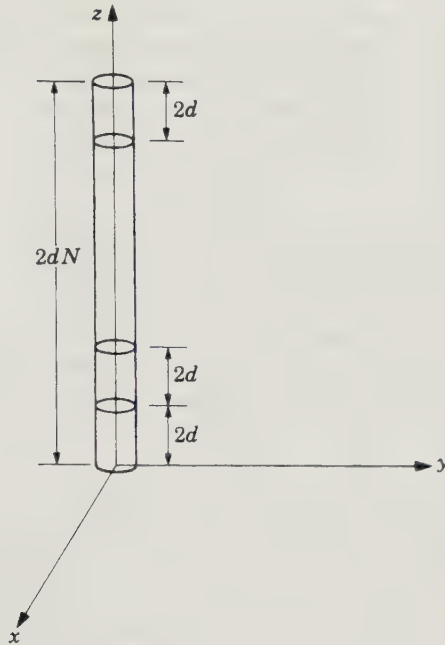
The next example uses the *method of moments*, which is in many ways similar to a finite element approach, to calculate the charge distribution on a long, thin cylindrical conductor.

EXAMPLE 3-6
Charge Distribution on a
Long, Thin Cylindrical
Conductor

The problem is to find the distribution of charge on a long, thin conducting cylinder. If charge is placed on the conductor it will distribute in such a way as to make the conductor an equipotential. If this charge distribution can be determined, the potential at any point in space can be found by using Eq. (2-15). Finding the charge distribution by analytical methods is difficult. Consequently we use a numerical method generally known as the *method of moments*, but also called the *boundary element method* and sometimes referred to as a finite element method. The cylinder has a circular cross section of radius a and lies along the z -axis. We will divide the cylinder into N elements each of

FIGURE 3-11

A charged wire divided into N segments for computational purposes



length $2d$ as shown in Fig. 3-11, and assume $a \ll d$. The total charge on the cylinder is Q but its distribution is not known.

Solution: We denote by q_i the charge on the i th element and assume that it is distributed uniformly over the surface of the element. These charged cylindrical sections are the finite elements—in this case boundary elements—and the trial function is the set of charge values on the various segments. Consider now one element located on the z -axis with its center at z' (Fig. 3-12). The potential at a point z on the axis is

$$\varphi(z) = \frac{q'}{8\pi\epsilon_0 d} \int_{z'-d}^{z'+d} \frac{dz''}{\sqrt{(z''-z)^2 + a^2}} \quad (3-65)$$

The transformation $\zeta = z - z''$ converts the integral into a standard form whose value is

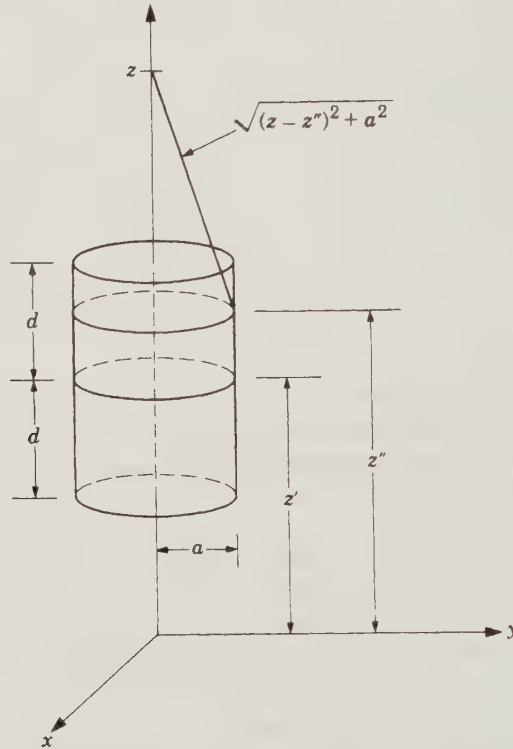
$$\varphi(z) = \frac{q'}{8\pi\epsilon_0 d} \ln \left[\frac{z - z' + d + \sqrt{(z - z' + d)^2 + a^2}}{z - z' - d + \sqrt{(z - z' - d)^2 + a^2}} \right] \quad (3-66)$$

The value of this integral at $z = z'$ —that is, the contribution of the charge on an element to its own potential—is approximately

$$\varphi(z') = \frac{q'}{4\pi\epsilon_0 d} \ln \left(\frac{2d}{a} \right) \quad (3-67)$$

FIGURE 3-12

The geometry for the potential produced by one charge element of a long straight wire. See Eq. (3-65).



where only the lowest order term in a/d has been kept. The contribution of the charge on an element at z' to the potential at z is (for $z' > z$)

$$\varphi(z) = \frac{q'}{8\pi\epsilon_0 d} \ln \left(\frac{z - z' + d}{z - z' - d} \right) \quad (3-68)$$

where a has been neglected compared to d . The potential of element one can now be written as

$$\begin{aligned} \varphi_1 &= \frac{q_1}{4\pi\epsilon_0 d} \ln \left(\frac{2d}{a} \right) + \frac{1}{8\pi\epsilon_0 d} \sum_{j=2}^N q_j \ln \left(\frac{2j-1}{2j-3} \right) \\ &= \sum_{j=1}^N A_{ij} q_j \end{aligned} \quad (3-69)$$

The first term is the contribution of the charge on the element itself and the remaining terms give the contribution of the other $N - 1$ elements.

For the second element the form is the same, with $A_{22} = A_{11}$ and $A_{2,1} = A_{1,2}$; $A_{2,j} = A_{1,j-1}$ for $j \geq 3$. In other words, the equation for φ_1 contains all of the values for the $A_{i,j}$'s appearing in all of the equations.

Generalizing to subsequent elements provides a useful recursion relation for calculating the matrix coefficient in the remaining equations. In addition, we note that by symmetry the charge on element j is the same as the charge on element $(N - j + 1)$. Thus, the number of equations can be reduced by a factor of two if N is even or by a slightly smaller factor if N is odd. We now have a set of simultaneous linear equations in the form of Eq. (I-1) in Appendix I, with the righthand sides (the D_i 's) all equal to the same value. Note again that this equality is because we are dealing with an equipotential surface (i.e., a conductor).

The BASIC program CYLCON, listed in Fig. 3-13, uses these results to set up the equations for the charges on the elements, assuming that the number of elements is even. The equations are solved by the program G-J2KA (called by line 430), which is listed in Appendix I. For

FIGURE 3-13
BASIC program to evaluate matrix coefficients for charged cylindrical wire (Example 3-6)

```

100 '*****
110 '*
120 '*          CYLCON: CONTROL PROGRAM TO CALCULATE CHARGE DISTRIBUTION
130 '*          ON A CYLINDRICAL CONDUCTOR - USE WITH G-J2KA
140 '*          LINE NUMBER 510 IS ESSENTIAL
150 '*
160 '*****
170 '
180 'Identifiers:
190 '   L   length of wire
200 '   A   radius of wire
210 '   N   number of segments (even)
220 '   2D = L/N length of segments
225 CLS
230 INPUT; "L = ", L:INPUT "   A = ", A
240 INPUT; "Even number of segments, N = ", N%
250 INPUT "   Array dimension M1>N/2, M1 = ", M1%
260 DIM A(M1%, M1%), D(M1%), W(M1%)
270 N1% = N% / 2: N2% = N% / 2
280 'Equation A(I,J)*Q(J) = V(I), V(I) = 1 for all I
290 '
300 'Compute coefficients, A(I,J)'S
310 FOR I = 1 TO N% / 2
320   FOR J = I + 1 TO N% / 2
330     A(I,J) = (1/2)*LOG( (J-I+1/2) / (J-I-1/2) ) +
                LOG( (N%-I-J+3/2) / (N% - I - J + 1/2) )
340   NEXT J
350   FOR J = 1 TO I - 1
360     A(I,J) = A(J,I)
370   NEXT J
380   A(I,I) = LOG(L/(N%*A))+(1/2)*LOG( (2*(N%-2*I)+3) / (2*(N%-2*I)+1) )
390   D(I) = 1
400 NEXT I
410 '
420 'Solve equations
430 GOSUB 2000
440 '
450 'Print results
460 IF N% / 2 > 20 THEN 510
470 FOR I = 1 TO N% / 2
480   PRINT "Q("I") = " TAB(11):PRINT USING "#.#####"; W(I)
490 NEXT I
500 END
510 FOR I = 1 TO 20
520   PRINT "Q("I") = " TAB(11):PRINT USING "#.#####"; W(I),
530   PRINT TAB(25) "Q("I + 20") = " TAB(36):PRINT USING "#.#####"; W(I + 20)
540 NEXT I
550 END

```

TABLE 3-3

Solution to the Charged Wire Problem (Example 3-6). ($i = 1$ is an end segment; $i = 20$ is one of center segments.)

i	q	i	q
1	0.20401	11	0.14614
2	0.17210	12	0.14545
3	0.16340	13	0.14487
4	0.15838	14	0.14440
5	0.15502	15	0.14400
6	0.15256	16	0.14369
7	0.15068	17	0.14344
8	0.14919	18	0.14326
9	0.14797	19	0.14314
10	0.14697	20	0.14309

Note: Wire is 1.0 m long and 0.001 m in diameter, with $N = 40$ segments. $D_i = \varphi_i = 1.0$ volt for all i . The values given are actually $q_i/(4\pi\epsilon_0 d)$; to obtain the charge on each segment in coulombs, multiply the value given by d (in m) divided by 9×10^9 , with $d = \text{length}/(2N)$.

a wire 1.0 meter long, 1.0 millimeter in diameter, at a potential of 1.0 volt, the numerical results using 40 elements are tabulated in Table 3-3. This calculation is time-consuming even with a fast computer, but remember that 210 matrix elements are being calculated and twenty linear equations with no zero coefficients are being solved. This method does not calculate the potential everywhere and then determine the charge by differentiating. Instead, it determines the charge directly. This approach is particularly useful in determining the capacitance of an array of conductors, as discussed in Chapter 6. Also, as already noted, knowing the charge distribution makes it possible to calculate the potential and the electric field at any point in space. This method works equally well in three dimensions as Problem 6-28 shows. As usual with finite element problems, the difficulties lie in setting up the equations and solving them. None of the coefficients is zero in the equations for this problem, and the solution process is thus even more computation intensive.

These two examples only scratch the surface. The finite difference method is sufficiently simple that general programs are seldom seen. Programs are usually tailored to specific classes of problems and specific input and output requirements. An exception is the general rectangular

coordinate program listed by Koonin,* which deals with arbitrary boundaries and arbitrary boundary values in a very interesting way. The finite element method, on the other hand, is difficult to set up, and it too requires the solution of large arrays of linear equations. As a result many programs are available. Sylvester and Ferrari† list a number of programs (in Fortran 77) for two-dimensional problems. Many commercial packages are also available and, in fact, the finite element method has become so important as an engineering design tool that some firms have developed proprietary software for their purposes.

There are, needless to say, many other valuable applications of computers to electrostatic problems. We will illustrate two. In the previous example, the method of moments was used to calculate the charge distribution on a long cylinder. Having this charge distribution, the resulting potential and electric field can be calculated using Eqs. (2-15) and (2-8), respectively. For a single point, this calculation involves adding the separate contributions from each of what may be a very large number of the elements, a laborious task that can be easily delegated to the computer. Finding equipotential surfaces or flux lines is even more tedious, but with a little ingenuity the computer can also be programmed to do these calculations. Thus the direct calculation of potentials and fields is one of our other applications. The computer can also be used to evaluate an analytical solution to Laplace's equation as shown in the following example.

EXAMPLE 3-7

Numerical Evaluation of the Fourier Series Expansion for the Potential in a Rectangular Region

The problem is to evaluate the Fourier series expansion of the potential in a rectangular region as determined in Problem 3-18. In that problem, the solution to Laplace's equation in a two-dimensional rectangular region, bounded by the lines $x = 0$, $y = 0$, $x = 11$, and $y = 6$ subject to the boundary conditions $\varphi(0, y) = 0$, $\varphi(x, 6) = 0$, $\varphi(11, y) = 0$, $\varphi(x, 0) = 1$, is found to be

$$\varphi(x, y) = \frac{4}{\pi} \sum_{i=0}^{\infty} \frac{1}{2i+1} \frac{\sin \left[\frac{(2i+1)\pi x}{11} \right] \sinh \left[\frac{(2i+1)\pi(6-y)}{11} \right]}{\sinh \left[\frac{(2i+1)6\pi}{11} \right]}$$

This solution is exact, but if numerical values of the potential are required at many points, the calculations are formidable.

* Koonin, Steven, J., *Computational Physics* (Reading, MA: Addison-Wesley, 1986), p. 376 ff.

† Sylvester, P. P., and R. L. Ferrari, *Finite Elements for Electrical Engineers*, 2nd ed. (Cambridge: Cambridge University Press, 1990).

FIGURE 3-14

BASIC program to calculate the potential at points on a grid inside a rectangle

```

100 /*****
110 /*
120 /*          FOUCCOEF2: POTENTIAL IN RECTANGULAR REGION USING FOURIER SERIES
130 /*          NO ESSENTIAL LINE NUMBERS
140 /*
150 /*****
160 'Region is 0<=X<=12, 0<=Y<=7; potential on X=0, X=12, and Y=7 boundaries is
161 'zero; potential on y=0 boundary is one.
170 PI = 4*ATN(1)
180 DIM S(50)
190 CLS
200 INPUT "Number of terms less than 26, M =";M
210 CLS
220 DEF FNSINH(Z) = (EXP(Z)-EXP(-Z))/2
230 FOR Y=6 TO 0 STEP -1
240   FOR X = 1 TO 10
250     S(1) = (4/PI)*SIN(PI*X/11)*FNSINH(PI*(6-Y)/11)/FNSINH(PI*6/11)
260     FOR N = 1 TO M - 1
270       S(N+1) = S(N) + (4/PI)*(1/(2*N+1))*SIN((2*N+1)*PI*X/11)*FNSINH(PI
                *(2*N+1)*(6-Y)/11)/FNSINH(PI*(2*N+1)*6/11)
280     NEXT N
290     PRINT USING "#.#### "; S(M),
300   NEXT X
310   PRINT
320 NEXT Y
330 PRINT
340 PRINT "Number of terms =";M
350 END

```

Solution: A very short program, shown in Fig. 3-14, will calculate the potential at each point on a grid with unit spacing inside the rectangle. The results for the sum of the first 25 terms in the series are shown in Fig. 3-15(a).

It is instructive to compare the results obtained using Fourier series with those obtained by the finite difference method. The rectangular region used in Example 3-7 is the same as the one used earlier in the finite element calculation, Example 3-5. Thus we can use the program of Fig. 3-9 to obtain an approximate solution. The result is shown in Fig. 3-15(b). The values obtained by the two methods generally agree to two significant figures. This agreement is reassuring, but a little disappointing. The difficulty is that we have in fact solved two different problems to a very high degree of accuracy. The finite difference equations are an approximation to Laplace's equation. In other words, we have defined a *new problem* that approximates the problem of solving Laplace's equation. The approximation can only be improved by defining another problem using a finer mesh and accepting the larger computational burden. The Fourier series solution is exact if an infinite number of terms are included, but it is approximate if the series is truncated. In other words truncation produces an approximation to the (known) exact solution to the specified problem. In the finite difference case we approximate the problem and solve the approximate problem to any desired degree of accuracy. The Fourier series approach finds an exact solution and then approximates the solution to any desired accuracy. The difference between approximating the problem and approximating the exact

FIGURE 3-15

Comparison of approximate potential in a two-dimensional rectangular region as calculated using a Fourier series expansion and using a finite difference method

(a) Expansion in Basis Functions										
	0	0	0	0	0	0	0	0	0	0
0	.0425	.0790	.1062	.1235	.1317	.1317	.1235	.1062	.0790	.0425
0	.0915	.1680	.2225	.2559	.2715	.2715	.2559	.2225	.1680	.0915
0	.1573	.2801	.3603	.4060	.4264	.4264	.4060	.3603	.2801	.1573
0	.2638	.4375	.5324	.5806	.6008	.6008	.5806	.5324	.4375	.2638
0	.4844	.6736	.7484	.7813	.7944	.7944	.7813	.7484	.6736	.4844
	1.003	1.023	.9933	.9882	1.008	1.008	.9882	.9933	1.023	1.003
(b) Finite Difference Method										
	0	0	0	0	0	0	0	0	0	0
0	.0433	.0797	.1063	.1230	.1309	.1309	.1230	.1063	.0797	.0433
0	.0935	.1694	.2224	.2547	.2698	.2698	.2547	.2224	.1694	.0935
0	.1615	.2818	.3592	.4037	.4237	.4237	.4037	.3592	.2818	.1615
0	.2706	.4369	.5292	.5771	.5976	.5976	.5771	.5292	.4369	.2706
0	.4842	.6661	.7433	.7781	.7919	.7919	.7781	.7433	.6661	.4842
	1	1	1	1	1	1	1	1	1	1

solution is frequently important when computational methods are used to solve problems.

Lest one think that expansion in a series of orthogonal basis functions is always a better approach, we note that for less convenient boundary conditions obtaining the series expansion is very laborious, whereas in the finite difference method, it is only necessary to enter the values on the boundary. The rest of the work is done by the computer. Also, in some regions, the Fourier series converges extremely slowly. The choice of method depends on the specific problem and the use to be made of the solution.

SOLUTIONS OF POISSON'S EQUATION

In the preceding sections, we have dealt exclusively with Laplace's equation and its solution. Laplace's equation is applicable to those electrostatic problems in which all the charge resides on surfaces of conductors or is concentrated in the form of point or line charges. (We shall see in the next

chapter that if the region between the conductors is filled with one or more simple dielectric media, then Laplace's equation still holds in these media.)

Let us consider, now, an electrostatic problem in which part of the charge (the prescribed charge) is given by $\rho(x, y, z)$, a known function, and the rest of the charge (the induced charge) resides on the surfaces of conductors. Such a problem requires the solution of Poisson's equation. The general solution to this problem may be written as an integral of the type Eq. (3-1) over the *prescribed charge* plus a general solution to Laplace's equation. The solution to Laplace's equation must be chosen, however, so that the entire potential satisfies all boundary conditions.

When all of the charge is prescribed—that is, when $dq = \rho(x, y, z) dv$ is known at all points in space—then Eq. (3-1) represents the entire solution to Poisson's equation, and this integral may be performed (either analytically or numerically). There is one case, however, where the solution to Poisson's equation may be obtained more directly than by means of the formal solution Eq. (3-1); it occurs when both ρ and φ are functions of only one independent variable. As an example of this case, let ρ be a function of the spherical coordinate r only, and let the entire charge be distributed in a spherically symmetric way. Then Eq. (3-5b) becomes

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d\varphi}{dr} \right) = -\frac{1}{\epsilon_0} \rho(r) \quad (3-70)$$

We shall assume that the total charge is bounded—that is, either the charge does not extend to infinity or the charge density drops off sufficiently rapidly at large radii. Equation (3-70) may then be integrated directly, assuming the function $\rho(r)$ given, and the two constants of integration may be determined (1) from Gauss's law for the electric field at some radius and (2) from the fact that $\varphi \rightarrow 0$ as $r \rightarrow \infty$.

3-14
SUMMARY

The fundamental first-order vector differential equations for the electrostatic field, $\nabla \times \mathbf{E} = 0$ and $\nabla \cdot \mathbf{E} = \rho/\epsilon_0$, can be combined into a single scalar second-order differential equation, **Poisson's equation**:

$$\nabla^2 \varphi = -\frac{\rho}{\epsilon_0}$$

where $\mathbf{E} = -\nabla\varphi$. If $\rho(\mathbf{r})$ is a given function in a region V , Poisson's equation has the particular solution

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int_V \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dv'$$

as can be seen by operating on the right side of this expression with ∇^2 inside the integral. To this particular solution can be added any solution of

the corresponding homogeneous equation, **Laplace's equation**:

$$\nabla^2\varphi = 0$$

The *appropriate solution* of Laplace's equation is one that will satisfy the *boundary conditions* on the boundary of V . If φ or $\partial\varphi/\partial n$ is specified on a closed boundary, the solution is unique to within an additive constant.

- If $\rho = 0$ everywhere inside V , the complete solution is the solution of Laplace's equation that is continuous inside V and satisfies the imposed conditions at the boundary of V . The latter are $\varphi = \text{constant}_1$, constant_2 , and so on, if the boundary is composed of conductors.

- Analytical solutions for Laplace's equation can be composed of **linear combinations from a set of basis functions**, which is chosen by considering any symmetry that appears in the particular problem. The set of zonal harmonics is useful for spherical boundaries and azimuthal symmetry, and the cylindrical harmonics for cylindrical boundaries.

- The **method of images** is a technique where one devises a configuration of fictitious image charges outside of V such that the potential field of these image charges plus the real charges inside V satisfies the boundary conditions for φ on the given boundaries of V . The uniqueness theorem assures that the field inside V is the correct one. The technique can be applied in only a few situations where the symmetry is adequate. The simplest example is a point charge in front of a conducting plane.

- When all the charges reside on conducting surfaces, Laplace's equation requires that the potentials of the conductors are linear functions of their charges:

$$\varphi_i = \sum_j p_{ij} Q_j$$

- Two useful methods for solving Laplace's equation numerically are the **finite difference method** and the **finite element method**. In the finite difference method, the potential is determined at mesh points in the region by an iterative process; in the finite element method, the region is divided into subregions in which analytic approximations are postulated and then parameterized and optimized.

REFERENCES

The following texts are recommended for (1) a more complete discussion of Legendre's equation, (2) the general form of Laplace's equation in orthogonal, curvilinear coordinates, and (3) a more complete discussion of the solution to Laplace's equation:

- Mathematical Physics* by E. Butkov (Reading, Mass: Addison-Wesley, 1968).
- Electromagnetic Theory* by J. A. Stratton (New York: McGraw-Hill, 1941).
- An Introduction to Differential Equations and Their Applications*, 2nd ed., by S. L. Campbell (Belmont, CA: Wadsworth, 1990).

PROBLEMS

3-1 Two spherical conducting shells of radii r_a and r_b are arranged concentrically and are charged to the potentials φ_a and φ_b , respectively. If $r_b > r_a$, find the potential at points between the shells and at points $r > r_b$.

3-2 Two long cylindrical shells of radii r_a and r_b are arranged coaxially and are charged to the potentials φ_a and φ_b , respectively. Find the potential at points between the cylindrical shells.

3-3 If φ_1 is a solution to Laplace's equation, prove that the partial derivative of φ_1 with respect to one or more of the rectangular coordinates (e.g., $\partial\varphi_1/\partial x$, $\partial^2\varphi_1/\partial x^2$, $\partial^2\varphi_1/\partial x \partial y$, etc.) is also a solution.

3-4 Suppose φ satisfies Laplace's equation throughout a region V_0 . (a) Prove that the value of φ at any point O is the average of its values over the surface of any sphere centered at O that lies entirely in V_0 :

$$\varphi(O) = \frac{1}{4\pi R^2} \oint \varphi da$$

where R is the radius of the sphere. [Hint: Let $\psi = 1/r$ in Eq. (1-57).] (b) Show that consequently φ has no maxima or minima inside V_0 .

3-5 Expand the function

$$F(u) = (1 - 2xu + u^2)^{-1/2}$$

in Taylor series up to the term in u^3 . Note that the coefficients are the first four Legendre polynomials $P_n(x)$. In fact, $F(u)$ is a *generating function* for all the Legendre polynomials:

$$F(u) = \sum_{n=0}^{\infty} P_n(x)u^n$$

3-6 Show that half the zonal harmonics are generated by differentiating r^{-1} successively with respect to the rectangular coordinate z ($z = r \cos \theta$).

3-7 Obtain $\nabla^2\varphi$ in cylindrical coordinates, Eq. (3-8), from the rectangular form, Eq. (3-6), by direct substitution: $x = r \cos \theta$, $y = r \sin \theta$.

3-8 (a) Find the potential of an axial quadrupole: point charges q , $-2q$, q placed on the z -axis at distances l , 0 , $-l$ from the origin. (b) Find the potential only at distances $r \gg l$. (c) Show that this potential is proportional to one of the zonal harmonics.

3-9 Suppose a point dipole is located at the center of a grounded spherical conducting shell. Find the potential inside the shell. [Hint: Use zonal harmonics that are regular at the origin to satisfy the boundary conditions on the shell.]

3-10 Consider an uncharged conducting sphere placed in an initially uniform electric field. (a) Show that the potential due to the sphere is that of a point dipole. (b) Find the induced dipole moment.

3-11 A conducting sphere of radius a bearing total charge Q is placed in an initially uniform electric field E_0 . Find the potential at all points exterior to the sphere.

3-12 A long cylindrical conductor of radius a bearing no net charge is placed in an initially uniform electric field E_0 . The direction of E_0 is perpendicular to the cylinder

axis. (a) Find the potential at points exterior to the cylinder. (b) Find the charge density on the cylindrical surface.

3-13 An infinitely long cylindrical shell (of radius R) is approximated by a large number of parallel wires that can be maintained at arbitrary potentials. (a) If the potential of the shell has the value $V_0 \sin \theta$, find the potential and electric field inside and outside the shell. (b) Sketch the lines of force.

3-14 A point charge q is situated a distance d from a grounded conducting plane of infinite extent. Obtain the total charge induced on the plane by direct integration of the surface charge density.

3-15 Two point charges, q_1 and q_2 , are located near a grounded conducting plane of infinite extent. (a) Find the image charges that are needed to make the plane a surface of constant potential. (b) From the result obtained in part (a), can you predict the image charge distribution required for the case of a body of arbitrary shape with charge density ρ situated near a conducting plane of infinite extent?

3-16 Two grounded conducting planes intersect at 60° and a point charge q lies between them. Determine the positions of the image charges that will give the electric field between the planes.

3-17 A point charge q is located between two parallel, grounded, conducting planes that are separated by a distance d . (a) Find the locations of the infinite number of image charges. (b) Express the force on the charge q by an infinite series.

***3-18** (a) Use the method described in Section 3-7 to find a solution to Laplace's equation in the 2-D rectangular region bounded by the lines $x = 0$, $y = 0$, $x = 11$, $y = 6$, with the boundary conditions $\varphi(0, y) = \varphi(x, 6) = \varphi(11, y) = 0$, $\varphi(x, 0) = 1$. [Hint: Use the $\varphi = 0$ condition to show that the solution can be written as a sum of terms consisting of products of sine functions of x and hyperbolic sine functions of y , in which each term satisfies all three of the $\varphi = 0$ boundary conditions.] (b) Determine the coefficients by requiring that $\varphi(x, 0) = 1$ for all x in this standard Fourier series problem.

3-19 Show that the problem of an uncharged conducting sphere in an initially uniform electric field E_0 may be solved by means of images. [Hint: A uniform electric field in the vicinity of the origin may be approximated by the field of two point charges Q and $-Q$ placed on the z -axis at $z = -L$ and $z = +L$, respectively. The field becomes more nearly uniform as $L \rightarrow \infty$. It is evident that $Q/2\pi\epsilon_0 L^2 = E_0$.]

3-20 A point charge q is located inside and at distance r from the center of a spherical conducting shell. The inner radius of the shell is a . (a) Show that this problem can be solved by the image technique. (b) Find the charge density σ induced on the inside surface of the shell. (Note: The potential of the spherical shell cannot be completely specified in terms of q and its image, because exterior fixed charges can also contribute. Nevertheless, these exterior charges will add only a constant term to the potential.) (c) Find the total charge induced on the inside surface of the shell (i) by physical arguments, and (ii) by integration of σ over the surface.

3-21 A long conducting cylinder bearing a charge λ per unit length is oriented parallel to a grounded conducting plane of infinite extent. The axis of the cylinder is at distance x_0 from the plane, and the radius of the cylinder is a . (a) Find the location of the line image. (b) Find the constant M (which determines the potential of the cylinder) in terms of a and x_0 .

3-22 A spherical distribution of charge is characterized by a constant charge density ρ for $r \leq R$. For radii greater than R , the charge density is zero. (a) Find the potential $\varphi(r)$ by integrating Poisson's equation. (b) Check this result by evaluating the integral (3-1). [Hint: To perform (3-1), divide the charge region into spherical concentric shells of thickness dr .]

3-23 A dipole \mathbf{p} is oriented normal to and at distance d from an infinite conducting plane. The plane is grounded (i.e., at zero potential). Calculate the force exerted on the plane by the dipole.

3-24 Consider the one-dimensional example discussed in Section 3-12. (a) With $\varphi(0) = 0$ and $\varphi(1) = 1$, write a program for solving Laplace's equation by the finite difference method. (b) Explore the convergence in the first few iterations. (Take $N = 10$ and 20 , $w = 1$ and 1.5 .)

3-25 A two-dimensional region in the shape of a right triangle has two sides of 5 cm oriented parallel to the x - and y -axes and a hypotenuse of length $5\sqrt{2}$ cm. The hypotenuse is maintained at 50 V and the other two sides are maintained at zero. (a) Use the finite difference method with mesh points every 1 cm in each direction, and iterate to determine the potentials at each mesh point to the nearest volt. (b) Sketch in the equipotentials.

***3-26** A two-dimensional L -shaped region is bounded by conductors on two sides with potentials $\varphi = 0$ and 9 V as shown in Fig. 3-16a. (a) Use the finite difference method to determine the potential at interior mesh points and determine the electric field at 45° in the vicinity of the 270° interior angle. (Note: by symmetry the solution in one leg of the L -shape can be obtained from the other by reflection through the dotted line. Thus, we can carry out the solution in the configuration of Fig. 3-16b, except that after each iteration the mesh points to the right of the dotted line are to be given the values of their mirror images above the line.) (b) Solve first with a mesh of $1\text{ cm} \times 1\text{ cm}$. (c) Repeat with a mesh of $1/2\text{ cm} \times 1/2\text{ cm}$.

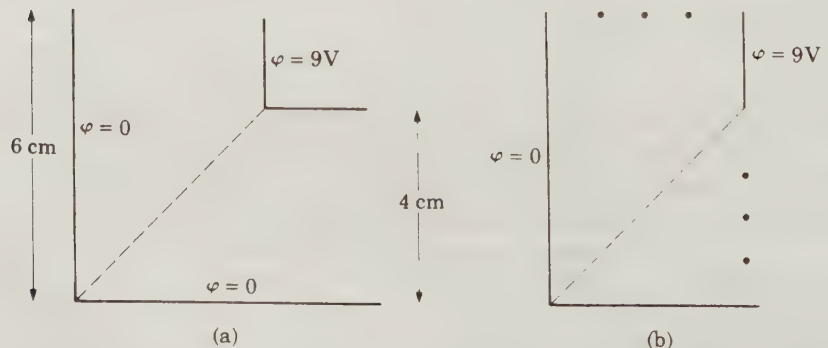


FIGURE 3-16

The Electrostatic Field in Dielectric Media

Thus far, we have ignored problems involving dielectric media and have dealt with cases in which the electric field is produced exclusively either by charges in a specified distribution or by free charge on the surface of conductors. In this chapter, we remedy this situation by considering a more general case.

An *ideal dielectric material* is one that has no free charges. Nevertheless, all material media are composed of molecules, which are composed of charged entities (atomic nuclei and electrons), and the molecules of the dielectric are certainly affected by the presence of an electric field. The electric field causes a force to be exerted on each charged particle, with positive particles being pushed in the direction of the field and negative particles oppositely, so that the positive and negative parts of each molecule are displaced from their equilibrium positions in opposite directions. These displacements, however, are limited (in most cases to very small fractions of a molecular diameter) by strong restoring forces that are set up by the changing charge configuration in the molecule. The term “bound charge,” in contrast to the “free charge” of a conductor, is sometimes used to emphasize that such molecular charges are not free to move very far or to be extracted from the dielectric material. The overall effect from the macroscopic point of view is most easily visualized as a displacement of the entire positive charge in the dielectric relative to the negative charge. The dielectric is said to be polarized.

A *polarized dielectric*, even though it is electrically neutral on the average, produces an electric field at exterior points and inside the dielectric as well. As a result, we are confronted with what appears to be an awkward situation: The polarization of the dielectric depends on the total electric field in the medium, but a part of the electric field is produced by the dielectric itself. Furthermore, the distant electric field of the dielectric may

modify the free charge distribution on conducting bodies, which, in turn, will change the electric field *in the dielectric*. The main purpose of this chapter is to develop general methods for handling this curious situation.

4-1

POLARIZATION

In this section, we consider a small volume element Δv of a dielectric medium that is, as a whole, electrically neutral. If the medium is polarized, then a separation of positive and negative charge has been effected, and the volume element is characterized by an electric dipole moment

$$\Delta \mathbf{p} = \int_{\Delta v} \mathbf{r} dq \quad (4-1)$$

According to Section 2-9, this quantity determines the electric field produced by Δv at distant points (i.e., at distances that are large compared with the dimensions of the volume element).

Since $\Delta \mathbf{p}$ depends on the size of the volume element, it is more convenient to work with \mathbf{P} , the electric dipole moment per unit volume. \mathbf{P} is usually called the *electric polarization*, or simply the *polarization*, of the medium.

$$\mathbf{P} = \frac{\Delta \mathbf{p}}{\Delta v} \quad (4-2)$$

Strictly speaking \mathbf{P} must be defined as the limit of this quantity as Δv becomes very small from the macroscopic viewpoint. In this way \mathbf{P} becomes a point function, $\mathbf{P}(x, y, z)$.

The dimensions of \mathbf{P} are charge per unit area; in mks units, C/m^2 .

It is apparent that $\mathbf{P}(x, y, z)$ is a vector quantity that, in each volume element, has the direction of $\Delta \mathbf{p}$. This quantity, in turn, has the direction of displacement of positive charge relative to negative charge (Fig. 4-1).

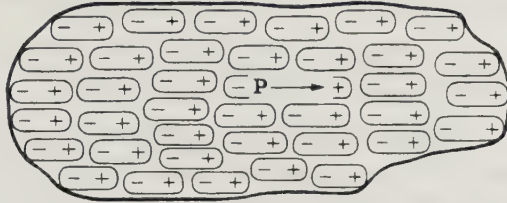
Although Δv is assumed very small from the macroscopic point of view, it still contains many molecules. Since a molecule is one of the small, electrically neutral entities that make up the dielectric material, it is sometimes desirable to speak about the electric dipole moment \mathbf{p}_m of a single molecule; that is,

$$\mathbf{p}_m = \int_{\text{molecule}} \mathbf{r} dq \quad (4-3)$$

It is evident from Eq. (4-1) that the dipole moment associated with Δv is given by $\Delta \mathbf{p} = \sum \mathbf{p}_m$, where the summation extends over all molecules

FIGURE 4-1

A piece of polarized dielectric material. Each volume element is represented as a dipole $\Delta\mathbf{p}$.



inside the element Δv . Hence,

$$\mathbf{P} = \frac{1}{\Delta v} \sum_m \mathbf{p}_m \quad (4-4)$$

This approach will be developed further in Chapter 5.

Although Fig. 4-1 represents each volume element of the polarized dielectric as a small dipole, it may be more instructive to visualize the dielectric in terms of its molecules and to imagine that each dipole of Fig. 4-1 represents a single molecule.

4-2

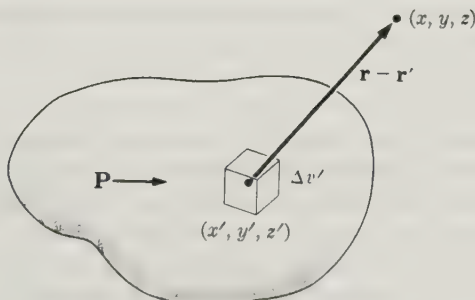
FIELD OUTSIDE OF A DIELECTRIC MEDIUM

We consider a finite piece of dielectric material that is polarized—that is, it is characterized at each point \mathbf{r}' by a polarization, $\mathbf{P}(\mathbf{r}')$. The polarization gives rise to an electric field, and our problem is to calculate this field at point \mathbf{r} , which is outside of the dielectric body (Fig. 4-2). As in Chapter 2, we shall find it more convenient to calculate first the potential $\varphi(\mathbf{r})$ and obtain the electric field as minus the gradient of φ .

Each volume element $\Delta v'$ of the dielectric medium is characterized by a dipole moment $\Delta\mathbf{p} = \mathbf{P} \Delta v'$, and since the distance between (x, y, z) and $\Delta v'$ is large compared with the dimensions of $\Delta v'$, this dipole moment

FIGURE 4-2

The electric field at (x, y, z) may be calculated by summing up the contributions due to the various volume elements $\Delta v'$ in V_0 . The surface of V_0 is denoted by S_0 .



completely determines the contribution of charges in $\Delta v'$ to the potential:

$$\Delta\varphi(\mathbf{r}) = \frac{\Delta\mathbf{p} \cdot (\mathbf{r} - \mathbf{r}')}{4\pi\epsilon_0 |\mathbf{r} - \mathbf{r}'|^3} = \frac{\mathbf{P}(\mathbf{r}') \cdot (\mathbf{r} - \mathbf{r}') \Delta v'}{4\pi\epsilon_0 |\mathbf{r} - \mathbf{r}'|^3} \quad (4-5)$$

Here $\mathbf{r} - \mathbf{r}'$ is the vector, directed out from $\Delta v'$, whose magnitude is given by

$$|\mathbf{r} - \mathbf{r}'| = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2} \quad (4-6)$$

The entire potential at point \mathbf{r} is obtained by summing the contributions from all parts of the dielectric:

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int_{V_0} \frac{\mathbf{P}(\mathbf{r}') \cdot (\mathbf{r} - \mathbf{r}') dv'}{|\mathbf{r} - \mathbf{r}'|^3} \quad (4-7)$$

This result is correct, and φ may be evaluated directly from Eq. (4-7) if the functional form of \mathbf{P} is known. It will be to our advantage, however, to express Eq. (4-7) in a rather different way by means of a simple mathematical transformation.

If $|\mathbf{r} - \mathbf{r}'|$ is given by Eq. (4-6), then

$$\nabla' \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) = + \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} \quad (4-8)$$

as may be seen by direct application of the gradient operator in Cartesian coordinates. The ∇' operator involves derivatives with respect to the primed coordinates. In certain circumstances, it may be desirable to use a gradient operator with respect to the unprimed coordinates; this operator will be indicated in the usual way by ∇ . Evidently, ∇' operating on a function of $|\mathbf{r} - \mathbf{r}'|$ is equal to $-\nabla$ operating on the same function. We shall require the ∇ operator later in order to get the electric field at point \mathbf{r} . However, in performing the integral Eq. (4-7) over the dielectric volume V_0 , the point \mathbf{r} is held fixed. Hence, the integrand of Eq. (4-7) may be transformed by means of Eq. (4-8):

$$\frac{\mathbf{P} \cdot (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} = \mathbf{P} \cdot \nabla' \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) \quad (4-9)$$

Equation (4-9) may be further transformed by means of the vector identity (1-1-7) of Table 1-1:

$$\nabla' \cdot (f\mathbf{F}) = f\nabla' \cdot \mathbf{F} + \mathbf{F} \cdot \nabla' f \quad (4-10)$$

where f is any scalar point function and \mathbf{F} is an arbitrary vector point function. Here again, the prime indicates differentiation with respect to the primed coordinates. Letting $f = (1/|\mathbf{r} - \mathbf{r}'|)$ and $\mathbf{F} = \mathbf{P}$, the integrand, Eq. (4-9), becomes

$$\frac{\mathbf{P} \cdot (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} = \nabla' \cdot \left(\frac{\mathbf{P}}{|\mathbf{r} - \mathbf{r}'|} \right) - \frac{1}{|\mathbf{r} - \mathbf{r}'|} \nabla' \cdot \mathbf{P} \quad (4-11)$$

Finally, the potential, Eq. (4-7), may be written as

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \oint_{S_0} \frac{\mathbf{P} \cdot \mathbf{n} da'}{|\mathbf{r} - \mathbf{r}'|} + \frac{1}{4\pi\epsilon_0} \int_{V_0} \frac{(-\nabla' \cdot \mathbf{P}) dv'}{|\mathbf{r} - \mathbf{r}'|} \quad (4-12)$$

where the volume integral of $\nabla' \cdot (\mathbf{P}/|\mathbf{r} - \mathbf{r}'|)$ has been replaced by a surface integral through application of the divergence theorem, and \mathbf{n} is the outward normal to the surface element da' (outward means out of the dielectric).

The quantities $\mathbf{P} \cdot \mathbf{n}$ and $-\nabla \cdot \mathbf{P}$ that appear in the integrals of Eq. (4-12) are two scalar functions obtained from the polarization \mathbf{P} . It is expedient to give these quantities special symbols. Since they have the dimensions charge per unit area and charge per unit volume, respectively, we write

$$\sigma_P \equiv \mathbf{P} \cdot \mathbf{n} = P_n \quad (4-13)$$

and

$$\rho_P \equiv -\nabla \cdot \mathbf{P} \quad (4-14)$$

We call σ_P and ρ_P *polarization charge densities*.

The surface density of polarization charge is given by the component of polarization normal to the surface, and the volume density of polarization charge is a measure of the nonuniformity of the polarization inside the material.

The potential due to the dielectric material

$$\begin{aligned} \varphi(\mathbf{r}) &= \frac{1}{4\pi\epsilon_0} \left[\oint_{S_0} \frac{\sigma_P da'}{|\mathbf{r} - \mathbf{r}'|} + \int_{V_0} \frac{\rho_P dv'}{|\mathbf{r} - \mathbf{r}'|} \right] \\ &= \frac{1}{4\pi\epsilon_0} \int \frac{dq'_P}{|\mathbf{r} - \mathbf{r}'|} \end{aligned} \quad (4-15)$$

is now written in such a way that it is evident that it arises from a charge distribution. In other words, the dielectric material has been replaced by an appropriate distribution of polarization charge.

Although Eq. (4-15) has been obtained by means of a mathematical transformation, it should be possible to understand σ_P and ρ_P on purely physical grounds. That a surface charge density σ_P exists is evident from Fig. 4-1, where it is seen that this charge is made up from the ends of similarly oriented dipoles. In this way, a charge density is developed on every surface that is not parallel to the polarization vector. Turning now to ρ_P , we expect that $\rho_P \Delta v'$ represents the *excess* or *net charge* in the volume element $\Delta v'$. That this is truly the case may be seen in the following way: Let us define two charge densities ρ^+ and ρ^- as representing the total

positive charge and the total negative charge per unit volume, respectively. That is, ρ^+ represents all the atomic nuclei in unit volume of the dielectric and, similarly, ρ^- counts all the electrons. In the unpolarized state, each volume element of the dielectric is electrically neutral. Hence,

$$\rho_0^+(x', y', z') + \rho_0^-(x', y', z') = 0 \quad (4-16)$$

where the subscript zero denotes densities in the unpolarized configuration. Let us assume that as a consequence of polarization the positive charge is displaced by $\delta^+(x, y, z)$ and the negative charge by $\delta^-(x, y, z)$. The positive charge crossing an element of area da' is $\rho_0^+ \delta^+ \cdot \mathbf{n} da'$. Thus, the *gain of positive charge* by the volume element $\Delta v'$ during the polarization process is

$$-\oint_{\Delta S} \rho_0^+ \delta^+ \cdot \mathbf{n} da' \quad (4-17)$$

where ΔS is the surface bounding $\Delta v'$. Similarly, the displacement of negative charge increases the charge (decreases the negative charge) in $\Delta v'$ by

$$\oint_{\Delta S} (-\rho_0^-) \delta^- \cdot \mathbf{n} da' \quad (4-18)$$

The total gain in charge by the volume element $\Delta v'$ is the sum of Eqs. (4-17) and (4-18), and as a consequence of Eq. (4-16) may be written as

$$-\oint_{\Delta S} \rho_0^+ (\delta^+ - \delta^-) \cdot \mathbf{n} da' = -\nabla \cdot [\rho_0^+ (\delta^+ - \delta^-)] \Delta v' \quad (4-19)$$

But $\delta^+ - \delta^-$ is just the relative displacement of positive and negative charge densities, and $\rho_0^+ (\delta^+ - \delta^-)$ is equivalent, therefore, to what we have called the polarization \mathbf{P} . Thus, $\rho_P \Delta v'$ is the net charge in a volume element of the polarized dielectric.

At first sight, it may seem rather strange that having started with electrically neutral volume elements of dielectric material, we end up with volume elements that bear a net charge. According to our original point of view, the dielectric is composed of elemental dipoles $\Delta \mathbf{p}$, and it was essential that each $\Delta \mathbf{p}$ be electrically neutral in order that Eq. (4-5) give the potential correctly. Now we find that so long as $\nabla \cdot \mathbf{P}$ does not vanish, the individual volume elements appear to be charged. The origin of this seeming paradox is found in the mathematical transformation Eq. (4-11). The contribution from each volume element is transformed to a different volume term and a surface term. The total charge in the volume and surface of the element is still zero, but when we stack various volume elements together to form a macroscopic piece of dielectric material, we find that the contributions to the potential from the various "internal surfaces" cancel out. We are left with effectively charged volume elements and a surface contribution from the actual boundary of the dielectric body.

The total polarization charge of a dielectric body

$$Q_P = \int_{V_0} (-\nabla' \cdot \mathbf{P}) dv' + \oint_{S_0} \mathbf{P} \cdot \mathbf{n} da' \quad (4-20)$$

must equal zero, since it was our premise that the dielectric, as a whole, is electrically neutral. This result is immediately obvious from the form of Eq. (4-20), which clearly vanishes as a consequence of the divergence theorem.

We now have two distinct expressions for the electrostatic potential $\varphi(\mathbf{r})$ due to a polarized dielectric specimen, namely, Eqs. (4-7) and (4-15). Both are correct, but we shall find the latter expression more convenient in most cases. The electric field \mathbf{E} may be obtained as minus the gradient of Eq. (4-15). Since φ is a function of the coordinates (x, y, z) , the appropriate gradient is $-\nabla$. The unprimed coordinates appear only in the function $1/|\mathbf{r} - \mathbf{r}'|$. Hence, noting that $\nabla(1/|\mathbf{r} - \mathbf{r}'|) = -\nabla'(1/|\mathbf{r} - \mathbf{r}'|)$ and using Eq. (4-8), we obtain

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \left[\int_{S_0} \sigma_P \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} da' + \int_{V_0} \rho_P \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} dv' \right] \quad (4-21)$$

THE ELECTRIC FIELD INSIDE A DIELECTRIC

Before we can write an expression for the electric field inside a polarized medium, it is necessary to define this electric field precisely. What we are interested in is the *macroscopic electric field*—that is, the average electric field in a small region of the dielectric that nevertheless contains a large number of molecules. An alternative and perhaps preferable approach is to define the electric field directly in terms of a macroscopic experiment as follows.

The (*macroscopic*) *electric field* is the force per unit charge on a test charge embedded in the dielectric, in the limit where the test charge is so small that it does not itself affect the charge distribution.

This test charge must be dimensionally small from the macroscopic point of view (what we shall call a “point” charge), but it will be large compared with the size of a molecule.

Although the above statement is the fundamental definition of the macroscopic electric field \mathbf{E} , it is difficult to use this definition directly to obtain an expression for the field, since we would have to calculate the force on a charged body of extended size and then go to the limit as the size of the object decreased. Hence, we find it expedient to use another property of the electric field to help us obtain the analytic expression we are seeking. In this way, we shall get \mathbf{E} in terms of the polarization charges of the medium.

FIGURE 4-3

The path $ABCD$ lies partly in the needle-shaped cavity and partly in the dielectric. In an isotropic dielectric (see Section 4-5), the polarization \mathbf{P} has the direction of \mathbf{E} , so that, for the orientation of the needle shown, $\sigma_p = 0$ on the cylindrical walls. In an anisotropic dielectric, σ_p is not necessarily zero, but its value does not affect the longitudinal component of electric field in the cavity.



Later, in Section 4-9 it will be shown that the quantity we have called \mathbf{E} is indeed in agreement with the fundamental “force definition.”

The electrostatic field in a dielectric must have the same basic properties that we found applied to \mathbf{E} in vacuum; in particular, \mathbf{E} is a conservative field and, hence, derivable from a scalar potential. Thus,

$$\nabla \times \mathbf{E} = 0$$

or, equivalently,

$$\oint \mathbf{E} \cdot d\mathbf{l} = 0$$

Let us apply the last equation to the path $ABCD$ shown in Fig. 4-3, where the segment AB lies in a needle-shaped cavity cut out of the dielectric and the segment CD lies in the dielectric proper. Since the segments AD and BC may be made arbitrarily small, the line integral reduces to

$$\mathbf{E}_v \cdot \mathbf{l} - \mathbf{E}_d \cdot \mathbf{l} = 0$$

or, equivalently,

$$E_{vt} = E_{dt} \quad (4-22)$$

where the subscripts v and d refer to vacuum and dielectric, respectively, and the subscript t stands for tangential component.

Equation (4-22) is valid regardless of the orientation of the needle-shaped cavity. If the “needle” is oriented along the direction of \mathbf{E} , then $E_{dt} = E_d$. Furthermore, by symmetry, the field in the cavity is along the

direction of the needle—that is, $E_{vt} = E_v$. We are thus led to the following important conclusion.*

The *electric field in a dielectric* is equal to the electric field inside a needle-shaped cavity in the dielectric, with the cavity axis oriented parallel to the direction of the electric field.

Evidently, the problem of calculating the electric field inside a dielectric reduces to calculating the electric field inside a needle-shaped cavity in the dielectric. But the electric field in the cavity is an external field and hence may be determined by means of the results of Section 4-2. Just as in Section 4-2, we assume here that the polarization of the dielectric is a given function $\mathbf{P}(x', y', z')$, and we calculate the potential and electric field arising from this polarization. Taking the field point \mathbf{r} at the center of the cavity and using Eq. (4-15), we obtain for the potential

$$\begin{aligned} \varphi(\mathbf{r}) = & \frac{1}{4\pi\epsilon_0} \int_{V_0-V_1} \frac{\rho_P(x', y', z') dv'}{|\mathbf{r} - \mathbf{r}'|} \\ & + \frac{1}{4\pi\epsilon_0} \int_{S_0+S'} \frac{\sigma_P(x', y', z') da'}{|\mathbf{r} - \mathbf{r}'|} \end{aligned} \quad (4-23)$$

where $V_0 - V_1$ is the volume of the dielectric excluding the “needle,” S_0 is the exterior surface of the dielectric, and $S' = S_1 + S_2 + S_c$ are the needle surfaces. But from Fig. 4-3 it is seen that $\sigma_P = 0$ on the cylindrical surface S_c of the needle. Furthermore, the needle may be made arbitrarily thin so that the surfaces S_1 and S_2 have negligible area. Thus, only the exterior surfaces of the dielectric contribute, and the surface integral of Eq. (4-23) becomes identical in form to the surface integral of Eq. (4-15). The volume integral of Eq. (4-23) excludes the cavity. However, the contribution of the cavity to this integral is negligible, as may readily be seen. The charge density ρ_P is bounded; the quantity $dv'/|\mathbf{r} - \mathbf{r}'|$ does not diverge at the field point (i.e., when $\mathbf{r}' = \mathbf{r}$) because the volume of a point is a higher-order zero than the $\lim |\mathbf{r} - \mathbf{r}'|$; and finally volume V_1 of the needle may be made arbitrarily small by making the cavity thin. Thus, we need not exclude the volume V_1 , and Eq. (4-23) becomes similar in form to Eq. (4-15). In other words, Eq. (4-15) gives the potential $\varphi(\mathbf{r})$ regardless of whether the point \mathbf{r} is located inside or outside the dielectric.

* This statement is strictly true only for isotropic dielectrics (see Section 4-5). For anisotropic dielectrics, the symmetry argument fails, and our conclusion must be generalized: The electric field in a dielectric is equal to the longitudinal component of the electric field inside a needle-shaped cavity in the dielectric provided the cavity axis is oriented parallel to the direction of the electric field in the dielectric.

The electric field $\mathbf{E}(\mathbf{r})$ may be calculated as minus the gradient of Eq. (4-23). But this field differs only by a negligible amount from Eq. (4-21). Thus, Eq. (4-21) gives the medium's contribution to the electric field at \mathbf{r} , independently of whether \mathbf{r} is inside or outside the medium.

The calculations indicated in Eqs. (4-15) and (4-21) are straightforward for cases in which $\mathbf{P}(x, y, z)$ is a known function of position. (Some examples of this type are to be found among the problems at the end of this chapter.) In most cases, however, the polarization arises in response to an electric field that has been imposed on the dielectric medium—that is, $\mathbf{P}(x', y', z')$ is a function of the *total macroscopic electric field* $\mathbf{E}(x', y', z')$ —and under these conditions, the situation is much more complicated. First, while it is necessary to know the functional form of $\mathbf{P}(\mathbf{E})$, this form is known experimentally in most cases and, hence, not a source of difficulty. The real complication arises because \mathbf{P} depends on the *total electric field*, including the contribution from the dielectric itself, and it is this contribution that we are in the process of evaluating. Thus, we cannot determine \mathbf{P} because we don't know \mathbf{E} , and conversely.

It is evident that a different approach to the problem is needed, and one will be provided in the following sections.

4-4

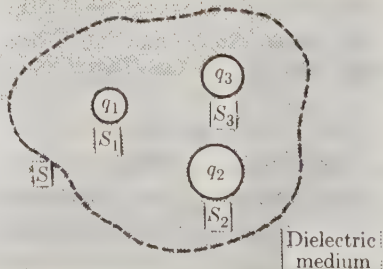
GAUSS'S LAW IN A DIELECTRIC: THE ELECTRIC DISPLACEMENT

In Chapter 2, we derived an important relationship between electric flux and charge, namely, Gauss's law. This law states that the electric flux across an arbitrary closed surface is proportional to the total charge enclosed by the surface. In applying Gauss's law to a region containing charges embedded in a dielectric, we must be careful to include all of the charge inside the gaussian surface—that is, the polarization charge as well as the charge that we embedded.

In Fig. 4-4, the dashed surface S is an imaginary closed surface located inside a dielectric medium. We embed a certain amount of charge Q in the

FIGURE 4-4

Construction of a gaussian surface S in a dielectric medium



volume bounded by S , and we assume that this charge exists on the surfaces of three conductors in amounts q_1 , q_2 , and q_3 . By Gauss's law,

$$\oint_S \mathbf{E} \cdot \mathbf{n} \, da = \frac{1}{\epsilon_0} (Q + Q_P) \quad (4-24)$$

where Q is the net embedded charge,

$$Q = q_1 + q_2 + q_3$$

and Q_P is the net polarization charge,

$$Q_P = \int_{S_1+S_2+S_3} \mathbf{P} \cdot \mathbf{n} \, da + \int_V (-\nabla \cdot \mathbf{P}) \, dv \quad (4-25a)$$

Here V is volume of the dielectric enclosed by S . There is no boundary of the dielectric material at S , so that the surface integral in Eq. (4-25a) *does not* contain a contribution from S .

If we transform the volume integral in Eq. (4-25a) to a surface integral by means of the divergence theorem, we must be careful to include contributions from all surfaces bounding V , namely, S , S_1 , S_2 , and S_3 . It is evident that that last three contributions will cancel the first term of Eq. (4-25a), so that

$$Q_P = -\oint_S \mathbf{P} \cdot \mathbf{n} \, da \quad (4-25b)$$

Combining this result with (4-24), we obtain

$$\oint_S (\epsilon_0 \mathbf{E} + \mathbf{P}) \cdot \mathbf{n} \, da = Q \quad (4-26)$$

Equation (4-26) states that the flux of the vector $\epsilon_0 \mathbf{E} + \mathbf{P}$ through a closed surface is equal to the net charge that we embedded in the volume enclosed by the surface. This vector quantity is important enough to warrant a name and a separate symbol.

We define, therefore, a new macroscopic field vector \mathbf{D} , the *electric displacement*:

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \quad (4-27)$$

which evidently has the same units as \mathbf{P} , charge per unit area.*

In terms of \mathbf{D} , Eq. (4-26) becomes

$$\oint_S \mathbf{D} \cdot \mathbf{n} \, da = Q \quad (4-28)$$

* In gaussian units, \mathbf{D} is defined as $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$; \mathbf{D} , \mathbf{E} , and \mathbf{P} all have the same units of charge per unit area and in vacuum $\mathbf{D} = \mathbf{E}$.

This result is usually referred to as *Gauss's law* for the electrical displacement, or simply Gauss's law.

Equation (4-28) is applicable to a region of space bounded by any closed surface S ; if we apply it to a small region in which the charge enclosed is distributed as a charge density ρ , then Gauss's law becomes

$$\oint_S \mathbf{D} \cdot \mathbf{n} \, da = \rho \Delta V$$

Dividing this equation by ΔV and proceeding to the limit, we obtain

$$\nabla \cdot \mathbf{D} = \rho \quad (4-29)$$

This result is sometimes called the differential form of Gauss's law.

The advantage of expressing the integral and differential forms of Gauss's law, Eqs. (4-28) and (4-29), in terms of the vector \mathbf{D} is that only the charge Q or the charge density ρ that we embedded in the dielectric medium appears explicitly. Henceforth we shall usually call the charge Q (or ρ) simply the charge (or the charge density). When it is necessary to distinguish it from the polarization charge Q_P of the medium or the *total* charge $Q + Q_P$, the charge Q will be called the *external charge*. By "external" we do not mean that it necessarily lies outside the physical boundary of the piece of material. We mean that it is in addition to the charges that make up the atomic constitution of the neutral material.* Since in many problems the external charges are given, it is an advantage that the total electrostatic field at each point in the dielectric medium is expressed as the sum of two parts

$$\mathbf{E}(x, y, z) = \frac{1}{\epsilon_0} \mathbf{D}(x, y, z) - \frac{1}{\epsilon_0} \mathbf{P}(x, y, z) \quad (4-30)$$

where the first term, $(1/\epsilon_0)\mathbf{D}$, is related to the external charge density

*The external charge is very often called the "free" charge, and polarization charge is sometimes used synonymously with bound charge. In electrostatics, this confusion does not cause too much trouble because the external charge on a conductor is free (i.e., free to move around) and the polarization charge in a dielectric is bound. The external charge in a dielectric is not free, however—if it were it would quickly move to the surface and leak off. Also, most conducting media contain, in addition to the free charges that determine the electrostatic behavior of a conducting object, some bound charges, which in other circumstances contribute to polarization (Chapter 7). With time-dependent fields (Chapter 19), it is very important not to confound the distinction between external and polarization charges with that between free and bound charges. Thus, we shall exclusively use the term *external charge* in this context.

through its divergence, and the second term, $(-1/\epsilon_0)\mathbf{P}$, is proportional to the polarization of the medium. In a vacuum, the electric field is given entirely by the first term in Eq. (4-30).

ELECTRIC SUSCEPTIBILITY AND DIELECTRIC CONSTANT

In the introduction to this chapter, it was stated that the polarization of a dielectric medium occurs in response to the electric field in the medium. The degree of polarization depends not only on the electric field, but also on the properties of the molecules that make up the dielectric material. From the macroscopic point of view, the behavior of the material is completely specified by an experimentally determined relationship, called a *constitutive equation*, $\mathbf{P} = \mathbf{P}(\mathbf{E})$, where \mathbf{E} is the macroscopic electric field. This expression is a point relationship, and if \mathbf{E} varies from point to point in the material, then \mathbf{P} will vary accordingly.

For most materials, \mathbf{P} vanishes when \mathbf{E} vanishes, which is the usual behavior. Consequently, we shall limit our discussion here to materials of this type. (Dielectrics with a permanent polarization will be discussed briefly in Section 5-4.) Furthermore, if the material is isotropic, then the polarization should have the same direction as the electric field that is causing it.

These results are summarized by the constitutive equation

$$\mathbf{P} = \chi(E)\mathbf{E} \quad (4-31)$$

where the scalar quantity $\chi(E)$ is called the *electric susceptibility* of the material.

A great many materials are electrically isotropic; this category includes fluids, polycrystalline and amorphous solids, and some crystals. A treatment of the electrical properties of anisotropic materials is beyond the scope of this text.

Combining Eq. (4-31) with Eq. (4-27), we obtain an expression for \mathbf{D} in isotropic media:

$$\mathbf{D} = \epsilon(E)\mathbf{E} \quad (4-32)$$

$$\epsilon(E) = \epsilon_0 + \chi(E) \quad (4-33)$$

where $\epsilon(E)$ is the *permittivity* of the material. It is evident that ϵ , ϵ_0 , and χ all have the same units.

Although we have been careful to write χ and ϵ in the form $\chi(E)$ and $\epsilon(E)$, nevertheless it is found experimentally that χ and ϵ are *frequently* independent of the electric field, except perhaps for very intense fields.

In other words, χ and ϵ are frequently constants characteristic of the material. Materials of this type will be called *linear dielectrics*, and they obey the relations

$$\mathbf{P} = \chi \mathbf{E} \quad (4-31a)$$

$$\mathbf{D} = \epsilon \mathbf{E} \quad (4-32a)$$

The electrical behavior of a material is now completely specified by either the permittivity ϵ or the susceptibility χ . It is more convenient, however, to work with a dimensionless quantity K defined by

$$\epsilon = K\epsilon_0 \quad (4-34)$$

K is called the dielectric coefficient, or simply the *dielectric constant*. From Eq. (4-33) it is evident that

$$K = \frac{\epsilon}{\epsilon_0} = 1 + \frac{\chi}{\epsilon_0} \quad (4-35)$$

The dielectric constants for some commonly encountered materials are given in Table 4-1. Except for a few examples in which the polarization \mathbf{P} of the material is specified, the problems in this book deal with linear dielectrics.

If the electric field in a dielectric is made very intense, it will begin to pull electrons completely out of the molecules, and the material will become conducting. The maximum electric field which a dielectric can withstand without breakdown is called its *dielectric strength*. The dielectric strengths, E_{\max} , of a few substances are also given in Table 4-1.

4-6

POINT CHARGE IN A DIELECTRIC FLUID

One of the simplest problems involving a dielectric that we might consider is that of a point charge q in a homogeneous isotropic medium of infinite extent. The dielectric medium will be assumed to be linear and characterized by a dielectric constant K . Although this problem is quite simple, it will nevertheless prove instructive.

EXAMPLE 4-1 Electric Field of a Point Charge in a Dielectric Fluid

Let the point charge q be located at the origin in a dielectric fluid of dielectric constant K . Find the electric field \mathbf{E} in the fluid.

TABLE 4-1
Properties of Dielectric
Materials

Material	Dielectric constant K	Dielectric strength E_{\max} (V/m)
Aluminum oxide	4.5	6×10^6
Glass*	5–10	9×10^6
Nylon	3.5	19×10^6
Polyethylene	2.3	18×10^6
Quartz (SiO ₂)	4.3	
Sodium chloride	6.1	
Sulfur	4.0	
Wood*	2.5–8.0	
Alcohol, ethyl (0°C)	28.4	
Benzene (0°C)	2.3	
Water (distilled, 0°C)	87.8	
Water (distilled, 20°C)	80.1	
Air (1 atm)	1.00059	3×10^6
Air (100 atm)	1.0548	
CO ₂ (1 atm)	1.000985	

* For materials such as glass and wood, the chemical composition varies; hence, the range of dielectric constants.

It is not to be inferred that the material is nonlinear.

Source: Data from the *Handbook of Chemistry and Physics*, 58th ed. (Cleveland, OH: CRC Press, Inc., 1978).

Solution: If the point charge q were situated in a vacuum, the electric field would be a pure radial field. But since \mathbf{E} , \mathbf{D} , and \mathbf{P} are all parallel to one another at each point, the radial nature of the field is not changed by the presence of the medium. Furthermore, from the symmetry of the problem, \mathbf{E} , \mathbf{D} , and \mathbf{P} can depend only on the distance from the point charge, not on any angular coordinate. Let us apply Gauss's law, Eq. (4-28), to a spherical surface of radius r that is located concentrically about q . Then

$$4\pi r^2 D = q$$

and

$$D = \frac{q}{4\pi r^2}$$

or

$$\mathbf{D} = \frac{q}{4\pi r^3} \mathbf{r} \quad (4-36)$$

The electric field and polarization may now be evaluated quite easily:

$$\mathbf{E} = \frac{q}{4\pi K \epsilon_0 r^3} \mathbf{r} \quad (4-37)$$

$$\mathbf{P} = \frac{(K-1)q}{4\pi K r^3} \mathbf{r} \quad (4-38)$$

Thus, the electric field is smaller by the factor K than would be the case if the medium were absent.

At this point, it will be instructive to look at the problem in more detail and try to see why the dielectric has weakened the electric field. The electric field has its origin in all of the charge—both external and polarization charge. The external charge is just the point charge q . The polarization charge, however, is made up from two contributions: a volume density $\rho_P = -\nabla \cdot \mathbf{P}$ and a surface density $\sigma_P = \mathbf{P} \cdot \mathbf{n}$ on the surface of the dielectric in contact with the point charge. Using Eq. (4-38), we find that $\nabla \cdot \mathbf{P}$ vanishes for $r \neq 0$, so there is no volume density of polarization charge in this case.

Our point charge q is a point in the macroscopic sense. Suppose it is large on a molecular scale and we can assign to it a radius b , which eventually will be made to approach zero. The total surface polarization charge is then given by

$$Q_P = \lim_{b \rightarrow 0} 4\pi b^2 (\mathbf{P} \cdot \mathbf{n})_{r=b} = -\frac{(K-1)q}{K} \quad (4-39)$$

The total charge

$$Q_P + q = \frac{1}{K} q \quad (4-40)$$

appears as a point charge from the macroscopic point of view, and it is now clear why the electric field is a factor K smaller than it would be if the medium were absent. A schematic diagram of the *point charge* q in a dielectric medium is shown in Fig. 4-5.

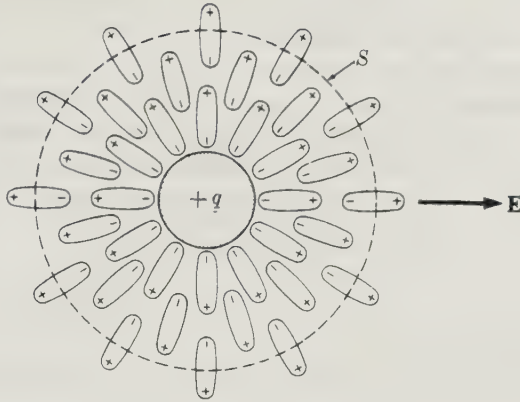
4-7 BOUNDARY CONDITIONS ON THE FIELD VECTORS

Before we can solve more complicated problems, we must know how the field vectors \mathbf{E} and \mathbf{D} change in passing an interface between two media. The two media may be two dielectrics with different properties, or a dielectric and a conductor. Vacuum may be treated as a dielectric with permittivity ϵ_0 .

Consider two media, 1 and 2, in contact as shown in Fig. 4-6. We shall assume that there is a surface density of external charge, σ , which may vary from point to point on the interface. Let us construct the small pillbox-shaped surface S that intersects the interface and encloses an area ΔS of the

FIGURE 4-5

Schematic diagram showing the orientation of polarized molecules in a dielectric medium surrounding a "point charge" q



interface, the height of the pillbox being negligibly small in comparison with the diameter of the bases. The charge enclosed by S is

$$\sigma \Delta S + \frac{1}{2}(\rho_1 + \rho_2) \times \text{volume}$$

but the volume of the pillbox is negligibly small, so that the last term may be neglected. Applying Gauss's law to S , we find

$$\mathbf{D}_2 \cdot \mathbf{n}_2 \Delta S + \mathbf{D}_1 \cdot \mathbf{n}_1 \Delta S = \sigma \Delta S$$

or

$$(\mathbf{D}_2 - \mathbf{D}_1) \cdot \mathbf{n}_2 = \sigma \quad (4-41a)$$

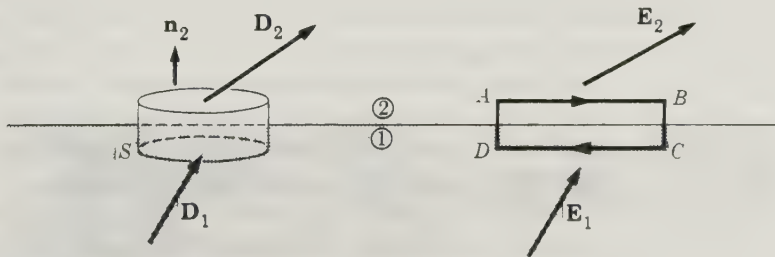
Since \mathbf{n}_2 may serve as the normal to the interface

$$D_{2n} - D_{1n} = \sigma \quad (4-41b)$$

Thus, the discontinuity in the normal component of \mathbf{D} is given by the surface density of external charge on the interface. Or, to put it another

FIGURE 4-6

Boundary conditions on the field vectors at the interface between two media, which may be obtained by applying Gauss's law to S and integrating $\mathbf{E} \cdot d\mathbf{l}$ around the path $ABCD$



way, if there is no charge on the interface between two media, the normal component of \mathbf{D} is continuous.

Because the electrostatic field \mathbf{E} may be obtained as minus the gradient of a potential, the line integral of $\mathbf{E} \cdot d\mathbf{l}$ around any closed path vanishes. Let us apply this result to the rectangular path $ABCD$ of Fig. 4-6. On this path, the lengths AB and CD will be taken equal to Δl and the segments AD and BC will be assumed to be negligibly small. Therefore,

$$\mathbf{E}_2 \cdot \Delta \mathbf{l} + \mathbf{E}_1 \cdot (-\Delta \mathbf{l}) = 0$$

or

$$(\mathbf{E}_2 - \mathbf{E}_1) \cdot \Delta \mathbf{l} = 0 \quad (4-42a)$$

Hence, the desired result is

$$E_{2t} = E_{1t} \quad (4-42b)$$

Thus, the tangential component of the electric field is continuous across an interface.

The preceding results have been obtained for two arbitrary dielectric media, but it is interesting to see what the equations would predict if one of the media were a conductor. Since there is no molecular restoring force on the free charges of a conductor, it would seem that we would have $\chi = \infty$ for a conductor in Eq. (4-31), and $\epsilon = \infty$ according to Eq. (4-33). If medium 1 is taken as the conductor, then we conclude that $\mathbf{E}_1 = 0$ whatever the (finite) values of \mathbf{P}_1 and \mathbf{D}_1 , as we have already inferred by different reasoning in Chapter 2. Since \mathbf{E}_1 vanishes, Eq. (4-42b) becomes

$$E_{2t} = 0 \quad (4-43)$$

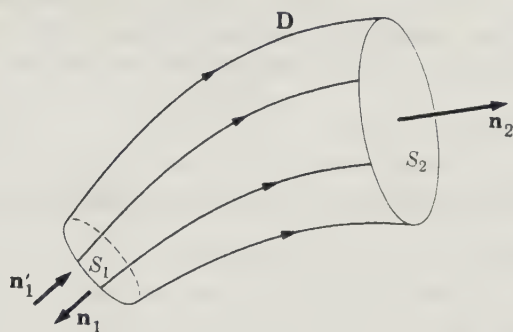
The displacement \mathbf{D}_1 is not determined by these considerations, however. If for our present purposes we arbitrarily take it to be zero, Eq. (4-41b) becomes

$$D_{2n} = \sigma \quad (4-44)$$

where σ then represents the *total* surface charge density on the conductor but does not include the polarization surface charge on the dielectric. An alternative approach is to solve the dielectric problem and then let K_1 go to infinity. (See Problems 4-12 and 4-14.) Then σ represents polarization charge of the conductor (plus any external charge on the conductor). Physically the result is the same. Equation (4-44) also follows from Eq. (4-28) by the same argument that led to Eq. (2-31). (See Fig. 2-10.) Note that the field in the dielectric is always perpendicular to the surface of the conductor according to Eq. (4-43).

It is evident on purely physical grounds that the potential φ must be continuous across an interface. Continuity of φ follows because the difference in potential, $\Delta\varphi$, between two closely spaced points is $-\mathbf{E} \cdot \Delta \mathbf{l}$, where $\Delta \mathbf{l}$ is the separation of the two points, and from what has been said

FIGURE 4-7
A tube of displacement
flux



above there is no reason to expect \mathbf{E} to become infinite at an interface. Actually, the continuity of the potential is a boundary condition, but not independent of those already derived. It is equivalent in most cases to Eq. (4-42b).

From the discussion given here and in preceding sections, it may be inferred that the electric displacement \mathbf{D} is closely related to external charge. We should now like to prove an important property of \mathbf{D} , namely, that the flux of \mathbf{D} is continuous in regions containing no external charge. To do so, we again resort to Gauss's law. Let us focus our attention on a region of space and construct *lines of displacement*, which are imaginary lines drawn in such a way that the direction of a line at any point is the direction of \mathbf{D} at that point. Next we imagine a tube of displacement, a volume bounded on the sides by lines of \mathbf{D} but not cut by them (Fig. 4-7). The tube is terminated at its ends by the surfaces S_1 and S_2 . Applying Gauss's law, we obtain

$$\int_{S_2} \mathbf{D} \cdot \mathbf{n} \, da - \int_{S_1} \mathbf{D} \cdot \mathbf{n}' \, da = Q \quad (4-45)$$

If there is no external charge in the region, then $Q = 0$, and the same amount of flux enters the tube through S_1 as leaves through S_2 . When external charge is present, it determines the discontinuity in displacement flux. Thus, lines of displacement terminate on external charges. The lines of force, on the other hand, terminate on either external or polarization charges.

BOUNDARY-VALUE PROBLEMS INVOLVING DIELECTRICS

The fundamental equation that has been developed in this chapter is

$$\nabla \cdot \mathbf{D} = \rho \quad (4-46)$$

where ρ is the external charge density. If the dielectrics with which we are concerned are linear, isotropic, and homogeneous, then $\mathbf{D} = \epsilon \mathbf{E}$, where ϵ is a constant characteristic of the material, and we may write

$$\nabla \cdot \mathbf{E} = \frac{1}{\epsilon} \rho \quad (4-47)$$

But the electrostatic field \mathbf{E} is derivable from a scalar potential φ :

$$\mathbf{E} = -\nabla\varphi$$

so that

$$\nabla^2\varphi = -\frac{1}{\epsilon} \rho \quad (4-48)$$

Thus, the potential in the dielectric satisfies Poisson's equation. The only difference between Eq. (4-48) and the corresponding equation for the potential in vacuum is that ϵ replaces ϵ_0 (and ρ means the external instead of the total charge density).

In most cases of interest, the dielectric contains no charge distributed throughout its volume—that is, $\rho = 0$ inside the dielectric material. The charge exists on the surfaces of conductors or is concentrated in the form of point charges, which may, to be sure, be embedded in the dielectric. In these circumstances, the potential satisfies Laplace's equation throughout the body of the dielectric:

$$\nabla^2\varphi = 0 \quad (4-49)$$

In some problems there may be a surface density of charge, σ , on the surface of a dielectric body or on the interface between two dielectric materials, but this condition does not alter the situation, and Eq. (4-49) still applies so long as $\rho = 0$.

An electrostatic problem involving linear, isotropic, and homogeneous dielectrics reduces, therefore, to finding solutions of Laplace's equation in each medium and joining the solutions in the various media by means of the boundary conditions of the preceding section. There are many problems that may be solved by this method. One example will be discussed here, and additional examples will be found in the problems at the end of the chapter.

EXAMPLE 4-2
Dielectric Sphere in a Uni-
form Electric Field

We should like to determine how the lines of force are modified when a dielectric sphere of radius a is placed in a region of space containing an *initially* uniform electric field \mathbf{E}_0 . Let us assume the dielectric to be linear, isotropic, and homogeneous and to be characterized by the dielectric constant K . Furthermore, it bears no charge. The origin of our coordinate system may be taken at the center of the sphere, and the direction of \mathbf{E}_0 as the polar direction (z -direction).

Solution: The potential may then be expressed as a sum of zonal harmonics. Just as in Section 3-5, all boundary conditions can be satisfied by means of the two lowest-order harmonics, and we write

$$\varphi_1(r, \theta) = A_1 r \cos \theta + C_1 r^{-2} \cos \theta \quad (4-50)$$

for the vacuum region (1) outside the sphere, and

$$\varphi_2(r, \theta) = A_2 r \cos \theta + C_2 r^{-2} \cos \theta \quad (4-51)$$

for the dielectric region (2). The constants A_1 , A_2 , C_1 , and C_2 are unknown and must be determined from the boundary conditions. The harmonic r^{-1} is not required, since its presence implies a net charge on the sphere. A constant term may be added to Eqs. (4-50) and (4-51), but since it turns out that the same constant is required in both equations, we may, without loss of generality, take it to be zero.

At distances far from the sphere, the electric field will retain its uniform character, and $\varphi_1 \rightarrow -E_0 r \cos \theta$. Hence $A_1 = -E_0$. Furthermore, unless $C_2 = 0$, the potential and associated electric field would become infinite at the center of the sphere, which would imply the existence of a point dipole at the center—that is, a dipole whose moment is not proportional to ΔV . No *macroscopic* point dipole exists here, for as has already been discussed in Section 4-3, the potential and macroscopic field do not become infinite in a dielectric devoid of point charges. Hence, $C_2 = 0$, and the remaining constants, A_2 and C_1 , may be obtained from the boundary conditions of Section 4-7.

Continuity of the potential across the interface between the dielectric and vacuum requires that $\varphi_1 = \varphi_2$ at $r = a$, or

$$-E_0 a + C_1 a^{-2} = A_2 a \quad (4-52)$$

Since the normal component of \mathbf{D} at the interface is $D_r = -\epsilon(\partial\varphi/\partial r)$, the continuity of D_r (there is no charge on the surface of the dielectric) requires that $D_{1r} = D_{2r}$, at $r = a$, or

$$E_0 + 2C_1 a^{-3} = -KA_2 \quad (4-53)$$

Continuity of E_t at $r = a$ is equivalent to Eq. (4-52).

Combining Eqs. (4-52) and (4-53), we obtain

$$A_2 = -\frac{3E_0}{K+2} \quad (4-54)$$

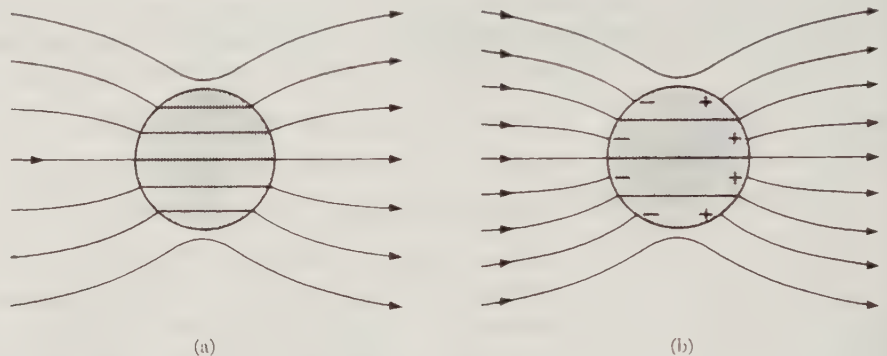
and

$$C_1 = \frac{(K-1)a^3 E_0}{K+2} \quad (4-55)$$

Thus, the problem has been solved. The potential is given by Eqs. (4-50) or (4-51), and the constants A_1 , C_1 , A_2 , and C_2 are all known. The components of \mathbf{E} and \mathbf{D} may be obtained at any point (r, θ, ϕ) by

FIGURE 4-8

A uniform electric field is distorted by the presence of a dielectric sphere: (a) lines of electric displacement, (b) lines of the electric field



differentiation. It is evident from Eq. (4-54) and because $C_2 = 0$ that the electric field inside the sphere has the direction of \mathbf{E}_0 and is given by

$$\mathbf{E}_2 = \frac{3}{K + 2} \mathbf{E}_0 \quad (4-56)$$

The lines of displacement and the lines of force are shown in Fig. 4-8.

***4-9**

FORCE ON A POINT CHARGE EMBEDDED IN A DIELECTRIC

We are now in a position to determine the force on a small, spherical, charged conductor embedded in a linear, isotropic dielectric. In the limit in which the conductor is negligibly small from the macroscopic viewpoint, this calculation gives the force on a point charge.

The electric field and surface charge density at a representative point of the conductor surface will be obtained by the boundary-value procedure of the preceding section, and the force \mathbf{F} may then be obtained from the integral over the surface:

$$\mathbf{F} = \oint_S \mathbf{E}' \sigma da \quad (4-57)$$

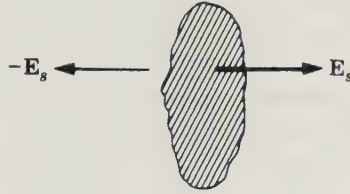
Here \mathbf{E}' stands for the electric field at the surface element da minus that part of the field produced by the element itself. In other words,

$$\mathbf{E}' = \mathbf{E} - \mathbf{E}_s \quad (4-58)$$

where \mathbf{E}_s is the electric field produced by the surface element of charge σda . It is important that \mathbf{E}_s not be included in the field \mathbf{E}' , because the quantity $\mathbf{E}_s \sigma da$ represents the interaction of the charge element σda with its own field. This self-interaction clearly produces no net force on the

FIGURE 4-9

Self-field of a charged surface element



element, but gives rise to a surface stress

$$\mathcal{F}_s = \sigma E_s \quad (4-59)$$

which is due to the mutual repulsion of the electrons (or of the excess positive ions) in the surface layer. This stress is balanced by strong cohesive forces in the material of which the element is composed. It should be pointed out that when calculating forces on charged objects in Chapters 2 and 3, we implicitly subtracted the self-field \mathbf{E}_s . Thus, when calculating the force on a point charge, the field produced by the point charge was not included. A further discussion of the forces on charged objects will be taken up in Section 6-8.

It may appear that the self-field of the charged surface element σda is negligible because the element is of infinitesimal size. This field, however, is finite. The element is small from the macroscopic point of view, to be sure, but one never quite goes to the limit. At a point directly on its surface, the element appears as a section of a plane (Fig. 4-9). The field \mathbf{E}_s is perpendicular to the element but oppositely directed on each side of the element. The total field inside the conductor is zero; hence,

$$\mathbf{E}' - \mathbf{E}_s = 0 \quad (\text{inside conductor}) \quad (4-60)$$

This equation provides us with a relationship between \mathbf{E}' and \mathbf{E}_s on the surface and, hence, between \mathbf{E}' and the macroscopic electric field \mathbf{E} at the surface:

$$\mathbf{E}' = \mathbf{E} - \mathbf{E}_s = \frac{1}{2} \mathbf{E} = \frac{\sigma}{2\epsilon} \mathbf{n} \quad (4-61)$$

The force on the conductor becomes

$$\mathbf{F} = \frac{1}{2} \int_s \mathbf{E} \sigma da \quad (4-57a)$$

Let us now fix our attention on a small spherical conductor embedded in a dielectric of infinite extent. The total charge on the conductor is Q , and its radius is a . Since we shall eventually go to the limit in which a becomes very small and since variations in the electric field (if they exist) are on a macroscopic scale, it is sufficient to consider the case in which the electric

field is initially uniform in the neighborhood of the conductor. Let us denote this uniform field by the symbol \mathbf{E}_0 . The picture is similar to that of the boundary-value problem we solved in Section 3-5, except that here the conducting sphere is embedded (or immersed) in a dielectric of permittivity ϵ , and in addition bears a net charge Q .

By analogy with Section 3-5 we easily determine the following.

The potential,

$$\varphi(r, \theta) = \varphi_0 - E_0 r \cos \theta + \frac{E_0 a^3}{r^2} \cos \theta + \frac{Q}{4\pi\epsilon r} \quad (4-62)$$

The electric field,

$$\begin{aligned} E_r &= E_0(1 + 2a^3/r^3) \cos \theta + Q/4\pi\epsilon r^2 \\ E_\theta &= -E_0(1 - a^3/r^3) \sin \theta \end{aligned} \quad (4-63)$$

The surface charge density on the surface of the sphere,

$$\sigma(\theta) = \epsilon E_r|_{r=a} = 3\epsilon E_0 \cos \theta + Q/4\pi a^2 \quad (4-64)$$

The force may now be determined from Eq. (4-57a). By symmetry, the only nonzero component of force is that in the direction $\theta = 0$, i.e., in the z -direction:

$$F_z = \frac{1}{2} \int_0^\pi (E_r)_{r=a} \cos \theta \sigma(\theta) 2\pi a^2 \sin \theta d\theta = E_0 Q \quad (4-65a)$$

or

$$\mathbf{F} = Q\mathbf{E}_0 \quad (4-65b)$$

This result is unchanged as we go to the limit of small a . Thus, the electric field in the dielectric \mathbf{E}_0 is in agreement with the fundamental definition, namely, the force on a small test charge Q divided by the magnitude of Q .

4-10

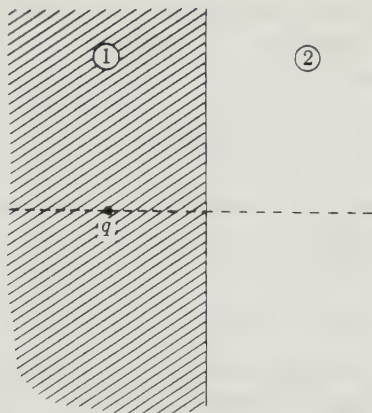
METHOD OF IMAGES FOR PROBLEMS INVOLVING DIELECTRICS

The method of images, discussed in the previous chapter, may be generalized to obtain solutions to Laplace's equation for the case where there are two or more dielectric media in the problem. Recall that when employing the image-charge technique, the potential in a region of space is envisioned as being produced by "point" charges and charged conductors in the region and by fictitious image charges outside the region. The image charges appear to be located *inside* of one or more conductors.

The only difference when we have two or more dielectrics present is that the fictitious image charges may also be located inside one of the other dielectrics (not the one where the potential is being calculated), and the boundary conditions at each dielectric-dielectric interface must be satisfied.

FIGURE 4-10

Two dielectric media with a point charge in medium 1

**EXAMPLE 4-3**

Field of a Point Charge Near a Dielectric-Dielectric Interface

Consider two dielectric media with constant permittivities, ϵ_1 and ϵ_2 , that are separated by a plane interface. There is no external charge on the interface. A point charge q is embedded in the medium characterized by ϵ_1 at a distance d from the interface (Fig. 4-10). The problem posed: What are the electric fields in media 1 and 2?

Solution: For convenience, we may take the yz -plane through the origin to be the interface, and we locate q on the x -axis at $x = -d$. If we use the problem described in Section 3-9 as a guide, we might expect the potential in each medium to be expressible as the sum of the potential of the point charge q and that of an image charge. Furthermore, we surmise the image charge to be located at the same distance d from the interface as the original point charge. Let

$$r = \sqrt{(x + d)^2 + y^2 + z^2} \quad \text{and} \quad r' = \sqrt{(x - d)^2 + y^2 + z^2}$$

Then in medium 1,

$$\varphi_1 = \frac{1}{4\pi\epsilon_1} \left[\frac{q}{r} + \frac{q'}{r'} \right] \quad (4-66)$$

In other words, the image charge q' is located in medium 2 at position $(d, 0, 0)$.

For the potential φ_2 in medium 2, the image charge must be located in medium 1, as is also the original charge q , both at position $(-d, 0, 0)$. We denote the *sum* of the charges by q'' . Then in medium 2,

$$\varphi_2 = \frac{q''}{4\pi\epsilon_2 r} \quad (4-67)$$

The quantities q' and q'' can be obtained from the boundary conditions at the interface. We calculate $D_{1n} = -\epsilon_1 \partial\varphi_1/\partial x|_{x=0}$, and so forth, and evaluate $D_{1n} = D_{2n}$ according to Eq. (4-41b):

$$\frac{(q - q')d}{[d^2 + y^2 + z^2]^{3/2}} = \frac{q''d}{[d^2 + y^2 + z^2]^{3/2}} \quad (4-68)$$

Next we evaluate $E_{1y} = -\partial\varphi_1/\partial y|_{x=0}$, and so on and then evaluate $E_{1y} = E_{2y}$ according to Eq. (4-42b):

$$\frac{(q + q')y}{\epsilon_1[d^2 + y^2 + z^2]^{3/2}} = \frac{q''y}{\epsilon_2[d^2 + y^2 + z^2]^{3/2}} \quad (4-69)$$

Nothing new is obtained from the continuity of φ at the interface. Solving Eqs. (4-68) and (4-69) for q' and q'' , we obtain

$$q' = \frac{\epsilon_1 - \epsilon_2}{\epsilon_1 + \epsilon_2} q, \quad q'' = \frac{2\epsilon_2}{\epsilon_1 + \epsilon_2} q \quad (4-70)$$

Laplace's equation is satisfied in both media and the boundary conditions are satisfied. The solution is unique.

4-11

SUMMARY

The electrostatic behavior of a dielectric medium is completely characterized by its dipole moment per unit volume or **polarization**:

$$\mathbf{P} = \lim \frac{\Delta\mathbf{p}}{\Delta v}$$

This procedure produces the **polarization charge density**

$$\rho_P = -\nabla \cdot \mathbf{P}, \quad (\sigma_P = \mathbf{n} \cdot \mathbf{P})$$

giving rise to the potential

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \left[\int \frac{\rho_P(\mathbf{r}') dv'}{|\mathbf{r} - \mathbf{r}'|} + \oint \frac{\sigma_P(\mathbf{r}') da'}{|\mathbf{r} - \mathbf{r}'|} \right]$$

The total E -field due to external charges plus the polarization charge satisfies

$$\nabla \cdot \mathbf{E} = \frac{1}{\epsilon_0} (\rho + \rho_P)$$

It is convenient to define the vector field, **electric displacement**,

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}$$

such that

$$\nabla \cdot \mathbf{D} = \rho$$

with only the external charges as sources. The curl equation

$$\nabla \times \mathbf{E} = 0$$

is unchanged because it does not contain the charge density. In order to solve the equations for the fields, the **constitutive equation**

$$\mathbf{P} = \mathbf{P}(\mathbf{E})$$

must also be known for the particular material. Then the last four equations, subject to the **boundary conditions**

$$D_{2n} - D_{1n} = \sigma \quad \text{and} \quad E_{2t} - E_{1t} = 0$$

are sufficient to determine \mathbf{E} and \mathbf{D} inside and outside dielectrics.

- The integral form of **Gauss's law** becomes

$$\oint_S \mathbf{D} \cdot \mathbf{n} \, da = Q$$

where Q includes only the *external* charge located inside the surface S . (Consequently, just outside the surface of a conductor embedded in a dielectric medium, $D_n = \sigma$.) The curl equation still allows the definition of the potential by

$$\mathbf{E} = -\nabla\varphi$$

- Most dielectric materials are linear, with constant **susceptibility** χ :

$$\mathbf{P} = \chi\mathbf{E}$$

This **constitutive equation** combined with the definition of \mathbf{D} gives

$$\mathbf{D} = \epsilon\mathbf{E}$$

where

$$\epsilon = \epsilon_0 + \chi$$

The **dielectric constant**

$$K = \frac{\epsilon}{\epsilon_0}$$

is between 1 and 100 for most common dielectrics; for all dielectrics, $K \geq 1$ ($\chi \geq 0$). For vacuum, $K = 1$ ($\chi = 0$). The electrostatic behavior of a conductor can be obtained by letting K be infinite.

- In a linear medium,

$$\nabla \cdot \mathbf{E} = \frac{1}{\epsilon} \rho \quad \text{and} \quad \nabla^2 \varphi = -\frac{\rho}{\epsilon}$$

The mathematical techniques for solving Poisson's and Laplace's equations are similar to those of Chapter 3, with the appropriate boundary conditions

at a dielectric interface,

$$K_2 \frac{\partial \varphi_2}{\partial n} = K_1 \frac{\partial \varphi_1}{\partial n} \quad \text{and} \quad \varphi_2 = \varphi_1$$

(with the last equation equivalent to $E_{2t} = E_{1t}$).

PROBLEMS

4-1 A thin dielectric rod of cross section A extends along the x -axis from $x = 0$ to $x = L$. The polarization of the rod is along its length and is given by $P_x = ax^2 + b$. (a) Find the volume density of polarization charge and the surface polarization charge on each end. (b) Show *explicitly* that the total polarization charge vanishes in this case.

4-2 A dielectric cube of side L has a radial polarization given by $\mathbf{P} = A\mathbf{r}$, where A is a constant and $\mathbf{r} = \mathbf{i}x + \mathbf{j}y + \mathbf{k}z$. The origin of coordinates is at the center of the cube. (a) Find all polarization charge densities. (b) Show *explicitly* that the total polarization charge vanishes.

4-3 A dielectric rod in the shape of a right circular cylinder of length L and radius R is polarized in the direction of its length. If the polarization is uniform and of magnitude P , calculate the electric field resulting from this polarization at a point on the axis of the rod.

4-4 Prove the following relationship between the polarization \mathbf{P} and the polarization charge densities ρ_p and σ_p for a dielectric specimen of volume V and surface S :

$$\int_V \mathbf{P} \, dv = \int_V \rho_p \mathbf{r} \, dv + \int_S \sigma_p \mathbf{r} \, da$$

Here, $\mathbf{r} = \mathbf{i}x + \mathbf{j}y + \mathbf{k}z$ is the position vector from any fixed origin. [*Hint*: Expand $\nabla \cdot (x\mathbf{P})$ according to Eq. (4-10).]

4-5 Two semi-infinite blocks of dielectric are placed almost in contact so that there exists a narrow gap of constant separation between them. The polarization \mathbf{P} is constant throughout all of the dielectric material, and it makes the angle γ with the normal to the planes bounding the gap. Determine the electric field in the gap.

4-6 A long cylindrical conductor of radius a , bearing the charge λ per unit length, is immersed in a dielectric medium of constant permittivity ϵ . Find the electric field at distance $r > a$ from the axis of the cylinder.

4-7 Two dielectric media with dielectric constants K_1 and K_2 are separated by a plane interface. There is no external charge on the interface. Find a relationship between the angles θ_1 and θ_2 , where these are the angles that an arbitrary line of displacement makes with the normal to the interface: θ_1 in medium 1, θ_2 in medium 2.

4-8 A coaxial cable of circular cross section has a compound dielectric. The inner conductor has an outside radius a , which is surrounded by a dielectric sheath of dielectric constant K_1 and of outer radius b . Next comes another dielectric sheath of dielectric constant K_2 and outer radius c . The outer conducting shell has an inner radius c . If a potential difference φ_0 is imposed between the conductors, calculate the polarization at each point in the two dielectric media.

4-9 A spherical conductor of radius R_1 is surrounded by a solid dielectric spherical

shell of outer radius R_2 , inner radius R_1 and dielectric constant K_1 . This composite entity is immersed in a fluid of dielectric constant K_2 and is subjected to an initially uniform electric field \mathbf{E}_0 . Determine the electric field in the two dielectric media.

4-10 A long dielectric cylinder of radius a and dielectric constant K is placed in a uniform electric field \mathbf{E}_0 . The axis of the cylinder is oriented at right angles to the direction of \mathbf{E}_0 . The cylinder contains no external charge. Determine the electric field at points inside and outside the cylinder.

4-11 A point dipole \mathbf{p} is placed at the center of a solid dielectric sphere of radius a and dielectric constant K . Find the electric field at points inside and outside the sphere. [*Hint*: The field outside is a dipole field; inside another term is needed in addition to a dipole field.]

4-12 Show that if K goes to infinity in the solution for a dielectric sphere in an initially uniform field as derived in Section 4-8, the result is the same as that for a conducting sphere derived in Section 3-5.

4-13 A plane slab of material with dielectric constant K_1 is bounded on both sides by material of dielectric constant K_2 . The electric field in medium 2, \mathbf{E}_2 , is given to be uniform and perpendicular to the boundaries. Find (a) the field \mathbf{E}_1 , (b) the polarization \mathbf{P}_1 , and (c) the polarization charge in medium 1.

4-14 Show that if K_1 goes to infinity in Problem 4-13, the result agrees physically with Eq. (4-44).

4-15 Two parallel conducting plates are separated by the distance d and maintained at the potential difference $\Delta\varphi$. A dielectric slab, consisting of dielectric constant K and of uniform thickness $t < d$, is inserted between the plates. Determine the field vectors \mathbf{E} and \mathbf{D} in the dielectric and also in the vacuum between dielectric and one plate. Neglect edge effects due to the finite size of the plates.

4-16 Two parallel conducting plates are separated by the distance d and maintained at the potential difference $\Delta\varphi$. A dielectric slab, of dielectric constant K and of uniform thickness d , is inserted snugly between the plates. However, the slab does not completely fill the volume between the plates. (a) Find the electric field (i) in the dielectric and (ii) in the vacuum region between the plates. (b) Find the charge density σ on that part of the plate (i) in contact with the dielectric and (ii) in contact with vacuum. (c) Find σ_p on the surface of the dielectric slab.

4-17 A conducting sphere of radius R floats half submerged in a liquid dielectric medium of permittivity ϵ_1 . The region above the liquid is a gas of permittivity ϵ_2 . The total charge on the sphere is Q . (a) Find a radial inverse-square electric field satisfying all boundary conditions, and determine the free, bound, and total charge densities at all points on the surface of the sphere. (b) Formulate an argument to show that this electric field is the actual one.

4-18 A uniform electric field \mathbf{E}_0 is set up in a medium of dielectric constant K . Prove that the field inside a spherical cavity in the medium is

$$\mathbf{E} = \frac{3K\mathbf{E}_0}{2K + 1}$$

***4-19** A dielectric sphere of radius R has a permanent polarization \mathbf{P} , which is uniform in direction and magnitude. The polarized sphere gives rise to an electric field. Determine this field (a) inside the sphere and (b) outside the sphere. Inside the

sphere the electric field, which is in the opposite direction to the polarization, is called a depolarizing field. [*Hint*: Since $\nabla \cdot \mathbf{P}$ vanishes at all points, the electrostatic potential satisfies Laplace's equation both inside and outside the sphere. Do *not* assume that the dielectric is characterized by a dielectric constant.]

4-20 In the text, it was shown that the polarization $\mathbf{P} = \rho_0^+(\delta^+ - \delta^-)$. Use this relation for the uniformly polarized sphere of Problem 4-19 to determine the external dipole field directly.

Microscopic Theory of Dielectrics

In the preceding chapter, we were concerned with the macroscopic aspects of dielectric polarization. It was shown how the polarization could, in many cases, be taken into account through the introduction of a dielectric constant. In this way, the electric field could be computed directly from a consideration of the external charge distribution. Although reference was made to the molecules of the dielectric several times in Chapter 4, a microscopic treatment of the material was not carried through in detail, and the overall picture that was presented was certainly from the macroscopic point of view. In this chapter, we examine the molecular nature of the dielectric and see how the electric field responsible for polarizing the molecule is related to the macroscopic electric field. Further, on the basis of a simple molecular model, it is possible to understand the linear behavior that is characteristic of a large class of dielectric materials.

MOLECULAR FIELD IN A DIELECTRIC

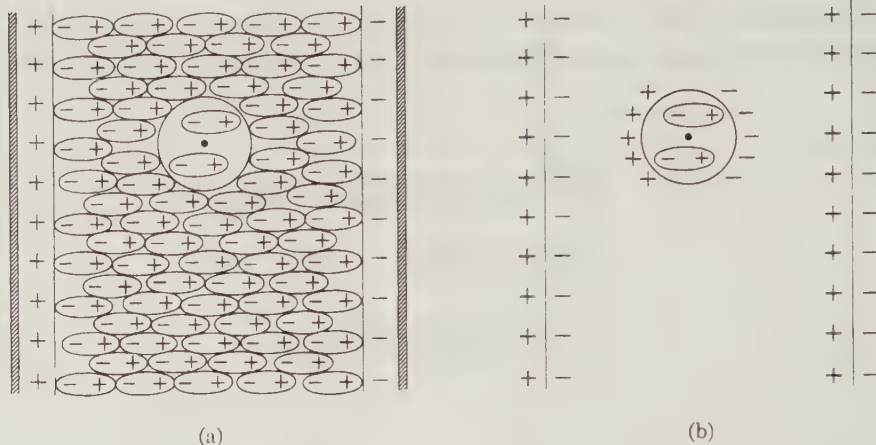
The electric field that is responsible for polarizing a molecule of the dielectric is called the molecular field, \mathbf{E}_m .

The *molecular field* is the electric field at a molecular position in the dielectric, and it is produced by all external sources and by all polarized molecules in the dielectric *with the exception* of the one molecule at the point under consideration.

It is evident that \mathbf{E}_m need not be the same as the macroscopic electric field because, as was discussed in Section 4-3, the latter quantity is related to the

FIGURE 5-1

Replacement of the dielectric outside the "cavity" by a system of polarization charges



force on a test charge that is large in comparison with molecular dimensions.

The molecular field may be calculated in the following way. First we cut out a small piece of the dielectric, leaving a spherical cavity surrounding the point at which the molecular field is to be computed. The dielectric that is left will be treated as a continuum—that is, from the macroscopic point of view. Now we put the dielectric back into the cavity, molecule by molecule, except for the molecule at the center of the cavity where we wish to compute the molecular field. The molecules that have just been replaced are to be treated, not as a continuum, but as individual dipoles. The procedure just outlined can be justified only if the result of the calculation is independent of the size of the cavity, which, as we shall see, is indeed the case under certain conditions.

Let us suppose that a thin dielectric sample has been polarized by placing it in the uniform electric field between two parallel plates that are oppositely charged, as shown in Fig. 5-1(a). It will be assumed that the polarization is uniform on a macroscopic scale (i.e., $\nabla \cdot \mathbf{P} = 0$) and that \mathbf{P} is parallel to the field producing it. The part of the dielectric outside the cavity may be replaced by a system of polarization charges as shown in Fig. 5-1(b), whence the electric field at the center of the cavity may be written as

$$\mathbf{E}_m = \mathbf{E}_x + \mathbf{E}_d + \mathbf{E}_s + \mathbf{E}' \quad (5-1)$$

Here \mathbf{E}_x is the primary electric field due to the charged parallel plates, \mathbf{E}_d is the depolarizing field due to polarization charge on the outside surfaces of the dielectric, \mathbf{E}_s is due to polarization charge on the cavity surface S , and \mathbf{E}' is due to all of the dipoles inside of S . Although we are not concerned with the explicit form of \mathbf{E}_x , it is evident that if the dimensions of the plates are large compared with their separation, $E_x = (1/\epsilon_0)\sigma$, where σ is the

surface charge density. The depolarizing field is also produced by two parallel planes of charge, this time with the density σ_P . Since $\sigma_P = P_n = \pm P$,

$$\mathbf{E}_d = -\frac{1}{\epsilon_0} \mathbf{P} \quad (5-2)$$

Let us write the macroscopic electric field *in the dielectric* without a subscript—that is, \mathbf{E} . Since the normal component of the electric displacement \mathbf{D} is continuous across the vacuum-dielectric interface, and since $\mathbf{D} = \epsilon_0 \mathbf{E}_x$ in the vacuum just outside the dielectric slab,

$$\mathbf{E} = \mathbf{E}_x + \mathbf{E}_d \quad (5-3)$$

Combining Eqs. (5-1), (5-2), and (5-3) yields

$$\mathbf{E}_m = \mathbf{E} + \mathbf{E}_s + \mathbf{E}' \quad (5-4)$$

This equation relates the molecular field to the macroscopic electric field in the dielectric material. The result is quite general and not restricted to the spherical-cavity and plane-electrode geometry of Fig. 5-1. Nevertheless, this simple derivation is instructive and will be useful in Section 5-4.

The field \mathbf{E}_s arises from polarization charge density, $\sigma_P = P_n$, on the spherical surface S . Using spherical coordinates and taking the polar direction along the direction of \mathbf{P} , as in Fig. 5-2, we obtain

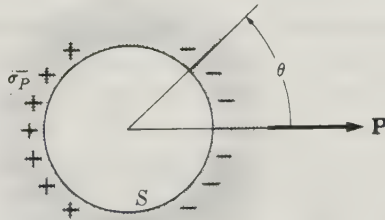
$$d\mathbf{E}_s = \frac{(-P \cos \theta)}{4\pi\epsilon_0 r^3} \mathbf{r} da \quad (5-5)$$

where \mathbf{r} is the vector from the surface to the center of the sphere. From symmetry, it is evident that only the component of $d\mathbf{E}_s$ along the direction of \mathbf{P} will contribute to the integral of Eq. (5-5) over the complete surface. Since $da = r^2 \sin \theta d\theta d\phi$,

$$\mathbf{E}_s = \frac{1}{4\pi\epsilon_0} \mathbf{P} \int_0^{2\pi} d\phi \int_0^\pi \cos^2 \theta \sin \theta d\theta = \frac{1}{3\epsilon_0} \mathbf{P} \quad (5-6)$$

FIGURE 5-2

Calculation of the "cavity" surface contribution to \mathbf{E}_m



Finally, we come to the last term in Eq. (5-4), that due to the electric dipoles inside S . There are several important cases for which this term vanishes. If there are a great many dipoles in the cavity, if they are oriented parallel but randomly distributed in position, and if there are no correlations between the positions of the dipoles, then $\mathbf{E}' = 0$.^{*} This situation might, for example, prevail in a gas or a liquid. Similarly, if the dipoles in the cavity are located at the regular atomic positions of a cubic crystal, then again $\mathbf{E}' = 0$. In this connection, the reader is referred to Problem 5-2.

In the general case, \mathbf{E}' is not zero, and if the material contains several species of molecule, \mathbf{E}' may differ at the various molecular positions. It is this term that gives rise to the anisotropic electrical behavior of calcite, for example. Since it is not our purpose to develop a theory of anisotropic material, we restrict further discussion to the rather large class of materials in which $\mathbf{E}' = 0$. Thus, Eq. (5-4) reduces to

$$\mathbf{E}_m = \mathbf{E} + \frac{1}{3\epsilon_0} \mathbf{P} \quad (5-7)$$

It is interesting to note that this result would be obtained directly by the above method if the spherical cavity were created by removing just one molecule. But under these conditions, the cavity would be so small that the replacement of the rest of the dielectric by a system of polarization charges could not be justified.

The dipole moment of a molecule is proportional to the electric field (the polarizing field) acting on it.

The ratio of the dipole moment of a molecule to the polarizing field is called the *molecular polarizability*, α . In other words,[†]

$$\mathbf{p}_m = \alpha \mathbf{E}_m \quad (5-8)$$

If there are N molecules per unit volume, then the polarization $\mathbf{P} = N\mathbf{p}_m$. Combining this result with Eqs. (5-7) and (5-8), we obtain

$$\mathbf{P} = N\alpha \left(\mathbf{E} + \frac{1}{3\epsilon_0} \mathbf{P} \right) \quad (5-9)$$

This equation may be rewritten in terms of the dielectric constant, K , since $\mathbf{P} = \chi \mathbf{E} = (K - 1)\epsilon_0 \mathbf{E}$.

^{*} See W. Panofsky and M. Phillips, *Classical Electricity and Magnetism*, 2nd ed. (Reading, Mass.: Addison-Wesley, 1962), Chap. 2.

[†] The concept of polarizability can also be used to describe polar molecules, as we shall see in Section 5-3.

In this way, Eq. (5-9) becomes

$$\alpha = \frac{3\epsilon_0(K-1)}{N(K+2)} \quad (5-10)$$

which is known as the *Clausius-Mossotti equation*.

It is evident that Eq. (5-10) defines a molecular property, namely, the molecular polarizability, in terms of quantities that can be determined on a macroscopic basis.

5-2
INDUCED DIPOLES: A SIMPLE MODEL

The molecules of a dielectric may be classified as polar or nonpolar. A *polar molecule* is one that has a permanent dipole moment, even in the absence of a polarizing field \mathbf{E}_m . In the next section, the response of a polar dielectric to an external electric field will be studied. Here we deal with the somewhat simpler problem involving *nonpolar molecules*, in which the "centers of gravity" of the positive and negative charge distributions normally coincide. Symmetrical molecules such as H_2 , N_2 , and O_2 and monatomic molecules such as He, Ne, and Ar fall into this category.

The application of an electric field causes a relative displacement of the positive and negative charges in nonpolar molecules, and the molecular dipoles so created are called *induced dipoles*. The simplest type of molecule that can be envisioned is composed of a single atom. It is possible to construct a simple classical model for the atom and from this model derive an expression for the induced dipole moment, and hence for its polarizability. Although specifically designed to treat monatomic molecules, the model may be used for symmetrical diatomic molecules by applying it separately to each of the atoms in the molecule to obtain the atomic polarizabilities. The molecular polarizability is then the sum of these, or twice the atomic polarizability.

An atom consists of an extremely small positively charged nucleus surrounded by orbital electrons that are in a state of continual motion. Since the electrons traverse their orbits in an exceedingly short time, of the order of 10^{-15} second, it is evident that in the equivalent "static" atom each electronic charge is "smeared" over its orbit. Quantum mechanics tells us that although this picture is essentially correct, it is somewhat naive; the electrons are not really localized on orbits, but have a finite probability of being situated in any part of the atom. Thus, the response of the atom to an electrostatic field or to slowly varying electric fields may be treated by considering the electron to be distributed over its orbit in the atom and each orbit to be smeared over a substantial part of the atomic volume. In short, a simple classical model of the atom consistent with this picture is a point positive charge (the nucleus) surrounded by a spherically symmetric cloud

of negative charge in which the density is essentially uniform out to the atomic radius R_0 and is zero at larger radii.

We are now in a position to compute the polarizability of this "atom." The nucleus will be assigned the charge Ze , where e is the absolute value of the electronic charge and Z is the atomic number. Since the atom is electrically neutral, the total charge in the electron cloud is $-Ze$. If the atom is placed in a polarizing field \mathbf{E}_m , the nucleus will be displaced *relative* to the center of the charge cloud by a distance that we shall call x . This displacement will be in the direction of \mathbf{E}_m . We shall assume that the charge cloud moves rigidly during this displacement—that is, there is no distortion of the cloud by the polarizing field. The displacement x may be determined from the equilibrium of forces on the nucleus: The force ZeE_m acts in the direction of the field, whereas an electrostatic force between the nucleus and charge cloud tends to restore the initial configuration. By Gauss's law, the negative charge attracting the nucleus is that part of the cloud within the sphere of radius x , and if the electronic density in the cloud is uniform, then this charge is Zex^3/R_0^3 . Hence,

$$\frac{(Ze)(Zex^3/R_0^3)}{4\pi\epsilon_0x^2} = ZeE_m \quad (5-11)$$

or

$$Zex = 4\pi\epsilon_0R_0^3E_m \quad (5-12)$$

Since the atomic dipole created in this process is $\mathbf{p}_m = Zex$, the last equation may be compared with Eq. (5-8), whence

$$\alpha = 4\pi\epsilon_0R_0^3 \quad (5-13)$$

The atomic model just described may be tested by comparing results obtained from it with results derived from other sources. For example, Eq. (5-13) may be combined with the Clausius-Mossotti equation (5-10) to eliminate α . The resulting equation predicts the atomic radius R_0 in terms of experimentally determined quantities. R_0 obtained in this way agrees reasonably well with results from other experiments in those cases for which the model is particularly suited. R_0 is of the order of magnitude of 1 angstrom unit—that is, 10^{-10} m. (See Problem 5-1.)

The polarizability derived in Eq. (5-13) is a constant, independent of the polarizing field. Hence, Eq. (5-13) leads to a constant value of K , and the dielectric so described is linear.

POLAR MOLECULES: THE LANGEVIN-DEBYE FORMULA

As mentioned in the preceding section, a polar molecule has a permanent dipole moment. A polar molecule consists of at least two different species of

FIGURE 5-3

A random distribution of permanent dipoles



atoms. During molecule formation, some of the electrons may be completely or partially transferred from one atomic species to the other, with the resulting electronic arrangement being such that positive and negative charge centers do not coincide in the molecule. In the absence of an electric field, a macroscopic piece of the polar dielectric is not polarized, since the individual dipoles are randomly oriented, as shown in Fig. 5-3. The polarization has been defined as

$$\mathbf{P} = \frac{1}{\Delta v} \sum \mathbf{p}_m \quad (5-14)$$

where the summation extends over all molecules in the volume element Δv . When the \mathbf{p}_m are oriented at random, the summation vanishes.

If the polar dielectric is subjected to an electric field, the individual dipoles experience torques that tend to align them with the field. If the field is strong enough, the dipoles may be completely aligned, and the polarization achieves the saturation value

$$\mathbf{P}_s = N\mathbf{p}_m \quad (5-15)$$

where N is the number of molecules per unit volume. This orientation effect is in addition to the induced dipole effects, which are usually present also. For the moment, we shall ignore the induced dipole contribution, but its effect will be added in later.

At field strengths normally encountered, the polarization of a polar dielectric is usually far from its saturation value, and if the temperature of the specimen is raised the polarization becomes even smaller. The lack of complete dipole alignment is due to the thermal energy of the molecules, which tends to produce random dipole orientations. The average effective dipole moment per molecule may be calculated by means of a principle from statistical mechanics, which states that at temperature T the probability of finding a particular molecular energy E is proportional to the Boltzmann factor

$$e^{-E/kT} \quad (5-16)$$

where k is Boltzmann's constant and T is the absolute temperature. A

complete discussion of the basis for this principle will not be given here; the reader familiar with the Maxwell velocity distribution in a perfect gas has already encountered the principle. According to the Maxwell distribution law, the probability of a molecular velocity v is proportional to $e^{-mv^2/2kT}$. But in Maxwell's perfect gas, the molecules have only kinetic energy, $\frac{1}{2}mv^2$. In the general case, E in Eq. (5-16) must include both kinetic energy E_k and potential energy U , and the factor becomes

$$e^{-E_k/kT}e^{-U/kT} \quad (5-17)$$

The potential energy of a permanent dipole p_0 in an electric field \mathbf{E}_m is

$$U = -\mathbf{p}_0 \cdot \mathbf{E}_m = -p_0 E_m \cos \theta \quad (5-18)$$

where θ is the angle between \mathbf{p}_0 and the electric field. Since the molecular kinetic energies do not depend on the electric field, we can ignore the velocity distribution completely in the following calculation. The *effective* dipole moment of a molecular dipole is its component along the field direction—that is, $p_0 \cos \theta$. Using the Boltzmann relationship, we find the average value of this quantity to be

$$\langle p_0 \cos \theta \rangle = \frac{\int p_0 \cos \theta e^{+p_0 E_m \cos \theta / kT} d\Omega}{\int e^{+p_0 E_m \cos \theta / kT} d\Omega} \quad (5-19)$$

where $d\Omega$ is an element of solid angle that may be replaced by $2\pi \sin \theta d\theta$ and the limits on θ are 0 and π . Since p_0 , E_m , and kT are constants in the integration, the integrals may be readily performed. It is convenient to define

$$y = \frac{p_0 E_m}{kT} \quad (5-20)$$

Equation (5-19) then yields

$$\langle p_0 \cos \theta \rangle = p_0 \left[\coth y - \frac{1}{y} \right] \quad (5-21)$$

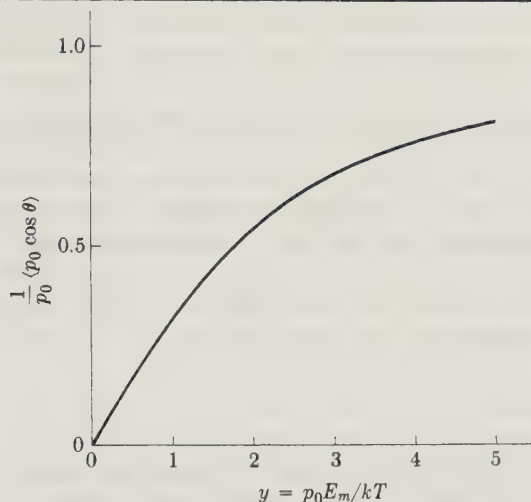
This equation is known as the *Langevin formula*. A plot of this function is given in Fig. 5-4.

It can be seen from the figure that Eq. (5-21) does indeed give a saturation effect at large field strengths. At small values of y , however, the curve is linear, and this linear region is important at ordinary temperatures. The molecular dipole moment p_0 of most polar materials is such that $y \ll 1$ for a full range of field strengths, even for those approaching the dielectric strength of the material, so long as the temperature is above about 250 K. Thus, a dielectric material containing polar molecules is, in general, *linear*.

Since it is the linear region of Eq. (5-21) that is important, it is

FIGURE 5-4

Plot of the Langevin function. The asymptotic value as $y \rightarrow \infty$ is one.



appropriate to expand $\coth y$ in a power series and keep only leading terms (see Problem 5-4). The first term cancels the last term in Eq. (5-21), with the result that

$$\langle p_0 \cos \theta \rangle \approx \frac{1}{3} p_0 y = \frac{p_0^2 E_m}{3kT} \quad (5-22a)$$

The term $\langle p_0 \cos \theta \rangle$ is the average effective dipole moment; therefore the polarization $P = N \langle p_0 \cos \theta \rangle$ and has the direction of \mathbf{E}_m . Hence, Eq. (5-22a) may be written in the form

$$\frac{1}{N} \mathbf{P} = \frac{p_0^2}{3kT} \mathbf{E}_m \quad (5-22b)$$

From a comparison of this equation with Eq. (5-8), it is evident that the polarizability α (i.e., the molecular dipole moment per unit polarizing field) is

$$\alpha = \frac{p_0^2}{3kT} \quad (5-23)$$

This result has been derived by neglecting induced dipole moments, and represents what we might call an *orientational* polarizability. Induced dipole effects, such as have been considered in the previous section, give rise to what might be termed a *deformation* polarizability, α_0 . In the general case, then, the total molecular polarizability is

$$\alpha = \alpha_0 + \frac{p_0^2}{3kT} \quad (5-24)$$

This expression is known as the *Langevin-Debye equation* and has been of great importance in interpreting molecular structures.

*5-4 PERMANENT POLARIZATION: FERROELECTRICITY

It was seen in Section 5-1 that it is the molecular field \mathbf{E}_m that is responsible for polarizing the individual molecules. The relationship between \mathbf{E}_m and the macroscopic electric field \mathbf{E} was given in Eq. (5-7). In most cases, the polarization is proportional to \mathbf{E} , so that \mathbf{E}_m vanishes when \mathbf{E} goes to zero. But under certain conditions, Eq. (5-7) is also compatible with a permanent (or spontaneous) polarization. When \mathbf{E} is set equal to zero,

$$\mathbf{E}_m = \frac{1}{3\epsilon_0} \mathbf{P}_0 \quad (5-25)$$

or, in words, if a polarization \mathbf{P}_0 exists, it will create an electric field at the molecule, which tends to polarize the molecule. To be sure, a polarizing field exists, but if this field gives rise to a polarization different from \mathbf{P}_0 , then the solution is not self-consistent. Therefore, if N is the number of molecules per unit volume,

$$\mathbf{P}_0 = N\alpha\mathbf{E}_m = \frac{N\alpha}{3\epsilon_0} \mathbf{P}_0 \quad (5-26)$$

This result is satisfied when either

$$\mathbf{P}_0 = 0$$

or

$$\frac{N\alpha}{3\epsilon_0} = 1 \quad (5-27)$$

Thus, the condition for a permanent polarization is Eq. (5-27).*

For most materials, $N\alpha/3\epsilon_0$ is less than one, and ordinary dielectric behavior results. In a few crystalline solids, however, condition Eq. (5-27) is met. Such materials are called *ferroelectric* because their electrical properties are analogous to the magnetic properties of ferromagnetic materials. The best known example of a ferroelectric material is barium titanate, BaTiO_3 , which exhibits a spontaneous dipole moment at temperatures below 120°C . This temperature is called the *Curie point* of the material.

* Strictly speaking, Eq. (5-27) has been derived for materials that are composed of only one species of molecule and for which the term \mathbf{E}' of Section 5-1 vanishes. In a quantitative theory applicable to the general case, Eq. (5-27) is replaced by a set of simultaneous equations. Such complications are not necessary for a fundamental understanding of the origin of ferroelectricity and, consequently, will not be discussed here.

The polarized state of a ferroelectric material is a relatively stable one and one that can persist for long periods of time. This statement may surprise us to some extent because a polarized specimen is subjected to its own depolarizing field and, depending on the geometry of the specimen, this depolarizing field may be rather large. The depolarizing field is largest for a specimen in the shape of a flat slab, polarized in a direction normal to its faces. As was seen in Section 5-1, if the dimensions of the slab face are large compared with the slab thickness, then

$$\mathbf{E}_d = -\frac{1}{\epsilon_0} \mathbf{P} \quad (5-28)$$

Actually, the high stability of a polarized ferroelectric is due to the fact that there is *no* depolarizing field on the specimen, even for the case of slab geometry. The specimen is polarized by placing it between parallel conducting plates that subsequently have a large potential difference applied to them. In this process, the *free* charge from the plates will, to a large extent, be neutralized by surface polarization charge, as is also the case during polarization of a conventional dielectric. If the parallel plates are now brought to the same potential by short-circuiting them, the polarized state of the ferroelectric is still energetically favorable, so the *free charge stays in place*, still neutralizing the polarization charge. The situation is something like that shown in Fig. 5-5, where the *free charge* is held in place by the surface polarization charge. The macroscopic field inside the ferroelectric is zero. Furthermore, the external electric field is zero, and it is difficult to distinguish the polarized specimen from a conventional unpolarized dielectric material.

If a large potential difference of the opposite sign is now applied to the plates surrounding the polarized ferroelectric, the specimen will change its polarization, and free charge of the opposite sign will flow to the plates from the external circuit, sufficient not only to neutralize the free charge already

FIGURE 5-5
A polarized piece of
ferroelectric material

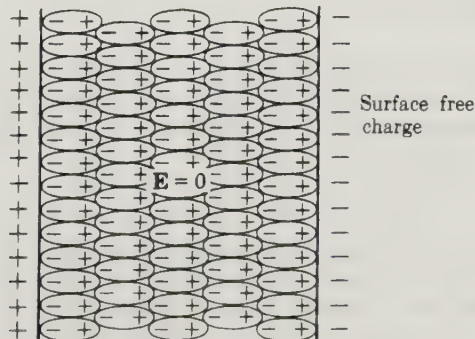
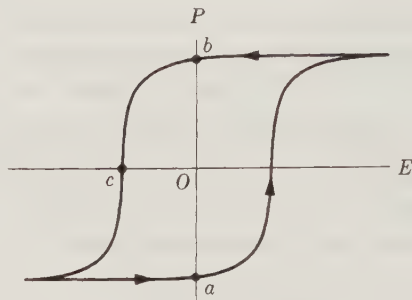


FIGURE 5-6
Hysteresis curve for a
ferroelectric specimen



there, but also to neutralize the new polarization charge. Thus, a ferroelectric slab between two parallel plates may serve as the basic element of a memory device. It is capable of storing charges on the two plates surrounding the ferroelectric as \pm or \mp , and its polarization persists in the absence of an external electric field. The “number” \pm or \mp may be read by applying a potential difference across the specimen. If the applied field is in the direction of the original polarization, no charge will pass through the external circuit: If the potential difference is opposite to the original polarization, a charge will flow through the external circuit as the polarization of the ferroelectric changes its direction.

A polarized ferroelectric is stable against a reversed electric field provided this electric field is not too large. Figure 5-6 shows the complete curve of polarization versus electric field. It is evident that for low fields there are two values of P for each value of E . A curve such as that in Fig. 5-6 is called a *hysteresis loop*. Hysteresis means “to lag behind,” and it is apparent that the polarization vector lags the electric field vector. Points b and a are the stable configurations at $E = 0$; they represent the polarizations \pm and \mp , respectively. Point c is the electric field that must be exceeded in order to reverse the polarization.

5-5

SUMMARY

The macroscopic polarization \mathbf{P} of an isotropic dielectric material depends on the molecular dipole moment (or its effective component) \mathbf{p}_m , which arises in response to the local electric field at the molecule—the **molecular field** \mathbf{E}_m :

$$\mathbf{P} = N\mathbf{p}_m$$

Usually \mathbf{p}_m is proportional to \mathbf{E}_m to a good approximation,

$$\mathbf{p}_m = \alpha\mathbf{E}_m$$

where α is the **molecular polarizability**. The molecular field depends on the

applied field \mathbf{E} and also on the polarization itself (i.e., the dipole fields of all the other molecules). In the simplest cases,

$$\mathbf{E}_m = \mathbf{E} + \frac{1}{3\epsilon_0} \mathbf{P}$$

In the usual case, \mathbf{E}_m can be eliminated from these equations to give

$$\mathbf{P} = \chi \mathbf{E}$$

with the constant susceptibility

$$\chi = \frac{N\alpha}{1 - \frac{N\alpha}{3\epsilon_0}}$$

But if the molecules are very polarizable ($N\alpha > 3\epsilon_0$), another solution is possible with $\mathbf{E} = 0$, $\mathbf{P} \neq 0$ (i.e., the material can be spontaneously polarized in zero applied field, as in a ferroelectric).

- All molecules exhibit an **induced dipole moment** in an electric field due to the deformation of the electronic charge distribution. A simple linear model leads to a constant atomic polarizability proportional to the atomic volume,

$$\alpha_0 = 4\pi\epsilon_0 R_0^3$$

- *Polar molecules*, which have a permanent dipole moment p_0 , display an additional **orientational polarizability** that is described by the Langevin function derived by statistical mechanics. At high temperature T , this contribution is also linear, with

$$\alpha = \frac{p_0^2}{3kT}$$

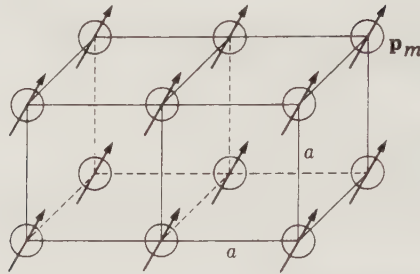
- The **Langevin-Debye equation** combines the orientational and deformation polarizabilities.

- A few materials, for example barium titanate, exhibit **ferroelectricity**.

PROBLEMS

5-1 (a) Use the Clausius-Mossotti equation to determine the polarizability of *atoms* in the air molecules: N_2 , O_2 . (Note that only the weighted average of the polarizabilities for nitrogen and oxygen may be obtained from Eq. (5-10).) (b) Combine this result with the theory of Section 5-2 to determine the average radius of the atom in an air molecule.

5-2 Figure 5-7 shows a simple cubic lattice of molecules all of which have the same (in direction and magnitude) dipole moment \mathbf{p}_m . Let us fix our attention on one particular molecule, call it j . It is evident that j has six nearest neighbors at distance a , twelve next-nearest neighbors at distance $\sqrt{2}a$, etc. Find the electric field at j due to the six \mathbf{p}_m on nearest-neighbor molecules for an arbitrary orientation of \mathbf{p}_m . (Let

**FIGURE 5-7**

Part of a simple cubic array of molecules, each with dipole moment \mathbf{p}_m

the lines joining j to its nearest neighbors define the x -, y -, and z -axes. For simplicity, take \mathbf{p}_m in the xz -plane, making an angle θ with the x -axis.)

5-3 (a) Using the result of Problem 5-1 for the atomic polarizability of nitrogen, compute the relative displacement of the nitrogen nucleus and electron cloud at a field strength $E_m = 3 \times 10^6$ V/m. (b) Compare this displacement with the radius of the atom found in Problem 5-1.

5-4 (a) By using the well-known series expansions for e^y , expand $\coth y$, and obtain Eq. (5-22a) from Eq. (5-21). (b) Go one step further and obtain another term in the series Eq. (5-22a).

5-5 Water is a polar molecule for which the Clausius-Mossotti equation is, strictly speaking, not applicable. Assume its validity, however, and determine p_0 for the water molecule.

Electrostatic Energy

Many problems in mechanics are greatly simplified by means of energy considerations. Hence, when the mechanical behavior of an electrical system is to be studied, it may prove advantageous to use energy methods. In general, the energy of a system of charges, just like that of any other mechanical system, may be divided into its potential and kinetic contributions. Under static conditions, however, the entire energy of the charge system exists as potential energy, and in this chapter we are particularly concerned with that potential energy that arises from electrical interaction of the charges, the so-called *electrostatic energy*.

In Section 2-4, it was shown that the electrostatic energy U of a point charge is closely related to the electrostatic potential φ at the position of the point charge. In fact, if q is the magnitude of a particular point charge, then the work done by the force on the charge when it moves from position A to position B is

$$\begin{aligned}\text{Work} &= \int_A^B \mathbf{F} \cdot d\mathbf{l} = q \int_A^B \mathbf{E} \cdot d\mathbf{l} \\ &= -q \int_A^B \nabla \varphi \cdot d\mathbf{l} = -q(\varphi_B - \varphi_A)\end{aligned}\quad (6-1)$$

Here the force \mathbf{F} has been assumed to be only the electric force $q\mathbf{E}$ at each point along the path. Under these conditions the charged particle would accelerate. If it does not accelerate, the electric force must be balanced at each point by an equal and opposite force applied by some other agency, so that the *total* work is zero and the kinetic energy does not change. The work done by this *other* force is

$$W = q(\varphi_B - \varphi_A)\quad (6-2)$$

This result is equal to the increase in electrostatic energy of the charge over the path interval $A \rightarrow B$.

Similar considerations may be applied to more complicated systems of charges. In fact, the electrostatic energy of an arbitrary charge distribution may be calculated as the work required to assemble this distribution of charge against the Coulomb interaction of the charges without imparting to it other forms of energy. This calculation and some of its consequences are the subject of this chapter.

6-1

POTENTIAL ENERGY OF A GROUP OF POINT CHARGES

By the electrostatic energy of a group of m point charges, we mean the potential energy of the system relative to the state in which all point charges are infinitely separated from one another. This energy may be obtained rather easily by calculating the work to assemble the charges, according to Eq. (6-2), bringing in one at a time. The first charge q_1 may be placed in position without any work, $W_1 = 0$. Placing the second charge q_2 requires

$$W_2 = \frac{q_2 q_1}{4\pi\epsilon_0 r_{21}}$$

where $r_{21} = |\mathbf{r}_2 - \mathbf{r}_1|$. For the third charge q_3 ,

$$W_3 = q_3 \left[\frac{q_1}{4\pi\epsilon_0 r_{31}} + \frac{q_2}{4\pi\epsilon_0 r_{32}} \right]$$

The work required to bring in the fourth charge, fifth charge, and so on, may be written in a similar fashion. The total electrostatic energy of the assembled m -charge system is the sum of the W 's, namely,

$$U = \sum_{j=1}^m W_j = \sum_{j=1}^m \left(\sum_{k=1}^{j-1} \frac{q_j q_k}{4\pi\epsilon_0 r_{jk}} \right) \quad (6-3)$$

Let us abbreviate this result for U as

$$U = \sum_{j=1}^m \sum_{k=1}^{j-1} W_{jk}$$

Now if we arrange the W_{jk} in the form of a matrix, noting that $W_{jk} = W_{kj}$ and putting $W_{jj} = 0$, it will be apparent that U can also be written as

$$U = \frac{1}{2} \sum_{j=1}^m \sum_{k=1}^m W_{jk} \quad (W_{jj} = 0)$$

That is, a factor of $\frac{1}{2}$ appears in this form of the summation, which is more symmetric, in order that the interaction between each pair of charges not be counted twice. Thus, an alternative and more convenient form of Eq. (6-3)

is

$$U = \frac{1}{2} \sum_{j=1}^m \sum_{k=1}^{m'} \frac{q_j q_k}{4\pi\epsilon_0 r_{jk}} \quad (6-4)$$

where the prime on the second summation means that the term $k = j$ is specifically excluded.

Equation (6-4) may be written in a somewhat different way by noting that the final value of the potential φ at the j th point charge due to the other charges of the system is

$$\varphi_j = \sum_{k=1}^{m'} \frac{q_k}{4\pi\epsilon_0 r_{jk}} \quad (6-5)$$

Thus, the electrostatic energy of the system is

$$U = \frac{1}{2} \sum_{j=1}^m q_j \varphi_j \quad (6-6)$$

If the point charges had been assembled in a linear dielectric medium of infinite extent, instead of in vacuum, then the permittivity ϵ would replace ϵ_0 in Eqs. (6-4) and (6-5), but Eq. (6-6) would remain unchanged. In the following section, it will be shown that this last equation has rather general validity. It applies to a group of point charges that are located in more than one dielectric medium, and it even applies to conductors of finite size. The only limitation on the validity of Eq. (6-6) is that all dielectrics in the electrical system be linear.

6-2 ELECTROSTATIC ENERGY OF A CHARGE DISTRIBUTION

In this section, we shall calculate the electrostatic energy of an arbitrary charge distribution with volume density ρ and surface density σ . Some of the charge may reside on the surfaces of conductors. In fact, it will be explicitly assumed that there are conductors in the system. In addition, it will be assumed that the dielectrics in the system are *linear*. This restriction is necessary in order that the work expended in bringing the system to its final charged state shall be independent of the way in which this final state is reached.

Suppose we assemble the charge distribution by bringing in charge increments δq from a reference potential $\varphi_A = 0$. If the charge distribution is partly assembled and the potential at a particular point in the system is $\varphi'(x, y, z)$, then, from Eq. (6-2), the work required to place δq at this point is

$$\delta W = \varphi'(x, y, z) \delta q \quad (6-7)$$

The charge increment δq may be added to a volume element located at

(x, y, z) , so that $\delta q = \delta\rho \Delta v$, or δq may be added to a surface element at the point in question, in which case $\delta q = \delta\sigma \Delta a$. The total electrostatic energy of the assembled charge distribution is obtained by summing contributions of the form Eq. (6-7).

Since the work required to assemble the charges is independent of the order in which things are done, we may choose a particular assembly procedure for which the summation of the δW 's is conveniently calculated. This procedure is one in which all parts of the system are brought to their final charge values in concert—that is, at any stage of the charging process, all charge densities will be at the same fraction of their final values. Let us call this fraction α . If the final values of the charge densities are given by the function $\rho(x, y, z)$ and $\sigma(x, y, z)$, then the charge densities at an arbitrary stage are $\alpha\rho(x, y, z)$ and $\alpha\sigma(x, y, z)$. Furthermore, the increments in these densities are $\delta\rho = \rho(x, y, z) d\alpha$ and $d\sigma = \sigma(x, y, z) d\alpha$. The total electrostatic energy, which is obtained by summing Eq. (6-7), is

$$U = \int_0^1 \delta d \int_V \rho(x, y, z) \varphi'(\alpha; x, y, z) dv \\ + \int_0^1 \delta d \int_S \sigma(x, y, z) \varphi'(\alpha; x, y, z) da$$

But since all charges are at the same fraction, α , of their final values, the potential $\varphi'(\alpha; x, y, z) = \alpha\varphi(x, y, z)$, where φ is the final value of the potential at (x, y, z) . Making this substitution, we find that the integration over α is readily done, and yields

$$U = \frac{1}{2} \int_V \rho(\mathbf{r}) \varphi(\mathbf{r}) dv + \frac{1}{2} \int_S \sigma(\mathbf{r}) \varphi(\mathbf{r}) da \quad (6-8)$$

This equation gives the desired result for the energy of a charge distribution. It is important to note that the volume of integration V must be large enough to include *all* of the charge density in the problem, and that the potential φ is just that due to the charge density ρ (and σ) itself. Normally the charge density vanishes outside some bounded region, in which case V may be taken to include all space. If all space is filled with a single dielectric medium except for certain conductors, the potential is given by

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon} \int_V \frac{\rho(\mathbf{r}') dv'}{|\mathbf{r} - \mathbf{r}'|} + \frac{1}{4\pi\epsilon} \int_S \frac{\sigma(\mathbf{r}') da'}{|\mathbf{r} - \mathbf{r}'|} \quad (6-9)$$

If several dielectrics are present, the proper boundary conditions must be satisfied, for example, by adding suitable solutions of Laplace's equation to Eq. (6-9). Equations (6-8) and (6-9) are generalizations of Eqs. (6-6) and (6-5) for point charges. The latter can be recovered as a special case by

letting

$$\rho(\mathbf{r}) = \sum_{j=1}^m q_j \delta(\mathbf{r} - \mathbf{r}_j)$$

$$\rho(\mathbf{r}') = \sum_{k=1}^{m'} q_k \delta(\mathbf{r}' - \mathbf{r}_k)$$

where the prime on the second summation indicates that the term $k = j$ must be excluded when the double sum is constructed. When ρ is a continuous distribution, the vanishing of the denominator in Eq. (6-9) does not cause the integral to diverge, and so it is unnecessary to exclude the point $\mathbf{r}' = \mathbf{r}$.

It was stipulated that conductors are present in the system. Although Eq. (6-8) covers this case very well, it is convenient to separate out the contribution from the conductors explicitly. The last integral involves, in part, integrations over the surface of these conductors. Since a conductor is an equipotential region, each of these integrations may be done:

$$\frac{1}{2} \int_{\text{conductor } j} \sigma \varphi da = \frac{1}{2} Q_j \varphi_j \quad (6-10)$$

where Q_j is the charge on the j th conductor.

Equation (6-8) for the *electrostatic energy of a charge distribution*, which includes conductors, becomes

$$U = \frac{1}{2} \int_V \rho \varphi dv + \frac{1}{2} \int_S \sigma \varphi da + \frac{1}{2} \sum_j Q_j \varphi_j \quad (6-11)$$

where the last summation is over all conductors, and the surface integral is restricted to nonconducting surfaces.

As we have seen in Chapter 3, in many problems of practical interest all of the charge resides on the surfaces of conductors. In these circumstances Eq. (6-11) reduces to

$$U = \frac{1}{2} \sum_j Q_j \varphi_j \quad (6-12)$$

We shall have occasion to develop this equation in a later section of this chapter.

For the present, we should like to compare Eq. (6-12) with Eq. (6-6), which was derived for an assembly of point charges. It appears at first sight that the two equations are identical. However, there is an important difference. Equation (6-12) was derived by starting with uncharged macroscopic conductors that were gradually charged by bringing in charge

increments. Thus, the energy described by Eq. (6-12) includes both interaction energy between different conductors and the self-energies of the charge on each individual conductor. If there is only a single conductor, its self-energy, $U = \frac{1}{2}Q_1\varphi_1$, is due to the interaction energy of the charges assembled on that conductor. In deriving Eq. (6-6), however, each point charge was brought in as a unit. Hence, the energy to assemble the point charge from *smaller* charge increments, the so-called self-energy of the point charge, is not included. An attempt to calculate it would give an infinite result if the charge is a mathematical point, but it is not included in the formulation of Coulomb's law of the force between point charges, and it should not be considered now. In order to see that Eq. (6-12) gives the same result in the limit where the conductors become very small, approaching "point" charges, the potential of the j th conductor may be written as the sum of two terms,

$$\varphi_j = \varphi_{j1} + \varphi_{j2} \quad (6-13)$$

where φ_{j1} is the contribution to the potential due to the charge on conductor j itself, and φ_{j2} is the contribution from charge on other conductors. Thus, Eq. (6-12) becomes

$$U = \frac{1}{2} \sum_j Q_j \varphi_{j1} + \frac{1}{2} \sum_j Q_j \varphi_{j2} \quad (6-14)$$

The first term of this equation represents the various self-energies of the conductors. Each self-energy, $\frac{1}{2}Q_j\varphi_{j1}$, depends on the environment of the conductor (since the charge distribution on each conductor adjusts itself to its environment). Furthermore, the only physically meaningful potential associated with conductor j is the total potential φ_j . Thus, the decomposition, Eq. (6-14), does *not* make a great deal of sense in general. However, if the conductors are so small that they may be treated as point charges from the macroscopic point of view, then redistribution of charge on the "point" cannot be important, and each self-energy may be taken to be independent of its environment. In addition, since by potential at the point charge j we mean φ_{j2} , the energy required to place a group of *previously charged, very small conductors* in position is the second summation in Eq. (6-14), which is equivalent to Eq. (6-6).

6-3 ENERGY DENSITY OF AN ELECTROSTATIC FIELD

In the preceding section, an expression was developed for the electrostatic energy of an arbitrary distribution of charge. This expression, Eq. (6-8), involves an explicit integration over the charge distribution. It is possible, however, to express the electrostatic energy of the system in a different way, and this alternate form is frequently rather useful. By means of a mathematical transformation (integration by parts), therefore, we convert Eq. (6-8) to an integral involving the field vectors \mathbf{E} and \mathbf{D} of the system.

We again consider an arbitrary distribution of charge characterized by the densities ρ and σ . For convenience, it will be assumed that the charge system is bounded—that is, it is possible to construct a closed surface, S' , of finite dimensions that encloses all of the charge. In addition, all surface densities of charge σ will be assumed to reside on conductor surfaces. The last statement is really no restriction at all, since a surface charge density on a dielectric-dielectric interface may be spread out slightly and then treated as a volume density ρ . The densities ρ and σ are related to the electric displacement:

$$\rho = \nabla \cdot \mathbf{D}$$

throughout the dielectric regions, and

$$\sigma = \mathbf{D} \cdot \mathbf{n}$$

on the conductor surfaces. Hence, Eq. (6-8) becomes

$$U = \frac{1}{2} \int_V \varphi \nabla \cdot \mathbf{D} \, dv + \frac{1}{2} \int_S \varphi \mathbf{D} \cdot \mathbf{n} \, da \quad (6-15)$$

The volume integral here refers to the region where $\nabla \cdot \mathbf{D}$ is different from zero, and this is the region external to the conductors. The surface integral is over the conductors.

The integrand in the first integral of Eq. (6-15) may be transformed by means of a vector identity that we have had occasion to use several times before, namely, Eq. (1-1-7) of Table 1-1:

$$\varphi \nabla \cdot \mathbf{D} = \nabla \cdot \varphi \mathbf{D} - \mathbf{D} \cdot \nabla \varphi$$

Of the two volume integrals resulting from this transformation, the first may be converted to a surface integral through the use of the divergence theorem. Finally, using the fact that $\mathbf{E} = -\nabla \varphi$, we may write Eq. (6-15) as

$$U = \frac{1}{2} \int_{S+S'} \varphi \mathbf{D} \cdot \mathbf{n}' \, da + \frac{1}{2} \int_V \mathbf{D} \cdot \mathbf{E} \, dv + \frac{1}{2} \int_S \varphi \mathbf{D} \cdot \mathbf{n} \, da \quad (6-16)$$

This equation may be simplified substantially. The surface $S + S'$ over which the first integral of Eq. (6-16) is to be evaluated is the entire surface bounding the volume V . It consists of S (the surfaces of all conductors in the system) and S' (a surface that bounds our system from the outside and that we may choose to locate at infinity). In both cases, the normal \mathbf{n}' is directed *out of* the volume V . In the last integral, the normal \mathbf{n} is directed out of the conductor and, hence, *into* V . Thus, the two surface integrals over S cancel each other. It remains to show that the integral over S' vanishes.

If our charge distribution, which is arbitrary but bounded, bears a net charge, then at large distances from the charge system, the potential falls off inversely as the distance—that is, as r^{-1} . \mathbf{D} falls off as r^{-2} . The area of a closed surface that passes through a point at distance r is proportional to r^2 .

Hence, the value of the integral over S' , which bounds our system at distance r , is proportional to r^{-1} , and when S' is moved out to infinity, its contribution vanishes.

If the charge distribution bears zero net charge, then the potential at large distances acts like some multipole and falls off *more* rapidly than r^{-1} . Again the contribution from S' may be seen to vanish. Thus, for the electrostatic energy, we have

$$U = \frac{1}{2} \int_V \mathbf{D} \cdot \mathbf{E} \, dv \quad (6-17)$$

where the integration is over the volume of the system external to the conductors—that is, over the various dielectrics in the system. The integration may, of course, be extended to include all space, since the electric field \mathbf{E} equals zero inside a conductor. If this formulation is applied to fields that are produced in part by point charges, it is essential to subtract their infinite “self-energy” explicitly. (See Problem 6-7.)

Where is the electrostatic energy of the electrical system located? The precise meaning of this question is difficult to pin down; nevertheless, it is convenient to imagine the energy to be stored in the electric field. Equation (6-17) shows that such a procedure is at least not unreasonable, and in addition it prescribes that the energy be distributed with a density $\frac{1}{2} \mathbf{D} \cdot \mathbf{E}$ per unit volume.

We are led to the concept of *energy density* in an electrostatic field:

$$u = \frac{1}{2} \mathbf{D} \cdot \mathbf{E} \quad (6-18a)$$

Since Eq. (6-17) was derived on the basis of linear dielectrics, each dielectric is characterized by a constant permittivity ϵ . Furthermore, the discussion in preceding chapters has been limited to isotropic dielectrics. Thus, Eq. (6-18a) is equivalent to

$$u = \frac{1}{2} \epsilon E^2 = \frac{1}{2} \frac{D^2}{\epsilon} \quad (6-18b)$$

6-4 ENERGY OF A SYSTEM OF CHARGED CONDUCTORS: COEFFICIENTS OF POTENTIAL

In Section 3-12, it was shown that a linear relationship exists between the potentials and charges on a set of conductors. In fact, in a system composed of N conductors, the potential of one of them is given by

$$\varphi_i = \sum_{j=1}^N p_{ij} Q_j \quad (3-52)$$

The derivation of Eq. (3-52) was carried out for N conductors in vacuum. However, it is clear that this derivation also holds when dielectrics are present in the system, so long as these dielectrics are both linear and devoid of external charge. The coefficient p_{ij} is the potential of the i th conductor due to a unit charge on conductor j . These coefficients are usually referred to as *coefficients of potential*.

In Section 6-2, an expression was developed for the electrostatic energy of a set of N charged conductors, namely, Eq. (6-12). Combining this result with Eq. (3-52), we obtain

$$U = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N p_{ij} Q_i Q_j \quad (6-19)$$

Thus, the energy is a quadratic function of the charges on the various conductors.

Three important general statements can be made about the coefficients p_{ij} : (1) $p_{ij} = p_{ji}$, (2) all of the p_{ij} are positive, and (3) $p_{ii} \geq p_{ij}$ for all j . The first of these statements follows from Eq. (6-19), which expresses U as $U(Q_1 \cdots Q_N)$. Thus,

$$dU = \left(\frac{\partial U}{\partial Q_1} \right) dQ_1 + \cdots + \left(\frac{\partial U}{\partial Q_N} \right) dQ_N$$

If dQ_1 only is changed, then

$$dU = \left(\frac{\partial U}{\partial Q_1} \right) dQ_1 = \frac{1}{2} \sum_{j=1}^N (p_{1j} + p_{j1}) Q_j dQ_1 \quad (6-20)$$

This increment in the electrostatic energy may also be calculated directly from Eq. (6-2). Bringing in dQ_1 from a zero potential reservoir, we obtain

$$dU = dW = \varphi_1 dQ_1 = \sum_{j=1}^N p_{1j} Q_j dQ_1 \quad (6-21)$$

Equations (6-20) and (6-21) must be equivalent for all possible values of the Q_j , which implies that

$$\frac{1}{2}(p_{1j} + p_{j1}) = p_{1j}$$

or

$$p_{j1} = p_{1j} \quad (6-22)$$

In order to prove the second and third statements, we note that all potentials are measured from a zero potential reservoir. We further assume that conductor i bears a positive charge Q_i and that all other conductors are uncharged. Since the origin of the displacement flux is the charge on i , every line of displacement leaving the i th conductor can be followed, possibly by way of other conductors, to the reservoir. Thus $\varphi_i > 0$, and $p_{ii} > 0$. Similarly, unless the j th conductor is shielded by one of the other conductors, every line of displacement impinging on j can be traced back to

the i th conductor, and every line leaving the j th conductor can be traced to the zero potential reservoir. Thus,

$$p_{ii} > p_{ij} > 0$$

It remains to consider the case of a conductor that is completely shielded by another conductor. Suppose the j th conductor, which is uncharged, lies entirely within the conducting shell that comprises the i th conductor. The electric field in the interior region is obviously zero; consequently, the potentials of the two conductors are the same, and $p_{ij} = p_{ii}$. This analysis proves statements (2) and (3), which can be combined into

$$p_{ii} \geq p_{ij} > 0 \quad (6-23)$$

The usefulness of the coefficients p_{ij} may be illustrated by means of a simple example.

EXAMPLE 6-1
Use of Coefficient of Potential to Calculate Electrostatic Potential

The problem is to find the potential of an *uncharged* spherical conductor in the presence of a point charge q at distance r from the center of the sphere, where $r > R$, and R is the radius of the spherical conductor.

Solution: The point charge and sphere are taken to be a system of two conductors, and use is made of the equality $p_{12} = p_{21}$. If the sphere is charged (Q) and the "point" uncharged, then the potential of the "point" is $Q/4\pi\epsilon_0 r$. Thus,

$$p_{12} = p_{21} = \frac{1}{4\pi\epsilon_0 r}$$

Evidently, when the "point" has charge q and the sphere is uncharged, the potential of the latter is $q/4\pi\epsilon_0 r$.

6-5 COEFFICIENTS OF CAPACITANCE AND INDUCTION

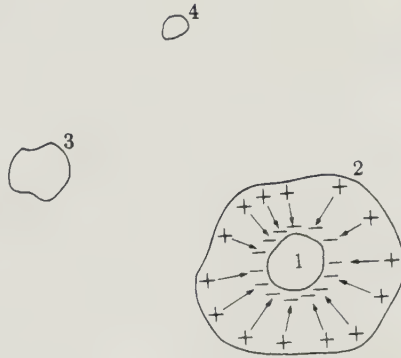
Equation (3-52), which was derived in Chapter 3 and discussed again in Section 6-4, is a set of N linear equations giving the potentials of the conductors in terms of their charges. This set of equations may be solved for the Q_i 's, yielding

$$Q_i = \sum_{j=1}^N c_{ij} \varphi_j \quad (6-24)$$

where c_{ii} is called a *coefficient of capacitance* and c_{ij} ($i \neq j$) is a *coefficient of induction*. The actual solution of Eq. (3-52), expressing each c in terms of the p_{ij} , can be done by matrix inversion.

FIGURE 6-1

Conductors 1 and 2 form a capacitor. Here $p_{13} = p_{23}$ since, by Gauss's law, when 1 and 2 are uncharged, they must be at the same potential, independently of the charge on 3. Similarly, $p_{14} = p_{24}$.



Properties of the c 's follow from those of the p 's, which we have already discussed. Thus: (1) $c_{ij} = c_{ji}$, (2) $c_{jj} > 0$, (3) the coefficients of induction are negative or zero. (See Problem 6-10.)

Equation (6-24) may be combined with Eq. (6-12) to give an alternative expression for the electrostatic energy of an N -conductor system:

$$U = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N c_{ij} \varphi_i \varphi_j \quad (6-25)$$

6-6**CAPACITORS**

In this section we discuss an important device for storing electrostatic energy—namely, the capacitor.

Two conductors that can store equal and opposite charges ($\pm Q$), with a potential difference between them that is independent of whether other conductors in the system are charged, form what is called a *capacitor*.

This independence of other charges* implies that one of the pair of conductors is shielded by the other. In other words, the potential contributed to *each* of the pair by other charges must be the same. Such a situation is depicted in Fig. 6-1 where conductors 1 and 2 form a device of this type. In general, if two conductors, 1 and 2, form a capacitor, we may write

$$\begin{aligned} \varphi_1 &= p_{11}Q - p_{12}Q + \varphi_x \\ \varphi_2 &= p_{12}Q - p_{22}Q + \varphi_x \end{aligned} \quad (6-26)$$

* What we have described is an ideal capacitor. Practical capacitors are affected to some extent by the charges on their surroundings.

where $+Q$ is the charge on 1 and $-Q$ is the charge on 2, and φ_x is the common potential contributed by other charges.

If Eqs. (6-26) are subtracted, we find

$$\Delta\varphi = \varphi_1 - \varphi_2 = (p_{11} + p_{22} - 2p_{12})Q \quad (6-27)$$

Thus, the difference in potential between the conductors of a capacitor is proportional to the charge stored, Q . (Obviously, the total charge stored is zero, but by convention, the absolute value of the charge on one of the two conductors is called the *charge* on the capacitor.)

Equation (6-27) may be written

$$Q = C \Delta\varphi \quad (6-28)$$

where $C = (p_{11} + p_{22} - 2p_{12})^{-1}$ is called the *capacitance of the capacitor*.

Evidently C is the charge stored per unit of potential difference. In the mks system, C is measured in coulombs per volt, or farads ($1 \text{ F} \equiv 1 \text{ C/V}$).

From Eq. (6-12), with Eq. (6-28), the energy of a charged capacitor may be written as

$$U = \frac{1}{2} Q \Delta\varphi = \frac{1}{2} C (\Delta\varphi)^2 = \frac{1}{2} \frac{Q^2}{C} \quad (6-29)$$

If the two conductors making up the capacitor have simple geometrical shapes, the capacitance may be obtained analytically. Thus, for example, it is easy to calculate the capacitance of two parallel plates, two coaxial cylinders, two concentric spheres, or that of a cylinder and a plane. The capacitance of a parallel-plate capacitor (Fig. 6-2) will be derived here. Other simple cases are taken up in the problems at the end of the chapter.

FIGURE 6-2

The electric field between oppositely charged parallel plates of finite area

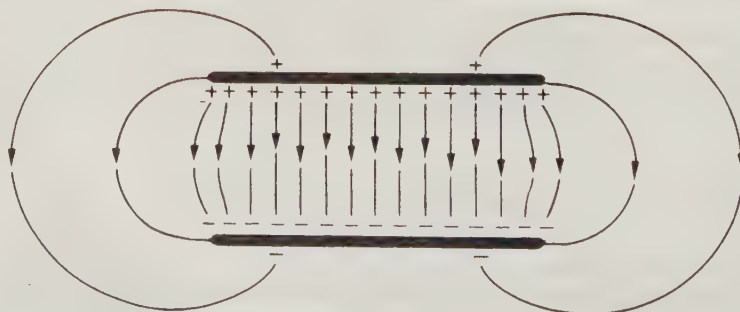
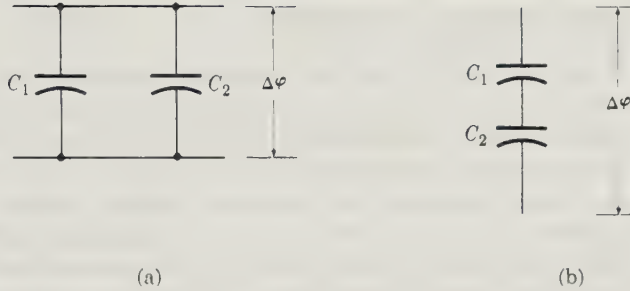


FIGURE 6-3

(a) Parallel and
(b) series connection of
two capacitors



Except for the fringing field at the edge of the parallel plates, the electric field between them is uniform. An ideal parallel-plate capacitor is one in which the plate separation d is very small compared with dimensions of the plate. Thus, the fringing field may be neglected in the ideal case. If the region between the plates is filled with dielectric of permittivity ϵ , then the electric field between the plates is

$$E = \frac{1}{\epsilon} \sigma = \frac{Q}{\epsilon A}$$

where A is the area of one plate. The potential difference $\Delta\varphi = Ed$. Therefore,

$$C = \frac{Q}{\Delta\varphi} = \frac{\epsilon A}{d} \quad (6-30)$$

is the capacitance of this capacitor.

When a capacitor is depicted as part of an electric circuit, it is usually indicated by the symbol $\nabla\epsilon$. Two or more capacitors may be joined together by connecting one of the conductors of the first capacitor to a conductor of the second one. Possible ways of joining two capacitors are by parallel connection (Fig. 6-3a), or by series connection (Fig. 6-3b). After the capacitors are joined, it is usually desirable to talk about the equivalent capacitance of the combination. In the parallel case, the same voltage $\Delta\varphi$ appears across each capacitor and also across the combination. Hence, the equivalent capacitance is given by

$$C = \frac{Q_{\text{total}}}{\Delta\varphi} = \frac{Q_1}{\Delta\varphi} + \frac{Q_2}{\Delta\varphi} = C_1 + C_2 \quad (6-31a)$$

If two uncharged capacitors are connected in series and subsequently charged, *conservation of charge* requires that each capacitor acquire the same charge. Thus, the equivalent capacitance C of the combination is related to C_1 and C_2 by the expression

$$\frac{1}{C} = \frac{\Delta\varphi}{Q} = \frac{\Delta\varphi_1 + \Delta\varphi_2}{Q} = \frac{1}{C_1} + \frac{1}{C_2} \quad (6-31b)$$

6-7 FORCES AND TORQUES

Thus far in this chapter, we have developed a number of alternative procedures for calculating the electrostatic energy of a charge system. We shall now see how the force on one of the objects in the charge system may be calculated from a knowledge of this electrostatic energy.

Let us suppose we are dealing with an isolated system composed of a number of parts (conductors, point charges, dielectrics), and we allow one of these parts to make a small displacement $d\mathbf{r}$ under the influence of the electrical forces \mathbf{F} acting upon it. The work performed by the electrical force on the system in these circumstances is

$$dW = \mathbf{F} \cdot d\mathbf{r} = F_x dx + F_y dy + F_z dz \quad (6-32)$$

Because the system is isolated, this work is done at the expense of the electrostatic energy U . In other words, according to Eq. (6-1),

$$dW = -dU \quad (6-33)$$

Combining Eqs. (6-32) and (6-33) yields

$$-dU = F_x dx + F_y dy + F_z dz$$

and

$$F_x = -\frac{\partial U}{\partial x} \quad (6-34)$$

with similar expressions for F_y and F_z . That is, in this case \mathbf{F} is a conservative force, and $\mathbf{F} = -\nabla U$.

If the object under consideration is constrained to move in such a way that it rotates about an axis, then Eq. (6-32) may be replaced by

$$dW = \boldsymbol{\tau} \cdot d\boldsymbol{\theta} \quad (6-35)$$

where $\boldsymbol{\tau}$ is the electrical torque and $d\boldsymbol{\theta}$ is the differential angular displacement. Writing $\boldsymbol{\tau}$ and $d\boldsymbol{\theta}$ in terms of their components, (τ_1, τ_2, τ_3) and ($d\theta_1, d\theta_2, d\theta_3$), and combining Eqs. (6-33) and (6-35), we obtain

$$\tau_1 = -\frac{\partial U}{\partial \theta_1} \quad (6-36)$$

and so on.

Thus, our goal has been achieved:

$$F_x = -\left(\frac{\partial U}{\partial x}\right)_Q \quad (6-34a)$$

$$\tau_1 = -\left(\frac{\partial U}{\partial \theta_1}\right)_Q \quad (6-36a)$$

where the subscript Q has been added to denote that the system is isolated,

and hence, its total charge remains constant during the displacement $d\mathbf{r}$ or $d\theta$. To exploit this method, it is necessary to express U in analytic form, and the specific dependence of U on the coordinate x , or θ_1 , must be given. An example showing the usefulness of the method will be presented shortly.

Equations (6-34a) and (6-36a) do not, however, cover all cases of interest because, as was mentioned in their derivation, they are limited to isolated systems in which the charge in the system remains constant. In another important class of problems, all of the charge exists on the surfaces of conductors, and these are maintained at fixed potentials by means of external sources of energy (e.g., by means of batteries). Here again we may allow one of the parts of the system to move under the influence of electrical forces acting upon it, and the work performed (this time by the system and the batteries) will still be related to the force by Eq. (6-32). In this instance, the work becomes

$$dW = dW_b - dU \quad (6-37)$$

where dW_b is the work supplied by the batteries. Before we can proceed to an expression linking U and the force on some part of the system for this case, it will be necessary to eliminate dW_b from Eq. (6-37).

The electrostatic energy U of a system of charged conductors has been given earlier in Eq. (6-12). Now, if some part of the system is displaced while at the same time the potentials of all conductors remain fixed,

$$dU = \frac{1}{2} \sum_j \varphi_j dQ_j \quad (6-38)$$

Furthermore, the work supplied by the batteries, dW_b , is the work required to move each of the charge increments dQ_j from zero potential to the potential of the appropriate conductor. By Eq. (6-2), this work is

$$dW_b = \sum_j \varphi_j dQ_j \quad (6-39)$$

Thus,

$$dW_b = 2 dU \quad (6-40)$$

Using this equation to eliminate dW_b from Eq. (6-37) and combining the result with Eq. (6-32), we obtain

$$dU = F_x dx + F_y dy + F_z dz$$

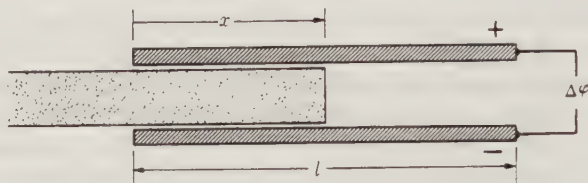
or

$$F_x = \left(\frac{\partial U}{\partial x} \right)_\varphi \quad (6-41)$$

Here the subscript φ is used to denote the fact that all potentials are maintained constant during the virtual displacement $d\mathbf{r}$. In a similar fashion,

FIGURE 6-4

Dielectric slab partially withdrawn from between two charged plates



we may derive

$$\tau_1 = \left(\frac{\partial U}{\partial \theta_1} \right)_\varphi \quad (6-42)$$

To illustrate the energy method, let us consider the following example.

EXAMPLE 6-2
Restoring Force on a Dielectric Block Partially Withdrawn from a Capacitor

A parallel-plate capacitor of plate separation d has the region between its plates filled by a block of solid dielectric of permittivity ϵ . The dimensions of each plate are length l , width w . The plates are maintained at the constant potential difference $\Delta\varphi$. If the dielectric block is withdrawn along the l dimension until only the length x remains between the plates (Fig. 6-4), calculate the force tending to pull the block back into place.

Solution: The energy of the system may be calculated by any of several methods. Thus, for example, since $E = \Delta\varphi/d$ is the same everywhere between the plates, let us use

$$U = \frac{1}{2} \int_V \epsilon E^2 dv$$

where the region of integration need include only those parts of space where $E \neq 0$. Neglecting fringing effects at the edge of the capacitor, we find

$$U = \frac{1}{2} \epsilon \left(\frac{\Delta\varphi}{d} \right)^2 dwx + \frac{1}{2} \epsilon_0 \left(\frac{\Delta\varphi}{d} \right)^2 dw(l-x)$$

The force may be calculated from Eq. (6-41):

$$F_x = \frac{1}{2} (\epsilon - \epsilon_0) w \frac{(\Delta\varphi)^2}{d} = \frac{1}{2} (K - 1) \epsilon_0 E^2 w d$$

in the direction of increasing x .

The case in which the plates are isolated (constant charge Q) is treated in Problems 6-19 and 6-24.

SUMMARY

The **electrostatic potential energy** of a system of point charges is calculated as the work that would have to be done by an outside agency against the Coulomb forces between the charges in order to assemble the given configuration. It is expressible as

$$U = \frac{1}{2} \sum q_j \varphi_j$$

where φ_j , the potential at the position of q_j due to all the *other* charges, is

$$\varphi_j = \sum' \frac{q_k}{4\pi\epsilon_0 r_{jk}}$$

with the term $k = j$ excluded. For a general charge distribution the electrostatic energy becomes, provided that all dielectrics present are linear,

$$U = \frac{1}{2} \int \rho \varphi dv$$

where the potential φ is that produced by the external charge density ρ in the presence of the dielectric media. (ρ may include charge concentrated in a surface distribution or point charges.) Integration by parts transforms the energy in linear dielectrics into an integral,

$$U = \int u dv$$

of the **energy density** of the electric field,

$$u = \frac{1}{2} \mathbf{E} \cdot \mathbf{D} = \frac{1}{2} \epsilon E^2 = \frac{1}{2} \frac{D^2}{\epsilon}$$

When this formulation is applied to point charges, their infinite "self-energy" must be subtracted.

• When all the charge is a surface distribution on conductors, whose surfaces are equipotentials, the electrostatic energy specializes to

$$U = \frac{1}{2} \sum Q_j \varphi_j$$

Then the coefficients in the linear functions

$$\varphi_i = \sum p_{ij} Q_j$$

and in the inverse functions

$$Q_i = \sum c_{ij} \varphi_j$$

are found to satisfy the conditions

$$p_{ij} = p_{ji}, \quad c_{ij} = c_{ji}$$

(In addition, $p_{ii} \geq p_{ij} > 0$ and $c_{ii} > 0 \geq c_{ij}$.)

- In the special case where two conductors form a **capacitor**,

$$U = \frac{1}{2} Q \Delta\varphi$$

with

$$Q = C \Delta\varphi$$

For a **parallel-plate capacitor**,

$$C = \frac{\epsilon A}{d}$$

- The **electric force** on a part of an isolated system, with constant charge on each conductor, is the negative gradient of the electrostatic energy.

$$F_x = - \left(\frac{\partial U}{\partial x} \right)_Q$$

If the system is not isolated, but instead the potential of each conductor is maintained constant by an outside agency (battery), the force is given by

$$F_x = + \left(\frac{\partial U}{\partial x} \right)_\varphi$$

PROBLEMS

- 6-1** A fast electron (kinetic energy = 5.0×10^{-17} J) enters a region of space containing a uniform electric field $E = 800$ V/m. The field is parallel to the electron's motion and in a direction such as to decelerate it. How far does the electron travel before it is brought to rest? (Charge of electron = 1.60×10^{-19} C.)
- 6-2** Given a spherical dielectric shell (inner radius a , outer radius b , dielectric constant K) and a point charge q , infinitely separated. Now let the point charge be placed at the center of the dielectric shell. Determine the change in energy of the system.
- 6-3** A dipole ql is positioned perpendicular to a conducting plane, so that the charge $-q$ is at a distance d and the charge $+q$ is at $d + l$. Calculate the electrostatic energy of the system of charges. [*Hint*: Consider the energy of the system consisting of the true charges plus the image charges in vacuum, where the image charges are chosen to give the correct E -field in front of the plane.]
- 6-4** Consider a spherical shell of charge, radius R , uniform surface charge density σ_0 . Determine the self-energy of the distribution in two ways: (a) by direct integration of Eq. (6-8); (b) by integration over the field, $\frac{1}{2} \int \mathbf{E} \cdot \mathbf{D} \, dv$.
- 6-5** Consider a spherical charge distribution of radius R , uniform charge density ρ_0 . Determine the self-energy of the distribution in two ways: (a) by direct integration of Eq. (6-8); (b) by an integration over the field, $\frac{1}{2} \int \mathbf{E} \cdot \mathbf{D} \, dv$.

6-6 Let us assume that an electron is a uniformly charged, spherical particle of radius R . Assume further that the rest energy, mc^2 (where m is the mass of the electron, and c is the velocity of light), is electrostatic in origin and given by the result of Problem 6-5. By putting in appropriate numerical values for the charge and mass of the electron, determine its "classical radius" R .

6-7 Two point charges q_1 and q_2 are separated by a distance d . If their respective fields at a point \mathbf{r} are \mathbf{E}_1 and \mathbf{E}_2 ,

$$E^2 = E_1^2 + E_2^2 + 2\mathbf{E}_1 \cdot \mathbf{E}_2$$

(a) Show that the integrals of E_1^2 and E_2^2 over space are divergent. This is the infinite "self-energy" that must be subtracted from the energy U . (b) Set up an integral for the contribution of $2\mathbf{E}_1 \cdot \mathbf{E}_2$ to U and show that it is not divergent.

6-8 (a) What is the capacity of a capacitor that can store 1000 J at 1000 V? (b) Assuming the capacitor has parallel plates separated by 10^{-5} m and filled with a material of dielectric constant 2, what is the necessary area of the plates?

6-9 An electrophorus consists of a flat circular plate of wax and a similar plate of metal with an insulating handle. The wax plate is given a bound charge Q by rubbing it with fur. The metal plate is laid on the wax plate and temporarily grounded, so that it acquires a charge $-Q$. The metal plate is finally removed from the wax plate, retaining its charge $-Q$. Suppose the radius of the plates is 10 cm, $Q = 0.3 \mu\text{C}$, and the initial separation of the two plates is 10^{-5} m. Find the potential difference between the plates and the stored energy when the separation is (a) $d = 10^{-5}$ m, (b) $d = 0.02$ m.

6-10 A system of conductors consists of only two conductors. Find the coefficients of capacity and induction explicitly in terms of the coefficients of potential p_{ij} .

6-11 Two spherical conductors are located in vacuum. Conductor 1, of radius R , is grounded (i.e., at zero potential). Conductor 2 is so small that it may be treated as a point charge. It bears the charge q and is located at distance d from the grounded sphere. What is the charge induced on the grounded sphere? (Use the concept of coefficient of potential.)

6-12 Consider a system of two conducting objects in a linear dielectric medium. Conductor 1 is uncharged, and conductor 2 is grounded. Prove that conductor 1 is also at ground potential.

6-13 A parallel-plate capacitor is made with a composite dielectric. A sheet of dielectric of permittivity ϵ_1 , thickness d_1 , is placed on top of a second dielectric sheet (permittivity ϵ_2 , thickness d_2). The combination is placed between parallel conducting plates that are separated by the distance $d_1 + d_2$. What is the capacitance per unit plate area of the capacitor?

6-14 A long conducting cylinder of radius a is oriented parallel to and at distance h from an infinite conducting plane. Show that the capacitance of the system, per unit length of the cylinder, is given by

$$C = 2\pi\epsilon_0/\cosh^{-1}(h/a)$$

(See Section 3-10.)

6-15 Two identical air capacitors are connected in series, and the combination is maintained at the constant potential difference of 50 V. If a dielectric sheet of

dielectric constant 10 and thickness equal to one-tenth of the air gap is inserted into one of the capacitors, calculate the voltage across this capacitor.

6-16 The capacitance of a gold-leaf electroscope is not quite constant because the leaf moves closer to the case as $\Delta\varphi$ increases. The expected form for the capacitance is

$$C = a + b(\Delta\varphi)^2$$

How would you determine the constants a and b for a particular instrument? What is the energy of the electroscope when it is charged? Is the energy entirely electrical?

6-17 Two concentric, spherical, conducting shells of radii r_1 and r_2 are maintained at potentials φ_1 and φ_2 , respectively. The region between the shells is filled with a dielectric medium. Show by direct calculation that the energy stored in the dielectric is equal to $C(\varphi_1 - \varphi_2)^2/2$, and thereby determine C , the capacitance of the system.

6-18 Two coaxial, cylindrical conductors of approximately the same radius are separated in the radial dimension by the small distance d . The cylinders are inserted normally into a liquid dielectric of susceptibility χ and mass density ζ . The cylinders are maintained at the potential difference $\Delta\varphi$. To what height h does the dielectric rise between the conductors? (Neglect surface tension.)

6-19 A parallel-plate capacitor has the region between its plates filled with a dielectric slab of dielectric constant K . The plate dimensions are width w and length l , and the plate separation is d . The capacitor is charged while it is connected to a potential difference $(\Delta\varphi)_0$, after which it is disconnected. The dielectric slab is now partially withdrawn in the l dimension until only the length x remains between the plates. (a) What is the potential difference across the capacitor? (b) What is the force tending to pull the dielectric slab back to its original position?

6-20 The capacitance of a variable air capacitor changes linearly from 50 to 364 pF during a rotation from 0° to 180° . When set at 75° , a potential difference of 400 V is maintained across the capacitor. What is the direction and magnitude of electrostatic torque experienced by the capacitor?

***6-21** An uncharged, conducting, spherical shell of mass m floats with one fourth of its volume submerged in a liquid dielectric of dielectric constant K . To what potential must the sphere be charged to float half submerged? [*Hint*: Assume the electric field of the half-submerged, charged shell to be a purely radial field, and show later that the sum of $\sigma + \sigma_p$ over the spherical surface is such as to justify this assumption.]

6-22 A dielectric slab of thickness d and dielectric constant K fills the region between the plates of a parallel-plate capacitor. The plate area is A . Calculate the electrostatic force on one of the capacitor plates (a) on the assumption that the dielectric is in direct contact with the plate and (b) on the assumption that there is a narrow air space between dielectric and plate. The plates are maintained at the potential difference $\Delta\varphi$ in both cases.

6-23 Solve the example of Fig. 6-4 with the electrostatic energy in the form $U = \frac{1}{2}C(\Delta\varphi)^2$, where $C = C(x)$ is the capacitance of the capacitor with the dielectric block inserted a distance x .

6-24 Suppose the electrostatic energy of a system is $U = \frac{1}{2}Q \Delta\varphi$, with $Q = C \Delta\varphi$. Let $C = C(x)$ depend on some geometrical parameter x which specifies the position

of a part of the system. Show that the force on that part, for a particular value of Q and $\Delta\varphi$, is the same according to Eq. (6-34a) for constant Q as according to Eq. (6-41) for constant $\Delta\varphi$.

6-25 Solve the example of Fig. 6-4 if the thickness of the dielectric slab t is much less than the separation of the plates d . [*Hint*: In this case, according to the result of Problem 4-15, it is D rather than E that is nearly the same everywhere between the plates.]

6-26 A long cylindrical wire of radius R_1 is surrounded coaxially by a long cylindrical shell of radius R_2 (with $R_2 > R_1$). These two conductors form the "plates" of a capacitor, and the region between them is filled with a nonconducting material of permittivity ϵ . (a) Determine the capacitance per unit length of the capacitor. (b) Show that this result depends only on the ratio R_2/R_1 .

6-27 The maximum charge Q that a capacitor can store is determined by both the capacitance of the capacitor and the dielectric strength of the dielectric E_{\max} (see Table 4-1). (a) Determine Q_{\max} in terms of ϵ and E_{\max} for a parallel plate capacitor. (b) Which dielectric, aluminum oxide or glass, will store more charge in a parallel plate capacitor of the same plate area?

6-28 (a) Use the method of moments described in Section 3-13 to calculate the capacitance of a square plate capacitor. Do at least 1×1 , 2×2 , 3×3 and 4×4 arrays. (b) Compare results with the capacitance obtained from the infinite plate approximation.

Electric Current

Up to this point, we have been dealing with charges at rest; now we wish to consider charges in uniform motion. This statement implies that we shall be dealing with conductors of electricity because, by definition, a conductor is a material in which the charge carriers are free to move under steady electric fields. (See Section 2-5.) The preceding definition includes not only the conventional conductors, such as metals and alloys, but also semiconductors, electrolytes, ionized gases, imperfect dielectrics, and even vacuums in the vicinity of a thermionic emitting cathode. In many conductors, the charge carriers are electrons, and in other cases, the charge may be carried by positive or negative ions.

Moving charge constitutes a *current*, and the process whereby charge is transported is called *conduction*. To be precise, the current I is defined as the rate at which charge is transported through a given surface in a conducting system (e.g., through a given cross section of a wire). Thus,

$$I = \frac{dQ}{dt} \quad (7-1)$$

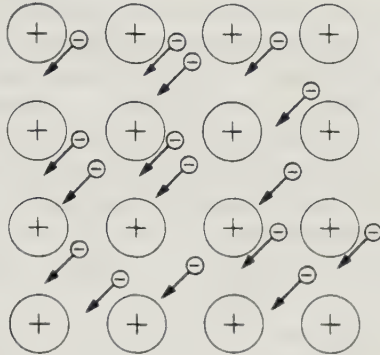
where $Q = Q(t)$ is the net charge transported in time t . The unit of current in the mks system is the ampere (A), named for the French physicist André Marie Ampère. Evidently,

$$1 \text{ ampere} = 1 \frac{\text{coulomb}}{\text{second}}$$

NATURE OF THE CURRENT

In a metal, current is carried entirely by electrons, while the heavy positive ions are fixed at regular positions in the crystal structure (Fig. 7-1). Only the

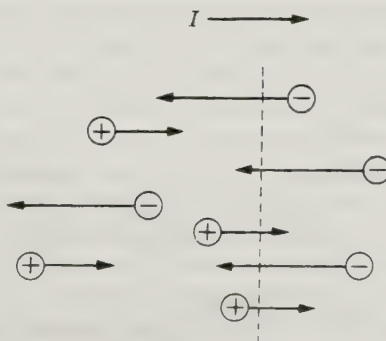
FIGURE 7-1
Schematic diagram of the
motion of conduction
electrons in a metal



valence (outermost) atomic electrons are free to participate in the conduction process. The other electrons are tightly bound to their ions. Under steady-state conditions, electrons may be fed into the metal at one point and removed at another, producing a current, but the metal as a whole is electrostatically neutral. Strong electrostatic forces keep excess electrons from accumulating at any point in the metal. Similarly, a deficiency of electrons is remedied by electrostatic forces of the opposite sign. We shall see later that excess charge is dissipated extremely rapidly in a conductor. Thus, we note that it is possible to study the subject of electric current without taking into account detailed electrostatic effects associated with the charge carriers.

In an electrolyte, although the current is carried by both positive and negative ions, because some ions move faster than others, conduction by one type of ion usually predominates. It is important to note that positive and negative ions traveling in *opposite* directions (Fig. 7-2) contribute to the current in the *same* direction. The basis for this fact is evident from Eq. (7-1), since the net charge transported through a given surface depends on

FIGURE 7-2
Current produced by the
motion of both positive
and negative charge
carriers



both the sign of the charge carrier and the direction in which it is moving. Thus, in Fig. 7-2, both the positive and negative carrier groups produce currents to the right. By convention, the direction in which the positive carrier moves (or, equivalently, the direction opposite to that in which the negative carrier moves) is taken as the direction, or *sense*, of the current. In general, an electric current arises in response to an electric field. If an electric field is imposed on a conductor, it will cause positive charge carriers to move in the general direction of the field and negative carriers in a direction opposite to the field. Hence, all currents produced in the process have the same direction as the field.

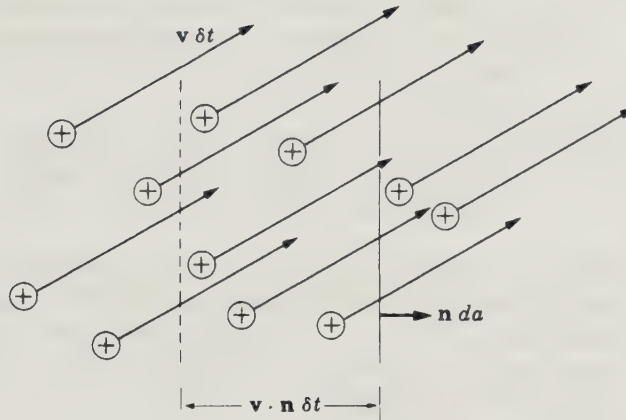
In a gas discharge, the current is carried by both electrons and positive ions, but because the electrons are so much more mobile than the heavy ions, practically all of the current is carried by electrons. Gas conduction is somewhat complicated, because the electronic and ionic populations vary greatly with the experimental conditions (they are determined primarily by the gas pressure and the potential drop across the gas). Under certain conditions *cascading* occurs, in which process the few ions that are initially present accelerate and make inelastic collisions with neutral atoms, thereby producing additional ions and electrons. The additional ions can also give rise to ionizing collisions, with the result that the carrier density builds up enormously.

In Figs. 7-1 and 7-2, we have pictured the charge carriers as falling into groups, each of which has a common motion, called the *drift motion* of the group. The picture has been greatly oversimplified, however. Each group of charge carriers actually represents an assembly of particles in thermal equilibrium with its environment, and so each particle has thermal motion as well as drift motion. But the thermal motion, although it may be large, is also random and, hence, gives rise to no organized transport of charge. The drift motion, on the other hand, is not random. In considering the conduction process, then, it is permissible to forget about the random motion, which yields no net current, and to use the simple picture presented in Figs. 7-1 and 7-2. For certain other transport processes, however, such as conduction in a thermal gradient (which gives rise to the thermoelectric effect), it is necessary to take the thermal motion into account in a detailed way in order to understand the phenomena fully.

The currents we have described thus far in this section are known as conduction currents. These *conduction currents* represent the drift motion of charge carriers in a *neutral* medium, which as a whole, may be and usually is at rest. Liquids and gases may also undergo hydrodynamic motion, and if the medium has a charge density, this hydrodynamic motion will produce currents. Such currents, arising from mass transport of a *charged* medium, are called *convection currents*. Convection currents are important to the subject of atmospheric electricity. In fact, the upward convection currents in thunderstorms are sufficient to maintain the normal potential gradient of the atmosphere above the earth. The motion of charged particles in vacuum

FIGURE 7-3

The drift motion of charge carriers across the plane da in time δt



(such as electrons in a vacuum diode) also constitutes a convection current. A characteristic feature of the convection current is that it is not electrostatically neutral, and its electrostatic charge must usually be taken into account.

In the rest of this chapter we shall deal exclusively with conduction currents.

7-2**CURRENT DENSITY: EQUATION OF CONTINUITY**

In this section, we first consider a conducting medium that has only one type of charge carrier, of charge q . The number of these carriers per unit volume will be denoted by N . In accordance with the preceding section, we ignore their random thermal motion and assign the same drift velocity \mathbf{v} to each carrier. We are now in a position to calculate the current through an element of area da such as is shown in Fig. 7-3. During the time δt , each carrier moves a distance $\mathbf{v} \delta t$. From the figure it is evident that the charge δQ that crosses da during time δt is q times the number of all charge carriers in the volume $\mathbf{v} \cdot \mathbf{n} \delta t da$, where \mathbf{n} is a unit vector normal to the area da . From Eq. (7-1), the current through da is

$$dI = \frac{\delta Q}{\delta t} = \frac{qN\mathbf{v} \cdot \mathbf{n} \delta t da}{\delta t} = Nq\mathbf{v} \cdot \mathbf{n} da \quad (7-2)$$

If there is more than one kind of charge carrier present, there will be a contribution of the form (7-2) from each type of carrier. In general, the current through the area da is

$$dI = \left[\sum_i N_i q_i \mathbf{v}_i \right] \cdot \mathbf{n} da \quad (7-3)$$

Here the summation is over the different carrier types.

The quantity in brackets is a vector that has dimensions of current per unit area. This quantity is called the *current density* and is given the symbol \mathbf{J} :

$$\mathbf{J} = \sum_i N_i q_i \mathbf{v}_i \quad (7-4)$$

The current density may be defined at each point in a conducting medium and is, therefore, a vector point function. It is a useful quantity, one that enters directly into the fundamental differential equations of electromagnetic theory. The mks unit of \mathbf{J} is ampere per meter² (A/m²). Equation (7-3) may be written as

$$dI = \mathbf{J} \cdot \mathbf{n} da$$

and the current through the surface S , an arbitrarily shaped surface area of macroscopic size, is given by the integral

$$I = \int_S \mathbf{J} \cdot \mathbf{n} da \quad (7-5)$$

The current density \mathbf{J} and the charge density ρ are not independent quantities, but are related at each point through a differential equation, the so-called *equation of continuity*. This relationship has its origin in the fact that charge can neither be created nor destroyed. The equation is most easily derived by applying Eq. (7-5) to an arbitrary *closed* surface S . The electric current entering V , the volume enclosed by S , is given by

$$I = -\oint_S \mathbf{J} \cdot \mathbf{n} da = -\int_V \nabla \cdot \mathbf{J} dv \quad (7-6)$$

In this equation, the last integral is obtained through the use of the divergence theorem. The minus sign comes about because \mathbf{n} is the outward normal, and we wish to call I positive when the net flow of charge is from the outside of V to within. But from Eq. (7-1), I is equal to the rate at which charge is transported into V :

$$I = \frac{dQ}{dt} = \frac{d}{dt} \int_V \rho dv \quad (7-7a)$$

Since we are dealing with a fixed volume V , the time derivative operates only on the function ρ . However, ρ is a function of position as well as of time, so that the time derivative becomes the partial derivative with respect to time when it is moved inside the integral. Hence,

$$I = \int_V \frac{\partial \rho}{\partial t} dv \quad (7-7b)$$

Equations (7-6) and (7-7b) may now be equated:

$$\int_V \left(\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} \right) dv = 0 \quad (7-8)$$

But V is completely arbitrary, and the only way that Eq. (7-8) can hold for an *arbitrary volume segment* of the medium is for the integrand to vanish at each point, giving the following result.

The equation of continuity is

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0 \quad (7-9)$$

7-3 OHM'S LAW: CONDUCTIVITY

It is found experimentally that in a metal at constant temperature the current density \mathbf{J} is linearly proportional to the electric field.

Thus, the constitutive equation is

$$\mathbf{J} = g\mathbf{E} \quad (7-10)$$

The constant of proportionality g is called the *conductivity*. Equation (7-10), which is called *Ohm's law*, is a very good approximation for a large number of the common conducting materials.

In the general isotropic case, however, Eq. (7-10) must be replaced by

$$\mathbf{J} = g(E)\mathbf{E}$$

where $g(E)$ is a function of the electric field. Materials for which Eq. (7-10) holds are called *linear isotropic media* or *ohmic media*. Here again, as with dielectrics, we shall be most concerned with the linear case.

The reciprocal of the conductivity is called the *resistivity* η .^{*} Thus,

$$\eta = \frac{1}{g} \quad (7-11)$$

The unit of η in the mks system is volt-meter per ampere, or simply

^{*} The common symbols for resistivity and conductivity are ρ and σ , respectively, but to avoid the possibility of confusion with volume charge density ρ and surface charge density σ , we shall use the symbols η and g .

TABLE 7-1

Resistivity η and Temperature Coefficient of Resistance α of Some Common Materials at 20°C

Material	$\eta, \Omega \text{ m}$	$\alpha = \frac{1}{\eta} \frac{d\eta}{dT}, (^{\circ}\text{C})^{-1}$
Aluminum	2.65×10^{-8}	0.0043
Copper	1.67×10^{-8}	0.0039
Gold	2.35×10^{-8}	0.004
Iron	9.71×10^{-8}	0.0065
Nickel	6.84×10^{-8}	0.0069
Silver	1.59×10^{-8}	0.0041
Mercury	95.8×10^{-8}	0.0009
Tungsten	5.51×10^{-8}	0.0045
Constantin (Cu 60, Ni 40)	49.0×10^{-8}	0.0000
Nichrome	100.0×10^{-8}	0.0004
Germanium (pure)	0.46	-0.048
Germanium ($5 \times 10^{-6}\%$ As)	0.011	
Graphite	1.4×10^{-5}	
NaCl solution (saturated)	0.044	-0.005
Aluminum oxide	1×10^{14}	
Glass	$10^{10} - 10^{14}$	
Iodine	1.3×10^7	
Quartz (SiO_2)	1×10^{13}	
Sulfur	2×10^{15}	
Wood	$10^8 - 10^{11}$	

Source: Data from the *Handbook of Chemistry and Physics*, 58th ed. (Cleveland, OH: CRC Press, Inc., 1978).

ohm-meter, where the ohm (Ω) is defined by

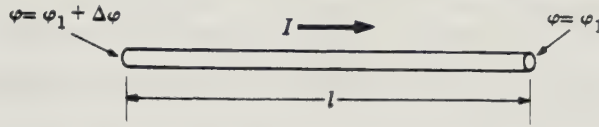
$$1 \text{ ohm} = \frac{1 \text{ volt}}{1 \text{ ampere}}$$

The unit of conductivity g is $\Omega^{-1} \text{ m}^{-1}$ or S/m. One siemens (S) is one reciprocal ohm. (A siemens was formerly called a mho.)

Resistivities of a number of common materials are given in Table 7-1. It is apparent from this table that all materials conduct electricity to some extent, but that the materials we have called insulators (dielectrics) are much poorer conductors than the metals by a tremendous factor (as much as 10^{23}). The distinction between a conductor and an insulator will be discussed in a more quantitative way in Section 7-7.

Consider a conducting specimen obeying Ohm's law, in the shape of a straight wire of uniform cross section with a constant potential difference, $\Delta\phi$, maintained between its ends. (See Fig. 7-4.) The wire is assumed to be homogeneous and characterized by the constant conductivity g . Under these conditions an electric field will exist in the wire, the field being related to

FIGURE 7-4
Geometry of the straight
wire segment



$\Delta\varphi$ by the relation

$$\Delta\varphi = \int \mathbf{E} \cdot d\mathbf{l} \quad (7-12a)$$

It is evident that there can be no steady-state component of electric field at right angles to the axis of the wire, since by Eq. (7-10) this would produce a continual charging of the wire's surface. Thus, the electric field is purely longitudinal.

Furthermore, because of the geometry, the electric field must be the same at all points along the wire. Therefore, Eq. (7-12a) reduces to

$$\Delta\varphi = El \quad (7-12b)$$

where l is the length of the wire. But an electric field implies a current of density $\mathbf{J} = g\mathbf{E}$. The current through any cross section of the wire is

$$I = \int_A \mathbf{J} \cdot \mathbf{n} \, da = JA \quad (7-13)$$

where A is the cross-sectional area of the wire. Combining Eq. (7-13) with Eqs. (7-10) and (7-12b), we obtain

$$I = \frac{gA}{l} \Delta\varphi \quad (7-14)$$

which provides a linear relationship between I and $\Delta\varphi$.

The quantity l/gA is called the *resistance* of the wire and will be denoted by the symbol R . Using R , we may rewrite Eq. (7-14):

$$\Delta\varphi = RI \quad (7-15)$$

which is the familiar form of Ohm's law (R is evidently measured in units of ohms). Equation (7-15) may be considered to be a definition of the resistance of an object or device that is passing a constant current. In the general case, R will depend upon the value of this current. However, as was mentioned earlier, we are primarily interested in linear materials, which are called *ohmic* materials, and here R is independent of the current.

The work done by the field when a charge dQ moves through the potential difference $\Delta\varphi$ is $dW = dQ \Delta\varphi$. The corresponding power is

$$P = I \Delta\varphi = I^2 R = (\Delta\varphi)^2 / R$$

where the last two forms are obtained by combining with Ohm's law. This power is dissipated as Joule heating of the material.

7-4 STEADY CURRENTS IN CONTINUOUS MEDIA

There is a very close analogy between an electrostatic system of dielectrics bounded by equipotential surfaces, on the one hand, and a system that conducts a steady current, on the other. This analogy is the subject of the present section.

Let us consider a homogeneous, ohmic (linear), conducting medium under conditions of steady-state conduction. Since we are dealing specifically with the steady state, the local charge density $\rho(x, y, z)$ is at its equilibrium value, and $\partial\rho/\partial t = 0$ for each point in the medium. Hence, the equation of continuity (Eq. 7-9) reduces to

$$\nabla \cdot \mathbf{J} = 0 \quad (\text{steady currents}) \quad (7-16)$$

Using Ohm's law in combination with (7-16), we obtain

$$\nabla \cdot g\mathbf{E} = 0$$

which for a homogeneous medium reduces to

$$\nabla \cdot \mathbf{E} = 0$$

But since $\nabla \times \mathbf{E} = 0$ for a static field, \mathbf{E} is derivable from a scalar potential:

$$\mathbf{E} = -\nabla\varphi$$

Combination of the last two equations yields

$$\nabla^2\varphi = 0 \quad (7-17)$$

which is Laplace's equation.

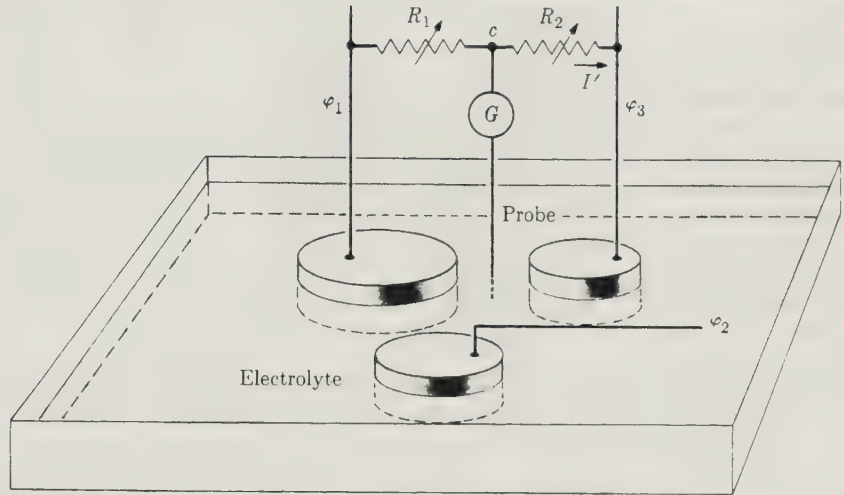
We see, therefore, that a steady-state conduction problem may be solved in the same way as electrostatic problems. Laplace's equation is solved by one of the techniques discussed in Chapter 3, the appropriate solution being determined, as always, by the boundary conditions. Boundary conditions that are sufficient to solve the problem are those that specify either φ or \mathbf{J} at each point on the surface of the conducting medium. Specifying \mathbf{J} at the surface is equivalent to specifying \mathbf{E} , since the two vectors are connected by Ohm's law. Once the appropriate solution to Laplace's equation has been found, \mathbf{E} (and hence \mathbf{J}) may be determined at each point inside the medium from the gradient operation.

Under steady-state conduction the current which crosses an interfacial area between two conducting media may be computed in two ways: in terms of the current density in medium 1 or in terms of the current density in medium 2. Since the two procedures must yield the same result, the normal component of \mathbf{J} must be continuous across the interface:

$$J_{1n} = J_{2n} \quad (7-18a)$$

FIGURE 7-5

Two-dimensional electrolytic tank. The three metallic conductors are maintained at potentials φ_1 , φ_2 , and φ_3 , where for convenience it is assumed that $\varphi_1 > \varphi_2 > \varphi_3$. The symbol $\text{---}/\text{---}$ stands for a resistor whose resistance may be varied, and G is a galvanometer. The wires are taken to be of negligible resistance. If the resistors R_1 and R_2 are adjusted so that there is no current through G ,



then $\varphi_{\text{probe}} = \varphi_c$, and the same current I' passes through both R_1 and R_2 . In these circumstances,

$$\varphi_{\text{probe}} = \varphi_1 - I'R_1 = \varphi_3 + I'R_2, \text{ or } \varphi_{\text{probe}} = \frac{\varphi_1 - (\varphi_1 - \varphi_3)R_1}{(R_1 + R_2)}.$$

or

$$g_1 E_{1n} = g_2 E_{2n} \quad (7-18b)$$

This equation is the analog to the equation for the continuity of D_n across dielectric interfaces in electrostatic problems.

Since the field is static in each medium,

$$\oint \mathbf{E} \cdot d\mathbf{l} = 0$$

for a closed path that links both media, and

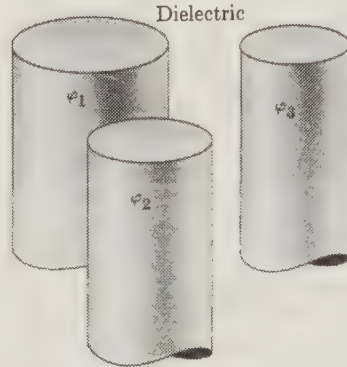
$$E_{1t} = E_{2t} \quad (7-19)$$

by the derivation of Section 4-7. This equation is evidently the same for both types of problems (electrostatic and steady conduction).

An illustration of the ideas presented above is found in the "electrolytic tank" shown in Fig. 7-5. Here a number of metal electrodes, which are connected to external sources of potential, are placed in a liquid conducting medium (ideally of infinite extent) with moderate conductivity (such as a salt solution). Since the conductivity of the salt solution is *much* smaller

FIGURE 7-6

The equivalent electrostatic problem to the conduction problem of the preceding figure. Since Fig. 7-5 depicted two-dimensional conduction, the electrostatic problem is also two-dimensional, and each conductor is an infinitely long cylinder.



than that of a metal (see Table 7-1), the electric field in the metal (for the same current density) is *much* smaller than that in the solution. The ratio of fields is so small that \mathbf{E} in the metal may be neglected, and each metal electrode may be assumed to be an equipotential surface. A small probe may be used, as shown in Fig. 7-5, to explore the potential in the solution, and in this way, a plot of the equipotential surfaces can be made. A possible use of this experimental approach is that it provides a numerical solution to Laplace's equation that, in the case of complicated geometry, might be difficult to determine analytically. The solution found is not limited to the conduction problem but applies equally well to the equivalent electrostatic problem in which the same metal electrodes are surrounded by a dielectric medium (Fig. 7-6). The "electrolytic tank" method was extensively used in the past to solve two-dimensional electrostatic problems with complicated geometry. Today, however, such problems are usually solved on the computer.

As a second illustration of the relation between conduction and electrostatics, we consider two metal electrodes in an infinite, homogeneous, ohmic medium of moderate conductivity g . If the electrodes are maintained at the potentials φ_1 and φ_2 , the current I between them is

$$I = \frac{\varphi_1 - \varphi_2}{R}$$

which defines the resistance R between the two electrodes.* This current may be written in terms of the current density \mathbf{J} in the medium:

* Since I is analogous to Q in the electrostatic problem, I is proportional to $\Delta\varphi$, and $1/R$ is defined as the proportionality constant. See Eq. (7-22).

$$I = \oint_S \mathbf{J} \cdot \mathbf{n} da$$

where S is any closed surface that completely surrounds one of the electrodes (except for an insulated metal wire to lead current onto the electrode so as to maintain its potential constant). But

$$\mathbf{J} = g\mathbf{E}$$

Combining the last three equations, we obtain

$$\frac{\varphi_1 - \varphi_2}{R} = g \oint_S \mathbf{E} \cdot \mathbf{n} da \quad (7-20)$$

If the identical electric field were produced by electrostatic charges on the two metal electrodes in a *dielectric* medium, then by Gauss's law

$$\oint_S \mathbf{E} \cdot \mathbf{n} da = \frac{1}{\epsilon} Q \quad (7-21)$$

where Q is the charge on the electrode surrounded by the surface S and ϵ is the permittivity of the medium. In these circumstances, the two electrodes form a capacitor:

$$Q = C(\varphi_1 - \varphi_2) \quad (7-22)$$

Insertion of Eqs. (7-21) and (7-22) into Eq. (7-20) yields

$$RC = \frac{\epsilon}{g} \quad (7-23)$$

This result relates the resistance between two conductors in a weakly conducting medium and the capacitance of the equivalent electrostatic problem.

This relation is in fact more than an analogy between conducting and dielectric media. It also holds for any single medium with conductivity g and dielectric constant ϵ . Since there is no ideal dielectric, every real dielectric has some nonvanishing g , no matter how small. From the other extreme, even a good conductor has its own value of ϵ . Thus, a capacitor has a leakage resistance, and a resistor has a small associated capacitance, and in each case R and C for the device are related by Eq. (7-23) (approximately, since the medium is not infinite).

7-5 APPROACH TO ELECTROSTATIC EQUILIBRIUM

In Chapter 2, it was shown that the excess charge on a conductor resides on its surface. This situation is the equilibrium state. The approach to

equilibrium was not studied in Chapter 2, but it was stated that for good (metallic) conductors equilibrium is attained extremely rapidly. The poorer the conductor, the slower is the approach to electrostatic equilibrium. In fact, if the conductivity of the material is extremely low, it may take years or even longer for electrostatic equilibrium to obtain.

Consider a homogeneous, isotropic medium characterized by conductivity g and permittivity ϵ , which has a volume density of prescribed charge $\rho_0(x, y, z)$. If this conducting system is suddenly isolated from applied electric fields, it will tend toward the equilibrium situation where there is no excess charge in the interior of the system. According to the equation of continuity,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0 \quad (7-9)$$

which, with the aid of Ohm's law, becomes

$$\frac{\partial \rho}{\partial t} + g \nabla \cdot \mathbf{E} = 0 \quad (7-24)$$

But $\nabla \cdot \mathbf{E}$ is related to the sources of the field. In fact, $\nabla \cdot \mathbf{E} = \rho/\epsilon$, so that

$$\frac{\partial \rho}{\partial t} + \frac{g}{\epsilon} \rho = 0 \quad (7-25)$$

The solution to this partial differential equation is, for constant g and ϵ ,

$$\rho(x, y, z, t) = \rho_0(x, y, z) e^{-gt/\epsilon} \quad (7-26)$$

and it is seen that the equilibrium state is approached exponentially.

From Eq. (7-26) it is evident that the quantity ϵ/g has the dimensions of time, and it is called the *time constant* or relaxation time t_c of the medium:

$$t_c = \frac{\epsilon}{g} = \epsilon \eta \quad (7-27)$$

The time constant is a measure of how fast the conducting medium approaches electrostatic equilibrium. More precisely, it is the time required for the charge in a specified region to decrease to $1/e$ of its original value.

A material will display conductorlike behavior in a particular applied field if its time constant is much shorter than the time characterizing a change in the applied field. For some applications a time constant of less

than 0.1 second is sufficient, and since most nonmetallic permittivities* fall into the range ϵ_0 to $10\epsilon_0$, a material with resistivity less than 10^9 or $10^{10} \Omega \cdot \text{m}$ is required. For high-frequency applications, a shorter time constant, and a correspondingly smaller resistivity, is required for true conductorlike behavior. In fact, the condition is

$$t_c \ll \frac{1}{f}$$

where f is the highest frequency involved in the experiment. Just the opposite condition holds for dielectriclike behavior.

7-6 RESISTANCE NETWORKS AND KIRCHHOFF'S LAWS

Thus far, we have discussed conduction primarily from the point of view of charge transport in a conducting medium and have approached the problem in terms of the differential equations that must apply at each point. In these cases, the important quantity to be determined is the current density, \mathbf{J} . But in many problems of practical interest, the electric charge carriers are constrained to follow a low resistance path called a *circuit*, and then the quantities of interest are the *currents* in each part of the circuit. In this section, we limit the discussion to circuits carrying steady currents—that is, to *direct current circuits*. A circuit may consist of several different branches, and in fact, a possible definition of a circuit is a network of conducting paths, each of which may contain applied voltages. The central problem of circuit analysis is, *given the resistance and applied voltage in each circuit branch, to find the current in each of these branches*.

In Chapter 2, it was shown that the integral of the tangential component of an electrostatic field around any closed path vanishes. That is, for an electrostatic field,

$$\oint \mathbf{E} \cdot d\mathbf{l} = 0 \quad (7-28)$$

For an ohmic material, $\mathbf{J} = g\mathbf{E}$. In the general isotropic case, this expression is modified to $\mathbf{J} = g(E)\mathbf{E}$, but $g(E)$ is always a positive quantity, so that Eq. (7-28) implies that $\oint \mathbf{J} \cdot d\mathbf{l} = 0$. Thus, it follows that a purely electrostatic force cannot cause a current to circulate in the same sense around an entire circuit. Or, in other words, a steady current cannot be maintained solely by means of electrostatic forces. A charged particle q may, however, experience other forces (mechanical, “chemical,” etc.) in

* We cannot apply Eq. (7-27) to a metal since we do not know the proper value of ϵ to use; in fact, $t_c \approx \tau \approx 10^{-14}$ s, where τ is the collision time to be discussed in Section 7-7. As will be seen, for times shorter than τ the assumption that $\mathbf{J} = g\mathbf{E}$ is not valid.

addition to the macroscopic electrostatic force, so that in part of the circuit the charges move opposite to the direction of \mathbf{E} . In previous sections, we sidestepped the question concerning the cause of electric current by assuming that two points on a conducting object were maintained at the constant potential difference $\Delta\varphi$ by means of external energy sources. It is still sufficient for our purpose here to assume the existence of such *applied voltages*,* but we shall make a brief digression to discuss how they may in fact be produced.

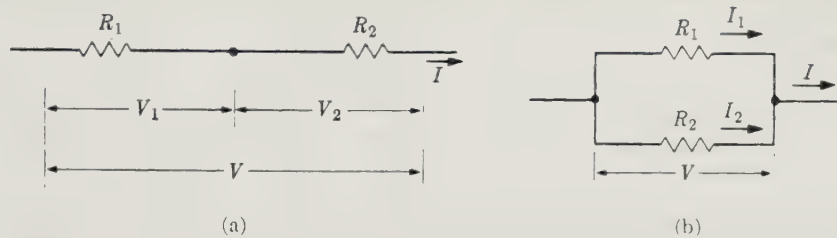
In the laboratory, a steady voltage is usually produced either by a battery or by an electronic power supply (which rectifies and smooths the line voltage), but it could be produced by a variety of other means, for example, a Van de Graaff generator. The last is conceptually the simplest case to analyze. In the Van de Graaff generator, charges are literally put onto a conveyor belt at one terminal and forcibly carried to another location of *higher* potential energy at the other terminal where they are removed from the belt. In steady-state operation, $\oint \mathbf{E} \cdot d\mathbf{l} = 0$ around any closed path. For example, the integral is negative along the belt and equally positive along an outside path between the terminals. A steady outside current can be allowed to flow through a resistance connected between the terminals. The power input is simply the mechanical power necessary to drive the belt that is transporting charges in opposition to an electric field. The operation of a battery is similar (except that the “forces” at work in a battery depend on the quantum mechanics of electrochemistry), and $\oint \mathbf{E} \cdot d\mathbf{l} = 0$ around any closed path, even one passing through the electrolyte in the battery. What is important for the purpose of circuit analysis, however, is simply that $\oint \mathbf{E} \cdot d\mathbf{l} = 0$ around a closed path containing the *terminals* of the voltage source—one leg of the path through the resistance network and the other directly across the terminals but *external* to the source. The purpose of electric circuit theory is to develop a procedure for analyzing the first leg, through the resistance network. We do not need to analyze the cause (mechanical, chemical, or whatever) of the voltage difference between the terminals of the power source, but simply refer to it as an applied voltage, \mathcal{V} . An ideal source would provide an applied voltage \mathcal{V}_0 that is independent of the current drawn from the source, but the terminal voltage of a real source depends to some degree on the current, $\mathcal{V} = \mathcal{V}(I)$. The simplest assumption, which is usually applicable, is that the dependence is linear:

$$\mathcal{V} = \mathcal{V}_0 - R_I I$$

* The applied voltage is usually called an *electromotive force* (or *emf*) in other books, although applied voltage is the term used in the laboratory for this potential difference. The historical term *emf* and the concept itself are rather confusing and unnecessary, so they will not be used here. We reserve the term *emf* for a somewhat different concept to be introduced later (Chapter 11).

FIGURE 7-7

(a) Series and (b) parallel connection of two resistors



The coefficient R_i is called the *internal resistance* and \mathcal{V}_0 is called the *open-circuit voltage* (or the emf, in most other books).

Before proceeding to the general network problem, we first review the elementary series and parallel connections of resistors. The resistance defined in Section 7-3 is a property of the material object under consideration, and it depends upon both the nature of the material from which the object is composed and its geometry. (The resistivity, on the other hand, depends only upon the nature of the conducting material.) A conducting object of convenient shape that is characterized primarily by its resistance is called a *resistor*, which is usually denoted by the symbol \sim . Resistors may be connected to form a resistance network, and the ways in which two resistors may be combined are illustrated in Fig. 7-7. Figure 7-7a shows a *series* connection, in which the same current I passes through both resistors. Applying Eq. (7-15) to each resistor and noting that the potential difference* $V = V_1 + V_2$, we find that

$$V = R_1 I + R_2 I = (R_1 + R_2) I$$

Thus, the equivalent resistance of the combination is

$$R = R_1 + R_2 \quad (\text{series connection}) \quad (7-29)$$

In the *parallel* connection (Fig. 7-7b), the potential difference across each resistor is the same, and the total current through the combination is $I = I_1 + I_2$. Applying Eq. (7-15), we find

$$I = \frac{V}{R_1} + \frac{V}{R_2} = \left(\frac{1}{R_1} + \frac{1}{R_2} \right) V$$

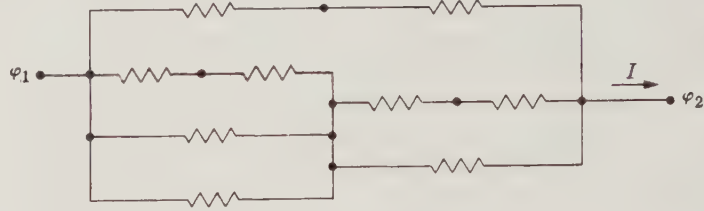
and the equivalent resistance R of the combination is obtained from

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} \quad (\text{parallel combination}) \quad (7-30)$$

* In this section we use the symbol V instead of $\Delta\phi$ for a potential difference, so as to agree with the most common electric circuit notation.

FIGURE 7-8

A resistor network



The equivalent resistance of a more complicated network like that in Fig. 7-8 can be determined by combining the resistors in pairs according to Eq. (7-29) or (7-30), and then repeating the process until only one equivalent resistance remains. While this procedure is not possible for every network, every two-terminal network can be reduced to one equivalent resistance by the procedure of the following paragraph.

Any network problem can be solved in a systematic way by means of two rules known as Kirchhoff's laws.* Before stating these laws, we define two terms. A *branch point* is a point of the circuit where three or more conductors are joined together, such as point *a*, *b*, *c*, or *d* in Fig. 7-9. A *loop* is any closed conducting path in the network.

Kirchhoff's laws may now be stated.

- I. *The algebraic sum of the currents flowing toward a branch point is zero:*

$$\sum I_j = 0 \quad (\text{I})$$

- II. *The algebraic sum of the voltage differences around any loop of the network is zero:*

$$\sum V_j = 0 \quad (\text{II})$$

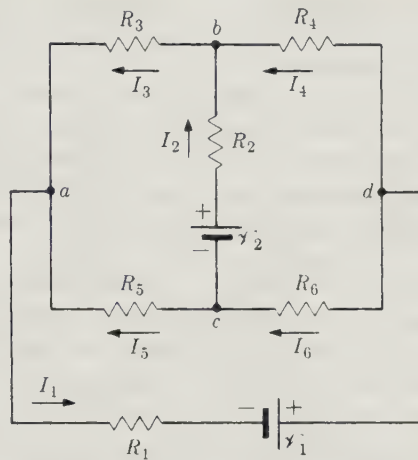
The first law is just a formal statement of the fact that charge does not accumulate at a branch point in the circuit as a result of the steady current. It is a restatement of the equation of continuity in the form of Eqs. (7-6) and (7-7), and as such it is equivalent to

$$\nabla \cdot \mathbf{J} = 0 \quad (\text{steady currents}) \quad (7-16)$$

* Named for Gustav Robert Kirchhoff (1824–1887).

FIGURE 7-9

A typical circuit requiring the application of Kirchhoff's laws. The symbol \mathcal{V} is used to designate an applied voltage. In a typical circuit problem, the \mathcal{V} 's and R 's are specified, and the currents are to be found. Two of the six equations for the currents in the above circuit are $-I_1 + I_3 + I_5 = 0$ and $\mathcal{V}_1 = I_6 R_6 + I_5 R_5 + I_1 R_1$.



The second law is merely a restatement of

$$\oint \mathbf{E} \cdot d\mathbf{l} = 0 \quad (\text{static fields}) \quad (7-28)$$

To apply Kirchhoff's laws, we need to recall Ohm's law:

The potential drop in a resistance R_j is

$$V_j = I_j R_j \quad (\text{resistor}) \quad (7-15)$$

where the higher potential is taken to be at the end at which the assumed current enters the resistance. Equation (7-15) is an equivalent form of

$$\mathbf{J} = g\mathbf{E} \quad (\text{linear medium}) \quad (7-10)$$

Finally, we label the applied voltages

$$V_j = -\mathcal{V}_j \quad (\text{applied voltage})$$

Combining this and Eq. (7-15), we can rewrite Kirchhoff's law II as

$$\sum \mathcal{V}_j = \sum I_j R_j \quad (\text{IIa})$$

If the internal resistances of the sources must be taken into account, they can be transferred to the right side of (IIa).

Before applying Kirchhoff's laws to a specific problem, it is necessary to assume directions for the currents in each of the branches. These directions should be indicated in the circuit diagram. The formulation of Eqs. (I) and (IIa) is then carried out on the basis of the assumed directions. If the numerical solution of these equations yields a negative value for a particular current, the correct direction of this current is opposite to that assumed. In

the problem illustrated in Fig. 7-9, there are six unknown currents. These are designated by the symbols I_1 , I_2 , I_3 , I_4 , I_5 , and I_6 , each having been given an assumed direction.

Kirchhoff's law I may be applied at each branch point of the circuit, but the equations so obtained are not all independent. The general rule is that if there are n branch points, only $n - 1$ of these will produce independent equations. In the problem shown in Fig. 7-9, there are six unknown currents. The solution requires three branch-point equations and three loop equations.

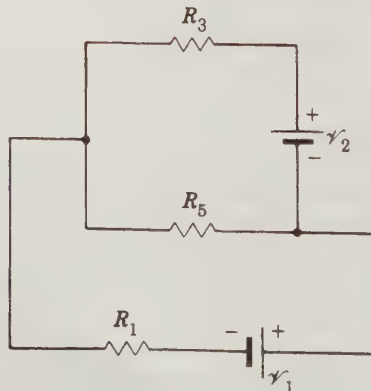
The summations in (I) and (IIa) are algebraic sums. In (I), the current is considered positive if its assumed direction points toward the branch point in question, or it is taken with the negative sign if its assumed direction points away from the junction. In applying the loop equations, some direction (either clockwise or counterclockwise) must be taken as the traversal direction. An applied voltage is taken with the positive sign if the voltage (by itself) would produce a positive current in the traversal direction. An IR term is taken with the positive sign if the current through the resistor in question is in the direction of traversal of the loop.

EXAMPLE 7-1
Current in a Resistive Network with Two Applied Voltages

Consider the resistive network shown in Fig. 7-10. The resistors R_1 , R_3 , and R_5 and the applied voltages \mathcal{V}_1 and \mathcal{V}_2 are specified. The problem posed is to find the currents in the resistors.

Solution: Figure 7-10 has been purposely drawn and labeled so that it conforms to the preceding figure, Fig. 7-9. We shall label the assumed currents, I_1 , I_3 , and I_5 through the resistors so that they also agree with

FIGURE 7-10
 Resistive network for Example 7-1



those in Fig. 7-9. Applying Kirchhoff's laws, we find

$$I_3 + I_5 - I_1 = 0$$

$$V_1 = I_1 R_1 + I_5 R_5$$

$$V_2 = I_3 R_3 - I_5 R_5$$

We use the first of these equations to eliminate I_3 from the last one. We then solve each of the remaining equations for I_5 and equate them. Solving for I_1 , we obtain

$$I_1 = \frac{V_2 + fV_1}{R_3 + fR_1}$$

where $f \equiv 1 + R_3/R_5$.

 7-7

MICROSCOPIC THEORY OF CONDUCTION

On the basis of a simple microscopic model of a conductor, it is possible to understand the linear behavior that is stated as Ohm's law, as well as some other experimental features of conduction. We consider a free particle of the medium, with charge q and mass m . Under the influence of the local electric force $q\mathbf{E}$, its drift velocity will increase according to $m d\mathbf{v}/dt = q\mathbf{E}$. If the charged particle were in a vacuum, it would continue to accelerate. In a material medium passing a steady current, however, the drift velocity is constant, and therefore, the total force on the particle must be zero. Another force, due to the medium, must be acting in addition to the electric force. The simplest possible assumption is that this retarding force is proportional to the velocity, so that the equation of motion is

$$m \frac{d\mathbf{v}}{dt} = q\mathbf{E} - G\mathbf{v} \quad (7-31)$$

It can be seen immediately that when $d\mathbf{v}/dt = 0$,

$$\mathbf{v}_d = \frac{q}{G} \mathbf{E} \quad (7-32)$$

This equation is the steady-state solution for the drift velocity. However, it is interesting to examine the complete solution of Eq. (7-31):

$$\mathbf{v}(t) = \frac{q}{G} \mathbf{E}(1 - e^{-Gt/m}) \quad (7-33)$$

if one makes the initial condition $\mathbf{v}(0) = 0$. This equation shows that the local drift velocity approaches its steady value exponentially, like $e^{-t/\tau}$, where the *relaxation time* τ is

$$\tau = \frac{m}{G} \quad (7-34)$$

Eliminating G between Eqs. (7-32) and (7-34), we find the steady-state drift velocity to be

$$\mathbf{v}_d = \frac{q\tau}{m} \mathbf{E} \quad (7-35)$$

Combining this with Eq. (7-4) for a single kind of charge carrier, we get the current density

$$\mathbf{J} = Nq\mathbf{v}_d = \frac{Nq^2\tau}{m} \mathbf{E} \quad (7-36)$$

proportional to the field according to Ohm's law. Comparison with Eq. (7-10) gives the conductivity

$$g = \frac{Nq^2\tau}{m} \quad (7-37)$$

or in case there are several kinds of charge carriers,

$$g = \sum \frac{N_i q_i^2 \tau_i}{m_i}$$

For a reasonably good electronic conductor such as a semiconductor or a metal (but not an electrolyte), we can interpret τ physically as the *mean time between collisions* of a conduction electron. In such a material, the electron accelerates for a short period, after which it makes a collision with one of the atoms of the material. As a result of this collision, the electron is thrown off in a random direction, so that the average effect of a collision is to reduce the drift velocity of the electron to zero again. If the mean collision time is τ , and the mean drift velocity is v , then the electrons lose the momentum mv_d after each time τ . In the steady state, the rate of momentum loss mv_d/τ is equated to the rate of momentum gain qE , and the result is identical to Eq. (7-35). The mean time τ is related to the *mean free path* of the electron by

$$l = v_T \tau \quad (7-38)$$

where v_T is the *thermal* velocity of the electrons. It is important to reemphasize that v_T is much greater* than the drift velocity v_d (though random in direction). For most metals, v_T is of the order of 10^6 m/s (nearly independent of temperature), and for a semiconductor, it is about an order of magnitude smaller at room temperature. On the other hand, the mean drift velocity v_d is not greater than about 10^{-2} m/s in normal metals. In metals and semiconductors, the mean free path is typically of the order of

* It is only because of this fact that τ can be taken to be independent of the accelerating field E .

10^{-8} m at room temperature, so that $\tau \approx 10^{-14}$ s in metals. In semiconductors, τ may be an order of magnitude larger. In either case, τ is also the time for the startup or decay of an Ohmic current. Thus, current changes practically instantaneously, on a local basis, after the field is applied or removed in resistors made of these materials. We note that in a metal the relaxation time for current decay τ , the collision time, and the time constant for dissipation of excess charge density t_c all happen to be the same, although they are conceptually different.*

It is evident from Table 7-1 that the group of materials with the highest electrical conductivity is that of the *metals*. These materials have high conductivity both because they contain a large density of charge carriers, of the order of one for each atom of the metal, and because the drift velocity per unit electric field is high. In metals, we deal with only one type of charge carrier, the electron. Hence, the conduction equations are simple in this case:

$$\mathbf{J} = -Ne\mathbf{v}_d \quad (7-39)$$

$$g = Ne(v/E) = Ne^2\tau/m \quad (7-40)$$

where e is the absolute value of the electronic charge. The drift velocity of the electron per unit electric field (v_d/E) is called the *mobility* of the electron. A large mobility implies a long collision time τ or, equivalently, a long mean free path. In order to get some feeling for the mean free path of electrons in a metal, we have to appeal to the dynamics of electron collisions. We know that the conductor is electrostatically neutral only on the average, that there are large variations in potential over distances of the order of one angstrom unit, and that a charged particle, such as an electron, ought to collide or be scattered by variations in potential. But we know also that the wave nature of the electron plays an important role in its motion on an atomic scale. A solution to the electron collision problem using wave-mechanical concepts is beyond the scope of this book. We merely state a result: *In a perfect crystal with a three-dimensional periodic potential, an electron wave makes no collision; its collision time τ is infinite.* Thus, the finite conductivity of metals arises from imperfections in the perfectly periodic structure. These imperfections are of two types: (1) impurities and geometric imperfections (such as grain boundaries in polycrystalline material) and (2) thermally induced imperfections arising from the thermal motion of the atoms in the structure. Both types often contribute independently to the resistivity (Matthiessen's rule), so that

$$\eta = \eta_1 + \eta_2(T) \quad (7-41)$$

where T is the absolute temperature.

* In a poor conductor, the collision time may have no meaning, or t_c may be *inversely* proportional to τ according to Eqs. (7-27) and (7-37).

In very pure metals, the dominant contribution to the resistivity at ordinary temperatures is the scattering of electron waves by thermally displaced atoms. Thus, $\eta \approx \eta_2(T)$. The scattering cross section of a displaced atom is proportional to the square of its vibration amplitude (x^2), in other words, to its maximum potential energy. Assuming elastic restoring forces operating on the displaced atoms,

$$(\text{Potential energy})_{\max} = (\text{Kinetic energy})_{\max} \propto kT$$

so that

$$\eta \approx \eta_2 \propto (\tau_2)^{-1} \propto x^2 \propto T \quad (7-42)$$

or, in words, the resistivity of a pure metal is proportional to the absolute temperature. The temperature coefficient of resistance, $(1/\eta) d\eta/dT$, for a very pure metal is, therefore,

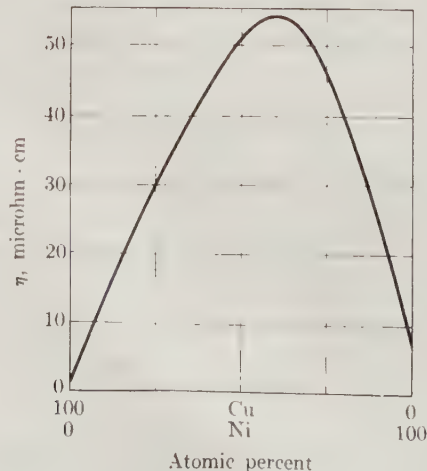
$$\alpha = \frac{1}{\eta} \frac{d\eta}{dT} \approx \frac{1}{T} \quad (7-43)$$

in approximate agreement with the metal entries of Table 7-1. Strictly speaking, the preceding argument is valid only for temperatures above the Debye temperature of the metal (the temperature above which all the atomic vibration modes are excited). At temperatures somewhat below the Debye temperature, η drops below the linear relationship predicted by Eq. (7-42). At very low temperatures the contribution from η_1 cannot be neglected.

The addition of small amounts of a soluble impurity always increases resistivity. An alloy, which may be regarded as an impure metal, always has a higher resistivity than that of the lower resistivity parent metal (Fig. 7-11).

FIGURE 7-11

Resistivity of copper-nickel alloys as a function of composition at 20°C



The temperature coefficient resistance, α , of an alloy is obviously lower than that of a pure metal just because its resistivity is higher, but certain alloys that have extremely small temperature coefficients have been developed.

7-8

SUMMARY

The most important technological uses of electricity depend on currents associated with moving charges; these are also of fundamental importance to magnetism, as described in the next chapter. Defined locally at a point in space, the **current density** is

$$\mathbf{J} = \sum N_i q_i \mathbf{v}_i$$

Since the charge density is

$$\rho = \sum N_i q_i$$

it may vanish even though the current does not. The total current through a surface S may be expressed as

$$I = \int_S \mathbf{J} \cdot \mathbf{n} \, da$$

or

$$I = \frac{dQ}{dt}$$

The conservation of charge is expressed locally by the **equation of continuity**

$$\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0$$

(For the present, we consider mainly steady currents, for which $\nabla \cdot \mathbf{J} = 0$, as well as $\partial \mathbf{J} / \partial t = 0$.) The conduction current in a medium is given by a constitutive equation, which in the simplest linear case defines the **conductivity** g :

$$\mathbf{J} = g \mathbf{E}$$

which is **Ohm's law**.

- The **resistance** of a straight conductor of uniform cross section is

$$R = l/gA$$

- In a continuous conducting medium with steady currents, the potential obeys the Laplace equation,

$$\nabla^2 \varphi = 0$$

The boundary conditions on \mathbf{E} are the same as in a dielectric medium, and those on \mathbf{J} are similar to the ones on \mathbf{D} . Consequently, between two

electrodes embedded in an infinite medium,

$$RC = \frac{\epsilon}{g} \quad (\text{mks units})$$

• If the volume charge density in a conducting medium is not zero initially, it disappears with a **time constant**

$$t_c = \frac{\epsilon}{g} \quad (\text{mks units})$$

For metals, this time is of the order of 10^{-14} s; for very poor conductors, it may be up to many months.

• For electric circuits, the static equations $\nabla \cdot \mathbf{J} = 0$ and $\nabla \times \mathbf{E} = 0$ become **Kirchhoff's laws**,

$$\sum I_j = 0 \text{ at a branch point} \quad (\text{I})$$

$$\sum V_j = 0 \text{ around a loop} \quad (\text{II})$$

In order to supply the power dissipated in resistors by steady currents, certain voltages must be applied by devices (e.g., batteries) whose operation cannot be described within the framework of electrostatics. Then, with Ohm's law, a complete solution of the circuit problem is straightforward.

• The microscopic theory of Ohmic conduction depends on the existence of a linear retarding force that acts on free charges in the medium, in addition to the accelerating electric force. Expressed in terms of a **relaxation time** τ , this gives

$$g = \frac{Nq^2\tau}{m}$$

The time τ is the time constant for the local establishment of an Ohmic current after the application of the field; in practical cases τ is short (10^{-14} s for metals). For good electronic conductors (metals, semiconductors), τ is interpreted as the mean time between collisions. In these cases, it depends on the electronic **mean free path** according to

$$\tau = l/v_T$$

where v_T is the random thermal velocity (not the net drift velocity).

PROBLEMS

7-1 The maximum rated current for a copper wire of cross-sectional area 2 mm^2 is 20 A. (a) What is the corresponding current density in A/m^2 ? (b) On the assumption that each copper atom contributes one conduction electron, calculate the electronic drift velocity corresponding to this current density. (Avogadro's number: $N_0 = 6.02 \times 10^{23}$ atoms per mole; atomic weight of copper: 63.5; density of copper:

8.92 g/cm³.) (c) Use the observed conductivity to calculate the average collision time for an electron in copper.

7-2 The conductivity of seawater is approximately $4.3 (\Omega\text{m})^{-1}$. Find the current density in a cell 2 cm long with 1 cm² cross-sectional area when 4 V is applied. Calculate the average drift velocity, assuming that the concentration of ions is 2 percent of the water molecule concentration.

7-3 Two infinite, plane, parallel plates of metal are separated by the distance d . The space between the plates is filled with two conducting media, the interface between the media being a plane that is parallel to the metal plates. The first medium (conductivity g_1 , permittivity ϵ_1) is of thickness a , and the second (conductivity g_2 , permittivity ϵ_2) is of thickness $d - a$. The metal plates are maintained at potentials φ_1 and φ_2 , respectively. In the steady state, (a) what is the potential of the interface separating the two media, and (b) what is the surface density of charge on this interface?

7-4 A system of charges and currents is completely contained *inside* the fixed volume V . The dipole moment of the charge-current distribution (see Section 2-9) is defined by

$$\mathbf{p} = \int_V \mathbf{r}\rho \, dv$$

where \mathbf{r} is the position vector from a fixed origin. Prove that

$$\int_V \mathbf{J} \, dv = \frac{d}{dt} \mathbf{p}$$

[*Hint:* First prove the identity $\int_V \mathbf{J} \, dv = \oint_S \mathbf{r}\mathbf{J} \cdot \mathbf{n} \, da - \int_V \mathbf{r}\nabla \cdot \mathbf{J} \, dv$, and note that \mathbf{J} vanishes on the surface S .]

7-5 A conducting medium is in a uniform field \mathbf{E}_0 . A spherical cavity of radius a is formed in the medium. (a) Find the potential inside and outside the cavity. (b) Find the surface charge that appears on the cavity. (c) Make a sketch of the field lines.

7-6 A parallel plate capacitor is filled with a material of dielectric constant K and conductivity g . It is charged with an initial charge Q . (a) Show that the charge leaks off the plates as an exponential function of time. (b) Show that the total joule heat production equals the electrostatic energy stored initially. (c) What is the time constant for the discharge if the material is silicon oxide? (See Tables 4-1 and 7-1.)

7-7 Two long cylindrical shells of metal (radii r_1 and r_2 , with $r_2 > r_1$) are arranged coaxially. The plates are maintained at the potential difference $\Delta\varphi$. (a) The region between the shells is filled with a medium of conductivity g . Use Ohm's law, $\mathbf{J} = g\mathbf{E}$, to calculate the electric current between unit lengths of the shells. (b) If the region between the shells is filled with a nonconducting medium of permittivity ϵ , the capacitance of the system may be computed from the definition $C = Q/\Delta\varphi$. Show explicitly for this geometry that the product of resistance per unit length and capacitance per unit length = ϵ/g .

7-8 The leakage resistance of a rubber cable insulation is measured in the following way: A length l of the insulated cable is immersed in a salt-water solution, a potential difference is applied between the cable conductor and the solution, and the resulting cable current is measured. In a particular case, 3 m of cable is immersed in

the solution; with 200 volts between cable conductor and solution, the current measured is 2×10^{-9} A. The insulation thickness is equal to the radius of the central conductor. What is the electrical resistivity of the insulation?

7-9 A long copper wire of radius a is stretched parallel to and at distance h from an infinite copper plate. The region above the plate and surrounding the wire is filled with a medium of conductivity g . Show that the electrical resistance between the two copper electrodes, per unit length of the wire, is given by

$$R = \frac{1}{2\pi g} \cosh^{-1} \frac{h}{a}$$

7-10 A homogeneous, isotropic sphere of conductivity g is subjected to a potential $\varphi_0 \cos \theta$ at all points on its surface. Here θ is the usual polar angle measured with respect to an axis through the center of the sphere. Determine the current density \mathbf{J} at all points inside the sphere.

7-11 Two cylindrical copper electrodes of radius a are oriented normal to a silicon wafer of thickness s and their axes are separated by the distance b . The electrodes are embedded in the wafer to the depth s ; in other words, they go completely through the wafer. The lateral dimensions of the wafer are large compared with b and may be considered infinite. Taking the conductivity of silicon to be g , find the current between the electrodes when their potential difference is $\Delta\varphi$.

***7-12** A square copper plate of length $20a$, thickness s , and conductivity g is subjected to a potential difference: Two opposite edges of the plate are maintained at the potentials φ_0 and $-\varphi_0$, respectively. (a) What is the electrical resistance of the plate? (b) If a small hole of radius a is drilled through the plate at its center, determine the *approximate* fractional change in resistance. [*Hint*: Find the potential distribution in the plate with the aid of the $\cos \theta$ cylindrical harmonics. Unfortunately, this distribution is not quite correct, because the two opposite edges of the square are not exact equipotentials. An approximate solution is obtained by taking the average potential of the two edges equal to $\pm\varphi_0$.]

7-13 Calculate the ratio of power dissipation to surface area for the conductors described in (a) Problem 7-1 and (b) Problem 7-2.

7-14 Given three resistors of 1Ω , 2Ω , and 3Ω , find the sixteen different resistances that can be made with these resistors.

7-15 A 0.4-watt lamp bulb is designed for operation with 2 volts across its terminals. A resistance R is placed in parallel with the bulb, and the combination is put in series with a 3-ohm resistor and a 3-volt battery (internal resistance, $\frac{1}{3}$ ohm). What should the value of R be if the lamp is to operate at design voltage?

7-16 In Problem 7-11, the radii of the copper electrodes are 0.5 mm, the axial separation of the electrodes $b = 1.5$ mm, $s = 3$ mm and η (the resistivity of Si) = $0.1 \Omega\text{m}$. If a 100 V battery is connected between the electrodes, determine the current passing through the disk.

7-17 Determine the length of nichrome wire, 0.2 mm in diameter, needed to dissipate 600 W when 120 V is applied to it.

***7-18** A resistance line, total resistance nR , is connected between the potential φ_0 and ground (ground is reference potential). The line is supported by $n - 1$ poles at equal resistance intervals such that the line resistance between poles is R . The

leakage resistance to ground at each pole is βR . If φ_m is the line potential at the m th pole, show that

$$\varphi_{m+1} - (2 + \beta^{-1})\varphi_m + \varphi_{m-1} = 0$$

7-19 Two batteries with open circuit voltages \mathcal{V}_1 and \mathcal{V}_2 and with internal resistances R_1 and R_2 , respectively, are connected in parallel with each other and with the load resistance R . (a) Find the current through the load. (b) If the load resistance is varied and other quantities kept fixed, what should R be in order that it dissipate maximum power?

7-20 A group of n identical cells of open circuit voltage \mathcal{V}_0 and internal resistance R_l is used to supply current to a load resistor R . Show that if the n cells are connected in series with each other and with R , then $I = n\mathcal{V}_0/(R + nR_l)$, whereas if the cells are connected in parallel and the combination put in series with R , then $I = \mathcal{V}_0/(R + R_l/n)$.

7-21 Six identical resistors (R) are joined to form a hexagon. Six more resistors (all again of the same resistance R) are connected between the six vertices and the center of the hexagon. (a) What is the equivalent resistance between opposite vertices; (b) between adjacent vertices?

7-22 Six resistors form the edges of a tetrahedron. Five of the resistors are identical (R), and the sixth is R_1 . A potential difference is applied across one of the resistors adjoining R_1 . Show that the joule heat production in R_1 is maximum when $R_1 = (5/3)R$.

7-23 A Wheatstone-bridge circuit is obtained from the circuit of Fig. 7-9 by making $\mathcal{V}_2 = 0$, and substituting a galvanometer R_g for R_2 . We shall also take $R_1 = 0$. The balance condition of the bridge (no current through the galvanometer) is obtained when $R_3R_6 = R_4R_5$. Thus, an unknown resistance, for example R_6 , may be determined in terms of known resistances: $R_6 = R_4R_5/R_3$ at balance. (a) Find the current through the galvanometer when the bridge is off balance. (b) Assume that the bridge is to be balanced by varying R_4 . The sensitivity of the bridge is defined by $S = CR_4(\partial I_2/\partial R_4)_0$, where C is the galvanometer deflection per unit current, and the subscript zero means that the derivative is to be evaluated at balance. Show that

$$S = \frac{C\mathcal{V}_1}{R_3 + R_4 + R_5 + R_6 + R_g(1 + R_5/R_6)(1 + R_4/R_3)}$$

***7-24** The Wheatstone bridge of the preceding problem is nearly balanced. Let $R_5/R_3 = \alpha$, and $R_6/R_4 = \alpha(1 - \epsilon)$, where $\epsilon \ll 1$. If the resistance R_g is negligible, show that $I_2/I_1 = \alpha\epsilon/(\alpha + 1)^2$.

***7-25** A resistance of approximately 10 ohms is to be measured in the Wheatstone-bridge circuit of Problem 7-23. A large selection of standard resistances is available. The maximum power allowed in the bridge is 5 W. If $R_g = 100 \Omega$, and the galvanometer will just detect a signal of 4×10^{-9} A, what is the highest precision one can obtain in measuring the unknown resistor? Assume that the standard resistors are exact, and do not limit the accuracy.

***7-26** A linear, conducting medium is connected at n points to electrodes with the fixed potentials: $\varphi_1, \varphi_2, \dots, \varphi_n$. Show that the joule heat production in the medium is given by $\sum_{i=1}^n \varphi_i I_i$, where I_i is the current *entering* the medium through electrode i .

The Magnetic Field of Steady Currents

The second kind of field that enters into the study of electricity and magnetism is the magnetic field. Such fields or, more properly, the effects of such fields have been known since ancient times, when the effects on the naturally occurring, permanent magnet magnetite (Fe_3O_4) were first observed. The discovery of the north- and south-seeking properties of this material had a profound influence on early navigation and exploration. Except for this application, however, magnetism was a little used and even less understood phenomenon until the early nineteenth century, when Oersted discovered that an electric current produced a magnetic field. This work, together with the later work of Gauss, Henry, Faraday, and others, brought the magnetic field into prominence as a partner to the electric field. The theoretical work of Maxwell and others (see Chapters 11 and 16) has shown that this partnership is real and that the electric and magnetic fields are inextricably intertwined. The efforts of practical people have resulted in the development of electrical machinery, communications equipment, and computers, which involve magnetic phenomena and which play such important roles in everyday life. In this chapter, the basic definitions of magnetism will be given, the production of magnetic fields by steady currents will be studied, and some important groundwork for future developments will be laid.

8-1

THE DEFINITION OF MAGNETIC INDUCTION

In Chapter 2, the Coulomb force on a charge q located at \mathbf{r} due to a charge q_1 at the origin was given by

$$\mathbf{F}_e = \frac{1}{4\pi\epsilon_0} \frac{qq_1}{r^2} \frac{\mathbf{r}}{r} \quad (8-1)$$

in which the assumption that the two charges are at rest is implicit.

If the charges are moving with constant velocities \mathbf{v} and \mathbf{v}_1 , respectively, an additional *magnetic force* \mathbf{F}_m is exerted on q by q_1 ,

$$\mathbf{F}_m = \frac{\mu_0 q q_1}{4\pi r^2} \mathbf{v} \times \left(\mathbf{v}_1 \times \frac{\mathbf{r}}{r} \right) \quad (8-2)$$

The number $\mu_0/4\pi$ plays the same role here as $1/4\pi\epsilon_0$ played in electrostatics—that is, it is the constant required to make an experimental law compatible with a set of units. In mks units, *by definition*,

$$\frac{\mu_0}{4\pi} = 10^{-7} \text{ N} \cdot \text{s}^2/\text{C}^2$$

exactly, and this stipulation leads to the primary definition of the coulomb. (See Section 8-3.) Also as in the case of the electrostatic force, it is convenient to abstract from the properties of the “test charge” by defining a *magnetic field*. In this case, not only the test charge q but also its velocity \mathbf{v} must be factored out:

$$\mathbf{F}_m = q\mathbf{v} \times \mathbf{B}, \quad (8-3)$$

where the *magnetic induction* \mathbf{B} is

$$\mathbf{B} = \frac{\mu_0 q_1}{4\pi r^2} \mathbf{v}_1 \times \frac{\mathbf{r}}{r} \quad (8-4)$$

If a number of moving source charges are present, the magnetic forces and fields are additive. Some sort of limiting process should be included in the definition of \mathbf{B} (as was done when the electric field was defined) to ensure that the test charge does not affect the sources of \mathbf{B} . The unit for magnetic induction in the mks system according to Eq. (8-3) is the newton-second per coulomb-meter, called the *tesla* (T).

If both an electric field and a magnetic field are present, the total force on a moving charge is $\mathbf{F}_e + \mathbf{F}_m$,

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (8-5)$$

This force is known as the *Lorentz force*.

The magnetic force between two charges is more complicated than the electric force because of the velocity dependence and the cross products. First, the similarities are that the magnitude of both forces depends on the product of the charges and the inverse square of their separation (in addition to a dimensional constant). The direction of the magnetic force, however, is not along the line joining the particles (i.e., it is not a central force), unless \mathbf{v} happens to be perpendicular to \mathbf{r} . The force is always in the

plane defined by \mathbf{r} and \mathbf{v}_1 . More importantly, the force is always perpendicular to \mathbf{v} ; from Eq. (8-3), $\mathbf{v} \cdot \mathbf{F}_m = 0$ for any field \mathbf{B} , so that a magnetic force never does any work on a charged particle. A further comparison between \mathbf{F}_m and \mathbf{F}_e is facilitated if we multiply numerator and denominator of Eq. (8-2) by ϵ_0 . Comparison of the result with Eq. (8-1) shows that $\epsilon_0\mu_0$ must have the dimensions of an inverse velocity squared. Let us write

$$\epsilon_0\mu_0 = \frac{1}{c^2} \quad (8-6)$$

where c has the dimensions of a velocity, so that

$$\mathbf{F}_m = \frac{1}{4\pi\epsilon_0} \frac{qq_1}{r^2} \frac{\mathbf{v}}{c} \times \left(\frac{\mathbf{v}_1}{c} \times \frac{\mathbf{r}}{r} \right)$$

Using the defined value of μ_0 and the experimental value of ϵ_0 , we find

$$c = 2.9979 \times 10^8 \text{ m/s}$$

which is numerically just the experimental velocity of light.* In Chapter 16, we shall see that this numerical coincidence is not accidental, but is a necessary consequence if light is an electromagnetic wave. Here we need not delve into the meaning of the relationship, but simply use the experimental fact. It means that for a given pair of particles

$$\frac{F_m}{F_e} \leq \frac{v}{c} \frac{v_1}{c}$$

That is, if the particle velocities are small compared to the velocity of light, the magnetic interaction is very much smaller than the electric interaction. In fact, Eqs. (8-1), (8-2), and (8-4) are only first approximations to the correct relativistic expressions to be derived in Chapter 21, and they hold only for $v_1 \ll c$. We may note that the fields produced by the uniformly moving charge q_1 are related by

$$\mathbf{B} = \frac{\mathbf{v}_1}{c} \times \frac{\mathbf{E}}{c}$$

* Equation (8-6) must hold in any consistent system of units. In gaussian units, where $\epsilon_0 = 1/4\pi$ by definition, $\mu_0/4\pi = 1/c^2$ is an experimental value. A more troublesome difference between the two systems of units is that in gaussian units the two c 's are split up with the two v 's, so that one defines

$$\mathbf{B} = \frac{q_1}{r^2} \frac{\mathbf{v}_1}{c} \times \frac{\mathbf{r}}{r} \quad \text{and} \quad \mathbf{F}_m = q \frac{\mathbf{v}}{c} \times \mathbf{B}$$

This definition has the advantage that \mathbf{B} is dimensionally the same as \mathbf{E} (and that the relativistic form v/c appears explicitly).

(This relation happens to hold for arbitrarily large velocities, even though \mathbf{E} and \mathbf{B} both become modified when v_1 is comparable to c .) Finally, it is remarkable that the magnetic force does not depend only on the *relative* velocity of the two charges—but is different in a moving coordinate system*—and that it does not simply change sign when the particle labels are exchanged. These aspects need not concern us now, however, since they cancel out in the applications that are to be made in this and the following chapters.

Since $F_m \ll F_e$, it might seem at first glance that the magnetic force could always be neglected in comparison to the electric one, but there are systems of particles where this is not so. Notably in a conduction current, where positive and negative charges are present in equal densities, the macroscopic electric field is zero, but the magnetic field of the moving charges is not. This is the case in electromagnets, motors, transformers, and other situations where magnetic forces have very great practical importance. We therefore begin with the study of magnetic interactions between conduction currents. In the next section, we discuss the force on a conduction current in an existing magnetic field, and in Section 8-3, we discuss the production of a magnetic field by a given conduction current.

8-2

FORCES ON CURRENT-CARRYING CONDUCTORS

From the Lorentz force (Eq. 8-5), an expression for the force on an element $d\mathbf{l}$ of a current-carrying conductor can be found. If $d\mathbf{l}$ is an element of the conductor with its sense taken in the direction of the current I that it carries, then $d\mathbf{l}$ is parallel to the drift velocity \mathbf{v} of the charge carriers in the conductor. If there are N charge carriers per unit volume in the conductor, the force on the element $d\mathbf{l}$ is

$$d\mathbf{F} = NA |d\mathbf{l}| q\mathbf{v} \times \mathbf{B} \quad (8-7)$$

where A is the cross-sectional area of the conductor and q is the charge per charge carrier. If several kinds of charge carriers are involved, then a summation must be included in Eq. (8-7). However, the final result, Eq. (8-8), is unchanged. Since \mathbf{v} and $d\mathbf{l}$ are parallel, an alternative form of Eq. (8-7) is

$$d\mathbf{F} = Nq |\mathbf{v}| A d\mathbf{l} \times \mathbf{B} \quad (8-7)$$

However, $Nq |\mathbf{v}| A$ is just the current for a single species of carrier.

* In particular, it vanishes in a coordinate system moving with \mathbf{v} . This dependence on the coordinate system contradicts the assumption of classical mechanics that forces are the same in all inertial coordinate systems. It is our first hint that the theory of relativity is necessary to accommodate electromagnetism.

Therefore, the expression

$$d\mathbf{F} = I d\mathbf{l} \times \mathbf{B} \quad (8-8)$$

is written for the force on an infinitesimal element of a current-carrying conductor.*

Equation (8-8) can be integrated to give the force on a complete (or closed) circuit. If the circuit in question is represented by the contour C ,

$$\mathbf{F} = \oint_C I d\mathbf{l} \times \mathbf{B} \quad (8-9)$$

So long as \mathbf{B} depends on position, the only simplification that can be made in Eq. (8-9) is to factor I from under the integral sign. If, however, \mathbf{B} is uniform—that is, independent of position—then it too can be removed from under the integral to give

$$\mathbf{F} = I \left\{ \oint_C d\mathbf{l} \right\} \times \mathbf{B}$$

The remaining integral is easy to evaluate. Since it is the sum of infinitesimal vectors forming a complete circuit, it must be zero. Thus,

$$\mathbf{F} = \oint_C I d\mathbf{l} \times \mathbf{B} = 0 \quad (\mathbf{B} \text{ uniform}) \quad (8-10)$$

Another interesting quantity is the torque on a complete circuit. Since torque is moment of force, the infinitesimal torque $d\boldsymbol{\tau}$ is given by

$$d\boldsymbol{\tau} = \mathbf{r} \times d\mathbf{F} = I \mathbf{r} \times (d\mathbf{l} \times \mathbf{B}) \quad (8-11)$$

The torque on a complete circuit is

$$\boldsymbol{\tau} = I \oint_C \mathbf{r} \times (d\mathbf{l} \times \mathbf{B}) \quad (8-12)$$

Once again, unless \mathbf{B} is uniform, no further simplification can be made. However, if it is uniform, a straightforward expansion is accomplished by writing

$$d\mathbf{l} \times \mathbf{B} = \mathbf{i}(dyB_z - dzB_y) + \mathbf{j}(dzB_x - dxB_z) + \mathbf{k}(dxB_y - dyB_x) \quad (8-13)$$

From these components, the components of $\mathbf{r} \times (d\mathbf{l} \times \mathbf{B})$ are readily found

* The original experiments that led to an understanding of magnetic forces were carried out on current-carrying circuits. These experiments produced Eqs. (8-8) and (8-9).

to be

$$\begin{aligned} [\mathbf{r} \times (d\mathbf{l} \times \mathbf{B})]_x &= y dx B_y - y dy B_x - z dz B_x + z dx B_z \\ [\mathbf{r} \times (d\mathbf{l} \times \mathbf{B})]_y &= z dy B_z - z dz B_y - x dx B_y + x dy B_x \\ [\mathbf{r} \times (d\mathbf{l} \times \mathbf{B})]_z &= x dz B_x - x dx B_z - y dy B_z + y dz B_y \end{aligned} \quad (8-14)$$

Since \mathbf{B} is assumed to be independent of \mathbf{r} (uniform field) the components of \mathbf{B} may be factored out of the integrals appearing in the expansion of Eq. (8-12). The spatial integrations that must be performed are of two general forms:

$$\oint \xi d\xi \quad (8-15a)$$

and

$$\oint \xi d\eta \quad (8-15b)$$

where ξ represents any coordinate and η represents any coordinate different from ξ . The first of these is trivial because it represents the integral from some lower limit ξ_1 to some upper limit ξ_2 of $\xi d\xi$, plus the integral from ξ_2 to ξ_1 of $\xi d\xi$. Since interchanging the limits introduces a minus sign, the result is zero, which eliminates six terms from Eq. (8-14). Integrals of the form (8-15b) involve only two variables, ξ and η ; hence, it makes no difference whether the integral is taken around the actual curve C or around its projection on the ξ, η -plane, as shown in Fig. 8-1. By using the projection on the ξ, η -plane it is easy to see what Eq. (8-15b) represents. In Fig. 8-2, the ξ, η -plane is shown along with the infinitesimal area $\xi d\eta$. The integral can be written

$$\oint \xi d\eta = \int_a^b \xi_1(\eta) d\eta + \int_b^a \xi_2(\eta) d\eta \quad (8-16)$$

FIGURE 8-1
Projection of the curve C
on the ξ, η -plane

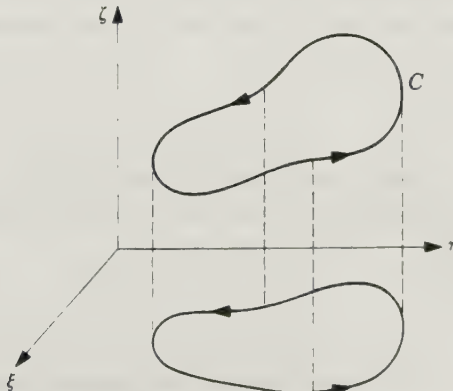
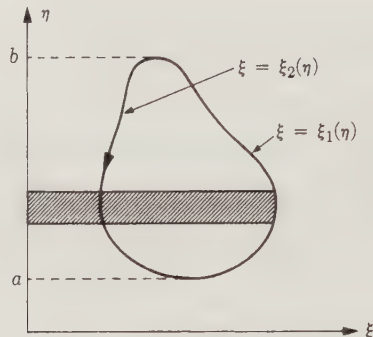


FIGURE 8-2

Evaluation of the integral
 $\int \xi d\eta$



This equation gives just the area enclosed by the projected curve, and in the figure, the result is positive. If ξ and η appear in cyclic order for a right-hand coordinate system, then the direction in which the contour is circled would give a normal in the positive ζ -direction. Thus, we may write

$$\oint \xi d\eta = A_{\zeta} \quad (8-17)$$

with ξ, η, ζ a cyclic permutation of x, y, z . Using this result to evaluate the integrals gives

$$\tau_x = I \oint_C [\mathbf{r} \times (d\mathbf{l} \times \mathbf{B})]_x = I(A_y B_z - A_z B_y) \quad (8-18)$$

with similar expressions for the y - and z -components. The three expressions are neatly summarized in the expression

$$\boldsymbol{\tau} = I \mathbf{A} \times \mathbf{B} \quad (8-19)$$

where \mathbf{A} is the vector whose components are the areas enclosed by projections of the curve C on the yz -, zx -, and xy -planes.* The quantity $I\mathbf{A}$ appears very frequently in magnetic theory and is referred to as the *magnetic dipole moment* of the circuit. The symbol \mathbf{m} will be used for magnetic dipole moment:

$$\mathbf{m} = I\mathbf{A} \quad (8-20)$$

with \mathbf{A} defined as in the text following Eq. (8-19).

It is easy to show, by the technique used above, that the integral of $\mathbf{r} \times d\mathbf{l}$ around a closed path gives twice the area enclosed by the curve:

$$\frac{1}{2} \oint_C \mathbf{r} \times d\mathbf{l} = \mathbf{A} \quad (8-21)$$

* Note that no restriction to plane curves has been imposed on C and that this definition of \mathbf{A} makes any such restriction unnecessary.

This expression can be used to obtain

$$\mathbf{m} = \frac{1}{2} I \oint_C \mathbf{r} \times d\mathbf{l} \quad (8-22)$$

as an alternative expression for the *magnetic dipole moment*.

If, instead of being confined to wires, the current exists in a medium, then the identification

$$I d\mathbf{l} \rightarrow \mathbf{J} dv \quad (8-23)$$

is appropriate, as has been shown earlier. We then write

$$d\mathbf{m} = \frac{1}{2} \mathbf{r} \times \mathbf{J} dv \quad (8-24)$$

which is useful in discussing the magnetic properties of matter.

8-3

THE LAW OF BIOT AND SAVART

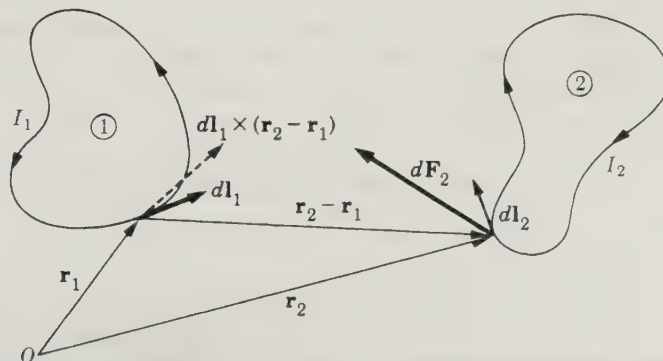
In 1820, just a few weeks after Oersted announced his discovery that currents produce magnetic effects, Ampere presented the results of a series of experiments that may be generalized and expressed in modern mathematical language as

$$\mathbf{F}_2 = \frac{\mu_0}{4\pi} I_1 I_2 \oint_1 \oint_2 \frac{d\mathbf{l}_2 \times [d\mathbf{l}_1 \times (\mathbf{r}_2 - \mathbf{r}_1)]}{|\mathbf{r}_2 - \mathbf{r}_1|^3} \quad (8-25)$$

This rather formidable expression can be understood with reference to Fig. 8-3. The force \mathbf{F}_2 is the force exerted on circuit 2 due to the influence of circuit 1, and the $d\mathbf{l}$'s and \mathbf{r} 's are explained by the figure. By definition,

$$\frac{\mu_0}{4\pi} = 10^{-7} \text{ N/A}^2$$

FIGURE 8-3
Magnetic interaction of two current circuits



in mks units, and Eq. (8-25) serves as the primary definition of the ampere, in terms of which the coulomb is defined. Equation (8-25) appears, superficially, to violate Newton's third law because of the lack of symmetry. However, by using some of the theorems of vector analysis, it can be shown that it is actually symmetric—that is, $\mathbf{F}_2 = -\mathbf{F}_1$. (See Problem 8-4.)

From Eq. (8-9), it is apparent that Eq. (8-25) implies

$$\mathbf{B}(\mathbf{r}_2) = \frac{\mu_0}{4\pi} I_1 \oint_1 \frac{d\mathbf{l}_1 \times (\mathbf{r}_2 - \mathbf{r}_1)}{|\mathbf{r}_2 - \mathbf{r}_1|^3} \quad (8-26)$$

This equation is a generalization of the *Biot and Savart law*,* which name will be used both for Eq. (8-26) and the differential form

$$d\mathbf{B}(\mathbf{r}_2) = \frac{\mu_0 I_1 d\mathbf{l}_1 \times (\mathbf{r}_2 - \mathbf{r}_1)}{4\pi |\mathbf{r}_2 - \mathbf{r}_1|^3} \quad (8-27)$$

Equation (8-27) is an immediate consequence of Eq. (8-4) as applied to a conductor, if one uses the same argument that led to Eq. (8-7). As a last point, Eqs. (8-26) and (8-27) take the forms

$$\mathbf{B}(\mathbf{r}_2) = \frac{\mu_0}{4\pi} \int_V \frac{\mathbf{J}(\mathbf{r}_1) \times (\mathbf{r}_2 - \mathbf{r}_1)}{|\mathbf{r}_2 - \mathbf{r}_1|^3} dv_1 \quad (8-28)$$

and

$$d\mathbf{B}(\mathbf{r}_2) = \frac{\mu_0 \mathbf{J}(\mathbf{r}_1) \times (\mathbf{r}_2 - \mathbf{r}_1)}{4\pi |\mathbf{r}_2 - \mathbf{r}_1|^3} dv_1 \quad (8-29)$$

for a continuous distribution of current described by the current density $\mathbf{J}(\mathbf{r})$.

It is an experimental observation that all magnetic induction fields can be described in terms of a current distribution. That is, \mathbf{B} always has the form of Eq. (8-28), with some $\mathbf{J}(\mathbf{r}_1)$. This observation implies that

$$\nabla \cdot \mathbf{B} = 0 \quad (8-30)$$

which, in turn, implies that there are no isolated magnetic poles. Equation (8-30) is true for any \mathbf{B} of the form (8-28) or (8-26), as can be verified mathematically: We take the divergence of Eq. (8-28). Using $\nabla \cdot (\mathbf{F} \times \mathbf{G}) = -\mathbf{F} \cdot \nabla \times \mathbf{G} + \mathbf{G} \cdot \nabla \times \mathbf{F}$ gives

$$\nabla_2 \cdot \mathbf{B}(\mathbf{r}_2) = -\frac{\mu_0}{4\pi} \int_V \mathbf{J}(\mathbf{r}_1) \cdot \nabla_2 \times \frac{\mathbf{r}_2 - \mathbf{r}_1}{|\mathbf{r}_2 - \mathbf{r}_1|^3} dv_1$$

* In passing, we mention that there has been some controversy over the naming of various laws. We do not wish to enter into this controversy, but refer the interested reader to E. T. Whittaker's excellent history, *History of the Theories of Aether and Electricity*, Vol. I (New York: Philosophical Library, 1951).

However, $(\mathbf{r}_2 - \mathbf{r}_1)/|\mathbf{r}_2 - \mathbf{r}_1|^3$ is the gradient of $-1/|\mathbf{r}_2 - \mathbf{r}_1|$ with respect to \mathbf{r}_2 . Since the curl of any gradient is zero, it follows that

$$\nabla_2 \cdot \mathbf{B}(\mathbf{r}_2) = 0$$

8-4

ELEMENTARY APPLICATIONS OF THE BIOT AND SAVART LAW

The range of problems to which Eq. (8-28) (or Eq. (8-26)) can be applied is limited primarily by the difficulty experienced in performing the integrations. Some of the tractable situations are considered in this section. In later sections, other techniques for obtaining \mathbf{B} will be considered.

EXAMPLE 8-1

The Magnetic Field of a Long, Straight Current-carrying Wire

The wire is imagined to lie along the x -axis from minus infinity to plus infinity and to carry a current I . The field is to be computed at a typical point \mathbf{r}_2 on the y -axis. The geometry is best explained by Fig. 8-4.

Solution: From the Biot-Savart law, the magnetic induction is

$$\mathbf{B}(\mathbf{r}_2) = \frac{\mu_0}{4\pi} I \int_{-\infty}^{\infty} \frac{dx \mathbf{i} \times (\mathbf{r}_2 - \mathbf{r}_1)}{|\mathbf{r}_2 - \mathbf{r}_1|^3} \quad (8-31)$$

Since $\mathbf{r}_2 - \mathbf{r}_1$ lies in the xy -plane,

$$\mathbf{i} \times (\mathbf{r}_2 - \mathbf{r}_1) = |\mathbf{r}_2 - \mathbf{r}_1| \sin \theta \mathbf{k} \quad (8-32)$$

Furthermore,

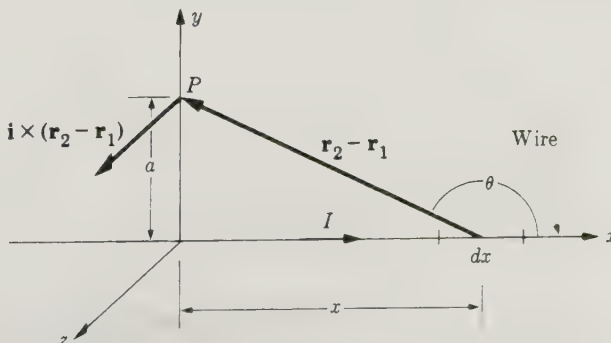
$$\frac{a}{x} = \tan(\pi - \theta) = -\tan \theta \quad (8-33)$$

and

$$|\mathbf{r}_2 - \mathbf{r}_1| = a \csc(\pi - \theta) = a \csc \theta \quad (8-34)$$

FIGURE 8-4

Magnetic field at point P due to a long straight wire



Using these relationships to convert Eq. (8-31) to an integral on θ from 0 to π gives

$$\mathbf{B}(\mathbf{r}_2) = \frac{\mu_0 I}{4\pi a} \int_0^\pi \sin \theta \, d\theta = \frac{\mu_0 I}{4\pi a} \mathbf{k}(-\cos \theta) \Big|_0^\pi = \frac{\mu_0 I}{2\pi a} \mathbf{k} \quad (8-35)$$

To use this result more generally, it is only necessary to note that the problem exhibits an obvious symmetry about the x -axis. Thus, we conclude that the lines of \mathbf{B} are circles everywhere, with the conductor as center. This is in complete agreement with the elementary result that gives the direction of \mathbf{B} by a right-hand rule.

EXAMPLE 8-2
Axial Magnetic Field of a
Circular Loop of Wire
Carrying a Current

Consider a circular loop of wire carrying a current I . The magnetic field produced by such a circuit at an arbitrary point is very difficult to compute; however, if only points on the axis of symmetry are considered, the expression for \mathbf{B} is relatively simple. In this example, a complete vector treatment will be used to demonstrate the technique. Figure 8-5 illustrates the geometry and the coordinates to be used. The field is to be calculated at point \mathbf{r}_2 on the z -axis; the circular turn lies in the xy -plane.

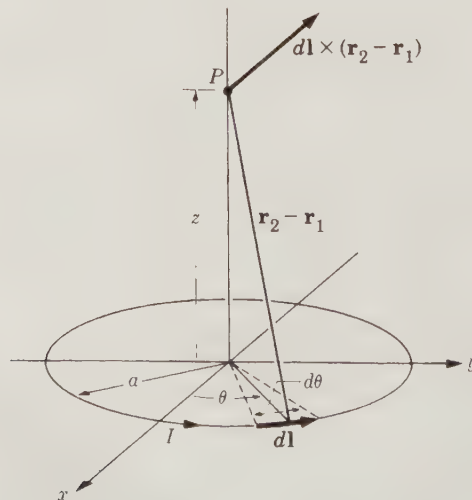
Solution: The magnetic induction is given by Eq. (8-26) in which, from Fig. 8-5, the following expressions are to be used:

$$d\mathbf{l} = a \, d\theta(-\mathbf{i} \sin \theta + \mathbf{j} \cos \theta)$$

$$\mathbf{r}_2 - \mathbf{r}_1 = -\mathbf{i}a \cos \theta - \mathbf{j}a \sin \theta + \mathbf{k}z \quad (8-36)$$

$$|\mathbf{r}_2 - \mathbf{r}_1| = (a^2 + z^2)^{1/2}$$

FIGURE 8-5
 Axial field of a circular
 turn of wire



Substituting these into Eq. (8-26) yields

$$\mathbf{B}(z) = \frac{\mu_0 I}{4\pi} \int_0^{2\pi} \frac{(i z a \cos \theta + \mathbf{j} z a \sin \theta + \mathbf{k} a^2)}{(z^2 + a^2)^{3/2}} d\theta \quad (8-37)$$

The first two terms integrate to zero, leaving

$$\mathbf{B}(z) = \frac{\mu_0 I}{2} \frac{a^2}{(z^2 + a^2)^{3/2}} \mathbf{k} \quad (8-38)$$

which is, of course, entirely along the z -axis.

A frequently used current configuration is the Helmholtz coil, which consists of two circular coils of the same radius, with a common axis, separated by a distance chosen to make the second derivative of \mathbf{B} vanish at a point on the axis halfway between the coils.

EXAMPLE 8-3
Axial Field of a
Helmholtz Coil

Figure 8-6 shows the configuration of a Helmholtz coil. We should like to determine the magnetic field at a point on the axis of the coil.

Solution: The magnetic induction at point P is

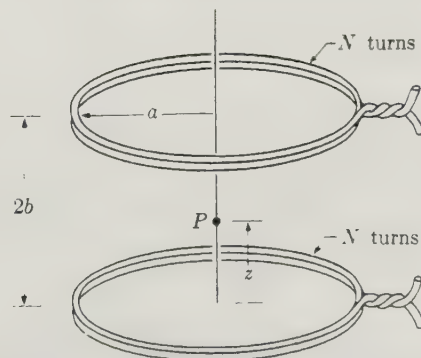
$$B_z(z) = \frac{N\mu_0 I a^2}{2} \left\{ \frac{1}{(z^2 + a^2)^{3/2}} + \frac{1}{[(2b - z)^2 + a^2]^{3/2}} \right\} \quad (8-39)$$

which is obtained by applying Eq. (8-38) to each of the two coils. The factor N is included to handle the situation where each coil contains N turns. The first derivative of B_z with respect to z is

$$\frac{dB_z}{dz} = \frac{\mu_0 N I a^2}{2} \left\{ -\frac{3}{2} \frac{2z}{(z^2 + a^2)^{5/2}} - \frac{3}{2} \frac{2(z - 2b)}{[(2b - z)^2 + a^2]^{5/2}} \right\} \quad (8-40)$$

At $z = b$, this derivative vanishes. The second derivative with respect

FIGURE 8-6
Axial field of a Helmholtz coil



to z is

$$\frac{d^2 B_z}{dz^2} = -\frac{3\mu_0 N I a^2}{2} \left\{ \frac{1}{(z^2 + a^2)^{5/2}} - \frac{5}{2} \frac{2z^2}{(z^2 + a^2)^{7/2}} \right. \\ \left. + \frac{1}{[(2b - z)^2 + a^2]^{5/2}} - \frac{5}{2} \frac{2(z - 2b)^2}{[(2b - z)^2 + a^2]^{7/2}} \right\}$$

At $z = b$, this derivative reduces to

$$\frac{d^2 B_z}{dz^2} \Big|_{z=b} = -\frac{3\mu_0 N I a^2}{2} \left\{ \frac{b^2 + a^2 - 5b^2 + b^2 + a^2 - 5b^2}{(b^2 + a^2)^{7/2}} \right\} \quad (8-41)$$

which vanishes if $a^2 - 4b^2 = 0$. Thus, the appropriate choice for b is

$$2b = a \quad (8-42)$$

That is, the coil separation should equal the radius. With this separation, the magnetic induction at the midpoint is

$$B_z = \frac{\mu_0 N I}{a} \frac{8}{5^{3/2}} \quad (8-43)$$

Helmholtz coils play an important role in scientific research, where they are frequently used to produce a relatively uniform magnetic field over a small region of space. Let us consider the magnetic field at a point on the axis near the midpoint between the coils. The field $B_z(z)$ can be developed in a Taylor's series about the point $z = \frac{1}{2}a$:

$$B_z(z) = B_z(\frac{1}{2}a) + (z - \frac{1}{2}a) \frac{\partial B_z}{\partial z} \Big|_{z=\frac{1}{2}a} + \dots$$

Since the first three derivatives vanish,

$$B_z(z) = B_z(\frac{1}{2}a) + \frac{1}{24}(z - \frac{1}{2}a)^4 \frac{\partial^4 B_z}{\partial z^4} \Big|_{z=\frac{1}{2}a} + \dots$$

If the fourth derivative is evaluated explicitly, $B_z(z)$ can be written as

$$B_z(z) = B_z(a/2) \left\{ 1 - \frac{144}{125} \left(\frac{z - a/2}{a} \right)^4 \right\} \quad (8-44)$$

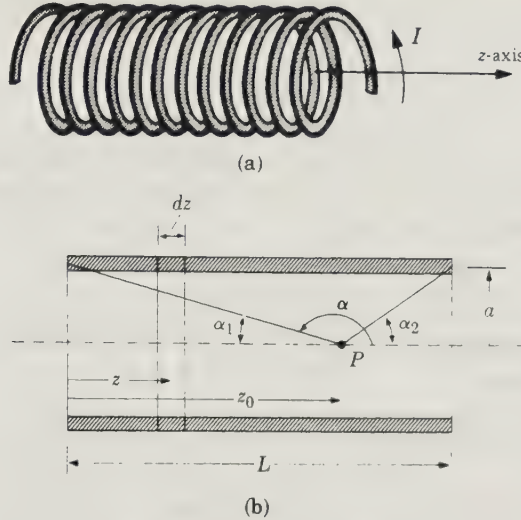
Thus, for the region where $|z - a/2|$ is less than $a/10$, $B_z(z)$ deviates from $B_z(a/2)$ by less than one and a half parts in ten thousand.

The tesla is a rather large unit for measuring most laboratory fields. Consequently, the unit *gauss* for B from the gaussian system* of units is commonly used: one gauss equals 10^{-4} tesla. For reference purposes,

* This system of units is discussed in Appendix III.

FIGURE 8-7

(a) A solenoid; (b) axial magnetic field of a solenoid



we give

$$B_z = \frac{32\pi N I}{5^{3/2} a^{3/2} 10}, \quad I \text{ in ampere, } a \text{ in cm, } B \text{ in gauss} \quad (8-43a)$$

for the induction at the midpoint of the Helmholtz coil. Of course N is still the number of turns in each of the two coils.

EXAMPLE 8-4
Axial Field of a Solenoid

Another device to which Eq. (8-38) can be applied is the solenoid. A solenoid may be described as N turns uniformly wound on a cylindrical form of radius a and length L . Such a configuration is shown in Fig. 8-7(a).

Solution: The magnetic induction at point z_0 is found by dividing the length L into elements dz , such as the one shown in Fig. 8-7(b), and applying Eq. (8-38) to each element and summing the results. Noting that the element dz contains $N dz/L$ turns, we find that

$$B_z(z_0) = \frac{\mu_0 N I a^2}{L} \int_0^L \frac{dz}{2 [(z_0 - z)^2 + a^2]^{3/2}} \quad (8-45)$$

The change of variable, $z - z_0 = a \cot \alpha$, leads to

$$\begin{aligned} B_z(z_0) &= \frac{\mu_0 N I}{2L} \int_{\alpha_2}^{\pi - \alpha_1} \sin \alpha \, d\alpha \\ &= \frac{\mu_0 N I}{2L} [-\cos(\pi - \alpha_1) + \cos \alpha_2] \end{aligned} \quad (8-46)$$

The angles α_1 and α_2 (both $< \pi/2$) are shown in Fig. 8-7(b). Equation (8-46) thus becomes

$$B_z(z_0) = \frac{\mu_0 NI}{L} \left[\frac{\cos \alpha_1 + \cos \alpha_2}{2} \right] \quad (8-47)$$

If the solenoid is long compared with its radius and z_0 is not too close to either zero or L , then α_1 and α_2 are both small angles and may be approximated by

$$\alpha_1 \cong \frac{a}{z_0}, \quad \alpha_2 \cong \frac{a}{L - z_0} \quad (8-48)$$

Maintaining quadratic terms in the expansion of $\cos \alpha_1$ and $\cos \alpha_2$, we obtain

$$B_z(z_0) \cong \frac{\mu_0 NI}{L} \left\{ 1 - \frac{a^2}{4z_0^2} - \frac{a^2}{4(L - z_0)^2} \right\} \quad (8-49)$$

From this, we conclude that if $z_0 = L/2$ and $L/a = 10$, a 2 percent error results from neglecting the quadratic terms and just using the elementary formula.

The general formula for \mathbf{B} , Eq. (8-26), is always a correct expression for the magnetic field of a current circuit, but it is difficult to use in the general case where the geometry is complex.

8-5
AMPERE'S CIRCUITAL LAW

For magnetic induction fields given by Eq. (8-26) or (8-28) that are due to steady currents—that is, to currents that satisfy

$$\nabla \cdot \mathbf{J} = 0, \quad (8-50)$$

a very important equation for the curl of \mathbf{B} can be derived by simply calculating the curl of Eq. (8-28). The curl involves differentiation with respect to \mathbf{r}_2 and hence operates only on the factor $(\mathbf{r}_2 - \mathbf{r}_1)/|\mathbf{r}_2 - \mathbf{r}_1|^3$:

$$\begin{aligned} \nabla_2 \times \mathbf{B}(\mathbf{r}_2) &= \frac{\mu_0}{4\pi} \int_V \left[\mathbf{J}(\mathbf{r}_1) \left(\nabla_2 \cdot \frac{\mathbf{r}_2 - \mathbf{r}_1}{|\mathbf{r}_2 - \mathbf{r}_1|^3} \right) \right. \\ &\quad \left. - \mathbf{J}(\mathbf{r}_1) \cdot \nabla_2 \frac{\mathbf{r}_2 - \mathbf{r}_1}{|\mathbf{r}_2 - \mathbf{r}_1|^3} \right] dv_1 \end{aligned}$$

The derivative may now be changed to differentiation with respect to \mathbf{r}_1 (with a minus sign) in the second term because of the symmetry between \mathbf{r}_2 and \mathbf{r}_1 :

$$\nabla_2 \times \mathbf{B}(\mathbf{r}_2) = \frac{\mu_0}{4\pi} \int_V \left[\mathbf{J}(\mathbf{r}_1) 4\pi \delta(\mathbf{r}_2 - \mathbf{r}_1) - \mathbf{J}(\mathbf{r}_1) \cdot \nabla_1 \frac{\mathbf{r}_1 - \mathbf{r}_2}{|\mathbf{r}_1 - \mathbf{r}_2|^3} \right] dv_1$$

The first term is expressed in terms of the Dirac delta function, as in Eq. (2-57); it integrates immediately to give $\mu_0 \mathbf{J}(\mathbf{r}_2)$. The second term can be

shown to vanish by means of an integration by parts:

$$\nabla_1 \cdot \left(\mathbf{J} \frac{x_1 - x_2}{|\mathbf{r}_1 - \mathbf{r}_2|^3} \right) = \frac{x_1 - x_2}{|\mathbf{r}_1 - \mathbf{r}_2|^3} \nabla_1 \cdot \mathbf{J} + \mathbf{J} \cdot \nabla_1 \frac{x_1 - x_2}{|\mathbf{r}_1 - \mathbf{r}_2|^3}$$

for the x -component and similarly for the other components. The term with $\nabla \cdot \mathbf{J}$ vanishes by the assumption (8-50), and the volume integral of the left side can be converted using the divergence theorem to a surface integral, which vanishes if the surface is chosen to lie outside a bounded region where \mathbf{J} is nonvanishing. (The same result follows directly from the identity 1-2-4 of Table 1-2.) Thus, the final result, which will be called the *differential form* of Ampere's law, is

$$\nabla \times \mathbf{B}(\mathbf{r}_2) = \mu_0 \mathbf{J}(\mathbf{r}_2) \quad (8-51)$$

In Chapter 9, this equation will be modified so as to be more useful when magnetic materials are present. However, Eq. (8-51) is still valid so long as \mathbf{J} is the *total* current and $\nabla \cdot \mathbf{J} = 0$.

Stokes's theorem can be used to transform Eq. (8-51) into an integral form, which is sometimes very useful. This application of Stokes's theorem is written

$$\int_S \nabla \times \mathbf{B} \cdot \mathbf{n} \, da = \oint_C \mathbf{B} \cdot d\mathbf{l} \quad (1-45)$$

Using Eq. (8-51) for $\nabla \times \mathbf{B}$ gives *Ampere's circuital law*,

$$\oint_C \mathbf{B} \cdot d\mathbf{l} = \mu_0 \int_S \mathbf{J} \cdot \mathbf{n} \, da \quad (8-52)$$

This equation simply says that the line integral of \mathbf{B} around a closed path is equal to μ_0 times the total current through the closed path.

It is instructive to verify Eq. (8-52) for a simple case. The long straight wire provides a particularly good example. In this case \mathbf{B} at a distance r from the conductor is given by $B(r) = \mu_0 I / 2\pi r$, and it is tangential to a circle of radius r with center at the conductor. Figure 8-8 illustrates the geometry. The current is directed upward, and C is described in the counterclockwise direction. From the figure,

$$\mathbf{B} \cdot d\mathbf{l} = |\mathbf{B}| |d\mathbf{l}| \cos \chi = |\mathbf{B}| r \, d\theta \quad (8-53)$$

With $|\mathbf{B}|$ as given above,

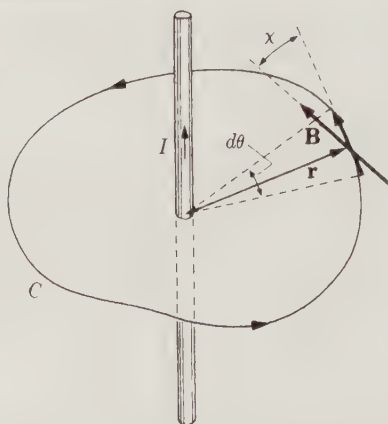
$$\oint_C \mathbf{B} \cdot d\mathbf{l} = \int_0^{2\pi} \frac{\mu_0 I}{2\pi r} r \, d\theta = \mu_0 I \quad (8-54)$$

which represents a special case of Eq. (8-52).

Ampere's circuital law is in many ways parallel to Gauss's law in electrostatics. That is, it can be used to obtain the magnetic field due to a certain current distribution of high symmetry without having to evaluate the complicated integrals that appear in the Biot law. An example follows.

FIGURE 8-8

Verification of Ampere's circuital law for long, straight-wire geometry



EXAMPLE 8-5
Magnetic Field of a Coaxial Cable Conductor

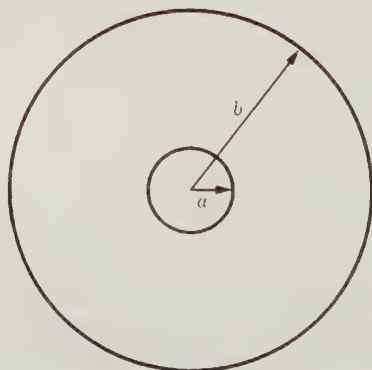
Consider a coaxial cable consisting of a small center conductor of radius a and a coaxial cylindrical outer cable conductor of radius b , as shown in Fig. 8-9. Assume that the two conductors carry equal total currents of magnitude I in opposite directions, the center being directed out of the paper. What is desired is to determine the magnetic field at points inside and outside of the cable.

Solution: From the symmetry of the problem it is clear that \mathbf{B} must be everywhere tangent to a circle centered on the center conductor and drawn through the point at which \mathbf{B} is being considered. Furthermore, \mathbf{B} cannot depend on the azimuthal angle. The appropriate curves to use in the application of Eq. (8-52) are circles centered on the center conductor. For such a circle of radius r ,

$$\oint \mathbf{B} \cdot d\mathbf{l} = 2\pi r B \quad (8-55)$$

FIGURE 8-9

Cross section through a coaxial cable



which must equal μ_0 times the total current through the circle. Thus,

$$\begin{aligned} 2\pi rB &= \mu_0 I, & a < r < b \\ 2\pi rB &= 0, & b < r \end{aligned} \quad (8-56)$$

This apparently trivial result can be obtained by integration of the Biot law only with considerable difficulty.

THE MAGNETIC VECTOR POTENTIAL

The calculation of electric fields was much simplified by the introduction of the electrostatic potential. The possibility of making this simplification resulted from the vanishing of the curl of the electric field. The curl of the magnetic induction does not vanish; however, its divergence does.

Since the divergence of any curl is zero, it is reasonable to assume that the magnetic induction may be written

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (8-57)$$

The vector field \mathbf{A} is called the *magnetic vector potential*.

The only requirement placed on \mathbf{A} is that

$$\nabla \times \mathbf{B} = \nabla \times \nabla \times \mathbf{A} = \mu_0 \mathbf{J} \quad (8-58)$$

Using the identity

$$\nabla \times \nabla \times \mathbf{A} = \nabla \nabla \cdot \mathbf{A} - \nabla^2 \mathbf{A} \quad (8-59)$$

and specifying that $\nabla \cdot \mathbf{A} = 0$ yields

$$\nabla^2 \mathbf{A} = -\mu_0 \mathbf{J} \quad (8-60)$$

Integrating each rectangular component and using the solution for Poisson's equation as a guide leads to

$$\mathbf{A}(\mathbf{r}_2) = \frac{\mu_0}{4\pi} \int_V \frac{\mathbf{J}(\mathbf{r}_1)}{|\mathbf{r}_2 - \mathbf{r}_1|} dv_1 \quad (8-61)$$

The integrals involved in this expression are much easier to evaluate than those involved in the Biot law. However, they are also more complicated than those used to obtain the electrostatic potential.

An alternative way of obtaining Eq. (8-61) is by the direct transformation of Eq. (8-28) to the form of Eq. (8-57). This calculation begins by noting that

$$\frac{\mathbf{r}_2 - \mathbf{r}_1}{|\mathbf{r}_2 - \mathbf{r}_1|^3} = -\nabla_2 \frac{1}{|\mathbf{r}_2 - \mathbf{r}_1|} \quad (8-62)$$

where ∇_2 indicates that the differentiation is with respect to \mathbf{r}_2 . The vector

identity

$$\nabla \times (\varphi \mathbf{F}) = \varphi \nabla \times \mathbf{F} - \mathbf{F} \times \nabla \varphi \quad (8-63)$$

which is valid for any vector \mathbf{F} and any scalar φ , gives

$$\nabla_2 \times \left\{ \frac{1}{|\mathbf{r}_2 - \mathbf{r}_1|} \mathbf{J}(\mathbf{r}_1) \right\} = -\mathbf{J}(\mathbf{r}_1) \times \nabla_2 \frac{1}{|\mathbf{r}_2 - \mathbf{r}_1|} \quad (8-64)$$

since $\mathbf{J}(\mathbf{r}_1)$ does not depend on \mathbf{r}_2 . Combining these results in Eq. (8-28) leads to

$$\begin{aligned} \mathbf{B}(\mathbf{r}_2) &= \frac{\mu_0}{4\pi} \int_V \mathbf{J}(\mathbf{r}_1) \times \frac{(\mathbf{r}_2 - \mathbf{r}_1)}{|\mathbf{r}_2 - \mathbf{r}_1|^3} dv_1 = -\frac{\mu_0}{4\pi} \int_V \mathbf{J}(\mathbf{r}_1) \times \nabla_2 \frac{1}{|\mathbf{r}_2 - \mathbf{r}_1|} dv_1 \\ &= \frac{\mu_0}{4\pi} \int_V \nabla_2 \times \frac{\mathbf{J}(\mathbf{r}_1)}{|\mathbf{r}_2 - \mathbf{r}_1|} dv_1 \end{aligned} \quad (8-65)$$

The curl can be taken outside the integral, which puts Eq. (8-65) into exactly the form of Eq. (8-57). Thus, this approach also yields

$$\mathbf{A}(\mathbf{r}_2) = \frac{\mu_0}{4\pi} \int_{V_1} \frac{\mathbf{J}(\mathbf{r}_1)}{|\mathbf{r}_2 - \mathbf{r}_1|} dv_1 \quad (8-61)$$

To avoid leaving a false impression, namely, that the vector potential is as useful as the electrostatic potential in computing simple fields, it must be noted that there are essentially no cases where \mathbf{A} can be computed in *simple* closed form (although it could always be done numerically for bounded current distributions). The long straight wire gives an infinite result for \mathbf{A} when Eq. (8-61) is used.* The circular turn involves elliptic integrals, and so on. It should also be noted that evaluating the vector potential at a single point is not useful, because the magnetic induction is obtained by differentiation. The principal use of the vector potential is in approximations such as that discussed in the next section, and in problems involving electromagnetic radiation (see Chapters 16 and 20).

8-7 THE MAGNETIC FIELD OF A DISTANT CIRCUIT

The magnetic vector potential due to a small circuit at large distances (Fig. 8-10) can be evaluated relatively easily. The expression for the vector potential Eq. (8-61) may be applied to current circuits by making the substitution: $\mathbf{J} dv \rightarrow I d\mathbf{r}$. Thus

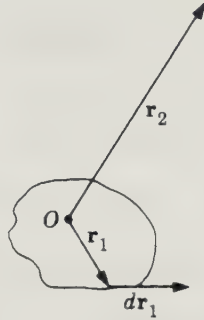
$$\mathbf{A}(\mathbf{r}_2) = \frac{\mu_0 I}{4\pi} \oint \frac{d\mathbf{r}_1}{|\mathbf{r}_2 - \mathbf{r}_1|} \quad (8-66)$$

For circuits whose dimensions are small compared with \mathbf{r}_2 , the denominator

* There is a finite vector potential for a long straight wire: $\mathbf{A} = -(\mu_0 I / 2\pi) \ln r \mathbf{k}$ in cylindrical coordinates for a wire on the z -axis carrying a current $I \mathbf{k}$. This result can be verified by direct calculation of $\nabla \times \mathbf{A}$. (See Appendix IV.)

FIGURE 8-10

The magnetic field at \mathbf{r}_2 due to a distant circuit. \mathbf{r}_1 ranges over the circuit. (The origin of \mathbf{r}_1 and \mathbf{r}_2 is within or near the circuit.)



can be approximated. To do so, we write, as in Eq. (2-46),

$$|\mathbf{r}_2 - \mathbf{r}_1|^{-1} = (r_2^2 + r_1^2 - 2\mathbf{r}_1 \cdot \mathbf{r}_2)^{-1/2} \quad (8-67)$$

and expand in powers of r_1/r_2 to get

$$|\mathbf{r}_2 - \mathbf{r}_1|^{-1} = \frac{1}{r_2} \left[1 + \frac{\mathbf{r}_1 \cdot \mathbf{r}_2}{r_2^2} + \dots \right] \quad (8-68)$$

to first order in r_1/r_2 . Using this expansion in Eq. (8-66) gives

$$\mathbf{A}(\mathbf{r}_2) = \frac{\mu_0 I}{4\pi} \left\{ \frac{1}{r_2} \oint d\mathbf{r}_1 + \frac{1}{r_2^3} \oint d\mathbf{r}_1 (\mathbf{r}_1 \cdot \mathbf{r}_2) + \dots \right\} \quad (8-69)$$

The first integral vanishes, and the second integrand is one term in the expansion

$$(\mathbf{r}_1 \times d\mathbf{r}_1) \times \mathbf{r}_2 = -\mathbf{r}_1(\mathbf{r}_2 \cdot d\mathbf{r}_1) + d\mathbf{r}_1(\mathbf{r}_1 \cdot \mathbf{r}_2) \quad (8-70)$$

To eliminate the first term on the right in Eq. (8-70), the differential of $\mathbf{r}_1(\mathbf{r}_2 \cdot \mathbf{r}_1)$ for a small change in \mathbf{r}_1 is written as

$$d[\mathbf{r}_1(\mathbf{r}_2 \cdot \mathbf{r}_1)] = \mathbf{r}_1(\mathbf{r}_2 \cdot d\mathbf{r}_1) + d\mathbf{r}_1(\mathbf{r}_2 \cdot \mathbf{r}_1) \quad (8-71)$$

which is exact. Adding Eqs. (8-70) and (8-71) and dividing by two yields

$$d\mathbf{r}_1(\mathbf{r}_1 \cdot \mathbf{r}_2) = \frac{1}{2}(\mathbf{r}_1 \times d\mathbf{r}_1) \times \mathbf{r}_2 + \frac{1}{2}d[\mathbf{r}_1(\mathbf{r}_2 \cdot \mathbf{r}_1)] \quad (8-72)$$

Since the last term is an exact differential, it contributes nothing to the second integral in Eq. (8-69). Thus, it follows that

$$\mathbf{A}(\mathbf{r}_2) = \frac{\mu_0}{4\pi} \left[\frac{I}{2} \oint \mathbf{r}_1 \times d\mathbf{r}_1 \right] \times \frac{\mathbf{r}_2}{r_2^3} \quad (8-73)$$

Equation (8-22) defines the quantity in brackets as the magnetic dipole moment, \mathbf{m} , of the circuit. Hence,

$$\mathbf{A}(\mathbf{r}_2) = \frac{\mu_0}{4\pi} \frac{\mathbf{m} \times \mathbf{r}_2}{r_2^3} \quad (8-74)$$

In this derivation, it has been assumed that all $r_1 \ll r_2$. Thus, Eq. (8-74) is not valid for an arbitrary origin, but only for an origin close to the circuit.

The magnetic induction can be determined by taking the curl of Eq. (8-74), which is readily accomplished by using vector identities. First,

$$\begin{aligned}\mathbf{B}(\mathbf{r}_2) &= \nabla \times \mathbf{A}(\mathbf{r}_2) = \frac{\mu_0}{4\pi} \nabla \times \left(\mathbf{m} \times \frac{\mathbf{r}_2}{r_2^3} \right) \\ &= \frac{\mu_0}{4\pi} \left[-(\mathbf{m} \cdot \nabla) \frac{\mathbf{r}_2}{r_2^3} + \mathbf{m} \nabla \cdot \frac{\mathbf{r}_2}{r_2^3} \right]\end{aligned}\quad (8-75)$$

The first term in the brackets can be transformed by noting that

$$m_x \frac{\partial}{\partial x_2} \left(\frac{\mathbf{r}_2}{r_2^3} \right) = \frac{m_x \mathbf{i}}{r_2^3} - 3m_x x_2 \frac{\mathbf{r}_2}{r_2^5}\quad (8-76)$$

Hence,

$$(\mathbf{m} \cdot \nabla) \frac{\mathbf{r}_2}{r_2^3} = \frac{\mathbf{m}}{r_2^3} - \frac{3(\mathbf{m} \cdot \mathbf{r}_2)\mathbf{r}_2}{r_2^5}\quad (8-77)$$

The second term involves only the calculation of

$$\nabla \cdot \frac{\mathbf{r}_2}{r_2^3} = \frac{3}{r_2^3} - \mathbf{r}_2 \cdot \frac{3\mathbf{r}_2}{r_2^5} = 0 \quad (r_2 \neq 0)\quad (8-78)$$

Finally,

$$\mathbf{B}(\mathbf{r}_2) = \frac{\mu_0}{4\pi} \left[-\frac{\mathbf{m}}{r_2^3} + \frac{3(\mathbf{m} \cdot \mathbf{r}_2)\mathbf{r}_2}{r_2^5} \right] \quad (\text{magnetic dipole})\quad (8-79)$$

Equation (8-79) shows that the magnetic field of a distant circuit does not depend on its detailed geometry, but only on its magnetic moment \mathbf{m} . Comparison with Eq. (2-36) shows that Eq. (8-79) is of the same form as the electric field due to an electric dipole, which explains the name *magnetic dipole* field.

8-8

THE MAGNETIC SCALAR POTENTIAL

Equation (8-51) indicates that the curl of the magnetic induction is zero wherever the current density is zero.

Therefore, the magnetic induction in such regions can be written as the gradient of a scalar potential:

$$\mathbf{B} = -\mu_0 \nabla \varphi^* \quad (8-80)$$

φ^* is called the *magnetic scalar potential*.

However, the divergence of \mathbf{B} is also zero, which means that

$$\nabla \cdot \mathbf{B} = -\mu_0 \nabla^2 \varphi^* = 0 \quad (8-81)$$

Thus φ^* satisfies Laplace's equation. Much of the work of electrostatics can be taken over directly and used to evaluate φ^* for various situations.

However, care must be taken in applying the boundary conditions. Furthermore, φ^* of a current-carrying circuit is not a single-valued function. A simple example is seen in Problem 8-25.

The expression for the scalar potential of a magnetic dipole is particularly useful. If it is noted that Eq. (8-79) can be written

$$\mathbf{B}(\mathbf{r}_2) = -\mu_0 \nabla \left(\frac{\mathbf{m} \cdot \mathbf{r}_2}{4\pi r_2^3} \right) \quad (8-82)$$

then it is clear that

$$\varphi^*(\mathbf{r}_2) = \frac{\mathbf{m} \cdot \mathbf{r}_2}{4\pi r_2^3} \quad (8-83)$$

for a magnetic dipole \mathbf{m} . Note the similarity between Eq. (8-83) and the potential of an electric dipole, Eq. (2-39).

A large circuit C can be divided into many small circuits by means of a mesh, as shown in Fig. 8-11. If each small loop formed by the mesh carries the same current as originally was carried by the circuit C , then, because of the cancellation of currents in the common branch of adjacent loops, the net effect is the same as if the charge flowed only in the circuit C . For any one of the small loops, the magnetic moment may be written as

$$d\mathbf{m} = I \mathbf{n} da \quad (8-84)$$

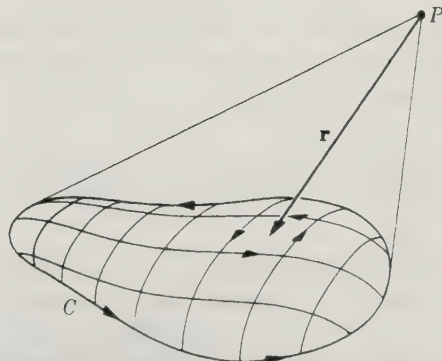
since each of the loops is sufficiently small to be regarded as planar. Here \mathbf{n} is the unit normal to the loop da . Using this expression in Eq. (8-83) and integrating over the surface bounded by C gives

$$\varphi^*(P) = \frac{I}{4\pi} \int_S \frac{\mathbf{r}_2 \cdot \mathbf{n} da}{r_2^3} \quad (8-85)$$

In this equation, \mathbf{r}_2 must be interpreted as the vector from da to the point P —that is, as $-\mathbf{r}$ in Fig. 8-11. Making the change $\mathbf{r}_2 = -\mathbf{r}$ results in

$$\varphi^*(P) = -\frac{I}{4\pi} \int_S \frac{\mathbf{r} \cdot \mathbf{n} da}{r^3} \quad (8-86)$$

FIGURE 8-11
A macroscopic current circuit constructed from elemental magnetic dipoles



The quantity $\mathbf{r} \cdot \mathbf{n} da$ is just r times the projection of da on a plane perpendicular to \mathbf{r} . Thus $\mathbf{r} \cdot \mathbf{n} da/r^3$ is the solid angle subtended by da at P . Equation (8-86) may then be written as

$$\varphi^*(P) = -\frac{I\Omega}{4\pi} \quad (8-87)$$

where Ω is the solid angle subtended by the curve C at the point P .

The magnetic scalar potential can be used for the calculation of the magnetic field due either to current-carrying circuits or to magnetic double layers (layers of dipoles). This procedure is occasionally useful in dealing with circuit problems; however, its principal use is in dealing with magnetic materials.

8-9 MAGNETIC FLUX

The quantity

$$\Phi = \int_S \mathbf{B} \cdot \mathbf{n} da \quad (8-88)$$

is known as the *magnetic flux* and is measured in webers (Wb).^{*} It is analogous to the electric flux discussed earlier, but it is of much greater importance.

The flux through a closed surface is zero, as can be seen by computing

$$\oint_S \mathbf{B} \cdot \mathbf{n} da = \int_V \nabla \cdot \mathbf{B} dv = 0 \quad (8-89)$$

It follows that the flux through a circuit is independent of the particular surface used to compute the flux. Use will be made of these results in Chapter 11, when electromagnetic induction is discussed.

8-10 SUMMARY

Magnetostatics is based on the addition of a magnetic force to the coulomb force when the charges are moving. In mks units, the **Lorentz force** on a test charge q with velocity \mathbf{v} is

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

The **magnetic field** of a source charge q_1 moving uniformly with velocity \mathbf{v}_1 is

$$\mathbf{B} = \frac{\mathbf{v}_1}{c} \times \frac{\mathbf{E}}{c}$$

^{*} Thus, a tesla is equal to a weber/m², which was formerly used as the mks unit for \mathbf{B} .

where \mathbf{E} is the electric field produced by q_1 and

$$c = 1/\sqrt{\epsilon_0\mu_0} \cong 3 \times 10^8 \text{ m/s}$$

is the velocity of light. (In gaussian units, B in these formulas is replaced by B/c .) The results are applied to conduction currents by writing

$$Nq \, dv \, \mathbf{v} = \mathbf{J} \, dv = I \, d\mathbf{l}$$

where $Nq = \rho$, $\rho\mathbf{v} = \mathbf{J}$ for the moving species of charged particle.

- The **force** on an element of wire $d\mathbf{l}$ in a magnetic field \mathbf{B} is

$$d\mathbf{F} = I \, d\mathbf{l} \times \mathbf{B}$$

The **torque** on a circuit is, for a uniform B -field,

$$\boldsymbol{\tau} = \mathbf{m} \times \mathbf{B}$$

- The **magnetic dipole moment** of the circuit is

$$\mathbf{m} = \frac{1}{2} I \oint_C \mathbf{r} \times d\mathbf{l}$$

The integral $\frac{1}{2} \oint_C \mathbf{r} \times d\mathbf{l}$ is the vector whose components are the areas enclosed by the projections of the curve C on the coordinate planes.

- The **magnetic field** produced by a current element $I \, d\mathbf{l}'$ is

$$d\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{I \, d\mathbf{l}' \times (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}$$

where $\mu_0/4\pi = 10^{-7} \text{ N/A}^2$ in mks units. The field of a complete circuit is calculated by integrating this around the circuit. For a general current distribution $\mathbf{J}(\mathbf{r}')$ as source,

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int_V \frac{\mathbf{J}(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} \, dv'$$

By differentiating this equation, we find there are no magnetic monopoles:

$$\boldsymbol{\nabla} \cdot \mathbf{B} = 0$$

In addition

$$\boldsymbol{\nabla} \times \mathbf{B} = \mu_0 \mathbf{J}$$

for a steady current distribution with

$$\boldsymbol{\nabla} \cdot \mathbf{J} = 0$$

These are the **fundamental differential equations** that must be satisfied locally at every point by all magnetostatic fields. (The divergence equation for \mathbf{B} is satisfied even by time dependent fields, and it is the second of the four fundamental Maxwell's equations.)

• **Ampere's law** follows from the curl equation by integrating both sides over an arbitrary surface S and applying Stokes's theorem:

$$\oint_C \mathbf{B} \cdot d\mathbf{l} = \mu_0 I$$

where $I = \int_S \mathbf{J} \cdot \mathbf{n} da$

is the total current through S bounded by C . This equation has practical usefulness for calculating B in a few special situations of high symmetry where \mathbf{B} is constant in magnitude and direction relative to a suitable curve C .

• The existence of the **vector potential function** $\mathbf{A}(\mathbf{r})$, such that

$$\mathbf{B} = \nabla \times \mathbf{A}$$

follows from the divergence equation. For a specified current distribution, $\mathbf{J}(\mathbf{r}')$,

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int_V \frac{\mathbf{J}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dv'$$

• At a large distance from the region where the source currents \mathbf{J} are located, the **multipole expansion** of \mathbf{A} gives

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{\mathbf{m} \times \mathbf{r}}{r^3} + \dots \quad (\text{There is no monopole term.})$$

• In regions where $\mathbf{J} = 0$, a **magnetic scalar potential** $\varphi^*(\mathbf{r})$ can be defined (since $\nabla \times \mathbf{B} = 0$) such that

$$\mathbf{B} = -\mu_0 \nabla \varphi^*$$

Like the electrostatic potential, φ^* satisfies Laplace's equation

$$\nabla^2 \varphi^* = 0$$

but the different boundary conditions may bring different sets of solutions into play. The dipole solution is the same as in the electrostatic case:

$$\varphi^*(\mathbf{r}) = \frac{\mathbf{m} \cdot \mathbf{r}}{4\pi r^3}$$

PROBLEMS

8-1 A charged particle of mass m and charge q moves in a uniform magnetic induction field \mathbf{B}_0 . Show that the most general motion of the particle traces out a helix, the cross section of which is a circle of radius $R = mv_{\perp}/qB$. (Here v_{\perp} is the component of velocity of the particle which is perpendicular to \mathbf{B}_0 .)

8-2 The Hamiltonian for a charged particle moving in a uniform magnetic induction field, \mathbf{B}_0 , which is parallel to the z -axis, is given by

$$\mathcal{H} = \frac{1}{2m} p^2 - \frac{qB_0}{2m} (xp_y - yp_x) + \frac{q^2 B_0^2}{8m} (x^2 + y^2)$$

Show that the equations of motion that may be derived from \mathcal{H} are consistent with the results of Problem 8-1.

8-3 A proton of velocity 10^7 m/s is projected at right angles to a uniform magnetic induction field of 0.15 T. (a) How much is the particle path deflected from a straight line after it has traversed a distance of 1 cm? (b) How long does it take the proton to traverse a 90° arc?

8-4 Show that the force law Eq. (8-25) can be transformed to

$$\mathbf{F}_2 = -\frac{\mu_0}{4\pi} I_1 I_2 \oint_1 \oint_2 d\mathbf{l}_2 \cdot d\mathbf{l}_1 \frac{\mathbf{r}_2 - \mathbf{r}_1}{|\mathbf{r}_2 - \mathbf{r}_1|^3}$$

which is clearly symmetric in that $\mathbf{F}_2 = -\mathbf{F}_1$.

8-5 Suppose a very long solenoid carries a (superconducting) current of 10 A and has 1000 turns per cm. Find the radial force per unit length f on one turn of the winding. Show that the tension in the wire T is $T = fa$, where a is the radius of the solenoid.

8-6 Show that the force between parallel wires carrying currents I_1 and I_2 , both in the same direction, is one of attraction. If the two parallel wires are very long and separated by distance a , find the magnetic force on segment $d\mathbf{l}_2$ of wire 2.

8-7 Consider a current circuit in the shape of a regular hexagon of side a . If the circuit carries the current I , find the magnetic induction at the center of the hexagon.

8-8 Consider a thin strip of metal of width w and very long. The current in the strip is uniform and directed along its length; the total current is I . Find the magnetic induction in the plane of the strip at distance b from the nearer edge.

8-9 A large number N of closely spaced turns of fine wire is wound in a single layer upon the surface of a wooden sphere of radius a , with the planes of the turns perpendicular to the axis of the sphere and completely covering its surface. If the current in the winding is I , determine the magnetic field at the center of the sphere.

8-10 A solenoid 15 cm long is wound in two layers. Each layer contains 100 turns; the first layer is 2 cm in radius, the second 2.05 cm. (a) If the winding carries a current of 3 A, find the magnetic induction at various points along the axis of the solenoid. (b) Make a plot of the axial magnetic induction as a function of distance, from the center to one end of the solenoid.

8-11 A solenoid of square cross section (i.e., a solenoid in which the individual turns are in the shape of a square) has N turns per unit length and carries current I . The cross sectional dimension is a . If the solenoid is very long, find the axial magnetic induction at its center.

8-12 The magnetic induction at a point on the axis (z -axis) of a circular turn of wire carrying current I is given in Eq. (8-38). Use the fact that $\nabla \cdot \mathbf{B} = 0$ to get an approximate expression for B_r (the radial component of the magnetic field) that is valid for points very near the axis.

8-13 The vertical component of the magnetic induction between the pole faces of a particle accelerator is given by $B_z = B_z(r, z)$, where $r = (x^2 + y^2)^{1/2}$ is the distance from the axis of the pole faces. (a) If $|B_z|$ is a decreasing function of r , show that the lines of magnetic intensity bow outward, as shown in Fig. 8-12, regardless of whether the upper pole is a north or south pole. [Hint: Use the fact that $\nabla \times \mathbf{B} = 0$, and that $B_r = 0$ on the median plane.] (b) If the lines of \mathbf{B} bow as shown in the figure, show

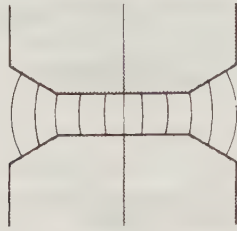


FIGURE 8-12

that accelerated particles that drift away from the median plane experience a force tending to restore them to the median plane, regardless of whether they are positively or negatively charged.

***8-14** It is evident from Eq. (8-30) that only vector fields of a certain class qualify as physically realizable magnetic induction fields. (a) Verify that $\mathbf{B} = (\mathbf{r}/r) \times \nabla g(x, y, z)$ with arbitrary $g(x, y, z)$ is an appropriate magnetic field. (b) Find the current density \mathbf{J} producing it, if g is a solution of Laplace's equation.

8-15 For a homogeneous, isotropic, nonmagnetic medium of conductivity g , in which there are steady currents, show that \mathbf{B} satisfies the vector Laplace equation: $\nabla^2 \mathbf{B} = 0$.

8-16 By using Ampere's circuital law, find the magnetic induction at distance r from the center of a long wire carrying current I . Do so for both $r > R$ and $r < R$, where R is the radius of the wire. Show explicitly that the magnetic induction vanishes on the axis of the wire.

8-17 A cylindrical conductor of radius b contains a cylindrical hole of radius a . The axis of the hole is parallel to the axis of the conductor and at a distance s away from it, $a < s < b - a$. The conductor carries a uniform current density \mathbf{J} . Find the B -field in the hole on the diameter that coincides with a diameter of the conductor. [Hint: Treat an equivalent current distribution with density \mathbf{J} throughout the hole as well as the conductor, plus $-\mathbf{J}$ in the hole.]

8-18 According to Problem 8-1, the orbit of a charged particle (q, m) in a plane perpendicular to the magnetic field \mathbf{B}_0 is a circle of radius $R = mv_{\perp}/qB_0$. Determine the periods of revolution of both an electron and a proton in a magnetic field of 0.5 T. (The masses of the electron and proton are 0.9110×10^{-30} kg and 1.6726×10^{-27} kg, respectively.)

8-19 A toroid is wound uniformly, as shown in Fig. 11-2. It has N turns of wire, which carry current I . The inner radius of the toroid is a , the outer b . (a) Find the magnetic induction at various points inside the toroidal winding. (b) Find the ratio b/a that will permit \mathbf{B} in the ring to vary by no more than 25 percent.

8-20 Show that the magnetic vector potential for two long, straight, parallel wires carrying the same current, I , in opposite directions is given by $\mathbf{A} = (\mu_0 I / 2\pi) \ln(r_2/r_1) \mathbf{n}$ where r_2 and r_1 are the distances from the field point to the wires, and \mathbf{n} is a unit vector parallel to the wires.

8-21 Consider the following set of conductors: an infinitely long straight wire surrounded by a thin cylindrical shell of metal (at radius b) arranged coaxially with the wire. The two conductors carry equal but opposite currents I . Find the magnetic vector potential for the system.

8-22 (a) Show that $\oint_C \mathbf{A} \cdot d\mathbf{l} = \Phi$, where Φ is the magnetic flux through a surface bounded by the circuit C . (b) Use this result together with Ampere's law to find \mathbf{A} at distance r both inside and outside a very long solenoid.

8-23 (a) Show by direct differentiation of Eq. (8-61) that $\nabla \cdot \mathbf{A} = 0$. (b) Show that $\mathbf{A} + \nabla\psi$, where ψ is an arbitrary function, is also a vector potential for the same B -field as \mathbf{A} . (c) Show that, by a suitable choice of ψ , the vector potential of \mathbf{B} can have any divergence desired.

8-24 (a) Show that the following are all possible vector potentials of the uniform field $\mathbf{B} = B\mathbf{k}$: $\mathbf{A}_1 = -By\mathbf{i}$, $\mathbf{A}_2 = Bx\mathbf{j}$, and $\mathbf{A}_3 = -\frac{1}{2}\mathbf{r} \times \mathbf{B}$. (b) For which of these is $\nabla \cdot \mathbf{A} = 0$? (c) Show that $\mathbf{A}_1 - \mathbf{A}_2$ is the gradient of a function $\nabla\psi$.

8-25 Show that the B -field outside of a long straight wire carrying a current I is derivable from the scalar potential $\varphi^* = -(I/2\pi)\theta$ in cylindrical coordinates and that φ^* satisfies Laplace's equation. Why is this φ^* not one of the cylindrical harmonics (as would be the case for the electrostatic potential of a line charge)?

8-26 The magnetic dip angle is defined as the angle between the direction of the earth's magnetic induction field and the tangent plane at the earth's surface. Derive an expression for the dip angle as a function of geomagnetic latitude, on the assumption that the induction is a dipole field.

8-27 A current-carrying circuit in the shape of a square (of side a) lies in the xy -plane (origin at center of square, its sides parallel to the axes). If the circuit carries the current I , calculate the magnetic field B at (a) the center of the square, (b) at point $(a/2, 0, a/2)$, (c) at point $(a, 0, 0)$, and (d) at point $(a/2, 0, -a/2)$.

8-28 (a) Make an accurate plot of the equipotential surfaces of the scalar potential for a magnetic dipole (in the $\phi = 0$ plane). Sketch in some magnetic field lines. (b) Determine the direction of the magnetic field as a function of r and θ . (c) Determine the r versus θ dependence of one particular magnetic field line.

***8-29** (a) Show that the magnetic scalar potential for a point on the axis (z -axis) of a circular loop, of radius a , is given by

$$\varphi^* = \frac{1}{2} I \left\{ 1 - \frac{z}{\sqrt{a^2 + z^2}} \right\}$$

(b) Expand this formula according to the binomial theorem to obtain a series expression valid for $z < a$. (c) The magnetic scalar potential φ^* should satisfy Laplace's equation; furthermore, by symmetry $\varphi^* = \varphi^*(r, \theta)$, where r is the distance from the center of the loop to the field point and θ is the angle between \mathbf{r} and the z -axis. Show that by using the zonal harmonics, Eq. (3-18), a solution for φ^* that reduces to the potential obtained in (b) on the symmetry axis, can be constructed. (d) Use the φ^* obtained in (c) to find B_r and B_θ at points off the symmetry axis of the loop.

***8-30** A sphere of radius a carrying surface charge density σ (rigidly attached) is rotated about an axis through its center with constant angular velocity ω . Show that the magnetic field at an external point is a dipole field and find the equivalent dipole moment.

8-31 Two dipoles \mathbf{m}_1 and \mathbf{m}_2 are in the same plane; \mathbf{m}_1 is fixed but \mathbf{m}_2 is free to rotate about its center. Show that, for equilibrium, $\tan \theta_1 = -2 \tan \theta_2$, where θ_1, θ_2 are the angles between \mathbf{r} and $\mathbf{m}_1, \mathbf{m}_2$ respectively (\mathbf{r} is the vector displacement between \mathbf{m}_2 and \mathbf{m}_1).

Magnetic Properties of Matter

In Chapter 8, we discussed techniques for finding the magnetic induction field due to a *specified* distribution of currents. Thus, for example, if we are dealing with a current-carrying circuit consisting of a closed loop of wire, the magnetic field in the vacuum region surrounding the wire may be calculated with the aid of Biot's Law. Now let the region surrounding the wire be filled with a material medium. Will the magnetic induction be altered by the presence of the matter? The answer is "yes." In this chapter, we explore the influence of matter on the magnetic field.

All matter consists ultimately of atoms, and each atom contains electrons in motion. These electron circuits, each of which is confined to a single atom, are what we shall call *atomic currents*. It thus appears that we have two kinds of current: (1) a conventional current, which consists of charge transport—that is, the motion of free electrons or charged ions—and (2) atomic currents, which are pure circulatory currents and give rise to no charge transport. However, both kinds of current may produce magnetic fields.

9-1

MAGNETIZATION

Each atomic current is a tiny closed circuit of atomic dimensions, so it seems reasonable that the "distant" magnetic field of the atom can be appropriately represented by a magnetic dipole field. In fact, a wide body of experimental studies as well as the formulation of quantum mechanics, which is our most accurate method for calculating atomic phenomena, tell us that the dominant part of the distant magnetic field due to a single atom is determined by specifying its magnetic dipole moment, \mathbf{m} .

Let the magnetic moment of the i th atom be \mathbf{m}_i . We now define a macroscopic vector quantity, the *magnetization* \mathbf{M} , by the same method

used to define polarization in Chapter 4.

We sum up vectorially all of the dipole moments in a small volume element Δv and then divide the result by Δv . The resulting quantity,

$$\mathbf{M} = \lim_{\Delta v \rightarrow 0} \frac{1}{\Delta v} \sum_i \mathbf{m}_i \quad (9-1)$$

is called the magnetic dipole moment per unit volume, or simply the *magnetization*.

The limit process in Eq. (9-1) is our usual macroscopic limiting process,—that is, Δv is made very small from the macroscopic point of view, but not so small that it does not contain a statistically large number of atoms. The quantity \mathbf{M} then becomes a vector point function. In the unmagnetized state, the summation $\sum \mathbf{m}_i$ will sum to zero as a result of random orientation of the \mathbf{m}_i , but in the presence of an external exciting field \mathbf{M} usually depends on this field. The specific dependence of \mathbf{M} on \mathbf{B} will be taken up in Section 9-6.

For the moment, we shall assume that $\mathbf{M}(x, y, z)$ is a known function, and shall compute the magnetized material's contribution to the magnetic field from the equations developed in Section 8-7.

The vector function \mathbf{M} provides us with a macroscopic description of the atomic currents inside matter. Specifically, \mathbf{M} measures the number of atomic current circuits per unit volume times the average or effective magnetic moment of each circuit. From the purely macroscopic point of view, all magnetic effects due to matter can be described adequately in terms of \mathbf{M} , or by its derivatives. One of these derivatives, $\nabla \times \mathbf{M}$, turns out to be the equivalent transport current density which would produce the same magnetic field as \mathbf{M} itself; it is called the *magnetization current density* \mathbf{J}_M . Before we derive this important relationship linking \mathbf{J}_M and \mathbf{M} , let us look at a simplified model of magnetized matter as though it consisted of atomic loop currents circulating in the same direction, side by side (Fig. 9-1). If the magnetization is uniform, the currents in the various loops tend

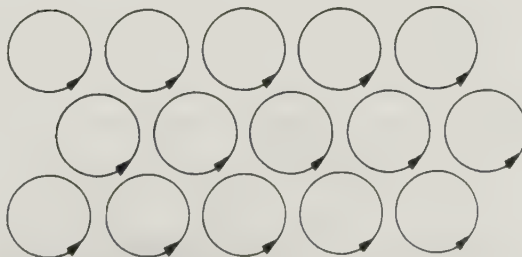
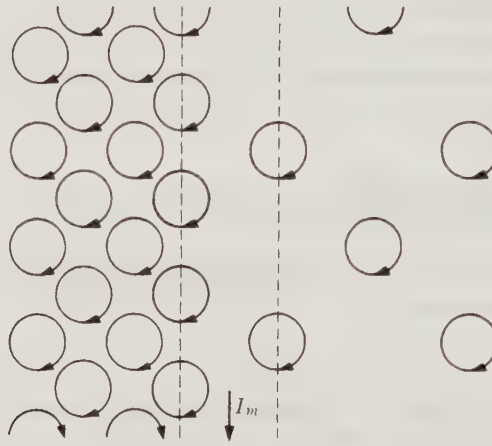


FIGURE 9-1
Simplified picture of magnetic material consisting of atomic loop currents circulating in the same direction

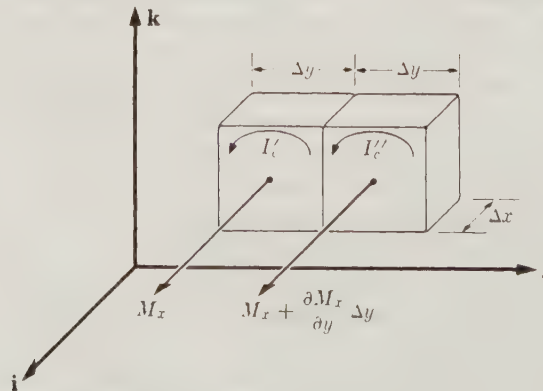
FIGURE 9-2
Example of abrupt
change in magnetization



to cancel each other out, and there is no net effective current in the interior of the material. If the magnetization is nonuniform, the cancellation will not be complete. As an example of nonuniform magnetization, consider the abrupt change in magnetization shown in Fig. 9-2. If we focus our attention on the region between the dotted lines, it is evident that there is more charge moving down than there is moving up. Thus, even though there is no charge transport, there is locally an effective motion of charge downward, and this “current” can produce a magnetic field. This current is called the *magnetization current*.

It remains for us to derive the relationship between \mathbf{J}_M and \mathbf{M} . Let us consider two small volume elements in a piece of magnetic material, each element of volume $\Delta x \Delta y \Delta z$, and located next to each other in the direction of the y -axis (Fig. 9-3). If the magnetization in the first volume

FIGURE 9-3
Replacement of volume
elements of magnetized
material by circulating
currents I'_c and I''_c



element is $\mathbf{M}(x, y, z)$, then the magnetization in the second element is

$$\mathbf{M}(x, y, z) + \frac{\partial \mathbf{M}}{\partial y} \Delta y + \text{higher-order terms}$$

The x -component of magnetic moment of the first element, $M_x \Delta x \Delta y \Delta z$, may be written in terms of a circulating current, I'_c :

$$M_x \Delta x \Delta y \Delta z = I'_c \Delta y \Delta z \quad (9-2)$$

Similarly, the x -component of magnetic moment of the second element, neglecting higher-order terms that vanish in the limit where each volume element becomes very small, is

$$\left(M_x + \frac{\partial M_x}{\partial y} \Delta y \right) \Delta x \Delta y \Delta z = I''_c \Delta y \Delta z \quad (9-3)$$

The net upward current in the middle region of the two volume elements is

$$I'_c - I''_c = -\frac{\partial M_x}{\partial y} \Delta x \Delta y \quad (9-4)$$

We next consider two adjacent volume elements along the x -axis and focus our attention on the y -component of the magnetization in each cell. In the middle region of the two cells, the net upward current due to the circulating currents that define the magnetic moments is

$$(I_c)_{\text{up}} = \frac{\partial M_y}{\partial x} \Delta x \Delta y \quad (9-5)$$

These are the only circulating currents of a particular cell that give rise to a net current in the z -direction. This net current, which comes about from nonuniform magnetization, is the magnetization current. This current is not a transport current but derives, as we have seen, from circulatory currents—that is, from atomic currents in the material. The effective area for each of the currents in (9-4) and (9-5) is $\Delta x \Delta y$. Thus,

$$(J_M)_z = \frac{\partial M_y}{\partial x} - \frac{\partial M_x}{\partial y} \quad (9-6a)$$

or

$$\mathbf{J}_M = \nabla \times \mathbf{M} \quad (9-6b)$$

The *magnetization current density* is the curl of the magnetization.

9-2 THE MAGNETIC FIELD PRODUCED BY MAGNETIZED MATERIAL

According to Eq. (9-1), each volume element $\Delta v'$ of magnetized matter is characterized by a magnetic moment

$$\Delta \mathbf{m} = \mathbf{M}(x', y', z') \Delta v' \quad (9-7)$$

Using the results of Section 8-7, we may write the contribution to the magnetic field at point (x, y, z) from each $\Delta \mathbf{m}$ (or, equivalently, from each $\Delta v'$). The magnetic field is then obtained as an integral over the entire volume of material, V_0 . This procedure is indicated schematically in Fig. 9-4.

Instead of calculating \mathbf{B} directly, we find it expedient to work with the vector potential \mathbf{A} , and to obtain \mathbf{B} subsequently by means of the curl operation. According to Section 8-7, the vector potential at (x, y, z) is given by

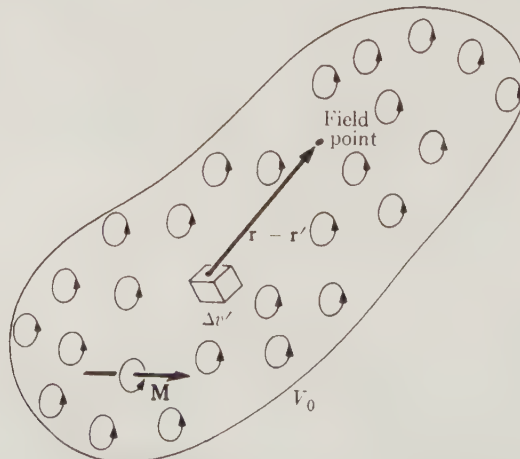
$$\begin{aligned} \mathbf{A}(x, y, z) &= \frac{\mu_0}{4\pi} \int_{V_0} \frac{\mathbf{M}(x', y', z') \times (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} dv' \\ &= \frac{\mu_0}{4\pi} \int_{V_0} \mathbf{M}(x', y', z') \times \nabla' \frac{1}{|\mathbf{r} - \mathbf{r}'|} dv' \end{aligned} \quad (9-8)$$

By means of the vector identities (1-1-9) and (1-2-3) in Tables 1-1 and 1-2, this integral may be transformed to

$$\mathbf{A}(x, y, z) = \frac{\mu_0}{4\pi} \int_{V_0} \frac{\nabla' \times \mathbf{M}}{|\mathbf{r} - \mathbf{r}'|} dv' + \frac{\mu_0}{4\pi} \int_{S_0} \frac{\mathbf{M} \times \mathbf{n}}{|\mathbf{r} - \mathbf{r}'|} da' \quad (9-9)$$

FIGURE 9-4

Contribution to the magnetic induction from a distribution of magnetized material



where S_0 is the surface of V_0 . Using Eq. (9-6b) and defining a surface magnetization current density \mathbf{j}_M (i.e., a magnetization current per unit length flowing in the surface layer) by the relation

$$\mathbf{j}_M = \mathbf{M} \times \mathbf{n} \quad (9-10)$$

we may write Eq. (9-9) as

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int_{V_0} \frac{\mathbf{J}_M(\mathbf{r}') dv'}{|\mathbf{r} - \mathbf{r}'|} + \frac{\mu_0}{4\pi} \int_{S_0} \frac{\mathbf{j}_M da'}{|\mathbf{r} - \mathbf{r}'|} \quad (9-11)$$

We might have ventured to predict the final expression, Eq. (9-11). Nevertheless, it is gratifying to see that it has come out of the mathematics in a natural way. Thus, the vector potential produced by a distribution of atomic currents inside matter has the same form as that produced by a distribution of true transport currents. We should point out that Eq. (9-10) is the proper expression for the surface current density that is consistent with $\mathbf{J}_M = \nabla' \times \mathbf{M}$. The \mathbf{j}_M must be introduced whenever \mathbf{M} changes abruptly, as it might at the interface between two media, but if the region of discontinuity in \mathbf{M} is imagined to be spread out over the distance $\Delta\xi$, then it can be shown that \mathbf{j}_M is contained in the term $\mathbf{J}_M \Delta\xi$. (Or, if the region is very thin, \mathbf{j}_M could be represented by a surface delta-function.)

Although Eq. (9-11) is both correct and of such form that it integrates nicely with the results of Chapter 8, it presents some practical difficulties when it comes to the task of computing \mathbf{B} from a specified distribution of magnetization. First, there is the $\nabla \times \mathbf{M}$ operation to perform, and second, another curl operation is involved in obtaining \mathbf{B} from the \mathbf{A} field. It certainly is preferable to work with scalar quantities if possible, and the gradient of a scalar field (such as we encountered in electrostatics) is easier to compute than the curl of a vector field. For this reason, we go back to Eq. (9-8) and try another approach. We are interested, after all, in \mathbf{B} , not \mathbf{A} , so let us formally take the curl:

$$\mathbf{B}(\mathbf{r}) = \nabla \times \mathbf{A} = \frac{\mu_0}{4\pi} \int_{V_0} \nabla \times \left[\mathbf{M} \times \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} \right] dv' \quad (9-12)$$

where the differential operators in the curl act on the unprimed coordinates.

As the reader may have anticipated, our next job is to transform the integrand of Eq. (9-12). To do so, we appeal to the vector identities of Table 1-1. According to (1-1-10),

$$\nabla \times (\mathbf{F} \times \mathbf{G}) = (\nabla \cdot \mathbf{G})\mathbf{F} - (\nabla \cdot \mathbf{F})\mathbf{G} + (\mathbf{G} \cdot \nabla)\mathbf{F} - (\mathbf{F} \cdot \nabla)\mathbf{G}$$

Letting $\mathbf{F} = \mathbf{M}(\mathbf{r}')$ and $\mathbf{G} = (\mathbf{r} - \mathbf{r}')/|\mathbf{r} - \mathbf{r}'|^3$, and noting that the differentiations are with respect to the unprimed coordinates, we find that the identity reduces to

$$\nabla \times \left[\mathbf{M} \times \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} \right] = \mathbf{M} \nabla \cdot \left[\frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} \right] - (\mathbf{M} \cdot \nabla) \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} \quad (9-13)$$

since $\nabla \cdot \mathbf{M}(x', y', z') = 0$, etc. Thus,

$$\mathbf{B}(\mathbf{r}) = \mathbf{B}_I(\mathbf{r}) + \mathbf{B}_{II}(\mathbf{r}) \quad (9-14)$$

where

$$\mathbf{B}_I(\mathbf{r}) = \frac{\mu_0}{4\pi} \int_{V_0} \mathbf{M} \nabla \cdot \left[\frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} \right] dv' \quad (9-14a)$$

$$\mathbf{B}_{II}(\mathbf{r}) = -\frac{\mu_0}{4\pi} \int_{V_0} (\mathbf{M} \cdot \nabla) \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} dv' \quad (9-14b)$$

We consider the simpler \mathbf{B}_I integral first. Using Eq. (2-57), we obtain

$$\mathbf{B}_I(\mathbf{r}) = \frac{\mu_0}{4\pi} \int_{V_0} \mathbf{M}(\mathbf{r}') 4\pi \delta(\mathbf{r} - \mathbf{r}') dv' = \mu_0 \mathbf{M}(\mathbf{r}) \quad (9-15)$$

We next consider the \mathbf{B}_{II} integral. The integrand may be transformed by means of a second identity (1-1-6), which becomes

$$\begin{aligned} \nabla \left[\mathbf{M} \cdot \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} \right] &= (\mathbf{M} \cdot \nabla) \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} \\ &+ \mathbf{M} \times \nabla \times \left[\frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} \right] \end{aligned} \quad (9-16)$$

The last term in (9-16) contains

$$\nabla \times \left[\frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} \right] = -\nabla \times \nabla \frac{1}{|\mathbf{r} - \mathbf{r}'|}$$

which vanishes identically. Hence,

$$\mathbf{B}_{II}(\mathbf{r}) = -\mu_0 \nabla \frac{1}{4\pi} \int_{V_0} \mathbf{M}(\mathbf{r}') \cdot \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} dv'$$

which may be written as in Eq. (8-80),

$$\mathbf{B}_{II}(\mathbf{r}) = -\mu_0 \nabla \varphi^*(\mathbf{r}) \quad (9-17)$$

The quantity $\varphi^*(\mathbf{r})$ is a scalar field, the magnetic scalar potential due to magnetic material:

$$\varphi^*(\mathbf{r}) = \frac{1}{4\pi} \int_{V_0} \mathbf{M}(\mathbf{r}') \cdot \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} dv' \quad (9-18)$$

Adding the two contributions, (9-15) and (9-17), we find for the magnetic induction field:

$$\mathbf{B}(\mathbf{r}) = -\mu_0 \nabla \varphi^*(\mathbf{r}) + \mu_0 \mathbf{M}(\mathbf{r}) \quad (9-19)$$

Thus, the magnetic induction due to a magnetized distribution of matter may be expressed as the sum of two terms: the gradient of a scalar field plus

a term proportional to the local magnetization. At an external point—that is, in vacuum— \mathbf{M} is zero, and the magnetic induction is then just the gradient of a scalar field which is the integral of the distant dipole fields given by Eq. (8-83).

MAGNETIC SCALAR POTENTIAL AND MAGNETIC POLE DENSITY

The expression for the magnetic scalar potential, Eq. (9-18), is similar in form to that for the electrostatic potential arising from polarized dielectric material. Here again mathematical transformation is suggested:

$$\frac{\mathbf{M} \cdot (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} = \mathbf{M} \cdot \nabla' \frac{1}{|\mathbf{r} - \mathbf{r}'|} = \nabla' \cdot \frac{\mathbf{M}}{|\mathbf{r} - \mathbf{r}'|} - \frac{1}{|\mathbf{r} - \mathbf{r}'|} \nabla' \cdot \mathbf{M} \quad (9-20)$$

so that Eq. (9-18) becomes

$$\varphi^*(\mathbf{r}) = \frac{1}{4\pi} \int_{S_0} \frac{\mathbf{M} \cdot \mathbf{n} da'}{|\mathbf{r} - \mathbf{r}'|} - \frac{1}{4\pi} \int_{V_0} \frac{\nabla' \cdot \mathbf{M}}{|\mathbf{r} - \mathbf{r}'|} dv' \quad (9-21)$$

where S_0 is the surface of the region V_0 .

By analogy with Section 4-2, it is expedient to define two scalar quantities:

$$\rho_M(\mathbf{r}') \equiv -\nabla' \cdot \mathbf{M}(\mathbf{r}') \quad (9-22)$$

called the *magnetic pole density*, and

$$\sigma_M(\mathbf{r}') \equiv \mathbf{M}(\mathbf{r}') \cdot \mathbf{n} \quad (9-23)$$

the *surface density of magnetic pole strength*.

These quantities are quite useful even though rather artificial. They play the same role in the theory of magnetism that ρ_P and σ_P play in dielectric theory. The units of ρ_M and σ_M are A/m² and A/m, respectively.

Consider, for example, a uniformly magnetized bar magnet. Since the magnetization is uniform, $\rho_M = 0$. The only surface densities that do not vanish are on those surfaces that have a normal component of the magnetization, and these are called the *poles* of the magnet. This example is somewhat idealized, yet not too different from the common laboratory bar magnet familiar to the reader. (In practice, the poles of a magnet exert a demagnetizing influence that destroys the uniformity of \mathbf{M} and thus spreads each pole over a somewhat larger region than just the surface.)

The total pole strength of every magnet is zero. This statement follows directly from the divergence theorem:

$$\int_{V_0} (-\nabla \cdot \mathbf{M}) dv + \int_{S_0} \mathbf{M} \cdot \mathbf{n} da = 0$$

We now complete the derivation that was started earlier. Equation (9-18) becomes

$$\varphi^*(\mathbf{r}) = \frac{1}{4\pi} \int_{V_0} \frac{\rho_M dv'}{|\mathbf{r} - \mathbf{r}'|} + \frac{1}{4\pi} \int_{S_0} \frac{\sigma_M da'}{|\mathbf{r} - \mathbf{r}'|} \quad (9-18a)$$

and $\mathbf{B}(x, y, z)$ is obtained as $-\mu_0$ times the gradient with respect to the unprimed coordinates, plus the term $\mu_0\mathbf{M}$:

$$\begin{aligned} \mathbf{B}(\mathbf{r}) = & \frac{\mu_0}{4\pi} \int_{V_0} \rho_M \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} dv' \\ & + \frac{\mu_0}{4\pi} \int_{S_0} \sigma_M \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} da' + \mu_0\mathbf{M}(\mathbf{r}) \end{aligned} \quad (9-19a)$$

This equation represents the contribution from the magnetized material in V_0 to the magnetic induction at (x, y, z) .

9-4

SOURCES OF THE MAGNETIC FIELD: MAGNETIC INTENSITY

In the preceding sections, we have seen how magnetized material produces a magnetic field. Furthermore, Chapter 8 dealt with the magnetic effects of conventional currents. In the general case, both types of magnetic sources are present: conventional or transport currents, which can be measured in the laboratory, and the atomic currents inside matter. It is important to realize that under certain conditions the same piece of matter may produce a magnetic field both because it is magnetized and because it is carrying a transport current of charge carriers. Thus, for example, one of our best magnetic materials, iron, may carry a transport current via its free electrons, but the fixed iron ions in the crystal contain atomic currents, which can be oriented to produce a strong magnetization.

In general, the expression for the magnetic field may be written as

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int_V \frac{\mathbf{J} \times (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} dv' - \mu_0 \nabla \varphi^*(\mathbf{r}) + \mu_0\mathbf{M}(\mathbf{r}) \quad (9-24)$$

where

$$\varphi^*(\mathbf{r}) = \frac{1}{4\pi} \int_V \frac{\rho_M dv'}{|\mathbf{r} - \mathbf{r}'|} + \frac{1}{4\pi} \int_S \frac{\sigma_M da'}{|\mathbf{r} - \mathbf{r}'|} \quad (9-25)$$

The volume V extends over all current-carrying regions and over all matter. The surface S includes all surfaces and interfaces between different media. The current density \mathbf{J} includes only conventional currents of the charge transport variety, whereas the effect of atomic currents is found in the magnetization vector \mathbf{M} (and potential φ^*).

Equation (9-24) may be solved for \mathbf{B} if \mathbf{M} and \mathbf{J} are specified at all

points. In most problems, however, \mathbf{J} is specified but $\mathbf{M}(x', y', z')$ depends on $\mathbf{B}(x', y', z')$, so that even if the functional form of $\mathbf{M}(\mathbf{B})$ is known, Eq. (9-24) provides at best an integral equation for \mathbf{B} .

To help get around this difficulty we introduce an auxiliary magnetic vector, the *magnetic intensity* \mathbf{H} , defined by

$$\mathbf{H} = \frac{1}{\mu_0} \mathbf{B} - \mathbf{M} \quad (9-26)$$

By combining (9-24) and (9-26) we obtain

$$\mathbf{H}(\mathbf{r}) = \frac{1}{4\pi} \int_V \frac{\mathbf{J} \times (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} dv' - \nabla \varphi^*(\mathbf{r}) \quad (9-27)$$

It appears that we have gained nothing by this maneuver because \mathbf{H} still depends on \mathbf{M} through ρ_M and σ_M ; but in the next section, we will see how \mathbf{H} is related to the conventional current density \mathbf{J} through a differential equation. The situation is similar to the electrostatic case, where the auxiliary vector \mathbf{D} is related to the prescribed charge density through its divergence.

The field vector \mathbf{H} plays an important role in magnetic theory, particularly in problems involving permanent magnets. These will be discussed in later sections of the chapter. The units of \mathbf{H} are the same as those of \mathbf{M} , namely, A/m.

9-5 THE FIELD EQUATIONS

In Chapter 8, the basic equations describing the magnetic effects of currents were expressed in differential form:

$$\nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{B} = \mu_0 \mathbf{J}$$

We should now like to see how these equations are modified when the magnetic field \mathbf{B} includes a contribution from magnetized material.

We saw in Section 8-3 that the divergence equation ($\nabla \cdot \mathbf{B} = 0$) is valid for all magnetic fields that are produced by a current distribution. This result is not limited to fields produced by conventional currents. We saw in Section 9-2 that the magnetic field produced by magnetized matter can be expressed as the curl of a vector (\mathbf{A}), and therefore $\nabla \cdot \mathbf{B} = 0$ is satisfied automatically. It is an experimental observation that

$$\nabla \cdot \mathbf{B} = 0 \quad (9-28)$$

for all magnetic induction fields. Thus, \mathbf{B} can always trace its origin to a current distribution and there is no evidence for isolated magnetic poles.

The “curl equation” is the differential form of Ampere’s circuital law. Here we must be careful to include all types of currents that can produce a magnetic field. Hence, in the general case, this equation is properly written as

$$\nabla \times \mathbf{B} = \mu_0(\mathbf{J} + \mathbf{J}_M) \quad (9-29)$$

where \mathbf{J} is the transport current density, and \mathbf{J}_M is the magnetization current density. Equation (9-6b) may be combined with Eq. (9-29) to yield

$$\nabla \times \left(\frac{1}{\mu_0} \mathbf{B} - \mathbf{M} \right) = \mathbf{J}$$

which, according to (9-26), is equivalent to the following expression:

$$\nabla \times \mathbf{H} = \mathbf{J} \quad (9-30)$$

That is, the auxiliary magnetic vector \mathbf{H} is related to the *transport current density* through its curl. This result also follows by taking the curl of Eq. (9-27). Equations (9-28) and (9-30) are the fundamental magnetic field equations in the presence of matter. These equations, together with appropriate boundary conditions and an experimental relationship between \mathbf{B} and \mathbf{H} , are sufficient to solve magnetic problems.

In some instances, it is preferable to use an integral formulation of the theory. With the aid of Stokes’s theorem, Eq. (9-30) may be converted to

$$\int_S \nabla \times \mathbf{H} \cdot \mathbf{n} \, da = \oint_C \mathbf{H} \cdot d\mathbf{l} = \int_S \mathbf{J} \cdot \mathbf{n} \, da$$

or

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = I \quad (9-31)$$

In other words, the line integral of the tangential component of the magnetic intensity around a closed path C is equal to the entire transport current through the area bounded by the curve C .

Because of the divergence theorem, Eq. (9-28) is equivalent to

$$\oint_S \mathbf{B} \cdot \mathbf{n} \, da = 0 \quad (9-32)$$

The magnetic flux through any closed surface is zero.

MAGNETIC SUSCEPTIBILITY, PERMEABILITY, AND HYSTERESIS

In order to solve problems in magnetic theory, it is essential to have a relationship between \mathbf{B} and \mathbf{H} or, equivalently, a relationship between \mathbf{M} and one of the magnetic field vectors. These relationships depend on the nature of the magnetic material and are usually obtained from experiment.

In a large class of materials, there exists an approximately linear relationship between \mathbf{M} and \mathbf{H} . If the material is isotropic as well as linear,*

$$\mathbf{M} = \chi_m \mathbf{H} \quad (9-33)$$

where the dimensionless scalar quantity χ_m is called the *magnetic susceptibility*.

If χ_m is positive, the material is called *paramagnetic*, and the magnetic induction is strengthened by the presence of the material. If χ_m is negative, the material is *diamagnetic*, and the magnetic induction is weakened by the presence of the material. Although χ_m is a function of the temperature and sometimes varies quite drastically with temperature, it is generally safe to say that χ_m for paramagnetic and diamagnetic materials is quite small. That is,

$$|\chi_m| \ll 1 \quad (\text{for paramagnetic, diamagnetic materials}) \quad (9-34)$$

The susceptibilities of some common materials are given in Table 9-1.

In most handbooks and tabulations of physical data, χ_m is not listed directly, but instead is given as the *mass susceptibility*, $\chi_{m,\text{mass}}$, or the *molar susceptibility*, $\chi_{m,\text{molar}}$. These are defined by

$$\chi_m = \chi_{m,\text{mass}} d \quad (9-35)$$

$$\chi_m = \chi_{m,\text{molar}} \frac{d}{A} \quad (9-36)$$

where d is the mass density of the material and A is the molecular weight. Since \mathbf{M} and \mathbf{H} both have the dimensions of magnetic moment per unit

* If the material is anisotropic but linear, Eq. (9-33) is replaced by the tensor relationships

$$M_x = \chi_{m,11}H_x + \chi_{m,12}H_y + \chi_{m,13}H_z$$

and so on. In these circumstances, \mathbf{M} does not necessarily have the same direction as \mathbf{H} . We shall restrict ourselves in this book to isotropic media.

TABLE 9-1
Magnetic Susceptibility
of Some Paramagnetic
and Diamagnetic Mate-
rials at Room
Temperature

Material	χ_m	$\chi_{m,\text{mass}}$ (m ³ /kg)
Aluminum	2.1×10^{-5}	0.77×10^{-8}
Bismuth	-16.4×10^{-5}	-1.68×10^{-8}
Copper	-0.98×10^{-5}	-0.11×10^{-8}
Diamond	-2.2×10^{-5}	-0.62×10^{-8}
Gadolinium chloride (GdCl ₃)	603.0×10^{-5}	133.3×10^{-8}
Gold	-3.5×10^{-5}	-0.18×10^{-8}
Magnesium	1.2×10^{-5}	0.68×10^{-8}
Mercury	-2.8×10^{-5}	-0.21×10^{-8}
Silver	-2.4×10^{-5}	-0.23×10^{-8}
Sodium	0.84×10^{-5}	0.87×10^{-8}
Titanium	18.0×10^{-5}	4.01×10^{-8}
Tungsten	7.6×10^{-5}	0.40×10^{-8}
Carbon dioxide (1 atm)	-1.19×10^{-8}	-0.60×10^{-8}
Hydrogen (1 atm)	-0.22×10^{-8}	-2.48×10^{-8}
Nitrogen (1 atm)	-0.67×10^{-8}	-0.54×10^{-8}
Oxygen (1 atm)	193.5×10^{-8}	135.4×10^{-8}

Source: Data obtained from the *Handbook of Chemistry and Physics*, 70th ed. (Boca Raton, FL: CRC Press, Inc., 1990). Practically all sources of data list magnetic susceptibilities in gaussian (cgs) units; if the superscript (1) is used to indicate the constant in the gaussian system, then $\chi_m = 4\pi\chi_m^{(1)}$ and $\chi_{m,\text{mass}} = 4\pi \times 10^{-3}\chi_{m,\text{mass}}^{(1)}$.

volume, it is evident that $\chi_{m,\text{mass}} \mathbf{H}$ and $\chi_{m,\text{molar}} \mathbf{H}$ give magnetic moment per unit mass and magnetic moment per mole, respectively. For convenience, the mass susceptibility is also listed in Table 9-1.

A linear relationship between \mathbf{M} and \mathbf{H} implies also a linear relationship between \mathbf{B} and \mathbf{H} :

$$\mathbf{B} = \mu \mathbf{H} \quad (9-37)$$

where the permeability μ is obtained from the combination of Eqs. (9-26) and (9-33):

$$\mu = \mu_0(1 + \chi_m) \quad (9-38)$$

The dimensionless quantity

$$K_m = \frac{\mu}{\mu_0} = 1 + \chi_m \quad (9-39)$$

is sometimes tabulated instead of χ_m . This quantity, K_m , is called the *relative permeability*.

For the paramagnetic and diamagnetic materials of Table 9-1, it is evident that K_m is very close to unity.

The *ferromagnetics* form another class of magnetic material. Such a material is characterized by a possible permanent magnetization and by the fact that its presence usually has a profound effect on the magnetic induction. Ferromagnetic materials are *not* linear, so that Eqs. (9-33) and (9-37) with constant χ and μ *do not* apply.* It has been expedient, however, to use Eq. (9-37) as the defining equation for μ —that is, with $\mu = \mu(\mathbf{H})$ —but the reader should be cautioned that this practice can lead to difficulty in certain situations. If the μ of a ferromagnetic material is defined by Eq. (9-37), then, depending on the value of \mathbf{H} , μ goes through an entire range of values from infinity to zero and may be either positive or negative. The best advice that can be given is to consider each problem involving ferromagnetism separately, try to determine which region of the \mathbf{B} - \mathbf{H} diagram is important for the particular problem, and make approximations appropriate to this region. The magnetic properties of some ferromagnetic materials are listed in Table 9-2. (The quantities denoted by M_s , H_c , and so on will be defined in the following paragraphs.)

First, let us consider an unmagnetized sample of ferromagnetic material. If the magnetic intensity, initially zero, is increased *monotonically*, then the \mathbf{B} - \mathbf{H} relationship will trace out a curve something like that shown in Fig. 9-5, which is the magnetization curve of the material. It is evident that μ 's taken from the magnetization curve, using the expression $\mu = B/H$, are always of the same sign (positive), but they show a rather large spectrum of values. The maximum permeability occurs at the “knee” of the curve. In some materials, this maximum permeability is as large as $10^5 \mu_0$, but in others it is much lower. The reason for the knee in the curve is that the magnetization \mathbf{M} approaches a maximum value in the material, and

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$$

continues to increase at very large \mathbf{H} only because of the $\mu_0\mathbf{H}$ term. The maximum value of \mathbf{M} is called the *saturation magnetization* of the material (See Table 9-2.)

Next consider a ferromagnetic specimen magnetized by the above procedure. If the magnetic intensity \mathbf{H} is decreased, the \mathbf{B} - \mathbf{H} relationship does not follow back down the curve of Fig. 9-5, but instead moves along the new curve of Fig. 9-6 to point r . The magnetization, once established, does not disappear with the removal of \mathbf{H} ; in fact, it takes a reversed magnetic intensity to reduce the magnetization to zero. If \mathbf{H} continues to build up in the reversed direction, then \mathbf{M} (and hence \mathbf{B}) will establish itself in the reversed direction, and Fig. 9-6 begins to show a certain symmetry.

* Certain kinds of iron, referred to as *soft iron*, can be treated as *approximately* linear nevertheless.

TABLE 9-2
Properties of Ferromagnetic Materials at Room Temperature

Material	Composition (%)	$\mu_0 M_s$ (T)	H_s (A/m)	K_m (max.)
Elements				
Iron (annealed)		2.15	1.6×10^5	5,500
Cobalt		1.79	7.0×10^5	
Nickel		0.61	5.5×10^5	
Alloys and Compounds				
			H_c (A/m)	
Iron-Silicon	96 Fe, 4 Si	1.97	56	8,000
Permalloy	55 Fe, 45 Ni	1.60	5.6	50,000
Mumetal	5 Cu, 2 Cr, 77 Ni, 16 Fe	0.75	1.2	100,000
Permendur	50 Co, 50 Fe	2.45	159	5,000
Mn ferrite	Mn Fe ₂ O ₄	0.49		2,500
Ni ferrite	Ni Fe ₂ O ₄	0.32		2,500
		B_r (T)		
Cobalt steel	52 Fe, 36 Co, 4 W, 6 Cr, 0.8 C	0.97	19×10^3	
Alnico V	51 Fe, 8 Al, 14 Ni, 24 Co, 3 Cu	1.25	52×10^3	
Platinum-cobalt	77 Pt, 23 Co	0.6	3.4×10^5	
Samarium-cobalt	Sm Co	0.84	6.7×10^5	
Neodymium-iron	13 Nd, 81 Fe, 6 B	0.80	1.2×10^6	

Note: M_s = saturation magnetization, H_s = magnetic intensity required for saturation, H_c = coercivity, B_r = remanence.

Source: Data from *American Institute of Physics Handbook*, 3rd ed. (New York: McGraw-Hill, 1972), and *Handbook of Chemistry and Physics*, 70th ed. (Boca Raton, FL: CRC Press, Inc., 1990). Data for Nd-Fe-B from J. J. Croat et al., *Journal of Applied Physics* **55**, 2079 (1985).

FIGURE 9-5
Magnetization curve and relative permeability of commercial iron (annealed)

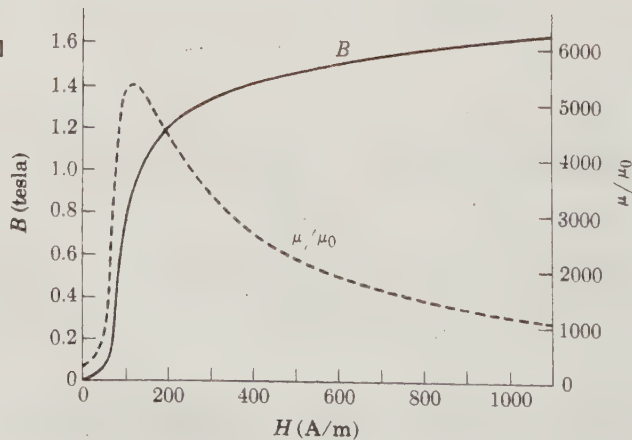
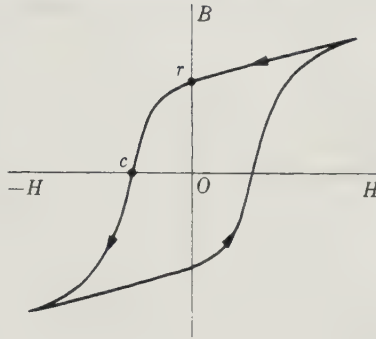


FIGURE 9-6
Typical hysteresis loop of
a ferromagnetic material



Finally, when \mathbf{H} once again increases, the operating point follows the lower curve of Fig. 9-6. Thus, the \mathbf{B} - \mathbf{H} curve for increasing \mathbf{H} is entirely different from that for decreasing \mathbf{H} . This phenomenon is called *hysteresis*, from the Greek word meaning “to lag”; the magnetization literally lags the exciting field.

The curve of Fig. 9-6 is called the *hysteresis loop* of the material. The value of \mathbf{B} at point r is known as the *retentivity* or *remanence*, and the magnitude of \mathbf{H} at point c is called the *coercive force* or *coercivity* of the material. From Fig. 9-6 it is evident that the value of μ , defined by Eq. (9-37), is negative in the second and fourth quadrants of the diagram. The shape of the hysteresis loop depends not only upon the nature of the ferromagnetic material (Fig. 9-7) but also on the maximum value of \mathbf{H} to which the material is subjected (Fig. 9-8). However, once \mathbf{H}_{\max} is sufficient to produce saturation in the material, the hysteresis loop does not change shape with increasing \mathbf{H}_{\max} . For soft iron, the hysteresis is relatively small.

For certain applications it is desirable to know the effective permeability of a material to a small alternating H -field superposed on a large constant field. Thus, if ΔB is the change in magnetic field produced by a change ΔH in the magnetic intensity, the *incremental permeability* μ_{in} is defined by

$$\mu_{\text{in}} = \frac{\Delta B}{\Delta H} \quad (9-40)$$

and is approximately equal to the slope of the hysteresis curve that goes through the point in question.

Ferromagnetic materials are used either (1) to increase the magnetic flux of a current circuit or (2) as sources of the magnetic field (permanent magnets). For use as a permanent magnet, the material is first magnetized to saturation by placing it in a strong magnetic field (i.e., by placing it between the poles of an electromagnet or by placing it in a solenoid subjected to a momentary large current). However, when the permanent magnet is removed from the external field, it will in general be subject to a

FIGURE 9-7

Comparison of the hysteresis curves of several materials. (Note that $\mu_0 H$ is plotted along the abscissa instead of just H . $\mu_0 = 4\pi \times 10^{-7} \text{ T} \cdot \text{m/A.}$) Data from R. M. Bozorth, *Ferromagnetism* (New York: Van Nostrand, 1951).

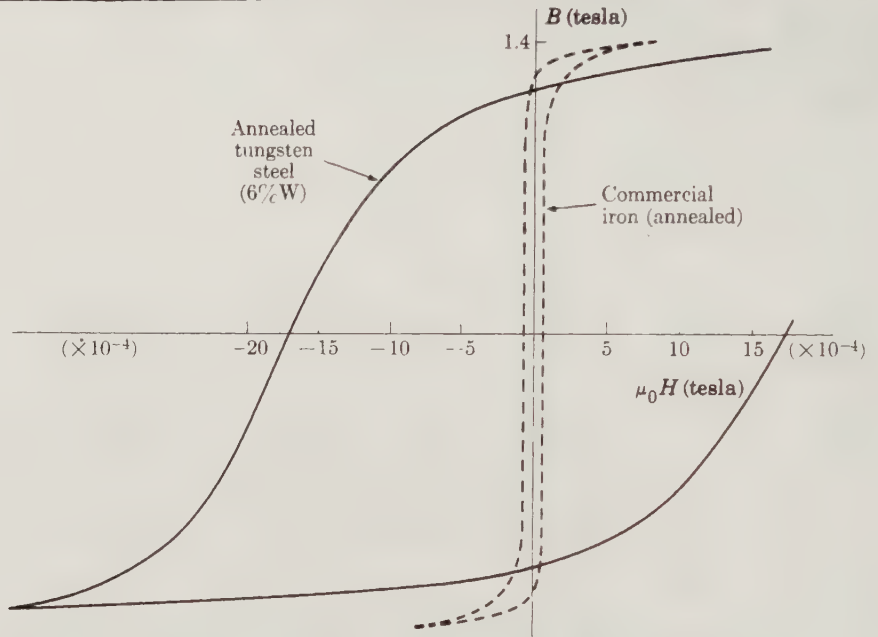


FIGURE 9-8

Major hysteresis loop and several minor hysteresis loops for a typical material

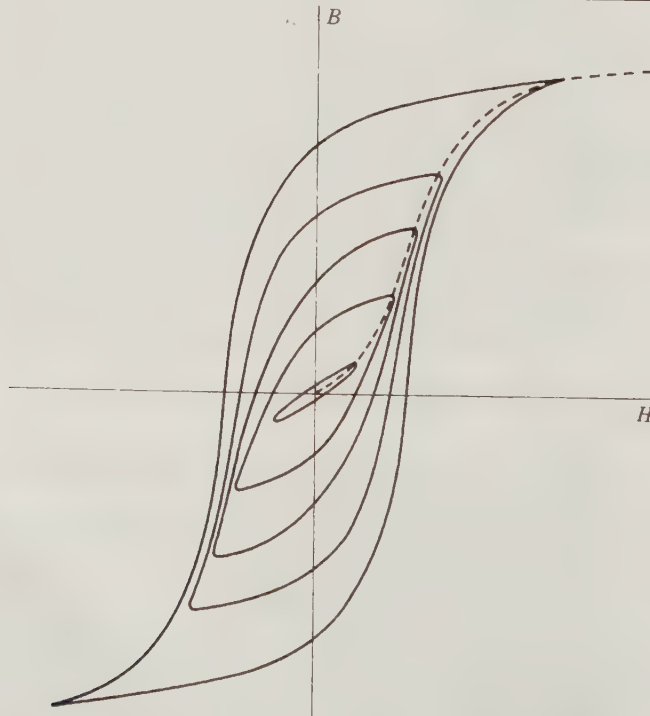
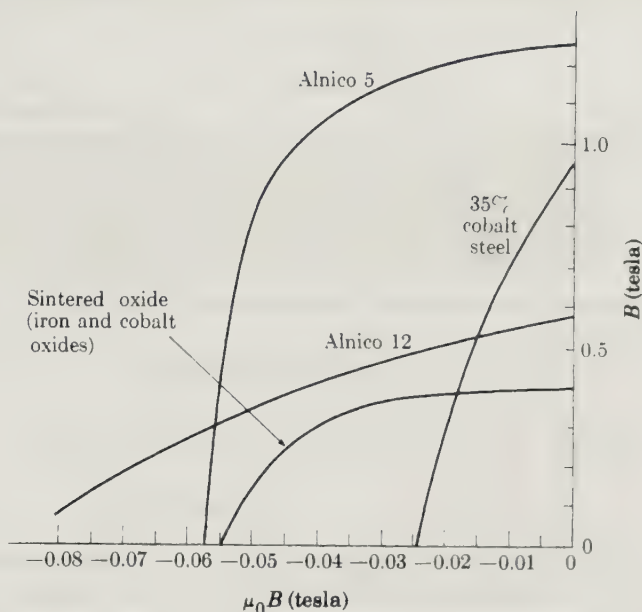


FIGURE 9-9

Hysteresis curves of permanent magnet materials. (Note that $\mu_0 H$ is plotted along the abscissa instead of just H .)



demagnetizing field, as will be discussed in detail in Sections 9-8 and 9-11. Thus, the second quadrant of the hysteresis loop diagram is the important part of the \mathbf{B} - \mathbf{H} relationship for a permanent magnet material (Fig. 9-9).

9-7**BOUNDARY CONDITIONS ON THE FIELD VECTORS**

Before we can solve magnetic problems, even simple ones, we must know how the field vectors \mathbf{B} and \mathbf{H} change in passing an interface between two media. The interface to be considered may be between two media with different magnetic properties or between a material medium and vacuum.

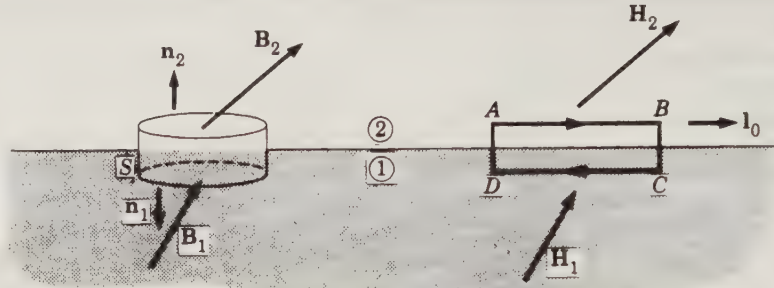
Consider two media, 1 and 2, in contact, as shown in Fig. 9-10. Let us construct the small pillbox-shaped surface S , which intersects the interface, the height of the pillbox being negligibly small in comparison with the diameter of the bases. Applying the flux integral, Eq. (9-32), to the surface S , we find

$$\mathbf{B}_2 \cdot \mathbf{n}_2 \Delta S + \mathbf{B}_1 \cdot \mathbf{n}_1 \Delta S = 0$$

where \mathbf{n}_2 and \mathbf{n}_1 are the outward-directed normals to the upper and lower surfaces of the pillbox. Since $\mathbf{n}_2 = -\mathbf{n}_1$ and since either of these normals

FIGURE 9-10

Boundary conditions on the field vectors at the interface between two media, obtained by applying Gauss's law to S and integrating $\mathbf{H} \cdot d\mathbf{l}$ around the path $ABCD$



may serve as the normal to the interface,

$$(\mathbf{B}_2 - \mathbf{B}_1) \cdot \mathbf{n}_2 = 0 \quad (9-41a)$$

or

$$\mathbf{B}_{2n} - \mathbf{B}_{1n} = 0 \quad (9-41b)$$

Thus, the normal component of \mathbf{B} is continuous across an interface.

A boundary condition on the \mathbf{H} -field may be obtained by applying Ampere's circuital law, Eq. (9-31), to the rectangular path $ABCD$ in Fig. 9-10. On this path, the lengths AB and CD will be taken equal to Δl , and the segments AD and BC will be assumed negligibly small. The current through the rectangle is negligible unless there is a true surface current. Therefore,

$$(\mathbf{H}_2 - \mathbf{H}_1) \cdot \mathbf{l}_0 = \mathbf{j} \cdot (\mathbf{n}_2 \times \mathbf{l}_0) = \mathbf{j} \times \mathbf{n}_2 \cdot \mathbf{l}_0$$

or

$$(\mathbf{H}_2 - \mathbf{H}_1)_t = \mathbf{j} \times \mathbf{n}_2 \quad (9-42a)$$

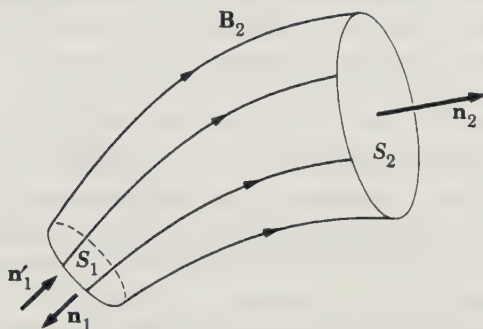
where \mathbf{j} is the *surface* current density (transport current per unit length in the surface layer) and \mathbf{l}_0 is a unit vector in the direction of Δl . Thus the tangential component of the magnetic intensity is continuous across an interface unless there is a true surface current. Finally, by taking the cross product of Eq. (9-42a) with \mathbf{n}_2 , the equation may be written as

$$\mathbf{n}_2 \times (\mathbf{H}_2 - \mathbf{H}_1) = \mathbf{j} \quad (9-42b)$$

This form is convenient for determining \mathbf{j} if \mathbf{H}_2 and \mathbf{H}_1 are known.

Before completing this section, we shall prove one other important property of the magnetic induction \mathbf{B} , namely, that its flux is everywhere continuous. Let us focus our attention on a region of space and construct *magnetic field lines*, which are imaginary lines drawn in such a way that the direction of a line at any point is the direction of \mathbf{B} at that point. Next we imagine a tube of flux, a volume bounded on the sides by lines of \mathbf{B} but not cut by them (Fig. 9-11). The tube is terminated on the ends by the surfaces

FIGURE 9-11
A tube of magnetic induction



S_1 and S_2 . Applying the divergence theorem, we obtain

$$\begin{aligned} \int_V \nabla \cdot \mathbf{B} \, dv &= 0 \\ &= \int_{S_2} \mathbf{B} \cdot \mathbf{n} \, da - \int_{S_1} \mathbf{B} \cdot \mathbf{n}' \, da \\ &= \Phi(S_2) - \Phi(S_1) \end{aligned} \quad (9-43)$$

Thus the same magnetic flux enters the tube through S_1 as leaves through S_2 . The flux lines can never terminate, but must eventually join back onto themselves, forming closed loops.

The previous statements apply to the \mathbf{B} -field, and it is perhaps worthwhile noting that they do not apply to the \mathbf{H} -field, since $\nabla \cdot \mathbf{H} = -\nabla \cdot \mathbf{M}$, which is not everywhere zero. Thus, from the divergence theorem applied to a tube of magnetic intensity, we find

$$\int_{S_2} \mathbf{H} \cdot \mathbf{n} \, da - \int_{S_1} \mathbf{H} \cdot \mathbf{n}' \, da = \int_V \rho_M \, dv \quad (9-44)$$

The discontinuity in the magnetic intensity flux is determined by the total magnetic pole strength intercepted by the flux tube.

BOUNDARY-VALUE PROBLEMS INVOLVING MAGNETIC MATERIALS

Since \mathbf{B} and \mathbf{H} obey boundary conditions similar to those for \mathbf{D} and \mathbf{E} , the problems of linear media or specified magnetization are similar to the dielectric problems discussed in Chapter 4. In this section, we shall be concerned with a particular class of problems, namely, calculation of magnetic fields inside magnetic material in which no transport current exists. This situation is formally identical to the dielectric with no external charge density.

When $\mathbf{J} = 0$, the fundamental magnetic equations, (9-28) and (9-30), reduce to

$$\nabla \cdot \mathbf{B} = 0 \quad (9-28)$$

$$\nabla \times \mathbf{H} = 0 \quad (9-45)$$

Equation (9-45) implies that \mathbf{H} can be derived as the gradient of a scalar field. This result should not surprise us because, according to the source equation, (9-27), the contribution to \mathbf{H} from magnetic material is already expressed in this form, and in Section 8-8, we showed that the field (actually the proof presented there must be generalized to the \mathbf{H} -field) produced by transport currents can be so derived when the local current density is zero. In accordance with Eq. (9-45), we write

$$\mathbf{H} = -\nabla\varphi^* \quad (9-46)$$

where φ^* is now the magnetic scalar potential due to all sources.

There are two types of magnetic material for which the magnetic field calculation reduces to a simple boundary-value problem: (1) *linear* or approximately linear magnetic material for which $\mathbf{B} = \mu\mathbf{H}$ and (2) a *uniformly* magnetized piece of material for which $\nabla \cdot \mathbf{M} = 0$. In both cases, Eq. (9-28) reduces to

$$\nabla \cdot \mathbf{H} = 0 \quad (9-47)$$

Combining this result with (9-46), we obtain

$$\nabla^2\varphi^* = 0 \quad (9-48)$$

which is Laplace's equation. Thus, the magnetic problem reduces to finding a solution to Laplace's equation that satisfies the boundary conditions. \mathbf{H} may then be calculated as minus the gradient of the magnetic potential, and \mathbf{B} obtained from

$$\mathbf{B} = \mu\mathbf{H}$$

or

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$$

whichever is appropriate.

Two examples serve to illustrate the usefulness of the method just described, and additional exercises of this type will be found in the problems at the end of the chapter.

EXAMPLE 9-1
Magnetically Permeable
Sphere in Uniform
Magnetic Field

Consider a sphere of linear magnetic material of radius a and permeability μ placed in a region of space containing an *initially* uniform magnetic field, \mathbf{B}_0 . We should like to determine how the magnetic field is modified by the presence of the sphere and, in particular, to determine the magnetic field in the sphere itself.

Solution: The problem is closely analogous to the case of a dielectric sphere in a uniform electric field, which was solved in Section 4-9. Thus, choosing the origin of our coordinate system at the center of the sphere and the direction of \mathbf{B}_0 as the polar direction (z -direction), we may express the potential as a sum of zonal harmonics. Again, all boundary conditions can be satisfied by means of the $\cos \theta$ harmonics:

$$\varphi_1^*(r, \theta) = A_1 r \cos \theta + C_1 r^{-2} \cos \theta \quad (9-49)$$

for the vacuum region (1) outside the sphere, and

$$\varphi_2^*(r, \theta) = A_2 r \cos \theta + C_2 r^{-2} \cos \theta \quad (9-50)$$

for the magnetic material region (2). The constants A_1 , A_2 , C_1 , and C_2 must be determined from the boundary conditions.

At distances far away from the sphere, the magnetic field retains its uniform character: $\mathbf{B} = B_0 \mathbf{k}$, and $\varphi_1^* \rightarrow -(B_0/\mu_0)r \cos \theta$. Hence, $A_1 = -(B_0/\mu_0)$. Since φ_2^* and its associated magnetic field cannot become infinite at any point, the coefficient C_2 must be set equal to zero. Having applied the boundary conditions at $r = \infty$ and at $r = 0$, we next turn our attention to the interface at $r = a$:

$$H_{1\theta} = H_{2\theta}, \quad B_{1r} = B_{2r}$$

or

$$-\left(\frac{B_0}{\mu_0}\right) \sin \theta + \frac{C_1}{a^3} \sin \theta = A_2 \sin \theta \quad (9-51)$$

$$B_0 \cos \theta + 2\mu_0 \frac{C_1}{a^3} \cos \theta = -\mu A_2 \cos \theta \quad (9-52)$$

Solving these two equations simultaneously yields

$$A_2 = -\frac{3B_0}{(\mu + 2\mu_0)},$$

and

$$C_1 = [(\mu/\mu_0) - 1] \frac{B_0 a^3}{(\mu + 2\mu_0)}$$

whence the magnetic fields inside and outside the sphere are given by

$$\mathbf{B}_2 = \frac{3B_0 \mathbf{k}}{1 + 2(\mu_0/\mu)} \quad (9-53)$$

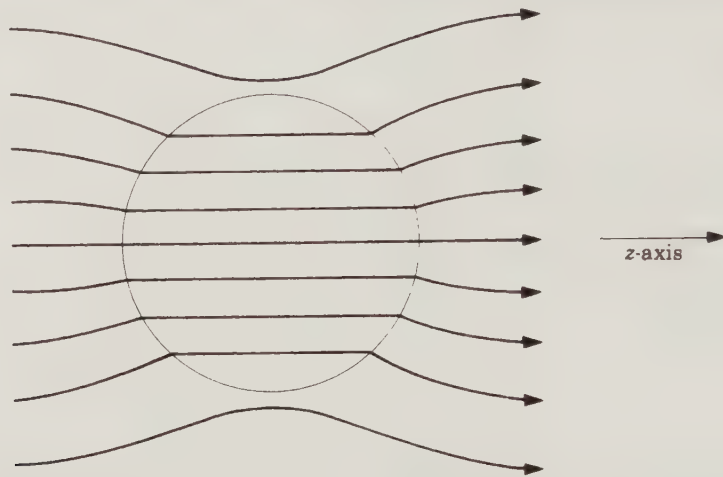
and

$$\mathbf{B}_1 = B_0 \mathbf{k} + \left[\frac{(\mu/\mu_0) - 1}{(\mu/\mu_0) + 2} \right] \left(\frac{a}{r} \right)^3 B_0 (2\mathbf{a}_r \cos \theta + \mathbf{a}_\theta \sin \theta) \quad (9-54)$$

The lines of magnetic induction are depicted in Fig. 9-12.

FIGURE 9-12

Lines of magnetic induction \mathbf{B} in the vicinity of a sphere of large permeability μ



The second example deals with a permanent magnet.

EXAMPLE 9-2
Field of a Uniformly Magnetized Sphere

The problem is to determine the magnetic field produced by a uniformly magnetized sphere of magnetization \mathbf{M} and radius a when no other magnetic fields are present.

Solution: Taking the magnetization along the z -axis and the origin of our coordinate system at the center of the sphere, we may expand the potential in zonal harmonics:

$$\varphi_1^*(r, \theta) = \sum_{n=0}^{\infty} C_{1,n} r^{-(n+1)} P_n(\theta) \quad (9-55)$$

for the vacuum region (1) outside the sphere, and

$$\varphi_2^*(r, \theta) = \sum_{n=0}^{\infty} A_{2,n} r^n P_n(\theta) \quad (9-56)$$

for the permanent magnet region (2). Here we have purposely left out the harmonics with positive powers of r from expansion (9-55) since these would become large at large distances, and we have left out the negative powers of r in (9-56) since these would become infinite at the origin. From the boundary conditions at $r = a$,

$$H_{1\theta} = H_{2\theta}$$

$$B_{1r} = B_{2r}$$

we obtain

$$\sum_{n=0}^{\infty} (C_{1,n}a^{-(n+1)} - A_{2,n}a^n)a^{-1} \frac{d}{d\theta} P_n(\theta) = 0 \quad (9-57)$$

or

$$\sum_{n=0}^{\infty} (C_{1,n}a^{-(n+1)} - A_{2,n}a^n)P_n(\theta) = \text{constant} \quad (9-57a)$$

and

$$\begin{aligned} \mu_0 C_{1,0}a^{-2} + \mu_0 \sum_{n=1}^{\infty} P_n(\theta)[C_{1,n}(n+1)a^{-(n+2)} + A_{2,n}na^{n-1}] \\ - \mu_0 M \cos \theta = 0 \end{aligned} \quad (9-58)$$

Since the $P_n(\theta)$ are orthogonal functions, each term of Eqs. (9-57a) and (9-58) must vanish individually. For $n = 0$,

$$C_{1,0}a^{-1} - A_{2,0} = \text{constant}, \quad \mu_0 C_{1,0}a^{-2} = 0$$

Therefore, $C_{1,0} = 0$, and $A_{2,0}$ is an arbitrary constant that may be set equal to zero without affecting \mathbf{H} or \mathbf{B} .

From the $n = 1$ terms,

$$C_{1,1}a^{-3} - A_{2,1} = 0$$

and

$$2C_{1,1}a^{-3} + A_{2,1} - M = 0$$

which may be solved simultaneously to yield

$$C_{1,1} = \frac{1}{3}Ma^3$$

and

$$A_{2,1} = \frac{1}{3}M$$

For all $n \geq 2$, the only $C_{1,n}$ and $A_{2,n}$ compatible with the two equations are $C_{1,n} = 0$ and $A_{2,n} = 0$.

Putting these results back into Eqs. (9-55) and (9-56), we obtain

$$\varphi_1^*(r, \theta) = \frac{1}{3}M(a^3/r^2) \cos \theta \quad (9-59)$$

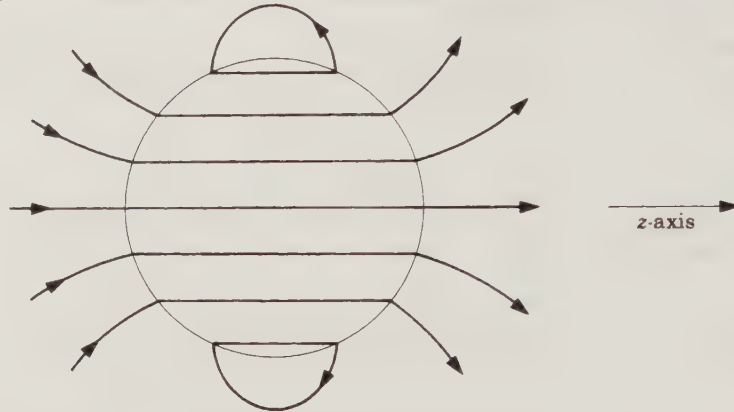
and

$$\varphi_2^*(r, \theta) = \frac{1}{3}Mr \cos \theta \quad (9-60)$$

The magnetic intensity \mathbf{H} may be calculated from the gradient opera-

FIGURE 9-13

Lines of magnetic induction for the uniformly magnetized sphere



tion, with the following result:

$$\mathbf{H}_1 = \frac{1}{3}M(a^3/r^3)[2\mathbf{a}_r \cos \theta + \mathbf{a}_\theta \sin \theta] \quad (9-61)$$

$$\mathbf{H}_2 = -\frac{1}{3}M\mathbf{k} \quad (9-62)$$

Thus, the external field of the uniformly magnetized sphere is exactly a dipole field, arising from the dipole moment $\frac{4}{3}\pi a^3 \mathbf{M}$. The magnetic intensity *inside* the sphere is a demagnetizing field, a result that is in accord with the E -field inside a uniformly polarized dielectric. We see, therefore, that the magnetized sphere is subjected to its own demagnetizing field. The factor $\frac{1}{3} = (1/4\pi)(4\pi/3)$ in Eq. (9-62) depends explicitly on the spherical geometry. The quantity $4\pi/3$ is known as the *demagnetization factor* of a sphere. The demagnetization factors for other geometrical shapes have been calculated and tabulated.*

The external magnetic field \mathbf{B}_1 is just μ_0 times Eq. (9-61). The magnetic induction in the sphere is

$$\mathbf{B}_2 = \frac{2}{3}\mu_0 M\mathbf{k} = \frac{2}{3}\mu_0 \mathbf{M} \quad (9-63)$$

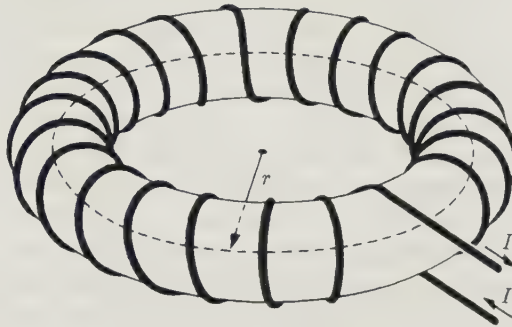
The lines of magnetic induction are shown in Fig. 9-13.

9-9 CURRENT CIRCUITS CONTAINING MAGNETIC MEDIA

In Chapter 8, we dealt with magnetic fields produced by current circuits *in vacuum*. One of the examples taken up in the problems (Problem 8-19) was that of a uniformly wound toroid of N turns carrying current I (Fig. 9-14). Let us solve the toroid problem again, but this time with the region inside the windings filled with a ferromagnetic material, which we shall assume to

* See, for example, *American Institute of Physics Handbook*, 3rd ed. (New York: McGraw-Hill, 1972), p. 5-247.

FIGURE 9-14
A toroidal winding



be homogeneous, isotropic, and originally unmagnetized. The field vector obtained most easily is the magnetic intensity, since it is related to the current in the windings by means of Ampere's circuital law, Eq. (9-31). If we apply Eq. (9-31) to a circular path that is coaxial with the hole in the toroid, such as the dotted path shown in the figure, symmetry arguments tell us that \mathbf{H} is the same at all points on the path:

$$H_t l = NI$$

or

$$H_t = \frac{NI}{l} \quad (9-64)$$

Here the subscript stands for the component tangent to the path, and $l = 2\pi r$ is the total path length. From Eq. (9-26),

$$B_t = \frac{\mu_0 NI}{l} + \mu_0 M_t \quad (9-65)$$

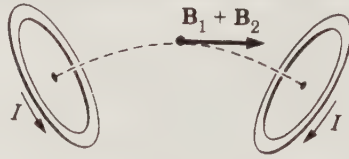
Thus, the magnetic field differs from that in the vacuum case by the additive term $\mu_0 M_t$.

Only the tangential component of \mathbf{B} (and of \mathbf{H}) is obtained by the above procedure. Nevertheless, this component is the only one we expect to be present. According to Eq. (9-27) there are two kinds of sources for the magnetic intensity: transport currents and magnetized material. It is easy to show that the current in the toroidal winding produces only a tangential field. The winding is equivalent to N circular loops of current; if we combine the loops in pairs (Fig. 9-15), it is evident that each pair of loops produces a tangential field at the point in question.

The second source of \mathbf{H} , the magnetized material itself, may possibly provide a contribution through the pole densities: $\rho_M = -\nabla \cdot \mathbf{M}$ and $\sigma_M = \mathbf{M} \cdot \mathbf{n}$. Since the ferromagnetic material in the toroid is isotropic, \mathbf{M} will have the same direction as \mathbf{H} . But \mathbf{M} arose in response to currents in the toroidal windings, and this field is a tangential one. Thus, we expect an M_t only and we can drop the subscript t . On this basis, there are no surfaces

FIGURE 9-15

Axial nature of the field in a toroidal winding, shown by combining the magnetic field of the current loops in pairs



in the toroidal specimen that are normal to \mathbf{M} , and hence no σ_M . Finally, ρ_M must equal zero. Although M may be a function of r (the distance from the axis of the toroid), the term $\partial M/\partial r$ does not contribute to $\nabla \cdot \mathbf{M}$. The interesting result is that the magnetized material provides no contribution to \mathbf{H} in this case, and Eq. (9-65) gives the entire magnetic field.

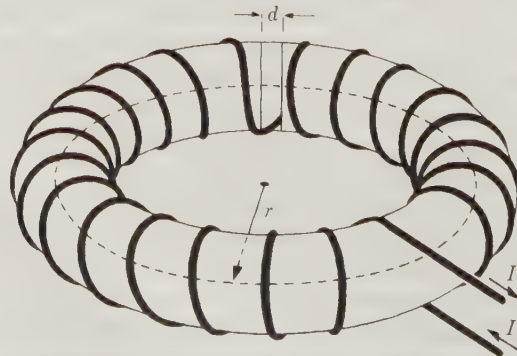
Another problem, somewhat more complicated than the preceding one, is that of a toroidal winding of N turns surrounding a ferromagnetic specimen in which a narrow air gap of width d has been cut (Fig. 9-16). We will not distinguish between an air gap and a vacuum gap, since it is evident from Table 9-1 that the permeability of air differs only slightly from μ_0 . In this problem, Ampere's circuital law does not suffice to determine \mathbf{H} , because symmetry arguments cannot be invoked to state that \mathbf{H} is the same at all points on a circular path. Thus, we first go to the source equation, Eq. (9-27).

Again we note that there are two contributions to the magnetic intensity, one from transport currents and one from the magnetization. Since the toroidal winding is identical to that of the preceding problem, the contribution to \mathbf{H} from the transport currents must be the same as before. Denoting this contribution by the subscript 1, we may write

$$H_1 = \frac{NI}{l} \quad (9-66)$$

FIGURE 9-16

A toroidal winding surrounding a ring of magnetic material with an air gap



Our problem, then, is to evaluate H_2 , or the $\nabla\varphi^*$ term. To keep the problem simple, we make the plausible assumption of uniform tangential magnetization M throughout the ferromagnetic material. This choice provides us with all the essential physics without complicating the algebra. Then ρ_M equals zero, but $\sigma_M = \pm M$ on the pole faces bordering the air gap. The situation here is strongly reminiscent of the electrostatic problem involving a charged parallel-plate capacitor. In fact, the mathematical formulation of the potential is identical in the two cases. If the air gap is exceedingly narrow, then, approximately,

$$\begin{aligned} H_2 &= M && \text{(in the gap)} \\ H_2 &= 0 && \text{(elsewhere)} \end{aligned} \quad (9-67)$$

However, this result is not consistent with Ampere's circuital law, since

$$\oint H \, dl = \oint (H_1 + H_2) \, dl = NI + Md \neq NI$$

unless d is negligibly small. For a narrow, but not negligibly small, air gap, a better approximation is

$$\begin{aligned} H_2 &= M\left(1 - \frac{d}{l}\right) && \text{(in the gap)} \\ H_2 &= -M\frac{d}{l} && \text{(in the material)} \end{aligned} \quad (9-68)$$

This result not only satisfies Ampere's circuital law, but also provides for the continuity of the normal component of \mathbf{B} across the pole faces.

Combining Eqs. (9-66) and (9-68), and substituting the result in Eq. (9-26),

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$$

we find

$$B = \frac{\mu_0 NI}{l} + \mu_0 M\left(1 - \frac{d}{l}\right) \quad (9-69)$$

both in the gap and in the magnetic material. In order to solve the problem completely, we have only to know the relation

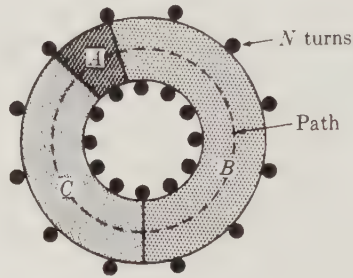
$$M = \chi_m(H)H$$

For "soft iron" χ_m can be taken as a constant.

*9-10 MAGNETIC CIRCUITS

The magnetic flux lines, as we have seen, form closed loops. If all the magnetic flux (or substantially all of it) associated with a particular

FIGURE 9-17
Magnetic circuit



distribution of currents is confined to a rather well-defined path, then we may speak of a *magnetic circuit*. Thus, the examples discussed in Section 9-9 are magnetic circuits, since the magnetic flux is confined to the region inside the toroidal winding. In the first example, the circuit consisted of just one material, a ferromagnetic ring; in the second case, however, we encountered a series circuit of two materials: a ferromagnetic material and an air gap.

Let us consider a more general series circuit of several materials surrounded by a toroidal winding of N turns carrying current I , such as that shown in Fig. 9-17. From the application of Ampere's circuital law to a path following the circuit (the dotted path in the figure),

$$\oint H dl = NI$$

It is convenient to express H at each point along the path in terms of the magnetic flux Φ ; using $B = \mu H$ and $\Phi = BA$, where A is the cross-sectional area of the circuit at the point under consideration, we find

$$\oint \frac{\Phi dl}{\mu A} = NI$$

Since we are dealing with a magnetic circuit, we expect Φ to be essentially constant at all points in the circuit; hence we may take Φ outside the integral:

$$\Phi \oint \frac{dl}{\mu A} = NI \quad (9-70)$$

This result is the basic magnetic circuit equation, which enables us to solve for the flux Φ in terms of the circuit parameters.

Equation (9-70) reminds us of the similar equation for a series current circuit: $IR = \mathcal{V}$. By analogy, we define the magnetomotive force (mmf):

$$\text{mmf} = NI \quad (9-71)$$

and the reluctance \mathcal{R} ,

$$\mathcal{R} = \oint \frac{dl}{\mu A} \quad (9-72)$$

Using these definitions, we may rewrite Eq. (9-70) as

$$\Phi = \frac{\text{mmf}}{\mathcal{R}} \quad (9-70a)$$

If the circuit is made from several homogeneous pieces, each of uniform cross section, the reluctance may be approximated:

$$\mathcal{R} = \sum_j \frac{l_j}{\mu_j A_j} = \sum_j \mathcal{R}_j \quad (9-72a)$$

Hence, the total reluctance of the series circuit is just the sum of the reluctances of the individual elements. The analogy between magnetic and current circuits is even closer than has been indicated, since the resistance of an electric circuit is given by

$$R = \oint \frac{dl}{gA}$$

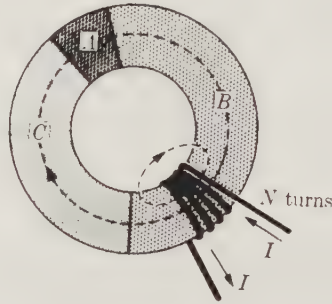
which differs from Eq. (9-72) only through the substitution of g for μ . Because of this analogy, it is apparent that series and parallel reluctance combinations may be combined in the same manner as series and parallel resistance combinations.

The magnetic circuit concept is of most use when applied to circuits containing ferromagnetic materials, but it is for just these materials that we experience a certain amount of difficulty. For a ferromagnetic material, $\mu = \mu(H)$, and we do not know H in the material until the circuit problem is completely solved and Φ determined. The situation is not hopeless, however. In fact, the problem can be solved rather easily by an iterative procedure: (1) As a first guess, we might take $H = NI/l_{\text{total}}$, where l_{total} is the total length of the circuit. (2) The permeability of each material in the circuit is obtained for this value of H from the appropriate magnetization curve. (3) The total reluctance of the circuit is computed, and (4) the flux Φ is calculated from Eq. (9-70a). (5) From Φ , the magnetic intensities in the various elements may be found and the permeabilities redetermined. (6) The procedure is repeated, starting again with item (3). One or two iterations are usually sufficient to determine Φ to within a few percent.

The reluctance \mathcal{R}_j is inversely proportional to the permeability μ_j . Since the permeability of ferromagnetic material may be 100 times μ_0 , $10^3 \mu_0$, or even $10^5 \mu_0$ in certain circumstances, it is apparent that ferromagnetic material forms a low-reluctance path for the magnetic flux. If the magnetic flux encounters two parallel paths, one high reluctance \mathcal{R}_h and the other low reluctance \mathcal{R}_l , then most of the flux will pass through the low reluctance

FIGURE 9-18

Magnetic circuit that is equivalent to the magnetic circuit of Fig. 9-17 if the permeabilities of A , B , and C are large



path, and the equivalent reluctance of the combination is given by $\mathcal{R} = \mathcal{R}_h \mathcal{R}_l / (\mathcal{R}_h + \mathcal{R}_l)$. Looking now at Fig. 9-18, we note that if materials A , B , and C are ferromagnetic, most of the flux will follow the ferromagnetic ring, because the air path between the ends of the solenoid is of relatively high reluctance. Thus, the magnetic circuits of Figs 9-17 and 9-18 are essentially equivalent.

If materials B and C are ferromagnetic, but A represents an air gap, the two circuits are no longer equivalent because there is some *leakage* of flux from the ends of the winding in Fig. 9-18. How much flux leaks out of the circuit depends on the reluctance ratio of magnetic circuit to leakage path. When the air gap A is small compared with the length of the winding, the leakage flux is small and in approximate calculations may be neglected. Reluctance of the leakage path has been worked out for many common geometries, and is given in a number of standard reference books.* The circuit concept is certainly a cruder approximation in the magnetic case than in the electrical one because (1) the ratio of circuit reluctance to leakage reluctance is not as small as the corresponding resistance ratio of the electrical case, and (2) the lateral dimensions of the magnetic circuit are usually not negligible in comparison with its length. Nevertheless, the magnetic circuit concept has proved to be extremely useful.

*9-11

MAGNETIC CIRCUITS CONTAINING PERMANENT MAGNETS

The magnetic circuit concept is useful also when applied to permanent-magnet circuits—that is, to flux circuits in which Φ has its origin in permanently magnetized material. We shall find it convenient to use the abbreviation P-M for permanent magnet. Because of the complicated B - H

* See, for example, S. A. Nasar and L. E. Unnewehr, *Electromechanics and Rotating Electric Machines* (New York: Wiley, 1978); and F. N. Bradley, *Materials for Magnetic Functions* (New York: Hayden Book Co., 1971), p. 162.

relationship in P-M material, the procedure outlined in the preceding section is not well suited to the problem at hand. Instead we start again with Ampere's circuital law, applied now to the flux path of the P-M circuit:

$$\oint H dl = 0$$

or

$$\int_a^b H dl = - \int_{b(\text{P-M})}^a H dl \quad (9-73)$$

In writing Eq. (9-73), we assume explicitly that the P-M material lies between the points b and a of the flux path, whereas from a to b the flux path encounters no P-M material. The use of $B = \mu H$ and $\Phi = BA$ in the left side of Eq. (9-73) yields

$$\Phi \int_a^b \frac{dl}{\mu A} = - \int_{b(\text{P-M})}^a H dl \quad (9-74a)$$

The magnetic flux Φ is continuous throughout the entire circuit, so $\Phi = B_m A_m$, where B_m is the magnetic field in the permanent magnet and A_m is its cross-sectional area. The right side of Eq. (9-74) may be written $-H_m l_m$, where H_m is the average magnetic intensity in the magnet and l_m is the length of the magnet. Thus,

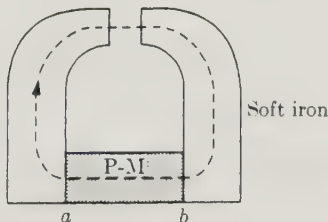
$$B_m A_m \mathcal{R}_{ab} = -H_m l_m \quad (9-74b)$$

is the equation that links the unknown quantities B_m and H_m . This equation can be solved simultaneously with the hysteresis curve of the magnet to yield both B_m and H_m .

As an example of a P-M circuit, consider the circuit composed of a magnet, an air gap, and soft iron (Fig. 9-19). It is important to realize that

FIGURE 9-19

A permanent-magnet circuit. For the circuit shown, the magnet has a rather large demagnetizing field acting upon it; the demagnetizing field can be reduced by increasing the length of P-M material, for example, by placing additional magnets in the side arms of the circuit.



soft iron is *not* a P-M material. Its hysteresis is actually negligible compared with that of the magnet, and $\mu_i = B_i/H_i$ is a positive quantity. The reluctance \mathcal{R}_{ab} is given by

$$\mathcal{R}_{ab} = \frac{l_i}{\mu_i A_i} + \frac{l_g}{\mu_0 A_g} \quad (9-75)$$

where the subscripts i and g refer to the soft iron and air gap, respectively. If the air gap is not too narrow, Eq. (9-75) may usually be approximated by

$$\mathcal{R}_{ab} \approx \frac{l_g}{\mu_0 A_g}$$

which, when combined with (9-74b), yields

$$B_m = -\frac{l_m A_g}{l_g A_m} \mu_0 H_m \quad (9-76)$$

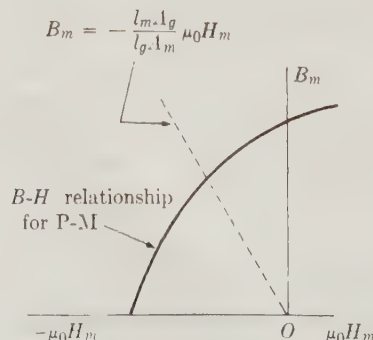
a linear relationship between B_m and H_m . This equation is plotted along with the hysteresis curve of the magnet in Fig. 9-20. The intersection of the two curves gives the operating point of the magnet. The problem is now essentially solved: From a knowledge of B_m , the flux Φ and the flux density B_g are easily determined.

Two points, however, deserve to be mentioned. The first is: What does one use for the effective area A_g of the gap? As a first approximation, we might take A_g equal to the pole-face area of the soft iron, and if the air gap is not too large, this approximation is adequate. We shall not enter into a detailed discussion of this point, but instead refer the interested reader to the references cited in the previous section. Second, the problem of leakage flux is just as important in P-M circuits as it is in other types of magnetic circuits. For the problems presented in this book, however, we shall usually make the assumption that leakage flux may be neglected.

Finally, we note that H_m as determined from Fig. 9-20 is negative—that is, the magnetic intensity in the magnet is a *demagnetizing* effect, which is a

FIGURE 9-20

Demagnetizing line for a magnetic circuit. (The subscript m stands for magnet.) Since $\mu_0 H_m$ is plotted instead of H_m , the slope of the demagnetizing line is just $-(l_m A_g / l_g A_m)$, in other words, a pure number.



general result. When the magnetic flux has its origin in a permanent magnet, then the magnet itself is subjected to a demagnetizing field.

SUMMARY

In Chapter 4, we saw that the response of a (dielectric) medium to an E -field was a polarization charge density $\rho_P = -\nabla \cdot \mathbf{P}$ ($\sigma_P = \mathbf{n} \cdot \mathbf{P}$), where $\mathbf{P} = \chi \mathbf{E}$; in Chapter 7, a (conducting) medium was seen to respond to an E -field with a transport current density $\mathbf{J} = g \mathbf{E}$. We now find that a (magnetic) medium responds to a B -field with another sort of current density, the **atomic magnetization current**,

$$\mathbf{J}_M = \nabla \times \mathbf{M} \quad (\mathbf{j}_M = -\mathbf{n} \times \mathbf{M})$$

where $\mathbf{M} = \Delta \mathbf{m} / \Delta v$ is the magnetic moment per unit volume of the material. The vector potential due to the magnetization is

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{J}_M(\mathbf{r}') dv'}{|\mathbf{r} - \mathbf{r}'|}$$

The **total B-field** due to steady transport currents *plus* magnetization currents satisfies

$$\nabla \times \mathbf{B} = \mu_0(\mathbf{J} + \mathbf{J}_M).$$

Note that $\nabla \cdot \mathbf{J}_M = 0$ identically. It is convenient to define the vector field

$$\mathbf{H} = \frac{1}{\mu_0} \mathbf{B} - \mathbf{M}$$

so that $\nabla \times \mathbf{H} = \mathbf{J}$, with only the conventional transport currents as sources. For a particular medium, the **magnetic susceptibility** in the constitutive equation

$$\mathbf{M} = \chi_m(H) \mathbf{H}$$

must be known. Combined with the definition of \mathbf{H} , the preceding equation gives

$$\mathbf{B} = \mu(H) \mathbf{H}$$

where $\mu = \mu_0[1 + \chi_m(H)]$. Then this relation, together with the differential equations

$$\nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{H} = \mathbf{J}$$

determines \mathbf{B} and \mathbf{H} , subject to the boundary conditions

$$\mathbf{B}_{2n} - \mathbf{B}_{1n} = 0, \quad \mathbf{H}_{2t} - \mathbf{H}_{1t} = \mathbf{j} \times \mathbf{n}_2$$

• Most materials are either **diamagnetic** ($\chi_m < 0$) or **paramagnetic** ($\chi_m > 0$). In either case, $|\chi_m| \ll 1$. The magnetic materials of practical

importance are **ferromagnetic**. For these, $|\chi_m|$ may be more than 1000, but $\mathbf{B} = \mathbf{B}(H)$ is not linear and not even single-valued (hysteresis).

• In problems with no transport currents, since $\nabla \times \mathbf{H} = 0$, it is convenient to use the **scalar potential**,

$$\mathbf{H} = -\nabla\varphi^*$$

Since $\nabla \cdot \mathbf{B} = 0$, $\nabla \cdot \mathbf{H} = -\nabla \cdot \mathbf{M}$, and so φ^* satisfies **Poisson's equation**

$$\nabla^2\varphi^* = \nabla \cdot \mathbf{M}$$

A solution is

$$\varphi^*(\mathbf{r}) = \frac{1}{4\pi} \int_{V_0} \frac{\rho_M(\mathbf{r}') dv'}{|\mathbf{r} - \mathbf{r}'|} + \frac{1}{4\pi} \int_{S_0} \frac{\sigma_M(\mathbf{r}') da'}{|\mathbf{r} - \mathbf{r}'|}$$

where $\rho_M = -\nabla \cdot \mathbf{M}$, $\sigma_M = \mathbf{n} \cdot \mathbf{M}$. (This solution is useful if \mathbf{M} is a given function.)

• In such problems with linear media or with uniform \mathbf{M} , $\nabla \cdot \mathbf{H} = 0$ and φ^* satisfies **Laplace's equation**. These problems are identical to the corresponding electrostatic problems with no external charge density.

• **Ampere's law** relates the \mathbf{H} field to the transport current:

$$\oint \mathbf{H} \cdot d\mathbf{l} = I$$

• The solution for \mathbf{H} divides into one part due to transport currents and another part due to the magnetic materials:

$$\mathbf{H}(\mathbf{r}) = \frac{1}{4\pi} \int \frac{\mathbf{J}(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} dv' - \nabla\varphi^*(\mathbf{r})$$

(Evaluation of the second term depends on knowing $\mathbf{M}(\mathbf{H})$.)

• In the presence of ferromagnetic materials with large μ , it may be a useful approximation to assume that all of the flux Φ is confined to a known volume. Then

$$NI = \Phi\mathcal{R}$$

where the **reluctance**

$$\mathcal{R} = \frac{l}{\mu A}$$

can be calculated for each element of the **magnetic circuit**.

PROBLEMS

9-1 A permanent magnet has the shape of a right circular cylinder of length L . The magnetization \mathbf{M} is uniform and has the direction of the cylinder axis. (a) Find the magnetization current densities, J_M and j_M . (b) Compare the current distribution with that of a solenoid.

9-2 (a) Find the distribution of magnetization currents corresponding to a uniformly magnetized sphere with magnetization \mathbf{M} . According to Eq. (9-63) the magnetic induction \mathbf{B} is uniform inside such a sphere. (b) Can you use this information to design a current winding that will produce a uniform magnetic field in a spherical region of space?

9-3 (a) The magnetic moment of a macroscopic body is defined as $\int_V \mathbf{M} dv$. Prove the relationship

$$\int_V \mathbf{M} dv = \int_V \mathbf{r} \rho_M dv + \oint_S \mathbf{r} \sigma_M da$$

where S is the surface bounding V . [Hint: Refer to the similar problem involving \mathbf{P} in Chapter 4.] (b) A permanent magnet in the shape of a sphere of radius R has uniform magnetization \mathbf{M}_0 in the direction of the polar axis. Determine the magnetic moment of the magnet from both the right and left sides of the equation in part (a).

9-4 (a) Consider a magnet with magnetization specified: $\mathbf{M}(x, y, z)$. Each volume element dv may be treated as a small magnetic dipole, $\mathbf{M} dv$. If the magnet is placed in a uniform magnetic field \mathbf{B}_0 , find the torque on the magnet in terms of its magnetic moment (defined in Problem 9-3). (b) A magnet in the shape of a right circular cylinder of length L and cross section A is uniformly magnetized parallel to the cylinder axis with magnetization \mathbf{M}_0 . The magnet is placed in a uniform magnetic field \mathbf{B}_0 . Find the torque on the magnet in terms of its pole densities.

9-5 An ellipsoid with principal axis of lengths $2a$, $2a$, and $2b$ is magnetized uniformly in a direction parallel to the $2b$ -axis. The magnetization of the ellipsoid is \mathbf{M}_0 . Find the magnetic pole densities for this geometry.

9-6 A spherical shell, with inside radius R_1 and outside radius R_2 , is uniformly magnetized in the direction of the z -axis. The magnetization in the shell is $\mathbf{M}_0 = M_0 \mathbf{k}$. Find the scalar potential φ^* for points on the z -axis, both inside and outside the shell.

9-7 A permanent magnet in the shape of a right circular cylinder of length L and radius R is oriented so that its symmetry axis coincides with the z -axis. The origin of coordinates is at the center of the magnet. If the cylinder has uniform axial magnetization M , (a) determine $\varphi^*(z)$ at points on the symmetry axis, both inside and outside the magnet, and (b) use the results of part (a) to find the magnetic induction B_z at points on the symmetry axis, both inside and outside the magnet.

9-8 A sphere of magnetic material of radius R is placed at the origin of coordinates. The magnetization is given by $\mathbf{M} = (ax^2 + b)\mathbf{i}$, where a and b are constants. Determine all pole densities and magnetization currents.

9-9 A toroid of annealed soft iron has a mean length of 0.1 m and is wound with a coil of 100 turns. Calculate the current required to magnetize the specimen to a field strength of (a) $B = 1.0$ T, (b) $B = 1.5$ T.

9-10 A soft-iron ring with a 1.0 cm air gap is wound with a toroidal winding such as is shown in Fig. 9-16. The mean length of the iron ring is 20 cm, its cross section is 4 cm², and its permeability, assumed constant, is 3000 μ_0 . The 200-turn winding carries a current of 10 A. Find B and H inside the iron ring and in the air gap.

9-11 A long cylinder of radius a and permeability μ is placed in a uniform magnetic field \mathbf{B}_0 such that the cylinder axis is at right angles to \mathbf{B}_0 . (a) Calculate the magnetic

induction inside the cylinder. (b) Make a semiquantitative sketch showing typical lines of induction through the cylinder. (Assume from the beginning that φ^* can be completely specified in terms of the $\cos \theta$ cylindrical harmonics. This assumption is justified, since all boundary conditions can be satisfied in terms of the $\cos \theta$ harmonics.)

9-12 A long straight copper wire and a long straight iron wire each carry the same current I in a uniform B -field, \mathbf{B}_0 . Show that the force on the iron wire is nearly twice the force on the copper wire. [Hint: Use the result of Problem 9-11.]

9-13 A wire carrying a current I is in a cylindrical iron conduit. The conduit has inner and outer radii a and b , has constant susceptibility χ , and is coaxial with the wire. Find the magnetization current density and the total magnetization current.

9-14 Two magnetic media are separated by a plane interface. Show that the angles between the normal to the boundary and the B -fields on either side satisfy

$$\mu_2 \tan \theta_1 = \mu_1 \tan \theta_2$$

9-15 A straight wire carrying a current I is parallel to an infinite plane, at a distance d from the plane. The wire is in air, and the medium on the far side of the plane boundary has constant permeability μ . Find an image current that gives the correct B -field in the air if (a) $\mu = \infty$, (b) $\mu = 0$. (Case (a) approximates a ferromagnetic material and (b) describes a superconductor.)

***9-16** A long cylindrical shell (outside radius b , inside radius a , relative permeability K_m) is oriented normal to a uniform magnetic induction field \mathbf{B}_0 . (a) Show that the magnetic induction \mathbf{B}_i in the vacuum region inside the shell is parallel to \mathbf{B}_0 . (b) Show that the magnetic shielding factor h_m is given by

$$h_m \equiv \frac{B_0}{B_i} = 1 + \frac{(K_m - 1)^2}{4K_m} \left(1 - \frac{a^2}{b^2}\right)$$

9-17 (a) Use the result of Problem 9-16 to calculate the shielding factor h_m of a cylindrical shield of relative permeability $K_m = 5000$ and outer radius $b = 2$ times inner radius a . (b) Show also that the magnetic field at $\theta = 90^\circ$ on the outer surface of the shield is increased to approximately 3.33 times the original field. (Assume the original field is directed along $\theta = 0$.)

9-18 A magnetic circuit in the shape shown in Fig. 9-21 is wound with 100 turns of

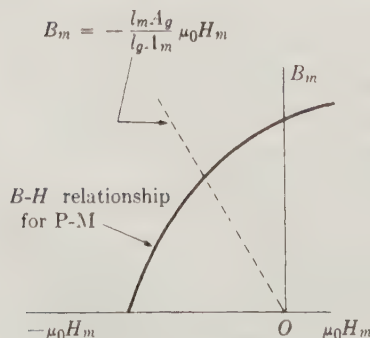


FIGURE 9-21

wire carrying a current of 1 A. The winding is located on the extreme left-hand leg of the circuit. The height of the circuit is 10 cm, its length is 20 cm, the cross section of each leg is 6 cm^2 , and its permeability, assumed constant, is $5000 \mu_0$. Neglecting leakage, calculate the magnetic flux through the central leg and also through the extreme right-hand leg of the circuit.

9-19 A magnetic circuit in the form shown in Fig. 9-19 has a permanent magnet of length 8 cm, a soft-iron path length of 16 cm, and an air gap of 0.8 cm. The cross section of the iron and of the magnet is 4 cm^2 on the average, whereas the effective cross-sectional area of the air gap is 3 cm^2 . The relative permeability of the iron is 5000. (a) Calculate the magnetic flux density in the gap for two different magnet materials: sintered oxide and 35 percent Co steel. Neglect leakage. (b) The dimensions of the magnetic circuit are altered in *one* respect: The air gap is decreased to 0.8 mm. Repeat the calculation called for in part (a).

9-20 Find the magnetic induction in a uniformly magnetized sphere for each of the materials shown in Fig. 9-9.

9-21 A magnetic circuit in the form shown in Fig. 9-19 has an Alnico V magnet of length 10 cm, a soft-iron path of 16 cm, and an air gap of 1 cm. It is also wound with 800 ampere-turns of wire (in a direction to aid the flux produced by the magnet). Find the magnetic flux density in the air gap. (Neglect leakage, take $K_m = 5000$ for the soft iron, and assume that the cross sections of the magnet, soft iron, and air gap are the same.)

Microscopic Theory of Magnetism

In the preceding chapter, we were concerned with the macroscopic aspects of magnetization. The magnetic properties of matter were introduced explicitly through the function \mathbf{M} , which was related to the magnetic induction by means of experimentally determined parameters. In this chapter, we look at matter from the microscopic point of view (i.e., as an assembly of atoms or molecules) and see how the individual molecules respond to an imposed magnetic field. If this procedure were carried through completely, we would end up with theoretical expressions for susceptibility and with B - H relationships for all types of materials. While such a procedure is certainly beyond the scope of this book, we can show rather simply how the various kinds of magnetic behavior come about and, in addition, derive expressions that predict the correct order of magnitude for susceptibility in certain cases. A more thorough discussion of the topics presented here is to be found in books on solid state physics.†

In the macroscopic formulations in Chapter 9, we dealt with two field vectors, \mathbf{B} and \mathbf{H} , which we related through the equation $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$. From the microscopic viewpoint, the distinction between \mathbf{B} and \mathbf{H} largely disappears because we deal with an assembly of molecules (i.e., with an assembly of magnetic dipoles or dipole groups) in vacuum. We are concerned with the magnetic field near a molecule in vacuum or at the position of a molecule when that molecule is removed from the system. Thus, $\mathbf{B}_m = \mu_0\mathbf{H}_m$. Here the subscript m denotes microscopic, but in the following sections of this chapter the symbol \mathbf{B}_m (and \mathbf{H}_m) will denote a

* This chapter may be omitted without loss of continuity.

† See, for example, C. Kittel, *Introduction to Solid State Physics*, 6th ed. (New York: Wiley, 1986), Chapters 14 and 15.

particular value of the microscopic field, namely, the field at the position of a molecule.

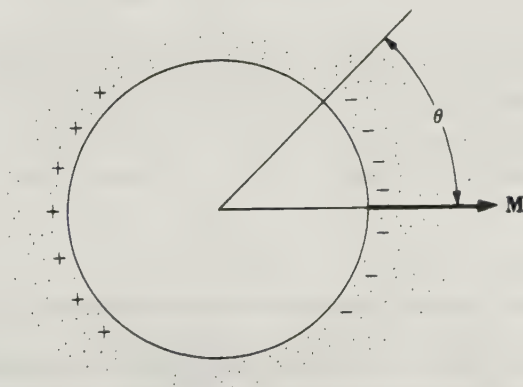
It is customary when discussing the microscopic field inside matter to relate \mathbf{H}_m to the macroscopic \mathbf{H} field, instead of \mathbf{B}_m to the \mathbf{B} field, because both \mathbf{H} and \mathbf{H}_m can be written simply in terms of integrals over the current and dipole distributions. It makes very little difference, however, whether we calculated \mathbf{H}_m or \mathbf{B}_m , since they differ from each other only by the scale factor μ_0 .

10-1 MOLECULAR FIELD INSIDE MATTER

The magnetic field that is effective in its interaction with atomic currents in an atom or molecule is the molecular field $\mathbf{B}_m = \mu_0 \mathbf{H}_m$. Called the *local field* in some textbooks, it is the magnetic field at a molecular (or atomic) position in the material. This field is produced by all external sources and by all molecular dipoles in the material *with the exception* of the one molecule (or atom) at the point under consideration. It is evident that \mathbf{B}_m need not be the same as the macroscopic magnetic induction field, since the latter quantity is related to the force on a current element whose dimensions are large compared with molecular dimensions.

The molecular field may be calculated by a procedure similar to that of Section 5-1 for the molecular electric field in a dielectric. We consider a material object of arbitrary shape, which, for convenience, we take to be uniformly magnetized with magnetization \mathbf{M} . Let us cut out a small piece of the object, leaving a spherical cavity surrounding the point at which the molecular field is to be computed (see Fig. 10-1). The material that is left is to be treated as a continuum—that is, from the macroscopic point of view. Next we put the material back into the cavity, molecule by molecule, except

FIGURE 10-1
Calculation of “cavity” surface contribution to H_m . + and - represent magnetic charge (i.e., N- and S-seeking poles).



for the molecule at the center of the cavity, where we wish to compute the molecular field. The molecules that have just been replaced are to be treated, not as a continuum, but as individual dipoles or dipole groups.

The *macroscopic* field \mathbf{H} , the magnetic intensity in the specimen, can be expressed, according to Eq. (9-27), as

$$\mathbf{H} = \frac{1}{4\pi} \int \frac{\mathbf{J} \times (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} dv' + \frac{1}{4\pi} \int \rho_M \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} dv' + \frac{1}{4\pi} \int_S \sigma_M \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} da'$$

where the integrals extend over all sources: \mathbf{J} , ρ_M , and σ_M . The molecular field \mathbf{H}_m may be expressed in a similar way, except that now there are additional contributions from the surface of the cavity and from the individual dipoles in the cavity. The integral of $\rho_M(\mathbf{r} - \mathbf{r}')/|\mathbf{r} - \mathbf{r}'|^3$ over the cavity volume need not be excluded specifically, since $\rho_M = -\nabla \cdot \mathbf{M} = 0$ in the uniformly magnetized specimen. Thus,

$$\mathbf{H}_m = \mathbf{H} + \mathbf{H}_s + \mathbf{H}' \quad (10-1)$$

where \mathbf{H} is the macroscopic magnetic intensity in the specimen, \mathbf{H}_s is the contribution from the surface pole density $\sigma_M = M_n$ on the cavity surface (refer to Fig. 10-1), and \mathbf{H}' is the contribution of the various dipoles inside the cavity.

From the corresponding derivation in Section 5-1, \mathbf{H}_s is seen to be

$$\mathbf{H}_s = \frac{1}{3}\mathbf{M} \quad (10-2)$$

Furthermore, the contribution from the dipoles in the cavity,

$$\mathbf{H}' = \frac{1}{4\pi} \sum_i \left[\frac{3(\mathbf{m}_i \cdot \mathbf{r}_i)\mathbf{r}_i}{r_i^5} - \frac{\mathbf{m}_i}{r_i^3} \right] \quad (10-3)$$

where \mathbf{r}_i is the distance from the i th dipole to the center of the cavity, is of the same form as the corresponding electric dipole term \mathbf{E}' in Section 5-1.

Thus, if we restrict our interest to the rather large class of materials for which Eq. (10-3) vanishes, Eq. (10-1) for the molecular field reduces to

$$\mathbf{H}_m = \mathbf{H} + \frac{1}{3}\mathbf{M} \quad (10-4)$$

and

$$\mathbf{B}_m = \mu_0 \mathbf{H}_m \quad (10-5)$$

Equations (10-4) and (10-5) give the molecular field in terms of the macroscopic magnetic intensity and the magnetization in the sample. For

most diamagnetic and paramagnetic materials the term $\frac{1}{3}\mathbf{M} = \frac{1}{3}\chi_m\mathbf{H}$ is negligibly small, but for ferromagnetic materials the correction is quite important.

10-2 ORIGIN OF DIAMAGNETISM

In order to calculate the diamagnetic susceptibility of an assembly of atoms, we must know something about the electronic motion in the atom itself. We shall assume that each electron circulates around the atomic nucleus in some kind of orbit, and for simplicity, we choose a circular orbit of radius R in a plane at right angles to the applied magnetic field. Quantum mechanics tells us that although this picture is approximately correct, the electrons do not circulate in well-defined orbits. To solve the problem properly, we would have to solve the Schrodinger equation for an atomic electron in a magnetic field. Nevertheless, our rather naive "classical" calculation will give the correct order of magnitude for the diamagnetic susceptibility.

Before the magnetic induction field is applied, the electron is in equilibrium in its orbit:

$$F_q = m_e\omega_0^2R, \quad (10-6)$$

where F_q is the electric force holding the electron to its atom, ω_0 is the angular frequency of the electron in its orbit, and m_e is the electron mass. Application of a magnetic field exerts an additional force $-e\mathbf{v} \times \mathbf{B}_m$ on the electron. Assuming that the electron stays in the same orbit, we find

$$F_q \pm e\omega RB_m = m_e\omega^2R$$

which, when combined with Eq. (10-6), yields

$$\pm e\omega B_m = m_e(\omega - \omega_0)(\omega + \omega_0) \quad (10-7)$$

The quantity $\Delta\omega = \omega - \omega_0$ is the change in angular frequency of the electron. Thus, the electron either speeds up or slows down in its orbit, depending on the detailed geometry (i.e., on the direction of $\mathbf{v} \times \mathbf{B}_m$ relative to \mathbf{F}_q), but in either case the *change* in orbital magnetic moment is in a direction *opposite* to the applied field. This statement may be easily verified by the reader.

Even for the largest fields that can be obtained in the laboratory (~ 100 T), $\Delta\omega$ is small compared with ω_0 , so that Eq. (10-7) may always be approximated by

$$\Delta\omega = \pm \frac{e}{2m_e} B_m \quad (10-8)$$

The quantity $(e/2m_e)B_m$ is known as the *Larmor frequency*.

Up to this point we have merely *assumed* that the electron stays in the same orbit. We have used this assumption together with the equilibrium of forces to derive Eq. (10-8). For the electron to stay in its orbit, the change

in its kinetic energy as determined from Faraday's law of induction must be consistent with Eq. (10-8).^{*} When the magnetic field is switched on, there is a change in flux through the orbit given by $\pi R^2 \Delta B_m$. This flux is linked by Δn electron loops, where Δn is the number of revolutions made by the electron during the time in which the field changes. The changing flux produces an emf according to Faraday's law:

$$\mathcal{E} = \pi R^2 \frac{dB_m}{dt} \Delta n = \pi R^2 \frac{dn}{dt} \Delta B_m \quad (10-9)$$

The energy given to the electron in this process is $\mathcal{E}e$, and it appears as a change in kinetic energy:

$$\frac{1}{2} m_e R^2 (\omega^2 - \omega_0^2) = e \pi R^2 \frac{dn}{dt} \Delta B_m \quad (10-10)$$

But ΔB_m is just the final value of the field B_m , and the average value of $dn/dt = (\omega + \omega_0)/4\pi$. Thus,

$$\Delta \omega = \frac{e}{2m_e} B_m$$

in agreement with Eq. (10-8). Thus, the assumption of a constant orbit leads to no contradiction between Eq. (10-9) and the force equation. Diamagnetism is the result of Lenz's law operating on an atomic scale. Upon the application of a magnetic field, the electronic currents in each atom are modified in such a way that they tend to weaken the effect of this field.

The change in angular velocity predicted by Eq. (10-8) produces a change in magnetic moment given by

$$\Delta \mathbf{m} = -\frac{e}{2\pi} \pi R^2 \frac{e}{2m_e} \mathbf{B}_m = -\frac{e^2}{4m_e} R^2 \mu_0 \mathbf{H}_m \quad (10-11)$$

In order to find the magnetization, this result must be summed over all electrons in a unit volume. For a substance containing N molecules per unit volume, all of the same molecular species,

$$\mathbf{M} = -\frac{Ne^2 \mu_0}{4m_e} \mathbf{H}_m \sum_i R_i^2 \quad (10-12)$$

where the summation is over the electrons in one molecule. For diamagnetic materials, \mathbf{H}_m differs very little from \mathbf{H} , so the diamagnetic susceptibility

$$\chi_m = -\frac{Ne^2 \mu_0}{4m_e} \sum_i R_i^2 \quad (10-13a)$$

This result has been obtained on the assumption that all electrons circulate

^{*} Faraday's law of induction and Lenz's law are discussed in Chapter 11, Section 11-1.

in planes perpendicular to the field \mathbf{H}_m . When the orbit is inclined, so that a normal to the orbit makes an angle θ_i with the field, only the component of \mathbf{H}_m along this normal ($H_m \cos \theta_i$) is effective in altering the angular velocity of the electron. Furthermore, the component of $\Delta \mathbf{m}$ parallel to the field is smaller by the factor $\cos \theta_i$. Hence a better approximation to the diamagnetic susceptibility is

$$\chi_m = -\frac{Ne^2\mu_0}{4m_e} \sum_i R_i^2 \cos^2 \theta_i \quad (10-13b)$$

Diamagnetism is presumably present in all types of matter, but its effect is frequently masked by stronger paramagnetic or ferromagnetic behavior that can occur in the material simultaneously. Diamagnetism is particularly prominent in materials that consist entirely of atoms or ions with "closed electron shells," since in these cases all paramagnetic contributions cancel out.

10-3 ORIGIN OF PARAMAGNETISM

The orbital motion of each electron in an atom or molecule can be described in terms of a magnetic dipole moment as follows directly from Eq. (8-22). In addition, it is known that the electron has an intrinsic property called *spin*, and an intrinsic magnetic dipole moment associated with this spinning charge. Each molecule, then, has a magnetic moment \mathbf{m}_i which is the vector sum of orbital and spin moments from the various electrons in the molecule. Briefly, paramagnetism results from the tendency of these molecular moments to align themselves with the applied field, just as the current circuit of Eq. (8-19) tends to align itself with the field.

The situation is not quite so straightforward as that for a current circuit, however. There are, in fact, two complications: (1) In the presence of a magnetic field, the electronic motions are quantized such that each orbital and spin moment has only a discrete set of orientations relative to the field direction. Furthermore, no two electrons in the molecule can occupy the same quantum state, so that if there are just enough electrons per molecule to fill "electron shells," then all possible orientations must be used and \mathbf{m}_i is zero. Of course, paramagnetism can occur only when $\mathbf{m}_i \neq 0$. (2) The electronic motion inside an atom that gives rise to \mathbf{m}_i also produces an angular momentum about the atomic nucleus. In fact, \mathbf{m}_i is linearly related to this angular momentum. Under these conditions the magnetic torque does not directly align the dipole moment \mathbf{m}_i with the field, but causes it to precess around the field at constant inclination.* The atoms (or molecules)

* A discussion of the precession of \mathbf{m}_i in a uniform magnetic field is given in many textbooks. See, for example, H. Goldstein, *Classical Mechanics* (Reading, Mass.: Addison-Wesley, 1950), pp. 176–177.

in our material system are in thermal contact with each other. In a gas or liquid, the atoms are continually making collisions with one another, and in a solid, the atoms are undergoing thermal oscillation. Under these conditions, the various \mathbf{m}_i can interchange magnetic energy with the thermal energy of their environment and make transitions from one precessional state to another of a different inclination. The thermal energy of the system tends to produce a completely random orientation of the \mathbf{m}_i , but orientations along or near the field direction have a lower magnetic energy and thus are favored. The situation is quite similar to that of polar molecules in an electric field, which was discussed in Section 5-3.

For a material composed entirely of one molecular species, each molecule having magnetic moment m_0 , the fractional orientation is given approximately by the Langevin function, Eq. (5-21), with

$$y = \frac{m_0 \mu_0 H_m}{kT} \quad (10-14)$$

The magnetization is given by

$$|\mathbf{M}| = Nm_0 \left[\coth y - \frac{1}{y} \right] \quad (10-15a)$$

where N is the number of molecules per unit volume. Except for temperatures near absolute zero, the Langevin function can be approximated by the first term in its power series:

$$\mathbf{M} = \frac{Nm_0^2}{3kT} \mu_0 \mathbf{H}_m \quad (10-15b)$$

which yields the paramagnetic susceptibility

$$\chi_m = \frac{Nm_0^2 \mu_0}{3kT} \quad (10-16)$$

According to atomic theory, m_0 is in the range of a few Bohr magnetons (1 Bohr magneton = $eh/4\pi m_e$, where h is Planck's constant). Equations (10-16) and (10-13b) account for the order of magnitude of the χ_m 's in Table 9-1.

We may summarize the results of this section briefly as follows: In order to exhibit paramagnetic behavior, the atoms (or molecules) of the system must have permanent magnetic moments, and these tend to orient in the applied field. The various molecular moments are decoupled—that is, they precess around the magnetic field as individuals (not in unison)—but they are able to exchange energy because of thermal contact with their environment. Except for temperatures near absolute zero and simultaneous large fields, the magnetization is far below the saturation value which would obtain when all the dipole moments are aligned.

THEORY OF FERROMAGNETISM

In ferromagnetic materials, the permanent atomic (or molecular) magnetic moments are very nearly aligned even in the absence of an applied field. The cause of this alignment is the molecular field \mathbf{H}_m which, according to Eq. (10-4), does not vanish when $\mathbf{H} = 0$ unless \mathbf{M} vanishes simultaneously. A magnetization \mathbf{M} does give rise to a molecular field, but unless this molecular field produces the same magnetization \mathbf{M} that is presumed to exist in the material, the solution is inconsistent. Our problem is to determine in what set of circumstances the magnetization can maintain itself via the molecular field.

It will prove necessary to generalize Eq. (10-4) to a certain extent. For the molecular field, let us write $\mathbf{H}_m = \mathbf{H} + \gamma\mathbf{M}$, which, for $\mathbf{H} = 0$, reduces to

$$\mathbf{H}_m = \gamma\mathbf{M} \quad (10-4a)$$

According to the simple theory of Section 10-1, $\gamma = \frac{1}{3}$. If the terms in Eq. (10-3) do not sum to zero, γ may be different from $\frac{1}{3}$; nevertheless, we expect γ to be of this order of magnitude.

Let us restrict our attention to a material composed entirely of one atomic species, each atom having magnetic moment m_0 . There are N atoms per unit volume. If the atomic moments are to be very nearly aligned, M must be a substantial fraction of Nm_0 . For the sake of definiteness, however, let us say

$$M > 0.7Nm_0 \quad (10-17)$$

According to Eq. (10-15), this result implies that $[\coth y - (1/y)] > 0.7$, or y [which is defined by Eq. (10-14)] > 3 . Thus,

$$y = \frac{m_0\mu_0 H_m}{kT} > 3$$

which, when combined with Eqs. (10-4a) and (10-17), yields

$$0.7 \frac{\gamma N \mu_0 m_0^2}{kT} > 3 \quad (10-18)$$

This result (approximately) is the condition for the occurrence of ferromagnetism.

In the previous section, it was stated that atomic theory predicts m_0 to be in the range of a few Bohr magnetons. On this basis, Eq. (10-18) requires a γ of about 10^3 in the vicinity of room temperature, which is orders of magnitude larger than can be accounted for in the derivation presented in Section 10-1. It would thus appear that the origin of ferromagnetism is considerably more complex than the corresponding situation in ferroelectrics (discussed in Section 5-4).

In 1907 Pierre Weiss* formulated his theory of ferromagnetism. Weiss appreciated the essential role played by the molecular field. He could not explain the large value of γ , but he accepted it as a fact and proceeded to develop this theory from his point. The predictions of his theory were found to be in close accord with experiment. For this reason, the molecular field of Eq. (10-4a) is often called the *Weiss molecular field*.

It was left to Heisenberg,† some twenty years later, to explain the origin of the large value of γ . Heisenberg showed, first, that it is only the spin magnetic moments that contribute to the molecular field and, second, that the field is produced basically by electrostatic forces. On the basis of quantum mechanics, he showed that when the spins on neighboring atoms change from parallel alignment to antiparallel alignment, there must be a simultaneous change in the electron charge distribution in the atoms.‡ The change in charge distribution alters the electrostatic energy of the system and, in certain cases, favors parallel alignment (i.e., ferromagnetism). A spin-dependent energy—that is, an energy that depends on the spin configuration of the system—can be viewed in terms of the force (or torque) that is produced on one of the atoms when the configuration is altered. The equivalent field turns out to be proportional to \mathbf{M} , but with a coefficient that depends in detail upon the charge distribution in the atom under consideration.

The Weiss-Heisenberg theory can be used to predict the way in which the magnetization of a ferromagnet changes with temperature. It is evident that the theory depicts ferromagnetism as the limiting case of paramagnetism in an extremely large magnetic field, but with this field coming from the magnetization itself. Combining Eq. (10-4a) with Eqs. (10-14) and (10-15) yields

$$M = Nm_0 \left[\coth y - \frac{1}{y} \right] \quad (10-19)$$

and

$$M = \frac{kTy}{\gamma\mu_0 m_0} \quad (10-20)$$

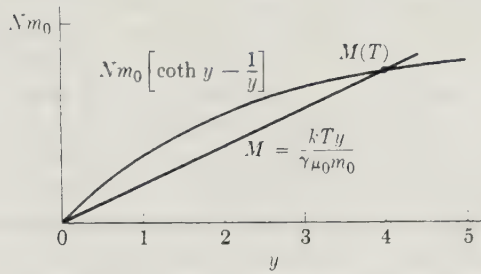
The *spontaneous magnetization*—that is, the magnetization at zero external field—for a given temperature is obtained from the simultaneous solution of Eqs. (10-19) and (10-20). This solution is easily obtained by a graphical procedure: We plot M versus y for *both* (10-19) and (10-20), as shown in Fig. 10-2. The intersection of the two curves gives a magnetization $M(T)$

* P. Weiss, *Journal de Physique*, **6**, 667 (1907).

† W. Heisenberg, *Zeitschrift für Physik*, **49**, 619 (1928).

‡ This change in charge distribution is a consequence of the Pauli exclusion principle.

FIGURE 10-2
Determination of the spontaneous magnetization $M(T)$ with the aid of the Langevin function



that is consistent with both equations. As the temperature is increased, the linear curve, Eq. (10-20), becomes steeper, but Eq. (10-19) is unchanged. Thus, the intersection point moves to the left in the figure, and a lower value for the spontaneous magnetization obtains. Finally, a temperature is reached at which Eq. (10-20) is tangent to Eq. (10-19) at the origin, and at this and higher temperatures, the spontaneous magnetization is zero. This temperature is the *Curie temperature*, T_c , above which the spontaneous magnetization vanishes and ordinary paramagnetic behavior results.

A plot of $M(T)$ versus temperature, obtained according to the above procedure, is displayed in Fig. 10-3. It is in approximate agreement with experimentally determined values of the spontaneous magnetization for a ferromagnetic material.

10-5 FERROMAGNETIC DOMAINS

According to the preceding section, a ferromagnetic specimen should be magnetized very nearly to saturation (regardless of its previous history) at

FIGURE 10-3
Magnetization of a ferromagnetic material as a function of temperature. T_c is called the Curie temperature. (The curve shown has been calculated with the aid of the classical Langevin function; quantum-mechanical corrections change the shape of the curve somewhat, bringing it into agreement with experimental data.)

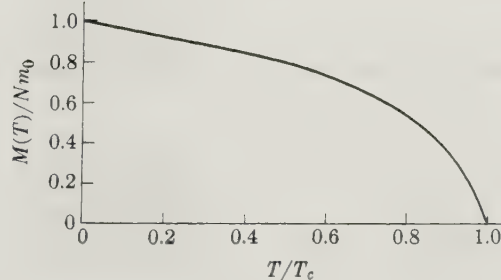
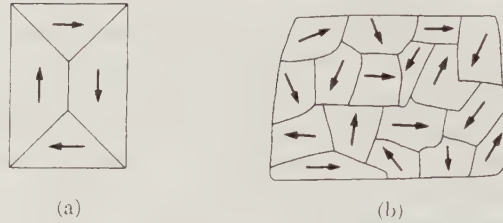


FIGURE 10-4

Ferromagnetic domain structures: (a) single crystal, (b) polycrystalline specimen. Arrows represent the direction of magnetization.

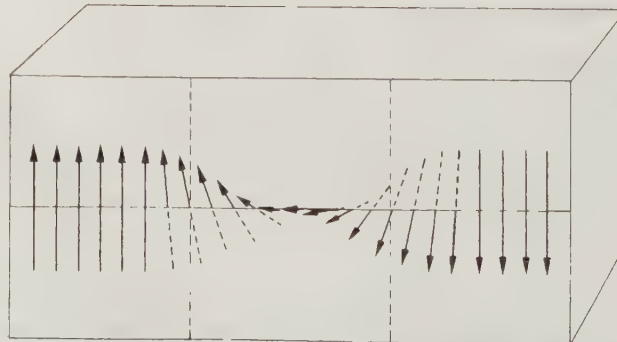


temperatures below the Curie temperature. This statement appears to be contrary to observation. We know, for example, that a piece of iron can exist in either a magnetized or unmagnetized condition. The answer to this apparent paradox is that a ferromagnetic material breaks up into *domains*. Each domain is fully magnetized in accord with the results of the preceding section, but the various domains can be randomly oriented (Fig. 10-4) and thus present an unmagnetized appearance from the macroscopic point of view. The presence of domains was first postulated by Weiss in 1907.

In passing from one domain to an adjacent one, the atomic moment vector \mathbf{m}_0 gradually rotates from its original to its new direction in the course of about 100 atoms (Fig. 10-5). This region between the two domains is called a *domain wall*. It would appear that an atomic spin moment in the wall region is subjected to a slightly lower molecular field than is an atomic spin moment inside the domain proper. This observation by itself would favor a single domain configuration. On the other hand, a specimen consisting of a single domain must maintain a large external magnetic field, whereas a multidomain specimen has a lower “magnetic energy” associated with its field structure. Thus, the multidomain structure is usually energetically favored.

FIGURE 10-5

Structure of transition region or “Bloch wall,” between domains in a ferromagnetic material



The macroscopic aspects of magnetization in ferromagnetic materials are concerned with changes in domain configuration. The increase in magnetization resulting from the action of an applied magnetic field takes place by two independent processes: by an increase in the volume of domains that are favorably oriented relative to the field at the expense of domains that are unfavorably oriented (domain wall motion), or by rotation of the domain magnetization toward the field direction. The two processes are illustrated schematically in Fig. 10-6.

In weak applied fields the magnetization usually changes by means of domain wall motion. In pure materials consisting of a single phase, the wall motion is to a large extent reversible in the weak-field region. In stronger fields the magnetization proceeds by irreversible wall motion, and finally by domain rotation. In these circumstances the substance remains magnetized when the external magnetic field is removed.

The experimental study of domains was made possible by a technique first developed by F. H. Bitter.* A finely divided magnetic powder is spread over the surface of the specimen, and the powder particles, which collect along the domain boundaries, may be viewed under a microscope. By means of this technique, it has even proved possible to observe domain wall motion under the action of an applied magnetic field. The size of domains varies widely, depending on the type of material, its previous history, and so on. Typical values are in the range from 10^{-6} to 10^{-2} cm³.

* F. H. Bitter, *Physical Review*, **41**, 507 (1932). For a brief discussion of the technique, see D. Jiles, *Magnetism and Magnetic Materials* (London: Chapman and Hall, 1991), p. 114.

FIGURE 10-6

Magnetization of a ferromagnetic material:
 (a) unmagnetized,
 (b) magnetization by domain wall motion,
 (c) magnetization by domain rotation

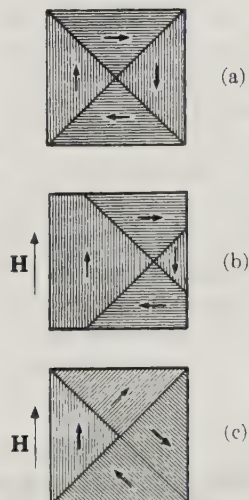
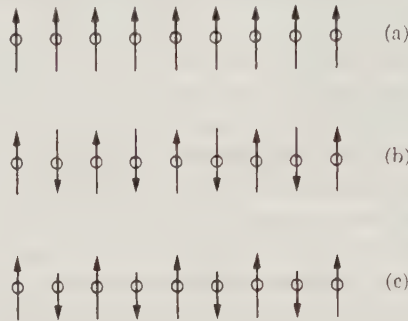


FIGURE 10-7

Schematic representation
of atomic spins in or-
dered spin structures:

(a) ferromagnetic, (b)
antiferromagnetic,
(c) ferrimagnetic

**10-6****FERRITES**

According to the Heisenberg theory of ferromagnetism, there is a change in electrostatic energy associated with the change from parallel to antiparallel spin alignment of neighboring atoms. If this energy change favors parallel alignment and is at the same time of sufficient magnitude, the material composed of these atoms is ferromagnetic. If the energy change favors antiparallel alignment, it is still possible to obtain an *ordered* spin structure, but with spins alternating from atom to atom as the crystal is traversed.

An ordered spin structure with zero net magnetic moment is called an *antiferromagnet* (Fig. 10-7b). The most general ordered spin structure contains both “spin-up” and “spin-down” components but has a net, nonzero magnetic moment in one of these directions. Such a material is called a *ferrimagnet* or simply a *ferrite*. The simplest ferrites of magnetic interest are oxides represented by the chemical formula $M\text{OFe}_2\text{O}_3$, where M is a divalent metal ion such as Co, Ni, Mn, Cu, Mg, Zn, Cd, or divalent iron. These ferrites crystallize in a rather complicated crystal structure known as the *spinel structure*. The classic example of a ferrite is the mineral magnetite (Fe_3O_4), which has been known since ancient times.

Ferrites are of considerable technical importance because, in addition to their relatively large saturation magnetization, they are poor conductors of electricity. Thus they can be used for high-frequency applications where the eddy-current losses in conducting materials pose serious problems. Typical resistivities of ferrites fall in the range from 1 to $10^4 \Omega \cdot \text{m}$; for comparison, the electrical resistivity of iron is approximately $10^{-7} \Omega \cdot \text{m}$.

10-7**SUMMARY**

The macroscopic magnetization M of a magnetic material results from the molecular magnetic dipole moment (or its component), which arises in response to the local field at the molecule—the **molecular field** H_m . The

molecular field depends on the applied field H and also on the magnetization itself. The latter contribution, which stems from the dipole magnetic fields of all the other molecules, to give

$$H_m = H + \frac{1}{3}M$$

in analogy to the dielectric case, is negligibly small for most linear materials because of the smallness of the magnetic susceptibility in

$$M = \chi_m H$$

Nevertheless spontaneous magnetization occurs in ferromagnetic materials because the contribution of the magnetization to the effective molecular field has a coefficient much greater than $\frac{1}{3}$.

- All molecules exhibit an induced magnetic dipole moment in a magnetic field due to the *deformation* of the electronic current distribution. The response is always such as to weaken the applied field—that is, the (diamagnetic) contribution to the susceptibility is always negative. A linear approximation leads to the constant **diamagnetic susceptibility**,

$$\chi_m = -\frac{Ne^2\mu_0}{4m_e} \sum_i R_i^2$$

- Molecules that have a permanent magnetic dipole moment m_0 show an additional *orientational* response. This response is described approximately by the Langevin function, as for polar molecules in an electric field. Except near absolute zero, the resulting **paramagnetic susceptibility** is

$$\chi_m = \frac{Nm_0^2\mu_0}{3kT}$$

- In order to understand **ferromagnetism**, one assumes that

$$H_m = H + \gamma M$$

with $\gamma \gg \frac{1}{3}$. (This contribution derives from a quantum mechanical energy that depends on the relative orientation of spin magnetic moments; it adds to the magnetic energy $\mathbf{m}_0 \cdot \mathbf{H}$ and can therefore be expressed in terms of an effective magnetic field, even though it is electrostatic in origin.) Then this equation and the Langevin equation admit a solution with $H = 0$, $M \neq 0$, so long as T is below the **Curie temperature**.

- Even below the Curie temperature, a macroscopic piece of ferromagnetic material may show no net magnetic moment because of its **domain structure**.

PROBLEMS

10-1 A Bohr magneton is defined as the magnetic moment of an electron circulating in the classic “Bohr orbit” of the hydrogen atom. This orbit is a circular orbit of

exactly one *de Broglie wavelength*, for which the coulomb attraction provides the centripetal acceleration. Show that 1 Bohr magneton = $eh/4\pi m_e$, where m_e is the mass of the electron and h is Planck's constant.

10-2 The Bohr magneton is a natural unit for measuring the magnetic moment of an atom. Calculate the magnetic moment per atom, in Bohr magneton units, for iron, nickel, and cobalt, under conditions of saturation magnetization. Use the data in Table 9-2.

10-3 Calculate the relative strength of the interaction between two typical magnetic dipoles, compared with the interaction between two typical electric dipoles. To be explicit, calculate the torque exerted on one dipole by the other when they are oriented perpendicularly to each other at a distance of one angstrom unit; take each magnetic dipole = 1 Bohr magneton, each electric dipole = $e \times 0.1$ angstrom. This calculation shows that the basic magnetic interaction is several orders of magnitude smaller than the electrical interaction in matter.

10-4 Calculate the diamagnetic susceptibility of neon at standard temperature and pressure (0°C, 1 atm) on the assumption that only the eight outer electrons in each atom contribute, and that their mean radius is $R = 4.0 \times 10^{-9}$ cm.

10-5 The magnetization of a ferromagnetic material drops essentially to zero at the Curie temperature. In Fig. 10-2, the Curie temperature is represented by the straight line that is tangent to the Langevin function at the origin. Use the experimental value of the Curie temperature of iron to determine γ for iron.

10-6 The gyromagnetic ratio of a current distribution is defined as the ratio of magnetic moment to angular momentum. Calculate the gyromagnetic ratio of a sphere of mass M and charge Q which is rotating with angular velocity ω about an axis through its center, on the assumption that the mass is distributed uniformly throughout and the charge is distributed uniformly on the surface of the sphere.

Electromagnetic Induction

The induction of electromotive force by changing magnetic flux was first observed by Faraday and by Henry in the early nineteenth century. From their pioneering experiments have developed modern generators, transformers, and so forth. This chapter is primarily concerned with the mathematical formulation of the law of electromagnetic induction and its exploitation in simple cases.

The equation that characterized electrostatics was

$$\nabla \times \mathbf{E} = 0$$

or, in integral form,

$$\oint \mathbf{E} \cdot d\mathbf{l} = 0$$

These equations follow immediately from Coulomb's law, and they are not spoiled by the magnetic force due to a steady current. They do not hold for more general time-dependent fields, however, and these cases are what we will now consider.

ELECTROMAGNETIC INDUCTION

Let us first introduce the concept of electromotive force.

We define the *electromotive force*, or *emf*, around a circuit by

$$\oint_C \mathbf{E} \cdot d\mathbf{l} = \mathcal{E} \quad (11-1)$$

With static E - and B -fields, this emf was always zero. We now take up cases where it is not zero. Since this E -field cannot be defined from Coulomb's law, it is legitimate to ask what does define it. It is defined so that the Lorentz force

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

is *always* the electromagnetic force on a test charge q .

The results of a large number of experiments can be summarized by associating an induced emf

$$\mathcal{E} = -\frac{d\Phi}{dt} \quad (11-2)$$

with a change in magnetic flux through a circuit. This result, which is known as *Faraday's law of electromagnetic induction*, is found to be independent of the way in which the flux is changed—the value of \mathbf{B} at various points inside the circuit may be changed in any way.

It is extremely important to realize that Eq. (11-2) represents an independent experimental law. It cannot be derived from other experimental laws, and it certainly is not, as is sometimes stated, a consequence of conservation of energy applied to the energy balance of currents in magnetic fields.

Since by definition

$$\mathcal{E} = \oint_C \mathbf{E} \cdot d\mathbf{l} \quad (11-1)$$

and

$$\Phi = \int_S \mathbf{B} \cdot \mathbf{n} \, da \quad (11-3)$$

Eq. (11-2) can be written

$$\oint_C \mathbf{E} \cdot d\mathbf{l} = -\frac{d}{dt} \int_S \mathbf{B} \cdot \mathbf{n} \, da \quad (11-4)$$

If the circuit is a rigid stationary circuit, the time derivative can be taken inside the integral, where it becomes a partial time derivative. Furthermore, Stokes's theorem can be used to transform the line integral of \mathbf{E} into the surface integral of $\nabla \times \mathbf{E}$. The result of these transformations is

$$\int_S \nabla \times \mathbf{E} \cdot \mathbf{n} \, da = -\int_S \frac{\partial \mathbf{B}}{\partial t} \cdot \mathbf{n} \, da \quad (11-5)$$

If Faraday's law in the form of Eq. (11-5) holds for all fixed surfaces S , it follows that

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (11-6)$$

This result is the *differential form of Faraday's law*.

It is the required generalization of $\nabla \times \mathbf{E} = 0$, which held for static fields. (Moving media and other subtleties require a more careful treatment, beyond the scope of this text.)

The negative sign in Faraday's law indicates, as can be easily demonstrated, that the direction of the induced emf is such as to oppose the change that produces it. Thus, if we attempt to increase the flux through a circuit, the induced emf tends to cause currents in such a direction as to decrease the flux. If we attempt to thrust one pole of a magnet into a coil, the currents caused by the induced emf set up a magnetic field which tends to repel the pole. All these phenomena are covered by the following law.

Lenz's law. In case of a change in a magnetic system, that thing happens which tends to oppose the change.

The utility of Lenz's law should not be underestimated. In many cases, it represents the quickest way of obtaining this information about electromagnetic reactions. Even if other methods are available, it affords a valuable check.

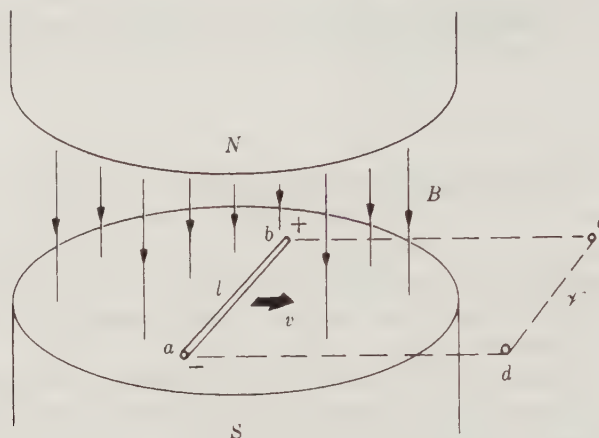
In order to get some insight into Faraday's law, it may be useful to consider an example that is usually regarded as an instance of the law but that *can* be analyzed completely in terms of the forces acting on charges, as developed in the preceding chapters. Suppose a straight metal wire of length l moves in a direction perpendicular to its length, with velocity \mathbf{v} . Let there be a magnetic field \mathbf{B} perpendicular to the plane in which the wire moves, as shown in Fig. 11-1. The free charges in the wire will experience the Lorentz force:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (8-5) \quad (11-7)$$

This force drives positive and negative charges to opposite ends of the wire because of the $q\mathbf{v} \times \mathbf{B}$ term. In the steady state, when the free charges are not moving with respect to the wire, the total force on a charge must equal zero—that is, the magnetic force must be balanced at each point in the wire by an equal and opposite electric force due to the charge separation,

$$E = vB \quad (11-8)$$

FIGURE 11-1
Voltage produced by a
wire moving in a mag-
netic field



If the B -field is uniform, then E is constant along the wire, and the potential difference between the ends is

$$\Delta\varphi = -\int_a^b \mathbf{E} \cdot d\mathbf{l} = El \quad (11-9)$$

If we call this potential difference \mathcal{V} , then combining Eqs. (11-8) and (11-9) gives

$$\mathcal{V} = Blv \quad (11-10)$$

In this example, the B -field is independent of time, and so $\nabla \times \mathbf{E} = 0$ and $\oint \mathbf{E} \cdot d\mathbf{l} = 0$, as in electrostatics. The integral $\int \mathbf{E} \cdot d\mathbf{l}$ is independent of path; in particular, if we imagine a circuit $abcda$ extending outside of the magnetic field, \mathcal{V} is also the potential difference along the path $bcda$. In fact, if b and c , and similarly d and a , are connected by perfectly conducting wires, \mathcal{V} will be the voltage between the terminals c and d outside the magnetic field.

The righthand side of Eq. (11-10) can be expressed in another way by noting that the flux Φ through the circuit $abcda$ is changing according to $d\Phi/dt = B dA/dt = Bl dx/dt = -Blv$. Thus,

$$\mathcal{V} = -\frac{d\Phi}{dt} \quad (11-11)$$

This equation has just the form of Faraday's law, Eq. (11-2), except that \mathcal{V} is *not* an emf in the sense defined by (11-1) because $\oint \mathbf{E} \cdot d\mathbf{l} = 0$ around every closed path in this problem. Equation (11-10) can be generalized by writing it in vector notation. If \mathbf{v} is arbitrarily oriented with respect to \mathbf{l} , then only the component of \mathbf{v} that is perpendicular to \mathbf{l} contributes to \mathcal{V} . Thus, \mathcal{V} is proportional to $\mathbf{l} \times \mathbf{v}$. For arbitrary \mathbf{B} , only the component perpendicular to the plane of \mathbf{l} and \mathbf{v} contributes to \mathcal{V} . Since $\mathbf{l} \times \mathbf{v}$ is

perpendicular to the \mathbf{l}, \mathbf{v} -plane, \mathcal{V} may be written as

$$\mathcal{V} = \mathbf{B} \cdot \mathbf{l} \times \mathbf{v} \quad (11-12)$$

The voltage in Eq. (11-12) is sometimes called *motional emf*.

Let us now look at the same problem from the standpoint of the wire—that is, let us imagine that we are in a coordinate system moving with the wire, so that in this system the wire is at rest and the magnet is moving toward the left in Fig. 11-1 with velocity v . One can readily believe that, moving with the wire, one would still observe the same charge separation and the same potential difference between the ends as before. The *explanation* is completely different, however. In this coordinate system, there can be no magnetic force since the wire is at rest. On the other hand, the magnetic field is no longer constant in time—at any point it changes from B to approximately zero as the edge of the moving magnet passes by that point. We shall see that the modified curl E , Eq. (11-6), is just sufficient to give the same result \mathcal{V} for the potential difference in this coordinate system also. In the steady state, the force on a free charge inside the wire must still be zero.

$$\mathbf{F} = q\mathbf{E} = 0$$

but there is no magnetic force since $\mathbf{v} = 0$. Thus, the electric force must vanish inside the wire,

$$\mathbf{E} = 0 = \mathbf{E}_1 + \mathbf{E}_2 \quad (11-13)$$

There is still a field \mathbf{E}_1 due to the charge separation, which is the same as in the previous case. This field is canceled inside the wire by a field \mathbf{E}_2 associated with the changing magnetic field,

$$\nabla \times \mathbf{E}_2 = -\frac{\partial \mathbf{B}}{\partial t}$$

If we again consider the closed curve $abcda$,

$$\mathcal{E} = \oint \mathbf{E} \cdot d\mathbf{l} = \int_a^b \mathbf{E} \cdot d\mathbf{l} + \int_b^a \mathbf{E} \cdot d\mathbf{l} = 0 + \mathcal{V}$$

The first term on the right is zero because \mathbf{E} vanishes inside the wire, and the second integral along the path $bcda$ is what we called \mathcal{V} in the previous case. Thus from this and Eq. (11-2), we again find that

$$\mathcal{V} = -\frac{d\Phi}{dt} \quad (11-11)$$

In conjunction with the Lorentz force, Eq. (11-7), the generalized curl E , Eq. (11-6), gives the same result, Eq. (11-11), in *either* inertial coordinate

system. Equation (11-6) is therefore generally valid.* Since the integral result Eq. (11-11) holds in both coordinate systems, it is not illogical to refer to it as Faraday's law in both cases, even though strictly speaking there is an emf, as defined by Eq. (11-1), only in the second case. In some situations, it may not be immediately obvious what circuit should be used to calculate Φ in Eq. (11-11), as for example in Problem 11-4. The equations that always apply to the E - and B -fields in any inertial coordinate system are Eqs. (11-6) and (11-7). No ambiguity arises in determining the emf or "motional emf" using them.

This example holds some further interest as a prototype of practical electrical generators. Refer again to Fig. 11-1. If the wire segment ab slides along two highly conducting wires, bc and ad (making excellent electrical contact to these wires), and if a resistance is connected between terminals c and d , a current I will flow through the circuit.† In this case, a mechanical force applied to the wire (or to the magnet in the second case) would be required in order to maintain a constant velocity v , so that the sum of the applied force and the magnetic force BIl on the wire would be zero. The power input of the applied mechanical force compensates for the I^2R power dissipation in the resistor. As far as the terminal voltage between c and d is concerned, it makes no difference whether the wire moves or the magnet moves in the generator (commonly the wire moves). In either case, $\oint \mathbf{E} \cdot d\mathbf{l} = 0$ around any path that does not encircle the magnetic field of the generator.

In our two examples, the generalized equation

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (11-6)$$

applies to both. It happened that in the magnet's coordinate system $\partial \mathbf{B} / \partial t = 0$, and so an electrostatic analysis was possible. It would be wrong to conclude, however, that one can always find a coordinate system where $\partial \mathbf{B} / \partial t$ vanishes. A third example based on Fig. 11-1 will illustrate this point. Suppose that *neither* the wire nor the magnet is moving, but suppose the magnet is an electromagnet, whose field can be increased or decreased in magnitude by increasing or decreasing the current in its windings. Now there is no coordinate system in which $\partial \mathbf{B} / \partial t$ vanishes. Still Eq. (11-6) applies, however, and Faraday's law, Eq. (11-2), gives the emf around any

* From another point of view, we shall find in Chapter 22 that the field $E_2 = vB$, which satisfies Eqs. (11-8) and (11-13), arises in the "moving" coordinate system from the relativistic Lorentz transformation of the E - and B -fields. The field E_2 replaces the magnetic force that vanishes in the "moving" system.

† This device could not be a practical direct-current generator because of the finite extent of the magnet; but if the wire moves back and forth, an alternating current is generated. (See Problem 11-5.)

circuit (e.g., *abcd*). This situation occurs in transformers and other practical devices with no mechanically moving parts and is the subject of the rest of this chapter.

11-2 SELF-INDUCTANCE

In this section, the relationship between the flux and current associated with an isolated circuit is considered and exploited for the purpose of introducing the practical circuit parameter: self-inductance. The magnetic flux linking an isolated circuit (flux produced by the current in the circuit itself) depends on the geometry of the circuit and, according to Eq. (8-26), is linearly dependent on the current in the circuit. Thus, for a rigid stationary circuit, the only changes in flux result from changes in the current. That is,

$$\frac{d\Phi}{dt} = \frac{d\Phi}{dI} \frac{dI}{dt} \quad (11-14)$$

This result is valid even when Eq. (8-26) is not. The only requirement is that Φ depend only on the current. If, however, Eq. (8-26) is valid or, more generally, if Φ is directly proportional to the current, then $d\Phi/dI$ is a constant, equal to Φ/I .

In any case, the *self-inductance*, L , is defined as

$$L = \frac{d\Phi}{dI} \quad (11-15)$$

When it is essential to distinguish between Φ/I and $d\Phi/dI$, the latter is called the *incremental* inductance. Unless otherwise specified, it is safest to associate the word inductance with Eq. (11-15). From Eqs. (11-14), (11-15), and (11-2), it follows that the *induced emf* is given by

$$\mathcal{E} = -L \frac{dI}{dt} \quad (11-16)$$

This equation is of considerable practical importance.

As an illustration of the use of Eq. (11-15) for the calculation of inductance, the self-inductance of a toroidal coil will be calculated.

EXAMPLE 11-1 Self-Inductance of a Toroidal Coil

Refer to the coil (with N turns) shown in Fig. 11-2. We want to calculate its self-inductance. Equation (11-15) applies to an entire circuit—that is, not only to the toroidal coil of Fig. 11-2, but also to the external circuit connected to terminals 1 and 2. By using twisted leads or a coaxial cable, which produce essentially no external magnetic field,

FIGURE 11-2

A toroidal winding



the field-producing portion of the external circuit can be removed to a sufficiently great distance that it does not contribute to the flux in the toroid.

Solution: If the toroid is thus isolated and if by emf we understand the voltage between terminals 1 and 2, then Eq. (11-15) can be used to obtain the inductance of the toroidal coil. From Ampere's circuital law, the magnetic induction inside the toroidal coil is

$$B = \frac{\mu_0 N I}{l} \quad (11-17)$$

where N is the number of turns, l the mean length, and I the current in the winding. (Equations (11-17) and (11-18) involve the approximation of neglecting the variation of the magnetic induction over the cross-sectional area. In Problem 11-10 the details of this approximation are considered.) The flux linking each turn is then

$$\Phi_1 = \frac{\mu_0 N I A}{l} \quad (11-18)$$

and the total flux linking the N turns is

$$\Phi = \frac{\mu_0 N^2 A}{l} I \quad (11-19)$$

The inductance is then simply

$$L = \frac{d\Phi}{dI} = \frac{\mu_0 N^2 A}{l} \quad (11-20)$$

The mks unit of inductance is the henry (H), which, from Eq. (11-16), is equal to one volt second/ampere since the unit of emf is the volt. Equation (11-20) indicates that the dimensions of μ_0 , which have been previously given as webers/ampere-meter or tesla-meters/ampere, can alternatively be given as henries/meter.

11-3 MUTUAL INDUCTANCE

In the preceding section, only isolated circuits were considered, so that all of the flux linking the circuit was due to the current in the circuit itself. This restriction can be lifted by assuming that there are n circuits, labeled 1, 2, The flux linking one of these circuits, say the one labeled i , can be written as

$$\Phi_i = \Phi_{i1} + \Phi_{i2} + \cdots + \Phi_{ii} + \cdots + \Phi_{in} = \sum_{j=1}^n \Phi_{ij} \quad (11-21)$$

That is, it may be written as a sum of fluxes due to each of the n circuits, Φ_{i1} being the flux through the i th circuit due to the circuit 1, and so on. The emf induced in the i th circuit, \mathcal{E}_i , can then be written as

$$\mathcal{E}_i = -\frac{d\Phi_i}{dt} = -\left\{ \frac{d\Phi_{i1}}{dt} + \cdots + \frac{d\Phi_{ii}}{dt} + \cdots + \frac{d\Phi_{in}}{dt} \right\} = -\sum_{j=1}^n \frac{d\Phi_{ij}}{dt} \quad (11-22)$$

If each of the circuits is a rigid stationary circuit, the only changes in the Φ_{ij} 's are those that result from changes in the currents. Thus,

$$\frac{d\Phi_{ij}}{dt} = \frac{d\Phi_{ij}}{dI_j} \frac{dI_j}{dt} \quad (11-23)$$

The coefficients $d\Phi_{ij}/dI_j$ are constants, independent of the current, if the magnetic media in the problem are linear.

In either case, linear or nonlinear,

$$M_{ij} = \frac{d\Phi_{ij}}{dI_j}, \quad i \neq j \quad (11-24)$$

is defined as the *mutual inductance* between circuit i and circuit j .

It will be seen later that $M_{ij} = M_{ji}$, and hence, there is no possibility of ambiguity in the subscripts. Of course $d\Phi_{ii}/dI_i$ is just the self-inductance of the i th circuit, for which L_i or M_{ii} is written. The units of mutual inductance are the same as those of self-inductance, namely, henries.

EXAMPLE 11-2 Mutual Inductance of Two Toroidal Windings

Consider the configuration of Fig. 11-2 with a winding of N_1 turns and a second toroidal winding of N_2 turns added. For this situation, a current I_1 in the first winding produces a magnetic induction

$$B = \frac{\mu_0 N_1 I_1}{l}$$

What is the mutual inductance between the two circuits?

Solution: The fluxes linking the two circuits are

$$\Phi_{11} = \frac{\mu_0 N_1^2 A I_1}{l} \quad \text{and} \quad \Phi_{21} = \frac{\mu_0 N_1 N_2 A I_1}{l}$$

From these fluxes, it follows that

$$L_1 = \frac{\mu_0 N_1^2 A}{l} \quad (11-25)$$

as we found in the preceding section, and using Eq. (11-24),

$$M_{21} = \frac{\mu_0 N_1 N_2 A}{l} \quad (11-26)$$

Reversing the procedure and considering a current I_2 gives

$$L_2 = \frac{\mu_0 N_2^2 A}{l} \quad (11-27)$$

and

$$M_{12} = \frac{\mu_0 N_1 N_2 A}{l} \quad (11-28)$$

thus demonstrating that for this case $M_{12} = M_{21}$. This result is a general one which we discuss further in Section 12-1. Furthermore, Eqs. (11-25), (11-26), and (11-27) may be combined to yield

$$M_{12} = \sqrt{L_1 L_2} \quad (11-29)$$

Equation (11-29) represents a limit that is imposed on the mutual inductance between two circuits; namely, it is always less than or equal to the square root of the product of the self-inductances of the two circuits. In view of this limit, a coupling coefficient k is often introduced and defined by

$$M = k\sqrt{L_1 L_2}, \quad 0 \leq k \leq 1 \quad (11-30)$$

A value of $k = 1$ implies that all of the magnetic flux produced by circuit 1 links circuit 2 and vice-versa.

11-4 THE NEUMANN FORMULA

For two rigid stationary circuits in a linear medium (vacuum for the present), the mutual inductance is just

$$M_{21} = \frac{\Phi_{21}}{I_1} \quad (11-31)$$

This equation is valid simply because Φ_{21} is proportional to I_1 , making Φ_{21}/I_1

and $d\Phi_{21}/dI_1$ equal. In this case, Eq. (8-26) can be used to calculate M_{21} . The flux is given by

$$\Phi_{21} = \frac{\mu_0}{4\pi} I_1 \int_{S_2} \left\{ \oint_{C_1} \frac{d\mathbf{l}_1 \times (\mathbf{r}_2 - \mathbf{r}_1)}{|\mathbf{r}_2 - \mathbf{r}_1|^3} \right\} \cdot \mathbf{n} da_2 \quad (11-32)$$

However,

$$\oint_{C_1} \frac{d\mathbf{l}_1 \times (\mathbf{r}_2 \times \mathbf{r}_1)}{|\mathbf{r}_2 - \mathbf{r}_1|^3} = \nabla_2 \times \oint_{C_1} \frac{d\mathbf{l}_1}{|\mathbf{r}_2 - \mathbf{r}_1|} \quad (11-33)$$

Hence,

$$M_{21} = \frac{\Phi_{21}}{I_1} = \frac{\mu_0}{4\pi} \int_{S_2} \nabla_2 \times \left\{ \oint_{C_1} \frac{d\mathbf{l}_1}{|\mathbf{r}_2 - \mathbf{r}_1|} \right\} \cdot \mathbf{n} da_2 \quad (11-34)$$

Using Stokes's theorem to transform the surface integral gives

$$M_{21} = \frac{\mu_0}{4\pi} \oint_{C_2} \oint_{C_1} \frac{d\mathbf{l}_1 \cdot d\mathbf{l}_2}{|\mathbf{r}_2 - \mathbf{r}_1|} \quad (11-35)$$

which is known as *Neumann's formula* for the mutual inductance. The symmetry alluded to earlier is apparent in Eq. (11-35).

Neumann's formula is equally applicable to self-inductance, in which case it is written as

$$L = \frac{\mu_0}{4\pi} \oint_{C_1} \oint_{C_1} \frac{d\mathbf{l}_1 \cdot d\mathbf{l}'_1}{|\mathbf{r}_1 - \mathbf{r}'_1|} \quad (11-36)$$

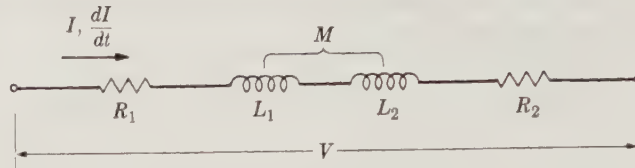
Some care must be used in the application of Eq. (11-36) because of the singularity at $\mathbf{r}_1 = \mathbf{r}'_1$; however, if care is taken, Eq. (11-36) is sometimes useful.

Equations (11-35) and (11-36) are usually difficult to apply to the calculation of inductance except for circuits in which the geometry is simple. But Eq. (11-35) in particular is very important in the study of forces and torques exerted by one circuit on another. This application will be exploited in Chapter 12.

INDUCTANCES IN SERIES AND IN PARALLEL

Inductances are often connected in series and in parallel, and as with resistors and capacitors, it is important to know the result of such connections. We could proceed with a derivation based simply on $\mathcal{E} = -L(dI/dt)$ and obtain formulas for the effective inductance of two inductances in series or in parallel. However, to do so would be to ignore the practical fact that an inductor always has a certain internal resistance. A perfect inductance is much more difficult to approximate practically than a perfect capacitance or a perfect resistance. For this reason, the series and

FIGURE 11-3
Series connection of two
inductors



parallel combinations of this section will always involve resistances as well as inductances.

For two inductors in series, the circuit of Fig. 11-3 is appropriate. In adding the voltage drops along the circuit, it is important to note that M can be either positive or negative (changing the direction in which either circuit C_1 or C_2 is traversed reverses the sign of M in Eq. (11-35)). Bearing this in mind, the sum of the voltage drops for the circuit of Fig. 11-3 is found to be

$$V + \mathcal{E}_1 + \mathcal{E}_2 = R_1 I + R_2 I$$

or

$$V = R_1 I + L_1 \frac{dI}{dt} + M \frac{dI}{dt} + R_2 I + L_2 \frac{dI}{dt} + M \frac{dI}{dt} \quad (11-37)$$

This result is equivalent to

$$V = (R_1 + R_2)I + (L_1 + L_2 + 2M) \frac{dI}{dt} \quad (11-38)$$

The circuit thus resembles a resistor of resistance $R_1 + R_2$ in series with an inductance $L_1 + L_2 + 2M$. The magnitude of the inductance is $L_1 + L_2 + 2|M|$ for positive coupling (i.e., for fluxes due to I_1 and I_2 in the same direction in each coil) and is $L_1 + L_2 - 2|M|$ for negative coupling. An alternative description of the mutual inductance is

$$M = k\sqrt{L_1 L_2}, \quad -1 \leq k \leq 1 \quad (11-39)$$

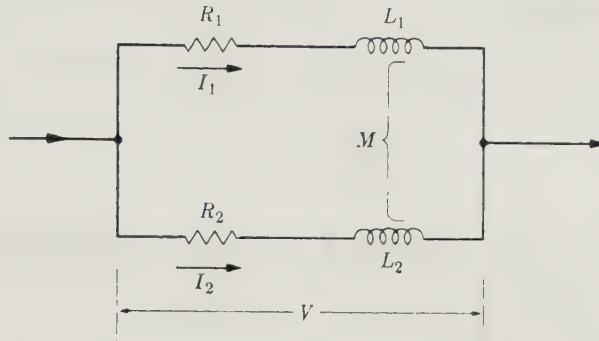
The effective inductance of the series circuit is then

$$L_{\text{eff}} = L_1 + 2k\sqrt{L_1 L_2} + L_2 \quad (11-40)$$

If k can be varied, then a variable inductance can be constructed. (In the early days of radio, this was a popular way of tuning resonant circuits; see Chapter 13.)

The parallel connection shown in Fig. 11-4 is not as simple as the series circuit. In fact, the circuit shown does not behave like a simple series L - R circuit. Thus, it is not possible to say that the effective inductance and effective resistance are certain functions of L_1 , L_2 , R_1 , and R_2 . If, however,

FIGURE 11-4
Parallel connection of
two inductors



R_1 and R_2 are negligible, then

$$V = L_1 \frac{dI_1}{dt} + M \frac{dI_2}{dt} \quad (11-41)$$

$$V = L_2 \frac{dI_2}{dt} + M \frac{dI_1}{dt}$$

If first dI_1/dt and then dI_2/dt are eliminated from between Eqs. (11-41), there results

$$V(L_2 - M) = (L_1 L_2 - M^2) \frac{dI_1}{dt} \quad (11-42)$$

$$V(L_1 - M) = (L_1 L_2 - M^2) \frac{dI_2}{dt}$$

Adding these gives

$$V = \frac{L_1 L_2 - M^2}{L_1 + L_2 - 2M} \frac{dI}{dt} \quad (11-43)$$

Thus the effective inductance of two inductors in parallel is

$$L_{\text{eff}} = \frac{L_1 L_2 - M^2}{L_1 + L_2 - 2M} \quad (11-44)$$

where again the sign of M depends on the way in which the inductors are connected.

The most important use of inductances is in alternating current circuits. For a circuit operating at a single frequency, an equivalent series circuit for Fig. 11-4 can be obtained; however, both the equivalent resistance and equivalent inductance are frequency dependent.

11-6 SUMMARY

In this chapter, we go a step beyond static fields to the so-called slowly varying case. The new generalization of the field equations is

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

This equation is the **third of Maxwell's four equations**, which always holds, along with the two divergence equations and the **Lorentz force**,

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

(At this point in our development of the fundamental equations of electricity and magnetism, we have three of the four Maxwell equations in final form. Only the curl H equation still needs to be generalized.) The differential form of **Faraday's law** (given above) can be cast in integral form,

$$\mathcal{E} = -\frac{d\Phi}{dt}$$

where the **emf** \mathcal{E} around a fixed circuit C is defined by

$$\mathcal{E} = \oint_C \mathbf{E} \cdot d\mathbf{l}$$

(It may happen in certain problems that a moving coordinate system can be found in which $\partial \mathbf{B} / \partial t = 0$, and the problem can be analyzed electrostatically, but such is not necessarily the case.)

- The easiest way to determine the correct polarity of an induced voltage is by means of **Lenz's law**.

- The "**motional emf**" of a straight wire moving in a magnetic field is

$$\mathcal{V} = \mathbf{B} \cdot \mathbf{l} \times \mathbf{v}$$

- The **self-inductance** of a fixed circuit (or circuit element) is defined as

$$L = \frac{d\Phi}{dI}$$

so that

$$\mathcal{E} = -L \frac{dI}{dt}$$

For a toroid (or long solenoid) L is easily found to be

$$L = \frac{\mu_0 N^2 A}{l}$$

- The **mutual inductance** of two circuits is defined as

$$M_{ij} = \frac{d\Phi_{ij}}{dI_j}$$

It follows that

$$M_{ii} = L_i$$

and

$$M_{12} = M_{21} = k\sqrt{L_1L_2}, \quad -1 \leq k \leq 1$$

- Pure inductances connected in series or parallel add according to the same formulas as resistances, assuming their mutual inductance and inherent resistance can be neglected.

PROBLEMS

11-1 A metallic conductor in the shape of a wire segment of length l is moved in a magnetic field \mathbf{B} with velocity \mathbf{v} . From a detailed consideration of the Lorentz force on the electrons in the wire, show that the ends of the wire are at the potential difference: $\mathbf{B} \cdot \mathbf{l} \times \mathbf{v}$.

11-2 A metal rod one meter long rotates about an axis through one end and perpendicular to the rod, with an angular velocity of 12 rad/s. The plane of rotation of the rod is perpendicular to a uniform magnetic field of 0.3 T. What is the motional emf induced between the ends of the rod?

11-3 In a betatron accelerator, an ion of charge q and mass m traverses a circular orbit at a distance R from the symmetry axis of the machine. The magnetic field has cylindrical symmetry—that is, the z -component is $B_z = B(r)$ in the plane of the orbit, where r is the distance from the symmetry axis. (a) Show that the ion's velocity is $v = qB(R)R/m$. (b) If the magnetic field is slowly increased in magnitude, show that the emf induced around the ion's orbit is such as to accelerate the ion. (c) Show that in order for the ion to stay in its same orbit, the radial variation of the B -field inside the orbit must satisfy the following condition: The spatial average of the increase in $B(r)$ (averaged over the area enclosed by the orbit) must equal twice the increase in $B(R)$ during the same time interval.

11-4 Faraday's homopolar generator consists of a metal disk that rotates in a uniform magnetic field perpendicular to the plane of the disk. Show that the potential difference produced between the center of the disk and its periphery is $\mathcal{V} = f\Phi$, where Φ is the flux through the disk and f is its frequency of rotation. What is the voltage if $f = 3000$ r/min and $\Phi = 0.1$ Wb?

11-5 An alternator consists of an N -turn coil of area A , which rotates in a field B about a diameter perpendicular to the field, with a frequency of rotation f . (a) Find the emf in the coil. (b) What is the amplitude of the alternating voltage if $N = 100$ turns, $A = 10^{-2}$ m², $B = 0.1$ T, and $f = 2000$ r/min?

11-6 A dielectric cylinder of permittivity ϵ rotates about its axis with angular velocity ω . If a uniform magnetic field \mathbf{B} exists parallel to the cylinder axis, find the induced polarization charge in the dielectric.

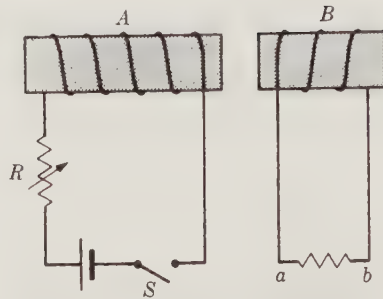


FIGURE 11-5

11-7 Two coupled circuits, A and B , are situated as shown in Fig. 11-5. Use Lenz's law to determine the direction of the induced current in resistor ab when (a) coil B is brought closer to coil A , (b) the resistance of R is decreased, (c) switch S is opened.

11-8 A 100-turn coil of circular cross section is wound compactly so that all loops lie in approximately the same plane. The average radius of the coil is 3 cm. The coil rotates about one of its diameters at 900 r/min. When the rotation axis is vertical, the root-mean-square induced motional emf in the coil is found to be 0.50 mV. What can be concluded about the earth's magnetic field at the location of the coil?

11-9 A circular disk rotates about its axis with angular velocity ω . The disk is made of metal with conductivity g , and its thickness is t . The rotating disk is placed between the pole faces of a magnet which produces a uniform magnetic field B over a small square area of size a^2 at the average distance r from the axis; B is perpendicular to the disk. Calculate the approximate torque on the disk. (Make a reasonable assumption about the resistance of the "eddy current circuit.")

11-10 A toroidal coil of N turns, such as is shown in Fig. 11-2, is wound on a nonmagnetic form. If the mean radius of the coil is b and the cross-sectional radius of the form is a , show that the self-inductance of the coil is given by $L = \mu_0 N^2 (b - \sqrt{b^2 - a^2})$.

11-11 A circuit consists of two coaxial cylindrical shells of radii R_1 and R_2 ($R_2 > R_1$) and common length L , connected by flat end plates. The charge flows down one shell and back up the other. What is the self-inductance of this circuit?

11-12 The toroidal coil of Problem 11-10 has 150 turns, $b = 4$ cm, and $a = 1.5$ cm. What is the self-inductance of the coil, in henries?

11-13 Two small circular loops of wire (of radii a and b) lie in the same plane at distance r apart. What is the mutual inductance between the loops if the distance r is sufficiently large that the dipole approximation may be used?

11-14 Two circular current loops with *parallel* axes are located at a distance r from each other that is sufficiently large so that the dipole approximation may be used. Show how one of the loops should be placed relative to the other so that their mutual inductance is zero.

11-15 Consider two circuits: a very long straight wire, and a rectangle of dimensions h and d . The rectangle lies in a plane through the wire, the sides of length h being parallel to the wire and at distances r and $r + d$ from it. Calculate the mutual inductance between the two circuits.

11-16 Consider two circuits: a very long straight wire, and a circle of radius a . The circle lies in a plane through the wire, with its center at a distance r . Calculate the mutual inductance between the two circuits.

11-17 A transmission line consists of two very long wires of radius a separated by a distance d . Calculate the self-inductance per unit length, assuming $d \gg a$, so that the flux inside the wires themselves can be ignored.

11-18 Consider two coaxial, circular loops of wire of radii a and b , separated by the axial distance x . Through the use of Neumann's formula, show that the mutual inductance of the loops is

$$M = \mu_0(ab)^{1/2} \left[\left(\frac{2}{k} - k \right) K(k) - \frac{2}{k} E(k) \right]$$

where

$$k^2 = \frac{4ab}{(a+b)^2 + x^2}$$

and $K(k)$ and $E(k)$ are complete elliptic integrals defined by

$$K(k) = \int_0^{\pi/2} \frac{d\phi}{(1 - k^2 \sin^2 \phi)^{1/2}}$$

and

$$E(k) = \int_0^{\pi/2} (1 - k^2 \sin^2 \phi)^{1/2} d\phi$$

11-19 Consider again the preceding problem. By expanding $1/|\mathbf{r}_2 - \mathbf{r}_1|$ in Neumann's formula according to the binomial theorem, integrate term by term to obtain

$$M = \frac{\mu_0 \pi a^2 b^2}{2h^3} \left(1 + 3 \frac{ab}{h^2} + \frac{75 a^2 b^2}{8 h^4} + \dots \right)$$

where $h^2 = x^2 + (a + b)^2$.

11-20 Two circuits with inductances L_1 and L_2 and resistances R_1 and R_2 are located near each other. If the mutual inductance between the circuits is M , show that a quantity of charge $Q = \mathcal{V}M/R_1 R_2$ will circulate through one of them if an applied voltage \mathcal{V} is suddenly connected in series with the other.

11-21 (a) Use your result from Problem 11-9 to calculate the torque on an aluminum disk rotating at 60 rad/s if it is subjected to a magnetic field of 0.3 T over an area of 4 cm² at a radius of 0.3 m. The thickness of the disk is 2 cm. (b) If the moment of inertia of the disk is 2.0 kg m², find the time required to decrease the angular speed to 30 rad/s.

11-22 Two flat, circular coils are coaxial and lie in the same plane. The large coil of radius 10 cm has 20 turns. The small coil of radius 1 cm has 10 turns. (a) Calculate the mutual inductance of these coils. (b) How does M change when the smaller coil is rotated about one of its axes?

11-23 Consider a nonmagnetic conducting medium of conductivity g , which is

subjected to a time-dependent magnetic field $\mathbf{B}(\mathbf{r}, t)$. (a) Starting with the differential form of Faraday's law, Eq. (11-6), show that on the assumption of no accumulation of charge (i.e., $\nabla \cdot \mathbf{J} = 0$) the induced eddy-current density in the medium satisfies the differential equation $\nabla^2 \mathbf{J} = g\mu_0(\partial \mathbf{J} / \partial t)$. (b) Show that \mathbf{E} and \mathbf{B} satisfy the same equation.

11-24 Show that the emf in a fixed circuit C is given by

$$-\frac{d}{dt} \oint_C \mathbf{A} \cdot d\mathbf{l}$$

where \mathbf{A} is the vector potential.

11-25 A "force-free field" is one such that $\mathbf{J} \times \mathbf{B} = 0$. (a) Show that such a field satisfies the equation

$$\nabla^2 \mathbf{B} = -\alpha^2 \mathbf{B}$$

where α is a constant. [*Hint*: In the force-free field $\nabla \times \mathbf{B}$ is parallel to \mathbf{B} .] (b) Show that \mathbf{J} satisfies the same equation. (c) Use the result of Problem 11-23 to find the time dependence of the current density and the fields.

11-26 The E -field induced by $\dot{\mathbf{B}} = \partial \mathbf{B} / \partial t$ can be written explicitly as

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi} \int \frac{(\mathbf{r} - \mathbf{r}') \times \dot{\mathbf{B}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} dv'$$

(a) Verify that $\nabla \times \mathbf{E} = -\partial \mathbf{B} / \partial t$ and $\nabla \cdot \mathbf{E} = 0$, by differentiating inside the integral. (b) Show that the gradient of any solution of Laplace's equation can be added to \mathbf{E} .

Magnetic Energy

Establishing a magnetic field requires the expenditure of energy. This conclusion follows directly from Faraday's law of induction: If a source of voltage \mathcal{V} is applied to a circuit, then, in general, the current through the circuit can be expressed by the equation

$$\mathcal{V} + \mathcal{E} = IR \quad (12-1)$$

where \mathcal{E} is the induced emf and R is the resistance of the current circuit. The work done by \mathcal{V} in moving the charge increment $dq = I dt$ through the circuit is

$$\mathcal{V} dq = \mathcal{V} I dt = -\mathcal{E} I dt + I^2 R dt = I d\Phi + I^2 R dt \quad (12-2)$$

the last form of which is obtained with the aid of Faraday's law, Eq. (11-2). The term $I^2 R dt$ represents the irreversible conversion of electrical energy into heat by the circuit, but this term absorbs the entire work input *only* in cases where the flux change is zero. The additional term, $I d\Phi$, is the work done against the induced emf in the circuit. It is that part of the work done by \mathcal{V} which is effective in altering the magnetic field structure. Disregarding the $I^2 R dt$ term, we write

$$dW_b = I d\Phi \quad (12-3)$$

where the subscript b indicates that this is work performed by external electrical energy sources (e.g., by batteries). The work increment Eq. (12-3) may be either positive or negative. It is positive when the flux change $d\Phi$ through the circuit is in the same direction as the flux produced by the current I .

For a rigid stationary circuit showing no energy losses other than Joule heat loss (e.g., no hysteresis), the term dW_b is equal to the change in *magnetic energy* of the circuit. In this chapter, we develop a number of

equivalent expressions for the magnetic energy of a system of coupled circuits and, in addition, derive equations for the force or torque on a rigid circuit. Hysteresis loss will be discussed in Section 12-4, but for the present we shall restrict our attention to reversible magnetic systems. The development here closely parallels that of Chapter 6.*

12-1
MAGNETIC ENERGY OF COUPLED CIRCUITS

In this section, we shall derive an expression for the magnetic energy of a system of interacting current circuits. If there are n circuits, then, according to Eq. (12-3), the electrical work done against the induced emf's is given by

$$dW_b = \sum_{i=1}^n I_i d\Phi_i \quad (12-4)$$

This expression is perfectly general; it is valid independently of how the flux increments $d\Phi_i$ are produced. We are particularly interested, however, in the case where the $d\Phi_i$ are produced by current changes in the n circuits themselves. In these circumstances, the flux changes are directly correlated with changes in these currents:

$$d\Phi_i = \sum_{j=1}^n \frac{d\Phi_{ij}}{dI_j} dI_j = \sum_{j=1}^n M_{ij} dI_j \quad (12-5)$$

If the circuits are rigid† and stationary, then no mechanical work is associated with the flux changes $d\Phi_i$, and dW_b is just equal to the change in magnetic energy, dU , of the system. Note that here we restrict our attention to stationary circuits, so that the magnetic energy can be calculated as a work term. Later we shall let the various circuits move relative to one another, but then we will not be able to identify dU with dW_b .

The magnetic energy U of a system of n rigid stationary circuits is obtained by integrating Eq. (12-4) from the zero flux situation (corresponding to all $I_i = 0$) to the final set of flux values. For a group of *rigid circuits* containing, or located in, *linear magnetic media*, the Φ_i are linearly related to the currents in the circuits, and the magnetic energy is independent of the way in which these currents are brought to their final set of values. Since

* A word of caution: Implicit in the derivations of this chapter is the assumption that the electrical circuits do *not radiate* electromagnetic energy. In other words, the currents in the circuits must be "slowly varying" (for a discussion of what is meant by this terminology, see Section 13-1). If one wants to treat the general case, with no restrictions on the currents or the circuit size, one must use energy relations derived from Maxwell's equations (see Section 16-3).

† By rigid circuit, we mean a circuit whose shape is fixed, but which can be moved around in its entirety.

this situation is of considerable importance, let us restrict our attention to the rigid-circuit, linear case.

Because the final energy is independent of the order in which the currents are varied, we may choose a particular scheme for which W is easily calculated. This scheme is one in which all currents (and hence all fluxes) are brought to their final values in concert—that is, at any instant of time all currents (and all fluxes) will be at the same fraction of their final values. Let us call this fraction α . If the final values of the current are given the symbols

$$I_1, I_2, \dots, I_n$$

then at any stage $I'_i = \alpha I_i$. Furthermore, $d\Phi_i = \Phi_i d\alpha$. Integration of Eq. (12-4) yields

$$\int dW_b = \int_0^1 d\alpha \sum_{i=1}^n I'_i \Phi_i = \sum_{i=1}^n I_i \Phi_i \int_0^1 \alpha d\alpha = \frac{1}{2} \sum_{i=1}^n I_i \Phi_i$$

Thus, the *magnetic energy* of n coupled circuits is

$$U = \frac{1}{2} \sum_{i=1}^n I_i \Phi_i \quad (\text{rigid circuits, linear media}) \quad (12-6)$$

With the aid of Eq. (12-5), which for a rigid-circuit, linear system may be integrated directly, the magnetic energy may be expressed in the following form:

$$\begin{aligned} U &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n M_{ij} I_i I_j \\ &= \frac{1}{2} L_1 I_1^2 + \frac{1}{2} L_2 I_2^2 + \dots + \frac{1}{2} L_n I_n^2 \\ &\quad + M_{12} I_1 I_2 + M_{13} I_1 I_3 + \dots + M_{1n} I_1 I_n \\ &\quad + M_{23} I_2 I_3 + \dots + M_{n-1,n} I_{n-1} I_n \end{aligned} \quad (12-7)$$

(rigid circuits, linear media)

Here we have used the results and notation of Sections 11-3 and 11-4: $M_{ij} = M_{ji}$; $M_{ii} = L_i$.

For *two* coupled circuits, the last equation reduces to

$$U = \frac{1}{2} L_1 I_1^2 + M I_1 I_2 + \frac{1}{2} L_2 I_2^2 \quad (12-8)$$

where, for simplicity, we have written M for M_{12} . The term $M I_1 I_2$ may be either positive or negative, but the total magnetic energy U must be positive (or zero) for any pair of current values: I_1 and I_2 . Denoting the current

ratio I_1/I_2 by x , we obtain

$$U = \frac{1}{2}I_2^2(L_1x^2 + 2Mx + L_2) \geq 0$$

The value of x that makes U a minimum (or maximum) is found by differentiating U with respect to x and setting the result equal to zero:

$$x = -\frac{M}{L_1} \quad (12-9)$$

The second derivative of U with respect to x is positive, which shows that Eq. (12-9) is the condition for a minimum. The magnetic energy $U \geq 0$ for any x ; in particular, the minimum value of U (defined by $x = -M/L_1$) is greater than or equal to zero. Thus,

$$\frac{M^2}{L_1} - \frac{2M^2}{L_1} + L_2 \geq 0$$

or

$$L_1L_2 \geq M^2 \quad (12-10)$$

a result that was stated, but not proved, in Section 11-3.

For a single circuit,

$$\Phi = LI$$

$$U = \frac{1}{2}I\Phi = \frac{1}{2}LI^2 = \frac{1}{2}\frac{\Phi^2}{L} \quad (12-11)$$

12-2 ENERGY DENSITY IN THE MAGNETIC FIELD

Equation (12-7) gives the magnetic energy of a current system in terms of circuit parameters: currents and inductances. Such a formulation is particularly useful because these parameters are capable of direct experimental measurement. On the other hand, an alternative formulation of the magnetic energy in terms of the field vectors \mathbf{B} and \mathbf{H} is of considerable interest because it provides a picture in which energy is stored in the magnetic field itself. This picture can be extended, as is done in Chapter 16, to show how energy moves through the electromagnetic field in nonstationary processes.

Consider a group of rigid current-carrying circuits, none of which extends to infinity, immersed in a medium with linear magnetic properties. The energy of this system is given by Eq. (12-6). For the present discussion it is convenient to assume that each circuit consists of only a single loop; then the flux Φ_i may be expressed as

$$\Phi_i = \int_{S_i} \mathbf{B} \cdot \mathbf{n} \, da = \oint_{C_i} \mathbf{A} \cdot d\mathbf{l}_i \quad (12-12)$$

where \mathbf{A} is the local vector potential. Substitution of this result into Eq. (12-6) yields

$$U = \frac{1}{2} \sum_i \oint_{C_i} I_i \mathbf{A} \cdot d\mathbf{l}_i \quad (12-13a)$$

We should like to make Eq. (12-13a) somewhat more general. Suppose that we do not have current circuits defined by wires, but instead each "circuit" is a closed path in the medium (assumed to be conducting) that follows a line of current density. Equation (12-13a) may be made to approximate this situation very closely by choosing a large number of contiguous circuits (C_i), replacing $I_i d\mathbf{l}_i \rightarrow \mathbf{J} dv$, and, finally, by the substitution of

$$\int_V \quad \text{for} \quad \sum_i \oint_{C_i}$$

Hence,

$$U = \frac{1}{2} \int_V \mathbf{J} \cdot \mathbf{A} dv \quad (12-13b)$$

The last equation may be further transformed by using the field equation $\nabla \times \mathbf{H} = \mathbf{J}$, and the vector identity (1-1-8):

$$\nabla \cdot (\mathbf{A} \times \mathbf{H}) = \mathbf{H} \cdot \nabla \times \mathbf{A} - \mathbf{A} \cdot \nabla \times \mathbf{H}$$

whence

$$U = \frac{1}{2} \int_V \mathbf{H} \cdot \nabla \times \mathbf{A} dv - \frac{1}{2} \int_S \mathbf{A} \times \mathbf{H} \cdot \mathbf{n} da \quad (12-14)$$

where S is the surface which bounds the volume V . Since, by assumption, none of the current "circuits" extend to infinity, it is convenient to move the surface S out to a very large distance so that all parts of this surface are far from the currents. The volume of the system must, of course, be increased accordingly. Now \mathbf{H} falls off at least as fast as $1/r^2$, where r is the distance from an origin near the middle of the current distribution to a characteristic point on the surface S ; \mathbf{A} falls off at least as fast as $1/r$; and the surface area is proportional to r^2 . Thus, the contribution from the surface integral in Eq. (12-14) falls off as $1/r$ or faster, and if S is moved out to infinity, this contribution vanishes.

By dropping the surface integral in Eq. (12-14) and extending the volume term to include all space, we obtain

$$U = \frac{1}{2} \int_V \mathbf{H} \cdot \mathbf{B} dv \quad (12-15)$$

since $\mathbf{B} = \nabla \times \mathbf{A}$. This result is completely analogous to the expression for electrostatic energy, Eq. (6-17). Equation (12-15) is restricted to systems containing linear magnetic media, since it was derived from Eq. (12-6).

By reasoning similar to that of Section 6-3, we are led to the concept of *energy density* in a magnetic field:

$$u = \frac{1}{2} \mathbf{H} \cdot \mathbf{B} \quad (12-16a)$$

For the case of isotropic, linear, magnetic materials, this equation reduces to

$$u = \frac{1}{2} \mu H^2 = \frac{1}{2} \frac{B^2}{\mu} \quad (12-16b)$$

12-3

FORCES AND TORQUES ON RIGID CIRCUITS

Up to this point, we have developed a number of alternative expressions for the magnetic energy of a system of current circuits. These are given by Eqs. (12-6) and (12-7), and (12-15). We will now see how the force, or torque, on one component of such a system may be calculated from a knowledge of the magnetic energy.

Suppose we allow one of the parts of the system to make a displacement $d\mathbf{r}$ under the influence of the magnetic forces acting upon it, all currents remaining constant. The mechanical work performed by the force \mathbf{F} acting on the system is

$$dW = \mathbf{F} \cdot d\mathbf{r} \quad (12-17)$$

as in Eq. (6-32). The work contains two contributions in these circumstances, as in Eq. (6-37):

$$dW = dW_b - dU \quad (12-18)$$

where dU is the change in magnetic energy of the system and dW_b is the work performed by external energy sources against the induced emf's to keep the currents constant.

Before we can proceed to an expression linking U and the force on a part of the system, it will be necessary to eliminate dW_b from Eq. (12-18), which is easily done for a system of rigid circuits in *linear* magnetic media. If the geometry of the system is changed by moving one or more parts as rigid units but all currents remain unaltered, then, according to Eq. (12-6),

$$dU = \frac{1}{2} \sum_i I_i d\Phi_i$$

But, from Eq. (12-4),

$$dW_b = \sum_i I_i d\Phi_i$$

Thus,

$$dW_b = 2 dU \quad (12-19)$$

Using this equation to eliminate dW_b from Eq. (12-18) and combining the result with Eq. (12-17), we obtain

$$dU = \mathbf{F} \cdot d\mathbf{r}$$

or

$$\mathbf{F} = \nabla U$$

$$F_x = \left(\frac{\partial U}{\partial x} \right)_I \quad (12-20)$$

The force on the circuit is the gradient of the magnetic energy when I is maintained constant.

If the circuit under consideration is constrained to move in such a way that it rotates about an axis, then Eq. (12-17) may be replaced by

$$dW = \boldsymbol{\tau} \cdot d\boldsymbol{\theta} = \tau_1 d\theta_1 + \tau_2 d\theta_2 + \tau_3 d\theta_3$$

where $\boldsymbol{\tau}$ is the magnetic torque on the circuit and $d\boldsymbol{\theta}$ is an angular displacement. Under these conditions,

$$\tau_1 = \left(\frac{\partial U}{\partial \theta_1} \right)_I \quad (12-21)$$

and so on. The results Eqs. (12-20) and (12-21) for constant current are analogous to the electrostatic case of constant potential, where work by the battery is needed to keep the potentials constant.

In some other interesting cases, the fluxes through the circuits can be treated as constant instead. Then, according to Eq. (12-4), $dW_b = 0$, and so the system can be taken as isolated.* Consequently,

$$\mathbf{F} \cdot d\mathbf{r} = dW = -dU$$

$$F_x = - \left(\frac{\partial U}{\partial x} \right)_\Phi \quad (12-22)$$

$$\tau_1 = - \left(\frac{\partial U}{\partial \theta_1} \right)_\Phi \quad (12-23)$$

Just as in the electrostatic case, in order to make use of the energy method it is necessary to express U in analytic form—that is, the specific dependence of U on the variable coordinates (x , y , z , θ_1 , θ_2 , or θ_3) must be given. When U can be expressed in this way, the energy method becomes a powerful technique for calculating forces and torques.

We shall illustrate the method by considering two examples. Additional

* We are disregarding the fact that in a normal circuit, a battery would still be needed to supply the I^2R power dissipation. If the wires were superconducting ($R = 0$), the system could in fact be isolated.

exercises of this type will be found in the problems at the end of the chapter.

EXAMPLE 12-1
Force Between Two
Current-Carrying Circuits

Let us calculate the force between two rigid circuits carrying constant currents, using the magnetic energy as given by Eq. (12-8).

Solution: The force on circuit 2 is

$$\mathbf{F}_2 = \nabla_2 U = I_1 I_2 \nabla_2 M,$$

where the mutual inductance M must be written so that it displays its dependence on \mathbf{r}_2 . Neumann's formula, Eq. (11-35), shows this dependence explicitly, so we may write

$$\begin{aligned} \mathbf{F}_2 &= \frac{\mu_0}{4\pi} I_1 I_2 \oint_{C_1} \oint_{C_2} (d\mathbf{l}_1 \cdot d\mathbf{l}_2) \nabla_2 \frac{1}{|\mathbf{r}_2 - \mathbf{r}_1|} \\ &= -\frac{\mu_0}{4\pi} I_1 I_2 \oint_{C_1} \oint_{C_2} (d\mathbf{l}_1 \cdot d\mathbf{l}_2) \frac{(\mathbf{r}_2 - \mathbf{r}_1)}{|\mathbf{r}_2 - \mathbf{r}_1|^3} \end{aligned} \quad (12-24)$$

an expression that evidently shows the proper symmetry—that is, $\mathbf{F}_2 = -\mathbf{F}_1$.

However, note that we already have an expression for the force between two circuits, Eq. (8-25), and it appears to be at variance with the formula just derived. Actually, the two expressions are equivalent, as may be easily verified. Let us expand the triple product in the integrand of Eq. (8-25):

$$d\mathbf{l}_2 \times [d\mathbf{l}_1 \times (\mathbf{r}_2 - \mathbf{r}_1)] = d\mathbf{l}_1 [d\mathbf{l}_2 \cdot (\mathbf{r}_2 - \mathbf{r}_1)] - (\mathbf{r}_2 - \mathbf{r}_1)(d\mathbf{l}_1 \cdot d\mathbf{l}_2)$$

The integral containing the last term on the right is identical with Eq. (12-24); that containing the first term may be written

$$\frac{\mu_0}{4\pi} I_1 I_2 \oint_{C_1} d\mathbf{l}_1 \oint_{C_2} \frac{d\mathbf{l}_2 \cdot (\mathbf{r}_2 - \mathbf{r}_1)}{|\mathbf{r}_2 - \mathbf{r}_1|^3} \quad (12-25)$$

Now $d\mathbf{l}_2 \cdot (\mathbf{r}_2 - \mathbf{r}_1)$ is $|\mathbf{r}_2 - \mathbf{r}_1|$ times the projection of $d\mathbf{l}_2$ on the vector $\mathbf{r}_2 - \mathbf{r}_1$. Let us denote $|\mathbf{r}_2 - \mathbf{r}_1|$ by r_{21} ; then the projection of $d\mathbf{l}_2$ is just dr_{21} . The integral over C_2 may be carried out at fixed $d\mathbf{l}_1$:

$$\oint_{C_2} \frac{dr_{21}}{r_{21}^2} = -\frac{1}{r_{21}} \Big|_a^a$$

the upper and lower limit being identical because of the complete circuit. Thus, the expression (12-25) vanishes, and Eq. (12-24) is equivalent to Eq. (8-25).

EXAMPLE 12-2
Force on Iron Rod
in a Solenoid

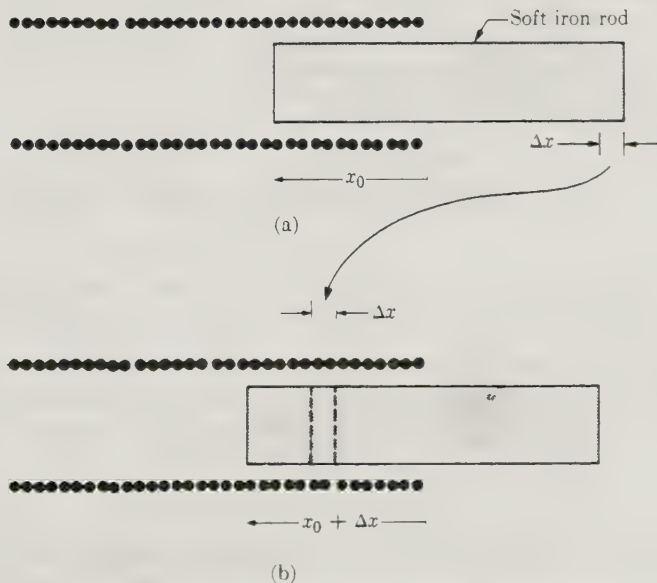
Consider a long solenoid of N turns, and length l carrying constant current I . An iron rod of constant permeability μ and cross-sectional area A is inserted along the solenoid axis. If the rod is withdrawn (Fig. 12-1a) until only one-half of its length remains in the solenoid, calculate approximately the force tending to pull it back into place.

Solution: The magnetic field structure associated with this problem is complicated if end-effects are included. Fortunately, however, we do not have to calculate the entire magnetic energy of the system but merely the difference in energy between the two configurations shown in Fig. 12-1(a) and (b). The field structure is relatively uniform far from the ends of the rod and the solenoid. The essential difference between configurations (a) and (b) is that a length Δx from the extreme right-hand end of the rod (outside the field region) is effectively transferred to the uniform field region inside the solenoid, at a place beyond the demagnetizing influence of the magnet pole. Thus, since \mathbf{H} is nearly longitudinal in the region Δx , and since the tangential component of \mathbf{H} is continuous at the cylindrical boundary of the rod, let us use

$$U = \frac{1}{2} \int \mu H^2 dv$$

where \mathbf{H} is constant inside and outside the rod because I is constant.

FIGURE 12-1
 Force on soft-iron rod
 inserted into a solenoid
 (by the energy method)



Consequently,

$$\begin{aligned} U(x_0 + \Delta x) &\approx U(x_0) + \frac{1}{2} \int_{A \Delta x} (\mu - \mu_0) H^2 dv \\ &= U(x_0) + \frac{1}{2} (\mu - \mu_0) \frac{N^2 I^2}{l^2} A \Delta x \end{aligned}$$

and from Eq. (12-20)

$$F_x \approx \frac{1}{2} (\mu - \mu_0) \frac{N^2 I^2 A}{l^2} = \frac{1}{2} \chi_m \mu_0 H^2 A \quad (12-26)$$

in the direction of increasing x_0 .

An example where Φ is constant is found in Problem 12-7.

***12-4**
HYSTERESIS LOSS

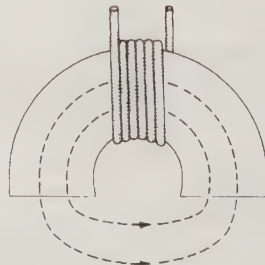
In the preceding sections, we have limited our discussion to reversible magnetic systems and in most instances to linear systems. We shall now say something about energy changes in systems containing permanent magnet material—that is, systems in which hysteresis plays a prominent role. Let us consider an electrical circuit, in the form of a closely wound coil of N turns, which surrounds a piece of ferromagnetic material (Fig. 12-2). If the coil is connected to an external source of electrical energy, the work done against the induced emf in the coil is given by Eq. (12-3). In Eq. (12-3), however, the flux change $\delta\Phi$ is the total flux change through the circuit; for the present purpose it is convenient to let the symbol $\delta\Phi$ stand for the flux change through a single turn of the coil. Thus, on the assumption that the same flux links every turn,

$$\delta W_b = NI \delta\Phi \quad (12-3a)$$

Let us treat the ferromagnetic specimen as forming part of a magnetic

FIGURE 12-2

A ferromagnetic specimen forming part of a magnetic circuit



circuit. Then NI may be replaced by $\oint \mathbf{H} \cdot d\mathbf{l}$ around a typical flux path, and Eq. (12-3a) becomes*

$$\delta W_b = \oint \delta \Phi \mathbf{H} \cdot d\mathbf{l} = \oint A \delta B \mathbf{H} \cdot d\mathbf{l}$$

where A is the cross section of the magnetic circuit appropriate to the length interval $d\mathbf{l}$. Since $d\mathbf{l}$ is always tangent to the flux path, the preceding equation may be written as

$$\delta W_b = \oint A \delta \mathbf{B} \cdot \mathbf{H} d\mathbf{l} = \int_V \delta \mathbf{B} \cdot \mathbf{H} dv \quad (12-27)$$

where V is the volume of the magnetic circuit—that is, the region of space in which the magnetic field is different from zero.

If the ferromagnetic material in the system shows reversible magnetic behavior, Eq. (12-27) may be integrated from $\mathbf{B} = 0$ to its final value, to yield the magnetic energy of the system. For linear material, the energy so obtained is identical with that expressed by Eq. (12-15). But Eq. (12-27) is much more general; it correctly predicts the work done on the magnetic system even for cases in which there is hysteresis.

According to Eq. (12-27), a change in the magnetic field structure implies a work input

$$dw_b = \mathbf{H} \cdot d\mathbf{B} \quad (12-28)$$

associated with each unit volume of magnetic material (or vacuum) in the system. Of particular interest is the case where the material is cycled, as it would be when the coil surrounding the specimen is subjected to alternating current operation. In one cycle, the magnetic intensity H (for a typical point in the specimen) starts at zero, increases to a maximum, H_{\max} , decreases to $-H_{\max}$, and then returns to zero. The magnetic induction B shows a similar variation, but for a typical ferromagnetic will lag behind H , thus tracing out a hysteresis curve (Fig. 12-3). The work input (per unit volume) required to change the magnetic induction from point a to b on the hysteresis curve,

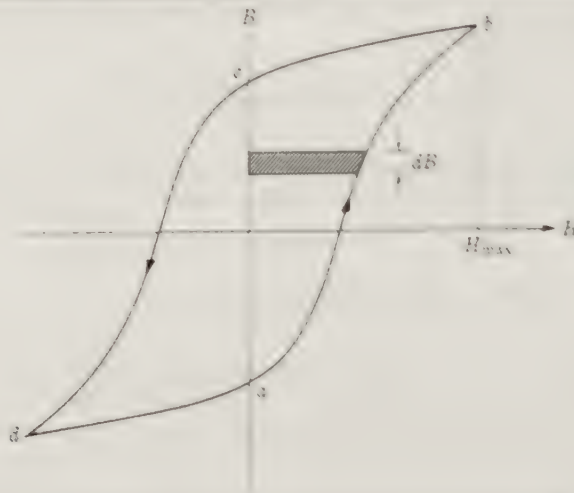
$$(w_b)_{ab} = \int_a^b H dB$$

* The analysis presented here may be put on a somewhat more rigorous basis by replacing the magnetic circuit with a large number of magnetic flux paths of various lengths (magnetic circuits in parallel). Equation (12-3a) then becomes

$$\delta W_b = NI \sum_j \delta \Phi_j = \sum_j \oint \delta \Phi_j \mathbf{H} \cdot d\mathbf{l}_j$$

where $\delta \Phi_j$ is the flux change associated with one of these paths. The final result, Eq. (12-27), is unchanged.

FIGURE 12-3
Work done per unit volume in cycling a ferromagnetic material



is just the area between the hysteresis segment ab and the B -axis. It is positive because both H and dB are positive. The contribution $(w_b)_{bc}$ is also the area between the appropriate hysteresis segment (bc) and the B -axis, but it must be taken negative, since H and dB are of opposite sign. Similar arguments can be made about $(w_b)_{cd}$ and $(w_b)_{da}$. Thus, in cycling the material once around the hysteresis loop, the work required per unit volume is

$$w_b = \oint H dB \quad (12-29)$$

which is the area enclosed by the hysteresis loop.

At the end of one complete cycle, the magnetic state of the material is the same as it was at the start of the cycle. Hence, the "magnetic energy" of the material is the same. It is evident, then, that Eq. (12-29) represents an energy loss. This loss appears as heat that is produced by irreversible changes in domain structure of the material. Hysteresis loss is an important factor in circuits subjected to alternating current operation. Equation (12-29) represents the energy loss per unit volume *per cycle*. Thus, the energy loss per unit time is directly proportional to the frequency of the alternating current.

According to Eq. (12-28), the work required to change the magnetic induction in a unit volume of material is

$$dw_b = \mathbf{H} \cdot d\mathbf{B} = \mu_0 H dH + \mu_0 \mathbf{H} \cdot d\mathbf{M} \quad (12-28a)$$

It is sometimes convenient to regard the $\mu_0 H dH$ term (the work done on the vacuum) as taking place whether the material is present or not. From

this point of view, then, the term $\mu_0 \mathbf{H} \cdot d\mathbf{M}$ is the specific work done on the material. This approach is usually taken in thermodynamics textbooks, and forms the basis for discussion of such processes as "magnetic cooling."

Since the integral of $H dH$ vanishes for a complete cycle, Eq. (12-29) is equivalent to

$$w_b = \mu_0 \oint H dM \quad (12-29a)$$

From $d(MH) \equiv H dM + M dH$, this can also be written

$$w_b = -\mu_0 \oint M dH \quad (12-29b)$$

12-5

SUMMARY

The work that is done by an outside agency (e.g., a battery) in altering the magnetic field of a system of current circuits is

$$dW_b = \sum_{i=1}^n I_i d\Phi_i$$

(exclusive of the work that supplies the Joule heat loss of resistive circuits). The **magnetostatic potential energy** of a system of current circuits and linear magnetic media is

$$U = \frac{1}{2} \sum_{i=1}^n I_i \Phi_i$$

where

$$\Phi_i = \sum_{j=1}^n M_{ij} I_j$$

For a continuous current distribution in linear media, the magnetic energy becomes

$$U = \frac{1}{2} \int \mathbf{J} \cdot \mathbf{A} dv$$

where the vector potential \mathbf{A} is that produced by the current density \mathbf{J} . Integration by parts transforms the energy in linear magnetic materials into an integral,

$$U = \int u dv$$

over the **energy density** of the magnetic field,

$$u = \frac{1}{2} \mathbf{H} \cdot \mathbf{B} = \frac{1}{2} \mu H^2 = \frac{1}{2} \frac{B^2}{\mu}$$

- For a single circuit,

$$U = \frac{1}{2} I \Phi$$

with

$$\Phi = LI$$

- The **magnetic force** on part of an isolated system, with constant flux through each circuit, is the negative gradient of the magnetostatic energy,

$$F_x = - \left(\frac{\partial U}{\partial x} \right)_{\Phi}$$

If the system is not isolated, but instead the current in each circuit is maintained constant by an outside agency (battery), the force is given by

$$F_x = + \left(\frac{\partial U}{\partial x} \right)_I$$

- In the presence of nonlinear material, including hysteresis,

$$dw_b = \mathbf{H} \cdot d\mathbf{B}$$

In a complete cycle of a cyclic process,

$$w_b = \oint H dB = \mu_0 \oint H dM = -\mu_0 \oint M dH$$

PROBLEMS

12-1 Consider a current circuit (not necessarily rigid) in a *prescribed* magnetic field. The magnetic force on each circuit element $d\mathbf{l}$ is given by $I d\mathbf{l} \times \mathbf{B}$. If the circuit is allowed to move under the influences of the magnetic forces, such that an element is displaced $\delta \mathbf{r}$ and at the same time the current I is held constant, show by direct calculation that the mechanical work done by the force is $\delta W = I \delta \Phi$, where $\delta \Phi$ is the additional flux through the circuit.

12-2 Consider a set of interacting current circuits in a linear magnetic medium. All circuits with the exception of circuit 1 are held stationary, but circuit 1 is allowed to move rigidly. The currents are all held constant by means of batteries. Show from the combination of Eqs. (12-4), (12-6), and (12-18) that the mechanical work done by the moving circuit is $dW = I_1 d\Phi_1$, where $d\Phi_1$ is the change in flux through circuit 1.

12-3 Consider two interacting current circuits characterized by the inductances $L_1 = \beta I_1^s$, $M_{12} = M_{21} = \beta I_1^{s/2} I_2^{s/2}$, and $L_2 = \beta I_2^s$, where β and s are constants. This system is a reversible magnetic system but not a linear one. Calculate the magnetic energy of the system in terms of the final currents I_1 and I_2 . Do this in two ways: (a) by bringing the currents to their final values in concert and (b) by keeping $I_2' = 0$ while I_1' is brought to its final value, then changing I_2' .

12-4 A circuit in the form of a circular turn of wire of radius b is placed at the center of a larger turn of radius a , $b \ll a$. The small circuit is fixed so that it is free to rotate about one of its diameters, this diameter being located in the plane of the larger

circuit. The circuits carry the steady currents I_b and I_a , respectively. (a) If the angle between the normals to the two circuits is θ , calculate the torque on the movable circuit. (b) In what direction is this torque when I_b and I_a circulate in the same sense?

***12-5** A U-shaped electromagnet of length l , pole separation d , and permeability μ has a square cross section of area A . It is wound with N turns of wire carrying a current I . Calculate the force with which the magnet holds a bar of the same material (of the same cross section) against its poles.

12-6 A permanent magnet with constant magnetization, and a circuit that is connected to a battery, form an isolated system. The circuit is allowed to move relative to the magnet, the current I in the circuit being maintained constant. The mechanical work done by the circuit is given in Problem 12-1. What conclusion can you draw about the change in magnetic energy of this system?

12-7 The magnetic induction field between the poles of an electromagnet is relatively uniform and is held at the constant value \mathbf{B}_0 . A thin paramagnetic slab, which is constrained to move vertically, is placed in the field as shown in Fig. 12-4. The susceptibility of the slab is χ_m and its cross-sectional area is A . (a) Calculate the force on the slab. (b) Obtain a numerical value for the force if the slab material is titanium, $A = 1 \text{ cm}^2$ and $B_0 = 0.25 \text{ T}$.

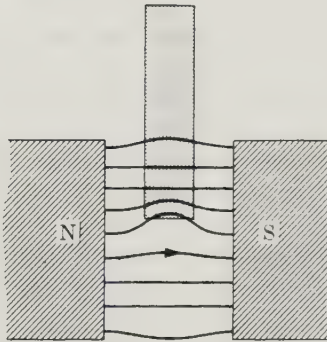


FIGURE 12-4

A paramagnetic slab inserted between the pole faces of a magnet

***12-8** From the result of Problem 12-1, the force on a current circuit in a prescribed magnetic field is given by $\mathbf{F} = I \nabla \Phi$. If the circuit is very small, the magnetic field \mathbf{B} may be treated as constant over the surface bounded by the circuit. Furthermore, the circuit itself may be characterized by its magnetic dipole moment \mathbf{m} . Show that when the prescribed magnetic field has no sources (i.e., $\mathbf{J}, \mathbf{J}_M = 0$) at the position of the dipole, the force on the dipole is

$$\mathbf{F} = (\mathbf{m} \cdot \nabla) \mathbf{B}$$

12-9 A rigid circuit consisting of a single loop of wire is located in a radial, inverse-square, magnetic induction field, $\mathbf{B} = K\mathbf{r}/r^3$. Show that the force on the

circuit is $\mathbf{F} = KI\nabla\Omega$, where Ω is the solid angle the circuit subtends at the field center, and I is the current in the circuit.

12-10 The center of a plane circular circuit of radius R and consisting of one turn lies on the x -axis at distance x from the origin. The circuit carries the current I , and its positive normal points in the $-x$ direction. Find the force exerted on the circuit by a radial induction field diverging from the origin, $\mathbf{B} = K\mathbf{r}/r^3$.

12-11 Consider a very long solenoid of N/l turns per unit length and radius R , such that the field inside is approximately uniform and the field outside is zero. Find the radial force on one turn of the winding, per unit length of circumference, from the magnetic energy. (a) Assume that the current I is maintained constant by a battery. (b) Repeat assuming that the flux remains constant and the system is isolated (with superconducting windings).

12-12 Solve the example of Fig. 12-1 with the magnetic energy in the form $U = \frac{1}{2}LI^2$, where $L = L(x_0)$ is the inductance of the solenoid with the iron rod inserted a distance x_0 . Assume that the rod's diameter is nearly as large as the solenoid's, and that both are so long that end effects are negligible.

12-13 For the toroid of Problem 11-10, (a) find the radial force on the coil if it carries a current I . (b) Does the force tend to expand the coil or collapse it?

12-14 Find the force between the straight wire and the rectangular circuit of Problem 11-15 if the currents are I_1 and I_2 .

12-15 Two isolated superconducting circuits carry certain currents when they are positioned so that their mutual inductance is zero. Now they are moved so that their mutual inductance is M . If the circuits are identical and had the same initial currents I_0 , find the final currents I .

12-16 (a) Prove that $\frac{1}{2}\int_V \mathbf{H} \cdot \mathbf{B} dv = 0$, where V is all space, if the fields are produced solely by magnets (i.e., no transport currents). (b) Is the magnetic energy zero?

12-17 (a) Estimate the areas enclosed by the two hysteresis curves shown in Fig. 9-8, and (b) calculate the power loss per unit volume due to hysteresis in these materials at 60 Hz operation.

12-18 The core of a generator armature is made of iron whose average hysteresis loop under operating conditions has an area of 2000 joules/m³. The core is cylindrical in shape, with a length of 0.4 m and a diameter of 0.15 m. If it rotates at 1800 r/min, calculate the rate at which heat is produced in the core.

12-19 A U-shaped permanent magnet (cross section 4 cm² and length 0.3 m) holds a piece of soft iron (of the same cross section) tightly against its poles. Use Eq. (12-28) to calculate the work required to pull the magnetic circuit apart. Specifically, calculate this work for each of the materials shown in Fig. 9-9. (Assume that initially $B = B_r$, and that after separation the magnet is subjected to a demagnetizing line of slope -4.0 .)

Slowly Varying Currents

In Chapter 7, the idea of an electrical circuit was introduced, and an analysis was made of the currents in such circuits when they are excited by constant applied voltages. In this chapter, these ideas will be expanded to include slowly varying voltages and the resulting slowly varying currents. To understand properly what is meant by “slowly varying,” Maxwell’s equations must be used. However, the general ideas can be understood without recourse to the details of these equations.

For sinusoidal voltage variations in circuits containing linear elements, the basis for elementary circuit theory, the behavior of a circuit is characterized by a frequency ω .^{*} An electromagnetic wave of this frequency in free space has a wavelength $\lambda = 2\pi c/\omega$, where c is the velocity of light. The principal restriction to be imposed in order that the current in the circuit may be called slowly varying is that the circuit should not radiate an appreciable amount of power. This restriction can be met by requiring that the maximum linear dimension of the system, ℓ_{\max} , be much smaller than the free space wavelength associated with the driving frequency. That is,

$$\ell_{\max} \ll \frac{2\pi c}{\omega} \quad \text{or} \quad \omega \ll \frac{2\pi c}{\ell_{\max}} \quad (13-1)$$

If this condition is satisfied, then for every element $d\mathbf{l}$ of the circuit carrying a current I there is, much less than one wavelength away, a corresponding element $-d\mathbf{l}$ carrying the same current. This pairing clearly ensures

^{*} The quantity ω is 2π times the frequency and is sometimes called the angular frequency. The use of ω instead of $2\pi f$ is of considerable advantage in many branches of physics. In particular, for the present discussion it eliminates a multitude of 2π 's from the circuit equations.

TABLE 13-1

f (Hz)	ω (rad/s)	λ (m)	ℓ_{\max} (m)
60	376	5×10^6	5×10^5 (300 miles)
10^6	6.28×10^6	300	30
10^8	6.28×10^8	3	0.3
10^{10}	6.28×10^{10}	0.03	0.003

cancellation of the fields produced by these elements at distances of the order of a few wavelengths in all directions and thus shows that the fields associated with the circuit are confined to the vicinity of the circuit. To see what practical restrictions are imposed by Eq. (13-1), $\ell_{\max} \sim \lambda/10$ has been used as the maximum circuit dimension in constructing Table 13-1. The frequencies chosen are a power line frequency, a low radiofrequency (AM broadcast band), a high radiofrequency (FM and TV), and a microwave frequency. It is clear that for the first three frequencies ordinary circuits satisfy the criterion. However, for the last one, the circuit must be built in a cube about 0.1 inch on a side, which limits the applicability to integrated circuits. It should also be noted that at 100 MHz the wavelength and circuit dimensions are of laboratory size and, hence, that care must be used in applying ordinary circuit theory at this and higher frequencies. In the balance of this chapter, it will be assumed that the slowly varying criterion is satisfied, without further explicit comment.

13-1

TRANSIENT AND STEADY-STATE BEHAVIOR

If a network of passive elements is suddenly connected to a source or sources of voltage, currents arise. Regardless of the nature of the applied voltages, the initial variation of the currents with time is nonperiodic. If, however, the voltages vary periodically with the time,* then a long time after the application of the voltages the currents will also be found to vary periodically with the time. (Actually they become strictly periodic only after infinite time; however, any desired approximation to periodicity can be attained by waiting a sufficiently long time.)

It is convenient to discuss the behavior of circuits in two stages, according to whether the periodic or nonperiodic behavior is important. The periodic behavior is referred to as the *steady-state* behavior, while the nonperiodic behavior is known as the *transient* behavior. Both aspects are governed by the same basic integro-differential equations; however, the elementary techniques used in solving them are radically different in the two cases.

* A constant voltage should be understood as a special case of a periodic voltage, in which the period is infinite or the frequency is zero.

The analysis presented here is restricted to elementary transient analysis (primarily excitation by constant voltages) and steady-state analysis for sinusoidal excitations. For further details, readers are referred to classic books by Guillemin and by Bode* and to other more recent engineering texts.†

13-2 KIRCHHOFF'S LAWS

In Chapter 7, Kirchhoff's laws were introduced for direct current (dc) circuits. They must now be generalized to include slowly varying currents. The first generalization is to note that not only resistors but also capacitors and inductors must be included as circuit elements. Each such element has a potential difference between its terminals, which must be included in Kirchhoff's loop law. Since the name "*IR-drop*" is no longer appropriate for all of these, the name *counter voltage* will be adopted to specify the difference in potential between the terminals of a passive element. The other generalization is to observe that both of Kirchhoff's laws must apply at each instant of time. That is, they must apply to the instantaneous values of the currents, applied voltages, and counter voltages. The laws may now be stated.

Kirchhoff's law I. The algebraic sum of the instantaneous currents flowing toward a junction is zero.

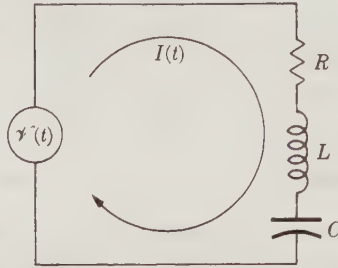
Kirchhoff's law II. The algebraic sum of the instantaneous applied voltages in a closed loop equals the algebraic sum of the instantaneous counter voltages in the loop.

The meaning of the first of these laws is clear: If currents directed toward a junction are called positive, then those oppositely directed should be called negative, and the law says that as much current enters the junction as leaves it. Basically, the second law represents the integral of the electric field around the loop; however, it is necessary to establish a sign convention. The sign convention to which we will adhere is best explained in terms of a single simple mesh, as shown in Fig. 13-1. In this figure, an applied voltage $\mathcal{V}(t)$ is shown connected in series with a resistance R , an inductance L , and a capacitance C . An arrow labeled $I(t)$ has been drawn to indicate the assumed (arbitrary) positive direction for the current. All signs are

* E. A. Guillemin, *Communication Networks*, 2 vols. (New York: Wiley, 1931 and 1935); and H. W. Bode, *Network Analysis and Feedback Amplifier Design* (Princeton, N.J.: D. Van Nostrand, 1945), (Huntington, N.Y.: Krieger, 1975, reprint of 1945 edition).

† For example, F. P. Yatsko and D. M. Hata, *Circuits: Principles, Analysis and Simulation* (New York: Saunders College Publishing, 1992).

FIGURE 13-1
Series circuit of circuit elements



ultimately referred to this direction. The voltage $\mathcal{V}(t)$ is positive if it tends to cause the current to move in the assumed direction—that is, if the top terminal in Fig. 13-1 is positive with respect to the bottom terminal. The resistive counter voltage is just IR , as in dc circuits. If dI/dt is positive, an emf will be induced in the inductance that tends to cause a current in the opposite direction to that assumed for I —that is, the upper terminal of L must be positive with respect to the lower terminal. Since the sense is the same as that of IR with respect to the direction of I , the counter voltage is just $L(dI/dt)$.* The capacitive counter voltage depends on the charge on the capacitor, which may be either positive or negative, depending on whether we consider the upper or the lower conductor. The difficulty is resolved by writing

$$Q = \int_{t_0}^t I(t) dt \quad (13-2)$$

where t_0 is chosen so that $Q(t_0)$ is zero. With this choice of Q , a positive Q makes the upper terminal of the capacitor positive and thus produces the capacitive counter voltage $+Q/C$. Kirchhoff's voltage law for the circuit of Fig. 13-1 is

$$\mathcal{V}(t) = RI + L \frac{dI}{dt} + \frac{1}{C} \int_{t_0}^t I dt \quad (13-3)$$

which is typical of the integro-differential equations of circuit theory.

13-3

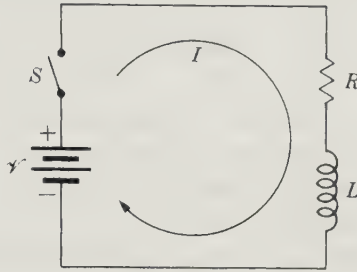
ELEMENTARY TRANSIENT BEHAVIOR

The only transient behavior we will consider is that associated with the sudden application of a constant voltage \mathcal{V} to a network of resistors,

* It is worth noting that the induced emf is written $-L(dI/dt)$; however, being an emf, it would normally be written on the other side of the equation from the counter voltages. Thus no inconsistency is introduced by writing $+L(dI/dt)$ for the counter voltage.

FIGURE 13-2

An R - L circuit. Transient response follows switch closure.



capacitors, and inductors. The first example is the simple R - L circuit shown in Fig. 13-2. For this circuit, Eq. (13-3) becomes

$$\mathcal{V} = RI + L \frac{dI}{dt} \quad (13-4)$$

after the switch S is closed. Before the switch is closed, the solution is trivial, being just $I = 0$. Equation (13-4) is a first-order linear differential equation with constant coefficients and, hence, can always be solved with one arbitrary constant in the solution. The solution is

$$I(t) = \frac{\mathcal{V}}{R} - Ke^{-tR/L} \quad (13-5)$$

with K the arbitrary constant. Since the circuit contains an inductance that prevents an abrupt change in the current, the current just after the switch is closed must be the same as the current just before the switch is closed—that is, zero. Thus, if the switch is closed at $t = t_0$,

$$\frac{\mathcal{V}}{R} - Ke^{-t_0R/L} = 0 \quad (13-6)$$

or

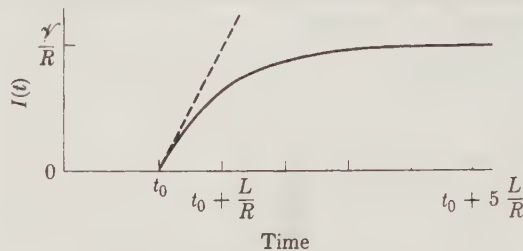
$$K = \frac{\mathcal{V}}{R} e^{t_0R/L} \quad (13-7)$$

The complete solution is then

$$I(t) = \frac{\mathcal{V}}{R} [1 - e^{-R(t-t_0)/L}] \quad (13-8)$$

which is plotted in Fig. 13-3. There are several useful, easily obtained facts that can be found from Eq. (13-8) and Fig. 13-3. First, L/R has the dimensions of time and is called the *time constant*. Since $1/e \cong 0.368$, the time constant is the time required for the current to reach 0.632 times its final value, \mathcal{V}/R . In five time constants, the current reaches 0.993

FIGURE 13-3
Transient response of an
R-L circuit



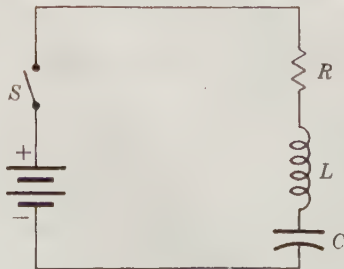
times its final value, which is conveniently remembered as 99 percent. The initial slope dI/dt is just the final current V/R divided by one time constant L/R —that is, a slope such that if the current continued to increase at this rate, it would reach its final value in one time constant. The usefulness of these facts is that by simply sketching a standard exponential curve, they enable evaluation of the exponential function involved in a simple transient problem to an accuracy of a few percent. Many other aspects of a resistance-inductance circuit can be explored, and a similar treatment can be applied to resistance-capacitance circuits. Several of the problems at the end of this chapter are devoted to accomplishing this end.

The second example to be considered is a series *R-L-C* circuit that is suddenly connected to a constant voltage V . Such a circuit is shown in Fig. 13-4. The appropriate equation after the switch is closed is

$$V = RI + L \frac{dI}{dt} + \frac{1}{C} \int_{t_0}^t I(t) dt \quad (13-9)$$

where again t_0 is a time at which the charge on the capacitor is zero. In the interest of simplicity, it will be assumed that the capacitor is initially uncharged and that the switch S is closed at $t_0 = 0$. Equation (13-9) may be unfamiliar; however, by simply differentiating it once with respect to the

FIGURE 13-4
An *R-L-C* circuit. Transient response follows switch closure.



time, it becomes

$$\frac{d\mathcal{V}}{dt} = R \frac{dI}{dt} + L \frac{d^2I}{dt^2} + \frac{I}{C} \quad (13-10)$$

This equation is an ordinary second-order linear differential equation with constant coefficients (the harmonic oscillator equation). The technique for solving such equations is well known, and in fact for the case at hand, $d\mathcal{V}/dt = 0$, the solution is*

$$I = \{Ae^{i\omega_n t} + Be^{-i\omega_n t}\}e^{-Rt/2L} \quad (13-11)$$

where

$$\omega_n = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}$$

so long as neither L nor C is zero. If either vanishes, an indeterminacy appears in Eq. (13-11); however, Eq. (13-10) can still be solved for $L = 0$; in fact, the solution is simpler than Eq. (13-11). Furthermore, the case $C = 0$ corresponds to the uninteresting case of an open circuit. To complete the discussion of this point, if $C = \infty$, which corresponds to short-circuiting the capacitor, Eq. (13-11) reduces to Eq. (13-5), with now two arbitrary constants to be obtained by fitting boundary conditions. This difference, of course, reflects the fact that all knowledge of \mathcal{V} was lost in going from Eq. (13-9) to Eq. (13-10).

We return now to the solution Eq. (13-11), where it remains to evaluate the constants A and B . For the current to be real, B must be the complex conjugate of A . Since the switch is closed at $t = 0$, at $t = 0$ the current must be zero, which means that the two imaginary exponentials must combine to give a sine function. These observations lead to

$$I(t) = De^{-Rt/2L} \sin \omega_n t \quad (13-12)$$

where D is a single real constant still to be evaluated. This evaluation is accomplished by noting that at $t = 0$, Q and I are both zero and, hence, that

$$\mathcal{V} = L \left. \frac{dI}{dt} \right|_{t=0} \quad (13-13)$$

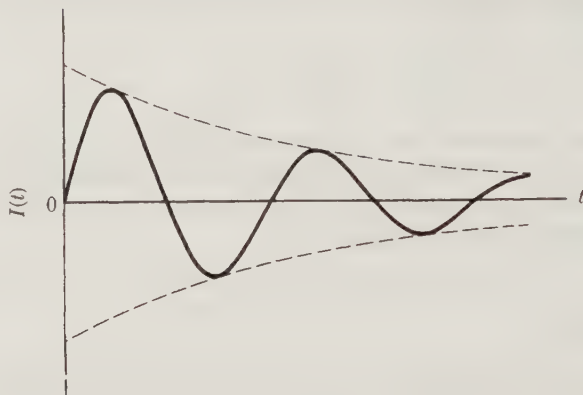
Using this initial condition gives

$$D = \frac{\mathcal{V}}{\omega_n L} = \frac{\mathcal{V}}{\sqrt{\frac{L}{C} - \frac{R^2}{4}}} \quad (13-14)$$

* Here i is the unit imaginary number—that is, $i \equiv \sqrt{-1}$.

FIGURE 13-5

Transient response of an
 R - L - C circuit



The solution is now complete. The current oscillates at the *natural frequency*

$$\omega_n = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}$$

with, however, an amplitude that decreases with time and is given by $De^{-Rt/2L}$. This behavior is shown in Fig. 13-5. If the time t_0 of closing the switch is not 0, it is only necessary to replace t with $t - t_0$.

The elementary transient analysis presented here is now complete. The balance of this chapter will be devoted to circuits excited by sinusoidal voltages in the steady state—that is, sufficiently long after the excitation has been applied to ensure that the transients are negligible.

13-4

STEADY-STATE BEHAVIOR OF A SIMPLE SERIES CIRCUIT

We now study the behavior of the circuit of Fig. 13-1 with the following excitation:

$$\mathcal{V}(t) = \mathcal{V}_0 \cos \omega t \quad (13-15)$$

where ω is a given frequency, not necessarily equal to ω_n . We could simply use this $\mathcal{V}(t)$ in Eq. (13-3) or Eq. (13-10) and solve the resulting equation. It is less cumbersome, however, to note that $\mathcal{V}_0 \cos \omega t$ is the real part of $\mathcal{V}_0 e^{i\omega t}$, and develop a method for finding the physical current from the solution to Eq. (13-10) with the complex excitation.

If a fictitious complex voltage $\mathcal{V}_1 + i\mathcal{V}_2$ were applied to the circuit, the resulting current would most certainly also be complex, $I_1 + iI_2$ (it is implied here that \mathcal{V}_1 , \mathcal{V}_2 , I_1 , and I_2 are all real). Putting these fictitious quantities

into Eq. (13-10) gives us

$$\begin{aligned} \frac{d\mathcal{V}_1}{dt} + i \frac{d\mathcal{V}_2}{dt} \\ = \left(L \frac{d^2 I_1}{dt^2} + R \frac{dI_1}{dt} + \frac{I_1}{C} \right) + i \left(L \frac{d^2 I_2}{dt^2} + R \frac{dI_2}{dt} + \frac{I_2}{C} \right) \end{aligned} \quad (13-16)$$

The only way this equation can be satisfied is if the real parts on the left and right are equal and the imaginary parts on the left and right are equal. Hence, if the physical excitation, $\mathcal{V}_p(t)$, is the real part of some complex function, $\mathcal{V}(t)$ is sufficient to solve Eq. (13-10) with the complex excitation $\mathcal{V}(t)$ and then take the real part of the resulting $I(t)$ to be the physical current. In some instances, it may be preferable to use $e^{i(\omega t + \phi)}$ in order to obtain the response to $\cos(\omega t + \phi)$, where ϕ is some given phase angle. In the rest of this chapter we will work primarily with complex voltages and currents. We will use the subscript P , as in $\mathcal{V}_p(t)$ above, where it is important to distinguish physical quantities from the corresponding complex quantities.

If $\mathcal{V}_0 e^{i\omega t}$ is used in Eq. (13-10), then the current will be $I_0 e^{i\omega t}$, with I_0 some complex constant. Direct substitution into the equation gives

$$i\omega \mathcal{V}_0 e^{i\omega t} = \left[-\omega^2 L + i\omega R + \frac{1}{C} \right] I_0 e^{i\omega t} \quad (13-17)$$

Dividing by $i\omega$ changes this to

$$\mathcal{V}_0 e^{i\omega t} = \left[R + i\omega L + \frac{1}{i\omega C} \right] I_0 e^{i\omega t} \quad (13-18)$$

which is in the form

$$\mathcal{V}_0 e^{i\omega t} = Z I_0 e^{i\omega t} \quad (13-19)$$

with

$$Z = R + i\omega L + \frac{1}{i\omega C} \quad (13-20a)$$

or

$$Z = R + i \left(\omega L - \frac{1}{\omega C} \right) \quad (13-20b)$$

The *impedance* Z of the circuit consists of two parts: the real part or *resistance* (R), and the imaginary part or *reactance* (X). The reactance is further divided into the *inductive reactance* $X_L = \omega L$ and the *capacitive reactance* $X_C = -1/\omega C$. The fact that the impedance is complex means that the current is not in phase with the applied voltage.

It is sometimes convenient to write the impedance in polar form:

$$Z = |Z| e^{i\theta} \quad (13-21)$$

with

$$|Z| = [R^2 + (\omega L - 1/\omega C)^2]^{1/2} \quad (13-22)$$

and

$$\theta = \tan^{-1}\left(\frac{\omega L - 1/\omega C}{R}\right) \quad (13-23)$$

Using this form for the impedance, we may write the complex current as

$$I(t) = \frac{\mathcal{V}_0}{|Z|} e^{i(\omega t - \theta)} \quad (13-24a)$$

and the physical current as

$$I_P(t) = \frac{\mathcal{V}_0}{|Z|} \cos(\omega t - \theta) \quad (13-24b)$$

If θ is greater than zero, the current reaches a specified phase later than the voltage and is said to lag the voltage. In the opposite case, the current leads the voltage. Currents that vary as $e^{i\omega t}$ or $\cos(\omega t - \theta)$ are called *alternating currents* (*ac*). This calculation formally completes the study of the simple series circuit, although in Section 13-7 we shall examine Eq. (13-24) in more detail to enhance our physical understanding of the situation.

13-5

SERIES AND PARALLEL CONNECTION OF IMPEDANCES

If two impedances are connected in series, then the same current flows through each of them. The voltages* across the two impedances are $V_1 = Z_1 I$ and $V_2 = Z_2 I$. The voltage across the combination is $V_1 + V_2 = (Z_1 + Z_2)I$. It is clear, then, that the connection of impedances in series adds the impedances, that is,

$$Z = Z_1 + Z_2 + Z_3 + \cdots \quad (\text{series connection}) \quad (13-25)$$

Thus, Eq. (13-20a) is the sum of the impedance of a resistance R ,

$$Z_1 = R$$

an inductance L ,

$$Z_2 = i\omega L$$

* In this and the remaining sections of the chapter, we shall use the symbol V in place of $\Delta\varphi$ for the potential difference across an element or group of elements.

and a capacitance C ,

$$Z_3 = \frac{1}{i\omega C}$$

all in series. It is important to note that the impedances add as complex numbers. If $Z_1 = R_1 + iX_1$ and $Z_2 = R_2 + iX_2$, then

$$Z = Z_1 + Z_2 = (R_1 + R_2) + i(X_1 + X_2) \quad (13-26)$$

In polar form,

$$Z = |Z| e^{i\theta}, \quad |Z| = [(R_1 + R_2)^2 + (X_1 + X_2)^2]^{1/2}$$

$$\theta = \tan^{-1} \frac{X_1 + X_2}{R_1 + R_2}. \quad (13-27)$$

Note that the magnitude of Z is **not** the sum of the magnitudes of Z_1 and Z_2 .

If impedances are connected in parallel, then the same voltage appears across each, and the currents are given by $I_1 = V/Z_1$, $I_2 = V/Z_2$, and so on. The total current is

$$I = I_1 + I_2 + \cdots = \frac{V}{Z_1} + \frac{V}{Z_2} + \cdots = V \left(\frac{1}{Z_1} + \frac{1}{Z_2} + \cdots \right)$$

from which it is clear that

$$\frac{1}{Z} = \frac{1}{Z_1} + \frac{1}{Z_2} + \cdots \quad (\text{parallel connection}) \quad (13-28)$$

Here, too, the addition is the addition of complex numbers. Equations (13-25) and (13-28) provide the basis for solving problems involving more complex configurations with a single applied voltage.

EXAMPLE 13-1 An Impedance Network

Consider the circuit of Fig. 13-6. The impedance consists of a resistor in series with the parallel combination of a capacitor and an inductor. Find the current through this network.

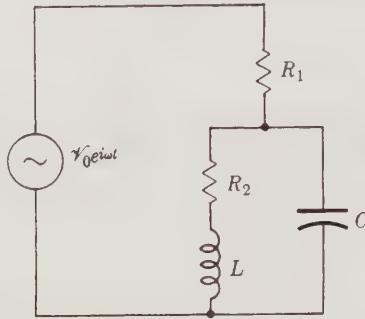
Solution: An expression for the impedance is conveniently written as

$$Z = R_1 + \frac{1}{\frac{1}{R_2 + i\omega L} + \frac{1}{1/i\omega C}} \quad (13-29)$$

Alternatively,

$$Z = R_1 + \frac{R_2 + i\omega L}{1 + i\omega C(R_2 + i\omega L)} \quad (13-30)$$

FIGURE 13-6
Typical ac circuit



or

$$Z = R_1 + \frac{(R_2 + i\omega L)[(1 - \omega^2 LC) - i\omega R_2 C]}{(1 - \omega^2 LC)^2 + \omega^2 R_2^2 C^2} \quad (13-31)$$

The only other worthwhile manipulation at this time is the separation into real and imaginary parts:

$$Z = R_1 + \frac{R_2}{(1 - \omega^2 LC)^2 + \omega^2 R_2^2 C^2} + i \frac{\omega L(1 - \omega^2 LC) - \omega R_2^2 C}{(1 - \omega^2 LC)^2 + \omega^2 R_2^2 C^2} \quad (13-32)$$

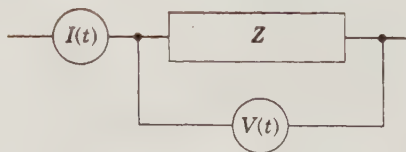
Having found Z , we now determine the current by dividing Z into $V_0 e^{i\omega t}$. The study of this circuit will be continued in Section 13-7 in connection with resonance phenomena.

13-6

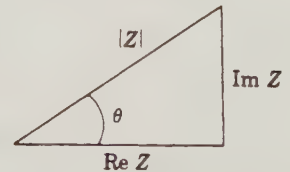
POWER AND POWER FACTORS

The power delivered to a resistor may be determined by multiplying the voltage across the resistor by the current through the resistor. However, for the more general case, such as the impedance shown in Fig. 13-7(a), a more

FIGURE 13-7
Measurement of power



(a)



(b)

subtle approach is required.

If $V(t)$ and $I(t)$ are the complex voltage and current as shown, then the *instantaneous power* is

$$P(t) = \text{Re } I(t) \text{ Re } V(t) \quad (13-33)$$

The *average power* is a more important quantity, with the average being taken over either one full period or a very long time (many periods). If the phases are chosen so that V_0 is real and, as usual, $Z = |Z| e^{i\theta}$, then it is straightforward to show (Problem 13-11) that

$$\bar{P} = \overline{\text{Re } I(t) \text{ Re } V(t)} = \frac{1}{2} |I_0| |V_0| \cos \theta \quad (13-34)$$

The factor one-half in Eq. (13-34) represents the fact that the average of $\sin^2 \omega t$ or $\cos^2 \omega t$ is one-half. The other interesting factor is cosine θ , which takes into account the fact that the current and voltage are not in phase. Cosine θ is frequently called the *power factor* of an ac circuit. In Section 17-3 it is shown that

$$\overline{\text{Re } (I_0 e^{i\omega t}) \text{ Re } (V_0 e^{i\omega t})} = \frac{1}{2} \text{Re } (I_0^* V_0) \quad (13-35)$$

where I_0^* is the complex conjugate of I_0 . This form is conveniently remembered and leads at once to Eq. (13-34).

Since $\cos \theta = \text{Re } Z / |Z|$ (refer to Fig. 13-7b), Eq. (13-34) can be expressed as

$$\bar{P} = \frac{1}{2} \frac{|V_0|^2}{|Z|^2} \text{Re } Z \quad (13-36)$$

As a final comment, we mention that the *effective values* of the voltage and current are often defined by

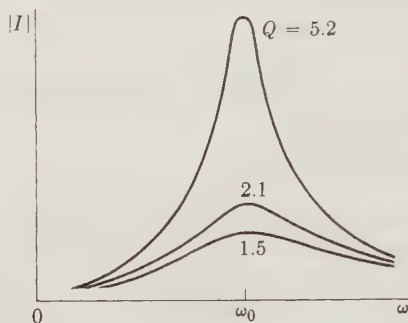
$$V_{\text{eff}} = \frac{\sqrt{2}}{2} |V_0|, \quad I_{\text{eff}} = \frac{\sqrt{2}}{2} |I_0| \quad (13-37)$$

The virtue of these definitions is that a given V_{eff} applied to a resistance dissipates the same power as a constant voltage of the same magnitude. The specification of effective values is very common; for example, 115-volt ac lines are 115 effective volt lines.

13-7 RESONANCE

Equation (13-22) shows that a simple series L - R - C circuit has a frequency-dependent impedance that is a minimum at $\omega^2 = \omega_0^2 \equiv 1/LC$. At this frequency the impedance is just R , the phase angle is zero, and the current is a maximum of magnitude V_0/R . This is a resonant phenomenon much like

FIGURE 13-8
Resonance curves for a
series R - L - C circuit



that observed in forced damped mechanical oscillators. If the magnitude of the steady-state current for the circuit in Fig. 13-1 (with $\mathcal{V}(t) = \mathcal{V}_0 \cos \omega t$) is plotted as a function of frequency, a curve of the form shown in Fig. 13-8 is obtained. Several curves are shown, all of which are based on the same values of L and C , but the series resistance varies from curve to curve. The curves are seen to be sharper for small than for large values of the series resistance.

A quantitative measure of the sharpness of the curve can be derived as follows. Let us define the “width” of the resonance curve as the frequency interval between the “half-power frequencies,” which are the two frequencies where the power dissipated is one-half of the power dissipated at the peak frequency ω_0 . Thus, we seek the values of ω that satisfy

$$\bar{P}(\omega) = \frac{1}{2} \bar{P}(\omega_0)$$

or by using Eqs. (13-22) and (13-36),

$$\frac{1}{2} \frac{|V_0|^2 R}{|Z(\omega)|^2} = \frac{1}{2} \frac{|V_0|^2}{R^2} R$$

Thus,

$$|Z(\omega)|^2 = 2R^2$$

or

$$\left| \omega L - \frac{1}{\omega C} \right| = R \quad (13-38)$$

For relatively sharply peaked responses, Eq. (13-38) will be satisfied at values of ω not far removed from ω_0 . We then write $\omega = \omega_0 + \Delta\omega$ and obtain

$$\left| \omega_0 L + \Delta\omega L - \frac{1}{\omega_0 C} \frac{1}{1 + \Delta\omega/\omega_0} \right| = R$$

Using $\omega_0^2 = 1/LC$ and $(1 + \Delta\omega/\omega_0)^{-1} \cong 1 - \Delta\omega/\omega_0$ gives

$$2|\Delta\omega|L = R$$

or

$$\frac{2|\Delta\omega|}{\omega_0} = \frac{R}{\omega_0 L} \quad (13-39)$$

The width of the resonance curve as defined above is thus approximated by $2|\Delta\omega|$. The quantity

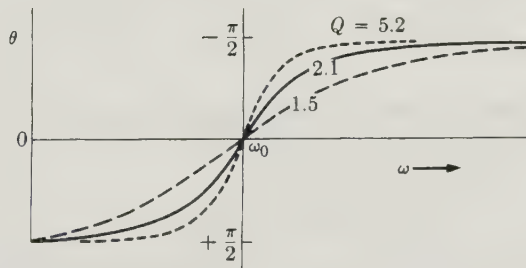
$$Q = \omega_0 L/R \quad \text{or} \quad Q = \frac{\omega_0}{2|\Delta\omega|} \quad (13-40)$$

characterizes the sharpness of the resonance and is known as the quality factor Q of the circuit.* For practical purposes, Q may be considered to be a property of the inductor only, since most of the unavoidable series resistance is associated with the wire with which the inductor is wound. However, a more refined treatment shows that the capacitor losses must also be included in computing Q 's. The curves of Fig. 13-8 are labeled with the appropriate Q values.

As the driving frequency is varied, not only the magnitude but also the phase of the current varies. This variation is shown in Fig. 13-9 for the same Q values used in Fig. 13-8. Below resonance, the phase angle of the impedance function is negative. Therefore, the phase of the current is positive, and it leads the voltage. Above resonance, the opposite is true, and the current lags the voltage.

It is interesting to note that radiofrequency resonant circuits found in communications equipment are often series resonant circuits, in spite of their parallel-circuit appearance. In the most simple case, the reason is that the driving power is inductively coupled into L and thus appears as an emf in series with L .

FIGURE 13-9
Phase angle of the
impedance in a typical
 R - L - C series circuit



* This Q has nothing to do with charge.

Resonance is not restricted to series circuits like those just discussed. Parallel circuits may also exhibit resonant characteristics. The circuit of Fig. 13-6 exhibits such a resonance. Defining the resonant frequency for a parallel resonant circuit is not as simple as it is for a series circuit. Some of the possibilities are: (1) $\omega_0 = 1/\sqrt{LC}$, (2) the frequency at which the magnitude of the impedance [calculated from Eq. (13-31)] is a maximum, or (3) the frequency at which the power factor is unity. Each of these three choices gives a different frequency. However, for high Q circuits they are very nearly the same. The first choice is by far the most useful in practice because it makes many series resonance results directly applicable to the parallel resonant case. One very interesting result is obtained by using Eq. (13-31) to evaluate Z , with $R_1 = 0$ and $\omega_0 = 1/\sqrt{LC}$. The result is

$$Z = \omega_0 L \left[\frac{\omega_0 L}{R} - i \right], \quad (\omega = \omega_0) \quad (13-41)$$

For a high Q circuit, the i can be neglected with the result that the impedance at resonance is Q times the inductive reactance at resonance.

The subject of resonant circuits can be pursued at great length. However, to do so here is unwarranted. Some of the problems at the end of the chapter extend this section.

***13-8**
MUTUAL INDUCTANCES IN ac CIRCUITS

Solving ac circuit problems involving mutual inductances presents a minor difficulty in assigning the correct sign to the mutual inductance. This difficulty can be readily resolved by noting that the sign to be associated with the mutual inductance depends on the assumed direction of the current in the two circuits involved and on the way in which the windings are connected. The notation M_{ij} will be used for the pure mutual inductance between two circuits.

It was shown in Chapter 11 that the emf in winding 2, due to a changing current in winding 1, is given in magnitude by

$$\mathcal{E}_2 = M_{21} \frac{dI_1}{dt} \quad (13-42)$$

For sinusoidal currents, using complex notation, we have

$$\mathcal{E}_2 = i\omega M_{21} \dot{I}_1 e^{i\omega t} \quad (13-43)$$

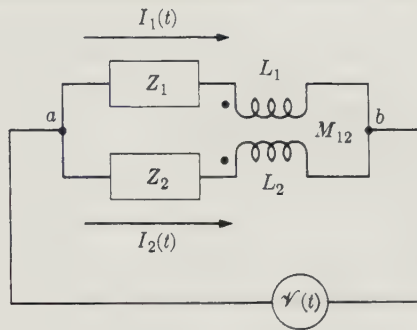
or

$$\mathcal{E}_2 = i\omega M_{21} I_1 \quad (13-44)$$

In what follows, the symbol M_{21} will be taken to be a positive quantity, and the sign of \mathcal{E}_2 will be displayed explicitly. In other words, M_{21} in Eq. (13-44) will be replaced by $\pm M_{21}$, with M_{21} being a positive quantity.

FIGURE 13-10

Circuit with mutual inductance



To demonstrate the technique for assigning signs, we now consider the circuit shown in Fig. 13-10, in which two impedances, Z_1 and Z_2 , are combined with a mutual inductance and connected to an applied voltage, $\mathcal{V}(t) = \mathcal{V}_0 e^{i\omega t}$. The mutual inductance is labeled M_{12} and is taken to be a positive number. The black dots in the figure indicate the ends of the two windings that are simultaneously positive—that is, if the lower winding is excited by a sinusoidal current that makes the left-hand terminal positive at some time t_1 , then the voltage induced in the upper winding makes the left-hand terminal of the upper winding positive at t_1 . The equation for the upper branch, in accordance with Kirchhoff's law, is

$$Z_1 I_1 + i\omega L_1 I_1 + i\omega M_{12} I_2 = \mathcal{V} \quad (13-45)$$

The plus sign is used with the mutual inductance because a positive I_2 gives a voltage in the upper branch that has the same sense as an $I_1 R$ drop. The second equation is

$$i\omega M_{21} I_1 + Z_2 I_2 + i\omega L_2 I_2 = \mathcal{V} \quad (13-46)$$

where $M_{12} = M_{21}$ has been written in the interests of symmetry.

The assignment of the sign is on the same basis as before and may be checked by noting that M_{12} should appear in the branch 1 equation with the same sign as M_{21} in the branch 2 equation. Equations (13-45) and (13-46) may be solved simultaneously by standard techniques to yield

$$I_1 = \mathcal{V} \frac{Z_2 + i\omega L_2 - i\omega M_{12}}{(Z_1 + i\omega L_1)(Z_2 + i\omega L_2) + \omega^2 M_{12}^2} \quad (13-47)$$

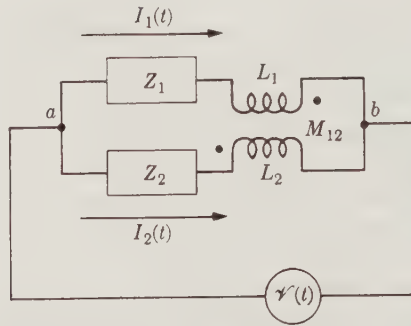
$$I_2 = \mathcal{V} \frac{Z_1 + i\omega L_1 - i\omega M_{12}}{(Z_1 + i\omega L_1)(Z_2 + i\omega L_2) + \omega^2 M_{12}^2}$$

Combining the two to obtain the total current $I_1 + I_2$ gives

$$I = I_1 + I_2 = \mathcal{V} \frac{Z_1 + i\omega L_1 + Z_2 + i\omega L_2 - 2i\omega M_{12}}{(Z_1 + i\omega L_1)(Z_2 + i\omega L_2) + \omega^2 M_{12}^2} \quad (13-48)$$

FIGURE 13-11

Circuit of Fig. 13-10 with the sign of the mutual inductance reversed



The coefficient of \mathcal{V} on the right side is the reciprocal of the impedance presented to the generator or the net impedance between points a and b . It is obvious that if M_{12} is zero, the impedance is the parallel combination of the two branch impedances. For the connection shown, as M_{12} increases, so does the impedance.

The circuit obtained by interchanging the leads on one winding of the mutual inductance is shown in Fig. 13-11. Note that the only difference is that the black dot has been moved from the left end of the upper winding to the right end. The result is to change the sign of the M_{12} term in Eqs. (13-45) and (13-46), with the result that

$$(Z_1 + i\omega L_1)I_1 - i\omega M_{12}I_2 = \mathcal{V}$$

and

$$-i\omega M_{12}I_1 + (Z_2 + i\omega L_2)I_2 = \mathcal{V}$$

(13-49)

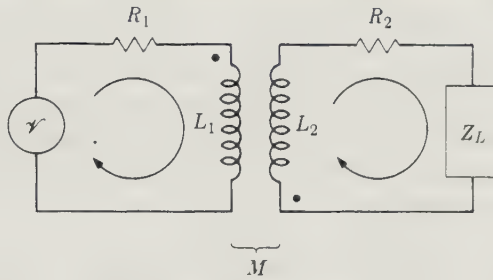
The currents are easily found and combined to obtain the impedance:

$$Z_{ab} = \frac{(Z_1 + i\omega L_1)(Z_2 + i\omega L_2) + \omega^2 M_{12}^2}{Z_1 + i\omega L_1 + Z_2 + i\omega L_2 + 2i\omega M_{12}} \quad (13-50)$$

which is the same as in the previous case when the mutual inductance is zero. The relationship between Z_{ab} for finite M_{12} and Z_{ab} for $M_{12} = 0$ depends on the parameter in a rather complicated way. We will state here only that $|Z_{ab}|$ may be larger or smaller than the $|Z_{ab}|$ for $M_{12} = 0$.

The basic circuit for the most common mutual inductance device, the transformer, is shown in Fig. 13-12. R_1 and R_2 are the resistances of the primary (driving) and secondary (driven) windings, L_1 and L_2 are their self-inductances, and M is the (positive) mutual inductance between them. Z_L is the impedance of the load connected to the secondary winding, and $\mathcal{V}(t) = \mathcal{V}_0 e^{i\omega t}$ is the voltage across the primary winding. If currents $I_1 e^{i\omega t}$ and $I_2 e^{i\omega t}$ are assumed to be in the directions indicated, then Kirchhoff's

FIGURE 13-12
Transformer circuit



voltage law requires that the equations

$$\mathcal{V}_0 = I_1 R_1 + i\omega L_1 I_1 + i\omega M I_2$$

and

$$0 = I_2 R_2 + i\omega L_2 I_2 + i\omega M I_1 + I_2 Z_L$$

(13-51)

be satisfied. The solutions to these equations are

$$I_1 = \frac{Z_L + R_2 + i\omega L_2}{(R_1 + i\omega L_1)(Z_L + R_2 + i\omega L_2) + \omega^2 M^2} \mathcal{V}_0$$

and

(13-52)

$$I_2 = \frac{-i\omega M}{(R_1 + i\omega L_1)(Z_L + R_2 + i\omega L_2) + \omega^2 M^2} \mathcal{V}_0$$

These relatively complicated equations represent an exact solution for the circuit of Fig. 13-12.

For many purposes, it is much more convenient to think in terms of an *ideal* transformer—that is, one that satisfies the relations

$$V_L = a \mathcal{V}_0, \quad I_2 = -\frac{I_1}{a} \quad (13-53)$$

where the constant a is independent of frequency, V_L is the voltage across Z_L , and all other quantities are as shown in Fig. 13-12. Multiplying Eqs. (13-53) together and assuming that the power factors in the two windings are equal, we find that the source power $\mathcal{V}_0 I_1 \cos \theta_1$ is all delivered to the load, $V_L I_2 \cos \theta_2$. In other words, there are no losses in the transformer.* The condition that must be satisfied to ensure the second of these relations is

* In the ideal transformer approximation, it can be shown that the power factors in the two windings are approximately equal (i.e., $\cos \theta_1 \approx \cos \theta_2$). If Z_L is purely resistive, then each power factor ≈ 1 .

$$\frac{Z_L + R_2 + i\omega L_2}{i\omega M} = a \quad (13-54)$$

which is satisfied if $\omega L_2 \gg |Z_L + R_2|$. From Eq. (13-54), $a \approx L_2/M - iZ_L/\omega M$ if $|Z_L| \gg R_2$. To satisfy the first Eq. (13-53), it is sufficient to require: $M \approx \sqrt{L_1 L_2}$, $\omega L_1 \gg R_1$, $|Z_L| \gg R_2$ and $|Z_L| \gg a^2 R_1$. Then it is easy to show that the second of Eqs. (13-52) together with (13-54) imply $V_L/\mathcal{V}_0 = a$, where $V_L \equiv -I_2 Z_L$:

$$\begin{aligned} \frac{V_L}{\mathcal{V}_0} &= -\frac{I_2 Z_L}{\mathcal{V}_0} \approx \frac{i\omega M Z_L}{i\omega L_1(i\omega M a) + \omega^2 M^2} \\ &= \frac{iZ_L}{\omega(M - L_1 a)} \approx \frac{iZ_L M}{\omega(M^2 - L_1 L_2) + iZ_L L_1} \\ &\approx M/L_1 \approx L_2 L_1 / (M L_1) \approx a \end{aligned}$$

The approximations made seem rather extreme, but practical transformers exist that approximate ideal transformers over relatively wide frequency ranges. For such devices,

$$I_2 = -\frac{I_1}{a}, \quad V_L = a\mathcal{V}_0$$

and

$$\frac{\mathcal{V}_0}{I_1} = -\frac{V_L}{a^2 I_2} = \frac{Z_L}{a^2} \quad (13-55)$$

The last of these relationships shows that the transformer acts also as an impedance transformer, with transformation ratio a^{-2} . It is left as an exercise to show that for very close coupling of the two windings $a = N_2/N_1$ —that is, the turns ratio.

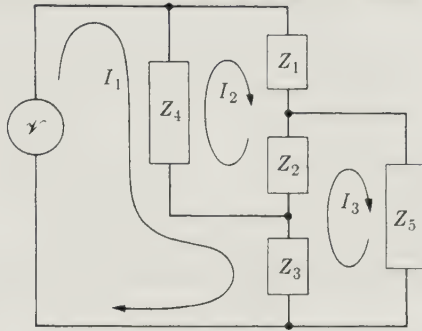
*13-9 MESH AND NODAL EQUATIONS

More complex ac circuits may be approached in two ways: one based on Kirchhoff's voltage law and known as *mesh analysis* and one based on Kirchhoff's current law and known as *nodal analysis*. Each method has its advantages and disadvantages. Since choosing the expedient method can greatly simplify some problems, both methods will be illustrated in this section.

The first step in applying mesh analysis is the assignment of meshes, which is accomplished by assuming closed loop currents such that at least one current goes through each element. With such a choice of currents, Kirchhoff's law I is satisfied automatically. For example, in Fig. 13-13 three meshes labeled I_1 , I_2 , and I_3 are shown. This set of meshes is, of course, not the only possible choice; several others are possible and useful. If

FIGURE 13-13

Illustration of the use of mesh analysis in ac circuits



Kirchhoff's voltage law II is applied to each of these meshes, we obtain

$$\begin{aligned} I_1(Z_3 + Z_4) - I_2Z_4 - I_3Z_3 &= \mathcal{V} \\ -I_1Z_4 + I_2(Z_1 + Z_2 + Z_4) - I_3Z_2 &= 0 \\ -I_1Z_3 - I_2Z_2 + I_3(Z_2 + Z_3 + Z_5) &= 0 \end{aligned} \quad (13-56)$$

Note that the minus signs appear because in mesh one, for example, I_2 flows through Z_4 counter to the direction of I_1 . Equations (13-56) can be solved most easily by matrix techniques, resulting in expressions for the set of mesh currents in the circuit. It is useful to note that the mesh equations can be written as

$$\sum_{j=1}^n Z_{ij}I_j = \mathcal{V}_i \quad (i = 1, 2, \dots, n) \quad (13-57)$$

(with $n = 3$ in the circuit above). In this notation, $Z_{ij} = Z_{ji}$, which is a useful check on the mesh equations.

As a second example of mesh analysis, consider the circuit of Fig. 13-14. The appropriate equations for this circuit are written as

$$\begin{aligned} I_1(Z_1 + Z_2) + I_2Z_2 &= \mathcal{V}_1 \\ I_1Z_2 + I_2(Z_2 + Z_3) &= \mathcal{V}_2 \end{aligned} \quad (13-58)$$

FIGURE 13-14

Further use of mesh equations

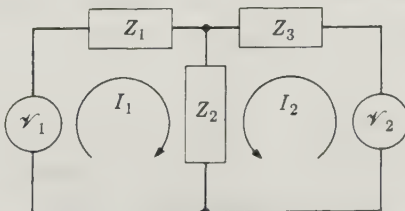
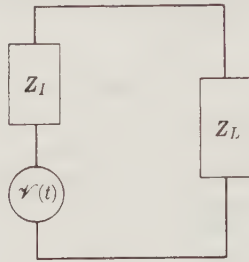


FIGURE 13-15

Practical generator connected to a load Z_L



There is no reason why \mathcal{V}_1 and \mathcal{V}_2 must be in phase. Usually they will not be in phase, but will be expressible as $\mathcal{V}_1 = |\mathcal{V}_{10}| e^{i\omega t}$, $\mathcal{V}_2 = |\mathcal{V}_{20}| e^{i(\omega t + \phi)}$. It is, however, very important to assign the phases correctly, and this task is most conveniently accomplished by examining the relative phases at $t = 0$ and assigning directions (senses) with respect to the assigned mesh currents. It is also important to note that unless all of the generators have the same frequency, the entire technique fails (more properly, the problem reduces to the superposition of two independent problems, each involving one generator and one frequency).

Before proceeding to discuss the alternative nodal equations, it is appropriate to discuss voltage and current generators. In the preceding sections, circuit problems have been phrased in terms of pure sources of applied voltage. Such idealized devices cannot be constructed, of course, since practical devices always have a certain internal impedance. Thus, a practical generator consists of a source of voltage, $\mathcal{V}(t)$, in series with an impedance Z_I , which is the internal impedance. Such a generator is shown in Fig. 13-15 connected to a load Z_L . Several observations may be made. First, for maximum power transfer to the external load, $Z_L = Z_I^*$; that is, Z_I and Z_L should have equal resistive parts and reactive parts that are equal in magnitude but opposite in sign. The proof of this statement is left as an exercise. Secondly, a voltage generator is equivalent to a current generator delivering a current $\mathcal{I}(t) = \mathcal{V}(t)/Z_I$ shunted by the internal impedance. This equivalence for the circuit of Fig. 13-15 is shown in Fig. 13-16. It is easy to show this equivalence if it is noted that an ideal current generator

FIGURE 13-16

A "current generator" that is equivalent to the voltage generator of Fig. 13-15

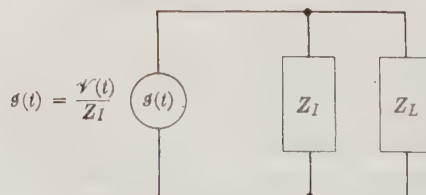
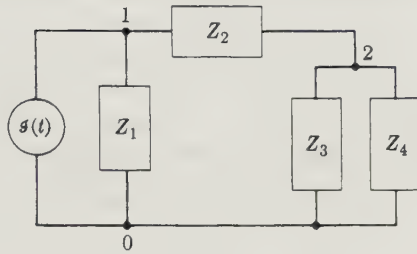


FIGURE 13-17
Illustration of the method
of nodal analysis in ac
circuits



delivers the current $\mathcal{I}(t)$ to *any* load connected to its terminals. The equivalence further means that in any circuit problem the generators may be taken either as voltage generators or as current generators, to suit the convenience of the situation.

The nodal equations for a circuit result from the application of Kirchhoff's current law I to each of the nodes, where a node is a point at which three or more elements join. In this procedure, Kirchhoff's voltage law II is satisfied automatically. As a simple example of the application of the nodal equations, we refer to the circuit of Fig. 13-17. The nodal equations are obtained by requiring that the algebraic sum of the currents to each node be zero. The nodes are numbered, starting with zero for the node whose potential is the reference for the circuit. If the potential at node 0 is taken to be zero, then at node 1

$$\mathcal{I}(t) = \frac{V_1}{Z_1} + \frac{V_1 - V_2}{Z_2} \quad (13-59)$$

where V_1 and V_2 are the potentials of nodes 1 and 2, respectively. At node 2,

$$0 = \frac{V_2 - V_1}{Z_2} + \frac{V_2}{Z_3} + \frac{V_2}{Z_4} \quad (13-60)$$

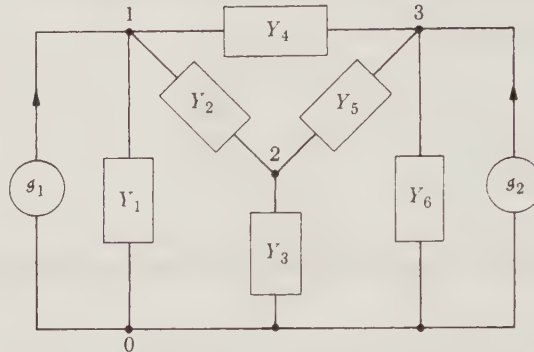
Before proceeding, we make the observation that a quantity that is the reciprocal of an impedance would be a great convenience. Such a quantity is the *admittance*, symbolized by Y : $Y = 1/Z$. Admittances in parallel add, while admittances in series combine by adding reciprocals. In terms of admittances, Eqs. (13-59) and (13-60) become

$$\begin{aligned} \mathcal{I}(t) &= (Y_1 + Y_2)V_1 - Y_2V_2 \\ 0 &= -Y_2V_1 + (Y_2 + Y_3 + Y_4)V_2 \end{aligned} \quad (13-61)$$

which are somewhat more convenient. The simultaneous solution of these equations yields the nodal voltages, V_1 and V_2 .

We shall consider one more example of the use of nodal equations, namely, for the circuit shown in Fig. 13-18. The nodal equations are simply

FIGURE 13-18
Another circuit illustrating nodal analysis



written down in the form

$$\begin{aligned} \mathcal{I}_1 &= Y_1 V_1 + Y_2(V_1 - V_2) + Y_4(V_1 - V_3) \\ 0 &= Y_2(V_2 - V_1) + Y_3 V_2 + Y_5(V_2 - V_3) \\ \mathcal{I}_2 &= Y_6 V_3 + Y_5(V_3 - V_2) + Y_4(V_3 - V_1) \end{aligned} \tag{13-62}$$

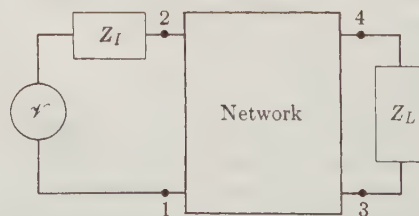
These equations may be solved by standard techniques to obtain the voltages at the nodes. The fact that voltages rather than currents are obtained when the equations are solved is a major advantage, particularly in communications circuits.

***13-10**

DRIVING POINT IMPEDANCE AND TRANSFER FUNCTIONS

We shall now present simple definitions for the driving point impedance and transfer functions of a four-terminal network. These definitions are presented because these terms appear in the technical literature and because they are sometimes a serious stumbling block to the uninitiated. Consider a four-terminal network, and call terminals 1 and 2 the input and 3 and 4 the output. If a generator of voltage \mathcal{V} and internal impedance Z_I is connected between terminals 1 and 2, and an impedance Z_L between terminals 3 and 4, as shown in Fig. 13-19, there will be a current I_I in Z_I , a current I_L in Z_L ,

FIGURE 13-19
Four-terminal network



and a voltage V_L across Z_L . The driving point impedance Z_D is

$$Z_D = \frac{\mathcal{V}}{I_t} \quad (13-63)$$

The four transfer functions that relate the input variables to the output variables are defined as follows.

Transfer impedance,

$$Z_T = \frac{\mathcal{V}}{I_L} \quad (13-64)$$

Transfer admittance,

$$Y_T = \frac{I_t}{V_L} \quad (13-65)$$

Current gain,

$$G_T = \frac{I_L}{I_t} \quad (13-66)$$

Voltage gain,

$$A_T = \frac{V_L}{\mathcal{V}} \quad (13-67)$$

It should be noted that the driving point impedance and the transfer functions depend on the source and load impedances as well as on the internal structure of the network. In many cases, however, adequate approximations can be made by using a zero impedance voltage source or an infinite impedance current source and an infinite impedance or zero impedance load.

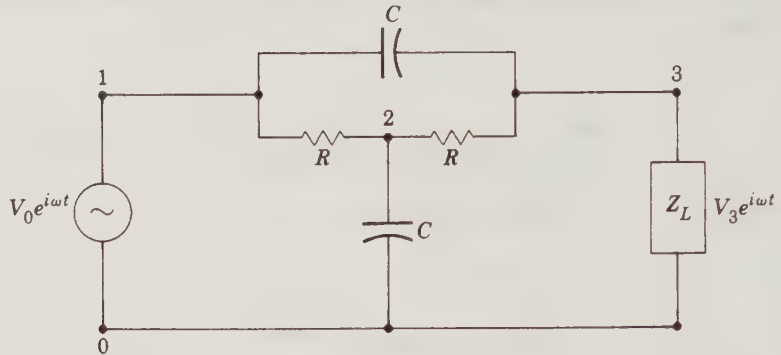
A brief treatment such as the above cannot do justice to the subject of network theory; classics such as that of Guillemin,* as well as the multitude of more recent books, should be consulted for the details of this complex subject.

13-11 SOLVING NETWORK EQUATIONS WITH THE COMPUTER

Finding the nodal voltages or loop currents in a complex network for a number of different frequencies is clearly a laborious task and one that is well suited to the use of a computer. To illustrate we consider a simple circuit known as the *bridged-T network*, as shown in Fig. 13-20. This circuit,

* E. A. Guillemin, *Communication Networks*, 2 vols. (New York: Wiley, 1931 and 1935).

FIGURE 13-20
Bridged-T network



which is a special case of the one shown in Fig. 13-13 or 13-18, attenuates frequencies in a band centered at $\omega_0 = 1/RC$ while passing frequencies some distance from ω_0 with relatively little attenuation. Thus, it can be used as a filter for removing unwanted frequencies. While much better filter circuits are available, the bridged-T is simple and demonstrates many aspects of circuit analysis. The nodal equations for this circuit can be written

$$\frac{1}{R}(V_2 - \mathcal{V}_0) + \frac{1}{R}(V_2 - V_3) + i\omega CV_2 = 0 \quad (13-68a)$$

$$\frac{1}{R}(V_3 - V_2) + i\omega C(V_3 - \mathcal{V}_0) + \frac{1}{Z_L} V_3 = 0 \quad (13-69a)$$

Introducing a dimensionless frequency, $F = \omega RC$ and rearranging the equations to put the known input on the right-hand side gives

$$(2 + iF)V_2 - V_3 = \mathcal{V}_0 \quad (13-68b)$$

$$-V_2 + \left(1 - iF + \frac{R}{Z_L}\right)V_3 = iF\mathcal{V}_0 \quad (13-69b)$$

Taking Z_L to be infinite, the solution for V_3/\mathcal{V}_0 is

$$\frac{V_3}{\mathcal{V}_0} = \frac{(1 - F^2)^2 + 6F^2 - iF(1 - F^2)}{(1 - F^2)^2 + 9F^2} = A_r + iA_i = Ae^{i\theta} \quad (13-70)$$

This closed form solution can be evaluated for any value of $F = \omega RC$. What is usually wanted, however, is a graph of V_3/\mathcal{V}_0 as a function of the frequency. For this purpose Eq. (13-70) must be evaluated repeatedly. This task is an ideal computer application. The simple program, BRIDGET2, in Fig. 13-21 computes the real and imaginary parts of the voltage gain and its magnitude and phase.

Since frequency responses are most conveniently plotted with a logarithmic frequency scale, the program first asks for the minimum value of F ,

FIGURE 13-21

BASIC program for calculating voltage gain of bridged-T network

```

100 '*****
110 '*
120 '*          TRANSFER FUNCTION FOR A SIMPLE BRIDGED T (BRIDGET2)
125 '*          NO ESSENTIAL LINE NUMBERS
130 '*
140 '*****
150 '
160 CLS
170 'Variable F = OMEGA*R*C is a dimensionless frequency.
180 '
190 INPUT "Low end of frequency range, FMIN = ", FMIN
200 INPUT "Number of points per decade, N = ", N
210 INPUT "Number of decades, M = ", M
220 'PRINT "Enter drive 'D' & file name 'BRIDT##' in format shown for data file"
230 'INPUT "D:\BRIDT## =", F$
240 PRINT
250 PRINT "For lineprinter output LP = 1": INPUT "LP = ", LP
260 '
270 PRINT "      F          AR          AI          A          THETA": PRINT
280 IF LP = 1 THEN LPRINT "      F          AR          AI          A          THETA"
290 IF LP = 1 THEN LPRINT
300 'OPEN F$ FOR OUTPUT AS #1
310 '
320 FOR K = 0 TO N * M
330 F = FMIN * (10 ^ (K / N))
340 AR = ((1 - F ^ 2) ^ 2 + 6 * F ^ 2) / ((1 - F ^ 2) ^ 2 + 9 * F ^ 2)
350 AI = (-F * (1 - F ^ 2)) / ((1 - F ^ 2) ^ 2 + 9 * F ^ 2)
360 A = SQR(AR ^ 2 + AI ^ 2)
370 THETA = ATN(AI / AR): 'radians
380 PRINT USING "###.###" ; F; AR; AI; A; THETA
390 'PRINT#1, USING "###.###" ; F, A
400 IF LP = 1 THEN LPRINT USING "###.###" ; F; AR; AI; A; THETA
410 NEXT K
420 '
430 'CLOSE #1
440 END

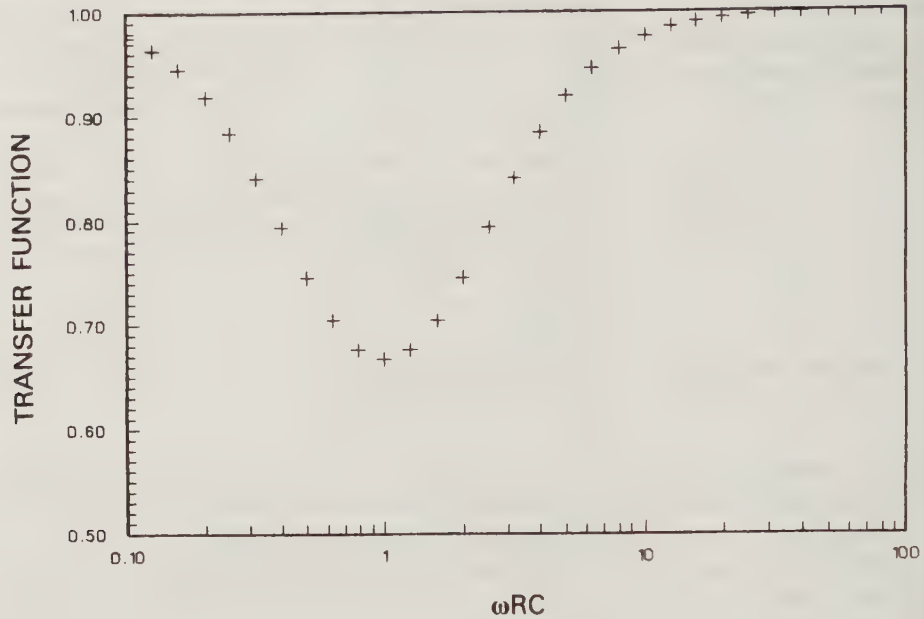
```

the number of points per decade of frequency, and the number of decades. The calculations and printing are done in lines 320 to 410. Line 330 generates a set of N frequencies per decade that are evenly spaced on a logarithmic scale. The balance of the program is standard simple calculation and printing.* The results for $F_{MIN} = 0.1$, $N = 10$ points per decade, $M = 3$ decades are plotted in Fig. 13.22. This graph was produced by generating a data file and using a commercial technical graphing package.† The program in Fig. 13-21 calculates 100 sets of values in about 20 seconds (on a computer with an 8088 processor chip, without coprocessor), an obvious enormous saving in time as compared to manual calculation. The avoidance of computational errors is another major advantage. In the future, we will refer to this approach as the *algebraic approach*.

* Lines 210 and 220 provide for a disk data file that is opened in line 300, printed in 390, and closed in line 450. Since these lines are used to make a data file for plotting the frequency response, they are disabled by leading apostrophes in Fig. 13-21. Lines 240 and 250 provide for lineprinter output. Only the response $LP = 1$ will cause the program to produce a printed copy of the results.

† The specific package used was TECH*GRAPH*PAD, Copyright Binary Engineering, Waltham, MA.

FIGURE 13-22
Magnitude of voltage
gain for bridged-T
network



If more than two equations are involved, obtaining the closed-form, algebraic solution is laborious and its evaluation requires a more complex program. Furthermore, a new program must be written for each new circuit. A useful alternative is to use a linear equation-solving program to obtain the required results. This approach will be called the *matrix approach*. The equations can be solved using the program G-J2KA, which we have already used in Chapter 3 (the listing is given in Fig. I-3 in Appendix I). There are two preliminary issues: The inputs to G-J2KA must be numbers and the numbers must be real. In Eqs. (13-68b) and (13-69b) the coefficients and the right-hand side, which comprise the inputs, depend on F and they are complex. A general scheme for solving such sets of equations would be first to pick a value for F and evaluate the coefficients and constants. The equations then involve complex numerical coefficients. Second, separate the equations into real and imaginary parts. A computer program to accomplish the separation is easy to devise. The number of equations is thus doubled but all of the coefficients and the right-hand sides are real. A standard program for solving linear equations, such as G-J2KA, can then be used to obtain a numerical solution. The next value of F is then selected and the process repeated.

For the simple case of the bridged-T network we can deal with the frequency dependent, complex coefficients easily by writing Eqs. (13-68b) and (13-69b) as four equations with real, frequency-dependent coefficients and constants as follows:

FIGURE 13-23

Control program (in BASIC) for solving networks by the matrix approach

```

1000 /*****
1010 /*
1020 /*          CONT1K: CONTROL PROGRAM FOR THE SOLUTION OF SIMULTANEOUS LINEAR
1030 /*          EQUATIONS WITH MATRIX ELEMENTS THAT DEPEND ON A PARAMETER
1040 /*          (E.G. FREQUENCY) USING GAUSS-JORDAN ELIMINATION
1050 /*          NO ESSENTIAL LINE NUMBERS
1060 /*
1070 /*****
1080 /
1090 'The equations are  A(I%,J%)W(J%) = D(I%)
1100 /
1110 'Identifiers:
1120 /      W      W      solution vector
1130 /      M1%    MAX%    maximum dimension of matrices
1140 /      N1%    NROW%   number of rows
1150 /      N2%    NCOL%   number of columns
1160 'End of identifiers.
1170 'Formats:
1180 A$ = "###.###^####": B$ = "=###.####": C$ = "###.####"
1190 INPUT "Maximum number of equations M1% = ", M1%
1200 DIM A(M1%,M1%), D(M1%), W(M1%)
1210 /
1220 GOSUB 4000:'Use SUB 4000 to read A1, A2, D1 and D2
1230 'Specify the frequency range in terms of lowest frequency,
1240 'number of points per decade and number of decades.
1250 INPUT "Low end of frequency range, FMIN = ", FMIN
1260 INPUT "Number of points per decade, N=", N
1270 INPUT "Number of decades, M = ", M
1280 PRINT "      F      W1      W2      W3      W4      W5      W6"
1290 FOR K = 0 TO N*M
1300   F = FMIN*(10^(K/N))
1310   GOSUB 5000:'Use SUB 5000 to compute matrix elements
1320   'Matrix elements and constants have been evaluated for F. Now solve.
1330   GOSUB 2000:'SUB 2000 is the Gauss-Jordan elimination program.
1340   PRINT USING C$; F,
1350   FOR I% = 1 TO N2%
1360     PRINT USING C$; W(I%),
1370   NEXT I%
1380   PRINT
1390 NEXT K
1400 END

```

$$2W_1 - FW_2 - W_3 = 1 \quad (13-71)$$

$$FW_1 + 2W_2 - W_4 = 0 \quad (13-72)$$

$$-W_1 + W_3 - FW_4 = 0 \quad (13-73)$$

$$-W_2 + FW_3 + W_4 = F \quad (13-74)$$

where $F = \omega RC$ as before, and $V_2/V_0 = W_1 + iW_2$; $V_3/V_0 = W_3 + iW_4$. These equations are in the form $\sum A_{ij}W_j = D_i$ and some of the quantities A_{ij} and D_i depend on F . Numerical values of these quantities must be obtained for each value of F before the equations can be solved for that value of F . The required instructions and inputs are provided by four short subroutines. The first is the control program, CONT1K, listed in Fig. 13-23. It sets the maximum dimensions of the arrays, calls subroutine INP4K, Fig. 13-24, which establishes the number of equations to be solved, and reads DAT6K, Fig. 13-25, to get the constants* $A1(I, J)$, $A2(I, J)$, $D1(I)$, and $D2(I)$ from

* We use A_{ij} for the symbolic quantity and $A(I, J)$, as it appears in computer programs, for the numerical value.

FIGURE 13-24

Subroutine (in BASIC)
for use with CONT1K
and G-J2KA

```

4000 /*****
4010 /**
4020 /** INP4K: SUBROUTINE TO INPUT A1(I%,J%); A2(I%,J%); C1(I%); C2(I%)
4030 /** USE WITH: CONT1K, INP5K, DAT6K
4040 /** LINE NUMBER 4000 IS ESSENTIAL
4050 /**
4060 /*****
4070 DIM A1(M1%,M1%), A2(M1%,M1%), C1(M1%), C2(M1%)
4080 INPUT "How many equations? N1% = ", N1%
4090 'N1 is the number of equations with real coefficients, i.e., twice
4100 'the number of equations with complex coefficients.
4110 IF (N1% > M1%) THEN PRINT "Subscript out of range": STOP
4120 IF (N1% < 2) THEN PRINT "Only one equation": STOP
4130 N2% = N1%
4140 'A(I%,J%) = A1(I%,J%) + F*A2(I%,J%); D(I%) = D1(I%) + F*D2(I%)
4150 PRINT
4160 PRINT "Coefficients A1, A2, D1 and D2 should be stored in lines 6000-"
4170 PRINT "If not, exit, type LIST 6000- and enter data."
4180 PRINT "Press CTRL BRK to exit. Press any key to continue."
4190 K$ = INPUT$(1)
4200 PRINT
4210 FOR I% = 1 TO N1%
4220 FOR J% = 1 TO N1%
4230 READ A1(I%,J%)
4240 READ A2(I%,J%)
4250 NEXT J%
4260 READ D1(I%)
4270 READ D2(I%)
4280 NEXT I%
4290 RETURN

```

which $A(I, J) = A1(I, J) + A2(I, J) * F$ and $D(I)$ will be calculated. Note that this subroutine and INP5K, also shown in Fig. 13-25, are both specific to this problem and must be modified if the circuit is changed. When control returns to the main program, the frequency range and number of points per

FIGURE 13-25

Subroutines (in BASIC)
for data entry and ma-
nipulation for use with
CONT1K

```

5000 /*****
5010 /**
5020 /** INP5K: SUBROUTINE TO COMPUTE A(I%,J%) and D(I%)
5030 /** LINE NUMBER 5000 IS ESSENTIAL
5040 /**
5050 /*****
5060 /
5070 FOR I% = 1 TO N1%
5080 FOR J% = 1 TO N1%
5090 A(I%,J%) = A1(I%,J%) + F*A2(I%,J%)
5100 NEXT J%
5110 D(I%) = D1(I%) + F*D2(I%)
5120 NEXT I%
5130 RETURN
6000 /*****
6010 /**
6020 /** DAT6K: DATA ENTRY FOR G-J INPUT SUBROUTINES INP4K and INP5K
6030 /** NO ESSENTIAL LINE NUMBERS
6040 /**
6050 /*****
6060 'Enter data A1(I%,J%), A2(I%,J%), D1(I%), D2(I%).
6070 'Use a separate line for each I.
6080 'There will be N1 lines with 2*N1 +2 data items per line.
6090 /
6100 'The following data are for the four equations for the Bridge T filter.
6110 /
6120 DATA 2.0, 0.0, 0.0, -1.0, -1.0, 0.0, 0.0, 0.0, 1.0, 0.0
6130 DATA 0.0, 1.0, 2.0, 0.0, 0.0, 0.0, -1.0, 0.0, 0.0, 0.0
6140 DATA -1.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, -1.0, 0.0, 0.0
6150 DATA 0.0, 0.0, -1.0, 0.0, 0.0, 1.0, 1.0, 0.0, 0.0, 1.0

```

decade are entered and the headings for the output table are printed. Subroutine INP5K (Fig. 13-25) is then used to calculate numerical values for $A(I, J)$ and $D(I)$. This procedure completes the determination of the parameters in the equations to be solved, and on return to the main program, control is immediately shifted to G-J2KA, which solves the equations, and then returns once again to the main program. The main program (CONT1K) prints the results, steps to the next F , and repeats the process. The output from this program is the set of real and imaginary components of the voltages at nodes two ($W1$ and $W2$) and three ($W3$ and $W4$). The pair for node three, $W3$ and $W4$ are just A_r and A_i as calculated above [Eq. (13-70)] using the algebraic approach. It takes about ten times as long to calculate the values for a single value of F . The programming is also more involved (even without using a subprogram to convert from equations with complex coefficients to equations with real coefficients). BRIDGET2 has fewer than 20 essential lines, while the matrix approach using G-J2KA has well over 100 essential lines. Why bother with the matrix approach at all? The answer to this question is that in the matrix approach, it is not necessary to find the algebraic solutions to the nodal equations. Instead the equations are solved numerically for each frequency by the computer. For two equations, avoiding the solution of the algebraic equations is not a particular advantage. The advantage of the matrix approach, however, rapidly becomes apparent as the number of equations increases. The other advantage is that a whole new program must be written to evaluate each algebraic solution (or else a lot of error-prone algebra must be done as a preliminary), whereas in the matrix approach it is only necessary to revise the relatively simple input and data programs. Some of the problems at the end of the chapter explore these points in more detail.

The third and final approach to solving network equations is to use commercial software. Most commonly used software is some version of SPICE (Simulation Program with Integrated Circuit Emphasis). As the name of the program indicates, SPICE is not limited to passive networks, but it deals with them very effectively. In SPICE, all of the nodes in a circuit are numbered and all the circuit elements named, for example, $R1$, $C4$, and so on. Each circuit element is entered into a "Netlist" that contains its name, the nodes between which it is connected, and the numerical parameters that characterize it. Driving voltages or currents are specified, as is the kind of analysis to be done: ac, dc, or transient. The output format is specified and the software does the rest. In an ac analysis, the equivalent of the matrix approach above, SPICE calculates all of the impedances, constructs the proper equations, and solves them for each value of frequency in the desired range. The output is presented in tabular or graphical form and data lists suitable for postprocessing can be written. Learning to use SPICE requires some effort, but if complex circuits must be analyzed—and especially if active circuits (circuits involving vacuum tubes or semiconducting devices) or nonlinear circuit elements must be

considered—the effort is well worthwhile. More information about SPICE can be found in Meares and Hymowitz.*

13-12 SUMMARY

As for dc circuits, the analysis of circuits carrying slowly varying currents depends on **Kirchhoff's current** and **voltage laws**, here applied at each instant of time. The current and voltage of a linear resistance are instantaneously related by **Ohm's law**:

$$V_R = RI$$

The analogous laws for a linear inductance and a linear capacitance are

$$V_L = L \frac{dI}{dt}, \quad V_C = \frac{1}{C} Q$$

where $I = dQ/dt$. For a single loop circuit containing an applied voltage $\mathcal{V}(t)$, Kirchhoff's laws result in the differential equation

$$L \frac{d^2 Q}{dt^2} + R \frac{dQ}{dt} + \frac{1}{C} Q = \mathcal{V}(t)$$

(For a more complicated network, the result is a system of such linear second-order differential equations.) The general solution is a superposition of a particular solution (steady-state solution), plus the general solution of the corresponding homogeneous equation obtained by putting $\mathcal{V}(t) = 0$ (transient solution). The arbitrary constants in the latter are chosen so as to satisfy the initial conditions imposed.

- The **transient solution** is exponential in t ; if the exponent is complex, the imaginary part physically represents an oscillation of Q and I . The real part represents a decaying behavior (transient), with a decay time which is small for large R . For most practical circuits, the decay time is typically a fraction of a second.

- The **steady-state behavior** is discussed only for a sinusoidal applied voltage (ac). (Constant voltage is the special zero-frequency case; an arbitrary time dependence could be represented as a Fourier synthesis.) If the applied voltage $\mathcal{V}_p(t) = \mathcal{V}_0 \cos \omega t$ is represented in the complex form

$$\mathcal{V} = \mathcal{V}_0 e^{i\omega t}$$

the steady-state current has the same frequency ω ,

$$I = I_0 e^{i\omega t}$$

* L. G. Meares and C. E. Hymowitz, *Simulating with SPICE* (San Pedro, CA: Intusoft, 1988).

The **complex impedance** Z is defined by

$$\mathcal{V} = ZI$$

When Z is expressed in polar form,

$$Z = |Z| e^{i\theta}$$

the modulus $|Z|$ gives the amplitude of the current

$$|I_0| = \mathcal{V}_0/|Z|$$

and θ gives the phase relative to the applied voltage. Ohm's law and its extensions yield

$$Z_R = R, \quad Z_L = i\omega L, \quad Z_C = \frac{1}{i\omega C}$$

since $I = i\omega Q$.

- The analysis of the steady-state behavior of linear ac circuits is exactly parallel to that of dc circuits, with the complex impedance serving as a generalization of the dc resistance.

- The instantaneous **power dissipation** is $P(t) = \text{Re } I(t) \text{Re } V(t)$. For ac excitation at a single frequency, the time-average power dissipation is

$$\bar{P} = \frac{1}{2} |I_0| |V_0| \cos \theta$$

The **power factor**, $\cos \theta$, is 1 for a pure resistance and 0 for a pure inductance of capacitance. The "effective" voltage and current are $|V_0|/\sqrt{2}$ and $|I_0|/\sqrt{2}$. An equivalent expression for the time-average ac power is

$$\bar{P} = \frac{1}{2} \text{Re}(I^*V)$$

- As a function of frequency, a series circuit exhibits a **resonance** near $\omega_0 = 1/\sqrt{LC}$, where $|Z|$ is a minimum and $|I_0|$ is a maximum. The sharpness of the resonance is given by

$$Q = \frac{\omega_0}{2|\Delta\omega|} = \frac{\omega_0 L}{R}$$

A parallel circuit also has a resonance near $\omega_0 = 1/\sqrt{LC}$, where $|Z|$ is a maximum and $|I_0|$ is a minimum.

- A **transformer** is the most common mutual inductance circuit element. An ideal transformer is one for which the ratio of secondary to primary current is the inverse of the ratio of secondary to primary voltage, and this ratio is frequency independent.

- Complex ac network problems are most easily solved by means of a computer. Two rather straightforward approaches that can be implemented on a personal computer are the **algebraic approach** and the **matrix approach**. More complex circuits such as those involving nonlinear or active circuit elements are best handled through the use of commercial software.

PROBLEMS

13-1 An inductance of 2 H and a resistance of 3 Ω are connected in series with a 5 V battery and a switch. Determine the current, and the rate of change of current (dI/dt) in the circuit at the following times after the switch is closed: (a) 0.3 s, (b) 1 s, (c) 4 s.

13-2 A circuit consisting of an inductance L_0 , a resistance R_0 , and a battery \mathcal{V}_0 has a steady current $I = \mathcal{V}_0/R_0$ through it. A switch in the circuit is opened at time $t = 0$, creating an arc across the switch. (a) If the arc resistance is given by k/I , where the constant $k < \mathcal{V}_0$, determine the current through the arc as a function of time. (b) What is the final steady value of current through the arc?

13-3 A capacitor C , a resistor R , and a battery \mathcal{V}_0 are connected in series with a switch. The switch is closed at time $t = 0$. (a) Set up the differential equation governing the charge Q on the capacitor. (b) Determine Q as a function of time.

13-4 A capacitor C with charge Q_0 is suddenly connected in series with a resistance R and inductance L . (a) Determine the current as a function of time. (b) Show that there are three different types of solution, depending upon whether $R^2 - 4L/C$ is less than, equal to, or greater than zero. The first of these conditions is called *underdamped*, the second *critically damped*, and the third *overdamped*.

13-5 A real capacitor C has a parallel leakage resistance R , and it is connected in series with an ideal inductance L . (a) Calculate $|Z|$. (b) Find the approximate values at high and low frequencies and at resonance, assuming that R is large. (c) Sketch a graph of $|Z|$ versus ω .

13-6 Repeat Problem 13-5 if the leaky capacitor is in parallel with the perfect inductor.

13-7 The circuit of Fig. 13-1 has an additional capacitor C' shunting the entire R - L - C combination. $R = 100 \Omega$, $L = 1$ H, $C = 100 \mu\text{F}$, and $C' = 10 \mu\text{F}$. Make a plot of the impedance $|Z|$ versus frequency from zero to $f = 10^4$ Hz.

13-8 The series combination of a resistance R and an inductance L is put in parallel with the series combination of resistance R and capacitance C . Show that if $R^2 = L/C$ the impedance is independent of frequency.

13-9 A wire-wound resistor has a dc resistance of 90.00 Ω and an inductance of 8 μH . (a) What is the phase angle of the impedance at 1000 Hz? (b) A capacitor is placed in parallel with the resistor to reduce the phase angle to zero at 1000 Hz without changing the resistance appreciably. Over what range of frequency is the phase angle less than it was before the capacitor was added?

13-10 (a) A capacitance C in parallel with a resistance R has an impedance Z . If a capacitance C' in series with a resistance R' has the same impedance Z , find the required values of C' and R' in terms of R and C for a given ω . (b) If the *dissipation factor* is defined as $D = \omega R' C'$, show that $D = 1/\omega RC$ and that the phase of the current is $\theta = \tan^{-1}(-1/D)$.

13-11 Prove Eq. (13-34) for the time-average power dissipation in a circuit that carries an ac current $I(t) = I_0 e^{i\omega t}$, with $V(t) = ZI(t)$.

13-12 An ac generator with internal impedance Z_i is connected in series with a variable load impedance Z_L . Prove that maximum power is transferred to the load when $Z_L = Z_i^*$.

13-13 Consider the circuit of Fig. 13-6, with $L = 4$ mH, $C = 2 \mu\text{F}$, $R_1 = 25 \Omega$,

$R_2 = 40 \Omega$. Find the following set of frequencies: (a) where $\omega = 1/\sqrt{LC}$, (b) where the impedance is maximum, (c) where the current through R_1 is in phase with the generator voltage.

13-14 Show that the quality factor Q defined in the text can be expressed as 2π times the maximum energy stored in the circuit, divided by the energy dissipated in one cycle. This statement is sometimes used as the definition of Q and is independent of specific circuit parameters.

13-15 A crossover network for a hi-fi set is to be designed so that two loudspeakers (each of resistance R) are connected to the output stage of an amplifier. One speaker is to receive predominantly high frequencies, the other predominantly low frequencies. The network is as shown in Fig. 13-26. The two capacitors are each of capacitance C and the two inductors each of inductance L . (a) Find a relationship between L and C for a given R such that the network presents a purely resistive load ($=R$) to the amplifier at all frequencies. (b) The crossover frequency ω_c is defined as the frequency at which each speaker receives half of the power delivered by the amplifier. For a given R and ω_c determine L and C .

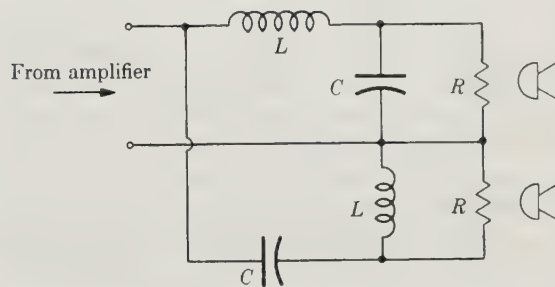


FIGURE 13-26

13-16 A $1 \mu\text{F}$ capacitor is first charged to 100 V by connecting it to a battery; it is then disconnected and immediately discharged through the 300-turn winding on a ring toroid. The toroid has a mean radius of 20 cm, 4 cm^2 cross-sectional area, and an air gap of 2 mm (see Fig. 9-16). Neglecting copper losses, hysteresis, and fringing, calculate the maximum magnetic field subsequently produced in the air gap. Take the relative permeability of the toroid equal to 5000.

13-17 A potential difference of 1 V at a frequency $f = 10^6/\pi$ Hz is impressed across the circuit of Fig. 13-27. The mutual inductance of the coils is such that they are in opposition. Find the current in the upper branch.

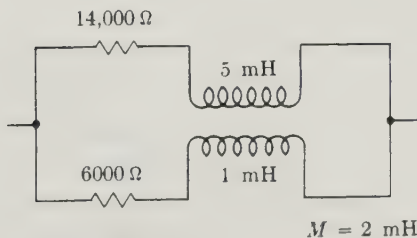


FIGURE 13-27

13-18 A 60 Hz power transformer (turns ratio 2:1) has a primary inductance of 100 H and a dc resistance of $20\ \Omega$. The coupling coefficient between primary and secondary is close to unity. If 1000 V is placed across the primary, calculate the current in the primary winding (a) when the secondary is open-circuited, (b) when a load resistance of $20\ \Omega$ is in the secondary circuit.

***13-19** Three identical capacitors and three identical inductors are connected as shown in Fig. 13-28. Find the natural frequencies of the system. [Hint: Use mesh analysis, with current of an assumed frequency ω , and show that the three equations obtained are compatible for certain ω only.]

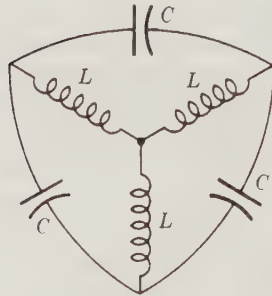


FIGURE 13-28

13-20 In the circuit shown in Fig. 13-14, $Z_1 = 2 + 5i$, $Z_2 = 8 - i$, $Z_3 = 4 + 3i$. The voltage generators are in phase with each other: $\mathcal{V}_1 = 10\text{ V}$, $\mathcal{V}_2 = 2\text{ V}$. Determine I_1 and I_2 .

13-21 Consider the circuit shown in Fig. 13-17. $Z_1 = R$; Z_2 is the parallel combination of resistance R and capacitance C ; and $Z_3 = R$ and $Z_4 = 1/(i\omega C)$. Z_3 and Z_4 together comprise the load impedance Z_L . Determine the transfer function G_T as a function of $F = \omega RC$ and plot its magnitude from $F = 0.1$ to 10.

13-22 For the circuit of Fig. 13-20, $Z_L = 10R$. Derive an equation for V_3/\mathcal{V}_0 in terms of $F = \omega RC$.

13-23 Consider the circuit of Fig. 13-20 with Z_L made up of the parallel combination of $2R$ and $3C$. Find equations of the form (13-71) to (13-74) for the voltage ratios V_2/\mathcal{V}_0 and V_3/\mathcal{V}_0 .

Physics of Plasmas

A *plasma* is a highly ionized gas, with properties so special that it is sometimes called a fourth state of matter. Gases that are highly ionized are good conductors of electricity. The charged particles in such a gas interact with the local electromagnetic field; furthermore, the organized motion of these charge carriers (currents, fluctuations in charge density) can *produce* magnetic and electric fields. When subjected to a static electric field, the charge carriers in the gas rapidly redistribute themselves in such a way that most of the gas is shielded from the field. To the relatively field-free regions of the gas where positive and negative space charges are nearly balanced, Langmuir† gave the name *plasma*, while to the space-charge or strong-field regions on the boundary of the plasma he gave the name *sheaths*.

Equivalently, we may say: An ionized gas that has a sufficiently large number of charged particles to shield itself, electrostatically, in a distance small compared with other lengths of physical interest, is a plasma. A somewhat more precise definition in terms of the shielding distance will be given in Section 14-2. The earliest interest in plasmas was in connection with gaseous electronics (electrical discharges through gases, arcs, and flames); recent interest has been directed toward problems in theoretical astrophysics and the problem of ion containment in thermonuclear (fusion) reactors. In this chapter, we shall look at some of the basic properties of a plasma and develop a few simple ideas for studying plasma applications. A more thorough discussion can be found in books on plasma physics.‡

* This chapter may be omitted without loss of continuity.

† I. Langmuir, *Physical Review*, vol. 33, p. 954 (1929).

‡ See, for example, J. A. Bittencourt, *Fundamentals of Plasma Physics* (Oxford: Pergamon Press, 1986); and F. F. Chen, *Introduction to Plasma Physics*, 2nd ed. (New York: Plenum Press, 1984).

14-1 THEORETICAL APPROACHES TO PLASMA DYNAMICS

The general area of study embracing the interaction of ionized gases with time-dependent electromagnetic fields is called *plasma dynamics*. For many of the problems in this area, and these are the more important and interesting ones, it is impossible to treat a plasma adequately in terms of a purely macroscopic formulation. Instead, it is necessary to use what is known conventionally as kinetic theory. The motions of individual ions and electrons must be studied; their collisions with other particles must be taken into account through solution of the Boltzmann transport equation. Thus a rigorous formulation for plasma problems exists, but their solution is extremely difficult in general, except for situations where it is permissible to neglect some of the terms in the Boltzmann equation. There are, however, three approximate formulations that provide considerable insight into what is happening inside the plasma.

The first of these methods is *equilibrium theory*, which rests on the premise that collisions between charged particles are sufficient to maintain the well-known Maxwell-Boltzmann velocity distribution for particles in the body of the plasma:

$$N_j(\mathbf{v}) dv_x dv_y dv_z = N_{0j} \left(\frac{m_p}{2\pi kT} \right)^{3/2} e^{-m_p v^2 / 2kT} dv_x dv_y dv_z$$

where N_{0j} is the number of particles of type j per unit volume in the plasma, v_x , v_y , and v_z are the components of velocity, m_p is the mass of type j particles, and T is the absolute temperature. Kinetic and transport properties may then be calculated in terms of this velocity distribution.

The second approximate method is *orbit theory*, which treats the motion of charged particles (ions and electrons) in prescribed electric and magnetic fields. These fields may be functions both of position and of time. Orbit theory is a good approximation to particle motion in a plasma when collisions between particles do *not* play the dominant role—that is, when the mean free path for collisions is large compared with characteristic dimensions of the orbit. Under these conditions the effect of collisions can be treated as a perturbation, and the primary problem centers around making the “prescribed” electromagnetic field self-consistent. In other words, the prescribed field must be the sum of the external field and the field produced by the orbiting particles.

The third approximate treatment is the *hydromagnetic formulation*. Here one uses the classical electromagnetic equations (Maxwell’s equations) in conjunction with the classical equations of fluid motion. Evidently, the hydromagnetic treatment is just a macroscopic description of the plasma. It becomes a good approximation when the mean free path for collisions is very small compared with distances of physical interest in the plasma system. The hydromagnetic picture forms a good starting point for

discussing the collective motion of particles in the plasma, for example, plasma oscillations.

The rigorous kinetic theory approach to plasma problems is beyond the scope of this book. On the other hand, many important properties of plasmas can be discussed in terms of the approximations outlined above. For simplicity, we assume that the plasma consists of electrons (charge, $-e$) and singly charged positive ions (charge, $+e$); neutral atoms may be present, but we shall ignore such complications as ionizing collisions and recombination of electrons and ions.

In Section 14-2, we encounter a plasma under stationary or steady-state conditions, for which equilibrium theory is well suited. In Sections 14-3 and 14-4, on the other hand, we shall be much concerned with individual particle motion, and here orbit theory is applicable. Finally, in Sections 14-5 through 14-8 we shall treat some dynamic aspects of the plasma, and we shall do this within the hydromagnetic framework.

14-2

ELECTRICAL NEUTRALITY IN A PLASMA

One of the most important properties of a plasma is its tendency to remain electrically neutral—that is, its tendency to balance positive and negative space charge in each macroscopic volume element. A slight imbalance in the space-charge densities gives rise to strong electrostatic forces that act, wherever possible, in the direction of restoring neutrality. On the other hand, if a plasma is deliberately subjected to an external electric field, the space-charge densities will adjust themselves so that the major part of the plasma is shielded from the field.

Let us consider a rather simple example. Suppose a point charge $+Q$ is introduced into a plasma, thereby subjecting the plasma to an electric field. Electrons find it energetically favorable to move closer to the charge, whereas positive ions tend to move away. Under equilibrium conditions (see Section 5-3), the probability of finding a charged particle in a particular region of potential energy U is proportional to the Boltzmann factor, $\exp(-U/kT)$. Thus the electron density N_e is given by

$$N_e = N_0 \exp\left(e \frac{\varphi - \varphi_0}{kT}\right) \quad (14-1)$$

where φ is the local potential, φ_0 is the reference potential (plasma potential), T is the absolute temperature of the plasma, and k is Boltzmann's constant. N_0 is the electronic density in regions where $\varphi = \varphi_0$.

If N_0 is also the positive ion density in regions of potential φ_0 , then the positive ion density N_i is given by

$$N_i = N_0 \exp\left(-e \frac{\varphi - \varphi_0}{kT}\right) \quad (14-2)$$

The potential φ is obtained from the solution of Poisson's equation:

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d\varphi}{dr} \right) = -\frac{1}{\epsilon_0} (N_i e - N_e e) = \frac{2N_0 e}{\epsilon_0} \sinh \left(e \frac{\varphi - \varphi_0}{kT} \right) \quad (14-3a)$$

This differential equation is nonlinear, and hence must be integrated numerically. On the other hand, an approximate solution to Eq. (14-3a), which is rigorous at high temperature, is adequate for our purposes here. If $kT > e\varphi$, then $\sinh(e\varphi/kT) \approx e\varphi/kT$, and

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d\varphi}{dr} \right) = \frac{2N_0 e^2}{\epsilon_0 kT} (\varphi - \varphi_0) \quad (14-3b)$$

The solution to this equation is

$$\varphi = \varphi_0 + \frac{Q}{4\pi\epsilon_0 r} \exp\left(-\frac{r}{h}\right) \quad (14-4)$$

where r is the distance from the point charge $+Q$, and h , the *Debye shielding distance* (or Debye length), is given by

$$h = \left(\frac{\epsilon_0 kT}{2N_0 e^2} \right)^{1/2} \quad (14-5)$$

Thus the redistribution of electrons and ions in the gas is such as to screen out Q completely in a distance of a few h .

An ionized gas is called a plasma if the Debye length, h , is small compared with other physical dimensions of interest. This restriction is not severe so long as ionization of the gas is appreciable. One finds, for example, that for an ionized gas at $T = 7000$ K and $N_0 = 10^{18}$ electrons or ions/m³, the Debye length is 2.2×10^{-6} m.

14-3 PARTICLE ORBITS AND DRIFT MOTION IN A PLASMA

The orbit of a charged particle q moving in a prescribed electric and magnetic field may be calculated directly from the force equation:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (14-6)$$

We shall find it convenient to start with relatively simple field configurations, and then to generalize to fields which are slowly varying in space.

A constant electric field applied to a plasma is not particularly interesting because the plasma adjusts itself by developing a thin sheath of space charge, which shields the main body of plasma from the field. On the other hand, a constant magnetic field causes the particles to gyrate about the field lines without altering the space-charge distribution.

Case 1: Uniform Magnetic Field ($\mathbf{E} = 0$)

This motion is the same as that described in Problem 8-1, but because it forms the basis for more complicated orbital motion in plasmas, we discuss it here in some detail. It should be emphasized, however, that Case 1 is applicable to many other situations besides plasmas; for example, it is fundamental to the operation of particle accelerators, such as the cyclotron and betatron.

The Lorentz force is always at right angles to the velocity \mathbf{v} of the charged particle; hence its kinetic energy remains constant:

$$K = \frac{1}{2}m_p v^2 = \text{constant} \quad (14-7)$$

where m_p is the mass of the particle. It is convenient to resolve the velocity \mathbf{v} into two components: \mathbf{v}_{\parallel} , parallel to \mathbf{B} , and \mathbf{v}_{\perp} , in the plane perpendicular to \mathbf{B} . Since \mathbf{v}_{\parallel} is unaffected by the field, $K_{\parallel} = \frac{1}{2}m_p v_{\parallel}^2$ remains constant also. It follows that

$$K_{\perp} = \frac{1}{2}m_p v_{\perp}^2 = K - K_{\parallel} \quad (14-8)$$

is also a constant of the motion.

The Lorentz force provides a centripetal acceleration. Thus

$$qv_{\perp}B = \frac{m_p v_{\perp}^2}{R}$$

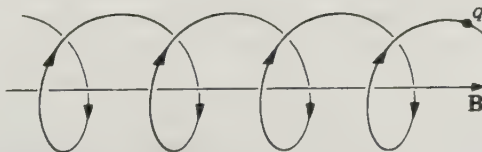
and R (the radius of the orbit) is given by

$$R = \frac{m_p v_{\perp}}{qB} \quad (14-9)$$

The radius R is frequently called the *Larmor radius* of the particle. The complete motion of the charged particle is described as a gyration of the particle in an orbit (the Larmor orbit) superimposed on the uniform motion of the orbit center, or *guiding center*, along a magnetic field line. The resulting helical motion is shown in Fig. 14-1.

The magnetic field acts to confine the plasma by bending the particles in circular orbits. Of course, no confinement is observed in the field direction. For ions and electrons of the same kinetic energy K_{\perp} , the electrons gyrate in much smaller orbits, the ratio of the two Larmor radii being equal to the square root of the mass ratio.

FIGURE 14-1
Particle motion in a uniform magnetic field



An interesting quantity, which we shall have occasion to use later, is the *magnetic moment* of the gyrating particle. By definition, the magnetic moment m is given by current \times area:

$$m = \frac{qv_{\perp}}{2\pi R} \pi R^2 = \frac{K_{\perp}}{B} \tag{14-10}$$

Inspection of Fig. 14-1 shows that m is directed opposite to the magnetic field and is thus a diamagnetic moment.

Case 2: Uniform Electric and Magnetic Fields ($\mathbf{E} \perp \mathbf{B}$)

If an electric and a magnetic field are simultaneously applied to a plasma, and \mathbf{E} is perpendicular to \mathbf{B} , then there is no tendency to produce a sheath. In fact, we shall see that positive and negative space charge drift together in the same direction. For convenience, let the particle velocity \mathbf{v} be written as

$$\mathbf{v} = \mathbf{u} + \mathbf{v}' \tag{14-11}$$

Then Eq. (14-6) may be written as

$$\mathbf{F} = q(\mathbf{E} + \mathbf{u} \times \mathbf{B} + \mathbf{v}' \times \mathbf{B}) \tag{14-12}$$

A particular choice for \mathbf{u} causes the first two terms on the right of this equation to cancel each other:

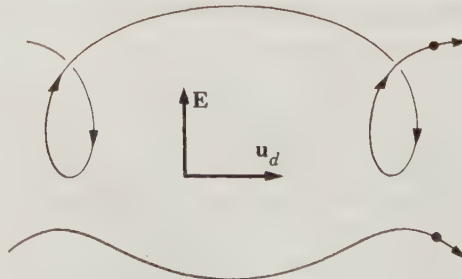
$$\mathbf{u} = \frac{\mathbf{E} \times \mathbf{B}}{B^2} \equiv \mathbf{u}_d \tag{14-13}$$

The remaining force, $q\mathbf{v}' \times \mathbf{B}$, is just what was studied under Case 1.

The total motion of the particle is thus made up of three terms: (a) constant velocity \mathbf{v}'_{\parallel} parallel to \mathbf{B} , (b) gyration about the magnetic field lines with angular frequency $v'_{\perp}/R = qB/m_p$, and (c) a constant drift velocity $u_d = E/B$ at right angles to both \mathbf{E} and \mathbf{B} . Some examples of this motion are shown in Fig. 14-2.

The velocity \mathbf{u}_d defined by Eq. (14-13) is called the *plasma drift velocity* or the *electric drift velocity*. It is important to note that \mathbf{u}_d does not depend on

FIGURE 14-2
Crossed electric and magnetic fields, with particle motion in plane perpendicular to the magnetic field. The figure shows the motion of ions of different initial momenta.



the charge, mass, or velocity of the particle. Thus all components of the plasma drift along together even though their individual gyrations may be vastly different.

Our derivation of Eq. (14-13) was obtained in a nonrelativistic fashion: If either \mathbf{u}_d or \mathbf{v} should approach c (the speed of light), then Eq. (14-11) must be replaced by an expression consistent with a Lorentz transformation. On the other hand, it turns out that Eq. (14-13) for the drift velocity is always correct* so long as $|\mathbf{E}| < c|\mathbf{B}|$. If $|\mathbf{E}| > c|\mathbf{B}|$, the magnetic field cannot prevent the particle from moving in the direction of \mathbf{E} .

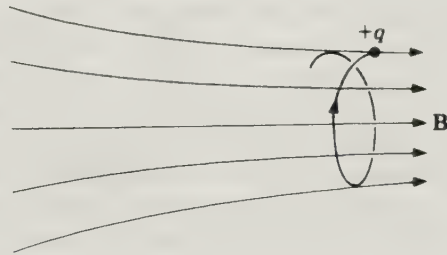
Case 3: Magnetic Field Constant in Time, but Space-Dependent ($\mathbf{E} = 0$)

Let us suppose that a charged particle is moving in a nearly uniform magnetic field, one in which the field lines are slowly converging in space. The particle motion may be treated as a perturbation of the helical orbit in Fig. 14-1.

The motion will be something like that shown in Fig. 14-3. We may easily verify that there is a force tending to push the particle into the weaker magnetic field region. To specify the problem precisely it will be assumed that the flux line through the guiding center coincides with the z -axis and that the magnetic field has azimuthal symmetry about the z -axis. Taking the z -component of Eq. (14-6), we obtain

$$F_z = m_p \frac{dv_z}{dt} = qv_\theta B_r|_{r=R} \quad (14-14)$$

FIGURE 14-3
Particle motion in a region in which the magnetic field is space-dependent



* The simplest way to treat the case where $|\mathbf{E}|$ is less than but not small compared with $c|\mathbf{B}|$ is to make a Lorentz transformation, transforming both the particle velocity and the fields. The velocity of the moving system is given by \mathbf{u}_d (Eq. 14-13), and the force in the moving system is given by

$$\mathbf{F}' = q(\mathbf{v}' \times \mathbf{B}) \left(\frac{c^2 - u_d^2}{c^2} \right)^{1/2}$$

But $\nabla \cdot \mathbf{B} = 0$ or, for the case in point,

$$\frac{1}{r} \frac{\partial}{\partial r} (rB_r) + \frac{\partial B_z}{\partial z} = 0$$

Since the field lines are converging slowly, $\partial B_z / \partial z$ may be taken constant over the orbit cross section, yielding

$$B_r|_{r=R} = -\frac{1}{2} R \frac{\partial B_z}{\partial z} \quad (14-15)$$

Furthermore, v_θ is analogous to the v_\perp of Case 1. Making these substitutions in Eq. (14-14) gives

$$m_p \frac{dv_\parallel}{dt} = -\frac{1}{2} q R v_\perp \frac{\partial B_z}{\partial z} = -m \frac{\partial B_z}{\partial z} \quad (14-16)$$

the last form being obtained through the use of Eq. (14-10).

The total kinetic energy K of the particle is unaltered in the magnetic field, since the Lorentz force, which is always at right angles to the velocity, can do no work. K_\perp , defined in (14-8), is not constant here; neither is K_\parallel , but we may write

$$\frac{d}{dt} \left(\frac{1}{2} m_p v_\parallel^2 \right) = \frac{d}{dt} (K - K_\perp) = -\frac{dK_\perp}{dt} = -\frac{d}{dt} (mB_z) \quad (14-17)$$

the last form coming from Eq. (14-10). On the other hand, we may multiply Eq. (14-16) by $v_\parallel = \partial z / \partial t$ to obtain

$$\frac{d}{dt} \left(\frac{1}{2} m_p v_\parallel^2 \right) = -m \frac{\partial B_z}{\partial z} \frac{\partial z}{\partial t} = -m \frac{dB_z}{dt} \quad (14-18)$$

where d/dt represents the time derivative taken along the dynamical path. By comparing Eqs. (14-17) and (14-18), we see that the magnetic moment m is a constant of the motion. It should be emphasized, however, that this is an approximate result that holds only so long as B_z varies slowly. If \mathbf{B} were to change substantially in distances of the order of R , the approximations used in the derivation of Eq. (14-18) would break down.

Of further interest is the fact that the particle is constrained to move on the surface of a flux tube. This follows because the magnetic flux through the orbit is

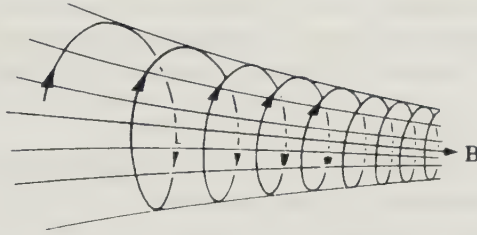
$$\Phi = B_z \pi R^2 = \pi B_z \frac{m_p^2 v_\perp^2}{q^2 B_z^2} = \frac{2\pi m_p}{q^2} \frac{K_\perp}{B_z} = \frac{2\pi m_p}{q^2} m \quad (14-19)$$

and m is constant. The motion of the particle is depicted schematically in Fig. 14-4.

The z -component (parallel component) of the force, Eq. (14-16), is always in such a direction as to accelerate particles toward the weaker part of the field. Gyrating particles that are approaching regions of stronger

FIGURE 14-4

The particle winds in a tighter and faster helix until it is reflected.



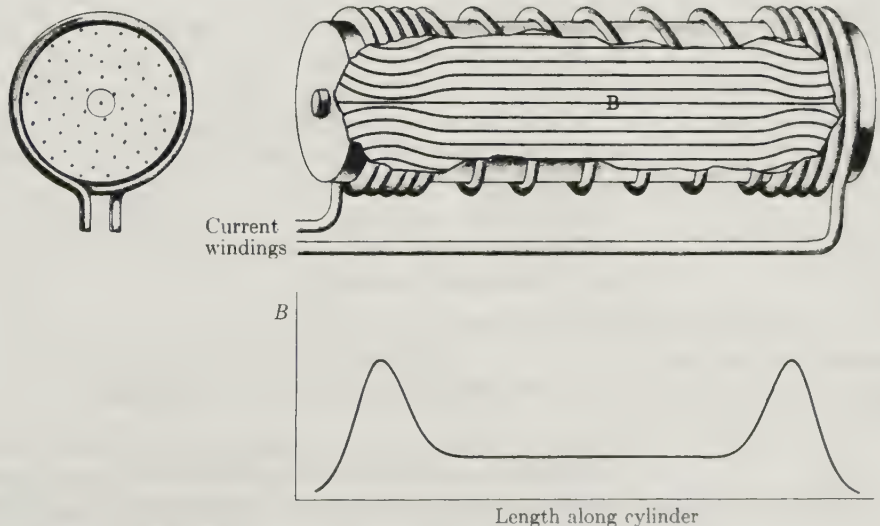
magnetic field are thus slowed down—that is, v_{\parallel} is decreased. On the other hand, conservation of energy requires that simultaneously the orbital motion v_{\perp} be speeded up. If the convergence of the magnetic field is sufficient, the particle will gyrate in an ever-tighter helical spiral until it is finally reflected back into the weaker field.

14-4 MAGNETIC MIRRORS

The results of the preceding section show that a slowly converging magnetic field can, in principle, confine a plasma. At right angles to the principal field direction the particles are bent into circular orbits; along the principal direction of the field the particles are slowed down and finally reflected by the converging field lines. Such a field configuration is called a *magnetic mirror*. At least two mirrors must be used in any confinement system; a system of this type is shown in Fig. 14-5.

FIGURE 14-5

Magnetic mirror system



Not all particles can be confined by the mirror system, however. The field lines cannot be made to converge to a point; thus there is a large but not infinite magnetic field B_m at the mirror. If the particle has too much "axial kinetic energy," it will not be turned back by the mirror field, and it will be able to escape.

Because the magnetic moment is a constant of the motion, we find, according to Eq. (14-10), that

$$\frac{K_{0\perp}}{B_0} = \frac{K_{1\perp}}{B_1}$$

Here the subscript 0 refers to the central region of Fig. 14-5, and the subscript 1 to the reflection point. At the reflection point, however, $K_{\perp} = K$. Furthermore, K , the total kinetic energy, is a constant of the motion. In order that the particle be reflected, the mirror field B_m must be greater than B_1 , that is,

$$B_m > B_1 = \frac{K}{K_{0\perp}} B_0$$

or

$$\frac{K_{0\perp}}{K} > \frac{B_0}{B_m} \quad (14-20a)$$

If the initial velocity v_0 makes an angle θ_0 with the field direction, then $v_{0\parallel} = v_0 \cos \theta_0$ and $v_{0\perp} = v_0 \sin \theta_0$. Equation (14-20a) then reduces to

$$\sin^2 \theta_0 > \frac{B_0}{B_m} \quad (14-20b)$$

as the criterion for reflection. For example, if the mirror field is one hundred times as intense as B_0 , then particles with velocities making an angle of less than 6° with the field direction escape from the system.

The collisions between particles in the central region of the mirror system tend to produce an isotropic velocity distribution. Thus the net result of collisions is that particles are continually scattered into a region of velocity space such that they can escape from the system. As a result of collisions particles can also "diffuse" at right angles to the field direction, and so eventually escape.

14-5

THE HYDROMAGNETIC EQUATIONS

Collective motions of the particles in a plasma, such as the "pinch effect" and plasma oscillations, are handled best in the hydromagnetic formulation. According to this description, the plasma is regarded as a classical fluid that obeys the conventional equations of hydrodynamics. The fluid, however, is

an electrical conductor, and thus electromagnetic forces must be taken into account explicitly.

The force on a unit volume of the plasma may be written as

$$\mathbf{F}_v = \mathbf{J} \times \mathbf{B} - \nabla p \quad (14-21)$$

where \mathbf{J} is the current density and p is the fluid pressure. Other forces, such as gravitational and viscous forces, may also be included, but are neglected here in the interest of simplicity. Because of the approximate electrical neutrality of the plasma, the term $\rho \mathbf{E}$ need not be included along with other force terms in Eq. (14-21). Deviations from neutrality must be considered, of course, in Poisson's equation, but they are usually ignored in the dynamical equations.

Momentum balance requires that

$$\zeta \frac{d\mathbf{v}}{dt} = \zeta \left[\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right] = \mathbf{F}_v = \mathbf{J} \times \mathbf{B} - \nabla p \quad (14-22)$$

which is the equation of motion, or the *Euler equation*, of the fluid. Here ζ is the mass density of the plasma and \mathbf{v} its fluid velocity. For problems in which the hydrodynamic motion is not particularly large, the term containing $(\mathbf{v} \cdot \nabla) \mathbf{v}$ can usually be neglected.*

It is sometimes convenient to interpret the $\mathbf{J} \times \mathbf{B}$ term of Eq. (14-21) as arising in part from a "magnetic pressure." This interpretation can be done with the aid of Ampere's circuital law, Eq. (9-29), which, specialized to the plasma case, is

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} \quad (14-23)$$

and the vector identity

$$\mathbf{B} \times \nabla \times \mathbf{B} = \nabla \left(\frac{1}{2} B^2 \right) - (\mathbf{B} \cdot \nabla) \mathbf{B} \quad (14-24)$$

Thus

$$\mathbf{J} \times \mathbf{B} = -\frac{1}{\mu_0} \mathbf{B} \times \nabla \times \mathbf{B} = -\nabla \left(\frac{B^2}{2\mu_0} \right) + \frac{1}{\mu_0} (\mathbf{B} \cdot \nabla) \mathbf{B} \quad (14-25)$$

and

$$\mathbf{F}_v = -\nabla p - \nabla \left(\frac{B^2}{2\mu_0} \right) + \frac{1}{\mu_0} (\mathbf{B} \cdot \nabla) \mathbf{B} \quad (14-21a)$$

The quantity $B^2/2\mu_0$, which is, of course, the magnetic energy density, also plays the role of a magnetic pressure, p_m :

$$p_m = \frac{B^2}{2\mu_0} \quad (14-26)$$

* Although it may not be neglected in steady-flow problems for which the term $\partial \mathbf{v} / \partial t$ vanishes explicitly.

It should be emphasized, however, that $-\nabla p_m$ gives in most cases only part of the magnetic force. The remaining force comes from the $(1/\mu_0)(\mathbf{B} \cdot \nabla)\mathbf{B}$ term. When $\mathbf{J} = 0$, the two terms on the right of Eq. (14-25) cancel each other.

As an example of the utility of the magnetic pressure concept, consider a *unidirectional* magnetic field. The equation $\nabla \cdot \mathbf{B} = 0$ guarantees that \mathbf{B} does not change along the field direction. Since space variations can occur only in directions at right angles to \mathbf{B} , it follows that $(\mathbf{B} \cdot \nabla)\mathbf{B} = 0$ for this case. Equation (14-21) reduces, therefore, to

$$\mathbf{F}_v = -\nabla(p + p_m)$$

and the condition for static equilibrium of each volume element is

$$p + p_m = \text{constant}$$

In other words, for this example the sum of the fluid pressure and the magnetic pressure must be space-independent.

In addition to Eq. (14-22) and the macroscopic equations governing the electric and magnetic fields,* we require two additional relationships to complete the hydromagnetic formulation. These are (1) the equation of continuity for the plasma fluid:

$$\frac{\partial \xi}{\partial t} + \nabla \cdot (\xi \mathbf{v}) = 0 \quad (14-27)$$

and (2) an equation relating \mathbf{J} to the field quantities. The latter relationship is simply a generalized form of Ohm's law that, under certain conditions, may be written as

$$\mathbf{J} = g(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (14-28a)$$

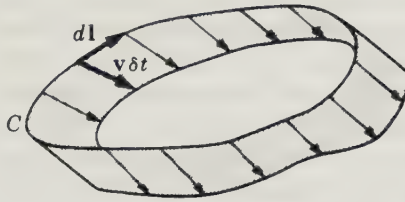
Here $\mathbf{v} \times \mathbf{B}$ is the "motional electric field" arising from hydrodynamic motion of the plasma in a magnetic field, and g is the conductivity of the plasma.

An approximation that is frequently made is that of infinite conductivity. The advantage of this approximation is that it permits a substantial simplification of the hydromagnetic equations, thus presenting a much clearer picture of the physical processes going on in the plasma. In some problems, particularly astrophysical ones, the approximation is quite good. For the case of infinite conductivity, Ohm's law reduces to

$$g \rightarrow \infty \\ \mathbf{E} + \mathbf{v} \times \mathbf{B} = 0 \quad (14-28b)$$

* The Maxwell equations are summarized in Section 16-2. The reader will note that in Eq. (16-10), the original Ampere's circuital law has been modified through inclusion of the displacement current, $\partial \mathbf{D} / \partial t$. Actually, the displacement current does not play an important role in *most* hydromagnetic phenomena.

FIGURE 14-6
Geometric interpretation
of $\oint_C d\mathbf{l} \times \mathbf{v}$



Infinite conductivity (or, for practical purposes, high conductivity) has an important consequence; namely, that the magnetic flux is frozen into the plasma. If Eq. (14-28b) is combined with the differential form of Faraday's law of induction, we obtain

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) \quad (14-29)$$

The normal component of this equation integrated over a fixed surface S yields

$$\frac{d}{dt} \int_S \mathbf{B} \cdot \mathbf{n} \, da = \int_S \nabla \times (\mathbf{v} \times \mathbf{B}) \cdot \mathbf{n} \, da$$

or

$$\frac{d\Phi}{dt} = \oint_C \mathbf{v} \times \mathbf{B} \cdot d\mathbf{l} = \oint_C \mathbf{B} \cdot (d\mathbf{l} \times \mathbf{v}) \quad (14-30)$$

where C is a fixed contour in space through which the plasma moves due to hydrodynamic motion. From Fig. 14-6, we see that $\oint_C d\mathbf{l} \times \mathbf{v}$ may be regarded as the increase in area, per unit time, of the cap surface that is bounded by C , and $\oint_C \mathbf{B} \cdot d\mathbf{l} \times \mathbf{v}$ is the magnetic flux associated with this increased area. Equation (14-30) simply states that the flux change per unit time through the contour C is just what we should calculate geometrically on the basis that all flux lines move along with the fluid. We conclude, therefore, that the lines of magnetic induction are "frozen" into the perfectly conducting material.

14-6 THE PINCH EFFECT

The tendency of a high-current discharge through a plasma to constrict itself laterally is known as the "pinch effect." The basic mechanism causing the pinch is the interaction of a current with its own magnetic field or, equivalently, the attraction between parallel current filaments. The pinch

effect was first predicted by Bennett, and later independently by Tonks.* A somewhat different picture of the pinch, showing its inherent instability, was given by Rosenbluth.†

Let us consider a current discharge of cylindrical symmetry through the plasma. The current in the discharge is axial, and its magnetic field is azimuthal. From Ampere's circuital law, the magnetic induction at distance r from the axis of the discharge is given by

$$B(r) = \frac{\mu_0}{r} \int_0^r J(r') r' dr' \quad (14-31)$$

From this it follows that

$$\frac{\partial B}{\partial r} = -\frac{\mu_0}{r^2} \int_0^r J(r') r' dr' + \mu_0 J(r) = -\frac{1}{r} B(r) + \mu_0 J(r) \quad (14-32)$$

The magnetic force per unit volume is

$$\mathbf{F}_v = \mathbf{J} \times \mathbf{B} = -J(r)B(r)\mathbf{a}_r \quad (14-33)$$

where \mathbf{a}_r is a unit vector in the r -direction. Eliminating $J(r)$ between Eqs. (14-32) and (14-33) yields

$$F_v = -\frac{1}{\mu_0} B \frac{\partial B}{\partial r} - \frac{1}{\mu_0 r} B^2 \quad (14-34)$$

This force can be converted to an equivalent pressure, p_{eq} , by writing $F_v = -\partial p_{\text{eq}}/\partial r$, and then integrating:

$$p_{\text{eq}} = \frac{1}{2\mu_0} B^2 + \frac{1}{\mu_0} \int_0^r \frac{B^2}{r} dr \quad (14-35)$$

We are particularly interested in the pressure on the lateral boundaries of the discharge. Following Rosenbluth, we restrict our attention to the high-conductivity case where the magnetic field lines cannot penetrate appreciably into the conducting fluid.‡ Here the integral in Eq. (14-35) contains no contribution from the discharge region. At the boundary of the discharge, $r = R$, and the pressure is just what we have called the *magnetic pressure*, p_m :

$$p_m = \frac{1}{2\mu_0} B^2(R) \quad (14-36)$$

* W. Bennett, *Physical Review*, vol. 45, p. 890 (1934); L. Tonks, *Physical Review*, vol. 56, p. 369 (1939).

† M. Rosenbluth, "Dynamics of a Pinched Gas," from *Magnetohydrodynamics*, edited by Rolf Landshoff (Stanford: Stanford University Press, 1957).

‡ The nonpenetration of the field lines follows from the results of the preceding section and the fact that both the current and magnetic field are initially very small in the discharge.

It is evident from Eq. (14-35) that the magnetic pressure is uniform in the outside region, but zero or very small inside the discharge. Thus the pinch effect can be viewed as coming about from the sudden buildup of magnetic pressure in the region external to the discharge.

The pinching of the discharge results in plasma compression. If the pinch could contract in a stable manner, it would proceed until the magnetic pressure in the external region was equal to the fluid pressure in the discharge. Let us treat the plasma as a perfect gas, whose fluid pressure $p = NkT$. Then, at the final radius R of the discharge,

$$\frac{1}{2\mu_0} B^2(R) = \frac{1}{2} \frac{\mu_0}{4\pi^2 R^2} I^2 = NkT$$

where I is the current in the discharge. This expression may be solved for the current:

$$I^2 = 2 \left(\frac{\mu_0}{4\pi} \right)^{-1} \pi R^2 N k T = 2 \left(\frac{\mu_0}{4\pi} \right)^{-1} A_0 N_0 k T,$$

since conservation of particles requires that $A_0 N_0 = \pi R^2 N$. Here A_0 is the initial cross section of the discharge, N_0 is its initial particle density, $\mu_0/4\pi = 10^{-7} \text{ T} \cdot \text{m/A}$, and Boltzmann's constant $k = 1.38 \times 10^{-23} \text{ J/K}$. In order to achieve the temperature of 10^8 K required for a thermonuclear (fusion) reactor, with $A_0 = 0.04 \text{ m}^2$ and $N_0 = 10^{21} \text{ particles/m}^3$, a pinch current of approximately one million amperes is required.

It is easy to see that the pinch is an inherently unstable phenomenon. The magnetic pressure on the boundary of the discharge depends on its radius as well as on its detailed geometry. Small perturbations will grow if the pressure changes that result are such as to enhance these perturbations. Figure 14-7 shows that small ripples on the bounding surface of the discharge, as well as kinks, fall into this category, producing the so-called sausage and kink instabilities of the pinched plasma.

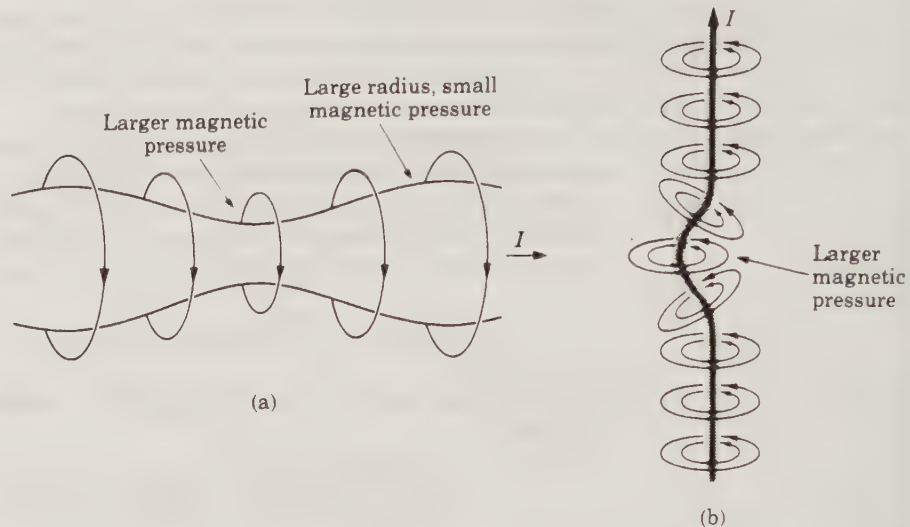
MAGNETIC CONFINEMENT SYSTEMS FOR CONTROLLED THERMONUCLEAR FUSION

Much of the current interest in plasma physics is motivated by our need to develop alternative sources of energy and the possibility of using a thermonuclear plasma of deuterium and tritium for an energy source. Both the pinch effect discussed in the preceding section and the magnetic mirror concept (Section 14-4) are being used in experimental fusion reactor systems in attempts to confine the thermonuclear plasma magnetically. Although the pinch effect is inherently unstable, its stability can be improved by supplying additional components to the magnetic field and by optimizing design parameters.

An important quantity for the design of fusion reactor systems is the

FIGURE 14-7

Instabilities in the pinched plasma:
 (a) sausage instability,
 (b) kink instability



ratio of kinetic pressure $p = NkT$ to the total pressure (kinetic plus magnetic p_m). This ratio is given the symbol β :

$$\beta = \frac{p}{[B^2/2\mu_0 + p]} = \frac{NkT}{[B^2/2\mu_0 + NkT]} \quad (14-37)$$

where N is the sum of the ion and electron densities in the plasma. Fusion reactors are generally characterized by their β -value. Low β refers to values less than 0.01 and high β values lie between 0.1 and 1.0. The plasma is either deuterium or a deuterium-tritium plasma with temperature in excess of 10^8 K; its density is in the range 10^{19} m^{-3} to 10^{22} m^{-3} . Confinement does not have to be absolute, but must be for a long enough period τ so that more energy is produced in the thermonuclear reaction than is consumed in establishing the plasma conditions. Confinement is believed to be adequate when the *Lawson condition* is met:

$$N_i \tau > 10^{20} \text{ m}^{-3} \text{ s} \quad \text{at} \quad T > 10^8 \text{ K} \quad (14-38)$$

where N_i is the ion density in the plasma.

The most active area of investigation relating to magnetic confinement involves a class of reactors designed to confine toroidal, or doughnut-shaped, plasmas. And within this class, the *tokamaks* show considerable promise. The name derives from a series of Tokamak experiments carried out at the I.V. Kurchatov Institute of Atomic Energy in Moscow, but it now applies to a generic class of experimental reactors that are characterized by a *toroidal, diffuse pinch*. The tokamak is a low to moderate β device. The

magnetic field in the plasma has a *poloidal** component due to current flowing in the plasma in response to a toroidal electric field, and a toroidal component produced by external coils, so that the resultant field traces out helical curves within the doughnut-shaped plasma. There are a number of tokamak devices in operation around the world: Several of the more advanced-design tokamaks operating in the mid-1980s were the Joint European Torus (JET), the Tokamak Fusion Test Reactor (TFTR) in Princeton, Alcator C at Massachusetts Institute of Technology, and GA Technologies' D III-D. These devices operated with a maximum plasma current of several million amperes, had magnetic fields of the order of 2 tesla, and had β 's in the range of 0.03 to 0.06, although techniques for achieving stable operation at higher β have since been developed. Some devices operated within a factor of ten of the Lawson break-even condition: for example, TFTR operated at $N_i\tau = 10^{19} \text{ m}^{-3} \text{ s}$, $T = 2 \times 10^8 \text{ K}$, and another device operated at $N_i\tau = 2 \times 10^{20} \text{ m}^{-3} \text{ s}$, $T = 2 \times 10^7 \text{ K}$.

The next stage in tokamak development is the Compact Ignition Tokamak (CIT) originally planned for operation in 1990, but delayed due to funding problems. Still contemplated, this device is to have a toroidal magnetic field of 10 tesla and a maximum plasma current of 9 million amperes. Its goal is a $N_i\tau$ of $2 \times 10^{20} \text{ m}^{-3} \text{ s}$ and a temperature of $2 \times 10^8 \text{ K}$ and the achievement of an actual ignition of a deuterium-tritium plasma—that is, a self-sustaining thermonuclear burn that continues for a while after all external heating is turned off. The CIT device is to be used to study the physics of plasma heating by alpha particles produced in the primary fusion reaction.

Examples of experimental devices with high β are the theta-pinch fusion reactor and the magnetic mirror reactor. The theta-pinch concept uses a toroidal-shaped plasma, but of substantially higher density than that of a tokamak. The magnetic mirror is not a true confinement system in that plasma leaks out the ends; however, it can amplify the power of an injected beam. The magnetic mirror is a high β device but uses a lower density plasma than the theta-pinch reactor.

For a more detailed discussion of experimental fusion reactor systems the reader is referred to one of the following reviews:

“Fusion Reactor Systems” by F. L. Ribe, in *Reviews of Modern Physics*, Vol. 47, p. 7 (1975).

“Fusion Energy in Context: Its Fitness for the Long Term,” by J. P. Holdren, in *Science*, vol. 200, p. 168 (April 1978).

“Progress in Toroidal Confinement and Fusion Research” by H. F.

* If the toroidal plasma were cut (for example, at the slot marked *d* in Fig. 9-15) and straightened into a cylinder, the poloidal component would become the azimuthal component and the toroidal component would become the axial component.

Furth in *From Particles to Plasmas*, edited by J. W. Van Dam (Reading: Addison-Wesley Publishing Co., 1989).

Fusion: The Search for Endless Energy by Robin Herman (Cambridge: Cambridge University Press, 1990).

14-8

PLASMA OSCILLATIONS AND WAVE MOTION

One of the interesting properties of a plasma is its ability to sustain oscillations and propagate waves. Various types of oscillatory behavior are possible, and because of the nonlinear character of the hydrodynamic equations, these oscillations can be quite complex. We find it expedient to restrict our attention to some rather simple cases, which, nevertheless, have been observed in controlled experiments.

Case 1: Electrostatic Plasma-Electron Oscillations

Electrostatic oscillations in a plasma were first discussed by Tonks and Langmuir.* Actually, there are two possible types of electrostatic oscillations: high-frequency oscillations that are too rapid for the heavy ions to follow, and oscillations of the ions that are so slow that the electrons are always distributed around the ions in a statistical manner. We discuss the first case only, the so-called electron oscillations.

Let us fix our attention on a region of plasma containing a uniform density of positive ions, N . There are no negative ions. Initially, the electrons also have uniform density N , but let us suppose that each electron is displaced in the x -direction by a distance ξ that is independent of the y - and z -coordinates and is zero on the plasma boundaries. The displacement of electrons disturbs the neutral plasma, producing a charge in each volume element $\Delta x \Delta y \Delta z$:

$$\begin{aligned} \delta\rho \Delta x \Delta y \Delta z &= -Ne \Delta y \Delta z \left[\xi - \left(\xi + \frac{\partial \xi}{\partial x} \Delta x \right) \right] \\ &= \Delta x \Delta y \Delta z Ne \frac{\partial \xi}{\partial x} \end{aligned}$$

The displacement of the electrons produces an electric field $\mathbf{E}(x, t)$ that, because of the symmetry of the problem, is in the x -direction. Thus

$$\nabla \cdot \mathbf{E} = \frac{1}{\epsilon_0} \delta\rho$$

or

$$\frac{\partial E}{\partial x} = \frac{1}{\epsilon_0} Ne \frac{\partial \xi}{\partial x}$$

* L. Tonks and I. Langmuir, *Physical Review*, vol. 33, p. 195 (1929).

which, when integrated, yields

$$E = \frac{Ne}{\epsilon_0} \xi \quad (14-39)$$

Here, the constant of integration has been taken equal to zero, since sheath formation will shield the plasma from a uniform electric field.

The force on each electron is $-eE$, which, according to Eq. (14-39), is proportional to the displacement ξ . It is also seen to be a restoring force. Thus each electron oscillates about its original position with simple harmonic motion. The equation of motion for each electron is

$$m_e \frac{d^2 \xi}{dt^2} + \frac{Ne^2}{\epsilon_0} \xi = 0 \quad (14-40)$$

The "plasma frequency," $f_p = \omega_p/2\pi$, is defined, therefore, by

$$\omega_p = \left(\frac{Ne^2}{m_e \epsilon_0} \right)^{1/2} \quad (14-41)$$

where m_e is the electron mass.

As a numerical example, we have $f_p = 9.0 \times 10^9 \text{ s}^{-1}$ for a particle density $N = 10^{18} \text{ electrons/m}^3$.

Case 2: Hydromagnetic or Alfvén Waves

Hydromagnetic waves represent true wave propagation in a conducting medium that is subjected to a constant magnetic field. This behavior, which was first predicted by Alfvén* in 1942, is consistent with the hydromagnetic formulation of a plasma discussed in Section 14-5.

Before proceeding to the differential equations, let us look at the physical processes in the plasma from as elementary a viewpoint as possible. Consider an infinite plasma subjected to a constant, uniform magnetic field \mathbf{B}_0 that is directed along the z -axis. If a segment of the plasma, the rectangular section $ABCD$ in Fig. 14-8 that extends parallel to the y -axis, is given a velocity \mathbf{v} directed parallel to the positive y -axis, then the charge carriers (ions and electrons) experience forces

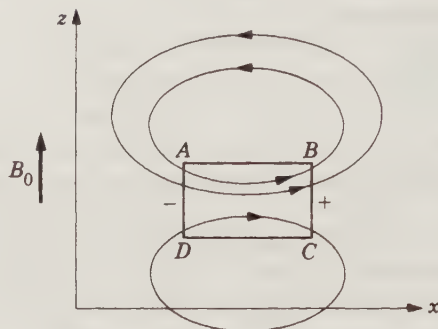
$$q_i(\mathbf{v} \times \mathbf{B}_0)$$

that tend to separate the positive and negative carriers. The segment $ABCD$ thus becomes a seat of motional emf, its right-hand end tending to charge up positively, its left end negatively. But since we are dealing with a

* H. Alfvén, *Cosmical Electrodynamics*, 2nd ed. (New York: Oxford University Press, 1963).

FIGURE 14-8

The segment, $ABCD$, of plasma moves in the positive y -direction. The currents that are generated are depicted schematically.



conducting medium, the plasma external to $ABCD$ completes the electrical circuit. A few of the current lines are shown in the figure.

The induced current now interacts with the magnetic field \mathbf{B}_0 . It is easy to verify that the force density $\mathbf{J} \times \mathbf{B}_0$ in the segment $ABCD$ is such as to oppose its motion, whereas the force on external parts of the plasma is such as to accelerate it in the positive y -direction. Eventually, $ABCD$ will have slowed down, and its motion will have been transferred to neighboring segments of the plasma. The mechanism is still operating, however, and the whole process is repeated, thus propagating the disturbance farther in the $\pm z$ -direction.

We now turn to the differential equations. Let $\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_1$, where \mathbf{B}_0 is the constant, uniform field parallel to the z -axis and \mathbf{B}_1 is the magnetic field set up by the induced currents. Using the results of the preceding paragraphs as a guide, we look for the simplest type of wave motion, characterized by v_y , E_x , J_x , and B_{1y} , other components vanishing. From Ampere's law,

$$-\frac{\partial B_{1y}}{\partial z} = \mu_0 J_x \quad (14-42)$$

and the Euler equation of the fluid, Eq. (14-22), gives the two relations

$$\zeta \frac{\partial v_y}{\partial t} = -J_x B_0 \quad (14-43a)$$

and

$$0 = J_x B_{1y} - \frac{\partial p}{\partial z} \quad (14-43b)$$

Equations (14-43) may be combined with (14-42) to yield

$$\frac{\partial v_y}{\partial t} = \frac{B_0}{\mu_0 \zeta} \frac{\partial B_{1y}}{\partial z} \quad (14-44)$$

and

$$\frac{\partial p}{\partial z} = -\frac{1}{2\mu_0} \frac{\partial(B_{1y}^2)}{\partial z} \quad (14-45)$$

The generalized Ohm's law may be written as

$$E_x = -v_y B_0 + \frac{1}{g} J_x = -v_y B_0 - \frac{1}{g\mu_0} \frac{\partial B_{1y}}{\partial z} \quad (14-46)$$

Finally, Faraday's law yields

$$\frac{\partial B_{1y}}{\partial t} = -\frac{\partial E_x}{\partial z} \quad (14-47)$$

If v_y is eliminated between Eqs. (14-44) and (14-46) and E_x is eliminated between the resulting equation and Eq. (14-47), we obtain, on the assumption of constant ζ ,

$$\frac{\partial^2 B_{1y}}{\partial t^2} = \frac{B_0^2}{\mu_0 \zeta} \frac{\partial^2 B_{1y}}{\partial z^2} + \frac{1}{g\mu_0} \frac{\partial^3 B_{1y}}{\partial z^2 \partial t} \quad (14-48)$$

which is the equation governing the propagation of Alfvén waves.

If the conductivity g of the plasma were infinite, then Eq. (14-48) would become identical with the wave equation whose solution is discussed in Sections 16-4 and 16-5. In these circumstances, Eq. (14-48) describes a plane, undamped wave moving parallel to the z -axis with phase velocity

$$v_p = \frac{B_0}{\sqrt{\mu_0 \zeta}} \equiv v_A \quad (14-49)$$

As a numerical example, take $B_0 = 0.01$ T, $\zeta = 10^{-5}$ kg/m³ = 10^{-8} g/cm³; then $v_p = 2800$ m/s.

In order to see what results for finite conductivity, we try a solution to Eq. (14-48) of the form

$$B_{1y} = b_1 \exp[\alpha z + i\omega t]$$

This solution is satisfactory, provided

$$\alpha^2 = \frac{-\omega^2}{v_p^2 + i\omega/g\mu_0} \quad (14-50)$$

with v_p as defined in Eq. (14-49). For small damping,

$$\alpha \approx \pm \left(i \frac{\omega}{v_p} + \frac{\omega^2}{2g\mu_0 v_p^3} \right) \quad (14-51)$$

Thus the solution to Eq. (14-48) is a damped plane wave propagating in the $\pm z$ -direction. The distance δ in which the amplitude of the wave is reduced

to $1/e$ of its original value is

$$\delta = \frac{2g\mu_0 v_p^3}{\omega^2} = \frac{2gB_0^3}{\mu_0^{1/2} \zeta^{3/2} \omega^2} \quad (14-52)$$

14-9
SUMMARY

Highly ionized gases are good electrical conductors; a **plasma** is a region of highly ionized gas in which the static electric field and net charge density are nearly zero. Three different approaches to the analysis of plasma behavior are the equilibrium kinetic theory, orbit theory, and the macroscopic hydromagnetic theory.

- The **equilibrium theory**, based on the statistical Boltzmann factor, shows that an external charge Q located in a plasma is shielded by the plasma in a distance called the *Debye length*. That is, the unshielded Coulomb potential $Q/4\pi\epsilon_0 r$ is replaced by

$$\varphi = \frac{Q}{4\pi\epsilon_0 r} e^{-r/h}$$

where the Debye length h is

$$h = \sqrt{\frac{\epsilon_0 kT}{2N_0 e^2}}$$

- **Orbit theory** is based on particle motions under the force

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

In a uniform magnetic field, a particle of mass m_p moves freely along the field line but gyrates around it. The orbit is a helix with the **Larmor radius**

$$R = \frac{m_p v_{\perp}}{qB}$$

The free particle has a diamagnetic moment. If the field is nonuniform, the particle gyrates in a tighter spiral as it moves into a stronger field along converging field lines; at the same time its axial motion slows down and is finally reversed. The result is a **magnetic mirror**.

- The **hydromagnetic approach** is based on the macroscopic force law for unit volume

$$\mathbf{F}_v = \mathbf{J} \times \mathbf{B} - \nabla p$$

where p is the fluid pressure. Sometimes the first term can be approximated by $-\nabla p_m$, where the "magnetic pressure" is equal to the energy density,

$$p_m = \frac{B^2}{2\mu_0}$$

The magnetic pinch effect can be treated as a compression of the plasma by the magnetic pressure outside it.

• The macroscopic approach also leads to **plasma waves**. The “electrostatic” waves are oscillations at the **plasma frequency** (for the electrons) of

$$\omega_p = \sqrt{\frac{Ne^2}{m_e \epsilon_0}}$$

with infinite wavelength. The “Alfvén” waves propagate in a uniform magnetic field B_0 with a phase velocity

$$v_A = \frac{B_0}{\sqrt{\mu_0 \zeta}}$$

PROBLEMS

14-1 The condition for orbit theory to be a good approximation to the motion of an electron in a plasma is that $\tau \gg 2\pi m_e / Be$, where τ is the mean collision time (see Chapter 7) and $2\pi m_e / Be$ is the cyclotron period in the magnetic field B . Show that this statement is equivalent to $\eta \ll \eta_H$, where $\eta_H \equiv B / N_0 e$ is the Hall resistivity.

14-2 Consider a steady-flow hydromagnetic problem in which \mathbf{v} , \mathbf{J} , and \mathbf{B} are mutually orthogonal. Assume that \mathbf{v} is in the x -direction and that \mathbf{v} , \mathbf{J} , and \mathbf{B} are functions of x only. Assume also that the channel cross section (perpendicular to x) is independent of x . Show that

$$v = v_0 - \frac{1}{2\zeta_0 v_0} \left[2B_0 \int J dx + \mu_0 \left(\int J dx \right)^2 \right]$$

where v_0 is the velocity when $\zeta = \zeta_0$, $B = B_0$.

14-3 Calculate the Debye shielding distance and the plasma frequency ω_p for (a) the sun's chromosphere ($T = 5000$ K, electron density = $3 \times 10^{19}/\text{m}^3$); (b) the ionosphere ($T = 300$ K, electron density = $10^{12}/\text{m}^3$); and (c) an arc discharge ($T = 4000$ K, electron density = $10^{22}/\text{m}^3$).

14-4 The electrical resistivity of a tenuous gaseous plasma in which the current is carried predominantly by electrons is given by [see Lyman Spitzer, *Physics of Fully Ionized Gases* (New York: Interscience, 1956)]:

$$\eta = 6.53 \times 10^5 Z \frac{\ln \Lambda}{T^{3/2}} \text{ ohm} \cdot \text{m}$$

with $\Lambda = 1.28 \times 10^7 T^{3/2} / (ZN_e^{1/2})$. Here T = temperature in K, N_e = electron density (per m^3), and Z is the charge number of the predominant ion. Take $Z = 1$, and calculate η for the three cases of Problem 14-3.

14-5 A uniform electric field of 10^4 V/m is directed at right angles to a uniform magnetic field $B = 0.1$ T. (a) Determine the plasma drift velocity for charged particles. (b) Plot the cycloidal path in a plane perpendicular to B for a proton with velocity $V_{\perp} = 10^5$ m/s and for a proton with $v_{\perp} = 10^6$ m/s.

14-6 An asteroid with no atmosphere has a magnetic dipole field. At a distance of 200 km (from its center) and at $\theta = 90$ the magnetic field is 10^{-6} T. A proton finds

itself trapped by the field at this point; it has an orbital speed of 10^5 m/s and the same speed parallel to the field. (a) Describe its subsequent motion. (b) At what angle θ is the particle reflected? (Use the results found in Problem 8-28.)

14-7 Determine the Alfvén velocity and attenuation distance for (a) chromosphere of the sun ($\zeta = 10^{-7}$ kg/m³, 5000 K, $B = 10^{-4}$ T), and (b) liquid mercury at 20°C ($\zeta = 13,600$ kg/m³, $B = 0.1$ T).

***14-8** A homogeneous sphere of radius a and electrical conductivity σ' moves with velocity $-\mathbf{v}_0$ in a nonviscous, incompressible fluid of conductivity σ in the presence of a uniform magnetic field \mathbf{B}_0 . The velocity \mathbf{v}_0 is parallel to \mathbf{B}_0 . (a) Calculate the Joule loss resulting from induced currents in the system, and (b) by equating this result to the rate at which mechanical energy is dissipated by the sphere ($F_1 v_0$), calculate the drag force F_1 . Assume potential flow in the fluid: In the coordinate system in which the sphere is at rest, the fluid velocity is given by

$$\mathbf{v} = \mathbf{v}_0 + \frac{1}{2}a^3 \nabla(\mathbf{v}_0 \cdot \mathbf{r}/r^3)$$

relative to an origin at the center of the sphere. [For a discussion of this and related problems, see J. R. Reitz and L. L. Foldy, *Journal of Fluid Mechanics*, vol. 11, p. 133 (1961).]

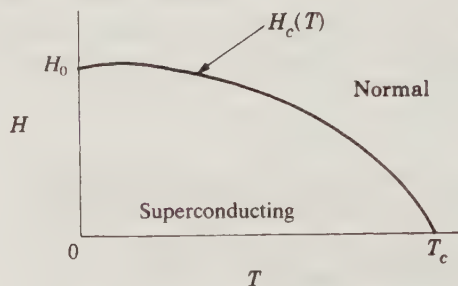
Electromagnetic Properties of Superconductors

In this chapter, we consider another special class of materials—superconducting solids—of very great theoretical interest and growing practical importance. Conductors that have zero resistance and can carry current without the power dissipation discussed in Chapters 7 and 13 have many potential applications, some of which have been realized. It is less obvious, but the magnetic behavior of superconductors is of equal importance because of the limitations it imposes on applications. Our focus here is mainly on the magnetic properties of the superconducting state.

THE HISTORY OF SUPERCONDUCTIVITY

Superconductivity was first discovered in 1911 by H. Kammerlingh Onnes at Leiden. He observed that as a sample of mercury was cooled, its resistance disappeared abruptly and apparently completely at 4.2 K. In a more sensitive experiment using a persistent current induced in a loop of superconducting wire, Kammerlingh Onnes estimated the resistance in the superconducting state to be at most 10^{-12} of the resistance in the normal state. In more recent experiments at the Massachusetts Institute of Technology, it was found that an induced current of several hundred amperes in a superconducting lead ring showed no change in the magnitude of the current for a period of at least one year, which provides strong evidence that the resistance in the superconducting state is indeed zero. The early experiments opened up a whole field of endeavor to characterize the new effect. It has been found that more than 20 elements and hundreds of alloys and intermetallic compounds are superconductors with transition temperatures ranging from substantially less than 1 K (e.g., 0.12 K for hafnium) to about 23 K, as well as perovskite compounds containing copper-oxygen bonds (the latter materials having transition temperatures

FIGURE 15-1
Plot of critical field **H**
versus temperature for a
type I superconductor



higher than 100 K in some cases). The transition, or critical, temperature is the temperature at which the transition from the normal state to the superconducting state takes place and is characteristic of the particular material being considered. The critical temperature depends to some extent on both the chemical purity and crystalline perfection of the sample being tested. Actually, inhomogeneities in the purity and strain of the sample generally tend to broaden the temperature range of the transition between the normal and superconducting states. A pure well-annealed sample may have a transition-temperature range as small as 0.001 K.

If a large enough magnetic field is applied parallel to a superconducting wire, it is found that the sample will become normal. The magnitude of the field that causes the transition is dependent on both the material and the temperature and is called the critical field. If a field is applied in some other direction, the sample will start to become normal when the actual field at any point on the surface reaches the critical field. The field-temperature plot that can be made has essentially the same thermodynamic significance as the pressure-temperature diagram for ordinary phase transitions, and the curve itself may be considered as the phase boundary between the normal and superconducting thermodynamic states (see Fig. 15-1). The shape of the curve is generally parabolic and is given to a good approximation by the equation

$$H_c = H_0[1 - (T/T_c)^2]$$

where H_c is the critical field, T is the absolute (or Kelvin) temperature of observation, and T_c and H_0 represent the sample characteristics (critical temperature at zero field and critical field at zero absolute temperature). In addition to broadening the transition, inhomogeneities may also have a marked effect on H_0 , sometimes increasing it by orders of magnitude. Such effects are of major importance in applications with high magnetic fields.

In the early history of superconductivity, the application of Maxwell's equations to a perfect conductor led to the conclusion that the time rate of change of the magnetic induction in the interior of the superconductor should be zero. Thus, depending on whether the sample was cooled to

below the transition temperature in the presence or absence of an applied magnetic field, magnetic flux should be trapped or excluded. This idea was so firmly believed that it was not until 1933 (22 years after the discovery of superconductivity) that W. Meissner and R. Ochsenfeld first tested it experimentally. The results of their experiments proved the hypothesis to be false and that, in all cases, irrespective of whether the sample is cooled in or out of a magnetic field, the magnetic induction in a superconductor is zero. This effect is called *flux exclusion* or, more commonly, the *Meissner effect*. An essentially equivalent statement is that a superconductor behaves as though it had zero permeability or a perfect diamagnetic susceptibility. The main significance of the Meissner effect is that it demonstrates that a superconductor is characterized by more complex electromagnetic properties than simple infinite conductivity. Any satisfactory explanation of superconductivity has to explain both zero resistivity and the Meissner effect in a natural way.

Because of the Meissner effect, for applied magnetic fields less than the critical field H_c , the field does not penetrate the specimen and it remains superconducting; for fields larger than H_c (and oriented parallel to a wire specimen), the field penetrates it and the superconducting state is destroyed. Strictly speaking, this statement describes the magnetic behavior of so-called type I superconductors. There is another type of superconductor (type II), and here the magnetic behavior is more complicated: At low magnetic fields the flux is completely excluded (just as it is in type I behavior), which persists up to a lower critical field H_{c1} . At fields above this value, the flux partially penetrates, but superconductivity is not destroyed until the flux penetrates completely at the upper critical field H_{c2} . Between H_{c1} and H_{c2} , the type II material does not exhibit a perfect Meissner effect. The magnetic behavior of both type I and type II materials is shown in Fig. 15-2.

Ever since the discovery of superconductivity by Kammerlingh Onnes there has been a dedicated effort to find superconducting materials with higher transition temperatures (it was hoped that a material could be found with a T_c above the boiling point of liquid nitrogen). But after decades of

FIGURE 15-2
Magnetization M versus
magnetic field H for
types I and II
superconductors

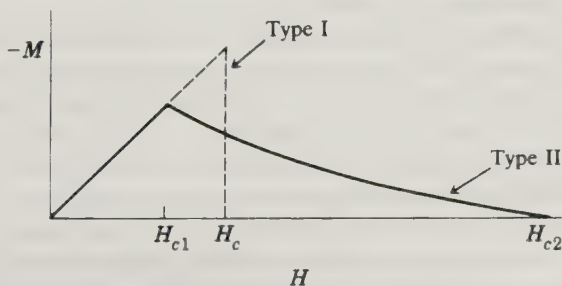


TABLE 15-1
Properties of Superconducting Materials

Material	T_c (K)	$\mu_0 H_0$ (tesla)*	λ_{exp} (10^{-10} m)
Aluminum	1.18	0.0105	505
Lead	7.20	0.0803	390
Mercury	4.15	0.0411	400
Tantalum	4.47	0.0829	345
Tin	3.72	0.0305	510
<i>Type II</i>			
<i>Superconductors</i>		$\mu_0 H_{c1}, \mu_0 H_{c2}$ (tesla)*	
Niobium	9.25	0.14, 0.42	380
Nb ₃ Sn	18.0	0.017, 7.0	400
V ₃ Ga	16.8	0.04, 35.0	
Nb _{0.6} Ti _{0.4}	9.3	—, >9.0	310
LaSrCuO _{4-δ}	38.5	0.15, —	2200
YBa ₂ Cu ₃ O _{6.9}	92.5	0.12, 15.0	1400 7500 \perp
Bi ₂ Sr ₂ CaCu ₂ O ₈	90	0.02, 60.0	>2500
Tl ₂ Ba ₂ Ca ₂ Cu ₃ O ₁₀	122	0.015, 100.0	>2500

* Critical field H_0 at $T = 0$ K; H_{c1} and H_{c2} , measured at 4.2 K except for V₃Ga (0 K), Nb₃Sn (14 K), and the high T_c , Cu–O superconductors (77 K). || means parallel to the Cu–O planes, and \perp means perpendicular to the Cu–O planes.

Source: Data from *Handbook of Chemistry and Physics*, 70th ed. (Boca Raton, FL: CRC Press, 1989); *Copper Oxide Superconductors* by C. P. Poole, T. Datta, and H. A. Farach (New York: John Wiley & Sons, 1988); and *Proceedings of the IEEE* 77 (No. 8), issue on High T_c Superconductivity (August 1989).

research, the highest T_c produced was in the vicinity of 23 K. Then in 1986 J. G. Bednorz and K. A. Müller discovered superconductivity in a new class of materials—a perovskite compound containing layers of copper and oxygen atoms—with a T_c above 30 K. This discovery stimulated intensive research on this new class of ceramic superconductors, and by early 1987, materials with T_c greater than 90 K were found. Bednorz and Müller were awarded the 1987 Nobel Prize for Physics for their discovery of this new class of superconducting materials.

Table 15-1 lists a number of superconducting materials along with their properties. It will be noted that some of the type II materials have very large critical fields (H_{c2}). Technological applications of superconductivity usually require a material that remains superconducting in large magnetic fields, and it is, therefore, the type II materials that are particularly useful. The quantity in the last column of the table is the experimentally determined value of the penetration depth, λ_{exp} (λ is defined in Section 15-4).

From a theoretical point of view, much has been done, starting with the application of thermodynamics to the superconducting transition by W. H.

Keesom as early as 1924. In 1934 a phenomenological explanation of the second-order transition and other properties, based on a two-fluid model, was developed by C. J. Gorter and H. B. G. Casimir. This work was followed (1935) by F. and H. London's account of the Meissner effect. In this chapter, we shall be mainly concerned with the so-called London equations (developed by the latter two scientists). But until the isotope effect was discovered in 1950,* there was little successful theoretical work directed toward a microscopic theory of superconductivity. In 1950, however, H. Fröhlich developed a theory based on the interaction of electrons with vibrating atoms in the crystal lattice, which explained the isotope effect but failed to predict other properties of the superconducting state. Then in 1957, J. Bardeen, L. N. Cooper and J. R. Schrieffer developed a microscopic or quantum-mechanical theory of superconductivity that has been quite successful. As a result of their work, Bardeen, Cooper, and Schrieffer were awarded the Nobel Prize in Physics in 1972.

The last-mentioned theory (BCS theory) accounts in a natural way for the second-order phase transition, the Meissner effect, and other thermodynamic and electromagnetic properties of superconductors. According to BCS theory, superconductivity manifests itself as a phase transition arising from the *pairing* of electrons. This pairing results from the interaction of electrons with lattice vibrations in the material. In some ways, superconductivity is analogous to a Bose-Einstein condensation of bound pairs of electrons; both effects (superconductivity and the Bose-Einstein condensation) are essentially quantum mechanical and have no simple classical interpretation. BCS theory appears capable of predicting, at least qualitatively, all of the experimental results relating to superconductivity, with the possible exception of superconductivity in the high- T_c , copper-oxide perovskites. But even in the latter case, the phase transition appears to arise from electron pairing, although the origin of pairing may be electron interaction with some other type of crystal excitation instead of with lattice vibrations.

The subject of superconductivity has developed into a very rich field of study. Nevertheless, the two complementary phenomenological theories—the Casimir-Gorter two-fluid theory and the London theory—together are adequate for the consideration of many problems involving superconductors. The Casimir-Gorter theory deals mainly with thermodynamic questions and is consequently only of peripheral interest here. The London theory, however, is for the most part an augmentation of Maxwell's equations for the purpose of constructing an electromagnetic theory that is

* Experiments on elemental superconductors of varying isotopic composition show that $T_c M^{1/2} \approx \text{constant}$, with M the isotopic mass. The first experimental work was done by E. Maxwell and by C. A. Reynolds *et al.* The effect is now known as the *isotope effect* and is a clue that the interactions between the superconducting electrons and the ion cores of the crystal lattice play an important role in superconductivity.

competent to deal with situations involving superconductors. (Strictly speaking, the London theory was developed for type I superconductors, but it can be applied qualitatively also to situations involving a type II material.) The balance of this chapter is concerned with the development of the London theory and its application to a few simple situations.

15-2

PERFECT CONDUCTIVITY AND PERFECT DIAMAGNETISM OF SUPERCONDUCTORS

We noted in the preceding section that superconductors exhibit two unique properties. They have essentially infinite conductivity, and they also exclude magnetic flux as demonstrated by the Meissner-Ochsenfeld experiment (so long as the magnetic field at the surface of the superconductor nowhere exceeds the critical field). These properties are independent in the sense that neither implies the other but, of course, both must and do emerge from satisfactory microscopic theories of superconductivity. To see more clearly what is meant by the independence of those two properties, we examine the behavior of a perfect conductor, but one that does *not* exhibit the Meissner effect in a magnetic field.

Consider a sphere whose conductivity can be switched from a finite value to infinity in some way, for example, by changing its temperature. When the conductivity is infinite the electric field is zero everywhere inside the material, and consequently its curl and $\partial\mathbf{B}/\partial t$ are also zero. Thus, if the sphere is cooled (attaining perfect conductivity) in a uniform field \mathbf{B}_0 , the flux density in the sphere *remains* \mathbf{B}_0 , even if the applied field is reduced to zero, until the perfect conductivity is destroyed. On the other hand, if the sphere is cooled in zero field, the flux density remains zero until the perfect conductivity is switched back to normal conductivity. Thus perfect conductivity does not imply flux exclusion and consequently $\mathbf{B} = 0$ is a postulate that must be introduced separately. Similarly $\mathbf{B} = 0$ does not imply perfect conductivity, for a material with susceptibility $\chi_m = -1$ would always have $\mathbf{B} = 0$, and this would not restrict the possible conductivity of the material.

In this chapter, we are primarily concerned with the magnetic aspects of superconductivity (infinite conductivity will be further discussed but will not play an important role in the problems considered here), and appropriate formalisms for this will be developed. The first approach, which represents the smallest deviation from what has already been done (cf. Chapter 9), is to say that inside a superconductor $\mathbf{B} = \mu_0[\mathbf{H} + \mathbf{M}] = 0$ and at boundaries between superconductors and other media, the tangential component of \mathbf{H} and the normal component of \mathbf{B} are continuous. This approach views the superconductor as a magnetic material with susceptibility $\chi_m = -1$; that is, a medium which exhibits perfect diamagnetism. At the surface of the superconductor, *magnetization* currents flow with surface density $\mathbf{j}_M = \mathbf{n} \times [\mathbf{M}_{\text{out}} - \mathbf{M}_{\text{in}}]$ with \mathbf{n} the outward drawn normal to the surface (note that \mathbf{M}_{out} is usually zero); inside the superconductor, volume magnetization

currents flow with density $\mathbf{J}_M = \nabla \times \mathbf{M}$ (cf. Chapter 9, particularly Section 9-1).

An alternative description puts $\mathbf{B} = \mathbf{H} = \mathbf{M} = 0$ inside the superconductor and invokes a *real* surface current $\mathbf{j}_S = \mathbf{n} \times \mathbf{H}_{out}$ (since \mathbf{H}_{in} is assumed to be zero). In this description there are no currents of any kind flowing in the interior of the superconductor. These two descriptions of a superconductor are so strikingly different that it is pertinent to ask how they are related. The usual statement is that they are equivalent when properly interpreted. However, it seems appropriate to consider the question in more detail. We first note that there are two differences between real transport currents and magnetization currents. The first of these is that transport currents are sources for \mathbf{H} while both transport and magnetization currents are sources for \mathbf{B} . Since \mathbf{B} is the magnetic field quantity that is accessible while \mathbf{H} was introduced mainly to have a magnetic field quantity determined by the transport currents, this first distinction between the two kinds of currents is clearly convenient but somewhat artificial. The second distinction is that transport currents in normal materials are dissipative (i.e., give rise to joule heating) while magnetization currents are not. But for superconductors, even this distinction disappears. Furthermore, since it can be shown that the magnetization of superconductors is not due to spins (and consequently is associated with orbital motions of charge carriers), the two descriptions appear to be equivalent. An alternative succinct statement is that since only \mathbf{B} is measurable, we can choose \mathbf{M} and \mathbf{H} according to relatively arbitrary rules so long as we apportion \mathbf{J} and \mathbf{J}_M accordingly and appreciate that they cannot be distinguished in a superconductor.

For most of what will be done, the $\mathbf{H}, \mathbf{M} \neq 0$ description will be convenient because it is a natural extension of what has been done earlier for normal materials and because this formulation leads to boundary value problems of a more conventional nature. In the next section, however, two problems will be considered, each in both formulations, in order to clarify the equivalence.

5-3

EXAMPLES INVOLVING PERFECT FLUX EXCLUSION

To reinforce the ideas presented in the preceding section, we will consider two elementary examples: a superconducting sphere in an asymptotically uniform field and an infinitely long, current-carrying, superconducting cylinder. Both formulations of Section 15-2 will be used to show explicitly that they are equivalent in these cases.

EXAMPLE 15-1

Superconducting Sphere in a Uniform Magnetic Field

Consider a superconducting sphere of radius a placed in a uniform external field $B_0\mathbf{k}$. We should like to determine the magnetic field \mathbf{B} both inside and outside of the sphere.

Solution: In the first formulation, which treats the superconductor as a magnetic material, the boundary value problem takes the form

$$\begin{aligned} \text{Outside: } \quad \mathbf{B} &\rightarrow B_0 \mathbf{k} \quad \text{as } r \rightarrow \infty \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{H} &= 0 \\ \mathbf{B} &= \mu_0 \mathbf{H} \end{aligned} \tag{15-1}$$

$$\begin{aligned} \text{Inside: } \quad \mathbf{B} &= 0, \quad \mathbf{H} = -\mathbf{M} \\ \nabla \times \mathbf{H} &= 0 \\ \nabla \cdot \mathbf{M} &= 0 \end{aligned} \tag{15-2}$$

$$\begin{aligned} \text{At } r = a: \quad B_r &\text{ continuous} \\ H_\theta &\text{ continuous} \end{aligned} \tag{15-3}$$

The only unusual equation is $\nabla \cdot \mathbf{M} = 0$, which is established on the basis that there are no magnetic poles inside the superconducting sphere. With these equations, two magnetic scalar potentials, φ_1^* outside and φ_2^* inside, may be introduced. Both satisfy Laplace's equation, and from them, the \mathbf{H} -field may be found by taking the negative gradient. Utilizing spherical coordinates and explicitly taking account of the first equation of Eq. (15-1), we have

$$\varphi_1^* = -\frac{B_0}{\mu_0} r \cos \theta + \sum_{\ell=0}^{\infty} c_\ell r^{-(\ell+1)} P_\ell(\cos \theta) \tag{15-4}$$

Thus

$$B_r = B_0 \cos \theta + \mu_0 \sum_{\ell=0}^{\infty} (\ell + 1) c_\ell r^{-(\ell+2)} P_\ell(\cos \theta) \quad (\text{outside}) \tag{15-5}$$

Since \mathbf{B} is zero inside and B_r is continuous across $r = a$, each c_ℓ except c_1 must be zero and $c_1 = -B_0 a^3 / 2\mu_0$. This procedure then completely solves the problem for $r > a$ without recourse to the boundary condition on the tangential component of \mathbf{H} ; the only thing that has entered is $\mathbf{B} = 0$ inside and the continuity of the normal component of \mathbf{B} at $r = a$. Inside the sphere, the potential φ_2^* must be regular at $r = 0$ and, in order to match the boundary conditions, can only involve $P_1(\cos \theta)$. Thus $\varphi_2^* = d_2 r \cos \theta$, with d_2 a constant to be determined, and by differentiation, $H_r = -d_2 \cos \theta$ and $H_\theta = d_2 \sin \theta$. Since outside $H_\theta = -\frac{3}{2}(B_0/\mu_0) \sin \theta$, it follows that $d_2 = -3B_0/2\mu_0$. There is no surface transport current, but a surface magnetization current $\mathbf{j}_M = -\frac{3}{2}(B_0/\mu_0) \sin \theta \mathbf{a}_\phi$ exists because of the discontinuity of \mathbf{M} . These

results may be summarized as follows:

$$\begin{aligned} \text{Outside: } \mathbf{B} &= \mu_0 \mathbf{H} = B_0 \mathbf{k} - B_0 \frac{a^3}{r^3} \cos \theta \mathbf{a}_r - \frac{1}{2} B_0 \frac{a^3}{r^3} \sin \theta \mathbf{a}_\theta \\ \text{Inside: } B &= 0, \quad \mathbf{H} = \frac{3 B_0}{2 \mu_0} \mathbf{k}, \quad \mathbf{M} = -\frac{3 B_0}{2 \mu_0} \mathbf{k} \\ \text{At } r = a: \quad \mathbf{j}_M &= -\frac{3 B_0}{2 \mu_0} \sin \theta \mathbf{a}_\phi \end{aligned} \quad (15-6)$$

The second formulation is identical for the outside region but takes the form $\mathbf{B} = \mathbf{H} = \mathbf{M} = 0$ for the inside. There is also a real transport current $\mathbf{j}_S = \mathbf{n} \times \mathbf{H}_{\text{out}} = -\frac{3}{2}(B_0/\mu_0) \sin \theta \mathbf{a}_\phi$ at the surface. This description can be summarized as follows:

$$\begin{aligned} \text{Outside: } \mathbf{B} &= \mu_0 \mathbf{H} = B_0 \mathbf{k} - B_0 \frac{a^3}{r^3} \cos \theta \mathbf{a}_r - \frac{1}{2} B_0 \frac{a^3}{r^3} \sin \theta \mathbf{a}_\theta \\ \text{Inside: } \mathbf{B} &= \mathbf{H} = \mathbf{M} = 0 \\ \text{At } r = a: \quad \mathbf{j}_S &= -\frac{3 B_0}{2 \mu_0} \sin \theta \mathbf{a}_\phi \end{aligned} \quad (15-7)$$

The relationship between the two descriptions is perhaps now clear. Outside, both are the same as they must be. Otherwise a simple experiment could be devised to select the correct description. Inside, both descriptions give $\mathbf{B} = 0$ but \mathbf{H} and \mathbf{M} are finite in one case and zero in the other. Neither \mathbf{H} nor \mathbf{M} , however, is experimentally observable, and consequently, this distinction is unimportant. Identical surface currents exist in the two cases: In one, however, it is considered a transport current, while in the other it is called a magnetization current. What the current is called is important only in that it must be consistent with \mathbf{H} and \mathbf{M} inside the superconductor. For example, when we calculate the magnetic moment of the superconducting sphere, either \mathbf{j}_M or \mathbf{M} may be used, but not both; however, a real \mathbf{j}_S always contributes to the magnetic moment.

EXAMPLE 15-2 Long Current-Carrying Wire

A second example further illustrates the indistinguishability of superconducting transport and magnetization currents. The case of an infinitely long, superconducting cylinder carrying a current. Before considering this problem in detail, however, we should note that inside a perfect superconductor the sum of \mathbf{J} and \mathbf{J}_M is always zero. This follows from $\mathbf{B} = 0$, which implies $\nabla \times \mathbf{B} = 0$ and thus, in turn, $\nabla \times \mathbf{H} + \nabla \times \mathbf{M} = \mathbf{J} + \mathbf{J}_M = 0$. At a surface of discontinuity this argument cannot be used, and a finite total surface current $\mathbf{j}_S + \mathbf{j}_M$ may exist. However, this argument shows clearly that the total current is

always a surface current. Let us return now to the wire, which will be taken to have radius a and to be carrying a current I_0 (in the positive z -direction).

Solution: We see from Ampere's law that outside the wire, $\mathbf{B} = \mu_0 \mathbf{H} = (\mu_0 I / 2\pi r) \mathbf{a}_\theta$ (cylindrical coordinates). If the first description, $\mathbf{M}, \mathbf{H} \neq 0$ inside, is used, we must make some assumption about the current density in the wire, and, because of the assumed continuity of the tangential component of \mathbf{H} , this assumption must not involve surface currents. The simplest possibility is uniform density: $\mathbf{J} = (I_0 / \pi a^2) \mathbf{k}$. Then inside,

$$\mathbf{H} = \frac{I_0 r}{2\pi a^2} \mathbf{a}_\theta \quad \text{and} \quad \mathbf{M} = -\frac{I_0 r}{2\pi a^2} \mathbf{a}_\theta$$

The magnetization current density is $\mathbf{J}_M = -(I_0 / \pi a^2) \mathbf{k}$, and at the surface there is a surface magnetization current density

$$\mathbf{j}_M = +\mathbf{a}_r \times \left(\frac{I_0}{2\pi a} \mathbf{a}_\theta \right) = \frac{I_0}{2\pi a} \mathbf{k}$$

which is just sufficient to carry the total current I_0 . The alternative description simply takes $\mathbf{B} = \mathbf{H} = \mathbf{M} = 0$ inside and thus requires that the current exist entirely on the surface with real surface current density $\mathbf{j}_S = (I_0 / 2\pi a) \mathbf{k}$. These two descriptions are summarized in Table 15-2. Unless a method for separating transport currents from magnetization currents in superconductors or a way of directly measuring \mathbf{H} or \mathbf{M} inside a superconductor can be found, the descriptions are equivalent.

In the two problems just considered the $\mathbf{M} = \mathbf{H} = 0$ formulation has an apparent advantage in simplicity. However, for more complicated problems, particularly those involving large demagnetizing factors, the distributed magnetization formulation is advantageous. Either method may be used and the results will be equivalent, but they must not be mixed in a single problem.

15-4
THE LONDON EQUATIONS

In the preceding section, flux exclusion was discussed on the basis of a highly idealized representation of a superconductor. This representation reproduces many of the observed features of superconductivity but fails to account adequately for some of the readily observable details. A more sophisticated theory can be developed by starting from the concept of perfect conductivity and making an appropriate modification to include the Meissner effect.

In a perfect conductor (*not* a superconductor), charge carriers would experience no retarding forces. Consequently, in an electric field \mathbf{E} they

TABLE 15-2
 Superconducting
 Current-Carrying Wire

	Formulation 1 (Superconductor as magnetic material with $\chi_m = -1$)	Formulation 2 (Flux exclusion by surface transport currents)
	$\mathbf{M} = -\mathbf{H} \neq 0$	$\mathbf{M} = \mathbf{H} = 0$
Outside:	$\mathbf{B} = \mu_0 \mathbf{H} = \frac{\mu_0 I_0}{2\pi r} \mathbf{a}_\theta$	$\mathbf{B} = \mu_0 \mathbf{H} = \frac{\mu_0 I_0}{2\pi r} \mathbf{a}_\theta$
Inside:	$\mathbf{B} = 0$	$\mathbf{B} = 0$
	$\mathbf{H} = \frac{I_0 r}{2\pi a^2} \mathbf{a}_\theta$	$\mathbf{H} = 0$
	$\mathbf{M} = -\frac{I_0 r}{2\pi a^2} \mathbf{a}_\theta$	$\mathbf{M} = 0$
	$\mathbf{J} = \frac{I_0}{\pi a^2} \mathbf{k}$	$\mathbf{J} = 0$
	$\mathbf{J}_M = -\frac{I_0}{\pi a^2} \mathbf{k}$	$\mathbf{J}_M = 0$
At $r = a$:	$\mathbf{j}_M = (I_0/2\pi a) \mathbf{k}$	$\mathbf{j}_M = 0$
	$\mathbf{j}_S = 0$	$\mathbf{j}_S = (I_0/2\pi a) \mathbf{k}$

would move according to

$$m_p \dot{\mathbf{v}} = q\mathbf{E} \quad (15-8)$$

where m_p is the mass of the charge carrier and $\dot{\mathbf{v}}$ is its acceleration. But if \mathbf{v} is the average velocity of the charge carriers and there are n of them per unit volume, then the current density is $\mathbf{J} = nq\mathbf{v}$. An alternative to Eq. (15-8) is then

$$\dot{\mathbf{J}} = (nq^2/m_p)\mathbf{E} \quad (15-9)$$

where $\dot{\mathbf{J}} = d\mathbf{J}/dt$. Taking the curl of this equation and using $\nabla \times \mathbf{E} = -\partial\mathbf{B}/\partial t$, we have

$$\nabla \times \dot{\mathbf{J}} = -(nq^2/m_p)\dot{\mathbf{B}} \quad (15-10)$$

Assuming that the fields are slowly varying* and using $\nabla \times \mathbf{H} = \mathbf{J}$ to eliminate $\dot{\mathbf{J}}$, we have

$$\nabla \times \nabla \times \dot{\mathbf{H}} = -(nq^2/m_p)\dot{\mathbf{B}} \quad (15-11)$$

Assuming that $\mathbf{B} = \mu_0 \mathbf{H}$ and using the definition of the Laplacian of a

* This assumption entails neglecting the displacement current, $\partial\mathbf{D}/\partial t$, which is discussed in Sections 16-1 and 16-2.

vector (with $\nabla \cdot \mathbf{B} = 0$) yields

$$\nabla^2 \dot{\mathbf{B}} = (\mu_0 n q^2 / m_p) \dot{\mathbf{B}} \quad (15-12)$$

The significance of this equation may be seen best by considering a semi-infinite perfect conductor bounded by the plane $z = 0$ and extending in the positive z -direction. Assume that at the surface $\dot{B}_y = \dot{B}_z = 0$, $\dot{B}_x = \dot{B}_{x0}$, and that \dot{B}_{x0} does not depend on x or y . The equation determining \dot{B}_x is then

$$\frac{d^2 \dot{B}_x}{dz^2} = \frac{\mu_0 n q^2}{m_p} \dot{B}_x \quad (15-13)$$

which has a general solution

$$\dot{B}_x = A e^{-\sqrt{\mu_0 n q^2 / m_p} z} + B e^{\sqrt{\mu_0 n q^2 / m_p} z}$$

The exponentially increasing solution may be rejected as being without any physical interpretation, and A may be chosen to give \dot{B}_x correctly at $z = 0$; then

$$\dot{B}_x = \dot{B}_{x0} e^{-\sqrt{\mu_0 n q^2 / m_p} z} \quad (15-14)$$

It is easy to verify that $(m_p / \mu_0 n q^2)^{1/2}$ has the dimensions of length and that for q and m_p appropriate to an electron and n corresponding to one electron per atom, this length is about 10^{-8} m. Equation (15-12) then indicates that, in the interior of the perfect conductor, the time derivative of \mathbf{B} goes to zero exponentially with distance from the surface. Thus, in the interior of the perfect conductor, $\dot{\mathbf{B}}$ is very small except in a thin surface layer, which is a reasonable refinement of the earlier conclusion that $\dot{\mathbf{B}} = 0$ everywhere in the interior of a perfect conductor.

The development just sketched shows again that perfect conductivity does not lead to flux exclusion. However, it also indicates how flux exclusion might be incorporated into a theory. If Eq. (15-12) described the behavior of \mathbf{B} instead of $\dot{\mathbf{B}}$, then \mathbf{B} itself would decrease exponentially from its value at the surface to zero in the interior of a superconductor. This idea was the motivation for the development of a theory of the electromagnetic behavior of superconductors by F. and H. London.*

In this theory, it is assumed that the total current can be broken up into a supercurrent \mathbf{J}_S , a dissipative current \mathbf{J}_{diss} , and a displacement current \mathbf{J}_{disp} :

$$\mathbf{J} = \mathbf{J}_S + \mathbf{J}_{\text{diss}} + \mathbf{J}_{\text{disp}} \quad (15-15)$$

The dissipative and displacement currents are governed by the equations $\mathbf{J}_{\text{diss}} = g\mathbf{E}$ and $\mathbf{J}_{\text{disp}} = \partial\mathbf{D}/\partial t$. The current \mathbf{J}_S can be related to the electromagnetic field by using Eq. (15-15), Maxwell's equations, and

* F. London and H. London, *Proceedings of the Royal Society*, vol. A 149, p. 71 (1935).

London's constitutive equation [similar in form to Eq. (15-10) but involving \mathbf{B} and \mathbf{J} rather than their derivatives]. If this procedure is carefully followed, it can be shown that for frequencies less than about 10^{11} Hz, both \mathbf{J}_{diss} and \mathbf{J}_{disp} are negligible compared to \mathbf{J}_S . We should like to assume this result; namely $\mathbf{J}_{\text{diss}} \approx 0$ and $\mathbf{J}_{\text{disp}} \approx 0$, without going through the detailed arguments. Such an assumption is at least a reasonable one for steady-current problems of the type considered in this chapter. The remaining current \mathbf{J}_S includes both transport and magnetization currents and consequently, from Maxwell's Eq. (8-51),

$$\mathbf{J}_S = (1/\mu_0)\nabla \times \mathbf{B} \quad (15-16)$$

To obtain an equation involving the magnetic field variables rather than their derivatives, London *postulated* that

$$\mu_0\nabla \times \mathbf{J}_S = -(1/\lambda^2)\mathbf{B} \quad (15-17)$$

This equation differs from Eq. (15-10) in that it involves \mathbf{J}_S and \mathbf{B} rather than $\dot{\mathbf{J}}$ and $\dot{\mathbf{B}}$. Therefore, it will lead to an equation analogous to Eq. (15-12) for the fields rather than for their derivatives. Also, a phenomenological penetration depth λ has been introduced as a specific parameter characteristic of the superconducting material (μ_0 occurs in order to make the dimensions of λ a length). Equation (15-17) will lead to the Meissner effect but, in order to include infinite conductivity, we must assume separately that

$$\mu_0\dot{\mathbf{J}}_S = (1/\lambda^2)\mathbf{E} \quad (15-18)$$

The latter equation will, however, play no further role in the problems considered here.

Equations (15-16) and (15-17) may be combined to give

$$\nabla \times \nabla \times \mathbf{B} = -(1/\lambda^2)\mathbf{B} \quad (15-19)$$

Since $\nabla \cdot \mathbf{B} = 0$ this may be written

$$\nabla^2\mathbf{B} = (1/\lambda^2)\mathbf{B} \quad (15-20)$$

Equation (15-20) may be solved in the case of a semi-infinite slab exactly as Eq. (15-13) was. The solution

$$B_x(z) = B_{x0}e^{-z/\lambda} \quad (15-21)$$

now indicates that \mathbf{B} rather than $\dot{\mathbf{B}}$ falls exponentially as the slab is penetrated and is the desired generalization of $\mathbf{B} = 0$ inside a superconductor.

The penetration depth λ has been introduced here as a phenomenological parameter; various theories have, however, been constructed that attempt to compute its size. We are more interested in the experimental determination of λ . An obvious approach would be to construct a solenoid with a superconducting core. The inductance of such a solenoid would be

very small if the superconductor were perfect and completely filled the volume enclosed by the solenoid. If, on the other hand, there were a finite penetration depth, then the inductance would be somewhat larger. If the penetration depth were a significant fraction of the radius of the solenoid, the penetration depth could be inferred from measurements of the inductance. The feasibility of such a determination depends on the ratio of the volume into which the field penetrates to the total volume of the sample. As it turns out, λ is typically a few millionths of a centimeter and, consequently, the simple experiment proposed above will not produce significant results. This difficulty can, however, be circumvented by using a sample with a large surface-to-volume ratio. The first successful experiments of this type, done with a mercury colloid, were performed by D. Shoenberg in 1939. These experiments demonstrated conclusively that the magnetic field penetrated into the small superconducting mercury spheres and that the penetration depended on temperature. The original Shoenberg experiments, which have been extended and supplemented, demonstrated that the concept of a penetration depth was valid and important.

Equations (15-15), (15-17), and (15-18), together with the four Maxwell equations, are often collectively called the *Maxwell-London equations* and are very useful for treating electromagnetic problems involving superconductors.

As is evident from the preceding discussion, the concept of perfect flux exclusion is an idealization. Instead, magnetic flux penetrates a thin layer at the surface of the superconductor and, according to the London theory, decreases exponentially toward the interior. The surface current density, \mathbf{j}_M (or as the case may be, \mathbf{j}_S), is also an idealization. Here again, the supercurrent density \mathbf{J}_S is spread out in a thin surface layer and decreases exponentially toward the interior. Because the supercurrent is confined within a few penetration depths of the interface, in most superconducting specimens it resembles a surface current. However, there is no \mathbf{j}_M in the London theory, only a total supercurrent density \mathbf{J}_S .

In order to understand the Maxwell-London equations better, they will now be used to obtain more refined solutions to the problems considered in Section 15-3.

***EXAMPLE 15-3**
London's Treatment of
Superconducting Sphere in
a Uniform Field

This example treats the same problem described in Example 15-1 by using the Maxwell-London equations.

Solution: The equations satisfied by the fields are

$$\begin{aligned} \text{Outside: } \nabla \cdot \mathbf{B} &= 0, & \nabla \times \mathbf{H} &= 0, & \mathbf{B} &= \mu_0 \mathbf{H} \\ \text{Inside: } \nabla^2 \mathbf{B} &= (1/\lambda^2)\mathbf{B}, & \nabla \cdot \mathbf{B} &= 0 \end{aligned} \quad (15-22)$$

where λ , the penetration depth, is considered as a phenomenological

parameter. The boundary conditions that must be satisfied are

$$\begin{aligned} \text{At } r = \infty: \quad \mathbf{B} &= B_0 \mathbf{k} \\ \text{At } r = a: \quad B_r \text{ and } B_\theta, &\text{ continuous} \end{aligned} \quad (15-23)$$

The only one of these boundary conditions that requires further comment is the continuity of B_θ at $r = a$, which follows from the assumption, in agreement with the discussion at the end of the preceding paragraph, that the supercurrents (both transport and magnetization) are never infinite—that is, there are no surface current densities \mathbf{j}_M or \mathbf{j}_S . In this case the tangential components of both \mathbf{H} and \mathbf{M} are continuous, and hence the tangential component of \mathbf{B} is also continuous.

The solution of the equations for the field outside the sphere poses no difficulties. A magnetic scalar potential satisfying Laplace's equation may be introduced exactly as was done in Section 15-3 and a general solution obtained. For the interior region, however, the equation $\nabla^2 \mathbf{B} = (1/\lambda^2)\mathbf{B}$ must be solved. If in spherical coordinates the Laplacian of a vector could be obtained simply by taking the Laplacian of each of its coordinates, then the solutions to this equation could be found readily. Such, however, is not the case, but rather the curl curl of the vector must be calculated. As a result, even in this simple problem, the r - and θ -components of $\nabla^2 \mathbf{B} = (1/\lambda^2)\mathbf{B}$ involve both B_r and B_θ . This complexity is relatively well known, and extensive techniques for solving the vector Helmholtz equation have been developed.* The development and application of the methods are, however, somewhat outside the scope of this book. Consequently, the results of Section 15-3 will be used to make a guess (*Ansatz*) at the form of the solution. The resulting final solution will be justified because it satisfies the equations and the boundary conditions and because these equations and boundary conditions have a unique solution. The uniqueness can, of course, be proved, but here it will be assumed.

In Section 15-3 it was found that only the $P_1(\cos \theta)$ -term in φ^* survived in the solution for the region outside the sphere. It will be assumed that this is also true for the Maxwell-London theory and that

$$\mathbf{B}(r, \theta) = B_0 \mathbf{k} - b \left(\frac{a}{r} \right)^3 \left[\cos \theta \mathbf{a}_r + \frac{1}{2} \sin \theta \mathbf{a}_\theta \right] \quad (\text{outside}) \quad (15-24)$$

This is very similar to the first of Eqs. (15-7), the only difference being that B_0 has been replaced by b in that part of the field that is due to the

* Cf. Morse and Feshbach, *Methods of Theoretical Physics* (New York: McGraw-Hill, 1953), Chapter 13.

magnetization of the sphere. The value of b will be determined by fitting the boundary conditions. For the interior of the sphere, Section 15-3 provides little by way of clues; however, from the form of \mathbf{M} found there and the fact that in Eq. (15-24) B_r depends on θ through $\cos \theta$ while B_θ depends on θ through $\sin \theta$, a reasonable assumption might be that

$$B_r = u(r) \cos \theta \quad (\text{inside}) \quad (15-25a)$$

$$B_\theta = v(r) \sin \theta \quad (\text{inside}) \quad (15-25b)$$

The two functions $u(r)$ and $v(r)$ must be determined so that $\nabla^2 \mathbf{B} = (1/\lambda^2) \mathbf{B}$ and also so that the boundary conditions at $r = a$ are satisfied. These boundary conditions are

$$u(a) = B_0 - b \quad (15-26a)$$

$$v(a) = -B_0 - b/2 \quad (15-26b)$$

By expanding $\nabla \times \nabla \times \mathbf{B}$ using the assumed forms (15-25), we find the equations

$$r \frac{dv}{dr} + v + u = -\frac{r^2}{2\lambda^2} u \quad (15-27a)$$

and

$$r^2 \frac{d^2v}{dr^2} + 2r \frac{dv}{dr} + r \frac{du}{dr} = \frac{r^2}{\lambda^2} v \quad (15-27b)$$

for u and v . Now, using Eqs. (15-25) in the $\nabla \cdot \mathbf{B} = 0$ equation, we obtain

$$u' + (2/r)(u + v) = 0 \quad (15-28)$$

Using this result to eliminate v and dv/dr from Eq. (15-27a) gives the equation for u :

$$r^2 \frac{d^2u}{dr^2} + 4r \frac{du}{dr} = \frac{r^2}{\lambda^2} u \quad (15-29)$$

In order to solve this equation, we introduce a new dependent variable

$$\psi \equiv r(u - 2v) \quad (15-30)$$

and differentiate it to obtain ψ' and ψ'' in terms of u and v and their derivatives. Now subtract two times Eq. (15-27b) from Eq. (15-29); the resulting equation is seen to be

$$r\psi'' = (r/\lambda^2)\psi \quad (15-31)$$

which has the solutions: $\sinh(r/\lambda)$ and $\cosh(r/\lambda)$, or $\exp(r/\lambda)$ and

$\exp(-r/\lambda)$. Since we want ψ/r to be regular at the origin, we choose

$$\psi = c_1 \sinh(r/\lambda) \quad (15-32)$$

Equation (15-28) can now be written as

$$u' + \frac{3u}{r} = \frac{\psi}{r^2} = \frac{c_1 \sinh(r/\lambda)}{r^2} \quad (15-33)$$

This equation suggests that u will have the form (which is regular at the origin if $n \leq 3$)

$$u(r) = c_2(\lambda/r)^n [\sinh(r/\lambda) - (r/\lambda) \cosh(r/\lambda)] \quad (15-34)$$

The constant n can be obtained by substituting Eq. (15-34) into (15-33) [or into (15-29), but the former course of action is less work]. It is found that $n = 3$. Putting this result in Eq. (15-28) and solving for v , we find that

$$v = \frac{c_2}{2} \left(\frac{\lambda}{r}\right)^3 \left[\left(1 + \frac{r^2}{\lambda^2}\right) \sinh\left(\frac{r}{\lambda}\right) - \left(\frac{r}{\lambda}\right) \cosh\left(\frac{r}{\lambda}\right) \right] \quad (15-35)$$

It still remains to determine the constants b and c_2 , using Eqs. (15-26), (15-34), and (15-35). These are found to be

$$c_2 = -3B_0(a/\lambda)/\sinh(a/\lambda) \quad (15-36)$$

$$b = B_0[1 + 3(\lambda/a)^2 - 3(\lambda/a) \coth(a/\lambda)] \quad (15-37)$$

It might be expected that for very small values of λ/a the fields would not be greatly different from those found in Section 15-3 for the perfect superconducting sphere. We can verify that such is the case by using the fact that $\coth x$ approaches unity exponentially for large values of x . Thus

$$b \approx B_0 \left(1 - 3\frac{\lambda}{a} + 3\frac{\lambda^2}{a^2} + \dots \right), \quad \frac{\lambda}{a} \ll 1 \quad (15-38)$$

and the first correction to the field outside the sphere is of order λ/a .

EXAMPLE 15-4
London's Treatment of
Long, Current-Carrying
Wire

The second example to illustrate using the London equations is the long current-carrying wire. The radius of the wire will be taken to be a , the penetration depth λ , and the total transport current I_0 .

Solution: Outside the wire \mathbf{H} is given by Ampere's law and $\mathbf{B} = \mu_0 \mathbf{H}$. Therefore,

$$H_r = H_z = B_r = B_z = 0, \quad B_\theta = \mu_0 H_\theta = \mu_0 \frac{I_0}{2\pi r} \quad (\text{outside}) \quad (15-39)$$

Inside, \mathbf{B} satisfies

$$\nabla^2 \mathbf{B} = \frac{1}{\lambda^2} \mathbf{B} \quad (15-40)$$

By symmetry, \mathbf{B} has only a θ -component, which depends only on r . Then Eq. (15-40) becomes

$$r^2 \frac{d^2}{dr^2} B_\theta + r \frac{d}{dr} B_\theta - \left(1 + \frac{r^2}{\lambda^2}\right) B_\theta = 0 \quad (15-41)$$

which is just Bessel's equation for index one and argument ir/λ . The solution that is not infinite at the origin is

$$B_\theta = AJ_1(ir/\lambda) \quad (15-42)$$

The coefficient A is determined by matching B_θ inside and B_θ outside at $r = a$. The result is

$$B_\theta = \frac{\mu_0 I_0 J_1(ir/\lambda)}{2\pi a J_1(ia/\lambda)} \text{ (inside)} \quad (15-43)$$

Since $J_1(ir/\lambda) = iI_1(r/\lambda)$, where I_1 is a modified Bessel function, Eq. (15-43) may be written in terms of standard tabulated functions.

Another way of dealing with Eq. (15-41) is direct numerical integration, a procedure that is easily carried out on a personal computer. We begin by introducing the dimensionless variable $w = r/\lambda$ and rewriting Eq. (15-41) as

$$B'' = \frac{1}{w^2} [(1 + w^2)B - wB'] \quad 0 \leq w \leq a/\lambda \quad (15-44)$$

B has been used for B_θ and $B'' = d^2B/dw^2$, $B' = dB/dw$. We divide the interval from 0 to a/λ into N equal segments of length $\Delta w = a/N\lambda$. Assume that B , B' , and B'' are known at w_{n-1} and w_n . Then we predict a set of values for B_{n+1} , B'_{n+1} , and B''_{n+1} as follows:

$${}^{(1)}B_{n+1} = B_{n-1} + 2(\Delta w)B'_n \quad (15-45)$$

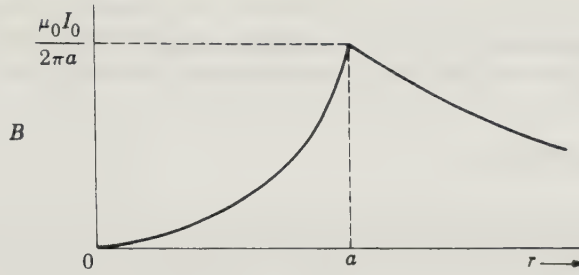
$${}^{(1)}B'_{n+1} = B'_{n-1} + 2(\Delta w)B''_n \quad (15-46)$$

$${}^{(1)}B''_{n+1} = \frac{1}{w_{n+1}^2} [(1 + w_{n+1}^2){}^{(1)}B_{n+1} - w_{n+1}{}^{(1)}B'_{n+1}] \quad (15-47)$$

We use these predicted values, together with the values at w_n , to compute approximate values for B' and B'' at a point midway between w_n and w_{n+1} . And then we recalculate:

$${}^{(2)}B_{n+1} = B_n + \frac{\Delta w}{2} (B'_n + {}^{(1)}B'_{n+1}) \quad (15-48)$$

FIGURE 15-3
 Penetration of the magnetic field into a superconductor according to the London theory. The radius of the wire $a = 5\lambda$.



$${}^{(2)}B'_{n+1} = B'_n + \frac{\Delta w}{2} (B''_n + {}^{(1)}B''_{n+1}) \quad (15-49)$$

$${}^{(2)}B''_{n+1} = \frac{1}{w_{n+1}^2} [(1 + w_{n+1}^2) {}^{(2)}B_{n+1} - w_{n+1} {}^{(2)}B'_{n+1}] \quad (15-50)$$

The last three quantities are taken to be good approximations to B , B' , and B'' at w_{n+1} . The entire process is repeated to find the values of B , B' , and B'' at the next meshpoint w_{n+2} .

The only remaining problem is how to find the values of B , B' , and B'' at two consecutive points to start the process. We do this by expanding B in a series:

$$B = w^m (c_0 + c_1 w + c_2 w^2 + c_3 w^3 + \dots)$$

and substituting in Eq. (15-44). Matching the coefficients of w gives $m = \pm 1$. The solution that is regular at the origin is $m = 1$. Matching higher powers of w determines c_1/c_0 , c_2/c_0 , and so on. The solution is

$$B = c_0 w (1 + \frac{1}{8} w^2 + \frac{1}{192} w^4 + \dots)$$

This series solution can be used to calculate B , B' , and B'' for an adjacent pair of points in the vicinity of the origin.

In this way, we evaluated the "exponential" flux penetration into a superconductor of radius $a = 5\lambda$; the result is shown in Fig. 15-3. We used $\Delta w = 0.1$ and started the numerical integration with a series solution in the vicinity of $w = 0.3$. The numerical accuracy of the procedure could be improved by making Δw smaller, but the accuracy with $\Delta w = 0.1$ was found to be quite good. (The curve in Fig. 15-3 is indistinguishable from the exact Bessel function solution.) The arbitrary constant c_0 is evaluated by making the value of B at the surface agree with the external value.

This discussion of the electromagnetic properties of superconductors has necessarily been fragmentary. In particular, problems involving time-dependent fields and the microscopic theory of superconductivity have been

ignored. The interested reader is referred to one of the many good introductory surveys of superconductivity.* Two of the early books that supplement this chapter are those by F. London† and D. Schoenberg.‡

15-5 SUMMARY

Superconductors constitute a fairly large class of materials that have a **phase transition** to the superconducting state at low temperatures, usually below about 20 K. The transition depends on the magnetic field as well as the temperature, with the material transforming back to the normal state in fields greater than the **critical field** H_c . Ideally,

$$H_c = H_0[1 - (T/T_c)^2]$$

The macroscopic electric and magnetic behavior in the superconducting state is most simply described by the **constitutive equations** $\mathbf{J} = g\mathbf{E}$ and $\mathbf{M} = \chi_m\mathbf{H}$ with extreme values of the material parameters,

$$g = 0, \quad \chi_m = -1 \quad (\mu = 0)$$

representing perfect conductivity and perfect diamagnetism. The latter requires that the flux density \mathbf{B} and the current density \mathbf{J} are zero inside a superconductor, and any superconducting currents are surface currents. A more refined description is given by the **London equations**, which replace the constitutive equations with the differential equations

$$\mu_0\dot{\mathbf{J}} = \left(\frac{1}{\lambda}\right)^2 \mathbf{E}, \quad \mu_0\nabla \times \mathbf{J} = -\left(\frac{1}{\lambda}\right)^2 \mathbf{B}$$

These, coupled with the Maxwell equations in vacuum, predict that the flux density and the current density fall off exponentially from the surface of a superconductor, with the **penetration depth** λ , instead of falling discontinuously to zero.

- The simple formulation with $\chi_m = -1$, $\mu = 0$, is most convenient for problems in which a superconductor is located in a magnetic field. The results are the same as those obtained by the techniques of Chapter 9, with the mere substitution of $\mu = 0$.

- An alternative formulation, which may be more convenient for problems in which the superconductor carries an external current, assumes that

* See for example, N. H. March and M. Parrinello, *Collective Effects in Solids* (Bristol: Adam Hilger, 1982), Chapter 5; and *Earlier and Recent Aspects of Superconductivity, Solid State Sciences 90*, edited by J. G. Bednorz and K. A. Müller (New York: Springer Verlag, 1990).

† F. London, *Superfluids. The Macroscopic Theory of Superconductivity*, Vol. I. (New York: Wiley, 1950; New York: Dover Publications, 1961).

‡ D. Shoenberg, *Superconductivity*, 2nd ed. (London: Cambridge University Press, 1965).

inside the superconductor $\chi_m = 0$, $\mu = \mu_0$. The condition $\mathbf{B} = 0$ is met with $\mathbf{H} = 0 = \mathbf{M}$ inside, and the boundary conditions are satisfied by assuming suitable superconducting surface transport currents. The two formulations are equivalent, but the choice of one *or* the other must be adhered to in a given problem.

PROBLEMS

15-1 Consider an infinitely long superconducting circular cylinder of radius a in a transverse magnetic field. At large distances from the cylinder the field is uniform and of magnitude B_0 . Compute (a) the fields inside and outside the cylinder and (b) the current density in the cylinder and on its surface. Assume that the superconducting properties are represented by perfect diamagnetism and perfect conductivity. Compare the two equivalent formulations.

15-2 Consider an infinite superconducting slab of thickness d , bounded by the planes $z = 0$ and $z = d$. Outside the slab the magnetic field is uniform and parallel to the surfaces, $B_x = B_0$. Find the field and the current density in the slab, using the London equations and the phenomenological penetration depth λ .

15-3 Do the computations of Problem 15-1, using the London equations and the phenomenological penetration depth λ .

15-4 Consider a superconducting sphere of radius a in a magnetic field that, at large distances from the sphere, is uniform and of magnitude B_0 . Using the formulation of Example 15-3 as a basis, provide the following: (a) An expansion of $\nabla \times \nabla \times \mathbf{B}$ and from it the equations satisfied by the components of \mathbf{B} inside the sphere; (b) a verification of Eq. (15-27); and (c) a quantitative discussion of the exponential fall-off of \mathbf{B} in the region close to the surface of the sphere.

15-5 A long, straight, superconducting wire of radius 0.2 mm carries a current I . (a) What is the magnetic field at the surface of the wire? (b) Calculate the value of the current I that will cause magnetic flux to penetrate the wire if the wire is composed of Pb, Ta, or Nb at 4.2 K.

15-6 A magnetic dipole is oriented parallel to and at distance d above a superconducting slab. (For simplicity assume that the slab is of infinite extent in its lateral dimensions.) (a) Consider the superconductor as a magnetic material with $\chi_m = -1$, and find an image dipole (or group of dipoles) that allows the boundary conditions to be satisfied at the interface between material and air [see also Problem 9-15(b)]. (b) Determine the magnetic force on the dipole. (c) If the mass of the dipole is m_D , at what distance d does the magnetic force balance the weight of the dipole?

15-7 Compare the "exponential" penetration of the magnetic field into the cylindrical wire of radius $= 5\lambda$ (Fig. 15-3) with that for the one-dimensional geometry of the semi-infinite slab. For the cylinder use the Bessel function solution based on $I_1(r/\lambda)$.

Maxwell's Equations

We are now ready to introduce the keystone of Maxwell's electromagnetic theory, the so-called *displacement current*. Although its observable effect is often negligible (indeed, it has been in all the phenomena treated up to this point), it constitutes an essential completion of the theory. And it has a crucial role in other topics—waves, optics, relativity—that make up the rest of the book.

16-1

THE GENERALIZATION OF AMPERE'S LAW: DISPLACEMENT CURRENT

In Chapter 9 we found that the magnetic field due to a current distribution satisfied Ampere's circuital law,

$$\oint \mathbf{H} \cdot d\mathbf{l} = \int \mathbf{J} \cdot \mathbf{n} da \quad (16-1)$$

We shall now examine this law, show that it sometimes fails, and find a generalization that is always valid.

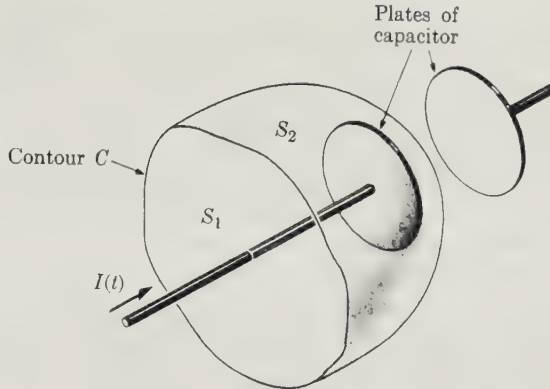
Consider the circuit shown in Fig. 16-1, which consists of a small parallel-plate capacitor being charged by a constant current I (we need not worry about what causes the current). If Ampere's law is applied to the contour C and the surface S_1 , we find

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = \int_{S_1} \mathbf{J} \cdot \mathbf{n} da = I \quad (16-2)$$

If, on the other hand, Ampere's law is applied to the same contour C and

FIGURE 16-1

The contour C and two surfaces, S_1 and S_2 , for testing Ampere's circuital law



the surface S_2 , then J is zero at all points on S_2 and

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = \int_{S_2} \mathbf{J} \cdot \mathbf{n} da = 0 \quad (16-3)$$

Equations (16-2) and (16-3) contradict each other and thus cannot both be correct. If C is imagined to be a great distance from the capacitor, it is clear that the situation is not substantially different from the standard Ampere law cases considered in Chapter 8. One is thus led to think that Eq. (16-2) is correct, since it is not dependent on the new feature, namely, the capacitor. Equation (16-3), on the other hand, requires consideration of the capacitor for its deduction. It would appear, then, that Eq. (16-3) requires modification.

This difficulty can be cast in a somewhat different form by *combining* Eqs. (16-2) and (16-3). S_2 and $-S_1$ together make a closed surface S . Using \mathbf{n} as the outward drawn normal everywhere we conclude that

$$\oint_S \mathbf{J} \cdot \mathbf{n} da = -I \quad (16-4)$$

with the minus sign coming from the change in the direction of the normal in Eq. (16-2). On the other hand, the surface integrals in Eqs. (16-2) and (16-3) are both equal to the line integral of \mathbf{H} around the same curve C . Pursuing this approach,

$$\oint_S \mathbf{J} \cdot \mathbf{n} da = \oint_C \mathbf{H} \cdot d\mathbf{l} - \oint_C \mathbf{H} \cdot d\mathbf{l} = 0 \quad (16-5)$$

where the minus sign arises from the change in the sense of traversal of C in the S_1 case. The contradiction now takes the form of a finite current I , which was assumed to be flowing into the volume bounded by S , being

equal to zero. Thus there is a conflict between Ampere's law, Eq. (16-1), from which the zero was obtained, and the initial assumption of a current I . The current flowing into the volume is in fact not zero, and indeed it is equal to the rate of change of the charge on the capacitor plate (according to the law of charge conservation).

This inconsistency can be explored and resolved more easily in the differential formulation where Ampere's law takes the form given by Eq. (9-30):

$$\nabla \times \mathbf{H} = \mathbf{J} \quad (9-30), (16-6)$$

Equation (16-2) is just the result of integrating Eq. (16-6) over the surface S_1 . If we take the divergence of Eq. (16-6), the result is zero since the divergence of any curl is zero. But we have another expression for the divergence of \mathbf{J} , namely, the charge conservation equation, Eq. (7-9):

$$\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0 \quad (7-9), (16-7)$$

Once again there is a significant inconsistency: The divergence of \mathbf{J} cannot be both zero and $-\partial\rho/\partial t$. There is no apparent problem with Eq. (16-7); it is difficult to imagine a way of modifying it to remove the inconsistency. What appears to be needed is a modification of Eq. (16-6) that will change the right-hand side into a vector with zero divergence. A way of doing so is to use Gauss's law in the form of Eq. (4-29)

$$\nabla \cdot \mathbf{D} = \rho \quad (4-29), (16-8)$$

to replace ρ in Eq. (16-7) by $\nabla \cdot \mathbf{D}$. Then

$$\nabla \cdot \left[\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \right] = 0 \quad (16-9)$$

where we have assumed that \mathbf{D} is a sufficiently continuous function of the space and time variables that the order of the derivatives can be interchanged. It is now clear that if $\partial\mathbf{D}/\partial t$ were added to the right-hand side of Eq. (16-6), the inconsistency would disappear—that is, the divergence of either side would be zero.

We are therefore led to *revise* Ampere's law and write it in the form

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \quad (16-10)$$

and refer to the time derivative of \mathbf{D} as the *displacement current*.

The introduction of the displacement current makes electromagnetic waves possible, as we shall soon see, and is the essence of Maxwell's great

contribution to electromagnetic theory. Our development is similar to his. We should emphasize that it is not a proof, but an hypothesis motivated by experimental observations. Further experiments have elevated the hypothesis to a principle on which subsequent development has been based.

In earlier chapters, we ignored the displacement current. Doing so was possible for one of three reasons: The fields were constant in time so that time derivatives were zero; the materials were good conductors of electricity so that the displacement current was small compared to the conduction current; or the displacement current was confined to small regions of space that did not need to be considered explicitly (as in capacitors). This constant field situation is straightforward, but the other two warrant a little more consideration. For a good conductor, a metal, the conductivity is of the order 10^8 S/m for frequencies below the far infrared. Thus the conduction current is of the order $10^8 \mathbf{E}$. The magnitude of the displacement current is dominated by the factor $\epsilon\omega$. Since $\epsilon_0 = 8.854 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2$, this factor is small except at very high (far infrared) frequencies where the simple discussion breaks down for other reasons. For frequencies up to, say, 10^{11} Hz the displacement current in metals can be ignored. In the case of good dielectrics, the conduction current is very small or zero; the displacement current can never be ignored. Even at 60 Hz, all of the current passing through a capacitor in an ac circuit is displacement current. It was not necessary to consider the displacement current explicitly in Chapter 13 because the time-varying fields in the interior of a capacitor were not examined when ac circuits were analyzed.

We turn now to the examination of the full set of Maxwell's equations and their implications.

MAXWELL'S EQUATIONS AND THEIR EMPIRICAL BASIS

Maxwell's equations consist of Eq. (16-10) plus three equations with which we are already familiar, namely:

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \quad (16-10)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (11-6), (16-11)$$

$$\nabla \cdot \mathbf{D} = \rho \quad (4-29), (16-12)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (8-30), (16-13)$$

Each of these equations represents a generalization of certain experimental observations: Eq. (16-10) represents an extension of Ampere's law; Eq. (16-11) is the differential form of Faraday's law of electromagnetic induction; Eq. (16-12) is Gauss's law, which in turn derives from Coulomb's

law; and Eq. (16-13) is usually said to represent the fact that single magnetic poles have never been observed.

It is clear that Maxwell's equations represent mathematical expressions of certain experimental results. In this light it is apparent that they cannot be proved; however, the applicability to any situation can be verified. As a result of extensive experimental work, Maxwell's equations are now believed to apply to all macroscopic situations, and they are used, much like conservation of momentum, as physical principles. They are the fundamental equations governing the electromagnetic fields produced by the source charge and current densities ρ and \mathbf{J} . If material bodies are present, in order to use the Maxwell equations, one must also know the applicable constitutive equations, $\mathbf{D} = \mathbf{D}(\mathbf{E})$ and $\mathbf{H} = \mathbf{H}(\mathbf{B})$ —either experimentally or from a microscopic theory of the particular kind of matter. The current density \mathbf{J} in a material includes a contribution given by a third constitutive equation, $\mathbf{J} = \mathbf{J}(\mathbf{E})$, which must likewise be known experimentally or theoretically. Coupled with the Lorentz force equation, $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$, which describes the action of the fields on charged particles, this set of laws gives the complete classical description of electromagnetically interacting particles.

We have just seen that the displacement current introduced in the preceding section is necessary in order to have charge conservation, and that when it is included in the Maxwell equations, they already imply the continuity equation, so that the latter need not be added to the set of fundamental equations. The Maxwell equations have two further interesting consequences, which are developed in the following sections. These will also be seen to depend crucially on the displacement current.

16-3 ELECTROMAGNETIC ENERGY

It was shown in Chapter 6 that the quantity

$$U_E = \frac{1}{2} \int_V \mathbf{E} \cdot \mathbf{D} \, dv \quad (16-14)$$

can be identified with the electrostatic potential energy of the system of charges producing the electric field. This was derived by computing the work done in establishing the field. In a similar way

$$U_M = \frac{1}{2} \int_V \mathbf{H} \cdot \mathbf{B} \, dv \quad (16-15)$$

was identified, in Chapter 12, with the energy stored in the magnetic field. The question of the applicability of these expressions to nonstatic situations now arises.

If the scalar product of Eq. (16-10) with \mathbf{E} is taken, and the resulting equation is subtracted from the scalar product of Eq. (16-11) with \mathbf{H} , the

resulting equation is

$$\mathbf{H} \cdot \nabla \times \mathbf{E} - \mathbf{E} \cdot \nabla \times \mathbf{H} = -\mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} - \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} - \mathbf{E} \cdot \mathbf{J} \quad (16-16)$$

The left side of this expression can be converted into a divergence by using the identity

$$\nabla \cdot (\mathbf{F} \times \mathbf{G}) = \mathbf{G} \cdot \nabla \times \mathbf{F} - \mathbf{F} \cdot \nabla \times \mathbf{G}$$

to obtain

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = -\mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} - \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} - \mathbf{E} \cdot \mathbf{J} \quad (16-17)$$

If, in the medium to which Eq. (16-17) is applied, $\mathbf{D}(t)$ is proportional to $\mathbf{E}(t)$ and $\mathbf{B}(t)$ is proportional to $\mathbf{H}(t)$ with constants of proportionality that do not depend explicitly on the time,* then the time derivatives on the right

* By proportionality we mean that g , ϵ , and μ do not depend explicitly on the field or on the time values. Such a medium is called linear and nondispersive. Both linearity and dispersion will be discussed further in Chapter 19. Chapters 17 and 18 are concerned with monochromatic radiation for which dispersion need not be considered explicitly. Also, we refer exclusively to linear media in these chapters. Even a monochromatic wave propagating in a *nonlinear* medium generates harmonic frequencies; we do not consider this phenomenon although it has significant practical importance. When additional restrictions are required they will be explicitly noted.

It should also be noted that anisotropy alone does not invalidate the expressions

$$\mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} = \frac{1}{2} \frac{\partial}{\partial t} (\mathbf{E} \cdot \mathbf{D}) \quad \text{and} \quad \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} = \frac{1}{2} \frac{\partial}{\partial t} (\mathbf{H} \cdot \mathbf{B})$$

In the case of anisotropic media, the relationship between \mathbf{E} and \mathbf{D} can be written as

$$D_i = \sum_{j=1}^3 \epsilon_{ij} E_j$$

Consequently,

$$\frac{1}{2} \frac{\partial}{\partial t} (\mathbf{E} \cdot \mathbf{D}) = \frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^3 \epsilon_{ij} \left(E_i \frac{\partial E_j}{\partial t} + \frac{\partial E_i}{\partial t} E_j \right)$$

A simple argument based on the conservation of energy (Wooster, *Crystal Physics*, Cambridge University Press, 1938, p. 277) shows that $\epsilon_{ij} = \epsilon_{ji}$. Using this result to interchange i and j in the last term, we have

$$\frac{1}{2} \frac{\partial}{\partial t} (\mathbf{E} \cdot \mathbf{D}) = \sum_{i=1}^3 \sum_{j=1}^3 E_i \epsilon_{ij} \frac{\partial E_j}{\partial t}$$

If $\{\epsilon_{ij}\}$ is a set of constants independent of \mathbf{E} and of t , then

$$\frac{1}{2} \frac{\partial}{\partial t} (\mathbf{E} \cdot \mathbf{D}) = \sum_{i=1}^3 E_i \frac{\partial}{\partial t} \sum_{j=1}^3 \epsilon_{ij} E_j = \sum_{i=1}^3 E_i \frac{\partial D_i}{\partial t} = \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t}$$

Thus it is seen that anisotropy alone does not restrict the derivation.

can be written as

$$\mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} = \mathbf{E} \cdot \frac{\partial}{\partial t} \epsilon \mathbf{E} = \frac{1}{2} \epsilon \frac{\partial}{\partial t} \mathbf{E}^2 = \frac{\partial}{\partial t} \frac{1}{2} \mathbf{E} \cdot \mathbf{D}$$

and

$$\mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} = \mathbf{H} \cdot \frac{\partial}{\partial t} \mu \mathbf{H} = \frac{1}{2} \mu \frac{\partial}{\partial t} \mathbf{H}^2 = \frac{\partial}{\partial t} \frac{1}{2} \mathbf{H} \cdot \mathbf{B}$$

Using this relationship, Eq. (16-17) takes the form

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = -\frac{\partial}{\partial t} \frac{1}{2} (\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H}) - \mathbf{J} \cdot \mathbf{E} \quad (16-18)$$

The first term on the right is the time derivative of the sum of the electric and magnetic energy densities; the second term is in many cases, in particular if $\mathbf{J} = g\mathbf{E}$, just the negative of the Joule heating rate per unit volume.

Integrating over a fixed volume V bounded by the surface S gives

$$\int_V \nabla \cdot (\mathbf{E} \times \mathbf{H}) dv = -\frac{d}{dt} \int_V \frac{1}{2} (\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H}) dv - \int_V \mathbf{J} \cdot \mathbf{E} dv \quad (16-19)$$

Applying the divergence theorem to the left side, we obtain

$$\oint_S \mathbf{E} \times \mathbf{H} \cdot \mathbf{n} da = -\frac{d}{dt} \int_V \frac{1}{2} (\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H}) dv - \int_V \mathbf{J} \cdot \mathbf{E} dv$$

Rewriting this equation

$$-\int_V \mathbf{J} \cdot \mathbf{E} dv = \frac{d}{dt} \int_V \frac{1}{2} (\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H}) dv + \oint_S \mathbf{E} \times \mathbf{H} \cdot \mathbf{n} da \quad (16-20)$$

makes it clear that the $\mathbf{J} \cdot \mathbf{E}$ term is comprised of two parts: the rate of change of electromagnetic energy stored in V , and a surface integral. The left side of Eq. (16-20) is the power transferred *into* the electromagnetic field through the motion of free charge in volume V . If there are no sources of emf in V , then the left side of Eq. (16-20) is negative and equal to minus the Joule heat production per unit time. In certain circumstances, however, the left side of Eq. (16-20) may be positive. Suppose that a charged particle q moves with constant velocity \mathbf{v} under the combined influence of mechanical, electric, and magnetic forces; the rate at which the mechanical force does work on the particle is

$$\mathbf{F}_m \cdot \mathbf{n} = -q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cdot \mathbf{v} = -q\mathbf{E} \cdot \mathbf{v}$$

But according to Eq. (7-4) the current density is defined by

$$\mathbf{J} = \sum_i N_i q_i \mathbf{v}_i$$

Thus the rate at which mechanical work is done (per unit volume) is

$$\sum_i N_i \mathbf{F}_m \cdot \mathbf{v}_i = -\mathbf{E} \cdot \mathbf{J}$$

and this power density is transferred into the electromagnetic field.

Since the surface integral in (16-20) involves only the electric and magnetic fields, it is feasible to interpret this term as the rate of energy flow across the closed surface. Equation (16-20) thus expresses the conservation of energy in a fixed volume V .

Let us return to the corresponding differential equation (16-18), which will express the local *conservation of energy* at a point. If we make the abbreviations

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \quad (16-21)$$

$$u = \frac{1}{2}(\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H}) \quad (16-22)$$

where \mathbf{S} is called the *Poynting vector*, then Eq. (16-18) implies that at any point

$$\nabla \cdot \mathbf{S} + \frac{\partial u}{\partial t} = -\mathbf{J} \cdot \mathbf{E} \quad (16-23)$$

There is no doubt that $\mathbf{J} \cdot \mathbf{E}$ is the work done by the local field on charged particles per unit volume. Previously u was interpreted as the energy density of the electric and magnetic fields. If $\nabla \cdot \mathbf{S} = 0$, Eq. (16-23) would therefore express local conservation of energy: The rate of change of field energy equals power dissipation per unit volume at each point. If, on the other hand, $\nabla \cdot \mathbf{S} \neq 0$ but $\mathbf{J} \cdot \mathbf{E} = 0$ (as for example in a nonconducting medium), then

$$\nabla \cdot \mathbf{S} + \frac{\partial u}{\partial t} = 0 \quad (16-24)$$

This expression has exactly the mathematical form of the equation of continuity (16-7) for charge, except that the energy density u takes the place of the charge density ρ . If Eq. (16-24) is still to describe conservation of energy, $\nabla \cdot \mathbf{S}$ must represent the divergence of an energy current density or, in other words, of a rate of energy flow per unit area. One usually treats $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ itself as the local energy flow per unit time per unit area.* We shall use these interpretations of u and \mathbf{S} , while recognizing that it is only their time derivative and divergence, respectively, whose interpretations are

* There is a continuing controversy over this point. For a good discussion, see W. H. Furry, *American Journal of Physics*, vol. 37, p. 621 (1969).

directly required by Maxwell's equations. It is usually only the latter that are physically measurable anyway. In any case, Eq. (16-23) expresses energy conservation locally, as does Eq. (16-20) in integral form.

16-4 THE WAVE EQUATION

One of the most important consequences of Maxwell's equations is the equations for electromagnetic wave propagation in a linear medium. The wave equation for \mathbf{H} is derived by taking the curl of (16-10):

$$\nabla \times \nabla \times \mathbf{H} = \nabla \times \mathbf{J} + \nabla \times \frac{\partial \mathbf{D}}{\partial t}$$

Putting $\mathbf{D} = \epsilon \mathbf{E}$ and $\mathbf{J} = g \mathbf{E}$ and assuming g and ϵ to be constants, we obtain

$$\nabla \times \nabla \times \mathbf{H} = g \nabla \times \mathbf{E} + \epsilon \frac{\partial}{\partial t} \nabla \times \mathbf{E}$$

The order of time and space differentiation can be interchanged if \mathbf{E} is a sufficiently well-behaved function, as we assume to be the case. Equation (16-11) can now be used to eliminate $\nabla \times \mathbf{E}$, yielding

$$\nabla \times \nabla \times \mathbf{H} = -g\mu \frac{\partial \mathbf{H}}{\partial t} - \epsilon\mu \frac{\partial^2 \mathbf{H}}{\partial t^2} \quad (16-25)$$

where $\mathbf{B} = \mu \mathbf{H}$, with μ a constant, has been used. The operator identity

$$\nabla \times \nabla \times = \nabla \nabla \cdot - \nabla^2 \quad (16-26)$$

is now used to obtain

$$\nabla \nabla \cdot \mathbf{H} - \nabla^2 \mathbf{H} = -g\mu \frac{\partial \mathbf{H}}{\partial t} - \epsilon\mu \frac{\partial^2 \mathbf{H}}{\partial t^2} \quad (16-27)$$

Since μ is a constant,

$$\nabla \cdot \mathbf{H} = \frac{1}{\mu} \nabla \cdot \mathbf{B} = 0$$

Consequently, the first term on the left side of Eq. (16-27) vanishes. The final wave equation is

$$\nabla^2 \mathbf{H} - \epsilon\mu \frac{\partial^2 \mathbf{H}}{\partial t^2} - g\mu \frac{\partial \mathbf{H}}{\partial t} = 0 \quad (16-28)$$

The vector \mathbf{E} satisfies the same wave equation, as is readily seen by first taking the curl of Eq. (16-11):

$$\nabla \times \nabla \times \mathbf{E} = -\nabla \times \frac{\partial \mathbf{B}}{\partial t}$$

Using Eq. (16-10) to eliminate the magnetic field and treating g , μ , and ϵ as constants yields

$$\nabla \times \nabla \times \mathbf{E} = -g\mu \frac{\partial \mathbf{E}}{\partial t} - \epsilon\mu \frac{\partial^2 \mathbf{E}}{\partial t^2}$$

Applying the vector identity (16-26) and restricting the application of the equation to a charge-free medium so that $\nabla \cdot \mathbf{D} = 0$ gives

$$\nabla^2 \mathbf{E} - \epsilon\mu \frac{\partial^2 \mathbf{E}}{\partial t^2} - g\mu \frac{\partial \mathbf{E}}{\partial t} = 0 \quad (16-29)$$

The wave equations derived above govern the electromagnetic field in a homogeneous, linear medium in which the charge density is zero, whether this medium is conducting or nonconducting. However, it is not enough that these equations be satisfied; Maxwell's equations must also be satisfied. It is clear that Eqs. (16-28) and (16-29) are a necessary consequence of Maxwell's equations, but the converse is not true. In solving the wave equations, special care must be used to obtain solutions to Maxwell's equations.

6-5 MONOCHROMATIC WAVES

Monochromatic waves are waves in which all of the fields are characterized by a single frequency. In this case, we can solve Eq. (16-29) to find \mathbf{E} and then use Maxwell's equations and the constitutive relations to find the other fields. This obviously guarantees that Maxwell's equations will be satisfied.

The methods of complex analysis provide a convenient way of implementing this procedure. The time dependence of the field (for definiteness we take the vector \mathbf{E}) is taken to be as $e^{-i\omega t}$, so that

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r})e^{-i\omega t} \quad (16-30)$$

It must be remembered that the physical electric field is obtained by taking the real part* of Eq. (16-30). Furthermore, $\mathbf{E}(\mathbf{r})$ is in general complex, so that the actual electric field is proportional to $\cos(\omega t + \phi)$ where ϕ is the phase of $\mathbf{E}(\mathbf{r})$. Using Eq. (16-30) in Eq. (16-29) gives

$$e^{-i\omega t} \{ \nabla^2 \mathbf{E} + \omega^2 \epsilon \mu \mathbf{E} + i\omega g \mu \mathbf{E} \} = 0 \quad (16-31)$$

* As discussed in Chapter 13, one goes from the convenient mathematical description in terms of complex variables to the physical quantities by taking either the real or imaginary part of the complex quantity. The choice of real or imaginary part is quite arbitrary. The two choices differ only by a phase shift of $\pi/2$; however, one must always make the same choice in a given problem. In this and the following chapters, the real part of complex quantities will represent the physical quantities unless otherwise explicitly noted.

for the equation governing the spatial variation of the electric field (the common factor $e^{-i\omega t}$ can, of course, be dropped). The next task is to solve Eq. (16-31) in various special cases of interest to determine the spatial variation of the electromagnetic field. This task will be treated in the following chapter; here let us merely look at some of the simplest possible cases.

First, suppose that the "medium" is empty space, so that $g = 0$, $\epsilon = \epsilon_0$, $\mu = \mu_0$. Further, suppose $\mathbf{E}(\mathbf{r})$ varies in only one dimension, say the z -direction, and is independent of x and y . Then Eq. (16-31) becomes

$$\frac{d^2\mathbf{E}(z)}{dz^2} + (\omega/c)^2\mathbf{E} = 0 \quad (16-32)$$

where we have written

$$\epsilon_0\mu_0 = 1/c^2$$

as was suggested in Chapter 8 for dimensional reasons; c has the dimensions of a velocity. This equation (Helmholtz equation) is mathematically the same as the harmonic oscillator equation and has solutions

$$\mathbf{E}(z) = \mathbf{E}_0 e^{\pm i\kappa z}$$

where \mathbf{E}_0 is a constant vector and

$$\kappa = \omega/c \quad (16-33)$$

Putting this $\mathbf{E}(\mathbf{r})$ into Eq. (16-30), we get the full solution

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 e^{-i(\omega t \mp \kappa z)} \quad (16-34)$$

or, taking the real part,

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 \cos(\omega t \mp \kappa z) \quad (16-35a)$$

With (16-33), an equivalent form is

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 \cos \omega(t \mp z/c) \quad (16-35b)$$

Eq. (16-35) represents a sinusoidal wave traveling to the right or left in the z -direction (depending on whether the minus or plus sign is used). The velocity of propagation of the wave is c . If light is a form of electromagnetic radiation, then the Maxwell equations *predict* that $c = 1/\sqrt{\epsilon_0\mu_0} = 2.9979 \times 10^8$ m/s is the velocity of light in vacuum. Although this result is what we anticipated, when Maxwell first announced the result, it was considered a great triumph of his theory, since up to that time the electromagnetic nature of light was only a speculation. The form of Eq. (16-35a) shows that the wave frequency is $f = \omega/2\pi$ and the wavelength is $\lambda = 2\pi/\kappa$. Thus Eq. (16-33) is the familiar result for a wave,

$$\lambda f = c$$

In a nonconducting, nonmagnetic dielectric, we still have $g = 0$,

$\mu = \mu_0$, but now $\epsilon = K\epsilon_0$. The preceding derivation will carry through just the same, except that now Eq. (16-33) becomes

$$\kappa = \sqrt{K} \omega / c \quad (16-33a)$$

Defining $n = \sqrt{K}$, we see that the results are the same as in vacuum,* except that the velocity of wave propagation is now c/n instead of c . The quantity n is called the *index of refraction* of the dielectric medium; for vacuum $n = 1$. This accounts for refractive effects in transparent materials, as will be seen in Chapter 18.

If the medium is conducting, $g > 0$, the third term in Eq. (16-31) must be retained. When g is small the result will be merely that the wave is damped, as we shall see in the next chapter. By small g , we mean that the third term of Eq. (16-31) is small compared with the second term, which led to the wave solution, or

$$\omega g \mu \ll \omega^2 \epsilon \mu, \quad g \ll \omega \epsilon$$

In the other extreme when $g \gg \omega \epsilon$, we may neglect the second term of Eq. (16-31). Again restricting attention to the one-dimensional case, we get

$$\frac{d^2 \mathbf{E}(z)}{dz^2} + i \omega g \mu \mathbf{E} = 0$$

We can make the coefficient of \mathbf{E} real if we assume that $\alpha = i\omega$ is real or, in other words, that the frequency is imaginary. Then, if

$$\kappa = \sqrt{\alpha g \mu}$$

the spatial dependence $\mathbf{E}(\mathbf{r})$ of the solution is just the same as before. The difference is that the time dependence (16-30) becomes

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r}) e^{-\alpha t}$$

That is, the field simply decays exponentially with time, instead of oscillating in a wave-like manner. The transition between the decaying and the wave behavior occurs when

$$|\omega| = |\alpha| \cong \left| \frac{g}{\epsilon} \right| = 1/t_c$$

where t_c is the relaxation time of the material that was discussed in Chapter 7. (We repeat that caution is needed when this condition is applied to a metal, since g/ϵ is itself strongly dependent on ω .)

Finally, by tracing the derivation of Eq. (16-31) back to Maxwell's equations, we notice that the second term, or $\partial^2 \mathbf{E} / \partial t^2$ in Eq. (16-29),

* If the relative permeability is not 1, then Eq. (16-33a) becomes $\kappa = \sqrt{K K_m} \omega / c$ and $n = \sqrt{K K_m}$. For many purposes, it is sufficient to consider only the case $K_m = 1$, and we shall do so in the remainder of the book.

derives from the displacement current $\partial \mathbf{D} / \partial t$ in Eq. (16-10), whereas the third term, or $\partial \mathbf{E} / \partial t$ in Eq. (16-29), derives from the transport current \mathbf{J} in Eq. (16-10). Thus the very existence of electromagnetic wave propagation depends on Maxwell's introduction of the displacement current. Without it, only exponential decay of the fields could occur.

16-6 BOUNDARY CONDITIONS

The boundary conditions that must be satisfied by the electric and magnetic fields at an interface between two media are deduced from Maxwell's equations exactly as in the static case. The most straightforward and universal boundary condition applies to the magnetic induction \mathbf{B} , which satisfies the Maxwell equation

$$\nabla \cdot \mathbf{B} = 0 \quad (16-36)$$

At any interface between two media a pillbox-like surface may be constructed as shown in Fig. 16-2. The divergence theorem may be applied to the divergence of \mathbf{B} over the volume enclosed by this surface, to obtain

$$\oint_S \mathbf{B} \cdot \mathbf{n} \, da = \int_{S_1} \mathbf{B} \cdot \mathbf{n}_1 \, da + \int_{S_2} \mathbf{B} \cdot \mathbf{n}_2 \, da + \int_{S_3} \mathbf{B} \cdot \mathbf{n}_3 \, da = 0 \quad (16-37)$$

If \mathbf{B} is bounded, letting h approach zero causes the last term to vanish and S_1 to approach S_2 geometrically. Taking account of the opposite directions of \mathbf{n}_1 and \mathbf{n}_2 , it is quickly concluded that

$$B_{1n} = B_{2n} \quad (16-38)$$

exactly as in the static case.

The tangential component of the electric field can be treated in an equally simple way. The basic equation is again one of Maxwell's equations:

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \quad (16-39)$$

Integration of this equation over the surface bounded by a rectangular loop

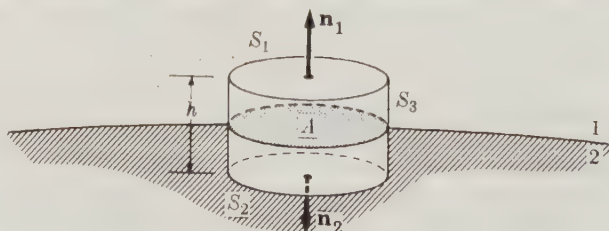
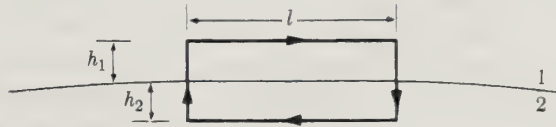


FIGURE 16-2

A pillbox-shaped surface at the interface between two media may be used to obtain boundary conditions on the field vectors.

FIGURE 16-3

The rectangular path shown at the interface between two media may be used to obtain boundary conditions on the field vectors.



such as that shown in Fig. 16-3 yields

$$\int_S \nabla \times \mathbf{E} \cdot \mathbf{n} \, da = - \int_S \frac{\partial \mathbf{B}}{\partial t} \cdot \mathbf{n} \, da \quad (16-40)$$

and applying Stokes's theorem to the left side gives

$$lE_{1t} - lE_{2t} + h_1E_{1n} + h_2E_{2n} - h_1E'_{1n} - h_2E'_{2n} = - \int_S \frac{\partial \mathbf{B}}{\partial t} \cdot \mathbf{n} \, da \quad (16-41)$$

If the loop is now shrunk by letting h_1 and h_2 go to zero, the last four terms on the left vanish, as does the right-hand side, provided only that $\partial \mathbf{B} / \partial t$ is bounded. The resulting equation contains l as a common factor that can be dropped, giving

$$E_{1t} = E_{2t} \quad (16-42)$$

Thus the tangential component of \mathbf{E} must be continuous across the interface.

The boundary condition on the normal component of the electric displacement is more complex; however, it too is derived from one of Maxwell's equations. The appropriate equation in this case is

$$\nabla \cdot \mathbf{D} = \rho \quad (16-43)$$

If we construct a pillbox-shaped volume, as shown in Fig. 16-2, and integrate Eq. (16-43) over this volume, we obtain

$$\int_V \nabla \cdot \mathbf{D} \, dv = \int_V \rho \, dv$$

Applying the divergence theorem and letting h go to zero, we find

$$(D_{1n} - D_{2n}) = \sigma \quad (16-44)$$

where σ is the surface charge density on the interface. The fact that, in general, σ is not zero introduces some complexity in this boundary

condition; however, by noting that charge must be conserved—that is, that

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t} \quad (16-45)$$

we can make certain simplifications. If we integrate this equation as we did Eq. (16-43) and shrink the pillbox in the same way, we obtain

$$J_{1n} - J_{2n} = -\frac{\partial \sigma}{\partial t} \quad (16-46)$$

If only monochromatic radiation is considered, the surface charge density must vary as $e^{-i\omega t}$; therefore the right side of Eq. (16-46) can be written as $i\omega\sigma$. Using the constitutive relations $\mathbf{D} = \epsilon\mathbf{E}$, $\mathbf{J} = g\mathbf{E}$ puts equations (16-44) and (16-46) in the form

$$\epsilon_1 E_{1n} - \epsilon_2 E_{2n} = \sigma \quad (16-47)$$

$$g_1 E_{1n} - g_2 E_{2n} = i\omega\sigma \quad (16-48)$$

Several cases of practical interest may be noted. If σ is zero, then

$$\frac{\epsilon_1}{g_1} = \frac{\epsilon_2}{g_2}$$

which can be true for appropriately chosen materials or, alternatively, if $g_1 = g_2 = 0$, or ∞ . The case where both conductivities are infinite is not of great interest; however, the case where both conductivities vanish is approximately realized at the boundary between two good dielectrics. If σ is not zero, which is perhaps a more common case, then it may be eliminated from Eqs. (16-47) and (16-48). The result of this elimination is

$$\left(\epsilon_1 + i\frac{g_1}{\omega}\right)E_{1n} - \left(\epsilon_2 + i\frac{g_2}{\omega}\right)E_{2n} = 0 \quad (16-49)$$

A final interesting case occurs when one conductivity, say g_2 , is infinite. In this case, E_{2n} must vanish and E_{1n} must equal σ/ϵ_1 in order for Eqs. (16-48) and (16-47) to be satisfied.

The final boundary condition is that imposed on the tangential component of the magnetic intensity \mathbf{H} . This boundary condition is obtained by integrating the Maxwell equation

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J} \quad (16-50)$$

over the area enclosed by a loop such as that shown in Fig. 16-3. If this is done and the loop is shrunk as before, the resulting boundary condition is

$$H_{1t} - H_{2t} = j_{\perp} \quad (16-51)$$

where j_{\perp} is the component of the surface current density perpendicular to

the direction of the \mathbf{H} -component that is being matched. The idea of a surface current density is closely analogous to that of a surface charge density—it represents a finite current in an infinitesimal layer. The surface current density is zero unless the conductivity is infinite; hence, for finite conductivity

$$H_{1t} = H_{2t} \quad (16-52)$$

That is, unless one medium has infinite conductivity the tangential component of \mathbf{H} is continuous. If the conductivity of medium 2 is infinite, then, as has already been shown, $E_{2n} = 0$. A more general result can be obtained by considering the Maxwell equation (16-50) as applied to medium 2:

$$\nabla \times \mathbf{H}_2 - \frac{\partial \mathbf{D}_2}{\partial t} = \mathbf{J}_2 \quad (16-53)$$

Using the constitutive relations and assuming that \mathbf{E}_2 varies with time as $e^{-i\omega t}$ yields

$$\mathbf{E}_2 = \frac{1}{g_2 - i\omega\epsilon_2} \nabla \times \mathbf{H}_2 \quad (16-54)$$

If the reasonable assumption that \mathbf{H}_2 is both bounded and differentiable is made, then Eq. (16-54) implies that \mathbf{E}_2 is zero in a medium of infinite conductivity. With the same assumptions as were made previously,

$$\mathbf{H}_2 = \frac{1}{i\omega\mu_2} \nabla \times \mathbf{E}_2 \quad (16-55)$$

and the vanishing of \mathbf{E}_2 also implies the vanishing of \mathbf{H}_2 . If \mathbf{H}_2 vanishes, then the boundary condition on the tangential component of \mathbf{H} at an interface at which one medium has infinite conductivity is

$$H_{1t} = j_{\perp} \quad (16-56)$$

The boundary conditions have now been obtained; for convenient reference they are tabulated in Table 16-1 for $g = 0$, $g = \infty$, and arbitrary g .

TABLE 16-1
Boundary Conditions

g	E_t	D_n	H_t	B_n
$g_1 = g_2 = 0$	$E_{1t} = E_{2t}$	$D_{1n} = D_{2n}$	$H_{1t} = H_{2t}$	$B_{1n} = B_{2n}$
$g_2 = \infty$	$E_{2t} = 0$ $E_{1t} = 0$	$D_{2n} = 0$ $D_{1n} = \sigma$	$H_{2t} = 0$ $H_{1t} = j_{\perp}$	$B_{2n} = 0$ $B_{1n} = 0$
$g_1, g_2 \text{ arb. } \neq \infty$	$E_{1t} = E_{2t}$	$(\epsilon_1 + i\frac{g_1}{\omega})E_{1n}$ $= (\epsilon_2 + i\frac{g_2}{\omega})E_{2n}$	$H_{1t} = H_{2t}$	$B_{1n} = B_{2n}$

16-7 THE WAVE EQUATION WITH SOURCES

In the preceding sections, it was shown that Maxwell's equations predict the propagation of monochromatic electromagnetic waves through a linear medium and also that the fields must match at an interface between two different media according to appropriate boundary conditions. It was assumed that the charge density ρ in the medium was zero and that the only current density \mathbf{J} arose from the passive response of an ohmic medium to the electric field of the wave. We did not inquire how these waves were produced, but we shall now find that they are fields produced by distant source charges that undergo accelerated motion.

The problem now is to consider prescribed charge and current distributions, $\rho(\mathbf{r}, t)$ and $\mathbf{J}(\mathbf{r}, t)$, and find the fields produced by them. There are several ways of approaching the problem, of which the most fruitful is the potential approach, which is developed analogously to the procedures used in electrostatics and magnetostatics. Since the magnetic induction has zero divergence, it may always be represented as the curl of a vector potential. That is,

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (16-57)$$

Using this expression for \mathbf{B} in Eq. (16-11) gives

$$\nabla \times \mathbf{E} + \frac{\partial}{\partial t} \nabla \times \mathbf{A} = 0 \quad (16-58)$$

Assuming sufficient continuity of the fields to interchange the spatial and temporal differentiations, Eq. (16-58) can be written

$$\nabla \times \left[\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right] = 0 \quad (16-59)$$

The vector $\mathbf{E} + \partial \mathbf{A} / \partial t$ thus has zero curl and can be written as the gradient of a scalar:

$$\mathbf{E} = -\nabla \varphi - \frac{\partial \mathbf{A}}{\partial t} \quad (16-60)$$

Equations (16-57) and (16-60) give the electric and magnetic fields in terms of a vector potential \mathbf{A} and a scalar potential φ . These potentials satisfy wave equations which are very similar to those satisfied by the fields. The wave equation for \mathbf{A} is derived by substituting the expressions given in Eqs. (16-57) and (16-60) for \mathbf{B} and \mathbf{E} into Eq. (16-10), with the result

$$\frac{1}{\mu} \nabla \times \nabla \times \mathbf{A} + \epsilon \frac{\partial}{\partial t} \left[\nabla \varphi + \frac{\partial \mathbf{A}}{\partial t} \right] = \mathbf{J} \quad (16-61)$$

Writing $\nabla \nabla \cdot - \nabla^2$ for $\nabla \times \nabla \times$ and multiplying by μ gives

$$-\nabla^2 \mathbf{A} + \epsilon \mu \frac{\partial^2 \mathbf{A}}{\partial t^2} + \nabla \nabla \cdot \mathbf{A} + \epsilon \mu \nabla \frac{\partial \varphi}{\partial t} = \mu \mathbf{J} \quad (16-62)$$

Until now only the curl of \mathbf{A} has been specified; the choice of the divergence of \mathbf{A} is still arbitrary. It is clear from Eq. (16-62) that imposing the so-called *Lorentz condition* (or Lorentz gauge, cf. p. 407),

$$\nabla \cdot \mathbf{A} + \epsilon \mu \frac{\partial \varphi}{\partial t} = 0 \quad (16-63)$$

results in a considerable simplification. If this condition is satisfied, then \mathbf{A} satisfies the wave equation

$$\nabla^2 \mathbf{A} - \epsilon \mu \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu \mathbf{J} \quad (16-64)$$

Furthermore, using Eq. (16-60) in (16-12) gives

$$-\epsilon \left[\nabla \cdot \nabla \varphi + \nabla \cdot \frac{\partial \mathbf{A}}{\partial t} \right] = \rho \quad (16-65)$$

Interchanging the order of the divergence and the time derivative operating on \mathbf{A} and using the Lorentz condition (Eq. 16-63) leads to

$$\nabla^2 \varphi - \epsilon \mu \frac{\partial^2 \varphi}{\partial t^2} = -\frac{1}{\epsilon} \rho \quad (16-66)$$

Thus, by imposing the Lorentz condition, both the scalar and vector potentials are forced to satisfy inhomogeneous wave equations of similar forms.

The problem of finding the general solution of the inhomogeneous scalar wave equation is analogous to finding the general solution of Poisson's equation. In the latter case, it will be recalled, the general solution consists of a particular solution of the inhomogeneous equation plus a general solution of the homogeneous equation. The inclusion of the solutions of the homogeneous equation provides the means for satisfying arbitrary appropriate boundary conditions, while the particular solution ensures that the total function satisfies the inhomogeneous equation. Exactly the same considerations apply to the inhomogeneous wave equation—the general solution consists of a particular solution plus a general solution of the homogeneous equation. Methods for finding certain solutions of the homogeneous equation will be treated in Chapter 17. These methods may be extended and supplemented to yield solutions to almost any solvable problem. Approximate methods are available for problems that cannot be solved in terms of known functions. It remains, then, to find the needed particular solution of the inhomogeneous equation. The inhomogeneous scalar wave equation (16-66) reduces in the static case,

$\partial\varphi/\partial t = 0$, to Poisson's equation, a particular solution of which we know from Eq. (3-1) (for vacuum):

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int_V \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dv' \quad (16-67)$$

The vector wave equation (16-64) has an analogous solution in the static (vacuum) case, given in Eq. (8-61). Unfortunately, we do not get solutions for the time-dependent case simply by inserting $\rho(\mathbf{r}', t)$ and $\mathbf{J}(\mathbf{r}', t)$ into the static solutions, for reasons that we shall see.

Let us rewrite Eq. (16-66) for vacuum, using $1/\sqrt{\epsilon\mu} = c/n$, with the refractive index $n = 1$:

$$\nabla^2\varphi - \frac{1}{c^2} \frac{\partial^2\varphi}{\partial t^2} = -\frac{\rho}{\epsilon_0} \quad (16-68)$$

can be solved most readily by finding the solution for a point charge, and then later summing over all the charge elements $\rho \Delta v$ in the appropriate charge distribution. The most convenient location for the point charge is at the origin of coordinates. Thus the equation

$$\nabla^2\varphi - \frac{1}{c^2} \frac{\partial^2\varphi}{\partial t^2} = 0 \quad (16-69)$$

must be satisfied everywhere except at the origin, whereas in a small volume Δv surrounding the origin,

$$\int_{\Delta v} dv \left[\nabla^2\varphi - \frac{1}{c^2} \frac{\partial^2\varphi}{\partial t^2} \right] = -\frac{1}{\epsilon_0} q(t) \quad (16-70)$$

must be satisfied. The function $q(t)$ is supposed to represent an assumed point charge of magnitude q that is located at the origin at time t , simply as a mathematical device to solve the equation, without any assumption about where an actual conserved charge was at an earlier or later time. (It does not represent a physical moving point charge, and the resulting solution for φ is not the correct potential for a moving point charge. The latter is more complicated and will be treated in Chapter 21.) It is clear from the symmetry of the charge distribution that the spatial dependence of φ must be only on r . With this clue, an attempt to solve Eq. (16-69) may be made. Since φ does not depend on either the azimuthal angle or the colatitude, Eq. (16-69) becomes

$$\frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial\varphi}{\partial r} - \frac{1}{c^2} \frac{\partial^2\varphi}{\partial t^2} = 0 \quad (16-71)$$

Now, by putting

$$\varphi(r, t) = \frac{\chi(r, t)}{r} \quad (16-72)$$

Eq. (16-71) is converted to

$$\frac{\partial^2 \chi}{\partial r^2} - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t^2} = 0 \quad (16-73)$$

This equation is the one-dimensional wave equation that is satisfied by any function of $r - ct$, or $r + ct$. To verify this, let

$$u = r - ct$$

and let $f(u)$ be any function of u that can be twice differentiated. Then

$$\frac{\partial f}{\partial r} = \frac{df}{du} \frac{\partial u}{\partial r} = \frac{df}{du}, \quad \frac{\partial^2 f}{\partial r^2} = \frac{d^2 f}{du^2} \frac{\partial u}{\partial r} = \frac{d^2 f}{du^2} \quad (16-74)$$

and

$$\frac{\partial f}{\partial t} = \frac{df}{du} \frac{\partial u}{\partial t} = -c \frac{df}{du}, \quad \frac{\partial^2 f}{\partial t^2} = c^2 \frac{d^2 f}{du^2} \quad (16-75)$$

Substituting the results of Eqs. (16-74) and (16-75) into Eq. (16-73) verifies that any function of $(r - ct)$ that is twice differentiable is a solution of Eq. (16-73). A similar calculation verifies that a function of $(r + ct)$ is a solution. Thus

$$\chi = f(r - ct) + g(r + ct) \quad (16-76)$$

is a very arbitrary solution of Eq. (16-73). It will be found that $g(r + ct)$ does not occur in our applications of the wave equation. For this reason, it will be dropped, and only the first term of Eq. (16-76) will be carried, since this procedure simplifies the ensuing equations and causes no particular omissions. It may be noted that $f(r - ct)$ represents a wave propagating outward from the source charge q at the origin, while $g(r + ct)$ represents a wave propagating inward toward the source charge from infinity. We keep the former and drop the latter for somewhat the same reason we would keep a plane wave solution propagating to the right if we were to the right of the source and drop the one propagating leftwards.

A spherically symmetric solution of Eq. (16-69),

$$\varphi = \frac{f(r - ct)}{r} \quad (16-77)$$

is now available. Furthermore, this solution contains an arbitrary function that may be chosen so that Eq. (16-70) is also satisfied. The proper choice is obtained by noting that for a static charge the potential compatible with Eqs. (16-69) and (16-70) is

$$\varphi = \frac{q}{4\pi\epsilon_0 r} \quad (16-78)$$

The two functional forms for φ given by Eqs. (16-77) and (16-78) may be

brought into concert by choosing

$$f(r - ct) = \frac{q(t - r/c)}{4\pi\epsilon_0} \quad (16-79)$$

The solution to Eqs. (16-69) and (16-70) is then

$$\varphi(r, t) = \frac{q(t - r/c)}{4\pi\epsilon_0 r} \quad (16-80)$$

With this result, we readily find that Eq. (16-66) is satisfied by

$$\varphi(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0} \int_V \frac{\rho(\mathbf{r}', t')}{|\mathbf{r} - \mathbf{r}'|} dv' \quad (16-81)$$

where $t' = t - |\mathbf{r} - \mathbf{r}'|/c$ is called the *retarded time*; φ is known as the *retarded scalar potential*.

The solution of Eq. (16-64) can be constructed in exactly the same way. The vectors \mathbf{A} and \mathbf{J} are first decomposed into rectangular components. The three resulting equations are closely analogous to Eq. (16-66), the x equation, for example, being

$$\nabla^2 A_x - \frac{1}{c^2} \frac{\partial^2 A_x}{\partial t^2} = -\mu_0 J_x \quad (16-82)$$

Each of these equations may be solved exactly as was Eq. (16-66), giving, for example,

$$A_x(\mathbf{r}, t) = \frac{\mu_0}{4\pi} \int_V \frac{J_x(\mathbf{r}', t')}{|\mathbf{r} - \mathbf{r}'|} dv' \quad (16-83)$$

These components are then combined to give

$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_0}{4\pi} \int_V \frac{\mathbf{J}(\mathbf{r}', t')}{|\mathbf{r} - \mathbf{r}'|} dv' \quad (16-84)$$

which is the *retarded vector potential*.

The physical interpretation of the retarded potentials is interesting. Equations (16-81) and (16-84) indicate that at a given point \mathbf{r} and a given time t the potentials are determined by the charge and current that existed at other points in space \mathbf{r}' at earlier times t' . The time appropriate to each source point is earlier than t by an amount equal to the time required to travel from source to field point \mathbf{r} with velocity c . If, for example, a charge

element located at the origin of coordinates were suddenly changed, then the effect of this change would not be felt at a distance r until a time r/c after the change was accomplished. The effect of the change propagates outward roughly as a spherical wavefront. (The actual situation is more complicated for a point charge because the charge density and the current density are intimately related through $\nabla \cdot \mathbf{J} + \partial\rho/\partial t = 0$; cf. Chapter 21.)

Having found the scalar and vector potentials, we find the fields by applying the gradient to φ , and the time derivative and curl to \mathbf{A} . These operations are in principle straightforward; however, it will be seen that they are relatively complicated in practice.

In the above procedure, it was essential to impose the Lorentz condition (16-63) on the potentials; otherwise it would not be the simple wave equations that they would have to satisfy. To see that we are always free to impose this condition, suppose that \mathbf{A} and φ are a particular choice of potential functions that give the correct \mathbf{E} - and \mathbf{B} -fields according to Eqs. (16-57) and (16-60). Then, if we were to choose new potentials

$$\mathbf{A}' = \mathbf{A} + \nabla\xi, \quad \varphi' = \varphi - \frac{\partial\xi}{\partial t} \quad (16-85)$$

these would give exactly the same \mathbf{E} - and \mathbf{B} -fields when substituted into Eqs. (16-57) and (16-60), whatever we use for the function ξ , which is completely arbitrary. This change to new but physically equivalent potentials is called a *gauge transformation*. Now substituting \mathbf{A}' and φ' into Eq. (16-63), we get, after rearranging terms, a scalar wave equation for ξ ,

$$\nabla^2\xi - \epsilon\mu \frac{\partial^2\xi}{\partial t^2} = -\left(\nabla \cdot \mathbf{A} + \epsilon\mu \frac{\partial\varphi}{\partial t}\right)$$

as the condition that \mathbf{A}' , φ' should satisfy the Lorentz condition. Thus, if the original potentials satisfied the Lorentz condition, the new ones will also, provided that ξ satisfies the homogeneous scalar wave equation. If the original \mathbf{A} , φ did not, we can still find new potentials that will, by choosing ξ as a solution of the inhomogeneous scalar wave equation with

$$\nabla \cdot \mathbf{A} + \epsilon\mu \frac{\partial\varphi}{\partial t}$$

as the source term. We have just seen how to find such a solution. A choice of potentials that satisfies the Lorentz condition is called the *Lorentz gauge*. Other choices of gauge (i.e., other choices of $\nabla \cdot \mathbf{A}$) are useful in other circumstances.

With the development of the retarded potentials the basic work on radiation is completed. It remains to apply this material to the solution of practical problems. These are the concern of the next five chapters.

16-8 SUMMARY

This chapter contains the foundations of all classical electromagnetic theory. **Maxwell's equations** are the differential equations that determine (together with boundary conditions for a particular situation) the fields produced by sources of charge and current:

$$\begin{aligned}\nabla \cdot \mathbf{B} &= 0, & \nabla \cdot \mathbf{D} &= \rho \\ \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} &= 0, & \nabla \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} &= \mathbf{J}\end{aligned}$$

The \mathbf{E} and \mathbf{B} fields are operationally defined by the **Lorentz force**

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

and the \mathbf{D} and \mathbf{H} fields are related to them by the **constitutive equations** of the medium, $\mathbf{D} = \mathbf{D}(\mathbf{E})$, $\mathbf{H} = \mathbf{H}(\mathbf{B})$.

The Maxwell equations have the following important consequences.

- Electric **charge is conserved**, according to the equation of continuity

$$\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0$$

- **Energy is conserved**, according to

$$\nabla \cdot \mathbf{S} + \frac{\partial u}{\partial t} = -\mathbf{J} \cdot \mathbf{E}$$

where the field energy density is (in a linear medium)

$$u = \frac{1}{2}(\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H})$$

and the energy flux per unit area is the **Poynting vector**

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}$$

- **Electromagnetic waves** propagate with the velocity of light, $c = 1/\sqrt{\epsilon_0\mu_0}$ in vacuum.

• **Boundary conditions** on the fields are determined at an interface between different media, the most important ones being that the tangential components of \mathbf{E} and \mathbf{H} are continuous.

- The fields \mathbf{E} and \mathbf{B} are derivable from **potential functions**:

$$\mathbf{B} = \nabla \times \mathbf{A}, \quad \mathbf{E} = -\nabla\varphi - \frac{\partial \mathbf{A}}{\partial t}$$

- The potentials satisfy the inhomogeneous wave equations

$$\nabla^2\varphi - \epsilon\mu \frac{\partial^2\varphi}{\partial t^2} = -\frac{1}{\epsilon}\rho, \quad \nabla^2\mathbf{A} - \epsilon\mu \frac{\partial^2\mathbf{A}}{\partial t^2} = -\mu\mathbf{J}$$

if the **Lorentz condition**

$$\nabla \cdot \mathbf{A} = -\epsilon\mu \frac{\partial \varphi}{\partial t}$$

is imposed. These will determine the generation of electromagnetic waves by prescribed charge and current distributions. Particular solutions (in vacuum) are

$$\varphi(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\mathbf{r}', t')}{|\mathbf{r} - \mathbf{r}'|} dv', \quad \mathbf{A}(\mathbf{r}, t) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{J}(\mathbf{r}', t')}{|\mathbf{r} - \mathbf{r}'|} dv'$$

where

$$t' = t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}$$

is the **retarded time**. These are called the **retarded potentials**.

PROBLEMS

16-1 A parallel-plate capacitor with plates having the shape of circular disks has the region between its plates filled with a dielectric of permittivity ϵ . The dielectric is imperfect, having a conductivity g . The capacitance of the capacitor is C . The capacitor is charged to a potential difference $\Delta\varphi$ and isolated. (a) Find the charge on the capacitor as a function of time. (b) Find the displacement current in the dielectric. (c) Find the magnetic field in the dielectric.

16-2 The Q of a dielectric medium is defined as the ratio of displacement current density to conduction current density. For monochromatic wave propagation we find that $Q = \omega\epsilon/g$. Determine Q for quartz and for sulfur, at the following frequencies: $f = 1, 10^6, 10^9$ Hz.

16-3 Two circular plates of radius a separated by distance d form an ideal capacitor: Assume that the dielectric is a perfect insulator with uniform D -field (i.e., neglect the fringing field at the edge of the plates). The capacitor is being charged by a constant current I . (a) Find the H -field at a point P on the cylindrical surface of the dielectric. (b) Find the magnitude and direction of the Poynting vector \mathbf{S} at P . (c) Integrate $\mathbf{S} \cdot \mathbf{n}$ over the cylindrical surface of the dielectric, and show that the result is equal to the time rate of change of the stored electrostatic energy.

16-4 A straight metal wire of conductivity g and cross-sectional area A carries a steady current I . Determine the direction and magnitude of the Poynting vector at the surface of the wire. Integrate the normal component of the Poynting vector over the surface of the wire for a segment of length L , and compare your result with the Joule heat produced in this segment.

16-5 Suppose that in a certain region there is an electrostatic field and also a magnetostatic field. Show that although the Poynting vector may be nonzero, the surface integral of $\mathbf{S} \cdot \mathbf{n}$ vanishes over an arbitrary closed surface inside the region.

16-6 Consider the one-dimensional wave equation

$$\frac{\partial^2 E}{\partial z^2} = \epsilon\mu \frac{\partial^2 E}{\partial t^2}$$

where E is the magnitude of the electric field vector. Assume that \mathbf{E} has a constant direction, namely, the y -direction. (a) By introducing the change of variables

$$\xi = t + \sqrt{\epsilon\mu} z, \quad \eta = t - \sqrt{\epsilon\mu} z$$

show that the wave equation assumes a form that is easily integrated. (b) Integrate the equation to obtain

$$E(z, t) = E_1(\xi) + E_2(\eta)$$

where E_1 and E_2 are arbitrary functions.

16-7 Consider the electromagnetic wave

$$\mathbf{E} = \mathbf{i}E_0 \cos \omega(\sqrt{\epsilon\mu} z - t) + \mathbf{j}E_0 \sin \omega(\sqrt{\epsilon\mu} z - t)$$

where E_0 is a constant. Find the corresponding magnetic field \mathbf{B} and the Poynting vector.

***16-8** Starting with an expression for the force per unit volume on a region of free space containing charges and currents,

$$\mathbf{F}_v = \rho \mathbf{E} + \mathbf{J} \times \mathbf{B}$$

and using the Maxwell equations and the vector identity of Eq. (14-24), show that

$$\begin{aligned} \mathbf{F}_v = & -\epsilon_0 \frac{\partial}{\partial t} (\mathbf{E} \times \mathbf{B}) + \epsilon_0 \mathbf{E} \nabla \cdot \mathbf{E} - \frac{1}{2} \epsilon_0 \nabla (E^2) + \epsilon_0 (\mathbf{E} \cdot \nabla) \mathbf{E} \\ & + \frac{1}{\mu_0} \mathbf{B} \nabla \cdot \mathbf{B} - \frac{1}{2\mu_0} \nabla (B^2) + \frac{1}{\mu_0} (\mathbf{B} \cdot \nabla) \mathbf{B} \end{aligned}$$

(The quantity $\epsilon_0 \mathbf{E} \times \mathbf{B}$ is sometimes referred to as the momentum density of the electromagnetic field.)

16-9 Given a plane wave characterized by an E_x , B_y , propagating in the positive z -direction,

$$\mathbf{E} = \mathbf{i}E_0 \sin \frac{2\pi}{\lambda} (z - ct)$$

(a) Show that it is possible to take the scalar potential $\varphi = 0$. (b) Find a possible vector potential \mathbf{A} for which the Lorentz condition is satisfied.

16-10 Show that in free space with $\rho = 0$, $\mathbf{J} = 0$, the Maxwell equations are correctly obtained from a single vector function \mathbf{A} satisfying

$$\nabla \cdot \mathbf{A} = 0, \quad \nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = 0$$

The gauge in which $\nabla \cdot \mathbf{A} = 0$ is called the *Coulomb gauge*.

16-11 Show that in a linear conducting medium a suitable gauge can be chosen so that \mathbf{A} and φ each satisfy the damped wave equation, Eq. (16-29). Assume that $\rho = 0$.

16-12 Consider a medium in which $\rho = 0$, $\mathbf{J} = 0$, $\mu = \mu_0$, but where the polarization \mathbf{P} is a given function of position and time: $\mathbf{P} = \mathbf{P}(x, y, z, t)$. Show that the

Maxwell equations are correctly obtained from a single vector function \mathbf{Z} (the Hertz vector), where \mathbf{Z} satisfies the equations

$$\nabla^2 \mathbf{Z} - \frac{1}{c^2} \frac{\partial^2 \mathbf{Z}}{\partial t^2} = -\frac{\mathbf{P}}{\epsilon_0}$$

and

$$\mathbf{E} = \nabla \times \nabla \times \mathbf{Z} - \frac{1}{\epsilon_0} \mathbf{P}, \quad \mathbf{B} = \frac{1}{c^2} \nabla \times \frac{\partial \mathbf{Z}}{\partial t}$$

16-13 Consider a medium in which $\rho = 0$, $\mathbf{J} = 0$, $\epsilon = \epsilon_0$, but where the magnetization $\mathbf{M}(x, y, z, t)$ is a given function. Show that the Maxwell equations are correctly obtained from a single vector function \mathbf{Y} , where \mathbf{Y} satisfies the equation

$$\nabla^2 \mathbf{Y} - \frac{1}{c^2} \frac{\partial^2 \mathbf{Y}}{\partial t^2} = -\mu_0 \mathbf{M}$$

and where

$$\mathbf{B} = \nabla \times \nabla \times \mathbf{Y}, \quad \mathbf{E} = -\nabla \times \frac{\partial \mathbf{Y}}{\partial t}$$

16-14 Show that Maxwell's equations for an isotropic, homogeneous, nonconducting, charge-free medium can be satisfied by taking either

$$1. \quad \mathbf{E} = \text{real part of } \nabla \times \nabla \times (F\mathbf{a}), \quad \mathbf{B} = \text{real part of } \epsilon\mu \frac{\partial}{\partial t} \nabla \times (F\mathbf{a})$$

or

$$2. \quad \mathbf{B} = \text{real part of } \nabla \times \nabla \times (F\mathbf{a}), \quad \mathbf{E} = \text{real part of } -\frac{\partial}{\partial t} \nabla \times (F\mathbf{a})$$

where \mathbf{a} is a constant unit vector and F satisfies the scalar wave equation.

Propagation of Monochromatic Electromagnetic Waves

Maxwell's equations have some special solutions that describe electromagnetic waves, as we saw by deriving the wave equation from the Maxwell equations. In this chapter, we consider these solutions in detail. We begin by considering the propagation of the waves through a linear medium, which we idealize to be infinite in extent. We leave until later chapters the question of how the waves are generated and how they enter the medium in the first place.

The term *electromagnetic wave propagation* covers a wide range of physical phenomena such as radio waves, visible light, and X-rays. In vacuum, all of the waves propagate with velocity c , but they are distinguished from one another by their frequency f (or their wavelength λ). Table 17-1 lists representative regions of the electromagnetic frequency spectrum from low-frequency radio waves (10^4 Hz) to hard gamma rays (10^{22} Hz).

17-1

PLANE MONOCHROMATIC WAVES IN NONCONDUCTING MEDIA

The most easily treated solutions of Eq. (16-31) are those known as plane wave solutions. A plane wave is defined as a wave whose phase is the same, at a given instant, at all points on each *plane* perpendicular to some specified direction.* If for example the specified direction is the z -direction, then \mathbf{E}

* It might seem more natural to define a plane wave as a wave for which the E - and B -fields are the same at all points on each plane perpendicular to some specified direction. For nonconducting media the two definitions are usually equivalent. For conducting media a real ω leads to a complex \mathbf{k} and one finds that there may be no true plane waves using this definition. There are always, however, plane waves using the constant phase definition, and we shall consistently use that definition.

TABLE 17-1

The Electromagnetic Spectrum

Type of EM radiation	Frequency f (Hz)	Wavelength λ (m)
Radio waves	10^4 to 10^8	3 to 3×10^4
TV, FM	10^8	3
Microwaves, RADAR	10^9 to 10^{11}	3×10^{-3} to 3×10^{-1}
Far infrared	10^{12}	3×10^{-4}
Infrared radiation	10^{13} to 4×10^{14}	8×10^{-7} to 3×10^{-5}
Visible light	4×10^{14} to 8×10^{14}	4×10^{-7} to 8×10^{-7}
Ultraviolet radiation	8×10^{14} to 10^{17}	3×10^{-9} to 4×10^{-7}
Soft X-rays	10^{17} to 10^{19}	3×10^{-11} to 3×10^{-9}
X-rays	10^{19} to 10^{20}	3×10^{-12} to 3×10^{-11}
Gamma rays	10^{20} to 10^{22}	3×10^{-14} to 3×10^{-12}

must have the same phase at all points that have the same z -value—that is, all points on a plane parallel to the xy -plane. Thus the solution in Eq. (16-34) that we already discussed is a plane wave solution, since $(\omega t - \kappa z)$ is a constant for a given t and z , no matter what the values of x and y . Plane waves traveling in the z -direction are adequate for problems in which the choice of z -direction is arbitrary. However, in many problems a system of axes is chosen for other reasons, for example, because of boundary conditions. In such cases it is necessary to construct plane waves with arbitrary directions of propagation. Suppose a plane wave solution with direction of propagation \mathbf{u} is to be constructed, where \mathbf{u} is a unit vector. Then the variable z in the exponent must be replaced by $\mathbf{u} \cdot \mathbf{r}$, the projection of \mathbf{r} in the \mathbf{u} direction. Thus a plane wave with direction of propagation \mathbf{u} is described by

$$e^{-i(\omega t - \kappa \mathbf{u} \cdot \mathbf{r})}$$

We define a vector, called the *propagation vector*,

$$\mathbf{k} = \kappa \mathbf{u}$$

and write the exponential space and time dependence of the plane wave as

$$e^{-i(\omega t - \mathbf{k} \cdot \mathbf{r})}$$

If $\mathbf{u} = \mathbf{k}$, the unit vector in the z -direction, then $\mathbf{u} \cdot \mathbf{r} = z$ as in the special case, but in every case the wavelength $\lambda = 2\pi/\kappa$.

The velocity of propagation of a plane monochromatic wave is precisely the velocity with which planes of constant phase move. Constant phase means, of course, that

$$\mathbf{k} \cdot \mathbf{r} - \omega t = \text{constant} \quad (17-1)$$

If $\mathbf{\kappa} \cdot \mathbf{r}$ is written $\kappa\xi$, with κ the magnitude of $\mathbf{\kappa}$ and ξ the projection of \mathbf{r} in the $\mathbf{\kappa}$ direction, then Eq. (17-1) becomes

$$\kappa\xi - \omega t = \text{constant}$$

Differentiating with respect to the time yields the *phase velocity*,

$$v_p = \frac{d\xi}{dt} = \frac{\omega}{\kappa} = \frac{c}{n} \quad (17-2)$$

where we have used the result of Eq. (16-33a) $\kappa = n\omega/c$. In free space,

$$v_p = c = \frac{1}{\sqrt{\epsilon_0\mu_0}} = 2.9979 \times 10^8 \text{ m/s}$$

Now to obtain the detailed plane wave solutions for \mathbf{E} and \mathbf{B} , we could go back to Eq. (16-31), but it is actually better to go all the way back to the Maxwell equations themselves. There are no prescribed charge or current distributions in the medium and the conductivity $g = 0$, so that the equations are

$$\nabla \cdot \mathbf{D} = 0 \quad (17-3)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (17-4)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (17-5)$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} \quad (17-6)$$

From our previous discussion of the wave equation, we already know the space and time dependence to be expected in a plane wave, so let us assume that the fields have the form

$$\mathbf{E}(\mathbf{r}, t) = \hat{\mathbf{E}}e^{-i(\omega t - \mathbf{\kappa} \cdot \mathbf{r})}, \dots \quad (17-7)$$

where $\hat{\mathbf{E}}$ is a complex constant vector amplitude of the plane wave (to be discussed in detail in the following section), and let us substitute the assumed solutions into the Maxwell equations (17-3) to (17-6). This substitution will impose conditions that the assumed constants $\mathbf{\kappa}$, $\hat{\mathbf{E}}$, and so on, will have to satisfy in order that the plane wave functions will actually be solutions of the Maxwell equations.

By differentiating a function of the form $\hat{\mathbf{E}}e^{-i\omega t}$ with respect to t , one sees that the operator $\partial/\partial t$ is equivalent to multiplying by $-i\omega$:

$$\frac{\partial}{\partial t} = -i\omega$$

for a function of this particular form. Similarly, one finds (Problem 17-1) that for a function of the form $\hat{\mathbf{E}}e^{i\mathbf{k}\cdot\mathbf{r}}$, the operator ∇ is equivalent to

$$\nabla = i\mathbf{k}$$

Thus, the Maxwell equations for plane waves (after i and the exponential are canceled) become

$$\mathbf{k} \cdot \hat{\mathbf{D}} = 0 \quad (17-8)$$

$$\mathbf{k} \cdot \hat{\mathbf{B}} = 0 \quad (17-9)$$

$$\mathbf{k} \times \hat{\mathbf{E}} = \omega \hat{\mathbf{B}} \quad (17-10)$$

$$\mathbf{k} \times \hat{\mathbf{H}} = -\omega \hat{\mathbf{D}} \quad (17-11)$$

If we assume that the medium is linear, the constitutive equations are

$$\hat{\mathbf{D}} = \epsilon \hat{\mathbf{E}} \quad (17-12)$$

$$\hat{\mathbf{H}} = \frac{1}{\mu} \hat{\mathbf{B}} \quad (17-13)$$

We also assume that the medium is homogeneous and isotropic, so that ϵ and μ are constant scalars. All of our applications will be to nonmagnetic media,* so for simplicity let us assume that $\mu = \mu_0$. Recalling that $\epsilon = K\epsilon_0$ and $\epsilon_0\mu_0 = 1/c^2$, we get the Maxwell equations in the form

$$K\mathbf{k} \cdot \hat{\mathbf{E}} = 0 \quad (17-14)$$

$$\mathbf{k} \cdot \hat{\mathbf{B}} = 0 \quad (17-15)$$

$$\mathbf{k} \times \hat{\mathbf{E}} = \omega \hat{\mathbf{B}} \quad (17-16)$$

$$\mathbf{k} \times \hat{\mathbf{B}} = -\frac{\omega}{c^2} K \hat{\mathbf{E}} \quad (17-17)$$

If we take ω to be a given frequency and K to be a given material constant, we must try to satisfy this set of algebraic vector equations by suitable choices of \mathbf{k} , $\hat{\mathbf{E}}$, and $\hat{\mathbf{B}}$. First, if we assume that $K \neq 0$, we see that $\mathbf{k} \cdot \hat{\mathbf{E}} = 0$; $\mathbf{k} \cdot \hat{\mathbf{B}} = 0$ always. That is, both \mathbf{E} and \mathbf{B} must be perpendicular to \mathbf{k} . Such a wave is called *transverse*. (The case $K = 0$ is actually possible and nontrivial, but we will defer its discussion.) Further, since $\hat{\mathbf{B}}$ is proportional to $\mathbf{k} \times \hat{\mathbf{E}}$, \mathbf{E} and \mathbf{B} are also perpendicular to each other. The vectors \mathbf{k} , \mathbf{E} , \mathbf{B} (in that order) form a right-handed orthogonal set. The relative magnitude of $\hat{\mathbf{E}}$ and $\hat{\mathbf{B}}$ is also determined by Eq. (17-16), $\hat{\mathbf{B}} = (\mathbf{k}/\omega)\hat{\mathbf{E}}$. Finally, we find

* The only media for which μ differs appreciably from μ_0 at low frequencies are the ferromagnetic ones, which are not linear anyway. For optical frequencies, $\mu \cong \mu_0$ for all materials. We exclude from consideration paramagnetic resonance, which is observable at radio and microwave frequencies under special circumstances.

the magnitude of κ by taking the vector product of κ with Eq. (17-16) and using $\kappa \times \hat{\mathbf{B}}$ from Eq. (17-17):

$$\kappa \times (\kappa \times \hat{\mathbf{E}}) = \omega \kappa \times \hat{\mathbf{B}} = -K(\omega/c)^2 \hat{\mathbf{E}}$$

With the vector identity

$$\kappa \times (\kappa \times \hat{\mathbf{E}}) = (\kappa \cdot \hat{\mathbf{E}})\kappa - \kappa^2 \hat{\mathbf{E}}$$

since $\kappa \cdot \hat{\mathbf{E}} = 0$ for the transverse wave,

$$-K(\omega/c)^2 \hat{\mathbf{E}} = -\kappa^2 \hat{\mathbf{E}}$$

which is equivalent to the wave equation for solutions of the form given by Eq. (17-7). $\hat{\mathbf{E}}$ can be dropped from both sides of this equation to yield

$$\kappa = \sqrt{K} \omega/c \quad (17-18)$$

This relation, called the *transverse dispersion relation*, determines the magnitude of the wave vector κ in terms of the assumed ω and K .

To recapitulate, a plane monochromatic transverse wave propagated in the plus \mathbf{u} direction is described by

$$\mathbf{E}(\mathbf{r}, t) = \hat{\mathbf{E}} e^{-i(\omega t - \kappa \cdot \mathbf{r})}, \quad \mathbf{B}(\mathbf{r}, t) = \hat{\mathbf{B}} e^{-i(\omega t - \kappa \cdot \mathbf{r})} \quad (17-7)$$

where $\kappa = \kappa \mathbf{u}$. The direction \mathbf{u} and the frequency ω are completely arbitrary. The amplitude $\hat{\mathbf{E}}$ is arbitrary except that it must be perpendicular to \mathbf{u} :

$$\mathbf{u} \cdot \hat{\mathbf{E}} = 0 \quad (17-19)$$

The magnitude of κ is determined, for the given frequency ω , by the refractive index of the material:

$$\kappa = n\omega/c \quad (17-20)$$

where n is defined as

$$n = \sqrt{K} \quad (17-21)$$

Then $\hat{\mathbf{B}}$ is completely determined in magnitude and direction:

$$\hat{\mathbf{B}} = \frac{n}{c} \mathbf{u} \times \hat{\mathbf{E}} \quad (17-22)$$

Note that in vacuum ($n = 1$), $c\hat{\mathbf{B}} = \hat{\mathbf{E}}$ in mks units.* The phase velocity of the wave is c/n . With these results it is possible to consider some extremely interesting and important optical problems. These, however, will be postponed until the next chapter.

In this discussion we have explicitly excluded the case $K = 0$. In

* In gaussian units, according to the discussion in Chapter 8, we replace B by B/c , so that $B = E$. That is, the E and B fields have equal magnitudes for a plane wave in vacuum.

addition by implicitly assuming that $\boldsymbol{\kappa}$ is real we have excluded a set of plane wave solutions that are needed in some cases to satisfy the boundary conditions at a plane interface between two nonconducting media. If $\boldsymbol{\kappa}$ is not real, we take

$$\boldsymbol{\kappa} = \boldsymbol{\kappa}_r + i\boldsymbol{\kappa}_i$$

so that Eq. (17-18) should be written

$$\boldsymbol{\kappa}_r \cdot \boldsymbol{\kappa}_r - \boldsymbol{\kappa}_i \cdot \boldsymbol{\kappa}_i + 2i\boldsymbol{\kappa}_r \cdot \boldsymbol{\kappa}_i = K \frac{\omega^2}{c^2}$$

The imaginary part of the left-hand side of this equation must vanish, and in order for the real parts to be equal, $\boldsymbol{\kappa}_r^2$ must be greater than $\boldsymbol{\kappa}_i^2$. There are two ways in which the imaginary part can vanish: The imaginary part of $\boldsymbol{\kappa}$, $\boldsymbol{\kappa}_i$, can vanish, or it can be perpendicular to $\boldsymbol{\kappa}_r$. The first case has been thoroughly discussed above. The second case consists of waves in which the planes of constant phase are perpendicular to the planes of constant amplitude, as will be explained in Section 17-4. (These waves will be considered in more detail in connection with total internal reflection in Section 18-4.) In either case, $\boldsymbol{\kappa} \cdot \boldsymbol{\kappa}$ is real, and consequently, a real magnitude of $\boldsymbol{\kappa}$ can be defined that satisfies Eq. (17-18). Complex wave vectors are further discussed in Section 17-4.

Although the plane wave solutions are only a restricted class of solutions of Maxwell's equations, they are very important since they form the basis of a much wider class of solutions. Because the equations are linear, a linear combination of solutions (superposition of plane waves) is also a solution. Thus we could form other solutions by taking sums of plane waves

$$\mathbf{E}(\mathbf{r}, t) = \sum_i \hat{\mathbf{E}}(\boldsymbol{\kappa}_i, \omega_i) \exp[-i(\omega_i t - \boldsymbol{\kappa}_i \cdot \mathbf{r})] \quad (17-23)$$

where each coefficient $\hat{\mathbf{E}}$ would depend on $\boldsymbol{\kappa}_i$ and ω_i . This superposition of plane waves has the form of a (complex) Fourier series and, therefore, could represent any solution that was periodic—not necessarily sinusoidal. Each term of the series would have to satisfy the conditions of Eqs. (17-14) to (17-17) separately. For a solution that is not even periodic, the sum in Eq. (17-23) can be converted to an integral—the Fourier integral†—with $\hat{\mathbf{E}}(\boldsymbol{\kappa}, \omega)$ a continuous function of $\boldsymbol{\kappa}$ and ω . The function $\hat{\mathbf{E}}(\boldsymbol{\kappa}, \omega)$ is called the *Fourier transform* of $\mathbf{E}(\mathbf{r}, t)$. In this case we would also have to consider the possibility that n depends on $\boldsymbol{\kappa}$ and ω as well,

$$n = n(\boldsymbol{\kappa}, \omega)$$

This latter effect, known as *dispersion*, will be discussed in Chapter 19.

† See Appendix VI.

17-2 POLARIZATION

There is more to be said about the complex vector amplitudes $\hat{\mathbf{E}}$ and $\hat{\mathbf{B}}$. In fact, we have not yet discussed fully what we mean by a complex vector. Two obvious meanings suggest themselves: A complex quantity whose real and imaginary parts are real vectors,

$$\hat{\mathbf{E}} = \mathbf{E}_r + i\mathbf{E}_i$$

or a vector whose components (with respect to real basis vectors) are complex scalars,

$$\hat{\mathbf{E}} = \hat{E}_p \mathbf{p} + \hat{E}_s \mathbf{s} + \hat{E}_u \mathbf{u}$$

We use the circumflex for quantities that are complex when it is necessary to distinguish them; in the second form, \mathbf{p} , \mathbf{s} , \mathbf{u} are a right-handed set of real orthogonal unit vectors. By writing the first form in terms of components and the second in terms of real and imaginary parts, it is easily seen that the two formulations are equivalent, provided that

$$E_{p_r} = E_{r_p}, \quad E_{p_i} = E_{i_p}, \quad E_{s_r} = E_{r_s}, \quad \text{and so on}$$

For our present purpose, the second form is more convenient. We take \mathbf{u} to be the propagation direction of the plane wave, so that $\hat{E}_u = 0$ according to the result $\mathbf{u} \cdot \hat{\mathbf{E}} = 0$ of Eq. (17-19), but \hat{E}_p and \hat{E}_s are arbitrary:

$$\hat{\mathbf{E}} = \hat{E}_p \mathbf{p} + \hat{E}_s \mathbf{s} \quad (17-24)$$

The unit vector \mathbf{p} can be chosen in any direction perpendicular to \mathbf{u} . In the next chapter, we will make a special choice that will account for the particular notation introduced here.

It is also more convenient to express the complex components in polar form instead of in terms of real and imaginary parts. Let

$$\hat{E}_p = E_p e^{i\phi_p}, \quad \hat{E}_s = E_s e^{i\phi_s} \quad (17-25)$$

Then, for example,

$$\hat{E}_s e^{-i(\omega t - \mathbf{k} \cdot \mathbf{r})} = E_s e^{-i(\omega t - \mathbf{k} \cdot \mathbf{r} - \phi_s)}$$

That is, ϕ_s is the phase of the E -field component in the \mathbf{s} direction. It is no restriction to let

$$\phi_p - \phi_s = \phi \quad \phi_s = 0$$

since $\phi_s = 0$ merely dictates a certain choice of the origin of t . With this choice,

$$\hat{\mathbf{E}} = E_p e^{i\phi} \mathbf{p} + E_s \mathbf{s}$$

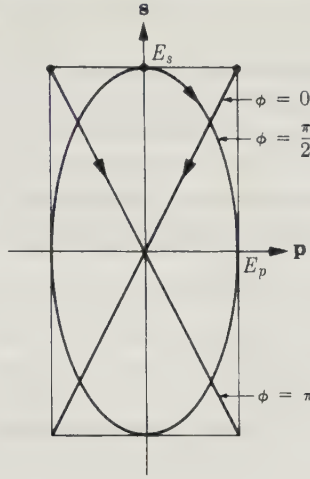
$$\mathbf{E}(\mathbf{r}, t) = E_p \mathbf{p} e^{-i(\omega t - \mathbf{k} \cdot \mathbf{r} - \phi)} + E_s \mathbf{s} e^{-i(\omega t - \mathbf{k} \cdot \mathbf{r})}$$

or the real part is

$$\mathbf{E}(\mathbf{r}, t) = E_p \mathbf{p} \cos(\omega t - \mathbf{k} \cdot \mathbf{r} - \phi) + E_s \mathbf{s} \cos(\omega t - \mathbf{k} \cdot \mathbf{r}) \quad (17-26)$$

FIGURE 17-1

Trace of the tip of the E -vector at a given point in space as a function of time. The propagation direction \mathbf{u} is pointing toward us. The traces for $\phi = 0$ and π are linearly polarized. The trace for $\phi = \pi/2$ is elliptically polarized, right-hand; for $-\pi/2$ (left-hand), it rotates in the opposite direction.



The E -field is resolved into components in two directions, with real amplitudes E_p and E_s , which may have any values. In addition, the two components may be oscillating out of phase by ϕ : That is, at any given point \mathbf{r} , the maximum of E in the \mathbf{p} direction may be attained at a different time from the maximum of E in the \mathbf{s} direction.

A detailed picture of the oscillating E -field at a certain point, say $\mathbf{r} = 0$, is best seen by considering some special cases. First, suppose that $\phi = 0$. Then,

$$\mathbf{E}(0, t) = (E_p \mathbf{p} + E_s \mathbf{s}) \cos \omega t$$

The E -field alternately decreases from $\sqrt{E_p^2 + E_s^2}$ through zero to $-\sqrt{E_p^2 + E_s^2}$ and back to the original value, always pointing along the direction $E_p \mathbf{p} + E_s \mathbf{s}$. This case is called *linear polarization** and is illustrated in Fig. 17-1. If $E_p = 0$ or $E_s = 0$, we also have linear polarization; then ϕ is undefined. For $\phi = \pi$,

$$\mathbf{E}(0, t) = (-E_p \mathbf{p} + E_s \mathbf{s}) \cos \omega t$$

again linear polarization, as shown in Fig. 17-1. For $\phi = \pi/2$,

$$\mathbf{E}(0, t) = E_p \mathbf{p} \sin \omega t + E_s \mathbf{s} \cos \omega t$$

The tip of the E -vector traces an elliptical path clockwise as shown. This

* This use of the term "polarization" is not the same as the use introduced in Chapter 4. It is unfortunate that the same word is conventionally used, but usually no confusion arises since one usage applies to a wave and the other to a medium.

case is called *right-hand* elliptical polarization*. For $\phi = -\pi/2$, the path is the same, but traced counterclockwise, called *left-hand elliptical polarization*. In the special case when $\phi = \pm\pi/2$ and $E_p = E_s$, we have (right- or left-hand) *circular polarization*. For other values of ϕ , we have elliptical polarization (even if $E_p = E_s$). The trace is still an ellipse inscribed in the box of Fig. 17-1, but the minor and major axes of the ellipse are at an angle to the p - and s -axes. With elliptical polarization the magnitude of the E -vector never goes to zero, although its component in any given direction does.

The complex amplitude of the B -vector is given by Eq. (17-22),

$$\hat{\mathbf{B}} = \frac{n}{c} \mathbf{u} \times \hat{\mathbf{E}} \quad (17-22)$$

Taking the dot product of this with $\hat{\mathbf{E}}$, and interchanging dot and cross, we find that

$$\hat{\mathbf{B}} \cdot \hat{\mathbf{E}} = 0$$

In general, the vanishing of the dot product of two *complex* vectors does not mean that their real parts are perpendicular, but in this case they are. From Eq. (17-26), writing $\mathbf{E}(0, t) = \mathbf{E}$, we have for the real part

$$\mathbf{E} = E_p \mathbf{p} \cos(\omega t - \phi) + E_s \mathbf{s} \cos \omega t \quad (17-27)$$

From $\mathbf{B}(0, t)$, the real part of $\hat{\mathbf{B}} e^{-i\omega t}$, with Eq. (17-22), we find

$$\mathbf{B} = (n/c)[E_p \mathbf{s} \cos(\omega t - \phi) - E_s \mathbf{p} \cos \omega t] \quad (17-28)$$

Since the dot product of Eq. (17-28) with (17-27) is zero, the real E - and B -vectors are perpendicular at each instant. Also $\text{Re } \hat{\mathbf{E}} = \mathbf{E}(0, 0)$, $\text{Re } \hat{\mathbf{B}} = \mathbf{B}(0, 0)$, so that the real parts of $\hat{\mathbf{E}}$ and $\hat{\mathbf{B}}$ are perpendicular. The trace of the tip of the B -vector is the same as Fig. 17-1 rotated 90° counterclockwise.

Since the p - and s -axes were chosen arbitrarily in the plane perpendicular to \mathbf{u} , any other choice could be made. A new choice would rotate the coordinate axes in Fig. 17-1 and introduce new values of E_p , E_s , and ϕ_s ; but the trace of the E -vector in Fig. 17-1 represents a physical reality, which would not be changed merely by the coordinate transformation. At this point, however, even the physical state of polarization itself is part of the arbitrariness admitted by our assumed plane wave solution for an infinite medium. In the next chapter, we shall see how a particular polarization can be produced and measured.

* Note that the "right-hand rule" does not apply to the motion of E in Fig. 17-1. The terminology derives instead from the path of the E -vector in space along the propagation direction, at a given instant of time. For right-hand polarization, the tip of the E -vector traces out a right-handed helix, or screw thread, as viewed from any direction.

17-3 ENERGY DENSITY AND FLUX

We have freely used complex expressions for the E - and B -fields, with the understanding that the actual physical quantities are given by the *real parts* of the complex quantity. The mathematical justification of this procedure is that the Maxwell equations are *linear* equations that are satisfied separately by the real and imaginary parts of a complex solution. The expressions

$$u = \frac{1}{2}(\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H}) \quad (17-29)$$

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \quad (17-30)$$

for the energy density and flux per unit area, however, are nonlinear in the fields. Therefore, in these expressions it is essential to take the real parts of the fields before multiplying them together. (See Problem 17-6.) We may again calculate representative values at $\mathbf{r} = 0$, since the origin is arbitrary. Squaring Eqs. (17-27) and (17-28) gives

$$E^2 = E_p^2 \cos^2(\omega t - \phi) + E_s^2 \cos^2 \omega t \quad (17-31)$$

$$B^2 = (n/c)^2 E^2 = \epsilon \mu_0 E^2 \quad (17-32)$$

Since $D = \epsilon \mathbf{E}$, $\mathbf{B} = \mu_0 \mathbf{H}$, we find that the electric and magnetic fields each make equal contributions to the energy density u :

$$\mathbf{B} \cdot \mathbf{H} = \mathbf{D} \cdot \mathbf{E}$$

$$u = \epsilon E^2 = \frac{1}{\mu_0} \left(\frac{n}{c}\right)^2 E^2 \quad (17-33)$$

Further, $\mathbf{E} \times \mathbf{H} = EH\mathbf{u}$, so that the Poynting vector points in the propagation direction, with magnitude

$$S = \frac{1}{\mu_0} \frac{n}{c} E^2 \quad (17-34)$$

The expressions for energy density and energy flux per unit area have assumed especially simple forms for plane waves. Further, the two expressions can be combined to give an interesting result, which is independent of the particular value of the E -field:

$$S = \frac{c}{n} u \quad (17-35)$$

If we write the phase velocity of the plane wave as a vector in the propagation direction, with magnitude

$$v_p = \frac{c}{n}$$

then

$$\mathbf{S} = u \mathbf{v}_p$$

This equation is analogous to the relation

$$\mathbf{J} = \rho \mathbf{v}$$

which defines the convective electric current density. This analogy reinforces the interpretation of \mathbf{S} as an energy current density—that is, an energy density u that is transported with the phase velocity \mathbf{v}_p of the plane wave.

The time dependence of u and \mathbf{S} is given by E^2 from Eq. (17-31) and depends on the polarization of the wave. For circular polarization ($\phi = \pm\pi/2$),

$$E^2 = E_p^2 \sin^2 \omega t + E_p^2 \cos^2 \omega t = E_p^2$$

is constant in time; for linear polarization ($\phi = 0, \pi$),

$$E^2 = (E_p^2 + E_s^2) \cos^2 \omega t$$

varies between zero and a maximum at twice the wave frequency. In every case, E^2 is always positive of course. At high frequencies the time dependence is not measurable, however, and so what is of more interest is the time average of E^2 . Since the time average over one period of $\cos^2(\omega t - \phi)$ is $\frac{1}{2}$, for any polarization

$$\overline{E^2} = \frac{1}{2}(E_p^2 + E_s^2) \quad (17-36)$$

This and similar results can be obtained more quickly by using a theorem that was introduced in Chapter 13 but not proved; we now give the proof.

If $f = f_0 e^{i\omega t}$ and $g = g_0 e^{i\omega t}$, where f_0 and g_0 may depend on other variables but not on the time, then

$$\overline{\operatorname{Re} f \operatorname{Re} g} = \frac{1}{2} \operatorname{Re} f^* g \quad (17-37)$$

The bar indicates the time average defined by

$$\bar{w} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T w(t) dt$$

To derive Eq. (17-37), let $f_0 = u + iv$ and $g_0 = \xi + i\eta$. Then,

$$\operatorname{Re} f \operatorname{Re} g = (u \cos \omega t - v \sin \omega t)(\xi \cos \omega t - \eta \sin \omega t) \quad (17-38)$$

while

$$\operatorname{Re} f^* g = u\xi + v\eta \quad (17-39)$$

The following integrals are easily verified:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \sin^2 \omega t dt = \frac{1}{2}$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \cos^2 \omega t dt = \frac{1}{2}$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \sin \omega t \cos \omega t dt = 0$$

By means of these integrals, it is easy to see that the time average of Eq. (17-38) is

$$\overline{\operatorname{Re} f \operatorname{Re} f} = \frac{1}{2}(u\xi + v\eta) \quad (17-40)$$

Comparison of Eqs. (17-40) and (17-39) proves the theorem of Eq. (17-37). The theorem holds for the product of any two complex quantities that depend harmonically on t with the same frequency, but not necessarily having the same phase. If the quantities are vectors, the product can be either the dot- or cross-product. It is easy to check that the expression (17-36) is immediately given by putting the complex \mathbf{E} of Eq. (17-26) into

$$\overline{E^2} = \frac{1}{2} \operatorname{Re} (\mathbf{E}^* \cdot \mathbf{E})$$

17-4 PLANE MONOCHROMATIC WAVES IN CONDUCTING MEDIA

The procedure described in Section 17-1 can again be used in the case of conducting media to obtain solutions that are formally very similar to those found in the nonconducting case. We may start with Eqs. (17-3) to (17-6) and the assumed solution, Eq. (17-7), except that Eq. (17-6) is replaced by

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + g\mathbf{E}$$

However, the physical interpretation of the solution is considerably more difficult here. There is another approach that does not simplify the physical interpretation, but it does provide some additional insight into electromagnetic waves, in conducting media in particular.

In Chapter 3, we used the technique of *separation of variables* to convert Laplace's equation into a set of ordinary differential equations. The same method can be applied to the wave equation

$$\nabla^2 \mathbf{E} - \epsilon\mu \frac{\partial^2 \mathbf{E}}{\partial t^2} - g\mu \frac{\partial \mathbf{E}}{\partial t} = 0 \quad (16-29)$$

The obvious differences here are that the dependent variable \mathbf{E} is a vector and one of the independent variables is the time t ; neither of these differences causes any problem. The interpretation to be given Eq. (16-29) is that each rectangular component of \mathbf{E} must satisfy the equation. We write $E(\mathbf{r}, t)$ for one such component and assume that it can be written in the form

$$E(\mathbf{r}, t) = R(\mathbf{r})T(t)$$

Inserting this product into Eq. (16-29) and dividing by $R(\mathbf{r})T(t)$ gives two terms: one independent of \mathbf{r} and one independent of t . Since the equation must be satisfied for all values of \mathbf{r} and t , each of the terms must be equal to

the *same constant*, which we denote as A . At this point, A is just a separation constant, but as the details of the solution develop, it will take on physical meaning. The result is two differential equations:

$$\epsilon\mu \frac{d^2T}{dt^2} + g\mu \frac{dT}{dt} + AT = 0 \quad (17-41a)$$

$$\nabla^2 R + AR = 0 \quad (17-41b)$$

It is again convenient to treat $R(\mathbf{r})$ and $T(t)$ as complex quantities, with the convention that the physical quantity, $E(\mathbf{r}, t)$, is the real part of a complex equivalent, $\hat{E}(\mathbf{r}, t) = R(\mathbf{r})T(t)$. In what follows we shall consistently use the circumflex to designate complex quantities. Equation (17-41a) is an ordinary second-order equation with constant coefficients, with the solution

$$\hat{T} = \hat{T}_1 e^{-i\hat{\omega}_1 t} + \hat{T}_2 e^{-i\hat{\omega}_2 t} \quad (17-42)$$

here $\hat{\omega}_1$ and $\hat{\omega}_2$ are independent solutions of the quadratic equation

$$\epsilon\mu\hat{\omega}^2 + ig\mu\hat{\omega} - \hat{A} = 0 \quad (17-43)$$

Equation (17-41b) is a partial differential equation (called the *Helmholtz equation*), which has a very rich set of solutions. From this set we choose those from which we expect that plane-wave-like solutions can be formed. In particular, we choose

$$\hat{R}(\mathbf{r}) = \exp(\pm i\hat{\mathbf{k}} \cdot \mathbf{r}), \quad \hat{\mathbf{k}} \cdot \hat{\mathbf{k}} = \hat{A} \quad (17-44)$$

Using Eq. (17-44) to eliminate \hat{A} from Eq. (17-43) gives

$$\epsilon\mu\hat{\omega}^2 + ig\mu\hat{\omega} - \hat{\mathbf{k}} \cdot \hat{\mathbf{k}} = 0 \quad (17-45)$$

Eq. (17-45) cannot be satisfied with $\hat{\mathbf{k}}$ and $\hat{\omega}$ both real. There are, however, interesting solutions in which one or the other is real. Monochromatic waves, in particular, correspond to real values of ω (which we identify with the angular frequency). We restrict our set of solutions in this way and write

$$\hat{E}(\mathbf{r}, t) = \hat{E} e^{i(\hat{\mathbf{k}} \cdot \mathbf{r} - \omega t)} \quad (17-46)$$

with $\hat{\mathbf{k}}$ and ω satisfying Eq. (17-45a):

$$\epsilon\mu\omega^2 + ig\mu\omega - \hat{\mathbf{k}} \cdot \hat{\mathbf{k}} = 0 \quad (17-45a)$$

Note that the circumflex on ω has been dropped. Eq. (17-46) is formally the same as the assumed harmonic solution, Eq. (17-7), for a nonconducting medium, except that $\hat{\mathbf{k}}$ is now a complex vector.

Equation (17-45a) is the general *dispersion relation* for monochromatic plane waves in a linear, homogeneous, isotropic medium. If the conductivity g is not zero and ω is real, Eq. (17-45a) cannot be satisfied unless $\hat{\mathbf{k}}$ is complex, and the real and imaginary parts of $\hat{\mathbf{k}}$ are not perpendicular.

Now, Maxwell's equations must be satisfied as well as the wave equation,

and this requirement places further restrictions on the vector fields. Maxwell's equations, in the form of Eqs. (17-14) to (17-16) and a modified (17-17) become (for a nonmagnetic medium, i.e., $\mu = \mu_0$):

$$K\hat{\mathbf{k}} \cdot \hat{\mathbf{E}} = 0 \quad (17-47)$$

$$\hat{\mathbf{k}} \cdot \hat{\mathbf{B}} = 0 \quad (17-48)$$

$$\hat{\mathbf{k}} \times \hat{\mathbf{E}} = \omega\hat{\mathbf{B}} \quad (17-49)$$

$$\hat{\mathbf{k}} \times \hat{\mathbf{B}} = -\frac{\omega}{c^2}K\hat{\mathbf{E}} - ig\mu_0\hat{\mathbf{E}} = -\frac{\omega}{c^2}\hat{K}\hat{\mathbf{E}} \quad (17-50)$$

Here we have used \hat{K} , the *complex dielectric constant*, defined as

$$\hat{K} = K + i\frac{g}{\epsilon_0\omega} \quad (17-51)$$

It is convenient to define the *complex index of refraction* \hat{n} by analogy to what was done earlier for the case of real index of refraction and real dielectric constant:

$$\hat{n}^2 = \hat{K} \quad (17-52)$$

Then

$$\hat{\mathbf{k}} \cdot \hat{\mathbf{k}} = \hat{n}^2\omega^2/c^2 \quad (17-52a)$$

Note that the combination of Eqs. (17-51), (17-52), and (17-52a) is equivalent to Eq. (17-45a) for a medium with $\mu = \mu_0$.

If

$$\hat{\mathbf{k}} = \boldsymbol{\kappa}_r + i\boldsymbol{\kappa}_i \quad (17-53)$$

then \mathbf{E} and \mathbf{B} may be written

$$\mathbf{E}(\mathbf{r}, t) = (\hat{\mathbf{E}}e^{-\boldsymbol{\kappa}_i \cdot \mathbf{r}})e^{-i(\omega t - \boldsymbol{\kappa}_r \cdot \mathbf{r})}$$

$$\mathbf{B}(\mathbf{r}, t) = (\hat{\mathbf{B}}e^{-\boldsymbol{\kappa}_i \cdot \mathbf{r}})e^{-i(\omega t - \boldsymbol{\kappa}_r \cdot \mathbf{r})}$$

This solution is a plane wave propagating in the direction $\boldsymbol{\kappa}_r$, with wavelength $\lambda = 2\pi/\kappa_r$; but instead of having a constant amplitude, it decreases in amplitude, most rapidly in the direction $\boldsymbol{\kappa}_i$. The surfaces of constant phase are planes perpendicular to the propagation direction $\boldsymbol{\kappa}_r$. There are also surfaces of constant amplitude that are planes perpendicular to $\boldsymbol{\kappa}_i$.

We write

$$\hat{n} = n + ik \quad (17-54)$$

to facilitate investigating the phase velocity and attenuation length of the wave. The real quantities n and k are called the *optical constants*.

The *dispersion relation* for a conducting medium is

$$\hat{k} = \hat{n}\omega/c \quad (17-52b)$$

with the scalar \hat{k} defined as

$$\hat{k} = \sqrt{\hat{\mathbf{k}} \cdot \hat{\mathbf{k}}} = \sqrt{\kappa_r^2 - \kappa_i^2 + 2i\kappa_r \cdot \kappa_i}$$

Let us assume to start that κ_r and κ_i have the same direction,* a restrictive assumption that fails to hold in some important cases to be discussed in the next chapter. It does hold, however, for the case of a plane wave entering a conducting medium at normal incidence to a plane boundary. Then we can write

$$\hat{\mathbf{k}} = (\kappa_r + i\kappa_i)\mathbf{u} = \hat{k}\mathbf{u}$$

where \mathbf{u} is the real unit vector in the common direction of κ_r and κ_i . Since \mathbf{u} is real, the equations

$$\mathbf{u} \cdot \hat{\mathbf{E}} = 0 = \mathbf{u} \cdot \hat{\mathbf{B}}$$

still mean that the wave has its E - and B -vectors perpendicular to the direction of propagation \mathbf{u} . Equation (17-22) becomes, however,

$$\hat{\mathbf{B}} = \frac{\hat{n}}{c} \mathbf{u} \times \hat{\mathbf{E}} \quad (17-55)$$

The complex \hat{n} means that \mathbf{E} and \mathbf{B} are not in phase with each other. It also means the real E - and B -vectors are not perpendicular to each other, except for linear polarization. (See Problem 17-9.) In terms of the complex refractive index

$$\hat{n} = n + ik$$

we have

$$\kappa_r = n\omega/c, \quad \kappa_i = k\omega/c \quad (17-56)$$

Writing $\mathbf{u} \cdot \mathbf{r} = \xi$, we get for the E -field in this special case

$$\mathbf{E}(\mathbf{r}, t) = (\hat{\mathbf{E}} e^{-k\omega\xi/c}) e^{-i\omega(t - n\xi/c)}$$

* Such waves are sometimes called uniform or homogeneous waves.

The wave propagates with *phase velocity* c/n and *attenuation constant* $k\omega/c$. The latter quantity determines how rapidly the amplitude of the oscillating field falls off with distance. Its reciprocal

$$\delta = c/k\omega \quad (17-57)$$

which is called the *skin depth*, measures the distance within which the field falls to $1/e$ of its value at a given point (say a surface where the wave enters the medium).

For a nonconducting medium ($k = 0$) the skin depth is infinite. Since the wavelength in the medium is $\lambda = 2\pi c/n\omega$ ($=1/n$ times the wavelength in vacuum at the same frequency), we can write the skin depth

$$\delta = \frac{n \lambda}{k 2\pi} \quad (17-58)$$

In a material where k is comparable in magnitude with n , the “wave” decays in about one wavelength; but if $k \ll n$, as in an imperfect dielectric with a small conductivity, the wave propagates many wavelengths without appreciable loss. In the latter case the material is *transparent*.

When κ_r and κ_i have different directions, the situation is more complicated. The real E - and B -fields do not have to be perpendicular to κ_r (nor to κ_i), although the wave is still called “transverse” if $\hat{K} \neq 0$, $\hat{\mathbf{k}} \cdot \hat{\mathbf{E}} = 0 = \hat{\mathbf{k}} \cdot \hat{\mathbf{B}}$. The wavelength and attenuation still depend on the material constants n and k , but in a much more complicated way than Eq. (17-56). The polarization and energy relations can be generalized, with due care for the complex quantities. We will have no need for these general results, however.

Since the wave propagation is determined by the optical constants n and k , it is important to explore their relation to the quantities K and g , through which the properties of the material were originally expressed. The definitions

$$\hat{n} = n + ik \quad (17-54)$$

$$\hat{K} = K + i \frac{g}{\epsilon_0 \omega} \quad (17-51)$$

are related by

$$\hat{K} = \hat{n}^2 \quad (17-52)$$

This relationship looks deceptively simple, but when it is expressed in terms of the real quantities n , k and K , g , it becomes fairly complicated. We

write*

$$\hat{K} = K_r + iK_i$$

where

$$K_r = K, \quad K_i = g/\epsilon_0\omega \quad (17-59)$$

Squaring $(n + ik)$ and equating real and imaginary parts in Eq. (17-52) gives

$$K_r = n^2 - k^2 \quad (17-60)$$

$$K_i = 2nk$$

These equations can be solved for n and k :

$$\begin{aligned} n &= \sqrt{\frac{1}{2}[K_r + \sqrt{K_r^2 + K_i^2}]} \\ k &= \sqrt{\frac{1}{2}[-K_r + \sqrt{K_r^2 + K_i^2}]} \end{aligned} \quad (17-61)$$

Positive square roots are chosen so that both n and k will always be real and positive, as is required by their meaning. It should be noted that $K_i = g/\epsilon_0\omega$ is inherently positive, since g is positive; the energy dissipation $\mathbf{J} \cdot \mathbf{E} = gE^2$ in a passive medium must be positive; K_i always represents an energy loss. On the other hand, $K_r = K$ could be, and indeed can be, either positive or negative. For nonconducting media and static fields, it is positive and greater than 1, but for alternating fields in metals it may be less than 1, zero, or negative. Specific examples can be found in Chapter 19; for the present, we simply take K and g as given, recognizing that in any material they may depend on the frequency ω .

The equations above are exact but complicated; hence it is convenient to examine certain approximations. Most often one or another of the following approximations is valid:

$$1. \quad K_i \ll |K_r|, \quad K_r > 0 \quad \left(\omega \gg \frac{g}{\epsilon} \right);$$

$$n \cong \sqrt{K_r}, \quad k = K_i/2n \ll n \quad (17-62)$$

$$2. \quad K_i \ll |K_r|, \quad K_r < 0 \quad \left(\omega \gg \frac{g}{-\epsilon} \right);$$

$$k \cong \sqrt{-K_r}, \quad n = K_i/2k \ll k \quad (17-63)$$

$$3. \quad K_i \gg |K_r| \quad \left(\omega \ll \frac{g}{|\epsilon|} \right);$$

$$n \cong k \cong \sqrt{K_i/2} \quad (17-64)$$

* More common notations are $\hat{K} = K_1 + iK_2$ or $\hat{K} = K' + iK''$, but we want to reserve the numerical subscripts and primes for other purposes in the next chapter.

For example, (1) holds for a good insulator down to very low frequencies, essentially dc (strictly dc only for a perfect insulator with $g = 0$), and gives $n = \sqrt{K}$ as for a nonconducting medium. Approximation (2) holds for metals in the upper infrared part of the frequency spectrum, while (3) holds for metals at microwave frequencies and below. The dividing frequency between the latter two cases is $\omega \approx 1/\tau$, where τ is the collision time for the free electrons. For pure metals at room temperature, $1/\tau \approx 10^{14} \text{ s}^{-1}$. Physically, in cases (1) and (2) the displacement current predominates and in case (3) the conduction current predominates.

Where the skin depth is important in electrical problems, usually case (3) applies. Equation (17-58) shows that the "wave" is very strongly attenuated in terms of the wavelength. In terms of absolute distance, Eq. (17-57) becomes for this case

$$\delta \equiv \frac{c}{\omega} \sqrt{\frac{2}{K_i}} = \sqrt{\frac{2}{\mu_0 \omega g}} \quad (17-65)$$

As ω approaches zero, δ becomes infinite, consistent with our earlier result that the E -field and current density are uniform in a conductor (not a superconductor*) for dc and also approximately for 60-Hz current. At higher frequencies δ becomes quite small within the range of validity of Eq. (17-65), however.

EXAMPLE 17-1
The Skin Depth of
Conducting Materials at
Radio Frequencies

We should like to calculate numerical values for the skin depth of some conducting materials with the aid of Eq. (17-65). The materials to be considered are silver at microwave frequencies and sea water in the low radio frequency range.

Solution: Fine silver has a conductivity

$$g = 3 \times 10^7 \text{ S/m}^\dagger$$

at microwave frequencies. At a frequency of 10^{10} Hz, which is a common microwave region, the skin depth is

$$\delta = \sqrt{\frac{2}{(2\pi \times 10^{10})(3 \times 10^7)(4\pi \times 10^{-7})}} = 9.2 \times 10^{-5} \text{ cm}$$

Thus at microwave frequencies the skin depth in silver is very small, and consequently, the difference in performance between a pure silver component and a silver-plated brass component would be expected to be negligible. This is indeed the case, and the plating technique is used to reduce the material cost of high-quality waveguide components.

* The skin depth has no connection with the penetration depth discussed in Chapter 15.

† One siemens (S) is one reciprocal ohm; it is a unit of conductance or reciprocal resistance (see Chapter 7). A siemens was formerly called a mho.

For the seawater case, we calculate the frequency at which the skin depth is one meter. For sea water, $\mu = \mu_0$ and $g \approx 4.3 \text{ S/m}$. The expression for the frequency corresponding to a given skin depth δ is

$$\omega = \frac{2}{g\mu_0\delta^2} = \frac{2}{4.3 \times 4\pi \times 10^{-7} \delta^2} \text{ s}^{-1} = \frac{3.70 \times 10^5}{\delta^2} \text{ s}^{-1}$$

which yields

$$f = 58.6 \times 10^3 \text{ Hz}$$

or a frequency of 60 kHz for a skin depth of one meter. If a submarine is equipped with a very sensitive receiver and if a very powerful transmitter is used, it is possible to communicate with a submerged submarine. However, a very low radiofrequency must be used, and even then an extremely severe attenuation of the signal occurs. At five skin depths (5 m in the case calculated above), only 1 percent of the initial electric field remains and only 0.01 percent of the incident power.

The anomalous case $\hat{K} = 0$, which we have so far excluded, allows the existence of a longitudinal, as opposed to a transverse, wave. (Problem 17-14.) In such a wave $\nabla \cdot \mathbf{E} \neq 0$ (even though $\nabla \cdot \mathbf{D} = 0$), so that the polarization charge density does not vanish as in the transverse wave case; only a charge density can give a longitudinal field. Such waves are of some importance in plasmas; the electrostatic oscillation discussed in Section 14-6 is an example.

***17-5**

SPHERICAL WAVES

As an example of a more difficult wave problem, where in fact it is not easy to find even the elementary waves, we consider the wave equation in spherical coordinates. The wave equation for the electric field in a vacuum is

$$\nabla^2 \mathbf{E} - \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0 \quad (17-66)$$

For monochromatic waves, the equation for the spatial portion becomes

$$\nabla^2 \mathbf{E}(\mathbf{r}) + \left(\frac{\omega}{c}\right)^2 \mathbf{E}(\mathbf{r}) = 0 \quad (17-67)$$

The difficulty in using spherical coordinates is that one would like to express the vector $\mathbf{E}(\mathbf{r})$ in terms of radial, azimuthal, and polar *components*, each expressed as functions of the radius, azimuth, and colatitude. If $\mathbf{E}(\mathbf{r})$ is so expressed, then it is not sufficient to use the expression for the Laplacian in spherical coordinates in Eq. (17-67); rather, it is necessary to define the

Laplacian of a vector by

$$\nabla^2 \mathbf{E} = -\nabla \times \nabla \times \mathbf{E} + \nabla \nabla \cdot \mathbf{E} \quad (17-68)$$

The divergence of \mathbf{E} is still zero. However, the radial component of $\nabla \times \nabla \times \mathbf{E}$ involves not only the radial component of \mathbf{E} , but also its azimuthal and polar components. The θ and ϕ components are similarly complicated, and the final result is three simultaneous partial differential equations involving the three components of \mathbf{E} . The separation that occurs for the vector Laplace equation in rectangular coordinates does not occur in spherical coordinates; it is in fact peculiar to rectangular coordinates. It should be pointed out, however, that rectangular components of \mathbf{E} may be used; in this instance they would be written: $E_x(r, \theta, \phi)$, $E_y(r, \theta, \phi)$, $E_z(r, \theta, \phi)$.

A simple procedure circumvents the difficulty discussed above. Consider the scalar Helmholtz equation:

$$\nabla^2 \psi + \left(\frac{\omega}{c}\right)^2 \psi = 0 \quad (17-69)$$

whose solutions are, as will be seen shortly, readily found. Suppose that ψ is any one of the solutions, then $\mathbf{E} = \mathbf{r} \times \nabla \psi$ satisfies the vector Helmholtz equation, Eq. (17-67):

$$-\nabla \times \nabla \times \mathbf{E} + \nabla \nabla \cdot \mathbf{E} + \left(\frac{\omega}{c}\right)^2 \mathbf{E} = 0 \quad (17-70)$$

To verify this, note the identity

$$\mathbf{r} \times \nabla \psi = -\nabla \times (\mathbf{r} \psi) \quad (17-71)$$

which follows from the vector identity

$$\nabla \times (\mathbf{F} \varphi) = \varphi \nabla \times \mathbf{F} - \mathbf{F} \times \nabla \varphi \quad (17-72)$$

and

$$\nabla \times \mathbf{r} = 0 \quad (17-73)$$

Since the divergence of any curl is zero, it is necessary to consider only the curl curl term in Eq. (17-70). The curl of \mathbf{E} can be found by using the vector identity

$$\nabla \times (\mathbf{F} \times \mathbf{G}) = \mathbf{F} \nabla \cdot \mathbf{G} - \mathbf{G} \nabla \cdot \mathbf{F} + (\mathbf{G} \cdot \nabla) \mathbf{F} - (\mathbf{F} \cdot \nabla) \mathbf{G} \quad (17-74)$$

to obtain

$$\nabla \times (\mathbf{r} \times \nabla \psi) = \mathbf{r} \nabla^2 \psi - (\nabla \psi) \nabla \cdot \mathbf{r} + (\nabla \psi \cdot \nabla) \mathbf{r} - (\mathbf{r} \cdot \nabla) \nabla \psi \quad (17-75)$$

As was shown in Problem 1-13, $(\mathbf{A} \cdot \nabla) \mathbf{r} = \mathbf{A}$ for any vector \mathbf{A} ; also, the divergence of \mathbf{r} is three (3). The first term of Eq. (17-75) can be reduced by using the fact that ψ satisfies the scalar Helmholtz equation, thus leaving

only the last term as a possible source of complication. The vector identity

$$\nabla(\mathbf{F} \cdot \mathbf{G}) = (\mathbf{F} \cdot \nabla)\mathbf{G} + (\mathbf{G} \cdot \nabla)\mathbf{F} + \mathbf{F} \times \nabla \times \mathbf{G} + \mathbf{G} \times \nabla \times \mathbf{F} \quad (17-76)$$

with $\mathbf{F} = \mathbf{r}$ and $\mathbf{G} = \nabla\psi$, gives

$$\nabla(\mathbf{r} \cdot \nabla\psi) = (\mathbf{r} \cdot \nabla)\nabla\psi + (\nabla\psi \cdot \nabla)\mathbf{r} \quad (17-77)$$

The last two terms of Eq. (17-76) vanish because the curl of any gradient is zero, as is the curl of \mathbf{r} . Using these relationships in Eq. (17-75) leads to

$$\nabla \times (\mathbf{r} \times \nabla\psi) = -\left(\frac{\omega}{c}\right)^2 \mathbf{r}\psi - 3\nabla\psi + \nabla\psi - \nabla(\mathbf{r} \cdot \nabla\psi) + \nabla\psi \quad (17-78)$$

Finally, taking the curl of Eq. (17-78), we obtain

$$\nabla \times \nabla \times (\mathbf{r} \times \nabla\psi) = -\left(\frac{\omega}{c}\right)^2 \nabla \times \mathbf{r}\psi = \left(\frac{\omega}{c}\right)^2 \mathbf{r} \times \nabla\psi \quad (17-79)$$

which is just the vector Helmholtz equation. No explicit use of the spherical coordinate system has been made. However, since \mathbf{r} is normal to a surface of constant radius in spherical coordinates, the solution $\mathbf{r} \times \nabla\psi$ would be expected to be particularly useful in this system. It is in fact not very useful in other coordinate systems.

Having found that $\mathbf{r} \times \nabla\psi$ is a solution of the vector Helmholtz equation, with ψ a solution of the scalar Helmholtz equation, it becomes pertinent to find out how such solutions can be used to construct electromagnetic waves. The procedure is very simple. The spatial variation of the electric field is taken as

$$\mathbf{E} = \mathbf{r} \times \nabla\psi \quad (17-71)$$

The magnetic field must be so chosen that it, together with \mathbf{E} , satisfies Maxwell's equations. To this end, we write Eq. (17-5) as

$$\nabla \times \mathbf{E} = i\omega\mathbf{B} \quad (17-80)$$

where the standard $e^{-i\omega t}$ time dependence has been assumed. Equation (17-78) gives the curl of \mathbf{E} explicitly or, in a shorter form,

$$\mathbf{B} = -i\frac{1}{\omega}\nabla \times (\mathbf{r} \times \nabla\psi) \quad (17-81)$$

Since the divergence of any curl vanishes, Eq. (17-4) is satisfied. That Eq. (17-6) is satisfied is obvious from the fact that \mathbf{E} and \mathbf{B} are both solutions of the wave equation, which in turn represents a combination of Eqs. (17-5) and (17-6).

The solution represented by Eqs. (17-71) and (17-81) is not the most

general solution that can be derived from a given ψ . Another solution is obtained by putting

$$\mathbf{B}' = \frac{1}{c} \mathbf{r} \times \nabla \psi \quad (17-82)$$

and obtaining the electric field from Eq. (17-6),

$$\mathbf{E}' = \frac{ic}{\omega} \nabla \times (\mathbf{r} \times \nabla \psi) \quad (17-83)$$

The considerations detailed above show that \mathbf{E}' , \mathbf{B}' form a solution to Maxwell's equations, just as \mathbf{E} , \mathbf{B} do. The solutions differ in that \mathbf{E} at any point is tangent to a spherical surface through the point with center at the origin of coordinates; on the other hand, \mathbf{B}' has the same property. These facts lead to the solution \mathbf{E} , \mathbf{B} being sometimes called *transverse electric* (TE), and \mathbf{E}' , \mathbf{B}' *transverse magnetic* (TM), transverse meaning perpendicular to the radial direction.

In the preceding paragraphs, the problem of solving the vector Helmholtz equation has been reduced to that of solving the scalar Helmholtz equation. In spherical coordinates, this equation can be solved by the technique of separation of variables already familiar from potential problems (Chapter 3) and Section 17-4. In spherical coordinates, the scalar Helmholtz equation is

$$\begin{aligned} \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \psi}{\partial \theta} \right) \\ + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} + \kappa^2 \psi = 0 \end{aligned} \quad (17-84)$$

where $\kappa^2 = (\omega/c)^2$ and ψ is assumed to have the form

$$\psi = R(r)\Theta(\theta)\Phi(\phi) \quad (17-85)$$

Substituting this assumed form for ψ in Eq. (17-84) and dividing by ψ gives

$$\begin{aligned} \frac{1}{R} \sin^2 \theta \frac{d}{dr} r^2 \frac{dR}{dr} + \frac{1}{\Theta} \sin \theta \frac{d}{d\theta} \sin \theta \frac{d\Theta}{d\theta} \\ + \frac{1}{\Phi} \frac{d^2 \Phi}{d\phi^2} + \kappa^2 r^2 \sin^2 \theta = 0 \end{aligned} \quad (17-86)$$

after multiplying by $r^2 \sin^2 \theta$. The third term depends only on ϕ , and this term is the only one that depends on ϕ . Consequently, this term must be a constant, which is chosen to be $-m^2$. In other words,

$$\frac{d^2 \Phi_m}{d\phi^2} + m^2 \Phi_m = 0 \quad (17-87)$$

where the subscript m serves to indicate that Φ depends on m . Rewriting

TABLE 17-2

Associated Legendre Polynomials, $P_l^m(u)$, where $u = \cos \theta$

Designation	Functional form
$P_0(u)$	1
$P_1(u)$	$u = \cos \theta$
$P_1^1(u)$	$(1 - u^2)^{1/2} = \sin \theta$
$P_2(u)$	$\frac{1}{2}(3u^2 - 1) = \frac{1}{4}(3 \cos 2\theta + 1)$
$P_2^1(u)$	$3u(1 - u^2)^{1/2} = \frac{3}{2} \sin 2\theta$
$P_2^2(u)$	$3(1 - u^2) = \frac{3}{2}(1 - \cos 2\theta)$
$P_3(u)$	$\frac{1}{2}(5u^3 - 3u)$
$P_3^1(u)$	$\frac{3}{2}(1 - u^2)^{1/2}(5u^2 - 1)$
$P_3^2(u)$	$15u(1 - u^2)$
$P_3^3(u)$	$15(1 - u^2)^{3/2}$

Eq. (17-86) using (17-87) gives

$$\frac{1}{R} \frac{d}{dr} r^2 \frac{dR}{dr} + \kappa^2 r^2 + \frac{1}{\Theta} \frac{1}{\sin \theta} \frac{d}{d\theta} \sin \theta \frac{d\Theta}{d\theta} - \frac{m^2}{\sin^2 \theta} = 0 \quad (17-88)$$

The first two terms depend only on r , while the last two depend only on θ . Thus the sum of the last two must be a constant, which is chosen as $-l(l + 1)$. The sum of the first two terms must, of course, be $l(l + 1)$. Thus there result two equations:

$$\frac{1}{\sin \theta} \frac{d}{d\theta} \sin \theta \frac{d\Theta_{lm}}{d\theta} + \left[l(l + 1) - \frac{m^2}{\sin^2 \theta} \right] \Theta_{lm} = 0 \quad (17-89)$$

and

$$\frac{d}{dr} r^2 \frac{dR_l}{dr} - [l(l + 1) - \kappa^2 r^2] R_l = 0 \quad (17-90)$$

The solutions of Eq. (17-87) are well known:

$$\Phi_m = e^{\mp im\phi} \quad (17-91)$$

The solutions of Eq. (17-89) are less well known, but some have already been met in Chapter 3, where solutions for $m = 0$ were discussed. These solutions* are the Legendre polynomials $P_l(\cos \theta)$. The solutions of (17-89) for arbitrary $m \leq l$ are known as the associated Legendre polynomials. They may be defined by

$$P_l^m(u) = (1 - u^2)^{m/2} \frac{d^m}{du^m} P_l(u) \quad (17-92)$$

with $u = \cos \theta$. It is clear that $P_l^0(u) = P_l(u)$, the ordinary Legendre

* In Chapter 3, these functions were written $P_l(\theta)$. Since, however, the Legendre polynomials are polynomials in $\cos \theta$, it is more common to write $P_l(\cos \theta)$; we follow this practice in the present chapter as well as in the succeeding ones.

polynomial. For $m \neq 0$ the functions are given in Table 17-2.

Finally, Eq. (17-90) must be considered. The change of variable from r to $\xi = \kappa r$ is readily accomplished, and the resulting equation is

$$\frac{d}{d\xi} \xi^2 \frac{d}{d\xi} R_l - [l(l+1) - \xi^2] R_l = 0 \quad (17-93)$$

The substitution $R_l = \xi^{-1/2} Z_l$ transforms this equation into

$$\xi^2 \frac{d^2 Z_l}{d\xi^2} + \xi \frac{dZ_l}{d\xi} - [(l + \frac{1}{2})^2 - \xi^2] Z_l = 0 \quad (17-94)$$

This equation, which is very familiar to mathematical physicists, is known as Bessel's equation. The solutions of the equation are also well known and have been extensively investigated and indeed tabulated. The common solutions are designated $J_{l+1/2}(\kappa r)$ and $N_{l+1/2}(\kappa r)$ and are known respectively as the Bessel function and the Neumann function, of order $l + \frac{1}{2}$. For purposes of the wave equation, it is extremely convenient to define spherical Bessel functions by

$$j_l(\kappa r) = \sqrt{\pi/2\kappa r} J_{l+1/2}(\kappa r), \quad n_l(\kappa r) = \sqrt{\pi/2\kappa r} N_{l+1/2}(\kappa r) \quad (17-95)$$

and from these in turn we obtain

$$h_l^{(1)}(\kappa r) = j_l(\kappa r) + in_l(\kappa r), \quad h_l^{(2)} = j_l(\kappa r) - in_l(\kappa r) \quad (17-96)$$

The functions $j_l(\kappa r)$, $n_l(\kappa r)$, $h_l^{(1)}(\kappa r)$, and $h_l^{(2)}(\kappa r)$ are all solutions of the radial equation, Eq. (17-90). These functions are tabulated for $l = 0, 1$, and 2 in Table 17-3. The h 's are particularly convenient for radiation problems

TABLE 17-3
Spherical Bessel and
Neumann Functions

Type	Functional form
$j_0(\rho)$	$(1/\rho) \sin \rho$
$n_0(\rho)$	$-(1/\rho) \cos \rho$
$h_0^{(1)}(\rho)$	$-(i/\rho)e^{i\rho}$
$h_0^{(2)}(\rho)$	$(i/\rho)e^{-i\rho}$
$j_1(\rho)$	$(1/\rho^2) \sin \rho - (1/\rho) \cos \rho$
$n_1(\rho)$	$-(1/\rho) \sin \rho - (1/\rho^2) \cos \rho$
$h_1^{(1)}(\rho)$	$-(1/\rho)e^{i\rho}(1 + i/\rho)$
$h_1^{(2)}(\rho)$	$-(1/\rho)e^{-i\rho}(1 - i/\rho)$
$j_2(\rho)$	$\left[\frac{3}{\rho^3} - \frac{1}{\rho}\right] \sin \rho - \frac{3}{\rho^2} \cos \rho$
$n_2(\rho)$	$-\frac{3}{\rho^2} \sin \rho - \left[\frac{3}{\rho^3} - \frac{1}{\rho}\right] \cos \rho$
$h_2^{(1)}(\rho)$	$(i/\rho)e^{i\rho} \left(1 + \frac{3i}{\rho} - \frac{3}{\rho^2}\right)$
$h_2^{(2)}(\rho)$	$-(i/\rho)e^{-i\rho} \left(1 - \frac{3i}{\rho} - \frac{3}{\rho^2}\right)$

because for large values of r , they behave as

$$h_l^{(1)}(\kappa r) \xrightarrow{\kappa r \rightarrow \infty} \frac{(-i)^{l+1} e^{i\kappa r}}{\kappa r} \quad h_l^{(2)}(\kappa r) \xrightarrow{\kappa r \rightarrow \infty} \frac{i^{l+1} e^{-i\kappa r}}{\kappa r}$$

and thus lead to outgoing and ingoing spherical waves.

A general form for ψ may be written as

$$\psi_{lm} = \sqrt{\pi/2\kappa r} Z_l(\kappa r) P_l^m(\cos \theta) e^{\mp im\phi} \quad (17-97)$$

The corresponding vector fields are computed by using Eqs. (17-71) and (17-81) for the TE waves, and Eqs. (17-82) and (17-83) for the TM waves.

The simplest interesting choice of ψ is ψ_{10} , which is just

$$\psi_{10} = \frac{1}{\kappa r} e^{i\kappa r} \left[1 + \frac{i}{\kappa r} \right] \cos \theta \quad (17-98)$$

The gradient of ψ_{10} is

$$\begin{aligned} \nabla \psi_{10} &= \mathbf{a}_r e^{i\kappa r} \left[\frac{i}{r} - \frac{2}{\kappa r^2} - \frac{2i}{\kappa^2 r^3} \right] \cos \theta \\ &\quad - \mathbf{a}_\theta e^{i\kappa r} \left[\frac{1}{\kappa r^2} + \frac{i}{\kappa^2 r^3} \right] \sin \theta \end{aligned} \quad (17-99)$$

The spatial portion of the electric field is

$$\mathbf{E} = \mathbf{r} \times \nabla \psi_{10} = -\mathbf{a}_\phi E_0 e^{i\kappa r} \left[\frac{1}{\kappa r} + \frac{i}{\kappa^2 r^2} \right] \sin \theta \quad (17-100)$$

where E_0 has been introduced to make the equation dimensionally correct. The surfaces of constant phase, $\kappa r = \text{constant}$, are spheres. (The surfaces of constant amplitude are not, however.) The spatial dependence of the magnetic induction is given by

$$\begin{aligned} \mathbf{B} &= -i \frac{1}{\omega} \nabla \times \mathbf{E} \\ &= i \frac{1}{\omega} E_0 e^{i\kappa r} \left[\frac{1}{\kappa r^2} + \frac{i}{\kappa^2 r^3} \right] 2 \cos \theta \mathbf{a}_r \\ &\quad - i \frac{1}{\omega} E_0 e^{i\kappa r} \left[\frac{i}{r} - \frac{1}{\kappa r^2} - \frac{i}{\kappa^2 r^3} \right] \sin \theta \mathbf{a}_\theta \end{aligned} \quad (17-101)$$

As will be seen later, these are just the fields (TE) produced by a radiating magnetic dipole. It is interesting to note that only the portions of \mathbf{E} and \mathbf{B} , which are proportional to $1/r$ contribute to the net radiation. All other terms give terms in the Poynting vector that fall off more rapidly than $1/r^2$ and consequently have integrals over spherical surfaces that vanish as the radii of these spherical surfaces go to infinity. The spherical wave solutions

are applicable in considering the radiation from bounded sources, which will be treated in Chapter 20 from another point of view.

17-6

SUMMARY

The transverse wave solutions of Maxwell's equations are most simply expressed in terms of **plane waves**:

$$\mathbf{E}(\mathbf{r}, t) = \hat{\mathbf{E}}e^{-i(\omega t - \boldsymbol{\kappa} \cdot \mathbf{r})}, \quad \mathbf{B}(\mathbf{r}, t) = \hat{\mathbf{B}}e^{-i(\omega t - \boldsymbol{\kappa} \cdot \mathbf{r})}$$

These solutions exist for any *frequency* ω and amplitude $\hat{\mathbf{E}}$, so long as $\boldsymbol{\kappa} \cdot \hat{\mathbf{E}} = 0$. The magnitude of $\boldsymbol{\kappa}$ is determined by the transverse dispersion relation, although its direction is arbitrary. The vector $\boldsymbol{\kappa}$ (real part) specifies the direction of propagation and *wavelength* ($\lambda = 2\pi/\kappa$). The B -field of the wave is determined by the other quantities.

- In vacuum, the **dispersion relation** is

$$\kappa = \omega/c$$

where $c = 1/\sqrt{\epsilon_0\mu_0}$ is the velocity of light. The vectors $\boldsymbol{\kappa}$, \mathbf{E} , \mathbf{B} form a right-handed orthogonal set. The magnitude of \mathbf{B} is

$$B = E/c$$

at every instant. If the two components of \mathbf{E} (in the plane perpendicular to $\boldsymbol{\kappa}$) are in phase, the polarization is linear; if they are out of phase, it is elliptical. The magnitude of the **Poynting vector** and the energy density u are related by

$$S = cu$$

and the direction of \mathbf{S} is along $\boldsymbol{\kappa}$.

- In a linear dielectric medium (nonconducting, nonmagnetic), all the above results still hold if c is replaced by c/n , where the **index of refraction** n is related to the dielectric constant K by

$$n = \sqrt{K}$$

(In a linear magnetic medium, $n = \sqrt{KK_m}$, but this generalization has little application.)

- In a conducting medium all of the same formal results hold if the **complex dielectric constant**

$$\hat{K} = K + ig/\epsilon_0\omega = K_r + iK_i$$

is used instead of K . This has the consequence that the refractive index

$$\hat{n} = n + ik$$

and the propagation vector

$$\hat{\mathbf{k}} = \boldsymbol{\kappa}_r + i\boldsymbol{\kappa}_i$$

are also complex. The wave is attenuated:

$$\mathbf{E}(\mathbf{r}, t) = (\hat{\mathbf{E}}e^{-\boldsymbol{\kappa}_i \cdot \mathbf{r}})e^{-i(\omega t - \boldsymbol{\kappa}_r \cdot \mathbf{r})}$$

In the simplest case, $\boldsymbol{\kappa}_r = n\boldsymbol{\omega}/c$, $\boldsymbol{\kappa}_i = k\boldsymbol{\omega}/c$; but in general all of the physical relations are more complicated.

- In general

$$\hat{K} = \hat{n}^2$$

The relation between the *real optical constants* and dielectric constants is

$$K_r = n^2 - k^2, \quad K_i = 2nk$$

and these can be inverted explicitly. The result is complicated, unless certain extreme cases are applicable, as is usually true.

- For conductors the **attenuation length**

$$\delta = c/\omega k = \frac{n \lambda}{k 2\pi}$$

also called the **skin depth**, is important. For at least moderately good conductors, below the infrared frequency range, the approximation

$$n \cong k \cong \sqrt{g/2\epsilon_0\omega}$$

can be expected to hold, so that

$$\delta \cong \sqrt{2/\mu_0\omega g}$$

Here g can be taken as the dc conductivity.

- Aside from plane waves, the most interesting set of transverse vector wave solutions is the *spherical waves*. They are much more complicated, but applicable in radiation problems.

PROBLEMS

17-1 Show that if $\mathbf{F}(\mathbf{r}) = \mathbf{A}e^{i\mathbf{k} \cdot \mathbf{r}}$, where \mathbf{A} is constant,

$$\nabla \cdot \mathbf{F} = i\mathbf{k} \cdot \mathbf{F}, \quad \nabla \times \mathbf{F} = i\mathbf{k} \times \mathbf{F}$$

17-2 Show that for a plane wave in vacuum

$$\frac{E}{H} = \sqrt{\frac{\mu_0}{\epsilon_0}} = 377 \Omega$$

This resistance is called the *impedance of free space*.

17-3 Two plane waves have the same ω , \mathbf{k} , and amplitude E , but opposite circular polarization (i.e., left and right). Show that the superposition of the two waves is linearly polarized, with amplitude $2E$.

- 17-4** Sketch a figure like Fig. 17-1 for $E_s = 2E_p$, with $\phi = 30^\circ$ and with $\phi = 60^\circ$.
- 17-5** The earth receives about 1300 W/m^2 radiant energy from the sun. Assuming the energy to be in the form of a plane polarized monochromatic wave and assuming normal incidence, compute the magnitude of the electric and magnetic field vectors in the sunlight.
- 17-6** Suppose \mathbf{A} and \mathbf{B} are complex vectors. (a) Calculate $\text{Re } \mathbf{A} \cdot \text{Re } \mathbf{B}$ and compare it with $\text{Re } (\mathbf{A} \cdot \mathbf{B})$. (b) Show that $\mathbf{A} \cdot \mathbf{B} = 0$ does not imply that $\text{Re } \mathbf{A} \cdot \text{Re } \mathbf{B} = 0$.
- 17-7** Consider two plane waves in vacuum with the same ω , \mathbf{k} , and polarization direction \mathbf{p} , but with different amplitudes and phases: E_1 , 0 and E_2 , ϕ . Calculate the time-average Poynting vector $\bar{\mathbf{S}}$ of the superposition of the two waves. Note the interference effect due to the phase difference ϕ , which would not occur if the two waves had perpendicular polarization directions.
- 17-8** Consider a standing wave in vacuum which is the superposition of two plane waves of the same frequency, amplitude, and linear polarization, but with opposite \mathbf{k} . (a) Calculate the Poynting vector $\mathbf{S}(\mathbf{r}, t)$ of the superposed waves. (Note that the origin is not arbitrary.) (b) What is $\bar{\mathbf{S}}$?
- 17-9** For a plane wave in a conducting medium

$$\mathbf{B} = \frac{\hat{n}}{c} \mathbf{u} \times \mathbf{E}$$

Suppose \mathbf{E} is elliptically polarized, with $\hat{\mathbf{E}} = E_p e^{i\phi} \mathbf{p} + E_s \mathbf{s}$. Prove that at each instant of time

$$\text{Re } \mathbf{E} \cdot \text{Re } \mathbf{B} = -\frac{k}{c} E_p E_s \sin \phi$$

- 17-10** For metals in the infrared spectral region, it happens that $K_i = -K_r$, at a frequency $\omega = g/(-\epsilon) \approx 10^{14} \text{ s}^{-1}$. Calculate the optical constants n and k for this case, in terms of K_i .
- 17-11** For a dielectric which becomes absorbing at high frequencies, or for a semiconductor, it happens that $K_i = K_r$. Calculate n and k for this case, in terms of K_r . Find δ/λ , the ratio of the attenuation length to the wavelength.
- 17-12** Show that in a nearly transparent medium of refractive index n , the attenuation length δ is related to the conductivity g by

$$\delta = \frac{2n}{g\sqrt{\mu_0/\epsilon_0}}$$

where $\sqrt{\mu_0/\epsilon_0} = 377 \Omega$.

- 17-13** A Poynting vector proportional to E^2 falls off as $e^{-\alpha z}$, where $\alpha = 2/\delta$ is called the *absorption coefficient*. The power loss is often expressed in decibels per meter (dB/m), where a *decibel* is defined as ten times the common logarithm of the ratio of the initial to the final energy flux per unit area. (a) Show that power loss = 4.34α dB/m. (b) From the result of Problem 17-12, calculate the (optical, not dc) conductivity a lightwave communications medium must have in order to achieve a loss as low as 1 dB/km, assuming a refractive index of $n = 1.5$.

17-14 Suppose the dielectric constant $K = 0$. (a) Show that Maxwell's equations with no external charge or current densities have a *longitudinal* solution in which $\mathbf{H} = 0$ and $\boldsymbol{\kappa} \times \mathbf{E} = 0$, but $\boldsymbol{\kappa} \cdot \mathbf{E} \neq 0$. (b) Show that there is a polarization charge density $\rho_p = i\boldsymbol{\kappa}\epsilon_0 E$. (The equation $K(\boldsymbol{\kappa}, \omega) = 0$ is called the *longitudinal dispersion relation*.)

17-15 According to quantum mechanics, the energy of a photon is hf where h is Planck's constant, 6.63×10^{-34} Js. A beam of visible light in vacuum ($f = 6 \times 10^{14}$ Hz) is focused to a cross-section of 1 square mm and is, approximately, a long, plane wave. If the electric field in the wave is 1.0 V/m, calculate the length of the wavetrain in one "photon."

17-16 What is the skin depth in mercury metal at $f = 10^5$ Hz?

Monochromatic Waves in Bounded Regions

The solutions of Maxwell's equations found in the preceding chapters will now be used to solve problems of practical interest. Two general classes of problems will be considered: boundary-value problems and radiation from prescribed charge-current distributions. In the first class of problems, solutions of the homogeneous wave equation are so combined as to satisfy the appropriate boundary conditions. In the second class, solutions of the inhomogeneous wave equation with specified sources are required and boundary conditions are largely ignored, except for such things as insistence on outgoing waves and that the fields fall off as $1/r$ at large distances. The first class of problems is the subject of this chapter; the second class of problems is considered in Chapter 20.

18-1

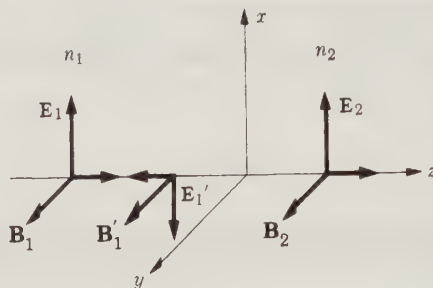
REFLECTION AND REFRACTION AT THE BOUNDARY OF TWO NONCONDUCTING MEDIA: NORMAL INCIDENCE

Electromagnetic waves propagating in materials* usually enter the material through a boundary between it and another medium, perhaps air or vacuum. Treating this problem requires the application of the boundary conditions derived in Chapter 16. We begin with the simplest possible case: a plane wave normally incident on a plane dielectric interface. Experience leads us to expect that the incident wave will be accompanied by both reflected and transmitted waves. In fact, we will see that the boundary conditions can be satisfied only if transmitted and reflected waves are

* In this chapter we assume that all of the materials are linear.

FIGURE 18-1

Reflection and transmission at normal incidence



present. We continue to consider only nonmagnetic media and describe the fields in terms of \mathbf{B} rather than \mathbf{H} . The three waves are each constrained to satisfy the relations between $\boldsymbol{\kappa}$, \mathbf{E} , and \mathbf{B} developed in Section 17-1. The situation is described in Fig. 18-1.

In this figure \mathbf{E}_1 , \mathbf{B}_1 describe the incident wave traveling in the plus z -direction, \mathbf{E}'_1 , \mathbf{B}'_1 describe the reflected wave traveling in the minus z -direction, and \mathbf{E}_2 , \mathbf{B}_2 describe the transmitted wave. The interface is taken as coincident with the xy -plane at $z = 0$, with medium 1 on the left and medium 2 on the right. The electric fields, which are at first assumed to be linearly polarized in the x -direction, are described by

$$\begin{aligned} \mathbf{E}_1 &= \mathbf{i}E_{1x}e^{i(\kappa_1 z - \omega t)} \\ \mathbf{E}'_1 &= -\mathbf{i}E'_{1x}e^{-i(\kappa_1 z + \omega t)} \\ \mathbf{E}_2 &= \mathbf{i}E_{2x}e^{i(\kappa_2 z - \omega t)} \end{aligned} \quad (18-1)$$

where

$$\kappa_1 = n_1 \frac{\omega}{c} \quad \text{and} \quad \kappa_2 = n_2 \frac{\omega}{c} \quad (18-2)$$

with n_1 and n_2 being the indices of refraction of the two media. From Eq. (17-22),

$$\hat{\mathbf{B}} = \frac{n}{c} \mathbf{u} \times \hat{\mathbf{E}}$$

where in this case $\mathbf{u} = \mathbf{k}$ for the incident and transmitted waves and $\mathbf{u} = -\mathbf{k}$ for the reflected wave, the magnetic fields associated with the electric fields of Eq. (18-1) are given by

$$\begin{aligned} c\mathbf{B}_1 &= \mathbf{j}n_1 E_{1x} e^{i(\kappa_1 z - \omega t)} \\ c\mathbf{B}'_1 &= \mathbf{j}n_1 E'_{1x} e^{-i(\kappa_1 z + \omega t)} \\ c\mathbf{B}_2 &= \mathbf{j}n_2 E_{2x} e^{i(\kappa_2 z - \omega t)} \end{aligned} \quad (18-3)$$

Clearly the reflected and transmitted waves must have the same frequency

ω as the incident wave if boundary conditions at $z = 0$ are to be satisfied for all t . Since the normal components of the fields vanish, only the boundary conditions on the tangential components need be considered. The E -field must be continuous at the boundary, so from Eq. (18-1) at $z = 0$,

$$E_{1x} - E'_{1x} = E_{2x} \quad (18-4)$$

The H -field must also be continuous, and for nonmagnetic media ($\mu_1 = \mu_2 = \mu_0$) so must the B -field:

$$n_1(E_{1x} + E'_{1x}) = n_2 E_{2x} \quad (18-5)$$

Equations (18-4) and (18-5) can be solved simultaneously for the amplitudes E'_{1x} and E_{2x} in terms of the given amplitude E_{1x} of the incident wave:

$$E'_{1x} = \frac{n_2 - n_1}{n_2 + n_1} E_{1x}, \quad E_{2x} = \frac{2n_1}{n_2 + n_1} E_{1x}$$

The ratios of the reflected and transmitted wave amplitudes to the incident amplitude are determined entirely by the refractive indices of the two media. These amplitudes in turn determine the amplitudes of the magnetic fields through Eq. (18-3).

Since only the field ratios are determined, it is convenient to introduce a special notation for them,

$$\frac{E'_{1x}}{E_{1x}} = r_{12}, \quad \frac{E_{2x}}{E_{1x}} = t_{12} \quad (18-6)$$

r_{12} and t_{12} are called *Fresnel coefficients for normal incidence* for reflection and transmission, respectively. The subscripts indicate that the wave is incident from medium 1 onto medium 2. Thus the solution is given as

$$r_{12} = \frac{n_2 - n_1}{n_2 + n_1}, \quad t_{12} = \frac{2n_1}{n_1 + n_2} \quad (18-7)$$

What is usually measurable is not the reflected and transmitted E -fields, but the reflected and transmitted average energy fluxes per unit area. The latter are given by the Poynting vector and are called the *intensities* of the waves. According to Eqs. (17-34) and (17-36), in each medium

$$\bar{S} = \frac{1}{2} \frac{n}{\mu_0 c} (E_p^2 + E_s^2) \quad (18-8)$$

for any polarization.

In this example, we have chosen $E_p = E_x$, $E_s = 0$.

We define the *reflectance* R_n and the *transmittance* T_n for normal incidence by the ratios of the intensities

$$\frac{\bar{S}'_1}{\bar{S}_1} = R_n, \quad \frac{\bar{S}_2}{\bar{S}_1} = T_n \quad (18-9)$$

Then from Eqs. (18-8) and (18-6),

$$R_n = r_{12}^2, \quad T_n = \frac{n_2}{n_1} t_{12}^2 \quad (18-10)$$

With the Fresnel coefficients given by Eq. (18-7), the reflectance and transmittance of Eq. (18-10) satisfy

$$R_n + T_n = 1 \quad (18-11)$$

for any pair of nonconducting media. This identity is an expression of energy conservation at the interface.

So far we have considered only linearly polarized radiation. If the incident wave is elliptically polarized, we must consider the perpendicular components $\hat{E}_y = \hat{E}_s$ in each medium, in addition to the x -components. They will also satisfy equations like (18-6), with the same Fresnel coefficients. The three y -components are all in phase with each other, although they are out of phase with the three x -components. From Eq. (18-8), we see that the intensities associated with the $x(p)$ and $y(s)$ components simply add,

$$\bar{S} = \bar{S}_p + \bar{S}_s \quad (18-12)$$

regardless of the phase difference between the p - and s -components—that is, regardless of the degree of elliptical polarization. Thus Eqs. (18-9) to (18-11) hold separately for each polarization component and also for the total intensities.

For a typical air–glass interface, where $n_2 = 1.5$ and $n_1 = 1$, the reflection and transmission coefficients are

$$R_n = 0.04, \quad T_n = 0.96.$$

Thus, as is expected from Eq. (18-11), all of the incident energy is either reflected or transmitted—there is no place to store energy in the interface. A further interesting fact is obtained by examining Eq. (18-6); namely, if n_2 is greater than n_1 the first ratio r_{12} is positive, and E'_{1x} is in phase with E_{1x} at the boundary.

As a second example, in water $n_2 = 1.33$ for visible light, so that $R_n = 0.02$. Below $\omega \approx 10^{11} \text{ s}^{-1}$, however, $K_2 = 81$. For radio frequencies, pure water approximates a nonconductor (Case 1 in Section 17-4) and so $n_2 = \sqrt{K_2} = 9$,

$$R_n = 0.64$$

For sea water, a better conductor, the reflectance is much larger yet, as we will find in Section 18-4.

18-2 REFLECTION AND REFRACTION AT THE BOUNDARY OF TWO NONCONDUCTING MEDIA: OBLIQUE INCIDENCE

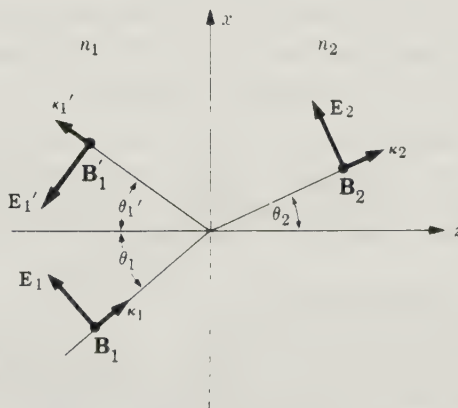
A more general case than that discussed in the preceding section is that of reflection of obliquely incident plane waves by a plane dielectric interface. Consideration of this case leads to three well-known optical laws: Snell's law, the law of reflection, and Brewster's law governing polarization by reflection.

The situation is described by Fig. 18-2, if the propagation vectors κ_1 , κ_1' , and κ_2 are coplanar and lie in the xz -plane, and the electric field vectors \mathbf{E}_1 , \mathbf{E}_1' , and \mathbf{E}_2 also lie in this plane.* The electric fields of the incident, reflected, and transmitted waves are given by

$$\begin{aligned} \mathbf{E}_1 &= \hat{\mathbf{E}}_{1p} e^{i(\kappa_1 \cdot \mathbf{r} - \omega t)} \\ \mathbf{E}_1' &= \hat{\mathbf{E}}'_{1p} e^{i(\kappa_1' \cdot \mathbf{r} - \omega t)} \\ \mathbf{E}_2 &= \hat{\mathbf{E}}_{2p} e^{i(\kappa_2 \cdot \mathbf{r} - \omega t)} \end{aligned} \quad (18-13)$$

where $\hat{\mathbf{E}}_{1p} = \hat{E}_{1p} \mathbf{p}_1$, $\hat{\mathbf{E}}'_{1p} = \hat{E}'_{1p} \mathbf{p}'_1$, $\hat{\mathbf{E}}_{2p} = \hat{E}_{2p} \mathbf{p}_2$. The propagation vectors are $\kappa_1 = \kappa_1 \mathbf{u}_1$, and so on, and the unit normal to the boundary is $\mathbf{n} = \mathbf{k}$. The

FIGURE 18-2
Reflection and refraction—oblique incidence. The xz -plane is the plane of incidence. The vectors \mathbf{B}_1 , \mathbf{B}_1' , and \mathbf{B}_2 are directed out of the paper.



* It will be proved that the propagation vectors are always coplanar. The most general electric field vector can be resolved into a component in the xz -plane (plane of incidence) and a component perpendicular to this plane. The reflection and transmission of these two components are governed by different laws. The choice illustrated gives Brewster's law.

plane defined by $\boldsymbol{\kappa}_1$ and \mathbf{n} is called the *plane of incidence*, and its normal is in the direction of $\mathbf{n} \times \boldsymbol{\kappa}_1$. In fact, $(\mathbf{n} \times \boldsymbol{\kappa}_1)/|\mathbf{n} \times \boldsymbol{\kappa}_1| = \mathbf{j}$, the unit vector in the y direction. The p -component of polarization has been chosen to lie parallel to the plane of incidence (p for "parallel"). In general there is also an s -component (not shown in Fig. 18-2) of the amplitude of each wave, $\hat{\mathbf{E}}_{1s} = \hat{E}_{1s}\mathbf{s}_1$, $\hat{\mathbf{E}}'_{1s} = \hat{E}'_{1s}\mathbf{s}'_1$, $\hat{\mathbf{E}}_{2s} = \hat{E}_{2s}\mathbf{s}_2$. For each of the three waves $\mathbf{s} = \mathbf{u} \times \mathbf{p}$ and $\mathbf{p} = \mathbf{s} \times \mathbf{u}$, so that

$$\mathbf{s}_1 = \mathbf{s}'_1 = \mathbf{s}_2 = \mathbf{j} \quad (18-14)$$

The s -components are perpendicular to the plane of incidence (s for "senkrecht," the German word for perpendicular).

With oblique incidence it is clear from Eqs. (18-13) that not only must the reflected and transmitted waves have the same frequency as the incident wave, but also the phases must match everywhere on the boundary:

$$\boldsymbol{\kappa}'_1 \cdot \mathbf{r}_0 = \boldsymbol{\kappa}_1 \cdot \mathbf{r}_0 = \boldsymbol{\kappa}_2 \cdot \mathbf{r}_0 \quad (18-15)$$

for any vector \mathbf{r}_0 on the boundary. This condition does not depend on the exact boundary conditions, and it has three interesting consequences.

The first of these is that the four vectors $\boldsymbol{\kappa}_1$, $\boldsymbol{\kappa}'_1$, $\boldsymbol{\kappa}_2$, and \mathbf{n} are *all coplanar*. To prove this statement, we use the vector identity

$$\mathbf{n} \times (\mathbf{n} \times \mathbf{R}) = \mathbf{n}(\mathbf{n} \cdot \mathbf{R}) - \mathbf{R}(\mathbf{n} \cdot \mathbf{n}) \quad (1-11)$$

which holds for any two vectors. If \mathbf{n} is a unit vector and \mathbf{R} is perpendicular to \mathbf{n} , as is the case for the normal to the boundary plane and \mathbf{r}_0 ,

$$\mathbf{n} \times (\mathbf{n} \times \mathbf{r}_0) = -\mathbf{r}_0$$

Writing \mathbf{r}_0 in this form in Eq. (18-15) and interchanging the scalar multiplication with the first vector multiplication gives

$$(\boldsymbol{\kappa}_1 \times \mathbf{n}) \cdot (\mathbf{n} \times \mathbf{r}_0) = (\boldsymbol{\kappa}'_1 \times \mathbf{n}) \cdot (\mathbf{n} \times \mathbf{r}_0) = (\boldsymbol{\kappa}_2 \times \mathbf{n}) \cdot (\mathbf{n} \times \mathbf{r}_0)$$

But since \mathbf{r}_0 is an arbitrary vector in the boundary plane, $\mathbf{n} \times \mathbf{r}_0$ is also arbitrary (it is the same length as \mathbf{r}_0 but is perpendicular to it). Consequently,

$$\mathbf{n} \times \boldsymbol{\kappa}_1 = \mathbf{n} \times \boldsymbol{\kappa}'_1 = \mathbf{n} \times \boldsymbol{\kappa}_2 \quad (18-16)$$

The unit vector $\mathbf{n}' = (\mathbf{n} \times \boldsymbol{\kappa}_1)/|\mathbf{n} \times \boldsymbol{\kappa}_1| = \mathbf{j}$ defines the plane of incidence—that is, the plane of incidence is the plane perpendicular to \mathbf{n}' or, equivalently, the plane containing \mathbf{n} and $\boldsymbol{\kappa}_1$. Since $\mathbf{A} \cdot (\mathbf{A} \times \mathbf{B}) = 0$ for any two vectors \mathbf{A} and \mathbf{B} , it follows from Eq. (18-16), by taking the scalar product with $\boldsymbol{\kappa}_1$, $\boldsymbol{\kappa}'_1$, $\boldsymbol{\kappa}_2$ and \mathbf{n} in turn, that

$$\boldsymbol{\kappa}_1 \cdot (\boldsymbol{\kappa}_1 \times \mathbf{n}) = \boldsymbol{\kappa}'_1 \cdot (\boldsymbol{\kappa}_1 \times \mathbf{n}) = \boldsymbol{\kappa}_2 \cdot (\boldsymbol{\kappa}_1 \times \mathbf{n}) = \mathbf{n} \cdot (\boldsymbol{\kappa}_1 \times \mathbf{n}) = 0$$

Thus, the four vectors $\boldsymbol{\kappa}$, $\boldsymbol{\kappa}'_1$, $\boldsymbol{\kappa}_2$, and \mathbf{n} are all perpendicular to \mathbf{n}' . Consequently all four vectors lie in the plane of incidence and are coplanar.

The second consequence of Eq. (18-15) is that *the angle of incidence is*

equal to the angle of reflection. The angle of incidence θ_1 is given by

$$\boldsymbol{\kappa}_1 \cdot \mathbf{n} = \kappa_1 \cos \theta_1$$

Also,

$$\begin{aligned}\boldsymbol{\kappa}'_1 \cdot \mathbf{n} &= -\kappa'_1 \cos \theta'_1 \\ \boldsymbol{\kappa}_2 \cdot \mathbf{n} &= \kappa_2 \cos \theta_2\end{aligned}\tag{18-17}$$

Therefore

$$\begin{aligned}|\boldsymbol{\kappa}_1 \times \mathbf{n}| &= \kappa_1 \sin \theta_1 \\ |\boldsymbol{\kappa}'_1 \times \mathbf{n}| &= \kappa'_1 \sin \theta'_1 \\ |\boldsymbol{\kappa}_2 \times \mathbf{n}| &= \kappa_2 \sin \theta_2\end{aligned}$$

and Eq. (18-16) then requires that

$$\kappa'_1 \sin \theta'_1 = \kappa_1 \sin \theta_1 = \kappa_2 \sin \theta_2$$

The magnitude κ'_1 of the reflected wave equals κ_1 of the incident wave, since they are propagating with the same frequency in the same medium. Thus,

$$\sin \theta_1 = \sin \theta'_1$$

or

$$\theta_1 = \theta'_1\tag{18-18}$$

The third consequence is *Snell's law*. It too follows from the preceding argument and use of the dispersion relation, Eq. (18-2), to change the κ 's to n 's. The result

$$n_1 \sin \theta_1 = n_2 \sin \theta_2\tag{18-19}$$

is Snell's law. (We note for future reference that none of these algebraic proofs depends upon whether the medium is a perfect insulator or conductor. In the latter case, however, the propagation vector is complex. Care must be used to interpret correctly the significance of the complex equations which result.)

None of these three consequences depends on the boundary conditions for the electric and magnetic fields derived from Maxwell's equations. The first two must hold for any kind of wave; the third, Snell's law, depends on the particular dispersion relation for the waves.

To derive the Fresnel coefficients that will give the ratios of the field amplitudes for oblique incidence, we do need the boundary conditions on the tangential components of the fields. (The conditions on the normal components turn out to be satisfied automatically.) To express the tangential component of the E -field amplitude in vector form, consider the vector identity

$$\mathbf{n} \times (\mathbf{n} \times \hat{\mathbf{E}}) = (\mathbf{n} \cdot \hat{\mathbf{E}})\mathbf{n} - \hat{\mathbf{E}}$$

or

$$\hat{\mathbf{E}} = (\mathbf{n} \cdot \hat{\mathbf{E}})\mathbf{n} - \mathbf{n} \times (\mathbf{n} \times \hat{\mathbf{E}})$$

Since $(\mathbf{n} \cdot \hat{\mathbf{E}})\mathbf{n}$ is just the normal component of $\hat{\mathbf{E}}$, the remainder has to be the tangential component, $-\mathbf{n} \times (\mathbf{n} \times \hat{\mathbf{E}})$, for any vector $\hat{\mathbf{E}}$. Then the boundary condition on the E -field becomes

$$\mathbf{n} \times (\hat{\mathbf{E}}_1 + \hat{\mathbf{E}}'_1) = \mathbf{n} \times \hat{\mathbf{E}}_2 \quad (18-20)$$

after canceling the exponentials (because of Eq. (18-15)) and $\mathbf{n} \times$. Since we are assuming a nonmagnetic medium, the continuity of the tangential B -field is also required:

$$\mathbf{n} \times (\hat{\mathbf{B}}_1 + \hat{\mathbf{B}}'_1) = \mathbf{n} \times \hat{\mathbf{B}}_2 \quad (18-21)$$

The Maxwell curl equations relate $\hat{\mathbf{E}}$ and $\hat{\mathbf{B}}$ to each other in each medium. With the dispersion relations Eq. (18-2) these take the form of Eq. (17-22):

$$\hat{\mathbf{B}} = \frac{n}{c} \mathbf{u} \times \hat{\mathbf{E}} \quad (18-22)$$

or its equivalent

$$\hat{\mathbf{E}} = -\frac{c}{n} \mathbf{u} \times \hat{\mathbf{B}} \quad (18-23)$$

If, for example, we substitute Eq. (18-22) into Eq. (18-21), it becomes

$$n_1 \mathbf{n} \times (\mathbf{u}_1 \times \hat{\mathbf{E}}_1 + \mathbf{u}'_1 \times \hat{\mathbf{E}}'_1) = n_2 \mathbf{n} \times (\mathbf{u}_2 \times \hat{\mathbf{E}}_2) \quad (18-24)$$

Equations (18-20) and (18-24) are a pair of vector equations, which must be solved for $\hat{\mathbf{E}}'_1$ and $\hat{\mathbf{E}}_2$ in terms of $\hat{\mathbf{E}}_1$. We can expand the vector triple products:

$$\mathbf{n} \times (\mathbf{u}_1 \times \hat{\mathbf{E}}_1) = (\mathbf{n} \cdot \hat{\mathbf{E}}_1)\mathbf{u}_1 - (\mathbf{n} \cdot \mathbf{u}_1)\hat{\mathbf{E}}_1$$

and similarly for the others. For the s -component $\hat{\mathbf{E}}_{1s}$, for which $\mathbf{n} \cdot \hat{\mathbf{E}}_{1s} = 0$, this expression simplifies considerably:

$$\mathbf{n} \times (\mathbf{u}_1 \times \hat{\mathbf{E}}_{1s}) = -\cos \theta_1 \hat{\mathbf{E}}_{1s}$$

since $\mathbf{n} \cdot \mathbf{u}_1 = \cos \theta_1$ from Eq. (18-17). Thus for the s -component, Eq. (18-24) becomes

$$n_1 (\cos \theta_1 \hat{\mathbf{E}}_{1s} - \cos \theta'_1 \hat{\mathbf{E}}'_{1s}) = n_2 \cos \theta_2 \hat{\mathbf{E}}_{2s}$$

and since $\theta'_1 = \theta_1$,

$$n_1 \cos \theta_1 (\hat{\mathbf{E}}_{1s} - \hat{\mathbf{E}}'_{1s}) = n_2 \cos \theta_2 \hat{\mathbf{E}}_{2s} \quad (18-25)$$

Taking the cross product of Eq. (18-20) with \mathbf{n} gives for the s -component

$$\hat{\mathbf{E}}_{1s} + \hat{\mathbf{E}}'_{1s} = \hat{\mathbf{E}}_{2s} \quad (18-26)$$

No such simplification occurs for the p -component, and so we are led to

consider the two polarization components separately. Substitution of Eq. (18-23) into Eq. (18-20) will make a corresponding simplification for the p -polarized case,

1. **s -polarization.** Equations (18-25) and (18-26) are easily solved simultaneously. We get

$$\hat{\mathbf{E}}'_{1s} = r_{12s}\hat{\mathbf{E}}_{1s}, \quad \hat{\mathbf{E}}_{2s} = t_{12s}\hat{\mathbf{E}}_{1s} \quad (18-27)$$

where

$$r_{12s} = \frac{n_1 \cos \theta_1 - n_2 \cos \theta_2}{n_1 \cos \theta_1 + n_2 \cos \theta_2} \quad (18-28)$$

$$t_{12s} = \frac{2n_1 \cos \theta_1}{n_1 \cos \theta_1 + n_2 \cos \theta_2} \quad (18-29)$$

2. **p -polarization.** When the E -vectors all lie in the plane of incidence, Eq. (18-22) shows that the corresponding B -vectors all lie along the s -direction. The choice of E -vectors in Fig. 18-2 was made so that all the corresponding B -vectors point in the $+\mathbf{j}$ direction. This time we substitute Eq. (18-23) into Eq. (18-20). Since $\mathbf{n} \cdot \hat{\mathbf{B}}_{1s} = 0 = \mathbf{n} \cdot \hat{\mathbf{B}}'_{1s} = \mathbf{n} \cdot \hat{\mathbf{B}}_{2s}$, the result simplifies to

$$\frac{1}{n_1} \cos \theta_1 (\hat{\mathbf{B}}_{1s} - \hat{\mathbf{B}}'_{1s}) = \frac{1}{n_2} \cos \theta_2 \hat{\mathbf{B}}_{2s} \quad (18-30)$$

Also, Eq. (18-21) simplifies to

$$\hat{\mathbf{B}}_{1s} + \hat{\mathbf{B}}'_{1s} = \hat{\mathbf{B}}_{2s} \quad (18-31)$$

We can write the solution of Eqs. (18-30) and (18-31) as

$$\hat{\mathbf{B}}'_{1s} = r_{12p}\hat{\mathbf{B}}_{1s}, \quad \hat{\mathbf{B}}_{2s} = \frac{n_2}{n_1} t_{12p}\hat{\mathbf{B}}_{1s} \quad (18-32)$$

where

$$r_{12p} = \frac{n_2 \cos \theta_1 - n_1 \cos \theta_2}{n_2 \cos \theta_1 + n_1 \cos \theta_2} \quad (18-33)$$

$$t_{12p} = \frac{2n_1 \cos \theta_1}{n_2 \cos \theta_1 + n_1 \cos \theta_2} \quad (18-34)$$

We write Eqs. (18-32) in this way so that when we calculate the magnitudes of the E -vectors from Eq. (18-23), we get

$$\hat{\mathbf{E}}'_{1p} = r_{12p}\hat{\mathbf{E}}_{1p}, \quad \hat{\mathbf{E}}_{2p} = t_{12p}\hat{\mathbf{E}}_{1p} \quad (18-35)$$

With these *Fresnel coefficients* (i.e., r_{12s} , t_{12s} , r_{12p} , and t_{12p}) we now have a complete solution of the boundary-value problem, since an incident wave of arbitrary polarization can be resolved into p - and

s -components. It should be emphasized that the terms p - and s -polarization always refer to the direction of the E -vector.*

For normal incidence $\theta_1 = 0$ and according to Snell's law $\theta_2 = 0$ as well. Thus, Eqs. (18-33) to (18-35) reduce to Eqs. (18-6) and (18-7) for normal incidence. Equation (18-29) becomes identical to Eq. (18-34), but from Eq. (18-28), we find that $r_{12s} = -r_{12p}$ for normal incidence. For normal incidence, the plane of incidence is undefined, and so the physical result has to be independent of polarization. The difference arises only because \mathbf{E}_{1p} and \mathbf{E}'_{1p} point in opposite directions in Fig. 18-2 when θ_1 goes to zero, whereas \mathbf{E}_{1s} and \mathbf{E}'_{1s} point in the same direction. For s -polarization, the Fresnel coefficients are relations between the E -vectors, Eq. (18-27); for p -polarization, they are relations only between their magnitudes, Eq. (18-35), since the E -vectors all point in different directions with oblique incidence. Using Snell's law, we could write

$$\cos \theta_2 = \sqrt{1 - (n_1/n_2)^2 \sin^2 \theta_1}$$

and thereby express the Fresnel coefficients entirely in terms of the material parameters n_1 , n_2 and the given incident angle θ_1 .

Relations between the intensities can again be obtained from the Fresnel coefficients, treating each polarization direction separately. We define the reflectance and transmittance as the component of the respective time-averaged Poynting vectors that is normal to the boundary, relative to the normal component of the incident Poynting vector,†

$$R_s = \frac{\mathbf{n} \cdot \bar{\mathbf{S}}'_{1s}}{\mathbf{n} \cdot \bar{\mathbf{S}}_{1s}}, \quad T_s = \frac{\mathbf{n} \cdot \bar{\mathbf{S}}_{2s}}{\mathbf{n} \cdot \bar{\mathbf{S}}_{1s}} \quad (18-36)$$

$$R_p = \frac{\mathbf{n} \cdot \bar{\mathbf{S}}'_{1p}}{\mathbf{n} \cdot \bar{\mathbf{S}}_{1p}}, \quad T_p = \frac{\mathbf{n} \cdot \bar{\mathbf{S}}_{2p}}{\mathbf{n} \cdot \bar{\mathbf{S}}_{1p}} \quad (18-37)$$

In terms of the Fresnel coefficients, we have‡

$$R_s = r_{12s}^2, \quad T_s = \frac{n_2 \cos \theta_2}{n_1 \cos \theta_1} t_{12s}^2 \quad (18-38)$$

* In older optics books, there occurs the term "plane of polarization," defined as the plane containing the B -vector and the propagation vector. We will avoid this term altogether.

† In medium 1, $\mathbf{S} = (\mathbf{E}_1 + \mathbf{E}'_1) \times (\mathbf{H}_1 + \mathbf{H}'_1)$. It is possible to show that

$$\mathbf{n} \cdot (\mathbf{E}_1 \times \mathbf{H}'_1 + \mathbf{E}'_1 \times \mathbf{H}_1) = 0$$

in a nonconducting medium (the only case we will consider), so that $\mathbf{n} \cdot \bar{\mathbf{S}} = \mathbf{n} \cdot (\bar{\mathbf{S}}_1 + \bar{\mathbf{S}}'_1)$ and it is meaningful to separate the Poynting vectors.

‡ Alternative definitions of the reflectance and transmittance are occasionally made, as ratios of the magnitudes of the Poynting vectors instead of normal components. This makes no difference in R , but removes the factor $(\cos \theta_2 / \cos \theta_1)$ from T .

$$R_p = r_{12p}^2, \quad T_p = \frac{n_2 \cos \theta_2}{n_1 \cos \theta_1} t_{12p}^2 \quad (18-39)$$

The identities

$$R_s + T_s = 1, \quad R_p + T_p = 1 \quad (18-40)$$

hold for oblique incidence on a nonconductor.

For certain purposes, it is more convenient to have the Fresnel coefficients in the following form:

$$r_{12s} = \frac{\sin(\theta_2 - \theta_1)}{\sin(\theta_2 + \theta_1)} \quad (18-41)$$

$$t_{12s} = \frac{2 \cos \theta_1 \sin \theta_2}{\sin(\theta_2 + \theta_1)} \quad (18-42)$$

$$r_{12p} = \frac{\tan(\theta_1 - \theta_2)}{\tan(\theta_1 + \theta_2)} \quad (18-43)$$

$$t_{12p} = \frac{2 \cos \theta_1 \sin \theta_2}{\sin(\theta_1 + \theta_2) \cos(\theta_1 - \theta_2)} \quad (18-44)$$

These are easily seen to be equivalent to the above forms by using trigonometric identities and Snell's law.

18-3 BREWSTER'S ANGLE: CRITICAL ANGLE

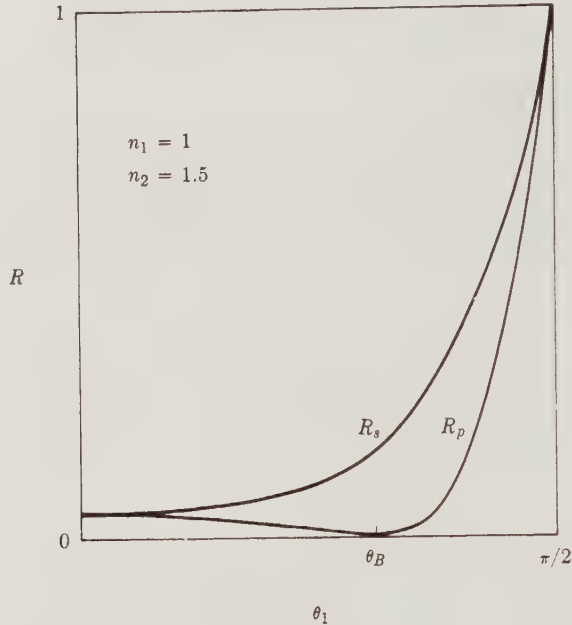
We next consider the dependence of R and T on the angle of incidence for the case of two nonconducting media, using the Fresnel coefficient derived in the preceding section. In every case, $T = 1 - R$, so we will discuss only R . We have already examined the case of normal incidence $\theta_1 = 0$: The polarization does not matter, and R is greater the more the ratio n_2/n_1 differs from unity. For grazing incidence $\theta_1 = \pi/2$, $\cos \theta_1 = 0$, and $R_s = 1 = R_p$, as can be seen most easily from Eqs. (18-28) and (18-33). Near grazing incidence the reflectance is large; this high reflectance is the reason a calm lake is like a mirror. For intermediate angles of incidence, there are two particularly interesting angles.

Can there be a case of zero reflectance? Equations (18-41) and (18-43) show that there can. If $\theta_1 = \theta_2$, then $\tan(\theta_1 - \theta_2) = 0 = \sin(\theta_2 - \theta_1)$ and there is no reflected wave. Unfortunately, if $\theta_1 = \theta_2$, Snell's law requires that $n_1 = n_2$ —that is, the two media are optically indistinguishable. If, on the other hand, $\theta_1 + \theta_2 = \pi/2$, then $\tan(\theta_1 + \theta_2)$ is infinite and the amplitude of the p -polarized reflected wave is again zero. In this case the media are optically distinguishable. Since the s -polarization, \mathbf{E} perpendicular to the plane of incidence, is partially reflected, unpolarized light incident at an angle satisfying $\theta_1 + \theta_2 = \pi/2$ will be polarized by

FIGURE 18-3

Reflectance for s - and p -polarization at an air-glass interface.

Brewster's angle is $\theta_B = 56^\circ$.



reflection. Snell's law

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

provides a means for determining the value of θ_1 . Using $\theta_2 = \pi/2 - \theta_1$ in Snell's law gives

$$n_1 \sin \theta_B = n_2 \sin \left(\frac{\pi}{2} - \theta_B \right) = n_2 \cos \theta_B$$

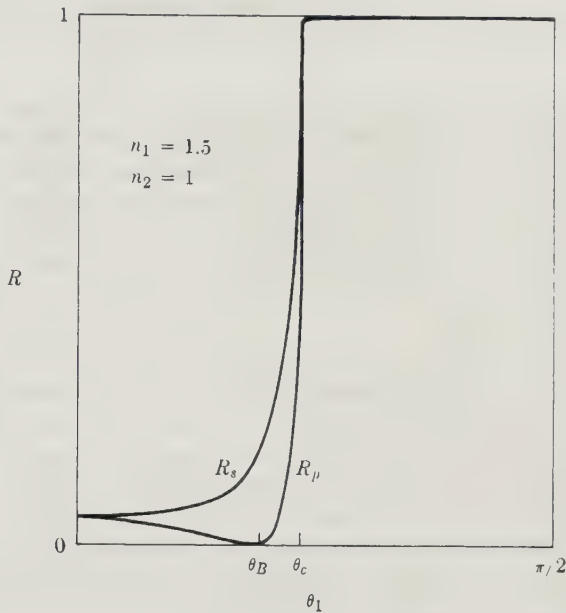
or

$$\tan \theta_B = \frac{n_2}{n_1} \quad (18-45)$$

The quantity θ_B is known as *Brewster's angle*; the relationship between it and the indices of refraction as given in Eq. (18-45) is known as *Brewster's law*. Polarization at the Brewster angle is a practical means of producing polarized radiation, although not the most common one. In Fig. 18-3, the values of R_s and R_p are plotted for all values of θ_1 , with $n_1 = 1$, $n_2 = 1.5$, as for an air-glass interface. The Brewster angle is $\theta_B = 56^\circ$ for this case. The generally lower reflectance for p -polarized light accounts for the usefulness of Polaroid sunglasses. Since most outdoor reflecting surfaces are horizontal, the plane of incidence for most reflected glare reaching the eyes is vertical. On the assumption that one's head is usually erect, the polarizing

FIGURE 18-4

Reflectance for s - and p -polarization at a glass–air interface. Brewster's angle $\theta_B = 34^\circ$ and the critical angle is $\theta_c = 42^\circ$.



lenses are oriented to pass light with E -vector in the vertical plane and eliminate the other more strongly reflected s -component.

There is another case besides grazing incidence in which $R_s = R_p = 1$. From Eqs. (18-28) and (18-33) we see that perfect reflection occurs for $\theta_2 = \pi/2$, as well as for $\theta_1 = \pi/2$. The incident angle for which $\theta_2 = \pi/2$ is called the *critical angle*, $\theta_1 = \theta_c$. From Snell's law,

$$\sin \theta_c = \frac{n_2}{n_1} \quad (18-46)$$

The critical angle is a real angle only if $n_2 < n_1$, but for any pair of transparent materials, this inequality will hold for incidence on one side or the other of their interface. From Eqs. (18-45) and (18-46),

$$\tan \theta_B = \sin \theta_c$$

Since $\tan \theta$ is not restricted in value, there is always a real Brewster's angle; furthermore, since $\tan \theta > \sin \theta$, $\theta_B < \theta_c$. The case of a glass–air interface with $n_1 = 1.5$, $n_2 = 1$ is plotted in Fig. 18-4. For incidence from the glass side, $\theta_B = 34^\circ$ and $\theta_c = 42^\circ$. For incident angles greater than the critical angle, $\theta_1 > \theta_c$, Snell's law gives

$$\sin \theta_2 = \frac{n_1}{n_2} \sin \theta_1 > \frac{n_1}{n_2} \sin \theta_c$$

Since $\sin \theta_c = n_2/n_1$, this equation requires

$$\sin \theta_2 > 1 \quad (18-47)$$

There is, of course, no real angle whose sine is greater than one, but this complication is not serious and is resolved in the next section. The result is that $R_s = R_p = 1$ for all $\theta_1 \geq \theta_c$. This perfect reflection is called *total internal reflection*. It is easily observed by looking into a glass prism or an aquarium or by looking up when swimming under water. It has a very important practical application in the optical fiber, a fine glass fiber through which a light beam is transmitted, as in a waveguide for microwaves. (See Problem 18-4.)

Our examples in this section have involved frequencies in the visible-light range, and transparent materials for which $n = \sqrt{K}$. For nonpolar materials, the same relations hold at all lower frequencies (but not in the ultraviolet and higher frequency ranges). They do not hold at all lower frequencies for optically transparent polar materials made of polar molecules (e.g., water) or ions (e.g., rocksalt), since K is frequency dependent.

18-4

COMPLEX FRESNEL COEFFICIENTS: REFLECTION FROM A CONDUCTING PLANE

The complication that arose in the last section for angles of incidence greater than the critical angle, namely $\sin \theta_2 > 1$, leads us to consider complex Fresnel coefficients. Since $\cos \theta = \sqrt{1 - \sin^2 \theta}$, a real value of $\sin \theta$ greater than 1 implies a pure imaginary value of $\cos \theta$, so that $\cos \hat{\theta}_2$ in the Fresnel coefficients is imaginary and they are complex. They would also be complex if medium 2 were conducting, since in that case \hat{n}_2 is complex. Snell's law

$$n_1 \sin \theta_1 = \hat{n}_2 \sin \hat{\theta}_2$$

shows that then $\sin \hat{\theta}_2$ has to be complex as well. We must, therefore, examine whether our derivation in Section 18-2 holds for complex angles and refractive indices. There is no way to draw Fig. 18-2 with a complex angle $\hat{\theta}_2$, and so we will have to be careful about the geometrical meaning of our conclusions. The derivation, however, did not appeal to the geometry of the figure. It was completely algebraic, and since all the algebraic vector relations hold as well for complex quantities as for real ones, the results are formally correct. We now consider the case where medium 1, the medium from which the wave is incident on the boundary, is nonconducting. This restriction is a physical one and excludes some cases of potential interest, but it is the case of most practical importance. Equation (18-16) then becomes

$$\mathbf{n} \times \mathbf{\kappa}_1 = \mathbf{n} \times \hat{\mathbf{\kappa}}_2 = \mathbf{n} \times \mathbf{\kappa}_{2r} + i\mathbf{n} \times \mathbf{\kappa}_{2i} \quad (18-48)$$

where $\hat{\mathbf{\kappa}}_2 = \mathbf{\kappa}_{2r} + i\mathbf{\kappa}_{2i}$. Since \mathbf{n} and $\mathbf{\kappa}_1$ are both real, the unit vector normal

to the plane of incidence is real and is still just \mathbf{j} . Then, as before,

$$\hat{\mathbf{k}}_2 \cdot \mathbf{j} = 0 \quad (18-49)$$

and both the real and imaginary parts of $\hat{\mathbf{k}}_2$ are perpendicular to \mathbf{j} —that is, in the plane of incidence. (Also, the right-hand side of Eq. (18-48) must be real so $\mathbf{n} \times \mathbf{k}_{2i} = 0$. Thus, the imaginary part of $\hat{\mathbf{k}}_2$ is a vector in the direction of \mathbf{n} —that is, normal to the interface.) The complex angle $\hat{\theta}_2$ is defined algebraically by

$$\hat{\mathbf{k}}_2 \cdot \mathbf{n} = \hat{k}_2 \cos \hat{\theta}_2 \quad (18-50)$$

Then Snell's law becomes

$$n_1 \sin \theta_1 = \hat{n}_2 \sin \hat{\theta}_2 \quad (18-51)$$

where

$$\sin \hat{\theta}_2 = \sqrt{1 - \cos^2 \hat{\theta}_2} \quad (18-52)$$

All of the algebraic manipulations with the boundary conditions on the E - and B -fields are valid, and so the complex Fresnel coefficients are still given by Eqs. (18-28), (18-29), (18-33), and (18-34), with complex \hat{n}_2 and $\cos \hat{\theta}_2$. If they are expressed in polar form,

$$\hat{r}_{12s} = |\hat{r}_{12s}| e^{i\alpha_s}, \quad \hat{r}_{12p} = |\hat{r}_{12p}| e^{i\alpha_p}, \dots$$

and used in Eqs. (18-27) and (18-35),

$$\hat{\mathbf{E}}'_{1s} = |\hat{r}_{12s}| e^{i\alpha_s} \hat{\mathbf{E}}_{1s}, \quad \hat{\mathbf{E}}'_{1p} = |\hat{r}_{12p}| e^{i\alpha_p} \hat{\mathbf{E}}_{1p}, \dots \quad (18-53)$$

it is clear that the reflected and transmitted E -fields are phase-shifted with respect to the incident E -field. The real reflectances for intensity defined by Eqs. (18-36) and (18-37) become

$$R_s = |\hat{r}_{12s}|^2, \quad R_p = |\hat{r}_{12p}|^2 \quad (18-54)$$

since the phases have no effect on the Poynting vectors given by Eq. (18-8). Considerable care is needed to obtain the correct transmittances from the complex quantities, but we shall have no use for these since in most cases they are not measurable in the conducting medium. In place of $R + T = 1$, the identities

$$\hat{r}_{12} = -\hat{r}_{21} \quad (18-55)$$

$$\hat{r}_{12}^2 + \hat{t}_{12}\hat{t}_{21} = 1 \quad (18-56)$$

are useful when one medium is conducting. These hold for both s - and p -polarization.

For normal incidence from air on a conducting medium, with $n_1 = 1$, $\hat{n}_2 = n + ik$, the reflectance is

$$R_n = \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2} \quad (18-57)$$

Since all the transmitted energy is eventually absorbed in a semi-infinite conducting medium, we define the *absorptance* as

$$A = 1 - R \quad (18-58)$$

For normal incidence,

$$A_n = \frac{4n}{(n+1)^2 + k^2} \quad (18-59)$$

The absorptance is small (high reflectance) if $n \ll 1$, or $n \gg 1$, or $k \gg 1$. When $n \cong k \gg 1$ (Case 3 in Section 17-14 with $K_i = g/\epsilon_0\omega \gg 1$),

$$A_n \cong \frac{2}{k} \ll 1 \quad (18-60)$$

In this case, $k \cong \sqrt{K_i/2} = \sqrt{g/2\epsilon_0\omega}$, so that

$$A_n \cong 2\sqrt{2\epsilon_0\omega/g} \quad (18-61)$$

The last expression is called the *Hagen-Rubens relation*; it should hold for moderately good conductors in the microwave region and below, and for metals into the infrared, with g the dc conductivity. With the same values used in calculating the skin depth from Eq. (17-56), we find that for silver at $f = 10^{10} \text{ s}^{-1}$ (3 cm wavelength)

$$A_n = 2\sqrt{2(8.854 \times 10^{-12})(2\pi \times 10^{10})/3 \times 10^7} = 3.9 \times 10^{-4}$$

$$R_n = 0.9996$$

For sea water at $f = 6 \times 10^4 \text{ s}^{-1}$,

$$A_n = 25 \times 10^{-4}$$

$$R_n = 0.9975$$

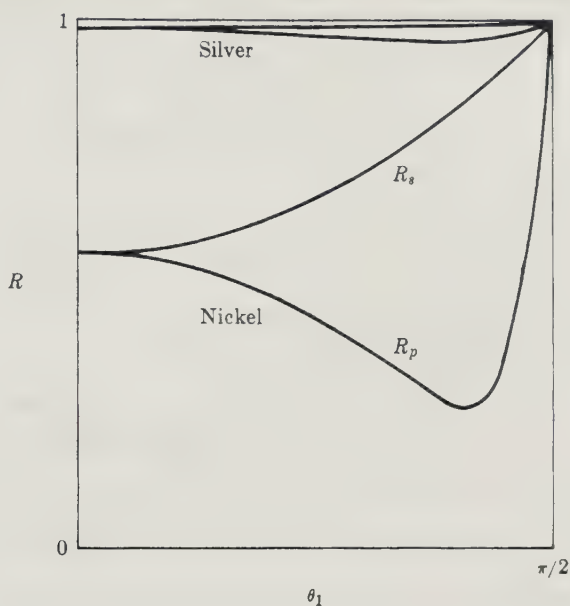
This high reflectance also manifests the problem in communicating with submarines. The absorptance is small in these cases because the skin depth is relatively small. From Eqs. (17-48) and (18-60),

$$A_n = 4\pi \frac{\delta}{\lambda_1}$$

where λ_1 is the wavelength in air. In the frequency range of visible light, for silver $n \cong 0.05$, $k \cong 3$. For these values, Eq. (18-60) is not valid, but Eq. (18-59) gives $R_n \cong 0.98$. More typical metallic values are perhaps those for nickel, $n \cong 2$, $k \cong 3$; these give $R_n \cong 0.56$. Such metals still look quite shiny, since the eye approximates a logarithmic detector. The corresponding values for silver and nickel are plotted as a function of the incident angle in Fig. 18-5. There is no Brewster's angle where R_p vanishes, but still R_p has a

FIGURE 18-5

Reflectance for s - and p -polarization at an air-metal interface. Representative values for visible light are $n \cong 0.05$, $k \cong 3$ (silver) and $n \cong 2$, $k \cong 3$ (nickel).



minimum and is always less than R_s . Some polarization by reflection still occurs. Equations (18-53) show that \hat{E}'_{1s} and \hat{E}'_{1p} have a relative phase difference $\alpha_s - \alpha_p$. Thus even if the incident wave is linearly polarized, the reflected wave may become elliptically polarized at oblique incidence.

The transmitted wave is important in problems such as those considered in the next section, even if it cannot be observed directly. Its amplitude and phase are given by \hat{t}_{12s} and \hat{t}_{12p} , and its propagation vector by the $\hat{\mathbf{k}}_2$ that satisfies Eqs. (18-48) to (18-52). The latter will define the planes of constant phase and the phase velocity, as well as the planes of constant amplitude and the attenuation constant. These results are obtained by comparing the two equivalent expressions for $\hat{\mathbf{k}}_2$:

$$\hat{\mathbf{k}} = \boldsymbol{\kappa}_r + i\boldsymbol{\kappa}_i \quad (18-62)$$

$$\hat{\mathbf{k}} = \hat{\kappa} \sin \hat{\theta} \hat{\mathbf{i}} + \hat{\kappa} \cos \hat{\theta} \hat{\mathbf{k}} \quad (18-63)$$

(In this discussion we will drop the subscript 2.) The second form is justified by the restriction Eq. (18-49). Eq. (18-48) requires that $\hat{\mathbf{k}} \times \mathbf{n}$ is real, and using Eq. (18-62) in Eq. (18-48) gives

$$\boldsymbol{\kappa}_r \times \mathbf{n} = \boldsymbol{\kappa}_i \times \mathbf{n} \quad (18-64)$$

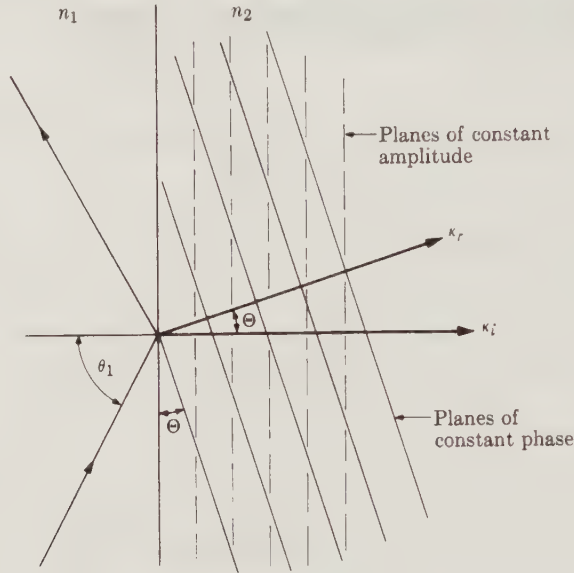
$$\boldsymbol{\kappa}_i \times \mathbf{n} = 0 \quad (18-65)$$

Equation (18-65) shows that $\boldsymbol{\kappa}_i$ is parallel to $\mathbf{n} = \mathbf{k}$, and Eq. (18-64) is

$$\kappa_r \sin \Theta = \kappa_i \sin \theta_1 \quad (18-66)$$

FIGURE 18-6

Refraction into a conductor. The planes of constant phase are perpendicular to the propagation direction and make an angle Θ with the boundary, but the planes of constant amplitude are parallel to the boundary.



where Θ is the *real* angle between $\mathbf{\kappa}_r$ and \mathbf{n} . (See Fig. 18-6.) The angle Θ could be called the real angle of refraction, in that it is the angle between the planes of constant phase and the boundary surface. The planes of constant amplitude, on the other hand, are parallel to the boundary surface, so that the wave is most rapidly attenuated directly into the conductor. We can rewrite Eq. (18-62)

$$\begin{aligned}\hat{\mathbf{k}} &= \kappa_r \sin \Theta \mathbf{i} + \kappa_r \cos \Theta \mathbf{k} + i\kappa_i \mathbf{k} \\ &= \kappa_1 \sin \theta_1 \mathbf{i} + (\kappa_r \cos \Theta + i\kappa_i) \mathbf{k}\end{aligned}$$

Comparing the components of this with those of Eq. (18-63), we see that

$$\kappa_1 \sin \theta_1 = \hat{\kappa} \sin \hat{\theta} \quad (18-67)$$

$$\kappa_r \cos \Theta + i\kappa_i = \hat{\kappa} \cos \hat{\theta} \quad (18-68)$$

The first is simply Snell's law again, but the second, together with Eq. (18-66), gives the relation of κ_r , κ_i , and Θ to n , k , and θ_1 , which we seek. Let us make the abbreviation

$$\hat{\kappa} \cos \hat{\theta} = \frac{\omega}{c} (p + iq)$$

so that

$$\hat{n} \cos \hat{\theta} = p + iq \quad (18-69)$$

Then from Eq. (18-68) $\kappa_r \cos \Theta = (\omega/c)p$, and with Eq. (18-66)

$$\kappa_r = \frac{\omega}{c} \sqrt{p^2 + n_1^2 \sin^2 \theta_1} \quad (18-70)$$

$$\kappa_i = \frac{\omega}{c} q \quad (18-71)$$

It remains only to find the values of p and q , which follow immediately by squaring Eq. (18-69):

$$\begin{aligned} p^2 - q^2 + 2ipq &= \hat{n}^2(1 - \sin^2 \hat{\theta}) = (n + ik)^2 - n_1^2 \sin^2 \theta_1 \\ &= n^2 - k^2 - n_1^2 \sin^2 \theta_1 + 2ink \end{aligned}$$

where we have used Snell's law $\hat{n} \sin \hat{\theta} = n_1 \sin \theta_1$ again. In terms of the dielectric constant $\hat{K} = \hat{n}^2$,

$$p^2 - q^2 + 2ipq = K_r - K_1 \sin^2 \theta_1 + iK_i$$

Equating real and imaginary parts, we get

$$K_r - K_1 \sin^2 \theta_1 = p^2 - q^2$$

$$K_i = 2pq$$

equations that are almost the same as Eq. (17-60) for n and k . The solutions

$$\begin{aligned} p &= \sqrt{\frac{1}{2}[(K_r - K_1 \sin^2 \theta_1) + \sqrt{(K_r - K_1 \sin^2 \theta_1)^2 + K_i^2}]} \\ q &= \sqrt{\frac{1}{2}[-(K_r - K_1 \sin^2 \theta_1) + \sqrt{(K_r - K_1 \sin^2 \theta_1)^2 + K_i^2}]} \end{aligned} \quad (18-72)$$

are the same as Eq. (17-61) for n and k except that K_r is replaced by $(K_r - K_1 \sin^2 \theta_1)$. Thus p and q play the role of generalized versions of n and k that depend on the angle of incidence θ_1 . When $\theta_1 = 0$, $p = n$ and $q = k$. The attenuation length is given directly by q according to Eq. (18-71). Equation (18-70) can be written

$$\kappa_r = N \frac{\omega}{c} \quad (18-73)$$

which defines a real refractive index $N(\theta_1)$,

$$N = \sqrt{p^2 + n_1^2 \sin^2 \theta_1} \quad (18-74)$$

and gives the phase velocity as c/N . From Eq. (18-66), N satisfies a real version of Snell's law:

$$N \sin \Theta = n_1 \sin \theta_1 \quad (18-75)$$

$$N \cos \Theta = p \quad (18-76)$$

Either of these can be used to find the real angle Θ . Although N is not the real part of \hat{n} , nor Θ the real part of $\hat{\theta}$, $N \cos \Theta$ is the real part of $\hat{n} \cos \hat{\theta}$. With this result, we have a complete solution for the complex case. The

results are a fairly simple generalization of the real case, and the application to any particular problem is straightforward. Equations (18-72) are so complicated, however, that not much more can be said in general about propagation into a conducting medium at oblique incidence, except in extreme cases. For example, when K_i is very large, the case that leads to the Hagen-Rubens formula,

$$p \cong n \cong q \cong k \gg 1$$

From Eq. (18-74),

$$N \gg 1$$

and from Eq. (18-75),

$$\Theta \cong 0$$

The propagation direction is almost directly into the medium for any angle of incidence, although the attenuation is very strong; the velocity and wavelength are much reduced. The skin depth that was defined for normal incidence holds approximately for any angle of incidence.

Returning to the case of total internal reflection, we can apply the results to complete the solution of that problem also. Here $n_2 = \sqrt{K_{2r}}$, $K_{2i} = 0$; but $\cos \hat{\theta}_2$ is imaginary when $\theta_1 > \theta_c$. (We now restore the subscript 2.)

$$\begin{aligned} \cos \hat{\theta}_2 &= \sqrt{1 - \sin^2 \theta_2} = \sqrt{1 - (n_1/n_2)^2 \sin^2 \theta_1} \\ &= i\sqrt{(\sin \theta_1/\sin \theta_c)^2 - 1} \end{aligned} \quad (18-77)$$

since

$$\sin \theta_c = n_2/n_1$$

Combining Eqs. (18-77) and (18-69), we find

$$n_2 \cos \hat{\theta}_2 = in_2\sqrt{(\sin \theta_1/\sin \theta_c)^2 - 1} = p + iq$$

so that

$$p = 0, \quad q = n_2\sqrt{(\sin \theta_1/\sin \theta_c)^2 - 1} \quad (18-78)$$

(The same result is given by Eq. (18-72).) With Eq. (18-78), the Fresnel coefficient for s -polarized reflection Eq. (18-28) becomes

$$\hat{r}_{12s} = \frac{n_1 \cos \theta_1 - iq}{n_1 \cos \theta_1 + iq}$$

The numerator is the complex conjugate of the denominator for any $\theta_1 > \theta_c$, so \hat{r}_{12s} has the form

$$\hat{r}_{12s} = \frac{z^*}{z}, \quad \hat{r}_{12s}^* = \frac{z}{z^*}$$

Therefore,

$$R_s = |\hat{r}_{12s}|^2 = \hat{r}_{12s} \hat{r}_{12s}^* = 1$$

Similarly, it is clear from Eq. (18-33) that $R_p = 1$ for all $\theta_1 > \theta_c$. We did not calculate T in the complex case, but energy conservation requires $T = 0$ when $R = 1$. On the other hand, the Fresnel transmission coefficients \hat{t}_{12} are not zero; there are nonvanishing E - and B -fields in medium 2. This seeming paradox is most easily resolved by finding κ_2 . With $p = 0$ in Eq. (18-74),

$$N = n_1 \sin \theta_1 = n_2 (\sin \theta_1 / \sin \theta_c) \quad (18-79)$$

The real refractive index N of medium 2 varies between n_2 and n_1 as θ_1 increases from θ_c to $\pi/2$,

$$n_2 \leq N \leq n_1 \quad (18-80)$$

From Eq. (18-76) with $p = 0$, $N \neq 0$,

$$\cos \Theta = 0 \quad (18-81)$$

for any angle of incidence $\theta_1 \geq \theta_c$. That is, κ_{2r} is always parallel to the boundary, and as a consequence there is no energy flow perpendicular to the boundary in medium 2. We already know that κ_{2i} is perpendicular to the boundary. Thus, in Fig. 18-6 the planes of constant phase are perpendicular to the boundary plane and the planes of constant amplitude are parallel to the boundary. These waves are known as *evanescent waves* and can be observed in appropriate experiments.

The evanescent wave is attenuated in the direction perpendicular to the boundary with an attenuation length $\delta = 1/\kappa_{2i}$. Using Eqs. (18-71) and (18-78), we can write the attenuation length:

$$\delta = \frac{c}{\omega q} = \frac{c}{n_2 \omega} \frac{1}{\sqrt{(\sin \theta_1 / \sin \theta_c)^2 - 1}} \quad (18-82)$$

Replacing n_2 with N from Eq. (18-79) and introducing the wavelength $\lambda_2/2\pi = c/N\omega$, we get

$$\delta = \frac{\lambda_2}{2\pi \sqrt{1 - (\sin \theta_c / \sin \theta_1)^2}}$$

As long as θ_1 is greater than the critical angle θ_c , δ is a real number, and the wave decays exponentially with depth into medium 2. As the angle of incidence decreases, the attenuation length increases until it becomes infinite at the critical angle. Thus, at this angle, ordinary unattenuated propagation becomes possible and instead of the evanescent wave the usual refracted wave propagates into medium 2. On the other hand, as θ_1 increases and approaches $\pi/2$, δ approaches $\lambda_2/2\pi \cos \theta_c$, which for glass-air is $\lambda_2/4.68$. Thus the behavior for $\theta_1 > \theta_c$ is a reasonable extension

of the behavior for $\theta_1 \leq \theta_c$: As θ_1 increases up to θ_c , R increases and θ_2 increases; at θ_c , $R = 1$ and $\theta_2 = \pi/2$. As θ_1 increases beyond θ_c , R remains 1 and the real angle of refraction Θ remains $\pi/2$, but the infinite penetration of medium 2 is gradually reduced to a fraction of a wavelength. At the same time the phase velocity and the wavelength in medium 2 decrease from that characteristic of medium 2 to that of medium 1. Two other interesting features are that a linearly polarized incident wave becomes elliptically polarized on reflection, because of the complex (but different) \hat{r}_{12s} and \hat{r}_{12p} , and that the wave in medium 2 is not transverse— \hat{E}_{2p} has a longitudinal component.

18-5 REFLECTION AND TRANSMISSION BY A THIN LAYER: INTERFERENCE

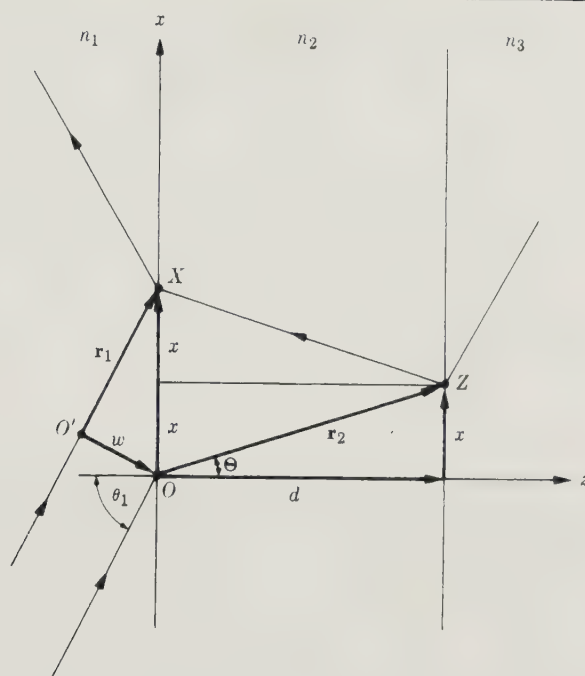
As a more complicated and more realistic boundary-value problem, we next consider *two* plane parallel infinite surfaces of discontinuity. This geometry represents a slab of material bounded on either side by semi-infinite media, which may have different properties from each other. To the left of the plane $z = 0$, we assume medium 1; to the right of the plane $z = d$, medium 3; and in between, medium 2. A straightforward application of the boundary conditions at each of the two planes of discontinuity, patterned after the calculation of Section 18-2, leads to results for the E - and B -fields in each of the three regions. (See Problem 18-11.) An alternative approach, which leads to the same answer, is based on the results already obtained in Section 18-2, and this approach is more informative in certain respects. The idea is to consider an incident wave in medium 1, which is partly reflected and partly transmitted at the first interface. The transmitted wave is partly reflected and partly transmitted at the second interface, and this reflected wave arrives back at the first interface, where it is partly reflected and partly transmitted, and so on. Since the Fresnel coefficients we already derived give the fractions reflected and transmitted at each interface, we have only to add up all the different contributions to the net wave reflected back into medium 1 and the net wave transmitted into medium 3. Although this procedure sounds infinite, it is in fact fairly simple.

The only new problem encountered before the waves are added up is that the different amplitudes must be added with their proper phase differences.* Each time the wave makes another pass through the layer, the phase shifts because of the change in $\mathbf{k}_2 \cdot \mathbf{r}$ in the exponent. The situation is

* We assume that the layer is actually thin enough and smooth enough that the coherent phase differences among all the multiple waves are meaningful. One of the advantages of the present approach is that if this assumption is not valid and the phase differences are more or less random, the summation procedure to be developed is still applicable, provided it is applied to the wave intensities instead of the amplitudes.

FIGURE 18-7

The ray that enters medium 2 at O is reflected from the back surface and reemerges to combine with the ray that is reflected at X .



shown in Fig. 18-7. Two incident rays perpendicular to the plane wave front in medium 1 strike the front surface of medium 2. One of them is partly reflected at X ; the other is partly refracted at O , partly reflected off the back surface at Z , and partly refracted at X to reemerge into medium 1 and combine with the first ray. Since the phase is the same at the two points on the wave front O and O' , we must calculate the phase difference between the two paths $O'X$ and OZX . This calculation is no more difficult when medium 2 is a conductor, and so we let the angle of refraction be Θ , the real angle found in the last section; for a nonconductor $\Theta = \theta_2$. The phase difference is

$$\hat{\beta} = 2\hat{\mathbf{k}}_2 \cdot \mathbf{r}_2 - \mathbf{\kappa}_1 \cdot \mathbf{r}_1$$

We resolve \mathbf{r}_2 into components

$$\mathbf{r}_2 = x\mathbf{i} + d\mathbf{k}$$

and \mathbf{r}_1 into

$$\mathbf{r}_1 = 2x\mathbf{i} - w\mathbf{p}_1$$

where

$$\mathbf{p}_1 = \mathbf{s} \times \mathbf{u}_1 = \mathbf{j} \times \mathbf{u}_1 \mathbf{u}_1$$

is perpendicular to $\boldsymbol{\kappa}_1 = \kappa_1 \mathbf{u}_1$. Then,

$$\hat{\beta} = 2x(\hat{\boldsymbol{\kappa}}_2 \cdot \mathbf{i} - \boldsymbol{\kappa}_1 \cdot \mathbf{i}) + 2d\hat{\boldsymbol{\kappa}}_2 \cdot \mathbf{k}$$

Now

$$\hat{\boldsymbol{\kappa}}_2 \cdot \mathbf{i} - \boldsymbol{\kappa}_1 \cdot \mathbf{i} = \hat{\kappa}_2 \sin \hat{\theta}_2 - \kappa_1 \sin \theta_1 = 0$$

according to Snell's law, and $\hat{\boldsymbol{\kappa}}_2 \cdot \mathbf{k} = \hat{\kappa}_2 \cos \hat{\theta}_2$, so that

$$\hat{\beta} = 2d\hat{\kappa}_2 \cos \hat{\theta}_2 = 2d \frac{\omega}{c} \hat{n}_2 \cos \hat{\theta}_2 \quad (18-83)$$

From Eq. (18-69),

$$\hat{\beta} = 2d \frac{\omega}{c} (p + iq) \quad (18-84)$$

In a nonconducting medium, $p = n_2 \cos \theta_2$ and $q = 0$ for all angles of incidence. In a conducting medium at normal incidence, $p = n$ and $q = k$. The real part of $\hat{\beta}$ gives the real phase shift, and the imaginary part of $\hat{\beta}$ gives the attenuation due to two traversals of the slab.

To add up all the contributions to the net amplitude reflection coefficient \hat{r} , we use the Fresnel coefficients for each boundary together with the phase shift $\hat{\beta}$. The Fresnel coefficients are different for s - and p -polarization, but we will drop the subscripts s and p for the present, remembering that the two polarizations have to be treated separately. From Fig. 18-8, we see that

$$\begin{aligned} \hat{r} &= \hat{r}_{12} + \hat{t}_{12}\hat{r}_{23}\hat{t}_{21}e^{i\hat{\beta}} + \hat{t}_{12}\hat{r}_{23}\hat{r}_{21}\hat{r}_{23}\hat{t}_{21}e^{2i\hat{\beta}} + \cdots \\ &= \hat{r}_{12} + \hat{t}_{12}\hat{r}_{23}\hat{t}_{21}e^{i\hat{\beta}}[1 + \hat{r}_{21}\hat{r}_{23}e^{i\hat{\beta}} + (\hat{r}_{21}\hat{r}_{23}e^{i\hat{\beta}})^2 + \cdots] \end{aligned}$$

Since

$$\begin{aligned} 1 + z + z^2 + \cdots &= \frac{1}{1 - z} \\ \hat{r} &= \hat{r}_{12} + \frac{\hat{t}_{12}\hat{t}_{21}\hat{r}_{23}e^{i\hat{\beta}}}{1 - \hat{r}_{21}\hat{r}_{23}e^{i\hat{\beta}}} = \frac{\hat{r}_{12} + \hat{r}_{23}(\hat{t}_{12}\hat{t}_{21} - \hat{r}_{12}\hat{r}_{21})e^{i\hat{\beta}}}{1 - \hat{r}_{21}\hat{r}_{23}e^{i\hat{\beta}}} \end{aligned}$$

Using the identities Eqs. (18-55) and (18-56), we get

$$\hat{r} = \frac{\hat{r}_{12} + \hat{r}_{23}e^{i\hat{\beta}}}{1 + \hat{r}_{12}\hat{r}_{23}e^{i\hat{\beta}}} \quad (18-85)$$

A similar calculation gives the net amplitude transmitted into medium 3:

$$\hat{t} = \frac{\hat{t}_{12}\hat{t}_{23}e^{(1/2)i\hat{\beta}}}{1 + \hat{r}_{12}\hat{r}_{23}e^{i\hat{\beta}}} \quad (18-86)$$

Note that the numerators give the effect of the front and back surfaces each acting alone, and the denominator accounts for all of the actual multiple

nevertheless important, because measurement of R and T for thin films of metal provides one method to determine experimentally the optical constants. Computer calculations are necessary to solve the equations for n and k in terms of experimental values of R and T . The transmittance T is proportional to $\hat{t}\hat{t}^*$, and $\hat{t}\hat{t}^*$ is proportional to

$$e^{(1/2)i\hat{\beta}}e^{-(1/2)i\hat{\beta}^*} = e^{(1/2)i(\hat{\beta}-\hat{\beta}^*)} = e^{-2d(\omega/c)q}$$

At normal incidence, $q = k$, so that T contains the factor

$$e^{-2d/\delta}$$

where $\delta = c/k\omega$ is the skin depth. If medium 1 is air, $\omega/c = 2\pi/\lambda_1$, so that

$$e^{-2d/\delta} = e^{-4\pi kd/\lambda_1}$$

For metals ($k \approx 2$) and visible light ($\lambda_1 \approx 5000 \text{ \AA}$), d must be less than about 1000 \AA for appreciable transmission of light. When this exponential factor is small, the denominator in Eqs. (18-85) and (18-86) is approximately 1.

For nonconductors, $q = 0$ (except for total internal reflection), so there is no attenuation due to this factor, but the equations still predict some interesting effects. For β and the Fresnel coefficients all real,

$$R = \frac{r_{12}^2 + r_{23}^2 + 2r_{12}r_{23} \cos \beta}{1 + r_{12}^2 r_{23}^2 + 2r_{12}r_{23} \cos \beta} \quad (18-89)$$

At normal incidence,

$$r_{12} = \frac{n_1 - n_2}{n_1 + n_2}, \quad r_{23} = \frac{n_2 - n_3}{n_2 + n_3}, \quad \beta = 2d \frac{\omega}{c} n_2$$

Suppose medium 1 is air with $n_1 = 1$, 3 is glass with $n_3 = 1.5$, and 2 is a thin layer of material with $n_2 = 1.3$. Then,

$$R = \frac{0.0221 + 0.0186 \cos \beta}{1.0001 + 0.0186 \cos \beta}$$

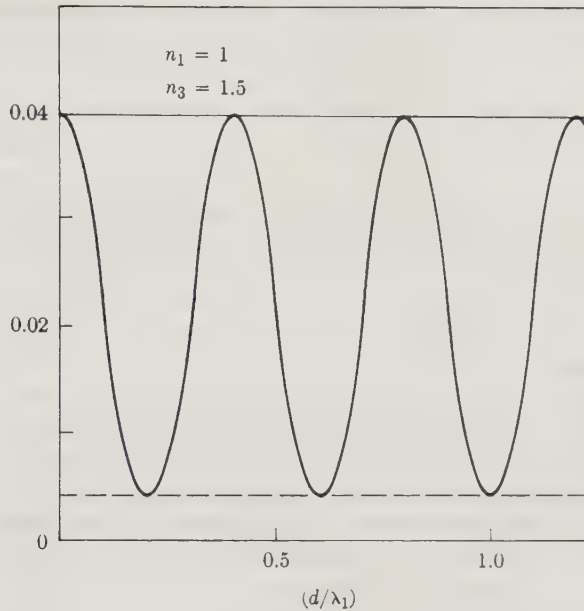
$$\beta = 4\pi n_2(d/\lambda_1) = 16.3(d/\lambda_1)$$

Thus R varies between 0.04 and 0.004, with the maxima* at integral multiples of $\beta = 2\pi$, or d/λ_1 integral multiples of 0.39, as shown in Fig. 18-9. The most interesting feature of this result is that R can be less than $r_{12}^2 = 0.017$, the reflectance from the front face alone; this kind of effect can happen only because of destructive interference. The variation of R is between $r_{13}^2 = 0.040$, the value without any covering layer, to something less than $r_{23}^2 = 0.005$, the value from the back face alone. In fact the

* Note that the positions of the extrema of Eq. (18-89) are just the same as in an elementary calculation that takes only two reflections into account.

FIGURE 18-9

Interference effect on the reflectance from an air-glass interface coated with a thickness d of material with $n_2 = 1.3$



minimum value of R can be made zero if a material can be found such that $n_2 = \sqrt{n_1 n_3}$. (See Problem 18-10.) The effect is exploited to produce nonreflecting lenses. Camera lenses are often coated to have nearly zero reflectance near the middle of the visible spectrum; the condition for a minimum does not hold, however, at the red and blue ends of the spectrum, and so the lens has a purple cast when seen by its reflected light. The colors also depend on the angle of viewing, since for oblique incidence $\beta = 4\pi n_2 \cos \theta_2 (d/\lambda_1)$. If n_2 is greater than n_1 and n_3 , R varies between a minimum of r_{13}^2 and a maximum that is greater than r_{12}^2 and r_{23}^2 . The wavelength at which a minimum or maximum occurs depends on the thickness d of the layer; if the thickness varies from point to point on the film, so will the predominantly reflected wavelengths. Such variation accounts for the colors seen in soap bubbles and oil films floating on water.

Another interesting effect occurs in nonconducting layers if n_2 is less than n_1 . This case is what leads to total internal reflection when there is only one boundary. One might first guess that if all the energy is reflected the presence of a second boundary could make no difference. This guess is wrong, however; the field always penetrates medium 2, a mean depth of δ . A second boundary spoils the perfect reflectance and there is a transmitted wave with $T = 1 - R$. This effect is called *frustrated total reflection*. Although n_1 , n_2 , and n_3 are real, $\hat{\theta}_2$ is complex when $\theta_1 > \theta_c$, making the

Fresnel coefficients complex. We found for *s*-polarization

$$\hat{r}_{12} = \frac{n_1 \cos \theta_1 - iq}{n_1 \cos \theta_1 + iq} \quad (18-90)$$

where

$$q = n_2 \sqrt{(\sin \theta_1 / \sin \theta_c)^2 - 1} = \sqrt{(n_1 \sin \theta_1)^2 - n_2^2} \quad (18-91)$$

from Eq. (18-78) (for either polarization), with $\sin \theta_c = n_2/n_1$. Writing \hat{r}_{12} in polar form, $\hat{r}_{12} = |\hat{r}_{12}|e^{i\alpha}$, we found that $|\hat{r}_{12}| = 1$. Thus,

$$\hat{r}_{12} = e^{i\alpha}$$

where

$$\tan \alpha = \frac{\text{Im}(\hat{r}_{12})}{\text{Re}(\hat{r}_{12})} = \frac{-2n_1 \cos \theta_1 q}{(n_1 \cos \theta_1)^2 - q^2} \quad (18-92)$$

By the same argument, $|\hat{r}_{23}| = 1$. For simplicity let us assume that medium 3 is the same as medium 1 (for example, a second glass prism with an air layer between). Then $\hat{r}_{23} = \hat{r}_{21}$, and since $\hat{r}_{21} = -\hat{r}_{12}$ identically,

$$\hat{r}_{23} = -\hat{r}_{12} = -e^{i\alpha} \quad (18-93)$$

From Eq. (18-84),

$$\hat{\beta} = 2d \frac{\omega}{c} (p + iq)$$

and in Eq. (18-78) we found that $p = 0$, so that $\hat{\beta} = i2dq(\omega/c)$ is purely imaginary. We therefore write

$$e^{i\hat{\beta}} = e^{-\gamma}$$

$$\gamma = 2dq(\omega/c) = 2 \frac{d}{\delta} \quad (18-94)$$

since $q(\omega/c) = 1/\delta$ from Eq. (18-82). Now substituting these into Eq. (18-85), we get

$$\hat{r} = \frac{e^{i\alpha}(1 - e^{-\gamma})}{1 - e^{2i\alpha}e^{-\gamma}}$$

Finally,

$$R = \hat{r}\hat{r}^* = \frac{(1 - e^{-\gamma})^2}{1 + e^{-2\gamma} - 2e^{-\gamma} \cos 2\alpha}$$

$$T = 1 - R = \frac{2(1 - \cos 2\alpha)e^{-\gamma}}{1 + e^{-2\gamma} - 2e^{-\gamma} \cos 2\alpha} \quad (18-95)$$

Note that for given n_1 and n_2 , α depends only on the angle of incidence θ_1 ; it varies from 0 to π as θ_1 increases from θ_c to $\pi/2$. The exponent γ depends

on the thickness d (as well as on θ , through δ). If γ is not too small,

$$T \cong 2(1 - \cos 2\alpha)e^{-2d/\delta} \quad (18-96)$$

Note also that δ is proportional to λ , so frustrated total reflection can be observed with microwaves on a larger scale.

18-6 PROPAGATION BETWEEN PARALLEL CONDUCTING PLATES

Guided waves are another problem that can be treated by considering the interference between an incident and a reflected wave, or alternatively by starting with a new boundary-value problem that involves simultaneously satisfying the conditions at multiple boundaries. We again begin with the first approach. Now we are interested in the wave propagating in a dielectric medium, say air, which is bounded by conducting surfaces. Waveguides for microwaves are an application of this problem. As a simplification we idealize the conductivity of the metal to be infinite. Infinite g means infinite K_i for the metal, which means infinite \hat{n}_z in Eqs. (18-28) and (18-33). Thus $\hat{r}_{12s} = -1$, $\hat{r}_{12p} = +1$, for reflection from a perfectly conducting plane at any angle of incidence. Actually we found $R_n = 0.9996$ for silver at 3 cm wavelength, so the approximation can be expected to be useful. In addition, the dielectric medium is assumed to be vacuum.

As a preliminary to the study of waveguides, we now consider the propagation of electromagnetic waves in the region between two parallel, perfectly conducting plates. The region in which wave propagation is to be treated is that shown in Fig. 18-10. Since the x and z directions are physically indistinguishable, no generality is lost by considering only waves with wave vectors in the yz -plane—in particular, those making an angle θ with the y -axis in the plane of incidence. Such waves will impinge on the perfectly conducting surface at $y = b$ and will be reflected as waves whose

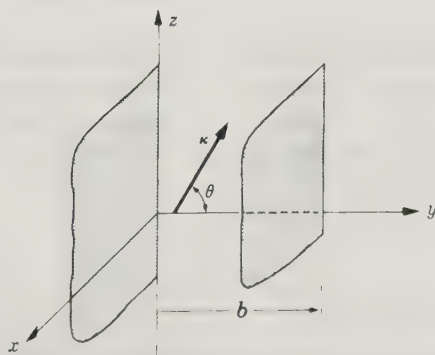


FIGURE 18-10
Wave propagation between two parallel, perfectly conducting planes

propagation vectors make the angle θ with the minus y -axis. When these waves are reflected a second time, by the surface at $y = 0$, they become waves of the first type again. Thus, the propagation between two parallel conducting planes can be described in terms of the exponential factors

$$e^{i[\kappa(y \cos \theta + z \sin \theta) - \omega t]}$$

and

$$e^{i[\kappa(-y \cos \theta + z \sin \theta) - \omega t]} \quad (18-97)$$

For such waves there are two possible polarizations, which may be described by saying for the s -polarized one that \mathbf{E} is parallel to the x -axis, and for the p -polarized one that \mathbf{H} is parallel to the x -axis. These are known respectively as TE, transverse electric, and TM, transverse magnetic, waves in waveguide terminology. Only TE waves will be considered here. The treatment of TM waves will be left to a problem at the end of the chapter.

The electric field in the region between the two conducting planes in the TE case is given by

$$\mathbf{E} = \mathbf{i}\{E_1 e^{i[\kappa(y \cos \theta + z \sin \theta) - \omega t]} + E'_1 e^{i[\kappa(-y \cos \theta + z \sin \theta) - \omega t]}\} \quad (18-98)$$

This electric field must vanish at $y = 0$, since E_t vanishes at the boundary of a perfect conductor. This condition is clearly satisfied for all z and all t if $E_1 = -E'_1 = E$, as given by $r_s = -1$. Then \mathbf{E} is given by

$$\mathbf{E} = \mathbf{i}E(e^{i\kappa y \cos \theta} - e^{-i\kappa y \cos \theta})e^{i(\kappa z \sin \theta - \omega t)} \quad (18-99)$$

In addition, \mathbf{E} must vanish at $y = b$ for all z and all t . This requirement imposes the condition

$$\kappa b \cos \theta = n\pi \quad (18-100)$$

where n is an integer. Thus for a given frequency ω , $\kappa = \omega/c$ and the angle that the waves make with the y -axis is fixed by Eq. (18-100). With this angle fixed, the apparent velocity in the z -direction is $v_p = c/\sin \theta$, which is always greater than the velocity of light in free space. This apparent contradiction with the special theory of relativity will be discussed in more detail later.

It is convenient to express the variation of the electric field in the y - and z -directions in terms of wavelengths. These wavelengths are

$$\lambda_g = \frac{2\pi}{\kappa \sin \theta} = \frac{\lambda_0}{\sin \theta} \quad \left(\lambda_0 = \frac{2\pi}{\kappa} = \frac{2\pi c}{\omega} \right) \quad (18-101)$$

for the z -direction, and

$$\lambda_c = \frac{2\pi}{\kappa \cos \theta} = \frac{\lambda_0}{\cos \theta} \quad (18-102)$$

for the y -direction. In terms of these wavelengths, the electric field, Eq.

(18-98), is*

$$\mathbf{E} = \mathbf{i}E_0 \sin \frac{2\pi y}{\lambda_c} e^{i[(2\pi z/\lambda_g) - \omega t]} \quad (18-103)$$

while Eq. (18-100) takes the form

$$\frac{b}{\lambda_c} = \frac{n}{2} \quad (18-104)$$

From Eqs. (18-101) and (18-102), it follows immediately that

$$\frac{1}{\lambda_g^2} + \frac{1}{\lambda_c^2} = \frac{1}{\lambda_0^2} \quad (18-105)$$

If the value $\lambda_c = 2b$, corresponding to $n = 1$ in Eq. (18-104), is considered, then as λ_0 increases, that is, as ω decreases, a point is reached where $1/\lambda_g^2$ must be negative in order to satisfy Eq. (18-105). In this case the coefficient of z in Eq. (18-103) is imaginary, and the exponential, instead of oscillating in z , becomes a decreasing exponential. To say this another way: If $\lambda_0 > 2b$ the electromagnetic wave will be exponentially damped in z , instead of propagating. If n is taken as 2, then $\lambda_c = 2b/2 = b$ and the longest wavelength propagated is b . The reason for the subscript c is now clear; it means "cutoff." The cutoff wavelength is the longest wavelength that can be propagated for a given mode (n value).

The velocity v_p , which was found earlier, always exceeds the velocity of light and, in fact, becomes infinite when the wavelength in free space equals λ_c —that is, when $\theta = 0$. This velocity is the phase velocity, by which is meant the velocity of a point of constant phase on the wave. Without dwelling on the relativistic aspects of the question, there seems to be a contradiction of the postulate that no signal can be propagated with a velocity greater than the velocity of light. The resolution of this apparent difficulty is that energy is propagated down the guide with a smaller velocity than the velocity of light, namely, with the so-called group velocity. Signals are transmitted with the group velocity; they are not transmitted with the phase velocity.

To determine the velocity of energy propagation, we shall calculate the energy density. This energy density times the *group velocity* gives the energy flux, or Poynting vector. Thus by dividing the Poynting vector by the energy density, the velocity of energy propagation can be obtained. This result is a generalization of Eq. (17-35).

The magnetic induction in the guide is readily obtained from

$$\nabla \times \mathbf{E} = - \frac{\partial \mathbf{B}}{\partial t} \quad (18-106)$$

* E_0 has been written for $2iE$.

By using Eq. (18-103) for \mathbf{E} and assuming that $\mathbf{B}(\mathbf{r}, t) = \mathbf{B}(\mathbf{r})e^{-i\omega t}$, we quickly find

$$\begin{aligned}\mathbf{B}(\mathbf{r}, t) &= \mathbf{j}E_0 \frac{2\pi}{\omega\lambda_g} \sin \frac{2\pi y}{\lambda_c} e^{i[(2\pi z/\lambda_g) - \omega t]} \\ &\quad + i\mathbf{k}E_0 \frac{2\pi}{\omega\lambda_c} \cos \frac{2\pi y}{\lambda_c} e^{i[(2\pi z/\lambda_g) - \omega t]}\end{aligned}\quad (18-107)$$

The energy density is

$$u = \frac{1}{2}(\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H}) \quad (18-108)$$

while the Poynting vector is

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \quad (18-109)$$

Complex notation has been used for \mathbf{E} and \mathbf{B} , with the tacit assumption that the real part of each expression is to be taken. In calculating u and \mathbf{S} then, the real parts should be taken and multiplied together. However, since the quantities to be used in calculating the group velocity are the time averages of Eqs. (18-108) and (18-109), Eq. (17-37) may be used to circumvent the taking of real parts.

The time average energy density is

$$\begin{aligned}\bar{u} &= \frac{1}{4} \operatorname{Re} [\mathbf{E}^* \cdot \mathbf{D} + \mathbf{B}^* \cdot \mathbf{H}] = \frac{1}{4} \operatorname{Re} \left[\epsilon_0 E_0^* E_0 \sin^2 \left(\frac{2\pi y}{\lambda_c} \right) \right. \\ &\quad \left. + \frac{1}{\mu_0} E_0^* E_0 \left(\frac{2\pi}{\omega\lambda_g} \right)^2 \sin^2 \left(\frac{2\pi y}{\lambda_c} \right) \right. \\ &\quad \left. + \frac{1}{\mu_0} E_0^* E_0 \left(\frac{2\pi}{\omega\lambda_c} \right)^2 \cos^2 \left(\frac{2\pi y}{\lambda_c} \right) \right]\end{aligned}\quad (18-110)$$

Integrating in the y direction, across the guide, effectively replaces each $\sin^2(2\pi y/\lambda_c)$ and $\cos^2(2\pi y/\lambda_c)$ by $b/2$. Thus

$$\int_0^b \bar{u} dy = \frac{1}{4} E_0^* E_0 \frac{b}{2} \left[\epsilon_0 + \frac{1}{\mu_0} \frac{4\pi^2}{\omega^2} \left(\frac{1}{\lambda_g^2} + \frac{1}{\lambda_c^2} \right) \right] = \frac{1}{4} E_0^* E_0 \epsilon_0 b \quad (18-111)$$

The time average of the z -component of the Poynting vector is

$$\begin{aligned}\bar{S}_z &= \frac{1}{2} \operatorname{Re} E_x^* H_y = \frac{1}{2} \operatorname{Re} \left[E_0^* \sin \left(\frac{2\pi y}{\lambda_c} \right) \frac{1}{\mu_0} E_0 \frac{2\pi}{\omega\lambda_g} \sin \left(\frac{2\pi y}{\lambda_c} \right) \right] \\ &= \frac{1}{2} E_0^* E_0 \frac{2\pi}{\mu_0 \omega \lambda_g} \sin^2 \left(\frac{2\pi y}{\lambda_c} \right)\end{aligned}\quad (18-112)$$

Integrating this expression from $y = 0$ to $y = b$ yields the total average

power (per unit length in the x -direction) traveling down the guide:

$$\int_0^b \bar{S}_z dy = \frac{1}{4} E_0^* E_0 \frac{2\pi}{\mu_0 \omega \lambda_g} b \quad (18-113)$$

The velocity of energy propagation is the quotient of Eq. (18-113) divided by (18-111). Thus,

$$v_g = \frac{2\pi}{\epsilon_0 \mu_0 \omega \lambda_g} = \frac{2\pi c^2}{\omega \lambda_g} = c \frac{\lambda_0}{\lambda_g} \quad (18-114)$$

From Eq. (18-101), we note that λ_g is greater than λ_0 , and hence v_g is less than c .

Our understanding of the difference between the group velocity v_g and the phase velocity v_p can be enhanced by noting that from Eq. (18-101) $\lambda_g = \lambda_0 / \sin \theta$. Using this result in Eq. (18-114), we find

$$v_g = c \sin \theta \quad (18-115)$$

and we have already seen that

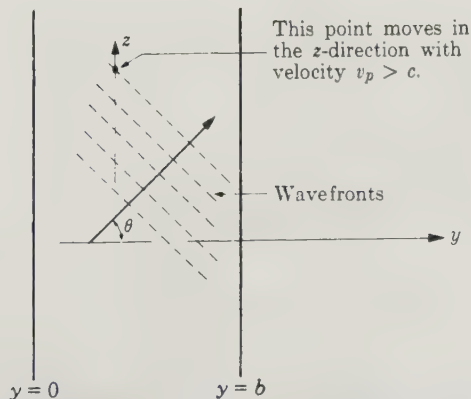
$$v_p = \frac{c}{\sin \theta} \quad (18-116)$$

It is readily apparent that

$$v_g v_p = c^2 \quad (18-117)$$

which is generally true for propagation in a waveguide. (Note that Eq. (18-117) does not necessarily apply to other kinds of wave propagation; in particular, it does not apply to plane waves in unbounded nondispersive media where the phase and group velocities are identical.) Recalling that θ is the angle between the direction of propagation of one of the component waves and the y -axis makes it a simple matter to draw Fig. 18-11, which

FIGURE 18-11
Detailed motion of the wavefronts during wave propagation between conducting planes



shows a section in the yz -plane of the region between the conducting planes. The intersection of a wavefront with the z -axis moves with the velocity $v_p = c/\sin \theta$; however, the component of c along the z -axis is $c \sin \theta = v_g$.

Many of the results obtained for the simple parallel plate waveguide persist for more complex cases. In particular, the common rectangular waveguide has very similar properties. In the next section, some general aspects of other waveguides will be considered, with particular reference to rectangular guides.

18-7 WAVEGUIDES

In Section 16-4, it was shown that \mathbf{E} and \mathbf{H} both satisfy the wave equation in free space; that is,

$$\nabla^2 \mathbf{E} - \epsilon_0 \mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0, \quad \nabla^2 \mathbf{H} - \epsilon_0 \mu_0 \frac{\partial^2 \mathbf{H}}{\partial t^2} = 0 \quad (18-118)$$

For monochromatic waves, that is, waves of the form $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r})e^{-i\omega t}$, these equations become

$$\nabla^2 \mathbf{E} + \frac{\omega^2}{c^2} \mathbf{E} = 0, \quad \nabla^2 \mathbf{H} + \frac{\omega^2}{c^2} \mathbf{H} = 0 \quad (18-119)$$

In addition to these wave equations, Maxwell's equations must be satisfied. For the transverse electric (TE) case propagating in the z direction, $E_z = 0$; furthermore, waves propagating in the z -direction have the remaining five field quantities proportional to $e^{i2\pi z/\lambda_g}$. Maxwell's curl equations in this case are

$$\nabla \times \mathbf{E} - i\mu_0 \omega \mathbf{H} = 0:$$

$$(a) \quad \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} - i\mu_0 \omega H_z = 0$$

$$(b) \quad E_x = + \frac{\mu_0 \omega \lambda_g}{2\pi} H_y \quad (18-120)$$

$$(c) \quad E_y = - \frac{\mu_0 \omega \lambda_g}{2\pi} H_x$$

$$\nabla \times \mathbf{H} + i\epsilon_0 \omega \mathbf{E} = 0;$$

$$(a) \quad \frac{\partial H_z}{\partial y} - \frac{2\pi i}{\lambda_g} H_y + i\epsilon_0 \omega E_x = 0$$

$$(b) \quad \frac{2\pi i}{\lambda_g} H_x - \frac{\partial H_z}{\partial x} + i\epsilon_0 \omega E_y = 0 \quad (18-121)$$

$$(c) \quad \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = 0$$

FIGURE 18-12
Wave propagation inside
a conducting cylinder



It is clear that (a) of Eq. (18-121) and (b) of Eq. (18-120) imply

$$\frac{\partial H_z}{\partial y} = \left(\frac{2\pi i}{\lambda_g} - i \frac{\epsilon_0 \mu_0 \omega^2 \lambda_g}{2\pi} \right) H_y \quad (18-122)$$

and therefore that H_y can be found if H_z is known. Similarly, from (c) of Eq. (18-120) and (b) of Eq. (18-121), H_x can be found from H_z . Finally, E_x and E_y are simply related to H_y and H_x by (b) and (c) of Eq. (18-120). Thus, if H_z is found, all the other field quantities may be found by differentiation. H_z itself must satisfy Eq. (18-118); therefore, taking cognizance of the $e^{i^2\pi z/\lambda_g}$ z -dependence, we write

$$\frac{\partial^2 H_z}{\partial x^2} + \frac{\partial^2 H_z}{\partial y^2} + \left(\frac{\omega^2}{c^2} - \frac{4\pi^2}{\lambda_g^2} \right) H_z = 0 \quad (18-123)$$

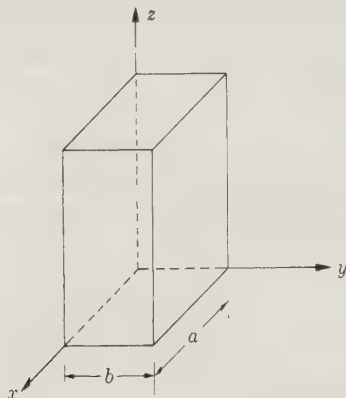
It remains only to determine the boundary conditions to be imposed on the solutions of Eq. (18-123).

If a general cylindrical guide with perfectly conducting walls, such as that shown in Fig. 18-12, is under consideration, then the appropriate boundary conditions are that the tangential component of \mathbf{E} and the normal component of \mathbf{B} should vanish on S . The tangential component of \mathbf{H} and the normal component of \mathbf{D} are arbitrary. Imposing these conditions gives rise to a relationship connecting λ_g , ω , and the dimensions of the guide, exactly as Eq. (18-105) does for the parallel plane case.

To better understand the procedure, consider the rectangular waveguide shown in Fig. 18-13. Equation (18-123) can be separated by the usual method of separation of variables. The general solution consists of a sum of terms of the form

$$\begin{aligned} H_z(x, y, z) = & (A \cos \kappa_x x \cos \kappa_y y + B \cos \kappa_x x \sin \kappa_y y \\ & + C \sin \kappa_x x \cos \kappa_y y + D \sin \kappa_x x \sin \kappa_y y) \\ & \times e^{2\pi i z / \lambda_g} \end{aligned} \quad (18-124)$$

FIGURE 18-13
A rectangular waveguide



with

$$-(\kappa_x^2 + \kappa_y^2) + [\omega^2/c^2 - (4\pi^2/\lambda_g^2)] = 0 \quad (18-125)$$

For this H_z , we obtain E_x :

$$E_x = -\frac{\mu_0\omega\lambda_g}{2\pi} \left(\frac{2\pi i}{\lambda_g} - i \frac{\epsilon_0\mu_0\omega^2\lambda_g}{2\pi} \right)^{-1} \frac{\partial H_z}{\partial y} \quad (18-126)$$

The partial differentiation changes every $\cos \kappa_y y$ to a $\sin \kappa_y y$, and conversely. However, since E_x must vanish at $y = 0$ and at $y = b$, only terms involving $\sin \kappa_y y$ may survive in E_x , and these terms must have $\kappa_y = n\pi/b$. Thus only $\cos \kappa_y y$ terms survive in Eq. (18-124). A similar argument shows that only $\cos \kappa_x x$ terms may survive, and that these must have $\kappa_x = m\pi/a$. The allowed solutions for H_z —that is, those that give vanishing tangential components of \mathbf{E} at the boundary—have the form

$$H_z = A \cos \frac{m\pi x}{a} \cos \frac{n\pi y}{b} e^{2\pi iz/\lambda_g} \quad (18-127)$$

Each possible pair of values of m and n is referred to as a mode. The notation TE_{mn} is used for modes of the form Eq. (18-127); TE means transverse electric, n and m count the number of half-waves in the narrower (n) and wider (m) dimensions.

Returning now to Eq. (18-125) and using $\kappa_x = m\pi/a$ and $\kappa_y = n\pi/b$, we obtain

$$\left(\frac{2\pi}{\lambda_g} \right)^2 = \left(\frac{2\pi}{\lambda_0} \right)^2 - \left(\frac{n\pi}{b} \right)^2 - \left(\frac{m\pi}{a} \right)^2 \quad (18-128)$$

which clearly indicates that for fixed λ_0 the guide wavelength, and consequently the guide velocity $v_g = c\lambda_0/\lambda_g$, depend on the mode. We also see that there are maximum wavelengths for the propagation of various

modes. Clearly, if λ_0 is sufficiently large $(2\pi/\lambda_0)^2$ will be smaller than $(n\pi/b)^2 + (m\pi/a)^2$. In this case, the right side of Eq. (18-128) becomes negative and consequently the value of λ_g is imaginary. This leads to attenuation rather than propagation.

Rectangular waveguides are extensively used for the transmission of microwave power. It is usual to choose a waveguide size such that only the TE_{10} mode at the desired frequency propagates in the guide. A common size of waveguide is 0.4 in. \times 0.9 in., inside dimensions. The maximum wavelength that will propagate in the TE_{10} mode is found by putting $m = 1$, $n = 0$, $a = 0.9$ in. = 2.28 cm and, $b = 0.4$ in. = 1.01 cm into Eq. (18-128). The result: $\lambda_{0,\max} = 4.57$ cm is obtained by putting $\lambda_g = \infty$; longer wavelengths will not propagate, but shorter wavelengths will. The mode with the next shorter cutoff wavelength is TE_{11} or TE_{20} , depending on the dimensions of the guide. If $b < a/\sqrt{3}$, the TE_{20} cutoff wavelength is greater than the TE_{11} . Calculation of the TE_{20} wavelength is very simple; it is just one-half the TE_{10} cutoff wavelength, or 2.28 cm. Imperfections in manufactured waveguides and high losses close to the TE_{10} cutoff wavelength make it necessary to restrict the TE_{10} band of commercial waveguides to the practical limits of 2.42 to 4.35 cm.

CAVITY RESONATORS

Another type of device closely related to waveguides and of considerable practical importance is the cavity resonator. Cavity resonators display the properties typical of resonant circuits in that they can store energy in oscillating electric and magnetic fields. Furthermore, practical cavity resonators dissipate a fraction of the stored energy in each cycle of oscillation. In this latter respect, however, cavity resonators are usually superior to conventional L - C circuits by a factor of about twenty—that is, the fraction of the stored energy dissipated per cycle in a cavity resonator is about 1/20 the fraction dissipated per cycle in an L - C circuit. An additional advantage is that cavity resonators of practical size have resonant frequencies which range upward from a few hundred megahertz—just the region where it is almost impossible to construct ordinary L - C circuits.

The simplest cavity resonator is a rectangular parallelepiped with perfectly conducting walls. For such a cavity, the appropriate boundary conditions are the vanishing of the tangential component of \mathbf{E} and the normal component of \mathbf{B} at the boundary. The tangential component of \mathbf{H} and the normal component of \mathbf{D} are arbitrary. The electric and magnetic fields must satisfy the wave equations (18-118). Thus, E_x must satisfy

$$\frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} + \frac{\omega^2}{c^2} E_x = 0 \quad (18-129)$$

If the cavity consists of the region bounded by the six planes $x = 0$, $x = a$;

$y = 0$, $y = b$; $z = 0$, $z = d$, then E_x must have the form

$$E_x = E_1 f_1(x) \sin \kappa_y y \sin \kappa_z z e^{-i\omega t} \quad (18-130)$$

with $\kappa_y = m\pi/b$ and $\kappa_z = n\pi/d$, in order that E_x vanish at $y = 0$, at $z = 0$, at $y = b$, and at $z = d$. For E_y and E_z the situation is similar, and the solutions take the forms

$$\begin{aligned} E_y &= E_2 \sin \kappa_x x f_2(y) \sin \kappa_z z e^{-i\omega t} \\ E_z &= E_3 \sin \kappa_x x \sin \kappa_y y f_3(z) e^{-i\omega t} \end{aligned} \quad (18-131)$$

with κ_y and κ_z as in Eq. (18-130), and $\kappa_x = l\pi/a$. If the divergence of \mathbf{E} is to vanish, then the equation

$$\begin{aligned} \left(E_1 \frac{df_1}{dx} \sin \kappa_y y \sin \kappa_z z + E_2 \sin \kappa_x x \frac{df_2}{dy} \sin \kappa_z z \right. \\ \left. + E_3 \sin \kappa_x x \sin \kappa_y y \frac{df_3}{dz} \right) e^{-i\omega t} = 0 \end{aligned} \quad (18-132)$$

must be satisfied. This requirement is satisfied if $f_1 = \cos \kappa_x x$, $f_2 = \cos \kappa_y y$, $f_3 = \cos \kappa_z z$, and

$$\kappa_x E_1 + \kappa_y E_2 + \kappa_z E_3 = 0 \quad (18-133)$$

The last equation is just the condition that $\boldsymbol{\kappa}$ be perpendicular to \mathbf{E} . Returning to the wave equation, it is apparent that the resonant frequencies of the cavity are given by

$$\kappa_x^2 + \kappa_y^2 + \kappa_z^2 - \frac{\omega^2}{c^2} = 0 \quad (18-134)$$

or

$$\frac{l^2}{a^2} + \frac{m^2}{b^2} + \frac{n^2}{d^2} - \frac{4f^2}{c^2} = 0 \quad (18-135)$$

A typical cavity constructed from a waveguide 0.4 in. \times 0.9 in. in size is characterized by $l = 1$, $m = 0$, $n = 2$ (a so-called TE₁₀₂ cavity). The resonant frequency of such a cavity is clearly determined by the z -dimension, d . Many other aspects of the rectangular cavity resonator problem can be treated in detail; some of these are left to the problems section.

Other forms of cavity resonators may be constructed. However, only the right circular cylinder and the rectangular parallelepiped are easily fabricated and amenable to an exact mathematical treatment. The treatment of the right circular cylinder involves functions that are more complicated than the sines and cosines, specifically the Bessel functions. Satisfying the boundary conditions requires finding the zeros of these functions in the same way the zeros of the sines entered the rectangular problem. Rather

than enter into the elaborate discussion that results, we refer the interested reader to the now-classic *Techniques of Microwave Measurements* by C. G. Montgomery (New York: McGraw-Hill, 1947), p. 297 ff., where a brief but very useful treatment of the cylindrical cavity resonator is given.

18-9 SUMMARY

Practical problems of wave propagation usually involve boundaries between different media, where the complex dielectric constant \hat{K} changes discontinuously. The conditions at a single plane boundary are expressed by Snell's law and the Fresnel coefficients, which depend on $\hat{n} = \sqrt{\hat{K}}$ and the angle of incidence. Problems involving multiple plane boundaries can be solved by superposition of the solutions for a single boundary, and show interference effects. Nonconducting media are a special case of conducting media in which the imaginary part of \hat{K} and \hat{n} vanishes.

- Boundary conditions on the amplitudes cannot be satisfied unless the wave frequency is the same on either side of the boundary and the phase matches everywhere on the boundary. Hence the *incident*, *reflected*, and *transmitted* wave propagation vectors are all **coplanar** with the normal to the boundary, and the **angle of reflection** equals the angle of incidence, $\theta'_1 = \theta_1$. The dispersion relation gives **Snell's law**

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

- Continuity of the tangential components of the E - and H -fields is expressed by the **Fresnel coefficients**, the ratios of the reflected and transmitted E -field amplitudes to the incident amplitude. These differ for s - and p -polarization (E -vector perpendicular and parallel to the plane of incidence, respectively.) For reflection

$$r_{12s} = \frac{n_1 \cos \theta_1 - n_2 \cos \theta_2}{n_1 \cos \theta_1 + n_2 \cos \theta_2}, \quad r_{12p} = \frac{n_2 \cos \theta_1 - n_1 \cos \theta_2}{n_2 \cos \theta_1 + n_1 \cos \theta_2}$$

- The reflected and transmitted intensities are calculated in terms of the Fresnel coefficients from the normal component of the Poynting vectors. The **reflectance** is

$$R = r_{12} r_{12}^*$$

For nonconducting media, the **transmittance** is

$$T = 1 - R$$

and for medium 2 conducting, the **absorptance** is

$$A = 1 - R$$

as a result of energy conservation. These equations are tractable for nonconductors or for normal incidence. They predict polarization by

reflection at Brewster's angle, total internal reflection at the critical angle, and the Hagen-Rubens formula for rf reflectance of a conductor.

- For **oblique incidence** on a conductor, the propagation and attenuation are described by $\kappa_r = N\omega/c$, $\kappa_i = q\omega/c$, with the planes of constant phase at an angle Θ with the boundary and the planes of constant amplitude parallel to the boundary:

$$N = \sqrt{p^2 + n_1^2 \sin^2 \theta_1}, \quad \sin \Theta = n_1 \sin \theta_1 / N$$

The quantities p and q are generalizations of n and k that are related to $(K_r - K_1 \sin^2 \theta_1)$ and K_i just as n and k are for $\theta_1 = 0$. They also explain total reflection for angles greater than critical.

- For **two plane parallel boundaries**, superposition gives the relative reflected amplitude as

$$r = \frac{r_{12} + r_{23}e^{i\beta}}{1 + r_{12}r_{23}e^{i\beta}}$$

where

$$\beta = 2d \frac{\omega}{c} n_2 \cos \theta_2 = 2d \frac{\omega}{c} (p + iq)$$

is the phase shift (and attenuation) on two traversals of the layer. The net reflectance is

$$R = rr^*, \quad 1 - R = T + A$$

where $A = 0$ if the layer is nonconducting. Through β , R shows interference effects that depend on (d/λ_1) . With conduction or frustrated total reflection,

$$T \sim e^{-2d/\delta}$$

if $d \gtrsim \delta$, where δ is the attenuation length.

- **Guided waves** propagate without attenuation (assuming perfectly reflecting boundaries) if the vacuum wavelength $\lambda_0 = 2\pi c/\omega$ is shorter than a cutoff wavelength λ_c , which is dependent on the dimensions of the waveguide:

$$\frac{1}{\lambda_g^2} = \frac{1}{\lambda_0^2} - \frac{1}{\lambda_c^2}$$

where λ_g is the wavelength of the guided wave. The group velocity (of energy propagation) is v_g , where

$$v_g v_p = c^2$$

For a rectangular guide in the TE_{mn} mode,

$$\frac{1}{\lambda_c^2} = \left(\frac{m}{2a}\right)^2 + \left(\frac{n}{2b}\right)^2$$

where $a > b$. Usually one restricts λ_0 so that only TE_{10} can propagate.

• For plane waves that are attenuated or guided (and also for spherical waves), the amplitude is not necessarily constant over a surface of constant phase. Such waves are called **inhomogeneous**. The Poynting vector is more complicated than for a homogeneous plane wave in a dielectric medium.

PROBLEMS

18-1 (a) Calculate the Fresnel reflection coefficient for an s -polarized wave incident from air onto a dielectric at the Brewster angle, $\theta_1 = \theta_B$. (b) Find the reflectance if $n = 1.5$.

18-2 A p -polarized wave is incident from air onto a dielectric surface near grazing incidence, $\theta_1 = \frac{1}{2}\pi - \delta$. Find the slope of the $R_p(\theta_1)$ curve as δ approaches zero, in terms of the dielectric constant K .

18-3 A p -polarized wave is incident from a transparent medium of dielectric constant K onto an air boundary at an angle slightly less than the critical angle, $\theta_1 = \theta_c - \delta$. Approximate R_p as a function of δ as δ approaches zero, and show that the slope of the $R_p(\theta_1)$ curve is infinite at θ_c .

18-4 Suppose an optical fiber has a refractive index $n = 1.55$. (a) Compute the largest angle between the fiber axis and a light ray that will propagate along the fiber, if the fiber is surrounded by air. (b) Compute the angle, if the fiber is surrounded by a cladding material with index 1.53.

18-5 A light wave with p -polarization in air is reflected from a metal surface. (a) Calculate R_p under the assumption that $\cos \theta_2 \cong 1$, as is often the case. (b) Find the value of θ_1 for which R_p is a minimum. (c) Evaluate this θ_1 and the corresponding R_p if $n = 1$, $k = 6$ (appropriate for aluminum).

18-6 A plane wave is normally incident from air onto the plane boundary of a metal. Assume that the frequency is in the range where $n \cong k \gg 1$. (a) From the Fresnel transmission coefficient, find $|E_2|^2$ just inside the metal surface. (b) Calculate the energy dissipation per unit volume near the surface, and (c) evaluate it if the incident amplitude is $E_1 = 10 \text{ V/cm}$ and the frequency is $f = 10^{10} \text{ Hz}$.

18-7 A wave in air is obliquely incident on a conducting surface at angle θ_1 , in the frequency range where the Hagen-Rubens relation holds. Show that Eq. (18-60) is replaced by

$$A_s = \frac{2 \cos \theta_1}{k}, \quad A_p = \frac{2}{k \cos \theta_1}$$

18-8 A wave in air is reflected at normal incidence from a conducting surface. (a) From \hat{r}_{12} , show that the phase shift of the E -vector is

$$\alpha_s = \tan^{-1} \frac{2k}{n^2 + k^2 - 1}$$

(b) Verify that this result goes to $\alpha_s = \pi$ for infinite conductivity.

18-9 Suppose a radio wave of $\omega = 10^7 \text{ s}^{-1}$ is reflected from the earth's surface at normal incidence. Compute the phase shift on reflection from the result of Problem 18-8, assuming that $K = 9$, $g = 10^{-4} (\Omega\text{m})^{-1}$ for this terrain.

18-10 A dielectric medium of refractive index n_3 is overcoated with a layer of index

n_2 , and a wave is incident from the dielectric medium n_1 . Show that $r_{12} = r_{23}$ for $n_2 = \sqrt{n_1 n_3}$ and thus $R = 0$ for $\cos \beta = -1$, at normal incidence.

18-11 A beam of monochromatic light (frequency ω) in air is incident normally on a dielectric film of refractive index n . The thickness of the film is d . Calculate the reflection and transmission coefficients as a function of d and n , by satisfying the boundary conditions at both faces. Let there be waves traveling toward the right and left inside the film, E_2 and E'_2 , in addition to the incident E_1 , reflected E'_1 , and transmitted E_3 waves.

18-12 Consider the matrix equation

$$\begin{pmatrix} E_m \\ E'_m \end{pmatrix} = (C_m) \begin{pmatrix} E_{m+1} \\ E'_{m+1} \end{pmatrix}$$

where

$$(C_m) = \frac{1}{t_{m,m+1}} \begin{pmatrix} e^{i(1/2)\beta_m} & r_{m,m+1} e^{i(1/2)\beta_m} \\ r_{m,m+1} e^{-i(1/2)\beta_m} & e^{-i(1/2)\beta_m} \end{pmatrix}$$

(a) Verify that the results for a single interface between two dielectrics are given by

$$\begin{pmatrix} E_1 \\ E'_1 \end{pmatrix} = (C_1) \begin{pmatrix} E_2 \\ 0 \end{pmatrix}$$

with $\beta_1 = 0$.

(b) Verify that the results for two interfaces are given by

$$\begin{pmatrix} E_1 \\ E'_1 \end{pmatrix} = (C_1)(C_2) \begin{pmatrix} E_3 \\ 0 \end{pmatrix}$$

with $\beta_2 = 2d_2(\omega/c)n_2 \cos \theta_2$. This approach is convenient to generalize to a system of multiple layers separated by plane boundaries.

18-13 A metal surface is overcoated by a dielectric layer. Calculate the resulting reflectance R , assuming that the metal is a perfect conductor ($g = \infty$).

18-14 Radiation is normally incident on an unsupported metal film in air. Assume that the film is thick enough that multiple reflections can be neglected. Calculate the transmittance T in terms of n and k of the metal.

18-15 Consider a dielectric slab that is too thick for the multiple reflected beams to interfere coherently. (a) Add up all the intensities to find the net reflectance in terms of the reflectances R_{12} and R_{23} of the individual boundaries. (b) Specialize the result to the case of identical media on each side of the slab.

18-16 Find the surface charge density and the current per unit width on the surface of a perfect conductor on which plane electromagnetic waves are incident, when the electric vector is (a) perpendicular to the plane of incidence, and (b) parallel to the plane of incidence.

18-17 Determine \mathbf{E} and \mathbf{B} for TM waves propagating in the yz -plane between two parallel, perfectly conducting plates, at $y = 0$ and at $y = b$.

18-18 Consider a TM wave in a rectangular waveguide ($H_z = 0$) propagating in the z -direction with wavelength λ_g . Show that

$$E_z = A \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{b} e^{2\pi iz/\lambda_g}$$

satisfies the wave equation, Eq. (18-123) and the boundary conditions. What is the cutoff frequency of the TM_{11} mode? Why is there no TM_{10} mode?

18-19 Determine the limiting values of the width a of a waveguide of square cross section that will transmit a wave of length λ in the TE_{10} mode but not in the TE_{11} or TM_{11} modes.

18-20 (a) Write down the \mathbf{E} and \mathbf{H} fields for the TE_{101} mode of a cubic cavity of side a . (b) Sketch the nature of the field distributions throughout the cube.

18-21 Indices of refraction for crown glass, flint glass, and diamond are 1.52, 1.65, and 2.42. (a) Calculate the reflectance R for waves of s - and p -polarization at the interface from air to each of the above materials, and (b) plot graphs analogous to Fig. 18-3. (c) Calculate Brewster's angle.

Dispersion and Oscillating Fields in Dispersive Media

How an electromagnetic wave propagates in a medium depends on the optical constants n and k . These in turn depend on the material constants: permittivity ϵ , permeability* μ , and conductivity g of the medium. So far these parameters have been treated as given, but notice has been made of the fact that their values may (in fact always do) vary greatly over the frequency range of the radiation from dc to X-rays. In order to use judgment in applying electromagnetic theory, one must know something of the principles behind this frequency dependence, which is called *dispersion*.

In this chapter, we first introduce a general definition of a linear system and then develop a microscopic model that predicts the optical behavior of materials. This model is an extension to oscillating fields of that treated briefly in Chapters 5 and 7 for static fields. The *Drude-Lorentz theory*, by which name this model is known, is based on treating the charged particles that constitute the material as harmonic oscillators or as free particles depending on the nature of the material.

19-1

LINEARITY AND CAUSALITY

In earlier chapters, we have frequently described media as linear, meaning by this that $\mathbf{D}(t)$, for example, was related to $\mathbf{E}(t)$ by a constant of proportionality called the permittivity. In dispersive media, such a relationship is not possible, but the concept of a linear medium is still very useful.

* In this chapter, we continue to consider only media for which $\mu = \mu_0$.

We define a linear medium (or system) as one that when subjected to an excitation (or input, or driving force) $I(t)$ yields a response (or output) $O(t)$ given by

$$O(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} R(t - t') I(t') dt' \quad (19-1a)$$

where $R(t)$ is called the response function and *does not depend on* $I(t)$. (The factor $1/\sqrt{2\pi}$ is introduced for later convenience.)

For real physical systems, there are, as we shall soon see, restrictions on $R(t - t')$. In particular, *causality* requires that $R(t - t') = 0$ for $t' > t$. By changing the integration variable from t' to $t'' = t - t'$, Eq. (19-1a) can be changed to the equivalent form:

$$O(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} R(t'') I(t - t'') dt'' \quad (19-1b)$$

Equation (19-1) describes a linear relationship between $O(t)$ and $I(t)$ in the sense that if $I(t)$ is multiplied by a constant, $O(t)$ is multiplied by that constant, and if two inputs are applied simultaneously the output is the sum of the outputs produced by the two inputs separately. These properties assure us that we can use the superposition principle to obtain the result of applying several simultaneous inputs, or to determine the response to a complicated input by decomposing it into simpler excitations and adding the responses. These are also the essential properties of a linear operator or a linear transformation, for which there is an enormous body of relevant mathematics. We do not invoke this mathematics, but we shall use Eqs. (19-1a) or (19-1b) as our definition of a linear medium. For example, if the displacement $D(t)$ is related to the electric field $E(t)$ by an equation of the form of Eq. (19-1), the medium would be described as a linear medium. Similar relationships between the current density $J(t)$ and electric field $E(t)$, or the magnetic induction $B(t)$ and the magnetic intensity $H(t)$, would be called linear if they satisfied equations of this form. We note that a material may, for example, have a linear dielectric behavior and a nonlinear magnetic behavior, as in some ferrites. These relationships are the generalizations of Eqs. (4-32a), (7-10), and (9-37) to general time-dependent fields and cases where the material cannot be described by material parameters ϵ , μ , and g that are simple constants of proportionality. In the case of a dielectric medium, the response function $R(t) = \sqrt{2\pi} \epsilon \delta(t)$ with a constant ϵ yields Eq. (4-32a); however, there is no such simple definition of permittivity for general time-dependent fields.

Equations (19-1) have two very important consequences. The first of

these is that the Fourier transforms* of $O(t)$, $R(t)$, and $I(t)$, which we designate by $\tilde{O}(\omega)$, $\tilde{R}(\omega)$, and $\tilde{I}(\omega)$, are related by

$$\tilde{O}(\omega) = \tilde{R}(\omega)\tilde{I}(\omega) \quad (19-2)$$

Conversely, if the Fourier transforms of the three functions are related by Eq. (19-2), then their inverse transforms are related by an equation of type Eq. (19-1). The Fourier transforms and inverses are defined by

$$\tilde{O}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} O(t)e^{i\omega t} dt, \quad O(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{O}(\omega)e^{-i\omega t} d\omega \quad (19-3)$$

and similarly for $\tilde{R}(\omega)$ and $\tilde{I}(\omega)$. Equation (19-2) can be obtained easily from Eq. (19-1) by using the definitions of the Fourier transforms and the representation of the delta function as

$$\delta(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} dt \quad (VI-8)$$

The fact that Eq. (19-1) implies Eq. (19-2), and conversely, is a general theorem called the *convolution theorem*.

Equation (19-2) is quite important because it tells us that in a linear medium the response to a monochromatic field is simply proportional to the impressed field and, conversely, that if such a proportionality exists, the medium is linear. The proportionality is, however, frequency dependent. The significance of proportionality is that the results obtained in Chapters 17 and 18 can be applied to more complicated linear media through the mechanism of Fourier transforms. We also note that Eq. (19-2) for zero frequency establishes the connection between this chapter and Chapters 4 through 12.

The second important consequence of Eq. (19-1) is that it implies that $O(t)$ and $I(t)$ are related by a linear differential equation with constant coefficients, which we write as

$$\sum_{n=0}^N a_n \frac{d^n}{dt^n} O(t) = I(t) \quad (19-4)$$

where N is the degree of the differential equation. The converse is easy to

* A brief summary of some of the fundamentals of Fourier integrals is presented in Appendix VI. We use the tilde notation to denote the specific Fourier transform of a corresponding time-dependent function: $\tilde{E}(\omega) = (2\pi)^{-\frac{1}{2}} \int E(t) \exp(i\omega t) dt$. Such transformations are generally complex. We use the circumflex to denote complex quantities in general. To keep the notation consistent with that of earlier chapters, we shall continue to use the circumflex for quantities such as the complex dielectric constant $\hat{K}(\omega)$, complex conductivity $\hat{g}(\omega)$, and so on, even though their dependence on ω implies that they are Fourier transforms.

prove by simply writing $O(t)$ and $I(t)$ as inverse transformations and substituting in Eq. (19-4). The n th time derivative becomes multiplication by $(-i\omega)^n$, and one finds

$$\sum_{n=0}^N a_n (-i\omega)^n \tilde{O}(\omega) = \tilde{I}(\omega) \quad (19-5a)$$

or

$$\tilde{R}(\omega) = \frac{1}{\sum_{n=0}^N a_n (-i\omega)^n} \quad (19-5b)$$

Equations (19-5) thus represent an equation of the form of Eq. (19-2), and from the convolution theorem, there is a relation between $O(t)$ and $I(t)$ of the form of Eq. (19-1). The proof that Eq. (19-1) implies Eq. (19-4) is more difficult, and presenting it would take us too far afield.* The importance of this result is that for a linear medium, the output is related to the input by a linear differential equation with constant coefficients and that a medium in which the output is related to the input by a linear differential equation is linear.

So far we have addressed only the question of linearity. There are other requirements that must be satisfied by the response function on physical grounds. The first of these is that a real input should produce a real output, and consequently $R(t)$ must be *real* for all real values of t . Note that $O(t)$ and $I(t)$ in Eq. (19-1) are not required to be real (although the actual physical inputs and output are real). The fact that $R(t)$ is real means that

$$\tilde{R}^*(\omega) = \tilde{R}(-\omega) \quad (19-6)$$

as can be shown easily from

$$\tilde{R}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} R(t) e^{i\omega t} dt$$

Thus, the a_n 's in Eq. (19-5b) must be real for even n 's and imaginary for odd n 's.

The second requirement is that Eq. (19-1) should be *causal* in the sense that an input at time t should not produce an output for times less (earlier)

* One method of proof is sketched in A. H. Zemanian, *Distribution Theory and Transform Analysis* (New York: McGraw-Hill, 1965; Reprinted, New York: Dover, 1987), (p. 130, Ex. 5 in the Dover edition).

than t . This condition requires* that

$$R(t'') = 0 \quad \text{for } t'' < 0 \quad (19-7)$$

Finally, we require, again on physical grounds, that a finite input produce only a *finite* output, and that an input at time t' produce an output at time t that decreases monotonically at large values of $(t - t')$.†

The ideas developed so far can be summarized as follows:

1. Linearity is defined by Eqs. (19-1).
2. The Fourier transforms of $O(t)$, $R(t)$, and $I(t)$ are related by Eq. (19-2), and if Eq. (19-2) holds, the medium (system) is linear.
3. For a linear medium (system), $O(t)$ and $I(t)$ are related by a linear differential equation with constant coefficients, Eq. (19-4), and conversely.

For actual physical systems, we must also require the following:

4. $R(t)$ is real for all t .
5. $R(t) = 0$ for $t < 0$.
6. $R(t)$ is square integrable over the interval $-\infty < t < \infty$.

The electric displacement in a linear dielectric medium would be related to the electric field by

$$\mathbf{D}(\mathbf{r}, t) = \frac{1}{\sqrt{2\pi}} \int_0^\infty R(\mathbf{r}, t') \mathbf{E}[\mathbf{r}, (t - t')] dt' \quad (19-8)$$

with $R(\mathbf{r}, t)$ a function of t that does not depend on \mathbf{E} . The lower limit on the integral can be taken to be zero because of Eq. (19-7). Equation (19-8) contains the additional assumption that $\mathbf{D}(\mathbf{r}, t)$ depends only on the value of \mathbf{E} at the same point. Such a relationship is called *spatially local*, and only such relations will be considered. We will also assume that $R(\mathbf{r}, t)$ changes only at sharp interfaces, for example, at the boundary between two semi-infinite homogeneous dielectric bodies. This assumption is a restrictive one and excludes some cases of practical importance.

* Equation (19-7) has implications for $R(\omega)$, namely, that the roots, ω_n , of the equation

$$\sum_{n=0}^N a_n (-i\omega)^n = 0$$

must have negative imaginary parts. This can be proved using Cauchy's Integral Theorem. (See Appendix VII.)

† This is equivalent to requiring that $R(t)$ be square integrable from which it follows that $\bar{R}(\omega)$ is also square integrable.

Then Eq. (19-8) can be written as

$$D(t) = 1/(2\pi)^{1/2} \int_0^{\infty} R(t')E(t - t') dt' \quad (19-8a)$$

with the understanding that the equation refers to a single point in space and that $D(t)$ and $E(t)$ are corresponding components of the vector quantities. Note that $D(t)$ depends not just on the value of E at time t , but on the value of E at all past times. Thus, there is generally no simple quantity that can properly be called the permittivity in the time domain; *permittivity is a frequency domain concept.*

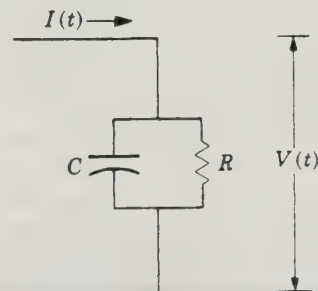
The exception to the preceding statement, as already noted, is the idealized case where $R(t') = (2\pi)^{1/2}\epsilon \delta(t')$ with ϵ independent of time. Then $D(t) = \epsilon E(t)$, so this case is the only one in which the static and slowly varying treatments of earlier chapters strictly apply. It is worth repeating that a medium in which the relationship between E and D is given by Eq. (19-8a) is a linear medium. If R depends explicitly on E , then the medium is nonlinear and further phenomena occur. The best known of these is *mixing*, in which a monochromatic excitation produces a response that is not monochromatic. We will not consider the response of nonlinear media to time-dependent fields.

To illustrate further the meaning of linearity, we consider the simple RC circuit shown in Fig. 19-1. The input is the known current $I(t)$ and the output is $V(t)$. This circuit is governed by the differential equation

$$\frac{1}{R} V(t) + C \frac{dV}{dt} = I(t) \quad (19-9)$$

This equation is an ordinary linear differential equation with constant

FIGURE 19-1
Simple RC parallel circuit
drawing the current $I(t)$



coefficients. Consequently, it describes a linear system and everything developed above for such systems applies. In particular, $V(t)$ can be written in the form of Eq. (19-1). An explicit form for $R(t)$ can be found by noting that $(1/C) \exp(t/RC)$ is an integrating factor that converts the left-hand side into an exact differential. With a little elementary manipulation, we obtain the particular solution

$$V(t) = \frac{1}{C} \int_0^{\infty} e^{-t'/RC} I(t-t') dt' \quad (19-10)$$

which is exactly the form of Eq. (19-1b) with the response function $R(t) = (1/C)(2\pi)^{1/2} \exp(-t/RC)$ for $t > 0$; $R(t) = 0$ for $t < 0$. The response function, $R(t)$, is *not* the impedance of the circuit; impedance is a concept associated with a single frequency. We can calculate $\tilde{R}(\omega)$ from $R(t)$ by using the inverse Fourier transform. We find

$$\tilde{R}(\omega) = \frac{1}{\frac{1}{R} - \frac{1}{(i\omega C)}} \quad (19-11)$$

This $\tilde{R}(\omega)$ would be the impedance of a resistor R in parallel with a capacitor C at frequency ω as given by Eq. (13-28) if the sign in the denominator of Eq. (19-11) were positive. The sign reversal has its origin in the choice of $\exp(i\omega t)$ for the time dependence in Chapter 13 (to conform with the usual convention in dealing with ac circuits) and the choice of $\exp(-i\omega t)$ in the Fourier transform representation of $I(t)$, $R(t)$, and $O(t)$ according to Eq. (19-3).^{*} With this minor issue resolved, we can write

$$\tilde{V}(\omega) = \tilde{R}(\omega)\tilde{I}(\omega) = \hat{Z}^*(\omega)\tilde{I}(\omega) \quad (19-12)$$

Thus, $\tilde{R}(\omega) = \tilde{V}(\omega)/\tilde{I}(\omega)$ is simply related to the impedance, showing that the concept of impedance has meaning in the frequency domain, but not in the time domain. Similar considerations apply to other linear relations between physical variables, and it is often possible to identify the frequency response function, $\tilde{R}(\omega)$, with physical parameters. It is important to understand that linearity does not imply simple proportionality between $I(t)$ and $O(t)$; linearity does imply proportionality between $\tilde{I}(\omega)$ and $\tilde{O}(\omega)$, but even here the proportionality depends on ω .

Linearity and causality as defined above have important consequences that apply to electromagnetic fields and many other physical systems. We will now investigate some of these consequences.

^{*} Either choice for the time dependence can be used. There will be no difference in the physical results. In comparing work using complex exponential time dependence, one must be sure that the same time dependence is used or make appropriate adjustments.

FREQUENCY RESPONSE AND DISPERSION RELATIONS

We consider a classical forced damped oscillator consisting of a particle of mass m and charge e bound to an equilibrium position by a linear restoring force and subject to an electric field $E(t)$. This system is governed by the equation

$$m \frac{d^2x}{dt^2} + G \frac{dx}{dt} + Cx = eE(t)$$

or

$$\frac{d^2x}{dt^2} + \gamma \frac{dx}{dt} + \omega_0^2 x = \frac{e}{m} E(t) \quad (19-13)$$

The electric field $E(t)$ provides the driving force for the oscillator. The natural frequency of the undamped oscillator is $\omega_0 = (C/m)^{1/2}$, and $\gamma = G/m$ is the damping constant. Equation (19-13) is a linear differential equation with constant coefficients, so the system is linear and the results of Section 19-1 can be used. The frequency response is found from Eq. (19-5b) with $a_0 = \omega_0^2$, $a_1 = \gamma$ and $a_2 = 1$, or by using the Fourier transform representation of $x(t)$ and $E(t)$. Either approach gives

$$\bar{x}(\omega) = \frac{e/m}{(\omega_0^2 - \omega^2) - i\gamma\omega} \bar{E}(\omega) = \bar{R}(\omega) \bar{E}(\omega) \quad (19-14)$$

$\bar{E}(\omega)$ and $\bar{x}(\omega)$ are complex. The frequency response function, $\bar{R}(\omega)$, can be resolved into real and imaginary parts:

$$\bar{R}(\omega) = \frac{e}{m} \left[\frac{\omega_0^2 - \omega^2}{(\omega_0^2 - \omega^2)^2 + \gamma^2 \omega^2} \right] + i \frac{e}{m} \frac{\gamma \omega}{(\omega_0^2 - \omega^2)^2 + \gamma^2 \omega^2} \quad (19-15)$$

This frequency response function is complex, and both its real and imaginary parts depend on frequency.

A remarkable and most useful property of the real and imaginary parts of $\bar{R}(\omega)$ is, as we will now demonstrate for a specific response function, that if the real part of it is known for all real values of ω , then the imaginary part can be determined, and conversely. The relationship between the real and imaginary parts of the response function could be demonstrated using the frequency response function given in Eq. (19-15); however, some of the mathematics is very laborious. For our purposes, it is equally instructive to consider an approximation to the frequency response function for the forced-damped harmonic oscillator. We first note that if γ is small, then $\bar{R}(\omega)$ is small except where $\omega \cong \omega_0$. So we approximate Eq. (19-15) by

$\tilde{T}(\omega)$ as follows:

$$\begin{aligned}\tilde{R}(\omega) &= \frac{e/m}{(\omega_0 + \omega)(\omega_0 - \omega) - i\gamma\omega} \\ &\cong \tilde{T}(\omega) = \frac{e/m}{2\omega_0 \left[(\omega_0 - \omega) - i\frac{\gamma}{2} \right]}\end{aligned}\quad (19-16)$$

where the approximation consists in replacing ω by ω_0 everywhere, except where it occurs in the factor $(\omega_0 - \omega)$. The approximation of $\tilde{R}(\omega)$ by $\tilde{T}(\omega)$ is valid only for $\omega > 0$; however, we now consider $\tilde{T}(\omega)$ for $-\infty < \omega < \infty$ as the approximate response function for our dynamical system.* We can easily show that $\tilde{T}(\omega)$ satisfies all of the conditions on physical systems except that $T(t)$ is not real. (Such a requirement is not needed for the present analysis.) Separating $\tilde{T}(\omega)$ into real and imaginary parts gives

$$\tilde{T}(\omega) = \frac{e}{2m\omega_0} \left[\frac{\omega_0 - \omega}{(\omega_0 - \omega)^2 + \frac{\gamma^2}{4}} + i \frac{\gamma/2}{(\omega_0 - \omega)^2 + \frac{\gamma^2}{4}} \right] \quad (19-17)$$

Equation (19-17) is much simpler than Eq. (19-15) and is easier to manipulate mathematically. As we have noted, it does not represent a classical forced-damped oscillator, but it does represent a system that approximates such an oscillator.

With this caveat, we compute a function $W(\omega)$ as follows:

$$W(\omega) = \lim_{\epsilon \rightarrow 0} \left\{ \int_{-\infty}^{\omega - \epsilon} + \int_{\omega + \epsilon}^{\infty} \frac{d\omega'}{\left[(\omega_0 - \omega')^2 + \frac{\gamma^2}{4} \right] (\omega' - \omega)} \right\} \quad (19-18)$$

This function will be seen shortly to be related to $\tilde{T}(\omega)$. The right-hand side of Eq. (19-18), which excludes from the range of integration a small interval around the position of the singularity in the integrand and then takes the limit as this small interval goes to zero, is known as the *Cauchy principal value* of the integral from minus infinity to plus infinity, and it is usually designated by placing a *P* in front of the integral. Changing the variable of integration to $x = \omega' - \omega_0$ converts the integrals in Eq. (19-18) to a

* The approximate system that yields $\tilde{T}(\omega)$ is governed by the differential equation

$$-2i\omega_0 \frac{dx}{dt} + (2\omega_0^2 - i\gamma\omega_0)x(t) = \frac{e}{m} E(t)$$

which is interesting because it is a first-order differential equation that approximates Eq. (19-13).

damped oscillator is, as we have noted in the last section,

$$m \frac{d^2x}{dt^2} + G \frac{dx}{dt} + Cx = eE_m \quad (19-22a)$$

or

$$\frac{d^2x}{dt^2} + \gamma \frac{dx}{dt} + \omega_0^2 x = \frac{eE_m}{m} \quad (19-22b)$$

where e and m are the charge and mass of the particle, and E_m is the “molecular field” discussed in Chapter 5. The damping constant $\gamma = G/m$ has the dimensions of frequency. The natural frequency of the undamped oscillator is ω_0 and is related to the force constant C by $m\omega_0^2 = C$. In Chapter 5, the force constant was expressed for bound outer-shell electrons in terms of the “radius” of the atom R_0 . In the static case, x is independent of t , so that Eq. (19-22) becomes identical to Eq. (5-12), with

$$\omega_0^2 = \frac{e^2}{4\pi\epsilon_0 m R_0^3} \quad (19-23)$$

For free electrons, we put $\omega_0 = 0$ in Eq. (19-22), which becomes identical to Eq. (7-31), with

$$\gamma = \frac{1}{\tau} \quad (19-24)$$

where τ is the mean time between collisions. Appropriate values of ω_0 and γ for other cases will be discussed later. To take some account of mutual interactions among the particles, the driving field E_m is assumed to depend on the E -field of the wave E according to

$$E_m = E + \frac{\nu}{\epsilon_0} P \quad (19-25)$$

where P is the polarization of the medium. It was shown in Eq. (5-7) that for an isotropic, nonpolar dielectric, $\nu = \frac{1}{3}$. For a metal, $\nu = 0$. It is not our main purpose here to deal with the difficult problem of the proper local field correction, and so we will just leave it as ν and try to eliminate it from consideration in the following.

We note that the spatial variation of E_m is quite small over $2x_0$, the distance that the oscillating charge moves. In the case of atoms, x_0 is a small fraction of R_0 , the “radius” of the atom, typically a few angstroms. Visible light has a wavelength in free space that is greater than about 4000 Å, and consequently, even with a dielectric constant as large as 400, the wavelength in the material would be greater than 10 Å. It would thus be reasonable to neglect the variation of E_m over the distance $2x_0$. The connection between the mechanical displacement x of the microscopic charged particles that make up the material and the macroscopic electrical response of the medium is given by calculating the polarization P . The dipole moment due

These relations enable us to eliminate negative values of ω from the range of integration in the dispersion relations. The result, as is shown in Appendix VII, is

$$R_r(\omega) = \frac{2}{\pi} P \int_0^{\infty} \frac{\omega' R_i(\omega')}{\omega'^2 - \omega^2} d\omega' \quad (19-21a)$$

$$R_i(\omega) = -\frac{2\omega}{\pi} P \int_0^{\infty} \frac{R_r(\omega')}{\omega'^2 - \omega^2} d\omega' \quad (19-21b)$$

If the response function is the electric susceptibility of a dielectric medium—that is, if $\tilde{R}(\omega) = \tilde{\chi}(\omega)$, then Eqs. (19-21) are the famous *Kramers-Kronig dispersion relations*.

The first application of the dispersion relations was the determination of the index of refraction of a solid material for X-rays. In this case, it is relatively easy to measure the absorption of the radiation by the material, but very difficult to measure its refraction. The index of refraction can, however, be determined by using the dispersion relations and the measured X-ray absorption. We now turn to an examination of dispersive media, where we shall find occasion to illustrate the use of the dispersion relations.

19-3 DRUDE-LORENTZ HARMONIC OSCILLATOR MODEL

All ordinary matter is composed of negative electrons and positive nuclei. If, for the purpose at hand, some of the electrons (more or fewer than Z , the nuclear charge) can be considered as tightly bound to the nucleus and moving with it, the composite entity is a charged ion. The electrons or ions will be treated as harmonic oscillators—that is, particles bound to an equilibrium position by a linear restoring force. For generality, we make it a damped harmonic oscillator, including a linear damping force proportional to the velocity. When an electromagnetic wave is present, the oscillator is driven by the electric field of the wave.* The response of the medium is obtained by adding up the motions of the particles; since the assumed forces are linear, the K and g that result from the model will be constant (i.e., independent of E , although they will depend on frequency). Applied to electrons, the model describes the bound electrons in atoms, but free electrons can be included as a special case simply by putting the restoring-force constant of the oscillator equal to zero.

The classical equation of motion† for the one-dimensional forced,

* The Lorentz force is $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$, but for a wave $B = (n/c)E$. The magnetic force is smaller by $n(v/c)$, and will be neglected. In any case the magnetic force does no work on the particle, since it is perpendicular to \mathbf{v} .

† Quantum mechanics should of course be used for electrons. The classical solution is useful because it happens to be the same as the quantum one, with a suitable reinterpretation of the natural frequency ω_0 .

damped oscillator is, as we have noted in the last section,

$$m \frac{d^2x}{dt^2} + G \frac{dx}{dt} + Cx = eE_m \quad (19-22a)$$

or

$$\frac{d^2x}{dt^2} + \gamma \frac{dx}{dt} + \omega_0^2 x = \frac{eE_m}{m} \quad (19-22b)$$

where e and m are the charge and mass of the particle, and E_m is the "molecular field" discussed in Chapter 5. The damping constant $\gamma = G/m$ has the dimensions of frequency. The natural frequency of the undamped oscillator is ω_0 and is related to the force constant C by $m\omega_0^2 = C$. In Chapter 5, the force constant was expressed for bound outer-shell electrons in terms of the "radius" of the atom R_0 . In the static case, x is independent of t , so that Eq. (19-22) becomes identical to Eq. (5-12), with

$$\omega_0^2 = \frac{e^2}{4\pi\epsilon_0 m R_0^3} \quad (19-23)$$

For free electrons, we put $\omega_0 = 0$ in Eq. (19-22), which becomes identical to Eq. (7-31), with

$$\gamma = \frac{1}{\tau} \quad (19-24)$$

where τ is the mean time between collisions. Appropriate values of ω_0 and γ for other cases will be discussed later. To take some account of mutual interactions among the particles, the driving field E_m is assumed to depend on the E -field of the wave E according to

$$E_m = E + \frac{\nu}{\epsilon_0} P \quad (19-25)$$

where P is the polarization of the medium. It was shown in Eq. (5-7) that for an isotropic, nonpolar dielectric, $\nu = \frac{1}{3}$. For a metal, $\nu = 0$. It is not our main purpose here to deal with the difficult problem of the proper local field correction, and so we will just leave it as ν and try to eliminate it from consideration in the following.

We note that the spatial variation of E_m is quite small over $2x_0$, the distance that the oscillating charge moves. In the case of atoms, x_0 is a small fraction of R_0 , the "radius" of the atom, typically a few angstroms. Visible light has a wavelength in free space that is greater than about 4000 Å, and consequently, even with a dielectric constant as large as 400, the wavelength in the material would be greater than 10 Å. It would thus be reasonable to neglect the variation of E_m over the distance $2x_0$. The connection between the mechanical displacement x of the microscopic charged particles that make up the material and the macroscopic electrical response of the medium is given by calculating the polarization P . The dipole moment due

to the displaced charge e is ex (assuming that the neutralizing charge $-e$ is remaining at rest). Then,

$$P = Nex \tag{19-26}$$

where N is the number of charges per unit volume.

Using Eqs. (19-25) and (19-26) converts Eq. (19-22) to

$$\frac{d^2P}{dt^2} + \gamma \frac{dP}{dt} + \omega_0^2 P - \nu \omega_p^2 P = \epsilon_0 \omega_p^2 E(t) \tag{19-27}$$

where

$$\omega_p^2 = Ne^2/\epsilon_0 m \tag{19-28}$$

Equation (19-27) is a linear differential equation with constant coefficients. Consequently the solution to Eq. (19-27) can be written in the form of Eq. (19-1). We also know that

$$\tilde{P}(\omega) = \tilde{R}(\omega)\tilde{E}(\omega)$$

with

$$\tilde{R}(\omega) = \frac{\epsilon_0 \omega_p^2}{(-i\omega)^2 + \gamma(-i\omega) + \omega_0^2 - \nu \omega_p^2} \tag{19-29}$$

Since $\tilde{\chi}(\omega)$ is the coefficient of proportionality between $\tilde{P}(\omega)$ and $\tilde{E}(\omega)$, we may write

$$\tilde{\chi}(\omega) = \tilde{R}(\omega) = \frac{\epsilon_0 \omega_p^2}{\omega_0^2 - \omega^2 - i\gamma\omega - \nu \omega_p^2} \tag{19-30}$$

It is easily verified that this susceptibility satisfies the requirements for a real response function, that it is causal as long as $\gamma > 0$, and that it is square integrable. Thus, as would be expected, it satisfies all of the requirements for the response function of a physical system. Susceptibility is defined in this way as a function of frequency.

The relationship between the susceptibility and the model parameters is often written in other forms. It requires only elementary algebra to obtain

$$\frac{\tilde{\chi}(\omega)}{1 + \nu \tilde{\chi}(\omega)/\epsilon_0} = \frac{\epsilon_0 \omega_p^2}{\omega_0^2 - \omega^2 - i\gamma\omega} \tag{19-31}$$

and using the definition of the dielectric constant, $\hat{K} = 1 + \tilde{\chi}/\epsilon_0$,

$$\frac{\hat{K} - 1}{1 + \nu(\hat{K} - 1)} = \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i\gamma\omega} \tag{19-32}$$

This equation gives the sought-for relation between the macroscopic dielectric constant \hat{K} and the microscopic properties of the charged particles that constitute the medium. Two important features of this result should be noted: First, \hat{K} is complex, and second, it is frequency dependent. Thus the simplest model we could set up automatically implies a conducting, dispersive medium.

Before discussing the consequences of the model, we should suggest a simple generalization, which extends its applicability to most real materials. Seldom do all the charged particles of a material have the same properties. For example, they may divide into electrons and ions, or into electrons in different inner and outer orbits of an atom, and so on. If there are N_i particles of charge e_i , mass m_i , natural resonant frequency ω_{0i} , and damping frequency γ_i , then following the same derivation as above leads to

$$\frac{\hat{K} - 1}{1 + \nu(\hat{K} - 1)} = \sum_i \frac{\omega_{pi}^2}{\omega_{0i}^2 - \omega^2 - i\gamma_i\omega} \quad (19-33)$$

If all of the particles have the same charge and mass (e.g., they are all electrons, but are distributed over different orbits so they have different restoring forces and different damping coefficients), this can be written

$$\frac{\hat{K} - 1}{1 + \nu(\hat{K} - 1)} = \omega_p^2 \sum_i \frac{f_i}{\omega_{0i}^2 - \omega^2 - i\gamma_i\omega} \quad (19-34)$$

where $f_i = N_i/N$ is the fraction of oscillators of type i . Since $\sum N_i = N$,

$$\sum f_i = 1 \quad (19-35)$$

(In the quantum mechanical interpretation of this result, f_i is called the *oscillator strength* and Eq. (19-35) is called the *f-sum rule*.) We shall limit our discussion to those cases that do not depend on subtleties of the local field correction ν . If the Lorentz local field $\nu = \frac{1}{3}$ applies, Eq. (19-34) becomes

$$\frac{\hat{K} - 1}{\hat{K} + 2} = \frac{1}{3} \omega_p^2 \sum \frac{f_i}{\omega_{0i}^2 - \omega^2 - i\gamma_i\omega} \quad (19-36)$$

Since ω_p^2 is proportional to N , this is a generalization of the Clausius-Mossotti equation. It is easy to solve Eqs. (19-33) or (19-36) for \hat{K} , and for some purposes the result is useful, but it does not clarify the frequency dependence of \hat{K} , which is the main subject of this chapter. However, if $\nu = 0$, Eq. (19-33) simplifies to

$$\hat{K} - 1 = \sum \frac{\omega_{pi}^2}{\omega_{0i}^2 - \omega^2 - i\gamma_i\omega} \quad (19-37)$$

This also holds, for any ν , at frequencies where \hat{K} is not too different from 1. In addition, at frequencies where one of the resonance peaks of Eq. (19-33) dominates all the others, we can eliminate ν from consideration:

$$(\hat{K} - 1)(\omega_0^2 - \omega^2 - i\gamma\omega) = [1 + \nu(\hat{K} - 1)]\omega_p^2$$

so that

$$\hat{K} - 1 = \frac{\omega_p^2}{(\omega_0^2 - \nu\omega_p^2) - \omega^2 - i\gamma\omega} \quad (19-38)$$

This result is the same as a single resonance peak with zero local field correction and effective resonance frequency $\sqrt{\omega_0^2 - \nu\omega_p^2}$. Therefore, we will be able to limit our detailed discussion of the frequency dependence of \hat{K} to the simple resonance peak

$$\hat{K} - 1 = \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i\gamma\omega} \quad (19-39)$$

and still be able to apply the results to many practical problems. Written in terms of real and imaginary parts, Eq. (19-39) is

$$K_r = 1 + \frac{\omega_p^2(\omega_0^2 - \omega^2)}{(\omega_0^2 - \omega^2)^2 + (\gamma\omega)^2} \quad (19-40a)$$

$$K_i = \frac{\omega_p^2\gamma\omega}{(\omega_0^2 - \omega^2)^2 + (\gamma\omega)^2} \quad (19-40b)$$

These quantities could then be substituted into Eqs. (17-61)

$$n = \sqrt{\frac{1}{2}[K_r + \sqrt{K_r^2 + K_i^2}]} \quad (19-41)$$

$$k = \sqrt{\frac{1}{2}[-K_r + \sqrt{K_r^2 + K_i^2}]}$$

in order to have explicit expressions for the dispersion of n and k as a function of frequency. The results would be too complicated, however, to reveal any general features of the dispersion. In the following sections we consider various cases of practical interest where simplifying assumptions can be made.

Before proceeding to these examples, we should also comment on a significant feature of this model—that \hat{K} is complex. Even though the model was set up for *bound* charges, the resulting complex \hat{K} is characteristic of a *conducting* medium. There is a nonvanishing $g = K_i\epsilon_0\omega$, without the intentional introduction of a conduction current density J . In addition, with $\omega_0 = 0$ for *free* charges, there is still a $K_r = K$ characteristic of a *dielectric* medium. The model automatically incorporates both real and imaginary parts of \hat{K} , corresponding to both the displacement current $\partial\mathbf{D}/\partial t$ and the conduction current \mathbf{J} in Maxwell's curl \mathbf{H} equation. In our model, we could have calculated

$$J = Ne \frac{dx}{dt} = -i\omega Nex$$

instead of $P = Nex$ and thus obtained an expression for the conductivity g . Clearly we would find that

$$J = \frac{dP}{dt} = -i\omega P$$

and

$$\tilde{g} = -i\omega\tilde{\chi} \quad (19-42)$$

For each of the groups of oscillators in Eq. (19-27), we have to calculate P or J but not both, since they are equivalent expressions of the fact that the particle displacement has a component in phase with the E -field and also a component 90° out of phase, or that the particle velocity has an out-of-phase component and also an in-phase component. For static fields, bound charges have displacement proportional to the field (in phase), and free charges have velocity proportional (in phase); but at high frequencies both bound and free charges can each have in-phase and out-of-phase components of displacement and velocity. All of the following discussion could instead be formulated in terms of the complex conductivity according to Eq. (19-42), and in certain contexts, it is more usual to do so. The relationship of \hat{g} to the complex dielectric constant \hat{K} is

$$\begin{aligned}\hat{g} &= -i\epsilon_0\omega(\hat{K} - 1) \\ g_r &= \epsilon_0\omega K_i, \quad g_i = -\epsilon_0\omega(K_r - 1)\end{aligned}\quad (19-43)$$

As expected, $g_r = g$ is the actual real conductivity, and g_i is related to the real dielectric constant.

19-4 RESONANCE ABSORPTION BY BOUND CHARGES

Let us apply the results of the preceding section to a system of bound charges, that is, to materials that are nonconductors for dc. To estimate the expected size of the parameters in Eqs. (19-40a) and (19-40b) for valence electrons, let us put $R_0 = 2 \text{ \AA}$ in Eq. (19-23):

$$\begin{aligned}\omega_0 &= \sqrt{\frac{(1.6 \times 10^{-19})^2}{4\pi(8.854 \times 10^{-12})(0.91 \times 10^{-30})(2 \times 10^{-10})^3}} \\ &= 5.6 \times 10^{15} \text{ s}^{-1}\end{aligned}$$

This electron resonance frequency corresponds to a wavelength in air of 3350 \AA (335 nm), which is in the ultraviolet region just beyond the visible spectrum. Larger or smaller R_0 would give a longer or shorter resonance wavelength. Combining Eq. (19-23) with Eq. (19-28), which defines ω_p , we find

$$\frac{\omega_p^2}{\omega_0^2} = 4\pi R_0^3 N = 3NV_a \quad (19-44)$$

where V_a is the volume of our "atom." Thus, if the atoms are closely packed, as when they are condensed in a liquid or solid,

$$\frac{\omega_p^2}{\omega_0^2} \approx 1 \quad (19-45)$$

When they are less densely aggregated, as in a gas (or a solution), this ratio is correspondingly less, since N is smaller but usually ω_0 is not strongly

affected by the state of aggregation. The ratio is also much less than 1 for the more tightly bound inner-shell electrons, which have smaller orbits. Reasonable values of the damping frequency γ are more difficult to estimate. To get further insight into its meaning, let us return to the equation of motion (19-22) and consider the case of the free oscillator with $E_m = 0$. The solution of this equation is also well known:

$$x = x_0 e^{-\gamma t/2} e^{-i\omega_0' t} \quad (19-46)$$

where

$$\omega_0' = \sqrt{\omega_0^2 - (\gamma/2)^2}$$

is nearly the same as ω_0 if γ is small. The amplitude of the oscillation decays as $e^{-\gamma t/2}$, and the energy of oscillation, which is proportional to the square of the amplitude, decays as $e^{-\gamma t}$. Thus we can write

$$\gamma = \frac{1}{\tau} \quad (19-47)$$

where τ is the mean decay time for the oscillator energy. The significance of τ in this equation is not the same as in Eq. (19-24) for free electrons, but it is analogous: In both cases, τ is the mean time of energy loss—energy of free harmonic oscillation or energy of free flight, respectively. The oscillating particle necessarily radiates electromagnetic energy at the expense of its energy of oscillation; in the next chapter, it is shown that the rate of radiation corresponds to a decay rate

$$\frac{\gamma}{\omega_0} = \frac{4\pi R_e}{3\lambda_0} \quad (19-48)$$

where

$$R_e = \frac{e^2}{4\pi\epsilon_0 mc^2} = 2.81 \times 10^{-15} \text{ m} \quad (19-49)$$

and $\lambda_0 = 2\pi c/\omega_0$ is the vacuum wavelength corresponding to ω_0 ; R_e is the so-called classical radius of the electron. (See Problem 6-6.)* Even for X-rays ($\lambda_0 \approx 1 \text{ \AA} = 10^{-10} \text{ m}$) the damping due to this mechanism is extremely small relative to the resonance frequency. Usually there are additional modes of decay due to interactions with other particles (collisions), which make the damping frequency enormously greater than the single-particle rate due to radiation. Still, however, we assume that

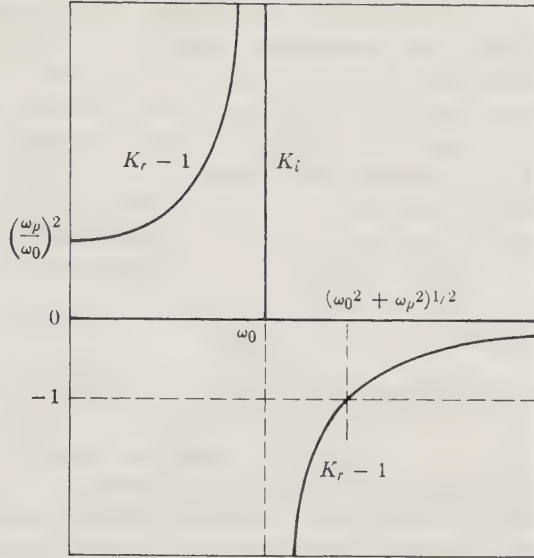
$$\frac{\gamma}{\omega_0} \ll 1$$

for the applications of this section.

* The charge distribution in an electron is not physically observable, so Eq. (19-49) is a more convenient definition than the one for a uniform charge distribution.

FIGURE 19-2

Dielectric constants as a function of frequency for an extremely narrow, strong absorption line at ω_0



Since γ is assumed to be small, a first approximation with $\gamma = 0$ is informative, even though unphysical. Equation (19-40a) with $\gamma = 0$ is

$$K_r - 1 = \frac{\omega_p^2}{\omega_0^2 - \omega^2} \quad (19-50)$$

as shown in Fig. 19-2. Equation (19-40b) with $\gamma = 0$ gives $K_i = 0$ at all frequencies except $\omega = \omega_0$, where it is infinite. Simply putting γ equal to zero does not give a useful result for K_i . Consequently, we use Eq. (19-40b) and take the limit as γ goes to zero. Then,

$$K_i = \lim_{\gamma \rightarrow 0} \left[\frac{\omega_p^2 \gamma \omega}{(\omega_0^2 - \omega^2)^2 + \gamma^2 \omega^2} \right] = \frac{\omega_p^2}{\omega} \lim_{\gamma \rightarrow 0} \frac{\gamma}{\left(\frac{\omega_0^2 - \omega^2}{\omega} \right)^2 + \gamma^2}$$

The latter limit is one of the representations given for the Dirac delta function in Appendix V. Thus, K_i can be written

$$K_i(\omega) = \frac{\omega_p^2}{\omega} \pi \delta\left(\frac{\omega_0^2 - \omega^2}{\omega}\right) = \frac{\omega_p^2 \pi}{\omega 2} \delta(\omega_0 - \omega) \quad (19-51)$$

for positive ω . The last step in Eq. (19-51) follows from the fact that $\delta[(\omega_0^2 - \omega^2)/\omega]$ is only nonzero in the vicinity of $\omega = \omega_0$ and hence can be approximated by $\delta(2\{\omega_0 - \omega\})$ and from $\delta(2\omega) = \frac{1}{2} \delta(\omega)$ [see Eq. (V-5)].

$K_i \neq 0$ describes energy dissipation, as could be seen by computing the rate of change of the energy stored in the electric field. Alternatively,

$K_i \neq 0$ is equivalent to $g_r \neq 0$, which gives a current in phase with the electric field and consequent dissipation of energy. The delta function approximation to $K_i(\omega)$ represents the case of an infinitely narrow absorption line at $\omega = \omega_0$, as is also indicated in Fig. 19-2. The most interesting feature of the figure is that K_r has values significantly different from 1 (for vacuum) at frequencies very far away from ω_0 , the only frequency where K_i differs from 0 (for vacuum). The effect of energy absorption in the optical region is manifested in the dielectric constant K (but not in the conductivity g) at all frequencies, even dc. For $\omega = 0$,

$$K_r(0) = K_0 = 1 + \frac{\omega_p^2}{\omega_0^2} \tag{19-52}$$

which may differ appreciably from 1 according to Eq. (19-45). The value $K_0 = 5.5$ for diamond, for example, is understandable, even though diamond is a nonconductor at frequencies below the ultraviolet. Conversely, a nonconductor can have K different from 1 only if it is conducting (absorbing) in some other frequency region; a medium in which $g(\omega) = 0$ for all frequencies necessarily has $K(\omega) = 1$ also, i.e., it is identical to vacuum, as we will see below. On the other hand, at frequencies far above the absorption frequency, K goes to 1 in all cases.

The optical constants n and k are obtained for this case by using the approximations (19-50) and (19-51) in (19-41). Except at $\omega = \omega_0$, $K_i = 0$, and so the approximations of (17-62) and (17-63) apply. From Eq. (19-50), $K_r = 0$ at $\omega = \sqrt{\omega_0^2 + \omega_p^2}$; above this frequency and below ω_0 , K_r is positive. In these regions,

$$n = \sqrt{K_r}, \quad k = 0 \tag{19-53}$$

In the intermediate region, K_r is negative, so that

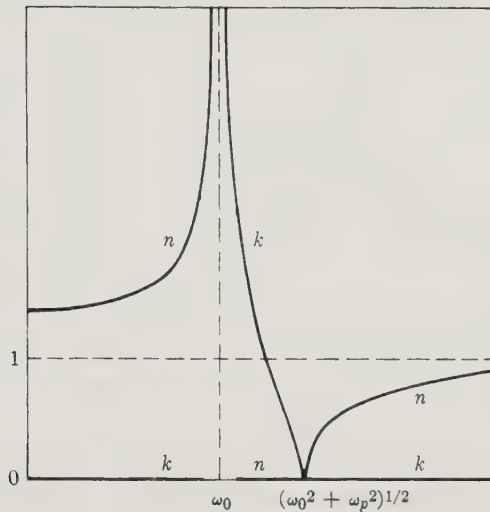
$$n = 0, \quad k = \sqrt{-K_r} \tag{19-54}$$

These functions are shown in Fig. 19-3. At frequencies below ω_0 and above $\sqrt{\omega_0^2 + \omega_p^2}$, the material is transparent ($k = 0$). Below the resonance, n is greater than 1 and increases with increasing frequency (shorter wavelength). This is the "normal dispersion" behavior of a glass prism in the visible region, which shows a greater refraction for blue than for red, and it is typical of all transparent materials. Above the resonance absorption, n increases with frequency, but it is less than 1, also characteristic of all materials in the X-ray region. Immediately above the resonance a wave is attenuated ($k > 0$), but not, however, because of absorption (except exactly at ω_0 where $K_i \neq 0$). The wave is attenuated because of perfect reflection at the surface where it enters the medium:

$$R = \frac{(n - 1)^2 + k^2}{(n + 1)^2 + k^2}$$

which for $n = 0$ gives $R = 1$. Although glass has a reflectance of only

FIGURE 19-3
Optical constants as a function of frequency, derived from the dielectric constants of Fig. 19-2



$R = 0.04$ in the visible region, in the ultraviolet, it should be highly reflecting.

To examine the detailed behavior in the immediate vicinity of ω_0 , suppose that $\gamma > 0$, but still $\gamma \ll \omega_0$. Near ω_0 (i.e., in range $\omega_0 \pm$ several times γ), the approximation

$$\omega_0^2 - \omega^2 = (\omega_0 + \omega)(\omega_0 - \omega) \cong 2\omega_0(\omega_0 - \omega)$$

is valid, and the dielectric constants Eqs. (19-40a) and (19-40b) simplify to

$$K_r - 1 = \frac{\omega_p^2}{\omega_0} \frac{\frac{1}{2}(\omega_0 - \omega)}{(\omega_0 - \omega)^2 + (\gamma/2)^2} \quad (19-55)$$

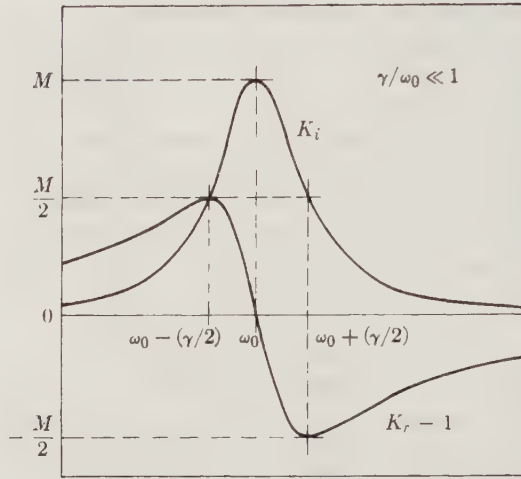
$$K_i = \frac{\omega_p^2}{\omega_0} \frac{\gamma/4}{(\omega_0 - \omega)^2 + (\gamma/2)^2} \quad (19-56)$$

The function K_i has the so-called *Lorentzian* line shape shown in Fig. 19-4, which also shows $K_r - 1$. These functions have especially simple properties: K_i is even with respect to $(\omega_0 - \omega)$, and $K_r - 1$ is odd. It is easy to show that the width of K_i at half-maximum is γ and the extremes of $K_r - 1$ occur at these same points. The integral of K_i over all positive frequencies can be carried out with the result $(\omega_p^2/\omega_0)(\pi/2)$.* The maximum value of K_i at

* This result follows directly from the approximate K_i of Eq. (19-51), but it also is obtained from the Lorentzian shape, Eq. (19-56).

FIGURE 19-4

Dielectric constants for a narrow line at ω_0 . (The origin of frequency is displaced far to the left.) For a weak line with $M \ll 1$, $n - 1$ and k are just one half of $K_r - 1$ and K_i , respectively.



resonance,

$$M = \frac{\omega_p^2}{\omega_0 \gamma} = \frac{\omega_p^2 \omega_0}{\omega_0^2 \gamma}$$

can be quite large. For a gas at atmospheric pressure, $(\omega_p/\omega_0)^2 \approx 10^{-3}$, but $(\omega_0/\gamma) \approx 10^5$, so that M may be greater than 100. This value gives a conductivity $g = \epsilon_0 \omega K_i$ at resonance that is comparable with the dc conductivity of a metal.

The shape of the n and k curves associated with the Lorentzian dielectric constants depends very much on the size of M . If M is large, $K_i \gg 1$ at ω_0 ; since $K_r = 1$ at ω_0 , the approximation

$$n \cong k \cong \sqrt{K_i/2}$$

holds at ω_0 . The shape of the curves, shown in Fig. 19-5, is rather different from the Lorentzian shape. It is a more realistic picture of the behavior near ω_0 than Fig. 19-3. In the other extreme when $M \ll 1$, $K_i \ll 1$ everywhere. Since $|K_r - 1| < K_i$ near the maximum, $|K_r - 1| \ll 1$, or $K_r \cong 1$. In this case,

$$n \cong \sqrt{K_r}, \quad k = K_i/2n$$

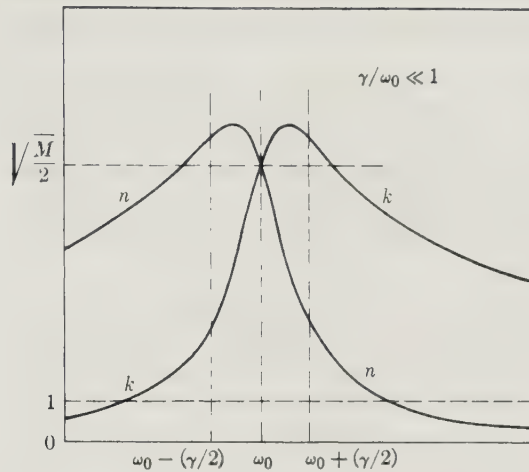
Since K_r differs from 1 by a small amount, the square root can be expanded to give

$$n - 1 \cong \frac{1}{2}(K_r - 1), \quad k \cong \frac{1}{2}K_i \quad (19-57)$$

The shape of the $n - 1$ and k curves is just the same Lorentzian shape as the $K_r - 1$ and K_i curves, respectively. This approximation may hold, for example, in a rather dilute solution or in a gas at a fairly low pressure.

FIGURE 19-5

Optical constants for a narrow, strong line, derived from the dielectric constants of Fig. 19-4 with $M = 100$. (The origin of frequency is displaced far to the left.)



Far from the resonance line, when $|\omega_0 - \omega| \gg \gamma$, γ can simply be neglected in the denominators of Eqs. (19-40a) and (19-40b), which then simplify to

$$K_r - 1 = \frac{\omega_p^2}{\omega_0^2 - \omega^2} \quad (19-58)$$

$$K_i = \frac{\omega_p^2 \gamma \omega}{(\omega_0^2 - \omega^2)^2}$$

In this approximation, $K_i \ll |K_r|$, and far on the low frequency side, $K_r > 0$. Equations (19-57) apply, so

$$\begin{aligned} n - 1 &\cong \frac{1}{2} \left(\frac{\omega_p}{\omega_0} \right)^2 \left[1 - \left(\frac{\omega}{\omega_0} \right)^2 \right]^{-1} \\ &\cong \frac{1}{2} \left(\frac{\omega_p}{\omega_0} \right)^2 \left[1 + \left(\frac{\omega}{\omega_0} \right)^2 \right] \end{aligned}$$

In terms of wavelength,

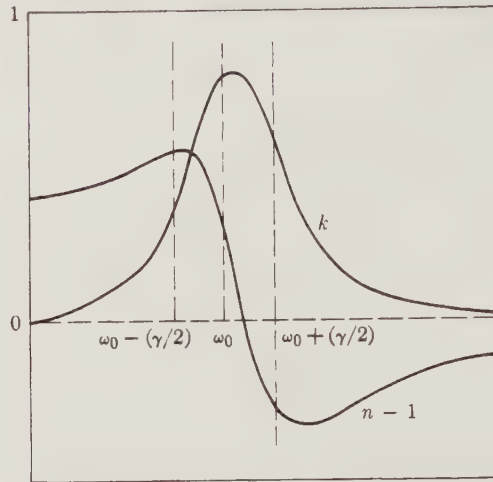
$$n - 1 \cong \frac{1}{2} \left(\frac{\lambda_0}{\lambda_p} \right)^2 \left[1 + \left(\frac{\lambda_0}{\lambda} \right)^2 \right] \quad (19-59)$$

This equation is known as the *Cauchy relation*, and it is a useful formula for fitting the refractive index of a transparent material.

If γ is not so small, none of the foregoing simple relations holds quantitatively, but the qualitative behavior of \hat{K} and \hat{n} is still similar. The case of $\omega_p/\omega_0 = 1$, $\gamma/\omega_0 = \frac{1}{2}$, is plotted in Fig. 19-6. The imaginary part always has a peak with its maximum near ω_0 . In the region of this peak, the

FIGURE 19-6

Optical constants for a wide, moderately strong absorption band at ω_0



real part always has a region of negative slope, called the *anomalous dispersion* region. As the frequency increases from zero, there are successive regions of transparency, absorption, high reflectance, and transparency, just as in the simplest delta-function approximation. In actual materials, there are always many electronic absorption peaks in the ultraviolet region, perhaps extending into the visible, and in solids they may be wide and strongly overlapping.

If the vibrating charged particles are heavy ions instead of electrons, the resonance frequency ω_0 is a few hundred times smaller: The linear restoring-force constant is about the same, since it also derives from the Coulomb force, but the ion mass is 4 or 5 orders of magnitude greater than the electron mass. The frequency ω_p is also proportionately smaller, since it too is inversely proportional to the square root of the mass. For example, in an ionic crystal,

$$\frac{\omega_p}{\omega_0} \cong 1, \quad \frac{\gamma}{\omega_0} \cong 0.1$$

The absorption peak is similar to the electronic one, but it is located in the infrared instead of the visible or the ultraviolet. The corresponding K , or n makes no contribution at higher frequency, but it does at lower frequency. Thus for rocksalt the static dielectric constant is about 6, in comparison to about 2 ($= 1.5^2$) in the visible. The latter is due to the electronic absorption in the ultraviolet; the former includes the effect of the ionic absorption in the infrared as well.

THE DRUDE FREE-ELECTRON THEORY

In some important states of matter, notably metals and plasmas, the electrons belonging to the outer atomic orbits are not localized (bound), but are *free* to contribute to dc conduction. At high frequency, their behavior is modified by inertial effects; but it is explained by our same microscopic model, if the assumed restoring force for these electrons is put equal to zero. Then, with $\omega_0 = 0$, Eq. (19-39) simplifies to*

$$\hat{K} - 1 = - \frac{\omega_p^2}{\omega(\omega + i\gamma)} \quad (19-60)$$

The real and imaginary parts are

$$K_r - 1 = - \frac{\omega_p^2}{\omega^2 + \gamma^2} \quad (19-61)$$

$$K_i = \frac{\omega_p^2 \gamma}{\omega(\omega^2 + \gamma^2)} \quad (19-62)$$

In the free-particle context,

$$\omega_p = \sqrt{\frac{Ne^2}{\epsilon_0 m}} \quad (19-28)$$

is called the *plasma frequency*,† and it is identical to that defined in Chapter 14. Now the frequency ranges in which interesting features of the dispersion occur are determined by ω_p . Some typical values are listed in Table 19-1. They span from the radio frequency to the ultraviolet regions. As noted previously, the damping constant γ is the reciprocal collision time,

$$\gamma = \frac{1}{\tau} \quad (19-24)$$

The time τ is also the time constant for the decay of a current with no driving field, as can be seen by putting $\omega_0 = 0$ and $E_m = 0$ in Eq. (19-22): $dx/dt = v$,

$$\frac{dv}{dt} + \gamma v = 0 \quad (19-63)$$

$$v = v_0 e^{-\gamma t} = v_0 e^{-t/\tau}$$

* Simply putting $\omega_0 = 0$ introduces a singularity in $\hat{K} - 1$ at $\omega = 0$. Although this behavior is not satisfactory for a frequency response function, it is usually possible to ignore the singularity. When it causes problems, a small positive ω_0 should be retained and the limit be taken as ω_0 goes to zero at some later stage in the calculation. Alternatively, the problem can be formulated in terms of the conductivity \hat{g} to avoid problems with singularities.

† The plasma frequency is usually seen expressed in gaussian units, in which $\omega_p = \sqrt{4\pi Ne^2/m}$.

TABLE 19-1

 Some Typical Particle
 Densities N and Plasma
 Frequencies ω_p for
 Electron Plasmas

	Density (m^{-3})	Plasma frequency (s^{-1})
Metal	10^{28}	10^{16}
Semiconductor (doped)	10^{24}	10^{14}
Semiconductor (pure)	10^{20}	10^{12}
Fusion experiment	10^{20}	10^{12}
Ionosphere	10^{11}	10^7
Interplanetary space	10^7	10^5

For metals at room temperature, $\gamma \approx 10^{14} \text{ s}^{-1}$, so that

$$\frac{\gamma}{\omega_p} \ll 1$$

This relation usually holds for semiconductor electronic materials and gaseous plasmas as well. Equation (19-60) assumes a simpler form when it is expressed in terms of the complex conductivity \hat{g} using Eq. (19-43) and $\gamma = 1/\tau$:

$$\hat{g} = \frac{g_0}{1 - i\omega\tau} \quad (19-64)$$

where

$$g_0 = \epsilon_0 \omega_p^2 \tau = \frac{Ne^2 \tau}{m}$$

is the dc conductivity. Nevertheless we will treat \hat{K} , since the aim is the find the optical constants n and k from it.

When γ is small, we may examine the limit as γ goes to zero, which is best done by putting $\omega_0 = 0$ in Eqs. (19-50) and (19-51), but retaining a small positive γ until the limit $\gamma \rightarrow 0$ is taken. The result is

$$K_r = 1 - \frac{\omega_p^2}{\omega^2}$$

$$K_i = \frac{\omega_p^2 \pi}{\omega} \delta(\omega) \quad (19-65)$$

where Eq. (19-51) has been used for K_i . Figures 19-2 and 19-3 still show the behavior, if the origin of ω is taken at $\omega_0 = 0$. The transparent region below ω_0 is eliminated, and the frequency where $K_r = 0$ is ω_p . Between $\omega_0 = 0$ and $\omega = \omega_p$ the material is perfectly reflecting ($n = 0$), and above ω_p it is transparent ($k = 0$). The transition between high reflectance and transparency at ω_p accounts for the well-known fact that the ionosphere reflects radio waves in the AM broadcast band ($f = \omega/2\pi \leq 1.5 \times$

10^6 s^{-1}), but is transparent to FM and TV ($f \approx 10^8 \text{ s}^{-1}$). In another frequency range, sodium metal is highly reflecting in the visible, but is transparent for ultraviolet wavelengths shorter than $\lambda_p = 2100 \text{ \AA}$ (210 nm) corresponding to its plasma frequency. Just at the plasma frequency, $n = 0$ and $k = 0$; the charges all move back and forth in phase (infinite wavelength) with no attenuation. This motion is the free plasma oscillation discussed in Chapter 14. It is an example of the longitudinal wave* that can occur when $\hat{K} = 0$. Although the no-damping approximation accounts for interesting effects near the plasma frequency, it is an oversimplification for lower frequencies, where it predicts infinite conductivity and infinite dielectric constant for dc. In particular it does not yield the Hagen-Rubens formula nor its associated skin depth.

To improve on the delta-function approximation, it is necessary to go back to Eqs. (19-61) and (19-62). We assume that the damping is small, $\gamma \ll \omega_p$, so there are three frequency ranges to consider.

1. $\omega \ll \gamma$:

$$K_r \cong \frac{-\omega_p^2}{\gamma^2} \quad (19-66)$$

$$K_i \cong \frac{\omega_p^2}{\omega\gamma} \quad (19-67)$$

The real dielectric constant K_r is a (large) negative constant; K_i goes to infinity as ω goes to zero, but

$$g = \epsilon_0 \omega K_i = \frac{\epsilon_0 \omega_p^2}{\gamma} = \frac{Ne^2 \tau}{m} = g_0$$

is constant in this range and equals the dc conductivity g_0 . Since $K_i/|K_r| = \gamma/\omega \gg 1$, Eq. (17-64) gives

$$n \cong k \cong \sqrt{K_i/2} = \frac{\omega_p}{\sqrt{2}\omega\gamma} \gg 1 \quad (19-68)$$

This case gives the Hagen-Rubens formula for the low frequency absorptance,

$$A \cong \frac{2}{k} = 2\sqrt{\frac{2\epsilon_0\omega}{g_0}}$$

* Since the wavelength is infinite and the propagation velocity zero, it is meaningless to say it is longitudinal or transverse in this case. We have calculated $\hat{K}(\kappa, \omega)$ only for $\kappa = 0$. A more sophisticated calculation including pressure effects shows the existence of longitudinal plasma waves with finite wavelength, $\kappa \neq 0$.

and the skin depth

$$\delta = \frac{c}{k\omega} \cong \sqrt{\frac{2}{\mu_0\omega g_0}}$$

2. $\gamma \ll \omega \ll \omega_p$:

$$K_r \cong -\frac{\omega_p^2}{\omega^2} \quad (19-69)$$

$$K_i \cong \frac{\omega_p^2\gamma}{\omega^3} \quad (19-70)$$

Since $K_i/|K_r| = \gamma/\omega \ll 1$,

$$k \cong \sqrt{-K_r} = \frac{\omega_p}{\omega} \quad (19-71)$$

$$n = \frac{K_i}{2k} \cong \frac{\gamma\omega_p}{2\omega^2}$$

In this region $k/n = 2\omega/\gamma \gg 1$, so that the absorptance Eq. (18-59) is

$$A \cong \frac{4n}{k^2} \cong \frac{2\gamma}{\omega_p} = \frac{2}{\omega_p\tau}$$

(sometimes called the Mott-Zener formula), and the skin depth is

$$\delta = \frac{c}{k\omega} = \frac{c}{\omega_p} = \lambda_p/2\pi$$

where λ_p is the vacuum wavelength corresponding to the plasma frequency. For metals, these results apply to the infrared region. Since $\omega_p\tau$ is large, $A = 1 - R$ is only a few percent; the skin depth is very small and independent of frequency.

3. $\omega_p \ll \omega$:

$$K_r \cong 1 \quad (19-72)$$

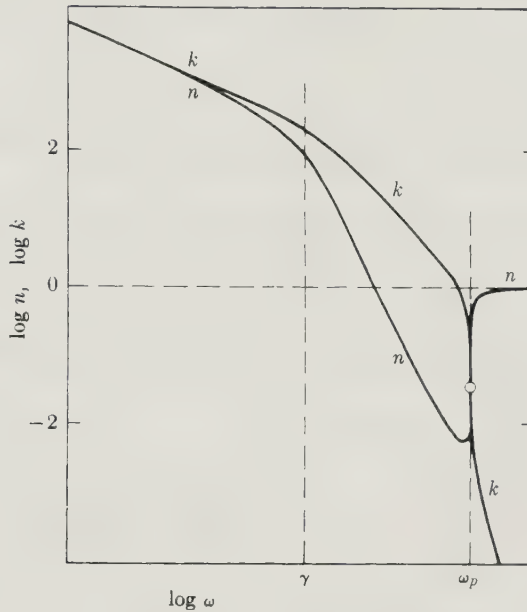
$$K_i \cong \frac{\omega_p^2\gamma}{\omega^3} \quad (19-73)$$

$$n \cong \sqrt{K_r} \cong 1, \quad k = \frac{K_i}{2n} \cong \frac{\omega_p^2\gamma}{2\omega^3} \quad (19-74)$$

Here the material is nearly transparent. In metals, however, except for the alkali metals, the onset of this transparency is obscured by the resonance absorption of bound electrons belonging to inner orbits. This additional absorption, discussed in the last section, also increases the values of K_i and n for most metals in the visible region, reducing the reflectance and accounting for the characteristic colors of copper and gold.

FIGURE 19-7

Log-log plot of the optical constants of free charges versus frequency, with $\gamma/\omega_p = 0.004$. Power-law approximations are valid over most of the frequency range.



The free-electron results for n and k are plotted on a log-log scale in Fig. 19-7. The parameter values chosen are $\omega_p = 9 \times 10^{15} \text{ s}^{-1}$, $\gamma = 3.6 \times 10^{13} \text{ s}^{-1}$; these are appropriate for sodium metal at room temperature, for which the free-electron theory gives good agreement with experiment. The three straight-line regions of the figure are plots of Eqs. (19-68), (19-71), and (19-74). The intermediate points at $\omega = \gamma$ and $\omega = \omega_p$ are easily calculated from Eqs. (19-61) and (19-62), since $K_i = -K_r$ and $K_i \gg K_r$, respectively. The region well above γ agrees with the delta-function approximation ($\gamma = 0$), with some rounding of the kinks at ω_p . Examples with $\gamma \gtrsim \omega_p$ are less common; their treatment is left as a problem.

***EXAMPLE 19-1**
Use of the Dispersion
Relation

The dispersion relation can be used to examine the low-frequency approximation of Case 1. This procedure will illustrate some of the problems encountered in the application of these integral relationships. We use the low frequency approximation to K_i given by Eq. (19-67) and attempt to derive Eq. (19-66) for K_r by means of the dispersion relation.

Solution: Using Eq. (19-67) in the Kramers-Kronig dispersion relation,

Eq. (19-21a), to get the real part of $\hat{K} - 1$, we have

$$K_r - 1 = \frac{2 \omega_p^2}{\pi \gamma} \int_0^\infty \frac{d\omega'}{\omega'^2 - \omega^2} \quad (19-75)$$

The upper limit of integration is taken to be A , which subsequently will become infinite. The principal value is then

$$P \int_0^\infty \frac{d\omega'}{\omega'^2 - \omega^2} = \lim_{\substack{A \rightarrow \infty \\ \delta \rightarrow 0}} \left\{ - \int_0^{\omega-\delta} \frac{d\omega'}{\omega^2 - \omega'^2} + \int_{\omega+\delta}^A \frac{d\omega'}{\omega'^2 - \omega^2} \right\} \quad (19-76)$$

Both integrals are well defined and can be evaluated without difficulty to get

$$P \int_0^\infty \frac{d\omega'}{\omega'^2 - \omega^2} = \lim_{\substack{A \rightarrow \infty \\ \delta \rightarrow 0}} \left\{ - \frac{1}{2\omega} \ln \left(\frac{\omega + \omega'}{\omega - \omega'} \right) \Big|_0^{\omega-\delta} + \frac{1}{2\omega} \ln \left(\frac{\omega' - \omega}{\omega' + \omega} \right) \Big|_{\omega+\delta}^A \right\} \quad (19-77)$$

At $\omega' = 0$, the logarithm in the first term is $\ln(\omega/\omega) = 0$. At $\omega' = A$, the logarithm in the second term is $\ln(1 + \omega/A)/(1 - \omega/A)$, which goes to zero in the limit as A goes to infinity. The two remaining terms give

$$P \int_0^\infty \frac{d\omega'}{\omega'^2 - \omega^2} = \lim_{\delta \rightarrow 0} \frac{1}{2\omega} \left(- \ln \frac{\delta}{2\omega + \delta} - \ln \frac{2\omega - \delta}{\delta} \right) = 0 \quad (19-78)$$

which vanishes in the limit as δ goes to zero. Thus,

$$P \int_0^\infty \frac{d\omega'}{\omega'^2 - \omega^2} = 0 \quad (19-79)$$

and

$$K_r = 1 \quad (19-80)$$

This result contains some instructive features. First, the singularity in the integrand does not make the integral infinite. (In fact, it is zero, but that is a peculiarity of this special case.) It is clear that the singularity will cause no trouble so long as K_i is continuous at ω . Secondly the result Eq. (19-80) for K_r is a constant, but the constant disagrees with the constant $-(\omega_p/\gamma)^2$ required by Eq. (19-66). This discrepancy emphasizes that $K_i(\omega)$ must be known over the entire frequency range from 0 to ∞ in order to get the correct result from Eq. (19-21a). In this example Eq. (19-67) is correct at frequencies well below γ , but wrong at higher frequencies. The true value of K_r at low frequency comes from this higher frequency behavior of K_i . The integral always gives the correct value of K_r when ω goes to infinity, namely, $K_r = 1$, as can be

seen directly in Eq. (19-21a). Thus if K_r is constant it must give the constant 1. In using the integral with experimental data, it is necessary to make some reasonable extrapolation above and below the frequency range in which the measurements were made, even to find K_r only within this range. A more obvious cause of error would be the omission of an unknown absorption peak above the range of measurement.

It is worth pointing out that the dispersion relations, Eqs. (19-21), can be modified to make them more tractable to numerical calculation. Consider Eq. (19-21a) with $\bar{R}(\omega) \equiv \hat{K} - 1$. The integral in Eq. (19-21a) can be transformed in various ways. Since the integral in Eq. (19-79) is zero, any constant multiple of it can be subtracted from Eq. (19-21a) without affecting the result. Subtracting $(2/\pi)\omega K_i(\omega)$ times it gives

$$K_r(\omega) - 1 = \frac{2}{\pi} \int_0^\infty \frac{[\omega' K_i(\omega') - \omega K_i(\omega)] d\omega'}{\omega'^2 - \omega^2}$$

This form is useful for numerical integrations because it has a nonsingular integrand, the numerator vanishing at the same point as the denominator. An integration by parts in Eq. (19-21a) gives

$$K_r(\omega) - 1 = \frac{1}{\pi} \int_0^\infty \frac{dK_i(\omega')}{d\omega'} \ln \frac{1}{|\omega'^2 - \omega^2|} d\omega'$$

which provides a qualitative idea of the behavior of K_r to be expected. The integrand is heavily weighted for frequencies near ω by the second factor, so that the magnitude of K_r is strongly affected by the *slope* of K_i at nearby frequencies.

All the results of this section also apply when the charged particles are heavy ions instead of electrons. Since the medium has been assumed to be electrically neutral, positive ions must always be present with an average number density N equal to that of the electrons. In metals the positive ions are not freely mobile, of course, but in gaseous plasmas they are, and their motion is often important. The plasma frequency ω_p is inversely proportional to the square root of the particle mass m , and so the ion plasma frequency is about two orders of magnitude smaller than the electron plasma frequency given in Table 19-1, for example.

*19-6 DIELECTRIC RELAXATION

Electrical resonance absorption does not usually occur in materials at frequencies lower than the heavy-ion peaks in the infrared (though it does in manmade structures, of course). There is, however, another kind of absorption mechanism, known as *dielectric loss*, which can occur at lower frequencies (but not at higher). It is frequently important as a loss mechanism at microwave frequencies and below, and its concomitant dispersion of the real dielectric constant explains, for example, the

difference between the static dielectric constant of water, 81, and the optical value of about 1.8 ($=1.33^2$). The effect is an extension to oscillating fields of the second type of static polarization discussed in Chapter 5—namely, the orientational polarization of permanent dipoles. This case can also be based on the model described by Eq. (19-22a), though with some strain on the physical interpretation of the quantities involved. It arises when the damping and restoring forces are both important, but inertial effects (acceleration) can be neglected. In this case, the equation of motion becomes

$$G \frac{dx}{dt} + Cx = eE_m \quad (19-81)$$

or

$$\gamma \frac{dx}{dt} + \omega_0^2 x = eE_m/m$$

The appropriate solution is obtained from Eq. (19-39) by neglecting ω^2 ,

$$\hat{K} - 1 = \frac{\omega_p^2}{\omega_0^2 - i\gamma\omega}$$

The parameters in this solution all contain m ; however, m does not really enter because the inertial term has been neglected in obtaining the solution. It is better to rewrite the solution as

$$\hat{K} - 1 = \frac{Ne^2/\epsilon_0}{C - iG\omega}$$

where m is eliminated entirely. From Eq. (19-81) with $E_m = 0$, the return to equilibrium is described by

$$x = x_0 e^{-Ct/G}$$

Therefore, we write

$$\frac{C}{G} = \frac{1}{\tau} \quad (19-82)$$

where τ in this context is called the *relaxation time*.* In terms of it,

$$\hat{K} - 1 = \frac{K_0 - 1}{1 - i\omega\tau} \quad (19-83)$$

where

$$K_0 - 1 = \frac{Ne^2}{\epsilon_0 C} \quad (19-84)$$

* Note that in this process $1/\tau \neq \gamma$. We are using τ consistently to mean the time constant for return to equilibrium, regardless of its relation to γ or the determining physical quantities.

Clearly K_0 is the static value of K , for $\omega = 0$; it is the same as the static value for resonance absorption, except that C has a different meaning. This meaning can be obtained from the calculation in Chapter 5:

$$K_0 - 1 = \frac{\chi_0}{\epsilon_0} = \frac{N p_0^2}{\epsilon_0 3k_B T}$$

where k_B is Boltzmann's constant. If we let the permanent dipole moment be

$$p_0 = ea$$

where a is the effective separation of charges $\pm e$, then

$$K_0 - 1 = \frac{Ne^2 a^2}{3\epsilon_0 k_B T} \quad (19-85)$$

Numerically this can be, at the maximum, of the order of 10^2 at room temperature. Comparing Eq. (19-85) with Eq. (19-84) shows that

$$C = \frac{3k_B T}{a^2} \quad (19-86)$$

The restoring force is therefore due to the thermal energy $k_B T$ rather than to a mechanical elastic potential energy.*

What is new in this discussion is the relaxation time τ , which determines the behavior for oscillating fields. Its value can be estimated most easily for certain *solid* polar materials in which the dipoles consist of mobile pairs of positive and negative ions of charge e separated by a distance a . The ions have certain fixed equilibrium positions in the solid, but they can make jumps to other neighboring allowed positions. By means of such jumps the dipole can rotate into alignment with an applied field, or return to a random orientation when the field is removed. The time τ is therefore the mean time for a jump and is given by

$$\frac{1}{\tau} = \frac{1}{\tau_0} e^{-\Delta U/k_B T} \quad (\text{solid}) \quad (19-87)$$

The factor $1/\tau_0$ is approximately the vibrational frequency of the ion around its equilibrium position, and $\exp(-\Delta U/k_B T)$ is the statistical Boltzmann factor. A jump is attempted with each vibration, but only a fraction succeeds, depending on the energy ΔU required in order to squeeze through the barrier to the neighboring equilibrium position. The frequency $1/\tau_0 = \omega_0$ is the ionic vibrational frequency considered previously, $\omega_0 \approx 10^{13} \text{ s}^{-1}$.

* The restoring force is a generalized "thermodynamic force," not a mechanical one. What relaxes after an applied force is removed (with the relaxation time τ) is not the energy but rather the entropy.

The barrier energy ΔU must be appreciable in order for this model to hold at all, since the very existence of equilibrium positions between which jumps occur depends on the barriers. Thus,

$$\frac{1}{\tau} \lesssim 10^{11} \text{ s}^{-1}$$

and $1/\tau$ will be very much smaller than this at temperatures much below the melting temperature. For liquids the calculation of τ is more difficult. The *Debye relaxation time* is given by

$$\frac{1}{\tau} = \frac{k_B T}{4\pi\eta R_0^3} \quad (\text{liquid}) \quad (19-88)$$

where η is the viscosity (not electrical resistivity) and R_0 is the radius of the molecule. This result agrees reasonably well with experimental values of η and predicts a frequency in the microwave region for most polar liquids at room temperature, since the dipoles in a liquid are more freely rotating than in a solid. The viscosity is temperature dependent, but still Eq. (19-88) is less strongly temperature dependent than Eq. (19-87).

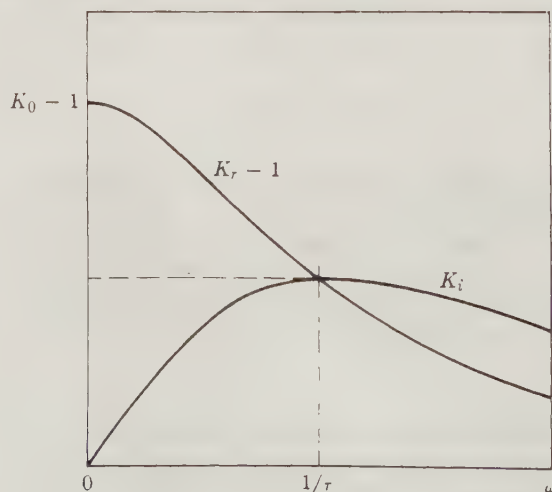
The frequency dependence of Eq. (19-83) is seen by separating it into real and imaginary parts, giving the *Debye equations*:

$$K_r - 1 = \frac{K_0 - 1}{1 + (\omega\tau)^2} \quad (19-89)$$

$$K_i = \frac{(K_0 - 1)\omega\tau}{1 + (\omega\tau)^2} \quad (19-90)$$

These are plotted in Fig. 19-8, which shows an absorption peak in K_i with

FIGURE 19-8
Dielectric constants as a
function of frequency for
Debye relaxation



maximum at $\omega = 1/\tau$ and height $\frac{1}{2}(K_0 - 1)$, and K_r falling from K_0 to 1 with the halfway point at $\omega = 1/\tau$. The dispersion of K_r has a shape characteristically different from that for resonance absorption. The shape of the curves is quite similar to that for a strongly overdamped oscillation. For frequencies much below or much above $1/\tau$, $K_i \ll K_r$ ($K_r > 0$), so that $n = \sqrt{K_r}$, $k = K_i/2n$. As an example, these results can be applied to fresh water in which the low frequency or dc conductivity can be neglected; the reflectance is given by Eq. (18-57). For electrical measurements on laboratory-sized samples, the results are usually expressed in terms of the *loss angle* θ :

$$\tan \theta = \frac{K_i}{K_r} = \frac{(K_0 - 1)\omega\tau}{K_0 + (\omega\tau)^2} \quad (19-91)$$

The physical reason for this kind of frequency dependence is that above $1/\tau$ the relaxation cannot follow the applied field.

*19-7 OSCILLATING FIELDS IN DISPERSIVE MEDIA

We have now shown that for monochromatic fields the real permittivity and real conductivity of a linear medium generally depend on the frequency.* Furthermore, in the complex representation, this frequency dependence can be described either by a permittivity or by a conductivity that is a complex function of the frequency. For many materials there are, however, large ranges over which the frequency dependence is small enough to be neglected in most applications. For example, in an insulating material, by which we mean a material with (nearly) zero real conductivity at zero frequency, the real permittivity may be constant, though different from ϵ_0 , up to $\omega \approx 10^{14} \text{ s}^{-1}$, with the real conductivity very nearly zero over a similar range. In metals, the real conductivity is very nearly the static conductivity, g_0 , for frequencies up to $\approx 10^{11} \text{ s}^{-1}$.

In the physical world electromagnetic fields come in wave trains that are, in some cases, sufficiently long that they may be treated as monochromatic to an adequate approximation. Often, however, the wave train is so short that the monochromatic approximation is not adequate. One way of dealing with short pulses is to write the electric field, for example, as a superposition of monochromatic fields. Solutions can be built up this way because Maxwell's equations are linear as long as we consider only linear materials.

* A similar treatment of the magnetic susceptibility would produce yet another frequency-dependent factor, but fortunately there is seldom a significant overlap in the frequency ranges where ϵ and μ are strongly frequency dependent. Consequently, we will continue to take $\mu = \mu_0$.

We now present two examples of time-dependent fields in insulators, in which the frequency dependence of the permittivity plays a significant role.

EXAMPLE 19-2
Propagation of an
Electromagnetic Pulse:
The Group Velocity

Consider a planar (independent of x and y) pulse of electromagnetic radiation moving in the z direction for which, at $t = 0$, the electric field is a specified function $E(z, 0)$. The problem is to find the electric field, $E(z, t)$, at subsequent times.

Solution: The electric field must satisfy

$$\frac{\partial^2 E(z, t)}{\partial z^2} - \mu_0 \frac{\partial^2 D(z, t)}{\partial t^2} = 0 \quad (19-92)$$

If the medium is linear, then from Section 19-1 we know that the monochromatic solutions to Eq. (19-92) satisfy

$$\frac{\partial^2 \tilde{E}(z, \omega)}{\partial z^2} + \mu_0 \epsilon(\omega) \omega^2 \tilde{E}(z, \omega) = 0 \quad (19-93)$$

and consequently, the monochromatic plane wave solutions to this equation have the form

$$E(z, t) = \tilde{E}_0 e^{i(\kappa z - \omega t)} \quad (19-94)$$

with

$$\kappa^2 - \mu_0 \epsilon(\omega) \omega^2 = 0 \quad (19-95)$$

We now construct a solution to Eq. (19-93) by superposing solutions of the form given by Eq. (19-94). We write this superposition as

$$E(z, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{E}_0(\kappa) e^{i(\kappa z - \omega t)} d\kappa \quad (19-96)$$

The integration in Eq. (19-96) is over κ but not also over ω because Eq. (19-95) associates a value of ω with each value of κ . The integration could equally well have been over ω but we prefer to use κ as the integration variable. Waves travelling in both the plus z and minus z directions are included. At $t = 0$, Eq. (19-96) is

$$E(z, 0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{E}_0(\kappa) e^{i\kappa z} d\kappa \quad (19-97)$$

so that $\tilde{E}_0(\kappa)$ is just the spatial Fourier transform of the pulse at $t = 0$. If the range of κ required to represent the pulse adequately is not too large, ω can be expanded as

$$\omega = \omega_0 + (\kappa - \kappa_0) v_g(\kappa_0) + \dots \quad (19-98)$$

with $v_g(\kappa_0) = (d\omega/d\kappa)|_{\kappa_0}$, and only the first two terms retained. Using

this approximation in Eq. (19-96), we obtain

$$\begin{aligned}
 E(z, t) &= e^{-i(\omega_0 - \kappa_0 v_g)t} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{E}(\kappa) e^{i\kappa(z - v_g t)} d\kappa \\
 &= E(z - v_g t, 0) e^{-i(\omega_0 - \kappa_0 v_g)t}
 \end{aligned} \tag{19-99}$$

In other words, if this approximation to the dependence of ω on κ is adequate, the pulse travels with unchanged shape with a velocity $v_g(\kappa_0)$ and a time-dependent phase shift that is dependent on wave number. The velocity v_g is called the *group velocity* and is the velocity with which energy is transmitted in the pulse.

The effects of dispersion are (1) $v_g(\kappa_0)$ is not ω_0/κ_0 as it would be in a nondispersive medium, and (2) there is a phase shift proportional to time (this would be zero in a nondispersive medium). The first of these effects means that two pulses, one represented by a superposition of plane waves with propagation vectors in a narrow band around κ_1 and the other with propagation vectors close to κ_2 , will move through the medium with different velocities. Thus, if they both start together at $z = 0, t = 0$, they will separate in space (or time) as they move through the medium.

In many cases the approximation of keeping only the first two terms in Eq. (19-98) is not adequate. Then the shape of the pulse will not be preserved; usually it spreads out but there are cases where the pulse becomes narrower as it propagates. In particular, the pulse has precursors that travel faster than the main pulse.*

Other common effects of dispersion arise from the dependence of the index of refraction on frequency (or on wavelength). This variation is shown for crown glass in Fig. 19-9. As a result of this frequency dependence, different colored light travels at slightly different velocities in the material. This phenomenon is responsible for the formation of spectra by prisms (“dispersion” of the light) and chromatic aberration in lenses, among other readily observable effects.

EXAMPLE 19-3
Polarization Response to a
Step-Function Electric
Field

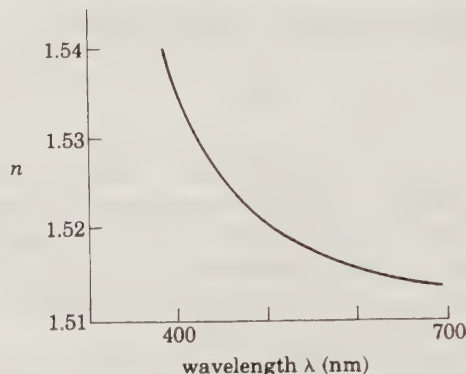
The polarization $P(t)$ due to an electric field $E(t)$ can be written in accordance with Eq. (19-1b) as

$$P(t) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} R(t'') E(t - t'') dt'' \tag{19-100}$$

* The propagation of electromagnetic radiation pulses in dispersive media is discussed in Julius A. Stratton, *Electromagnetic Theory* (New York: McGraw-Hill, 1941), pp. 333–340, and in J. David Jackson, *Classical Electrodynamics*, 2nd ed. (New York: John Wiley and Sons, 1975), pp. 313–326.

FIGURE 19-9

Index of refraction for crown glass



We want to find the response function $R(t'')$ for a medium characterized by a single sharp absorption peak at ω_0 and, from it, determine the response to an electric field that is zero for t less than zero and E_0 for t greater than zero.

Solution: We use Eqs. (19-50) and (19-51) and use χ_0 for $\chi(\omega = 0) = \epsilon_0 \omega_p^2 / \omega_0^2$ so that

$$\tilde{\chi}(\omega) = \chi_r(\omega) + i\chi_i(\omega) = \frac{\chi_0}{1 - \omega^2/\omega_0^2} + i\frac{\pi}{2}\chi_0\frac{\omega_0^2}{\omega}\delta(\omega - \omega_0) \quad (19-101)$$

The response function $R(t)$ can be obtained by using Eq. (19-3)—that is, the inverse Fourier transform of $\tilde{\chi}(\omega)$. Since $R(t)$ is real the transform may be expressed as an integral over positive frequencies only:

$$R(t'') = \sqrt{\frac{2}{\pi}} \int_0^\infty [\chi_r(\omega) \cos \omega t'' + \chi_i(\omega) \sin \omega t''] d\omega \quad (19-102)$$

Both terms integrate to give the same result. For example,

$$\sqrt{\frac{2}{\pi}} \int_0^\infty \chi_i(\omega) \sin \omega t'' d\omega = \sqrt{\frac{\pi}{2}} \chi_0 \omega_0 \sin \omega_0 t'' \quad (19-103)$$

Thus, Eq. (19-100) becomes

$$P(t) = \chi_0 \omega_0 \int_0^\infty E(t - t'') \sin \omega_0 t'' dt'' \quad (19-104)$$

It is convenient to change the integration variable to $t' = t - t''$ with the corresponding limits of integration $-\infty$ to t . The electric field is zero for

$t' < 0$ and is E_0 thereafter. The integral can be carried out, with the result that

$$P(t) = \chi_0 E_0 (1 - \cos \omega_0 t) \quad (t > 0) \quad (19-105)$$

Thus, after the sudden application of the electric field E_0 the medium “rings” at the resonant frequency ω_0 . If a damping term had been included, by using Eqs. (19-40a) and (19-40b) in place of Eq. (19-101), the oscillation of $P(t)$ would eventually die out, and the polarization would be just the static polarization, $\chi_0 E_0$. This calculation emphasizes the fact that there is no simple proportionality between $P(t)$ and $E(t)$ and that their time-dependent ratio is not a characteristic material property. The material parameters are properly considered as ratios between monochromatic responses and monochromatic excitations—that is, as ratios of the Fourier transforms of the time-dependent fields.

19-8 SUMMARY

A medium is **linear** if the output $O(t)$ and input $I(t)$ are related by

$$O(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t R(t-t') I(t') dt'$$

and the **response function** $R(t)$ *does not depend on* $I(t)$. The response is **causal** since the output depends on what has happened in the past, but not in the future.

The **complex dielectric constant** of a material is calculated as a function of the frequency of the electric field by treating the electrons and ions either as classical damped harmonic oscillators or free particles. The result is

$$\hat{K} - 1 = \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i\gamma\omega}$$

where ω_0 is the natural frequency, γ the damping frequency, and

$$\omega_p = \sqrt{\frac{Ne^2}{\epsilon_0 m}}$$

(the plasma frequency for free particles). On the basis of this analysis, the typical kinds of frequency dependence of the real dielectric constant and conductivity can be catalogued, depending on whether the inertial, damping, or restoring forces can be neglected. If local-field effects can be ignored, the dielectric responses of different groups of particles are additive. The frequency dependence of the optical constants n and k depends on that of the dielectric constants and also on the relative magnitudes of the real and imaginary parts. The real and imaginary parts are not independent of each other, but are related by the **Kramers-Kronig relations**.

- **Resonance absorption** occurs when inertial and restoring forces dominate. For small damping the dielectric constant has the simple Lorentzian shape. If K_i is small, the optical constants have the same shape. In any case, "anomalous dispersion" occurs in the region of the absorption peak. The peaks are in the visible or ultraviolet for electrons and in the infrared for ions.

- The **Drude free-electron theory** results by putting the restoring force (ω_0) equal to zero. The low-frequency skin depth and absorptance (Hagen-Rubens formula) with the dc conductivity result when $\omega \ll \gamma$. For $\omega \gg \gamma$, ω_p , the free particles make a very small contribution to conduction or absorption. If $\gamma \ll \omega_p$, there is an intermediate frequency region (in the infrared for metals) of low absorptance (high reflectance),

$$A = 2/\omega_p \tau$$

where τ is the collision frequency, and small skin depth,

$$\delta = \lambda_p/2\pi$$

where λ_p is the plasma wavelength. The plasma frequency is in the ultraviolet for metals, and at much lower frequencies in other common electron plasmas.

- Inertial effects are negligible in **dielectric relaxation**, which has an "over damped" frequency dependence, with no "normal dispersion" region. The mechanism is that of the orientational polarizability of permanent dipoles.

- The **Kramers-Kronig relation**

$$K_r(\omega) - 1 = \frac{2}{\pi} \int_0^\infty \frac{\omega' K_i(\omega') d\omega'}{\omega'^2 - \omega^2}$$

between the real and imaginary parts of the dielectric response function depends only on linearity of the medium (and on causality). It or variations of it have useful application in all linear systems.

PROBLEMS

19-1 The density and refractive index of liquid benzene at 20°C are 0.879 g/cm³ and 1.50 (for $\lambda = 589$ nm), respectively. From the Clausius-Mossotti equation, (a) compute the refractive index of benzene vapor at 20°C, where the vapor pressure is 0.1 atmosphere; (b) also compute it at the boiling point, 80°C.

19-2 Prove that the width of the Lorentzian curve Eq. (19-56) at half maximum is γ and that the area under the curve is

$$\frac{\pi \omega_p^2}{2 \omega_0}$$

19-3 The refractive index of diamond at $\lambda = 5893 \text{ \AA}$ is 2.417; take the static dielectric constant to be 5.50. Fit these data to the simple model that has a single δ -function absorption at λ_0 , in order to determine λ_0 .

19-4 Use the Cauchy formula to estimate the refractive index of hydrogen gas at standard conditions for wavelengths of 4000 and 7000 Å. Assume that $\lambda_0 = 1216$ Å (the Lyman- α line).

19-5 Experimental absorption peaks sometimes have a shape that is closer to a Gaussian curve than a Lorentzian. Plot a Gaussian and a Lorentzian with the same peak height and width-at-half-maximum on the same graph, so as to show the difference between them.

19-6 The dielectric function for bound ions oscillating with negligible damping can be written

$$K(\omega) = K_\infty + \frac{\omega_p^2}{\omega_T^2 - \omega^2}$$

where ω_T is the resonance frequency for the long wavelength transverse vibrations of the ions, and K_∞ approximates the contribution of electronic motions which resonate at frequencies much higher than ω_T . Note that $K(\omega) \rightarrow \infty$ for $\omega = \omega_T$. If $K(\omega) = 0$ for $\omega = \omega_L$, so that long wavelength longitudinal oscillations of frequency ω_L can occur, show that

$$\frac{\omega_L^2}{\omega_T^2} = \frac{K_0}{K_\infty}$$

where $K_0 = K(0)$ is the dc dielectric constant. This is called the *Lyddane-Sachs-Teller relation*.

19-7 Suppose a dilute solution consists of N atomic oscillators per unit volume dissolved in a transparent medium of refractive index n_∞ . Assuming that $K_0 \ll 1$ at the resonance frequency and $n = n_\infty$, (a) find k and the absorption coefficient α . Neglect local field effects. (b) Show that

$$N = \frac{\epsilon_0 m c}{e^2} n_\infty \gamma \alpha$$

Such a relation (known as Smakula's equation or Chako's equation) is often used to find N from the measured optical absorption height α and width γ .

19-8 Consider a medium containing free particles with collision time τ and dc conductivity g_0 . (a) Calculate the real and imaginary parts of the conductivity for the frequency $\omega = 1/\tau$. (b) What is the real dielectric constant K ?

19-9 For a free-electron plasma with $\gamma/\omega_p = 10^{-2}$, calculate approximate values of n and k at $\omega = \omega_p$.

19-10 Suppose that in the dielectric function of a free-electron plasma with negligible damping, the contribution from bound inner-orbit electrons which resonate at a higher frequency can be approximated by K_∞ :

$$K(\omega) = K_\infty - \frac{\omega_p^2}{\omega^2}$$

(a) Find a formula for the frequency of longitudinal oscillations. (b) In silver, the value calculated from the free (valence) electron density is $\omega_p = 13.8 \times 10^{15} \text{ s}^{-1}$, while longitudinal plasma oscillations are observed at $\omega = 5.8 \times 10^{15} \text{ s}^{-1}$. What is K_∞ ?

19-11 In a free-electron plasma, the longitudinal plasma waves occur at the frequency $\omega = \sqrt{Ne^2/\epsilon_0 m}$, where m is the electron mass. If the positive ions of mass M are also freely mobile, show that the longitudinal waves occur at $\omega = Ne^2/\epsilon_0 \mu$, where $\mu = mM/(M + m)$ is the reduced mass, assuming that the contributions to K are additive.

19-12 Discuss the dielectric behavior of free particles for the case $\gamma \gg \omega_p$. That is, find approximate expressions for \hat{K} and \hat{n} in the various frequency regions. Does the Hagen-Rubens relation hold, and if so in what frequency range?

19-13 The dielectric loss angle of a polar dielectric has a maximum as a function of frequency. Calculate the frequency at which the maximum occurs, and find K , and K_i at this frequency if $K_0 \gg 1$.

19-14 Use the Kramers-Kronig relations to prove the following formulas (called *sum rules*) for the real conductivity g and susceptibility χ :

$$\int_0^\infty \chi(\omega') d\omega' = \frac{\pi}{2} \lim_{\omega \rightarrow \infty} g(\omega) = 0$$

$$\int_0^\infty g(\omega') d\omega' = -\frac{\pi}{2} \lim_{\omega \rightarrow \infty} \omega^2 \chi(\omega) = \frac{\pi Ne^2}{2m}$$

19-15 Suppose a medium is characterized by a single absorption peak $\chi_i(\omega') = \frac{1}{2}\pi\omega_0\chi_0 \delta(\omega' - \omega_0)$. Find the response $P(t)$ to a pulsed E -field,

$$E(t) = \frac{E_0}{\omega_0} \delta(t)$$

19-16 By considering $J(t)$ instead of $P(t)$, show that the complex conductivity obeys the dispersion relations

$$g_r(\omega) = \frac{2}{\pi} \int_0^\infty \frac{\omega' g_i(\omega') d\omega'}{\omega'^2 - \omega^2}, \quad g_i(\omega) = -\frac{2}{\pi} \omega \int_0^\infty \frac{g_r(\omega') d\omega'}{\omega'^2 - \omega^2}$$

These hold for a conductor as well as an insulator, since $\hat{g}(\omega)$ has no singularity at $\omega = 0$. (a) Using the relation $\hat{g} = -i\omega\hat{\chi}$, compare these dispersion relations with Eqs. (19-21) where $\hat{R}(\omega) \equiv \hat{\chi}(\omega)$. (b) Derive the "sum rules"

$$\int_0^\infty \chi(\omega') d\omega' = -\frac{\pi}{2} g(0), \quad \int_0^\infty g(\omega') d\omega' = \frac{\pi Ne^2}{2m}$$

Note that they are consistent with those derived in Problem 19-14 for an insulator [$g(0) = 0$].

The Emission of Radiation

It is convenient to start the study of the generation of electromagnetic waves from the vector potentials satisfying the inhomogeneous wave equation with sources. In this chapter we consider several idealized radiation sources as well as more complicated systems. Approximations are made that limit the validity of the solutions to fields produced by slowly moving (nonrelativistic) charges—that is, those with velocity v small compared to the velocity of light, $v \ll c$. Nevertheless the results are applicable both to the emission of radio waves from antennas and of light from atoms. We begin with the simplest antenna example, and then develop a more general procedure.

RADIATION FROM AN OSCILLATING DIPOLE

A simple example of radiation from a prescribed, time-dependent charge-current distribution is provided by calculation of the radiation from an oscillating electric dipole. The dipole will be assumed to consist of spheres located at $z = \pm l/2$ connected by a wire of negligible capacitance, as shown in Fig. 20-1. The charge on the upper sphere is q , and that on the lower sphere is $-q$. Conservation of charge requires that the current in the connecting wire be given by

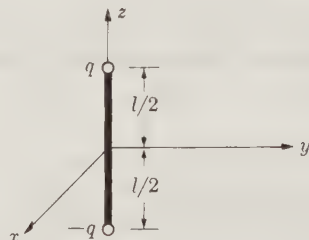
$$I = +\dot{q} \quad (20-1)$$

where I is positive in the plus z direction. It must be noted that the condition of negligible capacitance of the wire and its concomitant uniform current can be satisfied only if the length l of the dipole is small compared with the wavelength of the radiation (see the discussion at the beginning of Chapter 13).

The vector potential due to the current distribution specified by Eq.

FIGURE 20-1

An oscillating electric dipole



(20-1) is, in vacuum, from Eq. (16-84),

$$A_z(\mathbf{r}, t) = \frac{\mu_0}{4\pi} \int_{-l/2}^{l/2} \frac{I(z', t - |\mathbf{r} - z'\mathbf{k}|/c) dz'}{|\mathbf{r} - z'\mathbf{k}|} \quad (20-2)$$

This rather cumbersome expression can be quickly simplified if we examine the quantity $|\mathbf{r} - z'\mathbf{k}|$. It is clear that

$$|\mathbf{r} - z'\mathbf{k}| = (r^2 - 2z'\mathbf{k} \cdot \mathbf{r} + z'^2)^{1/2} \quad (20-3)$$

If l is small compared with r , that is, if we consider the field only at large distances from the dipole, the right side of Eq. (20-2) can be expanded in the form

$$|\mathbf{r} - z'\mathbf{k}| = r - z' \cos \theta \quad (20-4)$$

where θ is the angle between \mathbf{r} and the z -axis. The quantity in Eq. (20-4) is involved twice in the expression for \mathbf{A} . In the denominator, $z' \cos \theta$ can simply be neglected if r is large enough. In the retardation term, however, $z' \cos \theta$ can be neglected only if $z' \cos \theta/c$ is negligible compared with the time during which the current changes significantly, for example, compared with the period for harmonically varying currents. Since $z' \cos \theta \approx l/2$, this means that $z' \cos \theta/c$ can be neglected in the retardation term only if

$$\frac{l}{2} \ll cT = \lambda \quad (20-5)$$

Thus, if the dipole is small compared with one wavelength, and the observation point is far, compared with l , from the dipole, then \mathbf{A} is given by

$$A_z(\mathbf{r}, t) = \frac{\mu_0 l}{4\pi r} I\left(t - \frac{r}{c}\right) \quad (20-6)$$

The scalar potential φ can be found either by applying the Lorentz condition or by using the appropriate expression for the retarded potential. Both methods give the same final result. However, because the electric potential due to a dipole is the difference between two large terms, great

care must be used in approximating the retarded potential. Since this difficulty is circumvented in the Lorentz condition calculation, the scalar potential will be obtained by solving

$$\nabla \cdot \mathbf{A} + \frac{1}{c^2} \frac{\partial \varphi}{\partial t} = 0 \quad (20-7)$$

with \mathbf{A} as given by Eq. (20-6). Thus,

$$\begin{aligned} \frac{\partial \varphi}{\partial t} &= -\frac{l}{4\pi\epsilon_0} \frac{\partial}{\partial z} \frac{1}{r} I\left(t - \frac{r}{c}\right) \\ &= \frac{l}{4\pi\epsilon_0} \left[\frac{z}{r^3} I\left(t - \frac{r}{c}\right) + \frac{z}{r^2 c} I'\left(t - \frac{r}{c}\right) \right] \end{aligned} \quad (20-8)$$

where I' represents the derivative of I with respect to its argument. This equation is readily integrated by noting that $I = +q'$ and thus that

$$\varphi(\mathbf{r}, t) = \frac{l}{4\pi\epsilon_0} \frac{z}{r^2} \left[\frac{q(t - r/c)}{r} + \frac{I(t - r/c)}{c} \right] \quad (20-9)$$

Having obtained the scalar and vector potentials, we now need only to differentiate them to obtain the electromagnetic field. Before doing so, it is convenient to specialize the charge-current distribution to one which varies harmonically with the time. The particular choice

$$\begin{aligned} q\left(t - \frac{r}{c}\right) &= q_0 \cos \omega\left(t - \frac{r}{c}\right) \\ I &= I_0 \sin \omega\left(t - \frac{r}{c}\right) = -\omega q_0 \sin \omega\left(t - \frac{r}{c}\right) \end{aligned} \quad (20-10)$$

will be made. Resolving \mathbf{A} into spherical components, we obtain

$$\begin{aligned} A_r &= \frac{\mu_0 I_0 l}{4\pi r} \cos \theta \sin \omega\left(t - \frac{r}{c}\right) \\ A_\theta &= -\frac{\mu_0 I_0 l}{4\pi r} \sin \theta \sin \omega\left(t - \frac{r}{c}\right) \\ A_\phi &= 0 \end{aligned} \quad (20-11)$$

and it becomes obvious that only the ϕ -component of \mathbf{B} is different from zero. This component is

$$\begin{aligned} B_\phi &= \frac{1}{r} \frac{\partial}{\partial r} (r A_\theta) - \frac{1}{r} \frac{\partial A_r}{\partial \theta} \\ &= \frac{\mu_0 I_0 l}{4\pi r} \sin \theta \left[\frac{\omega}{c} \cos \omega\left(t - \frac{r}{c}\right) + \frac{1}{r} \sin \omega\left(t - \frac{r}{c}\right) \right] \end{aligned} \quad (20-12)$$

The computation of the electric field is somewhat more complex, since not only \mathbf{A} but also φ is involved. The result of performing the differentiations is

$$\begin{aligned} E_r &= -\frac{\partial\varphi}{\partial r} - \frac{\partial A_r}{\partial t} = \frac{2I_0 \cos\theta}{4\pi\epsilon_0} \left[\frac{\sin\omega(t-r/c)}{r^2c} - \frac{\cos\omega(t-r/c)}{\omega r^3} \right] \\ E_\theta &= -\frac{1}{r} \frac{\partial\varphi}{\partial\theta} - \frac{\partial A_\theta}{\partial t} \\ &= -\frac{I_0 \sin\theta}{4\pi\epsilon_0} \left[\left(\frac{1}{\omega r^3} - \frac{\omega}{rc^2} \right) \cos\omega\left(t - \frac{r}{c}\right) - \frac{1}{r^2c} \sin\omega\left(t - \frac{r}{c}\right) \right] \\ E_\phi &= -\frac{1}{r \sin\theta} \frac{\partial\varphi}{\partial\phi} - \frac{\partial A_\phi}{\partial t} = 0 \end{aligned} \quad (20-13)$$

It is interesting to compute the rate at which the dipole radiates energy. This is done by integrating the normal component of the Poynting vector over a sphere of radius R . Thus

$$\oint \mathbf{S} \cdot \mathbf{n} \, da = \frac{1}{\mu_0} R^2 \int_0^\pi E_\theta B_\phi 2\pi \sin\theta \, d\theta \quad (20-14)$$

Equations (20-12) and (20-13) make it possible to evaluate completely the integral that appears in Eq. (20-14). However, it is perhaps more instructive to evaluate only the portion that does not vanish as $R \rightarrow \infty$, by selecting the term proportional to $1/r$ in E_θ and B_ϕ . The result is

$$\oint \mathbf{S} \cdot \mathbf{n} \, da = \frac{(I_0 l)^2 \omega^2}{6\pi\epsilon_0 c^3} \cos^2\omega\left(t - \frac{r}{c}\right) \quad (20-15)$$

This equation gives the instantaneous radiated power; the average radiated power (since the average of \cos^2 is one-half) is

$$\bar{P} = \frac{I^2 \omega^2}{6\pi\epsilon_0 c^3} \frac{I_0^2}{2} \quad (20-16)$$

A more conventional form of Eq. (20-16) is obtained by introducing $\lambda = 2\pi c/\omega$ and $c = 1/\sqrt{\epsilon_0\mu_0}$. The result is

$$\bar{P} = \frac{2\pi}{3} \sqrt{\frac{\mu_0}{\epsilon_0}} \left(\frac{l}{\lambda}\right)^2 \frac{I_0^2}{2} \quad (20-17)$$

A resistance R carrying a current $I_0 \cos\omega t$ dissipates energy at an average rate $\bar{P} = RI_0^2/2$.

Comparing this with Eq. (20-17), we see that it is sensible to define the *radiation resistance* of a dipole by

$$R_r = \frac{2\pi}{3} \sqrt{\frac{\mu_0}{\epsilon_0}} \left(\frac{l}{\lambda}\right)^2 \quad (20-18)$$

or

$$R_r = 789 \left(\frac{l}{\lambda} \right)^2 \text{ ohms (free space)}$$

For a material medium μ_0 and ϵ_0 are replaced by μ and ϵ , and $\lambda = 2\pi/\omega\sqrt{\epsilon\mu}$.

One might be tempted to use Eq. (20-18) to describe the radiation from a radio antenna. Unfortunately, several defects prevent obtaining good results in this way. The principal defects are (1) the effect of the proximity of the earth is neglected, (2) ordinarily antennas are not capacitively loaded at the ends, and (3) antennas are very seldom short compared with the wavelength they radiate. Removal of the last two defects will be discussed in the next section; however, discussion of the perturbing effect of the earth is beyond the scope of this text.

20-2 RADIATION FROM A HALF-WAVE ANTENNA

The restriction to lengths small compared with one wavelength can be removed in some cases by relatively simple means. In particular, a wire that is just one-half wavelength in length can be broken into infinitesimal elements, to each of which the method of the preceding section can be applied. Let the wire lie along the z -axis from $-\lambda/4$ to $+\lambda/4$ and carry a current

$$I(z', t) = I_0 \sin \omega t \cos \left(\frac{2\pi z'}{\lambda} \right) \quad (20-19)$$

This current vanishes at the ends of the wire. The nonuniformity of the current requires a varying charge density, which is largest at the ends of the wire. An element dz' at z' contributes, in vacuum,

$$dE_\theta = I_0 \frac{\sin \theta}{4\pi\epsilon_0 R c^2} \omega \cos \omega \left(t - \frac{R}{c} \right) \cos \left(\frac{2\pi z'}{\lambda} \right) dz' \quad (20-20)$$

to E_θ . Here R is the distance from dz' to the point of observation, and terms of order $1/R^2$ have been neglected. In the same way,

$$dB_\phi = \frac{\mu_0 I_0 \omega}{4\pi R c} \sin \theta \cos \omega \left(t - \frac{R}{c} \right) \cos \left(\frac{2\pi z'}{\lambda} \right) dz' \quad (20-21)$$

The problem in calculating E_θ and B_ϕ is reduced to evaluating

$$K = \int_{-\pi/2}^{\pi/2} \frac{1}{R} \cos \omega \left(t - \frac{R}{c} \right) \cos u \, du \quad (20-22)$$

where $u = 2\pi z'/\lambda$. As before, $R = r - z' \cos \theta$, and hence by choosing r sufficiently large, $z' \cos \theta$ can be made negligible. In the argument of the

cosine, however, more care is required, and K is written as

$$K = \frac{1}{r} \int_{-\pi/2}^{\pi/2} \cos \left[\omega \left(t - \frac{r}{c} \right) + u \cos \theta \right] \cos u \, du$$

The cosine can be expanded to give

$$K = \frac{1}{r} \cos \omega \left(t - \frac{r}{c} \right) \int_{-\pi/2}^{\pi/2} \cos (u \cos \theta) \cos u \, du \\ - \frac{1}{r} \sin \omega \left(t - \frac{r}{c} \right) \int_{-\pi/2}^{\pi/2} \sin (u \cos \theta) \cos u \, du$$

The second integral vanishes, and the first can be evaluated by expressing the cosines as exponentials or by using standard tables. The result is

$$K = \frac{2}{r} \cos \omega \left(t - \frac{r}{c} \right) \frac{\cos [(\pi/2) \cos \theta]}{\sin^2 \theta} \quad (20-23)$$

Having evaluated K , we find that

$$E_\theta = \frac{I_0}{2\pi\epsilon_0 r c} \cos \omega \left(t - \frac{r}{c} \right) \frac{\cos [(\pi/2) \cos \theta]}{\sin \theta} \\ B_\phi = \frac{\mu_0 I_0}{2\pi r} \cos \omega \left(t - \frac{r}{c} \right) \frac{\cos [(\pi/2) \cos \theta]}{\sin \theta} \quad (20-24)$$

The integrated average Poynting vector is

$$\bar{P} = \frac{1}{4\pi} \sqrt{\frac{\mu_0}{\epsilon_0}} I_0^2 \int_0^\pi \frac{\cos^2 [(\pi/2) \cos \theta]}{\sin^2 \theta} \sin \theta \, d\theta \quad (20-25)$$

The remaining integral can be evaluated as an infinite series or by numerical integration. Its value is 1.219 and so for a half-wave antenna result is

$$\bar{P} = 73.1 \text{ ohms} \frac{I_0^2}{2} \quad (\text{free space}) \quad (20-26)$$

This method can be applied to more complicated problems; however, the technical details become rather formidable.

This treatment of the half-wave antenna assumes a current distribution and then calculates the fields. An actual half-wave antenna is driven at its midpoint by a sinusoidally varying current or voltage. The excitation causes time-dependent charge and current distributions in the antenna, and these in turn are sources for the fields. However, the charge and current distributions must be consistent with these fields. This problem can be solved by solving the corresponding problems for a slender spheroid or a biconical array (in both cases the wave equation separates) and taking the limit as the minor axis of the spheroid or the cone angle goes to zero. In either case, the problem is complicated. Alternatively, an approach similar

to that used in Section 3-12 to compute the charge distribution on a long wire can be developed and used to obtain a numerical solution for the current distribution in the antenna. Once this current distribution is known, the fields can be calculated, as we have done above. Unfortunately, this too is a complicated procedure. All of these approaches have been used, and for the most part, they give equivalent results. The most useful remark that we can make here is that, for a long thin antenna, the fields calculated by the approximation of Section 20-2 are accurate to within a few percent.

20-3

RADIATION FROM A GROUP OF MOVING CHARGES

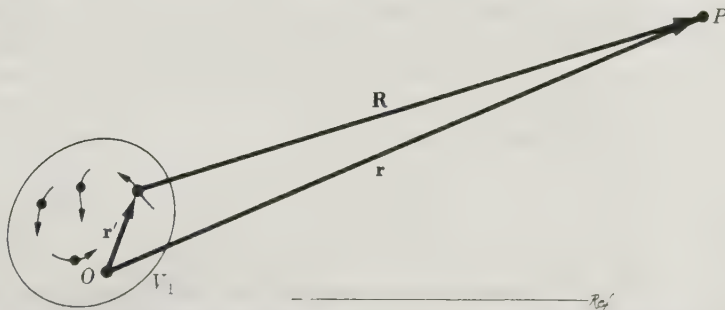
In this section, we shall derive an expression for the power radiated by a group of moving charges, or equivalently, by a charge-current distribution. The motion of the charges is *arbitrary* except for the following restrictions: During the time it takes for radiation to propagate from the neighborhood of the charges to the observation point, we can imagine that all charges and currents in the distribution are contained in a volume V_1 whose dimensions are small compared to the source-observer distance (see Fig. 20-2). Furthermore, the dimensions of V_1 are small compared to the dominant wavelengths of emitted radiation. The above restrictions also imply that, compared to the speed of light, the charges are moving slowly. It is assumed that the charges are moving in vacuum.

As a first step toward a solution to the problem, we must calculate the electromagnetic potentials. These are just the retarded potentials, which were discussed in Section 16-6. The origin of coordinates O is taken inside the volume V_1 and the position of a charge element is denoted by \mathbf{r}' (see Fig. 20-2). The field point P is at distance \mathbf{r} from the origin. For convenience, the auxiliary distance \mathbf{R} , which denotes the position of the field point relative to a charge element, is introduced. Clearly,

$$\mathbf{r}' + \mathbf{R} = \mathbf{r} \quad (20-27)$$

FIGURE 20-2

Arbitrarily moving charges contained in a volume V_1 . The fields are to be calculated at P .



Since $r \gg r'$

$$R = |\mathbf{r} - \mathbf{r}'| \approx r - \frac{\mathbf{r}' \cdot \mathbf{r}}{r} \quad (20-28)$$

The retarded scalar potential φ at the field point P now takes the form

$$\begin{aligned} \varphi(\mathbf{r}, t) &= \frac{1}{4\pi\epsilon_0} \int_{V_1} \frac{\rho(\mathbf{r}', t - R/c) dv'}{R} \\ &= \frac{1}{4\pi\epsilon_0} \int_{V_1} \frac{\rho(\mathbf{r}', t - r/c + \mathbf{r}' \cdot \mathbf{r}/cr) dv'}{r - (\mathbf{r}' \cdot \mathbf{r})/r} \end{aligned} \quad (20-29)$$

Using the Binomial Theorem,

$$\left(r - \frac{\mathbf{r}' \cdot \mathbf{r}}{r}\right)^{-1} = r^{-1} + r^{-2} \frac{\mathbf{r}' \cdot \mathbf{r}}{r} + \dots \quad (20-30)$$

and Taylor's series expansion,

$$\rho\left(\mathbf{r}', t - \frac{r}{c} + \frac{\mathbf{r}' \cdot \mathbf{r}}{cr}\right) = \rho\left(\mathbf{r}', t - \frac{r}{c}\right) + \frac{\mathbf{r}' \cdot \mathbf{r}}{cr} \frac{\partial \rho}{\partial t} \Big|_{\mathbf{r}', t-r/c} + \dots \quad (20-31)$$

we obtain

$$\begin{aligned} \varphi(\mathbf{r}, t) &= \frac{1}{4\pi\epsilon_0 r} \int_{V_1} \rho\left(\mathbf{r}', t - \frac{r}{c}\right) dv' + \frac{1}{4\pi\epsilon_0 r^3} \mathbf{r} \cdot \int_{V_1} \mathbf{r}' \rho\left(\mathbf{r}', t - \frac{r}{c}\right) dv' \\ &\quad + \frac{1}{4\pi\epsilon_0 r^2 c} \mathbf{r} \cdot \frac{d}{dt} \int_{V_1} \mathbf{r}' \rho\left(\mathbf{r}', t - \frac{r}{c}\right) dv' \\ &\quad + \text{higher-order terms} \end{aligned} \quad (20-32)$$

The first integral in Eq. (20-32) is the total charge Q of the distribution and is a constant, independent of time. The second (and also the third) integral is the electric dipole moment \mathbf{p} of the distribution, evaluated at the time $t - r/c$. The higher-order terms fall off with a larger power of r'/r and depend on a higher multipole moment of the distribution. Because of the restrictions imposed at the beginning of this section, these terms do not contribute appreciably (see below) to the distant electromagnetic field of the charge distribution. Thus,

$$\varphi(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0} \left[\frac{Q}{r} + \frac{\mathbf{r} \cdot \mathbf{p}(t - r/c)}{r^3} + \frac{\mathbf{r} \cdot \dot{\mathbf{p}}(t - r/c)}{cr^2} \right] \quad (20-33)$$

with $\dot{\mathbf{p}} \equiv d\mathbf{p}/dt$. As a result of the Taylor series expansion only one retarded time appears in the terms explicitly retained.

The retarded vector potential \mathbf{A} at the field point is given by

$$\begin{aligned}\mathbf{A}(\mathbf{r}, t) &= \frac{\mu_0}{4\pi} \int_{V_1} \frac{\mathbf{J}(\mathbf{r}', t - r/c + \mathbf{r}' \cdot \mathbf{r}/cr) dv'}{r - (\mathbf{r}' \cdot \mathbf{r})/r} \\ &= \frac{\mu_0}{4\pi r} \int_{V_1} \mathbf{J}\left(\mathbf{r}', t - \frac{r}{c}\right) dv' + \text{higher-order terms}\end{aligned}\quad (20-34)$$

The higher-order terms do not have to be written out explicitly since again they depend on a higher multipole moment of the distribution. In other words, Eq. (20-34) is already consistent with Eq. (20-32). From the results of Problem 7-2 one may write the last equation as

$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_0}{4\pi r} \dot{\mathbf{p}}\left(t - \frac{r}{c}\right) \quad (20-35)$$

The scalar potential of Eq. (20-33) could also have been obtained from Eq. (20-35) and the Lorentz condition.

The electric and magnetic fields may be obtained from the usual relationships

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla \varphi, \quad \mathbf{B} = \nabla \times \mathbf{A}$$

We shall restrict our interest here to the radiation zone fields, i.e., to contributions to \mathbf{E} and \mathbf{B} which fall off as r^{-1} , since these contributions are sufficient to determine the power radiated by the charge distribution. The calculation of $\partial \mathbf{A}/\partial t$ is straightforward; to obtain $\nabla \varphi$, we note that since $\dot{\mathbf{p}}$ is a function of $t - r/c$,

$$\frac{\partial}{\partial r} \dot{\mathbf{p}} \equiv -\frac{1}{c} \ddot{\mathbf{p}} \quad (20-36)$$

Thus,

$$\begin{aligned}\mathbf{E}(\mathbf{r}, t) &= \frac{-\mu_0}{4\pi r} \ddot{\mathbf{p}}\left(t - \frac{r}{c}\right) + \frac{1}{4\pi\epsilon_0} \frac{\mathbf{r} \cdot \ddot{\mathbf{p}}(t - r/c)}{c^2 r^3} \mathbf{r} \\ &+ \text{terms that fall off more rapidly than } (1/r)\end{aligned}\quad (20-37)$$

To calculate $\mathbf{B}(\mathbf{r}, t)$ we must take the curl of Eq. (20-35):

$$\begin{aligned}\nabla \times \left[\frac{1}{r} \dot{\mathbf{p}}\left(t - \frac{r}{c}\right) \right] &= \left(\nabla \frac{1}{r} \right) \times \dot{\mathbf{p}} + \frac{1}{r} \nabla \times \dot{\mathbf{p}} \\ &= -\frac{1}{r^2} \mathbf{r} \times \dot{\mathbf{p}} + \frac{1}{r} \mathbf{r} \times \frac{\partial \dot{\mathbf{p}}}{\partial r} \\ &= -\frac{1}{r^2} \mathbf{r} \times \dot{\mathbf{p}} - \frac{1}{cr} \mathbf{r} \times \ddot{\mathbf{p}}\end{aligned}$$

The last step comes about through the use of Eq. (20-36). The first term in

the curl may be disregarded since it falls off more rapidly than r^{-1} , so that

$$\nabla \times \mathbf{A}(\mathbf{r}, t) = \frac{-\mu_0}{4\pi cr^2} \mathbf{r} \times \ddot{\mathbf{p}} \left(t - \frac{r}{c} \right) \quad (20-38)$$

The radiation zone fields therefore are given by

$$\mathbf{B}(\mathbf{r}, t) = -\frac{\mu_0}{4\pi cr^2} \mathbf{r} \times \ddot{\mathbf{p}} \quad (20-39)$$

$$\begin{aligned} \mathbf{E}(\mathbf{r}, t) &= \frac{1}{4\pi\epsilon_0 c^2} \left[\frac{(\mathbf{r} \cdot \ddot{\mathbf{p}})\mathbf{r} - r^2 \ddot{\mathbf{p}}}{r^3} \right] = \frac{1}{4\pi\epsilon_0 c^2 r^3} \mathbf{r} \times (\mathbf{r} \times \ddot{\mathbf{p}}) \\ &= -\frac{c}{r} \mathbf{r} \times \mathbf{B}(\mathbf{r}, t) \end{aligned} \quad (20-40)$$

where $\ddot{\mathbf{p}}$ is evaluated at the retarded time.

It is evident that \mathbf{E} and \mathbf{B} are perpendicular to each other and are each perpendicular to \mathbf{r} . This result could have been anticipated from Eq. (17-22), since $\mathbf{r}/r = \boldsymbol{\kappa}/\kappa = \mathbf{u}$. Thus, it is sufficient to calculate only \mathbf{A} when only the radiation field is sought ($r \gg \lambda$). The Poynting vector $\mathbf{S} = (1/\mu_0)(\mathbf{E} \times \mathbf{B})$ has the direction of \mathbf{r} and is given by

$$\mathbf{S} = \frac{c}{\mu_0 r} \mathbf{B} \times (\mathbf{r} \times \mathbf{B}) = \frac{c}{\mu_0 r} \mathbf{r} B^2 \quad (20-41)$$

or

$$\mathbf{S} = \frac{1}{16\pi^2 \epsilon_0 c^3 r^5} \mathbf{r} (\mathbf{r} \times \ddot{\mathbf{p}})^2$$

If the z -axis is taken in the direction of $\ddot{\mathbf{p}}$,

$$\mathbf{S} = \frac{\ddot{p}^2 \sin^2 \theta}{16\pi^2 \epsilon_0 c^3 r^2} \frac{\mathbf{r}}{r} \quad (20-42)$$

The maximum power is radiated at 90° to $\ddot{\mathbf{p}}$. The total radiated power is obtained by integrating the Poynting vector over a closed surface surrounding the charge distribution. A convenient choice for such a surface is a sphere, centered in the charge distribution, with sufficiently large radius so that all parts of its surface are in the radiation zone. Then

$$\begin{aligned} P_R &= -\frac{dW}{dt} = \oint_s \mathbf{S} \cdot \mathbf{n} da \\ &= \frac{\ddot{p}^2}{16\pi^2 \epsilon_0 c^3} \int \frac{\sin^2 \theta}{r^3} \mathbf{r} \cdot \frac{\mathbf{r}}{r} r^2 \sin \theta d\theta d\phi \end{aligned}$$

from which one readily obtains the important result

$$P_R = -\frac{dW}{dt} = \frac{1}{4\pi\epsilon_0} \frac{2}{3} \frac{\ddot{p}^2}{c^3} \quad (20-43)$$

for the power radiated by a group of charges moving slowly compared to the speed of light.

Equation (20-43) is an expression for the power radiated by arbitrarily moving charges in terms of their electric dipole moment \mathbf{p} . The previously derived expression for the radiation from an oscillating dipole, Eq. (20-16), is a special example of (20-43); in that case $p = (I_0/\omega) \cos \omega(t - r/c)$. Now it could happen that due to a special symmetry of the system, the electric dipole moment vanishes or is independent of time. In this case, the power radiated is not necessarily zero, but more terms would have to be retained in the expansions of φ and \mathbf{A} (Eqs. 20-32 and 20-34) to make the calculation. In fact, we would find that the power radiated depends on some higher multipole moment of the system in this case. The various multipole radiations get progressively less intense as the multipole order is increased; for example, quadrupole radiation is roughly a factor $(a/\lambda)^2$ smaller than dipole radiation, where a is the dimension of the system and λ the wavelength of the emitted radiation. Thus, if \ddot{p} does not vanish for the system under consideration, Eq. (20-43) gives the major contribution to the radiated power.

Equations (20-39), (20-40), and (20-43) can also be applied to the radiation from a single accelerated charge q . The dipole moment of the charge is $q\mathbf{r}'$, where \mathbf{r}' is measured from an arbitrary origin. Then,

$$\dot{\mathbf{p}} = q\dot{\mathbf{r}}' = q\mathbf{v}$$

where \mathbf{v} is the velocity of the charge. This result is independent of the origin. Finally

$$\ddot{\mathbf{p}} = q\dot{\mathbf{v}}$$

where $\dot{\mathbf{v}}$ is the acceleration of the charge. Substituting this last result into Eq. (20-43), we obtain

$$P_R = -\frac{dW}{dt} = \frac{q^2}{4\pi\epsilon_0} \frac{2\dot{\mathbf{v}}^2}{3c^3} \quad (20-44)$$

for the power radiated by a *slowly* moving accelerated charge.

NEAR AND INTERMEDIATE ZONE FIELDS

Only terms in the fields proportional to $1/r$ contribute to the radiated energy, but higher powers of $1/r$ dominate in the region close to the radiating charges. These fields are of interest for comparison to the static case, and also in practical problems in the immediate vicinity of antennas. In this section, we shall calculate the entire E - and B -fields of a point electric dipole. To do so, we start with the dipole potentials Eqs. (20-33) and (20-35), and we retain the terms in the differentiations that were discarded in the preceding section. The contribution to \mathbf{E} from $-\partial\mathbf{A}/\partial t$ is

part of the radiation field and has already been calculated. Applying the vector identity (1-1-6) to Eq. (20-33) with $Q = 0$, we find

$$-\nabla\varphi = \frac{-1}{4\pi\epsilon_0} \left[\frac{\mathbf{r}}{cr^2} \cdot \nabla\dot{\mathbf{p}} + \frac{\mathbf{r}}{cr^2} \times (\nabla \times \dot{\mathbf{p}}) + \frac{\mathbf{r}}{r^3} \cdot \nabla\mathbf{p} \right. \\ \left. + \frac{\mathbf{r}}{r^3} \times (\nabla \times \mathbf{p}) + \dot{\mathbf{p}} \cdot \nabla \frac{\mathbf{r}}{cr^2} + \mathbf{p} \cdot \nabla \frac{\mathbf{r}}{r^3} \right]$$

with the terms arranged in order of increasing powers of $1/r$. The first terms, when added to $-\partial\mathbf{A}/\partial t$, give the *radiation field* of Eq. (20-40),

$$\mathbf{E}_1 = \frac{1}{4\pi\epsilon_0} \frac{1}{c^2 r} \left[\frac{\mathbf{r}}{r} \cdot \ddot{\mathbf{p}} \frac{\mathbf{r}}{r} - \ddot{\mathbf{p}} \right] = \frac{1}{4\pi\epsilon_0} \frac{1}{c^2 r r} \times \left(\frac{\mathbf{r}}{r} \times \ddot{\mathbf{p}} \right) \quad (20-40)$$

which varies as $1/r$ and depends on $\ddot{\mathbf{p}}$. The last term, when expanded, is just the *static dipole field* of Eq. (2-36),

$$\mathbf{E}_3 = \frac{1}{4\pi\epsilon_0} \frac{1}{r^3} \left[3 \frac{\mathbf{r}}{r} \cdot \mathbf{p} \frac{\mathbf{r}}{r} - \mathbf{p} \right] = \frac{1}{4\pi\epsilon_0} \frac{1}{r^3} \left[\frac{\mathbf{r}}{r} \times \left(\frac{\mathbf{r}}{r} \times \mathbf{p} \right) + 2 \frac{\mathbf{r}}{r} \cdot \mathbf{p} \frac{\mathbf{r}}{r} \right] \quad (2-36)$$

which varies as $1/r^3$ and depends on \mathbf{p} . The dipole \mathbf{p} is varying with time in this case, of course, but the spatial dependence of \mathbf{E}_3 is instantaneously that of a static dipole field. The third, fourth, and fifth terms (which are similar in form to the first, second, and last terms) lead to the transition field or the *induction field*,

$$\mathbf{E}_2 = \frac{1}{4\pi\epsilon_0} \frac{1}{cr^2} \left[3 \frac{\mathbf{r}}{r} \cdot \dot{\mathbf{p}} \frac{\mathbf{r}}{r} - \dot{\mathbf{p}} \right] = \frac{1}{4\pi\epsilon_0} \frac{1}{cr^2} \left[\frac{\mathbf{r}}{r} \times \left(\frac{\mathbf{r}}{r} \times \dot{\mathbf{p}} \right) + 2 \frac{\mathbf{r}}{r} \cdot \dot{\mathbf{p}} \frac{\mathbf{r}}{r} \right] \quad (20-45)$$

which varies as $1/r^2$ and depends on $\dot{\mathbf{p}}$. The total E -field is then

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3$$

Here \mathbf{p} and its derivatives are to be evaluated at the retarded time, $t' = t - r/c$. The B -field is obtained by taking the curl of Eq. (20-35). If we retain the term that was neglected in the derivation of Eq. (20-39), we see from Eqs. (20-40) and (20-45) that

$$\mathbf{B} = \frac{\mathbf{r}}{cr} \times (\mathbf{E}_1 + \mathbf{E}_2)$$

Thus, \mathbf{B} contains a radiation and an induction field, but there is no "static" B -field.

From the second form given above for each of $\mathbf{E}_{1,2,3}$, it is clear that all three fields have a transverse component, which is perpendicular to \mathbf{r} and in the plane defined by \mathbf{r} and \mathbf{p} . The radiation field is purely transverse, but

\mathbf{E}_2 and \mathbf{E}_3 have in addition a longitudinal component in the direction of \mathbf{r} . If we choose spherical coordinates with the polar axis in the direction of \mathbf{p} , then the transverse components are E_θ and are proportional to $\sin \theta$; the longitudinal components are E_r and are proportional to $\cos \theta$. The B -field has only a ϕ -component (TM).

Let us specialize to the case of a dipole in the \mathbf{k} direction with a sinusoidally oscillating magnitude: $\mathbf{p}(t) = p\mathbf{k}e^{-i\omega t}$. Then,

$$\mathbf{p}(t') = p\mathbf{k}e^{-i(\omega t - \kappa r)}$$

where κ is the propagation constant of the wave in the radiation zone,

$$\kappa = \omega/c$$

Inserting this $\mathbf{p}(t')$ into Eqs. (20-40), (2-36), and (20-45), we find for the components of $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3$:

$$E_\theta = \frac{p\kappa^3}{4\pi\epsilon_0} \sin \theta \left[-\frac{1}{\kappa r} - i\frac{1}{(\kappa r)^2} + \frac{1}{(\kappa r)^3} \right] e^{-i(\omega t - \kappa r)}$$

$$E_r = \frac{p\kappa^3}{4\pi\epsilon_0} 2 \cos \theta \left[-i\frac{1}{(\kappa r)^2} + \frac{1}{(\kappa r)^3} \right] e^{-i(\omega t - \kappa r)}$$

The radiation field or the static field dominates when $\kappa r \gg 1$ or $\kappa r \ll 1$, respectively—that is, when r is large or small compared to the wavelength of the emitted radiation. The transition between the “static” dipole field and the transverse radiation field is shown schematically by the field lines of Fig. 20-3. In the radiation zone the field is an outward propagating spherical wave.

RADIATION DAMPING: THOMSON CROSS SECTION

The radiated power calculated in Section 20-3 is lost by the system of charges, and in a steady-state situation, it must be made up by another source. For an antenna, this source is the transmitter, and the loss is expressed by the radiation resistance. For an electron in a material medium through which a wave is propagating, the power source is the E -field of the wave and the loss is expressed by the damping frequency. We now want to relate the damping frequency used in the last chapter to the rate of energy loss due to radiation by a charged particle, as given in Eq. (20-44),

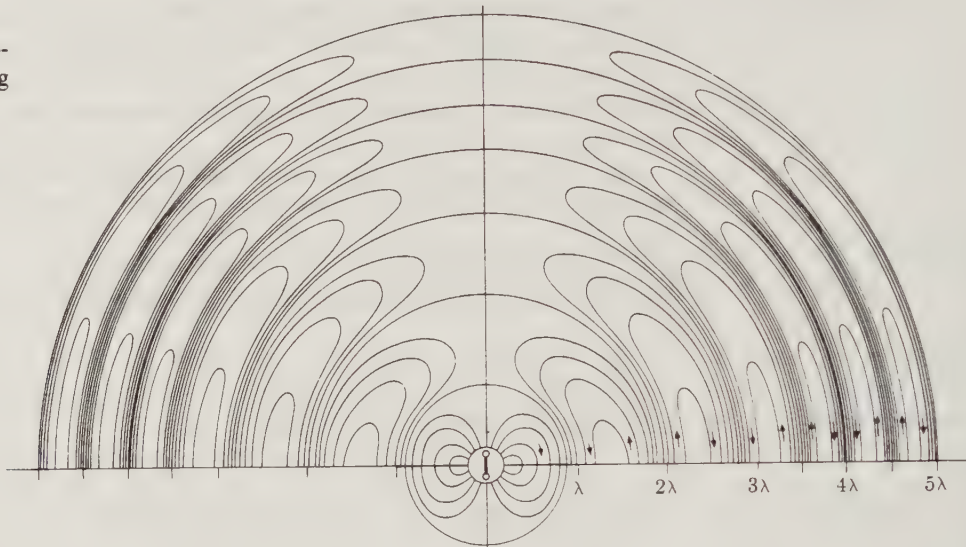
$$P = \frac{e^2}{4\pi\epsilon_0} \frac{2}{3} \frac{\dot{v}^2}{c^3} \quad (20-44)$$

The power loss due to a force F is $P = -Fv$. Equating this to the radiation loss of Eq. (20-44), we have for the damping force

$$F = -\frac{2}{3} \frac{e^2}{4\pi\epsilon_0 c^3} \frac{\dot{v}^2}{v} \quad (20-46)$$

FIGURE 20-3

Electric field lines produced by an oscillating electric dipole



The assumed linear damping force in Eq. (19-22) was $F = -Gv$, so that the damping frequency $\gamma = G/m$ is

$$\gamma = \frac{2}{3} \frac{e^2}{4\pi\epsilon_0 mc^3} \frac{\dot{v}^2}{v^2}$$

For an oscillating charge, $v = v_0 \sin \omega t$, and

$$\gamma = \frac{2}{3} \frac{e^2 \omega^2}{4\pi\epsilon_0 mc^3} \frac{\cos^2 \omega t}{\sin^2 \omega t}$$

The trouble with this result is that γ is not a constant, but varies from zero to infinity over a cycle of the harmonic motion. Nevertheless the average effect of the damping is very small in one cycle, and since $\bar{P} = m\gamma \overline{v^2} \propto \overline{v^2}$ and

$$\overline{\cos^2 \omega t} = \overline{\sin^2 \omega t}$$

the effective damping frequency is taken to be*

$$\gamma = \frac{2}{3} \frac{e^2 \omega^2}{4\pi\epsilon_0 mc^3} \quad (20-47)$$

* This last step of the argument is admittedly somewhat crude. The problem is not due to a failure of classical mechanics, since a quantum mechanical calculation gives the same result. The problem of a correct incorporation of the radiation reaction force Eq. (20-46) into dynamical theory has been discussed many times since Lorentz's presentation of Eq. (20-47) in 1909. It is a fundamental problem, but we cannot go into it here.

In terms of the "classical radius of the electron," $R_e = e^2/4\pi\epsilon_0 mc^2 = 2.81 \times 10^{-15}$ m,

$$\frac{\gamma}{\omega} = \frac{4\pi R_e}{3 \lambda} \quad (20-48)$$

Since $\gamma = \Delta\omega$ is the width of a narrow absorption peak and since $\Delta\omega/\omega = \Delta\lambda/\lambda$,

$$\Delta\lambda = \frac{4\pi}{3} R_e = 1.16 \times 10^{-4} \text{ \AA}$$

This *natural line width* is narrower than that ordinarily seen in absorption spectra, even in gases at low pressure, because there are other damping mechanisms that are usually greater than the radiation damping. The natural line width is, however, a lower limit on the possible damping.

Another immediate consequence of Eq. (20-44) that is more commonly observable is the Thomson cross section for the scattering of X-rays. X-ray frequencies (photon energies) are large compared to the resonance frequencies (binding energies) of many of the electrons in matter. These can be treated as free electrons accelerated by the E -field of the X-rays,

$$m\dot{v} = eE \quad (20-49)$$

so that

$$P = \frac{1}{4\pi\epsilon_0} \frac{2}{3} \frac{e^4}{m^2 c^3} E^2$$

is the total power radiated by one electron. The *Thomson scattering cross section* σ_T is defined as P divided by the incident Poynting vector (power per unit area)

$$S_0 = \frac{1}{\mu_0 c} E^2 \quad (20-50)$$

Thus

$$\sigma_T = \frac{2}{3} \frac{1}{4\pi\epsilon_0^2} \frac{e^4}{m^2 c^4} = \frac{8\pi}{3} R_e^2 \quad (20-51)$$

where R_e is the classical radius of the electron. The smallness of this cross section compared to the size of an atom is the reason X-rays are penetrating. The angular dependence of the scattered radiation is given by the *differential cross section*, defined from Eq. (20-42):

$$d\sigma_T = \frac{dP}{S_0}$$

where

$$dP = S da = S r^2 d\Omega$$

is the power scattered (i.e., re-radiated) into the element of solid angle

$$d\Omega = \sin \theta d\theta d\phi$$

With S from Eq. (20-42), the differential cross section is

$$\frac{d\sigma_T}{d\Omega} = R_e^2 \sin^2 \theta \quad (20-52)$$

Here θ is the angle between the observation direction and the E -vector of the incident wave (perpendicular to its propagation direction). If the incident X-rays are unpolarized (the usual case), a more useful expression is the average of this over all polarization directions. The result (Problem 20-12) is

$$\frac{d\sigma_T}{d\Omega} = R_e^2 \frac{1 + \cos^2 \beta}{2} \quad (20-53)$$

where β is the angle between the observation direction and the incident propagation direction. This cross section is measurable in the *Compton effect*, which is the incoherent* scattering of X-rays at frequencies too high for resonance absorption. This classical formula ceases to hold at very high frequencies; the photon energy ($\hbar\omega$) must be much less than the electron rest-energy mc^2 , or $\lambda \gg 0.02 \text{ \AA}$.

The Thomson cross section Eq. (20-51) also follows from the theory of Chapter 19, if we investigate the attenuation of the incident Poynting vector instead of the magnitude of the scattered one. From Eq. (20-50),

$$S_0 = \frac{1}{\mu_0 c} E^2 = \frac{1}{\mu_0 c} E_0^2 e^{-2z/\delta}$$

where

$$\delta = \frac{c}{k\omega}$$

is the "skin depth." Here $\delta/2$ plays the role of a mean free path for the incident photons. Since the mean free path equals $1/N\sigma$, where N is the number of electrons per unit volume,

$$\sigma_T = \frac{2}{N\delta} = \frac{2k\omega}{Nc} \quad (20-54)$$

Now when $n \approx 1$, $k \ll 1$, as for X-rays,

$$k = \frac{1}{2} K_i = \frac{\omega_p^2 \gamma}{2\omega^3}$$

* Incoherent scattering is opposed to the coherent scattering (with the same individual cross sections) that gives X-ray diffraction in crystals, under special conditions on wavelength and angle of incidence (Bragg's law).

This equation is the high frequency result for free electrons, and it also holds for bound electrons when $\omega \gg \omega_0$. Using $\omega_p^2 = Ne^2/\epsilon_0 m$, $\gamma = (4\pi/3)(R_e/\lambda)\omega$ from Eq. (20-48), and $\lambda = 2\pi c/\omega$, we have

$$k = \frac{Ne^2}{2\epsilon_0 m} \frac{4\pi}{3} \frac{R_e \omega^2}{2\pi c \omega^3} = \frac{4\pi}{3} \frac{Ne^2}{4\pi\epsilon_0 m c \omega} R_e$$

and from Eq. (20-54),

$$\sigma_T = \frac{8\pi}{3} R_e^2 \quad (20-51)$$

The damping frequency γ , which gives the Thomson cross section, is the radiation damping frequency. At such high frequencies ($\omega \geq 10^{19} \text{ s}^{-1}$), it dominates the collision frequency $1/\tau$. This approach, of looking at the propagation of the incident radiation, does not of course give the angle dependence Eq. (20-52) of the scattered radiation.

SUMMARY

The fields produced by time-dependent charge and current distributions are calculated from the integral solutions for \mathbf{A} and φ . (It would be sufficient to treat only \mathbf{A} , since φ is related to it by the Lorentz condition.) To find solutions valid at large distances $r \gg a$, where a is the size of the region in which the source currents are located, the integrand is expanded by the multipole expansion. The **electric dipole** term treated here is the most important, unless it happens to vanish in which case higher-order terms would have to be considered. The integral is further simplified for a **point dipole** $a \ll \lambda$, since all the source current elements are approximately in phase. (This condition implies that $v \ll c$.) The result is

$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_0}{4\pi r} \dot{\mathbf{p}}(t - r/c)$$

where

$$\dot{\mathbf{p}}(t) = \int \mathbf{J}(\mathbf{r}', t) dv'$$

From this, with a prescribed \mathbf{J} , one finds \mathbf{B} , \mathbf{E} , and \mathbf{S} . Finding them is particularly easy in the *radiation field* at $r \gg \lambda$. Then

$$\mathbf{E} = -c \frac{\mathbf{r}}{r} \times \mathbf{B} \quad \text{and} \quad \mathbf{S} = \frac{c}{\mu_0} B^2 \frac{\mathbf{r}}{r}$$

locally as for a plane wave. (It is unnecessary to find φ .)

- The Poynting vector in the **radiation field** of a **point dipole** is

$$\mathbf{S} = \frac{\ddot{\mathbf{p}}^2 \sin^2 \theta}{16\pi^2 \epsilon_0 c^3 r^2} \frac{\mathbf{r}}{r}$$

Integrating this over the entire sphere gives the total power radiated as

$$P = \frac{1}{4\pi\epsilon_0} \frac{2}{3} \frac{\ddot{\mathbf{p}}^2}{c^3}$$

• The radiation from a short linear (**electric dipole**) **antenna** is given by this equation, with $p(t) = (I_0 l / \omega) \cos \omega t$. In terms of the **radiation resistance** R_r ,

$$\bar{P} = \frac{1}{2} R_r I_0^2$$

$$R_r = 789 \left(\frac{l}{\lambda} \right)^2 \text{ ohms} \quad (l \ll \lambda)$$

in free space.

• The radiation from a longer antenna can be found by integrating the results for a short antenna, assuming $v = c$, as is always true for a radio antenna. For a **half-wave antenna**,

$$R_r = 73.1 \text{ ohms}$$

• For a **slowly moving point charge**, $\ddot{\mathbf{p}} = q\ddot{\mathbf{v}}$,

$$P = \frac{1}{4\pi\epsilon_0} \frac{2}{3} \frac{q^2 \dot{v}^2}{c^3}$$

• For an oscillating electron, this leads to the **radiation damping** frequency

$$\gamma = \frac{4\pi R_e}{3} \frac{\omega}{\lambda}$$

and the **Thomson cross section** for a free electron

$$\sigma_T = \frac{8\pi}{3} R_e^2$$

where

$$R_e = \frac{1}{4\pi\epsilon_0} \frac{e^2}{mc^2} = 2.81 \times 10^{-15} \text{ m}$$

is the “classical radius of the electron.”

PROBLEMS

20-1 Suppose a spherically symmetric charge distribution is oscillating purely in the radial direction, so that it remains spherically symmetric at every instant. Prove that no radiation is emitted.

20-2 (a) Determine, as a function of the angles θ and ϕ , the average power density radiated into vacuum by an oscillating dipole. (b) Calculate the total power radiated by a dipole of length 3 m at a frequency of 500 kHz if the current in the dipole is 2 A

(effective value). (c) What is the radiation resistance of the dipole oscillator in part (b)?

20-3 A circular loop of wire carrying the current $I = I_0 \cos \omega t$ constitutes an *oscillating magnetic dipole*. Determine the radiation fields \mathbf{E} and \mathbf{B} for this oscillator and the total power radiated. [*Hint*: Start from Eq. (8-66), reformulated in terms of the retarded potential:

$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_0 I_0}{4\pi} \oint \frac{\cos \omega(t - |\mathbf{r} - \mathbf{r}_1|/c)}{|\mathbf{r} - \mathbf{r}_1|} d\mathbf{r}_1$$

By symmetry \mathbf{A} has only a ϕ component:

$$A_\phi(\mathbf{r}, t) = \frac{\mu_0 I_0 a}{4\pi} \int_0^{2\pi} \frac{\cos \omega(t - |\mathbf{r} - \mathbf{r}_1|/c)}{|\mathbf{r} - \mathbf{r}_1|} \cos \phi d\phi$$

Also $|\mathbf{r} - \mathbf{r}_1| = (r^2 + a^2 - 2ra \sin \theta \cos \phi)^{1/2}$ with $a =$ radius of the loop. The dipole approximation requires that $r \gg a$ and $\omega a \ll c$.]

20-4 As sources of electromagnetic radiation, determine the relative efficiency of an electric dipole of length 2 m compared with a magnetic dipole of the same diameter at a frequency of 1 MHz.

20-5 (a) Compute the maximum current to a half-wave antenna that radiates 1 kW. (b) What is the E -field corresponding to the average power density at a distance of 10 km from the antenna? Ignore effects of the earth.

20-6 Verify that \mathbf{A} and φ in Eqs. (20-35) and (20-33) satisfy the Lorentz condition.

20-7 A measure of the directivity of an antenna is the power per steradian in a direction in which this is a maximum, divided by $1/(4\pi)$ times the total power radiated. Calculate the directivity for an oscillating electric dipole.

20-8 Suppose an electric dipole \mathbf{p} rotates with a constant angular velocity ω about an axis perpendicular to the dipole moment. Find the radiation field and the Poynting vector. [*Hint*: Treat the rotating dipole as the superposition of two sinusoidally varying dipoles at right angles to each other.]

20-9 The classical model of the hydrogen atom has the electron revolving in a circular orbit of radius r and kinetic energy

$$E_k = \frac{1}{2} \frac{e^2}{4\pi\epsilon_0 r}$$

(a) Calculate the fractional energy radiated per revolution, PT/E_k , where T is the orbital period. (b) Quantum mechanics prescribes that in the n th level

$$\frac{v}{c} = \frac{1}{n} \frac{1}{137}$$

Evaluate PT/E_k for $n = 2$.

20-10 An electron has been accelerated in a synchrotron of radius 2 m to half the speed of light. (a) Calculate the power radiated by the electron (rest mass = 9.11×10^{-31} kg). (b) What fraction of its kinetic energy is radiated per second?

20-11 Determine the radiation resistance of the oscillating magnetic dipole of

Problem 20-3. Express your result in terms of a and λ , so that it can be compared with Eq. (20-18).

20-12 An *unpolarized* X-ray beam of intensity I_0 is incident on matter containing free electrons. Considering one electron only and using the expressions from Section 20-3, show that the intensity of the scattered beam is given by

$$I_s = I_0 \frac{(1 + \cos^2 \beta) R_e^2}{2 r^2}$$

where β is the angle between OP and the original X-ray beam. Point O is the position of the electron, and P is the point where the scattered beam is to be measured.

20-13 X-rays of wavelength 0.2 \AA are attenuated in aluminum primarily by Compton scattering. Calculate the absorption coefficient $\alpha = 2/\delta$ from the Thomson cross section. There are 6.06×10^{28} Al atoms/m³.

Electrodynamics

The field produced by a fast moving point charge can be calculated from the retarded potentials. There are, however, certain difficulties associated with this calculation, which are related to the retardation and reflect the fact that the present charge distribution (in space) must be extrapolated back to the appropriate retarded time. This procedure would be essentially trivial except that different portions of the charge distribution require different retarded times. Although one might expect this effect to disappear for point charges, it actually does not. The appropriate scalar and vector potentials for a moving point charge are the Lienard-Wiechert potentials, which we derive in this chapter.

21-1 THE LIENARD-WIECHERT POTENTIALS

The Lienard-Wiechert potentials are the scalar and vector potentials produced by a moving point charge. One might think that $q/4\pi\epsilon_0 R$, with R the appropriate retarded distance, would give the scalar potential due to a moving point charge. This, however, is not the case, as can be shown in several ways. One of the most instructive procedures is to consider a moving volume carrying with it a fixed charged distribution, for example, a uniformly charged spherical volume, moving through space along a prescribed trajectory. The field due to a point charge is the properly taken limit of the field due to such a distribution.

The scalar potential due to a moving charge distribution, at point ξ and time t , is given by the retarded potential*

$$\varphi(\xi, t) = \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\mathbf{r}', t')}{|\xi - \mathbf{r}'|} dv' \quad (21-1)$$

* Throughout this chapter only free space is considered.

The crux of the difficulty is now apparent; namely, t' is not fixed and hence the volume of integration—that is, the volume in which ρ is different from zero—cannot be readily specified. To obviate the difficulty, some fixed time t_1 may be chosen and the integration over \mathbf{r}' changed to an integration over \mathbf{r}_1 . The most convenient choice for t_1 is the retarded time for some point in the interior of the charge distribution. If at time t_1 the charged volume is moving with a velocity $\mathbf{v}(t_1)$, then the important relationships are

$$\rho(\mathbf{r}', t') = \rho(\mathbf{r}_1, t_1) \quad (21-2)$$

$$\mathbf{r}_1 = \mathbf{r}' - \mathbf{v}(t')(t' - t_1) - \frac{1}{2}\dot{\mathbf{v}}(t')(t' - t_1)^2 + \dots \quad (21-3)$$

where $\dot{\mathbf{v}}$ is the time derivative of \mathbf{v} . It is important to understand that t' in Eq. (21-3) is not constant, but depends on \mathbf{r}' . The remaining problem is that of relating dv' to dv_1 , which is, of course, accomplished through the Jacobian. The relationship is

$$dv_1 = \frac{\partial(x_1, y_1, z_1)}{\partial(x', y', z')} dv' \quad (21-4)$$

where the Jacobian, $\partial(x_1, y_1, z_1)/\partial(x', y', z')$, is given by

$$\frac{\partial(x_1, y_1, z_1)}{\partial(x', y', z')} = \begin{vmatrix} \frac{\partial x_1}{\partial x'} & \frac{\partial x_1}{\partial y'} & \frac{\partial x_1}{\partial z'} \\ \frac{\partial y_1}{\partial x'} & \frac{\partial y_1}{\partial y'} & \frac{\partial y_1}{\partial z'} \\ \frac{\partial z_1}{\partial x'} & \frac{\partial z_1}{\partial y'} & \frac{\partial z_1}{\partial z'} \end{vmatrix} \quad (21-5)$$

The derivatives are

$$\frac{\partial x_1}{\partial x'} = 1 - v'_x \frac{\partial t'}{\partial x'} - \dot{v}'_x(t' - t_1) \frac{\partial t'}{\partial x'} + \dots$$

and

$$\frac{\partial x_1}{\partial y'} = -v'_x \frac{\partial t'}{\partial y'} - \dot{v}'_x(t' - t_1) \frac{\partial t'}{\partial y'} + \dots \quad (21-6)$$

where v'_x is $v_x(t')$, the x -component of the velocity at the retarded time t' . The retarded time t' is related to the retarded position simply by

$$t' = t - \frac{|\mathbf{r}' - \boldsymbol{\xi}|}{c} \quad (21-7)$$

Hence,

$$\frac{\partial t'}{\partial x'} = -\frac{n'_x}{c} \quad (21-8)$$

where \mathbf{n}' is a unit vector in the direction $\mathbf{r}' - \boldsymbol{\xi}$. A straightforward, though

tedious, expansion of the Jacobian, using Eqs. (21-6) and (21-8), gives

$$\frac{\partial(x_1, y_1, z_1)}{\partial(x', y', z')} = 1 + \frac{\mathbf{v}' \cdot \mathbf{n}'}{c} + \frac{\dot{\mathbf{v}}' \cdot \mathbf{n}'(t' - t_1)}{c} + \dots \quad (21-9)$$

where the higher terms involve the second and higher derivatives of \mathbf{v}' .

Equation (21-9) can be used in Eq. (21-1) to obtain the scalar potential. However, since the principal interest is in small charged volumes (point charges), it is appropriate to note that if

$$\frac{\dot{\mathbf{v}}' \cdot \mathbf{n}'}{c}(t' - t_1) \cong \frac{\dot{v}d}{c^2} \ll 1$$

where d measures the size of the charge distribution, then this term can certainly be neglected in the limiting point charge case. Similar criteria exist for the terms involving higher derivatives; however, we need not consider these. Finally, then

$$\varphi(\xi, t) = \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\mathbf{r}_1, t_1)}{|\xi - \mathbf{r}'|} \frac{dv_1}{1 + \mathbf{v}' \cdot \mathbf{n}'/c} \quad (21-10)$$

Again, if $d \ll |\xi - \mathbf{r}'|$, then $|\xi - \mathbf{r}'|$ can be replaced by R_{t_1} , the distance from the interior point (chosen earlier) to the observation point at time t_1 . Thus,*

$$\varphi(\xi, t) = \frac{1}{4\pi\epsilon_0} \frac{1}{R_{t_1}(1 + \mathbf{v}' \cdot \mathbf{n}'/c)} \int \rho(\mathbf{r}_1, t_1) dv_1 \quad (21-11)$$

or, since the integral is now over a well-defined volume,

$$\varphi(\xi, t) = \frac{1}{4\pi\epsilon_0} \frac{q}{R_{t_1}(1 + \mathbf{v}' \cdot \mathbf{n}'/c)} \quad (21-12)$$

which is the scalar Lienard-Wiechert potential. The vector potential is found to be

$$\mathbf{A}(\xi, t) = \frac{\mu_0}{4\pi} \frac{q\mathbf{v}'}{R_{t_1}(1 + \mathbf{v}' \cdot \mathbf{n}'/c)} \quad (21-13)$$

These expressions are often written as

$$\varphi(\xi, t) = \frac{q}{4\pi\epsilon_0} \left[\frac{1}{R(1 + \mathbf{v} \cdot \mathbf{n}/c)} \right]_{\text{ret}} \quad (21-14)$$

and

$$\mathbf{A}(\xi, t) = \frac{\mu_0 q}{4\pi} \left[\frac{\mathbf{v}}{R(1 + \mathbf{v} \cdot \mathbf{n}/c)} \right]_{\text{ret}}$$

* Note that $\mathbf{v}' \cdot \mathbf{n}' = \mathbf{v}(t_1) \cdot \mathbf{n}(t_1)$ to the approximation involved in Eq. (21-11).

which simply means that the quantities in brackets must be evaluated at our t_1 .

21-2

THE FIELD OF A UNIFORMLY MOVING POINT CHARGE

The most direct application of the Lienard-Wiechert potentials is to the calculation of the field of a point charge moving in a straight line with constant velocity. The geometry of such a situation is shown in Fig. 21-1. The field at point P is to be calculated at time t , at which time the charge is at x . The retarded position x' and the retarded time t' are determined by

$$R'^2 = c^2(t - t')^2 = (x_0 - x')^2 + b^2 \tag{21-15}$$

The scalar potential is given by

$$\varphi(P, t) = \frac{q}{4\pi\epsilon_0 R' [1 + (\mathbf{v} \cdot \mathbf{n}'/c)]} \tag{21-16}$$

From the diagram it is clear that

$$R' \frac{\mathbf{v} \cdot \mathbf{n}'}{c} = -R' \frac{v x_0 - x'}{c R'} = -\frac{v(x_0 - x')}{c} \tag{21-17}$$

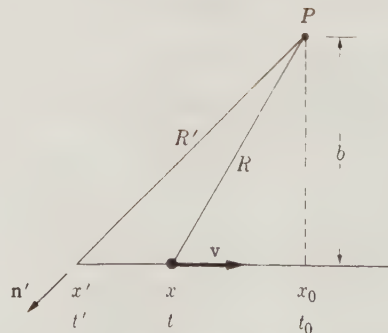
Even after Eq. (21-17) is substituted in Eq. (21-16), a multitude of variables appear in the expression for φ . In calculating the electric field by taking the gradient of φ and the time derivative of \mathbf{A} , these variables would have to be differentiated very carefully and would cause the calculation to be quite cumbersome. Rather than follow this procedure, it is preferable to eliminate the undesirable variables in φ and \mathbf{A} and obtain an expression that involves only the coordinates of P , the present time t , and parameters that describe the path of the charged particle.

Since the charge moves from x' to x_0 in time $t_0 - t'$, it is clear that

$$c^2(t - t')^2 = v^2(t_0 - t')^2 + b^2 \tag{21-18}$$

FIGURE 21-1

Diagram for calculating the electric field of a moving point charge



If this equation is solved for t' the result is

$$t' = \frac{c^2 t - v^2 t_0 \pm \sqrt{v^2 c^2 (t_0 - t)^2 + b^2 (c^2 - v^2)}}{c^2 - v^2} \quad (21-19)$$

The minus sign must be used in this equation to ensure that t' is retarded with respect to t , as can be verified by observing that at $t = t_0 = 0$, $t' = \pm \sqrt{b^2 (c^2 - v^2)} / (c^2 - v^2)$ and hence only the negative sign gives an earlier time. Having found t' , we find $x_0 - x'$ from

$$\begin{aligned} x_0 - x' &= v(t_0 - t') \\ &= v \left(\frac{t_0 (c^2 - v^2) - c^2 t + v^2 t_0 + \sqrt{v^2 c^2 (t_0 - t)^2 + b^2 (c^2 - v^2)}}{c^2 - v^2} \right) \end{aligned} \quad (21-20)$$

while R' is shown to be

$$R' = c \left(\frac{t(c^2 - v^2) - c^2 t + v^2 t_0 + \sqrt{v^2 c^2 (t_0 - t)^2 + b^2 (c^2 - v^2)}}{c^2 - v^2} \right) \quad (21-21)$$

Equations (21-20) and (21-21) may be used to evaluate the denominator which appears in Eq. (21-16). This denominator becomes

$$R^* = R' - \frac{v}{c} (x_0 - x') \quad (21-22)$$

through the use of Eq. (21-17), and subsequently, it becomes

$$\begin{aligned} R^* &= (c^2 - v^2)^{-1} \left[v^2 c (t_0 - t) + c \sqrt{v^2 c^2 (t_0 - t)^2 + b^2 (c^2 - v^2)} \right. \\ &\quad \left. - v^2 c (t_0 - t) - \frac{v^2}{c} \sqrt{v^2 c^2 (t_0 - t)^2 + b^2 (c^2 - v^2)} \right] \\ &= \sqrt{v^2 (t_0 - t)^2 + b^2 (1 - v^2/c^2)} \end{aligned} \quad (21-23)$$

through Eqs. (21-20) and (21-21). The scalar potential is

$$\varphi(P, t) = \frac{q}{4\pi\epsilon_0} \frac{1}{\sqrt{v^2 (t_0 - t)^2 + b^2 (1 - v^2/c^2)}} = \frac{q}{4\pi\epsilon_0 R^*} \quad (21-24)$$

while the vector potential is

$$\mathbf{A}(P, t) = \frac{\mu_0 q}{4\pi} \frac{\mathbf{v}}{\sqrt{v^2 (t_0 - t)^2 + b^2 (1 - v^2/c^2)}} = \frac{\mu_0 q}{4\pi} \frac{\mathbf{v}}{R^*} \quad (21-25)$$

It is important to realize that Eqs. (21-24) and (21-25) contain only the position and time of the observation point, and the parameters (\mathbf{v}, t_0) that describe the path of the charged particle.

To make this statement more concrete and to put the potentials in a form more suitable for calculating the fields, the coordinate system must be fixed more carefully. Since the charge moves along the x -axis and since this axis is one of symmetry for the problem, it is only necessary to specify the origin on the x -axis. It is convenient to take $x = 0$ as the position of the charge at $t = 0$. Then $x = vt$ and, in particular, $x_0 = vt_0$.

If the point P is specified by the cartesian coordinates ξ, η, ζ , then

$$\xi = x_0 = vt_0 \quad \text{and} \quad \eta^2 + \zeta^2 = b^2 \quad (21-26)$$

Using these results in Eq. (21-25) and letting $\xi = (\xi, \eta, \zeta)$, we obtain

$$\varphi(\xi, t) = \frac{q}{4\pi\epsilon_0} \frac{1}{\sqrt{(\xi - vt)^2 + (\eta^2 + \zeta^2)(1 - v^2/c^2)}} \quad (21-27)$$

and

$$\mathbf{A}(\xi, t) = \frac{\mu_0 q}{4\pi} \frac{\mathbf{v}}{\sqrt{(\xi - vt)^2 + (\eta^2 + \zeta^2)(1 - v^2/c^2)}}$$

It must be remembered that these equations apply only if \mathbf{v} is along the x -axis; other directions require modification of the formulas.

The important thing about Eqs. (21-27) is that they are in a form ideally suited for the calculation of the fields. Thus

$$\begin{aligned} \mathbf{E}(\xi, t) &= -\frac{\partial \mathbf{A}}{\partial t} - \nabla_{\xi} \varphi \\ &= -\frac{\mu_0 q}{4\pi} \mathbf{v} \frac{v(\xi - vt)}{R^{*3}} \\ &\quad + \frac{q}{4\pi\epsilon_0} \frac{1}{R^{*3}} [(\xi - vt)\mathbf{i} + \eta(1 - v^2/c^2)\mathbf{j} + \zeta(1 - v^2/c^2)\mathbf{k}] \end{aligned} \quad (21-28)$$

Noting that $\mathbf{v} = v\mathbf{i}$, $\epsilon_0\mu_0 = 1/c^2$, and $\xi - vt = x_0 - x$ makes it possible to rewrite Eq. (21-28) in the form

$$\mathbf{E}(\xi, t) = \frac{q}{4\pi\epsilon_0} \frac{\mathbf{R}}{R^{*3}} (1 - v^2/c^2) \quad (21-29)$$

where \mathbf{R} is a vector from the position of the charge at time t to the point P and

$$R^* = R' - \frac{\mathbf{R}' \cdot \mathbf{v}}{c}$$

The magnetic induction can be found by evaluating $\mathbf{B} = \nabla \times \mathbf{A}$; however, a much simpler procedure is to note that

$$\mathbf{A} = \mu_0\epsilon_0 \mathbf{v} \varphi \quad (21-30)$$

and hence that

$$\mathbf{B} = \mu_0 \epsilon_0 \nabla \times (\mathbf{v}\varphi) = -\mu_0 \epsilon_0 \mathbf{v} \times \nabla\varphi \quad (21-31)$$

Since \mathbf{v} is along the x -axis, only the y and z components of $\nabla\varphi$ are important in the cross product. These components are just the negatives of the y and z components of \mathbf{E} . In this way, we find

$$\mathbf{B} = \frac{1}{c^2} \mathbf{v} \times \mathbf{E} \quad (21-32)$$

which completes the computation of the fields.

It is interesting to note that although the source of the field is the retarded position, lines of \mathbf{E} are directed away from the instantaneous position of the charge. The lines of \mathbf{B} are circles with centers on the charge path. The E -field is not spherically symmetric as it is in the static case, but is stronger in the direction perpendicular to the velocity. (See Problem 21-1.)

Having obtained the field vectors, we are in a position to calculate other electromagnetic quantities; however, rather than pursue these possibilities, we refer the reader to more advanced texts* that deal at length with such problems.

THE FIELD OF AN ACCELERATED POINT CHARGE

If an accelerated point charge is to be considered, certain simplifications that appear in the constant-velocity case are no longer possible. The major difficulty here is a direct result of the fact that the Lienard-Wiechert potentials can no longer be expressed in terms of the present position of the charge; instead, the retarded position and time appear explicitly. The potentials

$$\varphi = \frac{q}{4\pi\epsilon_0} \left[\frac{1}{R(1 + \mathbf{v} \cdot \mathbf{n}/c)} \right]_{\text{ret}} \quad (21-33)$$

and

$$\mathbf{A} = \frac{q}{4\pi\epsilon_0} \left[\frac{\mathbf{v}/c^2}{R(1 + \mathbf{v} \cdot \mathbf{n}/c)} \right]_{\text{ret}}$$

are still correct. However, in differentiating them to obtain the fields, it must be noted that derivatives with respect to the position of the field point must be taken at constant observation time, and also derivatives with respect to the observation time at fixed field points. Since the retarded time appears explicitly in the potentials, care must be used to obtain the correct derivatives.

To clarify the differentiation problem, we note that the potentials are

* For example, J. D. Jackson, *Classical Electrodynamics*, 2nd ed. (New York: Wiley, 1975).

functions of the field point ξ , the observation time t , the retarded position \mathbf{r}' of the charge, and the retarded time t' . The trajectory of the particle is specified by giving \mathbf{r}' as a function of t' , so that the dependence on \mathbf{r}' can be removed. Furthermore, the retardation condition

$$(\xi - x')^2 + (\eta - y')^2 + (\zeta - z')^2 = c^2(t - t')^2 \quad (21-34)$$

provides a single relationship among the remaining variables. Thus, it is clear that although the potentials depend superficially on eight variables, only four of these are really independent. In computing the fields \mathbf{E} and \mathbf{B} it is necessary to differentiate the potentials with respect to each of ξ , η , ζ , and t , holding the other three fixed; for example, \mathbf{A} must be differentiated with respect to t holding ξ , η , and ζ constant. Since it is t' that appears explicitly in the potentials, the calculation of these derivatives causes some difficulty.

To keep track of the variables that are being held constant during various differentiations, the following notation will be adopted: A partial derivative in which all other variables, dependent or independent, are held constant will be designated by the usual partial derivative symbol. If not all other variables are held constant, then those that are will be indicated by subscripts. Thus the derivative of \mathbf{A} that is needed in computing \mathbf{E} is $(\partial\mathbf{A}/\partial t)_\xi$, while those of φ are $(\partial\varphi/\partial\xi)_{\eta,\zeta,t}$, etc. To transform $(\partial\mathbf{A}/\partial t)_\xi$ into a derivative with respect to t' , we write

$$\left(\frac{\partial\mathbf{A}}{\partial t}\right)_\xi = \left(\frac{\partial\mathbf{A}}{\partial t}\right) + \left(\frac{\partial\mathbf{A}}{\partial t'}\right)\left(\frac{\partial t'}{\partial t}\right)_\xi \quad (21-35)$$

and

$$\left(\frac{\partial\mathbf{A}}{\partial t'}\right)_\xi = \left(\frac{\partial\mathbf{A}}{\partial t'}\right) + \left(\frac{\partial\mathbf{A}}{\partial t}\right)\left(\frac{\partial t}{\partial t'}\right)_\xi \quad (21-36)$$

The retardation condition, Eq. (21-34), together with the equation specifying the trajectory, $\mathbf{x}' = \mathbf{x}'(t')$, is equivalent to an equation of the form

$$f(\xi, t, t') = 0$$

This relationship implies that $(\partial t/\partial t')_\xi = 1/(\partial t'/\partial t)_\xi$, which, when combined with Eqs. (21-35) and (21-36), gives

$$\left(\frac{\partial\mathbf{A}}{\partial t}\right)_\xi = \left(\frac{\partial\mathbf{A}}{\partial t'}\right)_\xi \left(\frac{\partial t'}{\partial t}\right)_\xi \quad (21-37)$$

In calculating the time derivatives of the potentials, this equation is just what is required to get the electric and magnetic fields. The other derivatives are all of the form $(\partial\varphi/\partial\xi)_t$. Such derivatives are readily evaluated by noting that

$$\left(\frac{\partial\varphi}{\partial\xi}\right)_t = \left(\frac{\partial\varphi}{\partial\xi}\right)_{t,t'} + \left(\frac{\partial\varphi}{\partial t'}\right)_{\xi,t} \left(\frac{\partial t'}{\partial\xi}\right)_t \quad (21-38)$$

in which all of the subscripts have been included to avoid any possibility of confusion.

From Eqs. (21-37) and (21-38) it is clear that the calculation of \mathbf{E} requires that the derivatives $(\partial t'/\partial t)_\xi$ and $(\partial t'/\partial \xi)_t$ must be evaluated. Each of these may be easily evaluated by differentiating the square root of Eq. (21-34),

$$[(\xi - x')^2 + (\eta - y')^2 + (\zeta - z')^2]^{1/2} = c(t - t') \quad (21-39)$$

in the appropriate way. If the derivative with respect to t (holding ξ constant) is taken, the equation

$$-\frac{1}{R'} \mathbf{R}' \cdot \left(\frac{\partial \mathbf{r}'}{\partial t} \right)_\xi = c \left[1 - \left(\frac{\partial t'}{\partial t} \right)_\xi \right] \quad (21-40)$$

results. In this equation, $\mathbf{r}' = \mathbf{i}x' + \mathbf{j}y' + \mathbf{k}z'$ and $\mathbf{R}' = \xi - \mathbf{r}'$. Since \mathbf{r}' depends explicitly only on t' , the derivative on the left is easily changed, to give

$$-\frac{1}{R'} \mathbf{R}' \cdot \mathbf{v}' \left(\frac{\partial t'}{\partial t} \right)_\xi = c \left[1 - \left(\frac{\partial t'}{\partial t} \right)_\xi \right] \quad (21-41)$$

where $\mathbf{v}' = \partial \mathbf{r}'/\partial t'$ is the velocity of the charge at the retarded time t' . Solving this equation for $(\partial t'/\partial t)_\xi$ leads to

$$\left(\frac{\partial t'}{\partial t} \right)_\xi = \frac{R'}{R' - \mathbf{R}' \cdot \mathbf{v}'/c} = \frac{R'}{R^*} \quad (21-42)$$

A similar calculation in which Eq. (21-39) is differentiated with respect to ξ at constant (η, ζ, t) gives

$$\left(\frac{\partial t'}{\partial \xi} \right)_t = -\frac{(\xi - x')}{(R' - \mathbf{R}' \cdot \mathbf{v}'/c)c} \quad (21-43)$$

Computing the other two components and writing the result as a vector equation, we obtain

$$(\nabla_\xi t')_t = -\frac{\mathbf{R}'/c}{R' - \mathbf{R}' \cdot \mathbf{v}'/c} = \frac{\mathbf{R}'}{-R^*c} \quad (21-44)$$

With these derivatives at hand, the electric field due to an accelerated point charge is readily computed from the Lienard-Wiechert potentials. Thus,

$$\begin{aligned} \mathbf{E}(\xi, t) &= -(\nabla_\xi \varphi)_t - \left(\frac{\partial \mathbf{A}}{\partial t} \right)_\xi \\ &= -(\nabla_\xi \varphi)_{t'} - \left(\frac{\partial \varphi}{\partial t'} \right)_{\xi t'} (\nabla_\xi t')_t - \left(\frac{\partial \mathbf{A}}{\partial t'} \right)_\xi \left(\frac{\partial t'}{\partial t} \right)_\xi \end{aligned} \quad (21-45)$$

The derivatives of the potentials that appear in this equation are easily

found to be

$$(\nabla_{\xi}\varphi)_{u'} = -\frac{q}{4\pi\epsilon_0} \frac{\mathbf{R}'/R' - \mathbf{v}'/c}{(R' - \mathbf{R}' \cdot \mathbf{v}'/c)^2} \quad (21-46)$$

$$\left(\frac{\partial\varphi}{\partial t'}\right)_{\xi} = \frac{q}{4\pi\epsilon_0} \left[\frac{\mathbf{R}' \cdot \mathbf{v}'}{R'} - \frac{v'^2}{c} + \frac{\mathbf{R}' \cdot \dot{\mathbf{v}}'}{c} \right] \frac{1}{R'^2} \quad (21-47)$$

$$\left(\frac{\partial\mathbf{A}}{\partial t'}\right)_{\xi} = \frac{q}{4\pi\epsilon_0} \left[\frac{\dot{\mathbf{v}}'}{R'^*c^2} + \frac{\mathbf{v}'}{c^2} \frac{1}{R'^*2} \left(\frac{\mathbf{R}' \cdot \mathbf{v}'}{R'} - \frac{v'^2}{c} + \frac{\mathbf{R}' \cdot \dot{\mathbf{v}}'}{c} \right) \right] \frac{1}{R'^*2} \quad (21-48)$$

Using these results in Eq. (21-45), we find

$$\begin{aligned} \mathbf{E}(\xi, t) = & \frac{q}{4\pi\epsilon_0} \left[\frac{1}{R'^*3} \left(\mathbf{R}' - \frac{R'\mathbf{v}'}{c} \right) \left(1 - \frac{v'^2}{c^2} \right) \right. \\ & \left. + \left(\mathbf{R}' - \frac{R'\mathbf{v}'}{c} \right) \frac{\dot{\mathbf{v}} \cdot \mathbf{R}'}{R'^*3c^2} - \frac{\dot{\mathbf{v}}'R'}{R'^*2c^2} \right] \end{aligned} \quad (21-49)$$

A similar calculation gives

$$\begin{aligned} \mathbf{B}(\xi, t) = & \frac{q}{4\pi\epsilon_0c^2} \left\{ \frac{\mathbf{v}' \times \mathbf{R}'}{R'^*3} \left(1 - \frac{v'^2}{c^2} \right) \right. \\ & \left. + \frac{1}{R'^*3c} \frac{\mathbf{R}'}{R'} \times \left[\mathbf{R}' \times \left(\left[\mathbf{R}' - \frac{R'\mathbf{v}'}{c} \right] \times \dot{\mathbf{v}}' \right) \right] \right\} \end{aligned} \quad (21-50)$$

These results may be used to explain many such important phenomena as radiation damping and the classical Bremsstrahlung. Most of these calculations are available in more advanced texts on electrodynamics and except for one example will be omitted here.

21-4

RADIATION FIELDS FOR SMALL VELOCITIES

The result of the calculation of the fields from the potentials for an arbitrarily moving point charge, which was carried out in the last section, may be rewritten as

$$\begin{aligned} \mathbf{E}(\xi, t) = & \frac{q}{4\pi\epsilon_0} \frac{1}{R'^*3} \left\{ \left(\mathbf{R}' - \frac{R'\mathbf{v}'}{c} \right) \left(1 - \frac{v'^2}{c^2} \right) \right. \\ & \left. + \frac{1}{c^2} \mathbf{R}' \times \left[\left(\mathbf{R}' - \frac{R'\mathbf{v}'}{c} \right) \times \dot{\mathbf{v}}' \right] \right\} \end{aligned} \quad (21-51)$$

$$\mathbf{B}(\xi, t) = \frac{\mathbf{R}' \times \mathbf{E}}{R'c} \quad (21-52)$$

From Eq. (21-52) we note that the B -field of a point charge in vacuum is

always perpendicular to the E -field at the same point and time and is also perpendicular to the line connecting the field point with the *retarded* position of the particle \mathbf{R}' . The fields contain two terms, with the second one proportional to the acceleration $\dot{\mathbf{v}}'$ and the first one independent of it. Both terms depend on v'/c . For uniform motion ($\dot{\mathbf{v}}' = 0$), the first term in Eq. (21-51) gives the previous result (21-29), since $R'/c = t - t'$. Even for nonuniform motion at high velocity, the first term does not contribute to the radiation from the charge, because its magnitude falls off with distance like $1/R'^2$. We have seen previously that the fields must fall off like $1/R'$ in order to contribute to the Poynting vector at large distances; the second, acceleration-dependent term does decrease like $1/R'$. We see that this part of the E -field is similarly perpendicular to \mathbf{R}' : In the radiation field, \mathbf{E} , \mathbf{B} , and \mathbf{R}' are mutually perpendicular, and $\mathbf{S} = (1/\mu_0 c)E^2(\mathbf{R}'/R')$.

For simplicity, we consider here only the case of a slowly moving charge. If the velocity of the charge is small compared with the velocity of light, that is, if $v'/c \ll 1$, then the approximations

$$\mathbf{R}' - \frac{R'\mathbf{v}'}{c} \approx \mathbf{R}'$$

and

$$R^* = R' - \frac{\mathbf{R}' \cdot \mathbf{v}'}{c} \approx R'$$

may be made in Eqs. (21-51) and (21-52). If, in addition, only the radiation field—that is, the part of the field proportional to $1/R'$ —is considered, then Eqs. (21-51) and (21-52) become

$$\mathbf{E}(\xi, t) = \frac{q}{4\pi\epsilon_0} \frac{\mathbf{R}' \times (\mathbf{R}' \times \dot{\mathbf{v}}')}{R'^3 c^2} \quad (21-53)$$

and

$$\mathbf{B}(\xi, t) = \frac{q}{4\pi\epsilon_0} \frac{\mathbf{R}' \times [\mathbf{R}' \times (\mathbf{R}' \times \dot{\mathbf{v}}')]}{R'^4 c^3} = \frac{q}{4\pi\epsilon_0 c^2} \frac{\dot{\mathbf{v}}' \times \mathbf{R}'}{R'^2 c} \quad (21-54)$$

From these field vectors the Poynting vector is found to be

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} = \frac{q^2}{16\pi^2 \epsilon_0^2 \mu_0 c^2} \frac{1}{R'^5 c^3} [\mathbf{R}' \times (\mathbf{R}' \times \dot{\mathbf{v}}')] \times [\dot{\mathbf{v}}' \times \mathbf{R}'] \quad (21-55)$$

which, through the use of vector identities, reduces to

$$\mathbf{S} = \frac{q^2}{16\pi^2 \epsilon_0 c^3} \frac{\mathbf{R}'(\mathbf{R}' \times \dot{\mathbf{v}}')^2}{R'^5} \quad (21-56)$$

The total radiated power is obtained by integrating this Poynting vector over a closed surface surrounding the charge. A convenient choice for such a surface is a sphere centered at the retarded position of the charge. If,

furthermore, the z -axis is chosen in the direction of $\dot{\mathbf{v}}'$, then

$$\begin{aligned} P_R &= -\frac{dW}{dt} = \int_S \mathbf{S} \cdot \mathbf{n} da \\ &= \frac{q^2}{16\pi^2\epsilon_0 c^3} \int \frac{R'^2 \dot{\mathbf{v}}'^2 \sin^2 \theta}{R'^5} \mathbf{R}' \cdot \frac{\mathbf{R}'}{R'} R'^2 \sin \theta d\theta d\phi \end{aligned} \quad (21-57)$$

from which one readily obtains

$$P_R = -\frac{dW}{dt} = \frac{q^2}{4\pi\epsilon_0} \frac{2}{3} \frac{\dot{\mathbf{v}}'^2}{c^3} \quad (21-58)$$

for the power radiated from a slowly moving, accelerated charge, in agreement with our earlier result of Eq. (20-44).

This completes our brief survey of radiation from moving charges. The basic ideas have been presented and some elementary applications have been given in detail. For the details of other calculations, reference should be made to various published works, and particularly to *Classical Electricity and Magnetism*, 2nd ed. by W.K.H. Panofsky and M. Phillips (Reading, Mass: Addison-Wesley, 1962); *The Classical Theory of Fields*, 2nd ed. by L. D. Landau and E. M. Lifshitz (Reading Mass.: Addison-Wesley, 1962); and *Classical Electrodynamics*, 2nd ed., by J. D. Jackson (New York: Wiley, 1975).

21-5 SUMMARY

The scalar and vector potentials of a *fast moving point charge* are obtained from the integral solutions for the **retarded potentials**. The results at the point P and time t are

$$\begin{aligned} \varphi(P, t) &= \frac{q}{4\pi\epsilon_0} \frac{1}{R^*} \\ \mathbf{A}(P, t) &= \frac{\mathbf{v}'}{c^2} \varphi(P, t) \end{aligned}$$

where

$$R^* = R' - \mathbf{R}' \cdot \mathbf{v}'/c$$

and where \mathbf{R}' is the vector to the field point P from the position of the particle of charge q at the *retarded* time $t' = t - R'/c$, and \mathbf{v}' is the particle velocity at the *retarded* time. The **fields** are obtained from the potentials by

differentiating them correctly:

$$\mathbf{E}(P, t) = \frac{q}{4\pi\epsilon_0 R^{*3}} \left\{ (\mathbf{R}' - R'\mathbf{v}'/c)(1 - v'^2/c^2) + \frac{1}{c^2} \mathbf{R}' \right. \\ \left. \times \left[\left(\mathbf{R}' - R' \frac{\mathbf{v}'}{c} \right) \times \dot{\mathbf{v}}' \right] \right\}$$

$$\mathbf{B}(P, t) = \mathbf{R}' \times \mathbf{E}(P, t)/R'c$$

- The E -field of a **uniformly moving** ($\dot{\mathbf{v}} = 0$) **point charge** is symmetric about a plane perpendicular to \mathbf{v} through the *present* position of the charge at t ; it is stronger in the direction perpendicular to \mathbf{v} than in the direction along \mathbf{v} . It is always pointing away from the *present* position of the charge.

- Only the **acceleration terms contribute to radiation**. In the radiation field, \mathbf{R}' , \mathbf{E} , and \mathbf{B} are mutually perpendicular. For $v'/c \ll 1$, the result is the same as the one derived in Chapter 20.

PROBLEMS

21-1 (a) For a charge in uniform motion with velocity \mathbf{v} , show that

$$R^* = R \sqrt{1 - \left(\frac{v}{c} \sin \theta \right)^2}$$

where \mathbf{R} is the vector from the present position of the charge to the field point and $\mathbf{R} \cdot \mathbf{v} = Rv \cos \theta$. (b) Plot a polar graph of $|\mathbf{E}|$ as a function of θ for a fixed R , assuming $v/c = 0.8$.

21-2 Use the result of Problem 21-1(a) to find the B -field of a charge dq moving uniformly with velocity v . Letting $v dq = I dx$, where I is the current in a long straight wire, integrate $d\mathbf{B}$ over the length of the wire. Note that the result agrees with that obtained from Ampere's law, even for $v/c \approx 1$.

21-3 Suppose the acceleration of a fast particle is in the same direction as its velocity. Show that the radiation is zero along the direction of motion.

The Special Theory of Relativity

As discussed in Chapters 2 and 8, the interaction between groups of charges (or currents) is usually described by separating the phenomenon into two parts: (1) the establishment of an electromagnetic field by the source and (2) the interaction of a second group of charges (and/or currents) with the field. The field itself may be sampled by an observer, using test charges and currents. This decomposition of the interaction is not unique; in fact, the detailed nature of the electromagnetic field depends on the state of motion of the observer.

For example, consider two observers, *A* and *B*. Observer *A* is at rest relative to a group of fixed charges and sees only an electric field associated with them. Observer *B* is in motion relative to *A*; Observer *B*, therefore, sees a group of moving charges and hence a magnetic as well as an electric field.

Are both observers entitled to use Maxwell's equations to describe their physical observations? And, if so, how does an observer transform the electric and magnetic fields in one frame of reference to give the field components in a second reference frame moving relative to the first? These are questions that we shall attempt to answer in this chapter.

22-1

PHYSICS BEFORE 1900

Maxwell's basic ideas about the electromagnetic field were first published in 1862. In the next forty years, the mathematical structure of the laws of electricity and magnetism was gradually developed (particularly by H. A. Lorentz), and many consequences of the theory were observed experimentally. Yet there were a number of problems that bothered theoretical physicists, particularly those concerned with the mathematical structure of physical laws.

FIGURE 22-1

Two coordinate systems moving relative to one another (in the x direction) with uniform velocity \mathbf{u}



All previous experience with wave motion indicated that a medium was required for the propagation of waves. In Chapter 16, we explained that Maxwell's equations in free space are compatible with, and in fact lead to, a wave equation in which the waves are propagated with velocity $c = 1/\sqrt{\epsilon_0\mu_0}$. Hence, it was natural to assume that some sort of ethereal medium pervaded all space (including vacuum) for the propagation of electromagnetic waves. Maxwell, himself, felt a need for this medium and called it the *ether*. But the existence of a medium poses a problem because it introduces a preferred reference frame, namely, the one in which the medium is at rest.

It was known that Newton's laws of motion were unaffected by a *Galilean transformation*—that is, by a coordinate transformation between two frames of reference (coordinate systems) in relative motion. For example, let Σ be a coordinate system at rest, and let Σ' be a coordinate system moving in the x -direction with uniform velocity \mathbf{u} (see Fig. 22-1). The relation between the coordinates and times in the two systems is given by (a Galilean transformation)

$$x' = x - ut, \quad y' = y, \quad z' = z, \quad t' = t \quad (22-1)$$

Newton's fundamental laws of motion are of the same form in Σ and Σ' . In fact, it is impossible to determine the absolute velocity of any reference frame by mechanical experiments.

What happens to Maxwell's equations under a Galilean transformation? We are not in a position to answer this question because we do not yet know how the fields transform, but we can sidestep the question by looking instead at the scalar wave equation (which is homogeneous and contains only one field component). In free space,

$$\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} + \frac{\partial^2 \varphi}{\partial z^2} = \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} \quad (22-2)$$

where φ stands for one of the field components. If we substitute

$$\frac{\partial}{\partial t} = \frac{\partial x'}{\partial t} \frac{\partial}{\partial x'} + \frac{\partial t'}{\partial t} \frac{\partial}{\partial t'} + \dots$$

into Eqs. (22-2), using Eq. (22-1) to evaluate $\partial x'/\partial t$, and so on, we find that the transformed wave equation is no longer of the same form as Eq. (22-2). This mathematical statement corresponds to the fact, which we know to be true for mechanical waves, that the wave motion is propagated with a fixed velocity relative to the stationary medium. But in a reference frame moving with respect to the medium, the wave propagation appears more complicated.

What, then, is the situation with regard to the laws of electricity and magnetism? A number of possibilities were proposed before the subject was clarified by Lorentz and Poincaré, and particularly by Einstein in 1905. Briefly stated, these possibilities are

1. Maxwell's equations are inadequate to explain electromagnetic phenomena.
2. There is a preferred reference frame, that of the stationary ether. Maxwell's equations require modification in other reference frames.
3. Maxwell's equations have the same form in all reference frames that are moving with uniform velocity relative to one another. The Galilean transformation is not adequate to connect different reference frames when electromagnetic fields are involved.

As we know, the third choice presented here is the correct one, and it is, in fact, a partial statement of the *principle of relativity*. The predictions of Maxwell's equations have been verified experimentally, at least in our terrestrial environment, and attempts to measure an absolute ether frame have not been successful. There is no *single* experiment that killed the "ether hypothesis" and forced us to accept relativity, but the combined results of a large number of experiments turned out to be compatible with no other possibility. Three fundamental experiments are*

1. The aberration of starlight (the small shift in apparent position of distant stars in the direction of the earth's orbital motion).
2. Measurement of the velocity of light in moving fluids (Fizeau, 1859).
3. The Michelson-Morley experiment (1887).

* The reader who would like to pursue the history of this subject in more detail is referred to R. S. Shankland, "Michelson-Morley Experiment," *American Journal of Physics*, vol. 32, p. 16 (1964); A. Einstein et al., *The Principle of Relativity* (New York: Dodd, Mead, 1923); E. T. Whittaker, *History of the Theories of Aether and Electricity*, vol. II (New York: Philosophical Library, 1951).

The Michelson-Morley experiment attempted to measure the velocity of the earth, relative to an absolute frame (in which light waves are propagated with velocity c). The results of the experiment indicated either that there is no preferred frame or that the earth is always in the preferred reference frame. This experiment in itself would appear to negate the absolute ether-frame hypothesis, since the earth is continually changing its velocity in its motion around the sun. However, it would be possible for the earth to remain in the preferred frame if it dragged the ether along with it—that is, a massive celestial object such as the earth could perhaps drag the ether along with it in its motion.

On the other hand, the first two experiments cited are not compatible with an "ether drag." From measurements made during the course of one year, it is seen that the apparent position of a star traces out a small elliptical path on the celestial sphere, the angular shift in position being of the order of v/c , where v is the earth's orbital speed. This aberration of starlight would presumably be absent if the ether were dragged along with the earth. The experiments with moving fluids can be made compatible with an ether-drag hypothesis only if one assumes (rather artificially) that objects less massive than the earth are only partially successful in dragging the ether along with them.

THE LORENTZ TRANSFORMATION AND EINSTEIN'S POSTULATES OF SPECIAL RELATIVITY

In 1904 H. A. Lorentz discovered a curious and remarkable transformation that leaves the form of Maxwell's equations unaltered, provided the field components are suitably changed. Let us again consider two coordinate systems Σ and Σ' that are moving relative to one another in the x direction with uniform velocity \mathbf{u} (see Fig. 22-1). Instead of the Galilean transformation, we now assume (a Lorentz transformation)

$$\begin{aligned}x' &= \frac{1}{\sqrt{1 - u^2/c^2}}(x - ut) \\y' &= y \\z' &= z \\t' &= \frac{1}{\sqrt{1 - u^2/c^2}}\left(t - \frac{u}{c^2}x\right)\end{aligned}\tag{22-3}$$

Again, we shall sidestep the question of how the electric and magnetic fields transform and look at the wave equation, Eq. (22-2). Let us substitute

$$\begin{aligned}\frac{\partial}{\partial x} &= \frac{\partial x'}{\partial x} \frac{\partial}{\partial x'} + \frac{\partial t'}{\partial x} \frac{\partial}{\partial t'} \\ \frac{\partial}{\partial y} &= \frac{\partial y'}{\partial y} \frac{\partial}{\partial y'}\end{aligned}$$

$$\frac{\partial}{\partial z} = \frac{\partial z'}{\partial z} \frac{\partial}{\partial z'}$$

$$\frac{\partial}{\partial t} = \frac{\partial x'}{\partial t} \frac{\partial}{\partial x'} + \frac{\partial t'}{\partial t} \frac{\partial}{\partial t'}$$

into the wave equation, using Eqs. (22-3) to evaluate the partial derivatives ($\partial x'/\partial x$), and so forth. For example,

$$\frac{\partial x'}{\partial x} = \frac{1}{\sqrt{1 - u^2/c^2}}, \quad \frac{\partial x'}{\partial t} = -\frac{u}{\sqrt{1 - u^2/c^2}}$$

If we make the indicated substitutions, combine terms, and cancel common factors on each side of the equation, we obtain

$$\frac{\partial^2 \varphi}{\partial x'^2} + \frac{\partial^2 \varphi}{\partial y'^2} + \frac{\partial^2 \varphi}{\partial z'^2} = \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t'^2} \quad (22-4)$$

Since this equation is homogeneous in φ , it is reasonable to expect that we can replace φ by φ' (its value in the primed coordinate system) in Eq. (22-4) without disturbing the equality. Thus the form of the wave equation is invariant to a Lorentz transformation.

Although the Lorentz transformation provides a basis for the development of special relativity, the far-reaching consequences of relativity were not discovered by Lorentz. He still believed in the ether hypothesis at this time, and tried very hard to fit his newly discovered transformation into the ether picture of electromagnetism. The development of special relativity as we now know it was left to H. Poincaré and A. Einstein.

As early as 1899, and again in 1900 and in 1904, Poincaré suggested that the experimental result of Michelson and Morley (i.e., their failure to observe an absolute ether frame) was a manifestation of a general principle—that absolute motion cannot be detected by laboratory experiments of any kind—which implied that the laws of nature must be the same for two observers in uniform motion relative to each other. He called this the *Principle of Relativity*. Poincaré also concluded that a new type of dynamics would have to be developed, which would be characterized, among other things, by the rule that no velocity can exceed the velocity of light. In 1905 Einstein published his *Electrodynamics of Moving Bodies*, in which he developed the *special theory of relativity* from two basic postulates: (1) the relativity principle and (2) the constancy of the speed of light. Einstein derived the way in which various physical quantities had to transform in going from one reference frame to another and showed how Newton's laws of mechanics had to be modified.

The *Einstein postulates* are

1. *The laws of nature are the same in all coordinate systems moving with uniform motion relative to one another.*

2. The velocity of light in empty space is the same in all reference systems and is independent of the motion of the emitting body.

Again we consider two coordinate systems Σ and Σ' , which are moving relative to one another in the x direction with uniform velocity \mathbf{u} (see Fig. 22-1). At times $t = 0$ and $t' = 0$, the origins of the two systems coincide, and at that instant a light source at the common origin emits a light pulse. An observer in the unprimed system, with appropriate detectors at various distances from the origin, sees the light signal propagate out as a spherical wave front. The coordinates of a point (x, y, z) on the wave front satisfy

$$x^2 + y^2 + z^2 - c^2t^2 = 0 \quad (22-5)$$

whereas for a point (x_1, y_1, z_1) in advance of the wave front (at the same time t),

$$x_1^2 + y_1^2 + z_1^2 - c^2t^2 > 0 \quad (22-6)$$

and for a point (x_2, y_2, z_2) behind,

$$x_2^2 + y_2^2 + z_2^2 - c^2t^2 < 0 \quad (22-7)$$

An observer in the primed coordinate system also sees the light signal propagate outward. And according to Einstein's two postulates, the observer sees a spherical wave front propagated with velocity c . Thus Eqs. (22-5), (22-6), and (22-7) also hold in the primed coordinates. Since the primed and unprimed coordinates are presumably related by a linear transformation, we are led to the result*

$$x^2 + y^2 + z^2 - c^2t^2 = x'^2 + y'^2 + z'^2 - c^2t'^2 \quad (22-8)$$

with (x, y, z, t) an arbitrary *space-time point* and (x', y', z', t') its transform in the primed system. Equivalently we might have written

$$\begin{aligned} (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - c^2(\Delta t)^2 &= (\Delta x')^2 + (\Delta y')^2 \\ &\quad + (\Delta z')^2 - c^2(\Delta t')^2 \end{aligned} \quad (22-9)$$

for the relationship between an arbitrary *space-time interval* in Σ and the corresponding interval in Σ' .

Having found a quantity that is invariant to a change of reference frame, we now look for a transformation that will leave the "invariant quantity" unaffected. The *Lorentz transformation* is just such a transforma-

* We have not ruled out the possibility that

$$x^2 + y^2 + z^2 - c^2t^2 = K(u)(x'^2 + y'^2 + z'^2 - c^2t'^2)$$

with $K(u)$ a constant of proportionality depending on u . Such a scale change can be ruled out by considering the inverse transformation, however, so that $K(u) = 1$.

tion; the quantity

$$x^2 + y^2 + z^2 - c^2t^2$$

is unaffected by the Lorentz transformation, Eq. (22-3), as can be verified by direct substitution. Thus, the application of Einstein's two postulates leads directly to the Lorentz transformation.

If the Lorentz transformation is the proper one for transforming coordinates from one reference frame to another, then the more intuitive Galilean transformation, Eq. (22-1), clearly cannot be correct. The Galilean transformation is never precisely correct, but it does form a valid approximation in the limit when all velocities are small compared to the speed of light. Newtonian mechanics, too, must be modified since the correct laws of motion must transform properly under a Lorentz transformation, not under a Galilean transformation.

In the following sections, the relativistic transformation will be discussed in more detail and the transformation laws for other physical quantities will be obtained. Before proceeding, however, we pause to discuss three simple consequences of the Lorentz transformation: (1) modification of the concept of simultaneity, (2) the Lorentz contraction, and (3) the time dilation.

Two events occur *simultaneously* if they happen at the same time. Since the events may occur at widely separated spatial positions, such a statement implies that we have a way of synchronizing clocks so that each event may be timed separately. Now let us suppose that two events at positions x_1 and x_2 in the Σ -frame occur simultaneously—that is, the times t_1 and t_2 at which the two events occur are the same. But according to the Lorentz transformation, (22-3), the times in the Σ' -system are not the same:

$$t'_1 - t'_2 = \frac{(u/c^2)}{\sqrt{1 - (u/c)^2}} [x_2 - x_1] \quad (22-10)$$

Thus, we must modify our intuitive concept of simultaneity: If two events occur simultaneously in one reference frame, they are not necessarily simultaneous in another. Having accepted the Lorentz transformation, we necessarily have to give up the concept of a "universal time."

A simple example will perhaps make this point more plausible. Suppose observer A is traveling in a space ship with velocity u relative to an earthbound observer B . Observer A wants to do an experiment involving the simultaneous detection of a light signal at two different positions and so places one detector in the front of the space ship, the other near the rear, and carefully measures the distance between the two detectors. Observer A then places a light source midway between the two detectors. Since the light signal propagates in spherical waves from the source, A 's two detectors do in fact detect the signal simultaneously. But what about observer B ? In the earthbound frame, B again sees the light signal propagate out from the

source in spherical waves, but the detector at the front end of the space ship is moving away from the expanding wave front whereas the detector in the rear is moving toward it. Thus the detection is not simultaneous in B 's system.

The apparent contraction of a moving object in the direction of its motion is called the *Lorentz contraction*. In a length measurement, the length of the object to be measured is compared with a standard scale. This measurement offers no special problem if the object and scale are at rest relative to one another. Suppose, however, that an observer in the Σ -frame wanted to measure the length of a moving object (an object at rest in the Σ' -frame). Because the object is in motion relative to the observer and scale, it is important to compare the two ends of the object with the scale at the same time—that is, if position x_1 is determined at time t_1 , and x_2 at time t_2 , then t_1 must equal t_2 in order for the length measurement to be meaningful. But according to the Lorentz transformation, Eq. (22-3),

$$x'_1 - x'_2 = \frac{1}{\sqrt{1 - \beta^2}}(x_1 - x_2) \quad (22-11)$$

with $\beta \equiv u/c$. Now $l' = x'_1 - x'_2$ may be regarded as the “true” length of the object (its length measured by an observer at rest relative to it). Its apparent length (the length seen by an observer in the Σ system)

$$l = l'\sqrt{1 - \beta^2} \quad (22-12)$$

appears contracted. It is easy to verify that the transverse dimensions, those in the y and z directions, are unaffected by the motion.

The *time dilation*, the apparent slowing down of temporal events associated with a moving object, can best be obtained from Eq. (22-9). This equation may be written

$$\begin{aligned} (dt)^2 \left[c^2 - \left(\frac{dx}{dt} \right)^2 - \left(\frac{dy}{dt} \right)^2 - \left(\frac{dz}{dt} \right)^2 \right] \\ = (dt')^2 \left[c^2 - \left(\frac{dx'}{dt'} \right)^2 - \left(\frac{dy'}{dt'} \right)^2 - \left(\frac{dz'}{dt'} \right)^2 \right] \end{aligned} \quad (22-13)$$

Let Σ' be the *rest system* of the object—that is, the system in which the object is at rest. Then (velocity) $'_x \equiv (dx'/dt') = 0$, and so forth, and

$$u^2 = \left[\left(\frac{dx}{dt} \right)^2 + \left(\frac{dy}{dt} \right)^2 + \left(\frac{dz}{dt} \right)^2 \right]$$

Equation (22-13) becomes

$$\Delta t = \frac{\Delta t'}{\sqrt{1 - \beta^2}} \quad (22-14)$$

Thus, the apparent time interval Δt (the interval measured by an observer

in the Σ system) appears to be lengthened over the intrinsic time interval $\Delta t'$. Another way of interpreting Eq. (22-14) is to say that clocks appear to slow down when they are in motion relative to an observer.

22-3

GEOMETRY OF SPACE-TIME

The Lorentz transformation of the preceding section is a linear transformation connecting the space coordinates and time in one frame of reference to the corresponding quantities in a second frame, which is in uniform motion relative to the first one. It thus appears that we might be able to construct a four-dimensional geometry in which the space coordinates and time show up on an equal footing, with the Lorentz transformation acting as some kind of geometric operation in this four-dimensional space. Now we have noted that a certain quadratic function of the coordinates and time, namely,

$$x^2 + y^2 + z^2 - c^2t^2$$

is an invariant—that is, it has the same value in all reference frames. This quadratic form recalls to mind that the length of a vector, specifically the length l of the position vector, with

$$l^2 = x^2 + y^2 + z^2$$

is invariant to rotations of the coordinate axes in ordinary (three-dimensional) space. (See Appendix II.)

In trying to extend this formalism to four dimensions and treat $x^2 + y^2 + z^2 - c^2t^2$ as the square of a “length” in space-time, we run into the obvious problem that the fourth component, ct , enters the expression with a minus sign. Thus, space-time is basically a non-Euclidean four-dimensional space. We can get around many of the difficulties that this implies by defining four “coordinates” as

$$x_1 = x, \quad x_2 = y, \quad x_3 = z, \quad x_4 = ict \quad (22-15)$$

with i the unit imaginary number. This four-dimensional space (which was introduced by H. Minkowski) is not, strictly speaking, Euclidean because it involves an imaginary coordinate. Many of its properties may, however, be derived by treating it as a Euclidean space. This approach will be used here. The quantity

$$\sum_{\mu} x_{\mu}^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2$$

is invariant to certain transformations. These transformations (which include, of course, the Lorentz transformations) have many of the properties of rotations or orthogonal transformations, but because they may have imaginary components, they are properly called *complex orthogonal transformations*. In what follows, however, this distinction does not have

important consequences and the Lorentz transformation in Minkowski space will be treated as an orthogonal transformation.*

The quantity defined by (x_1, x_2, x_3, x_4) is a four-dimensional vector. We shall have occasion to define other four-component vectors (i.e., quantities whose components transform as (x_1, x_2, x_3, x_4) under a Lorentz transformation). Four-dimensional vectors are called *four vectors* or *world vectors* to distinguish them from ordinary three-dimensional vectors. A quantity that is unchanged by a Lorentz transformation is called a *world scalar*. A point in space-time is called a *world point*, and the trajectory of a particle in space-time is called a *world line*.

22-4

THE LORENTZ TRANSFORMATION AS AN ORTHOGONAL TRANSFORMATION

The results of the formalism of orthogonal transformations applied to ordinary three-dimensional vectors (which is developed in Appendix II) may be directly applied to four-dimensional space-time by adding the fourth component, $x_4 = ict$. All summations now go from 1 to 4. It is customary to use Greek indices to describe four-dimensional quantities and reserve Latin indices for three-dimensional entities. Thus F_i represents the i th component of an ordinary three-dimensional vector, but $T_{\mu\nu}$ represents the μ, ν th component of a four-dimensional tensor.

The Lorentz transformation (22-3) for transforming the unprimed to the

* Because of the common use of abstract spaces in contemporary physics, it is appropriate to point out the source of the difficulty in using "Euclidean" to describe Minkowski space and "orthogonal" to describe the Lorentz transformations. The source of the difficulty stems from the fact that both *Euclidean* and *orthogonal* represent ideas that were developed to deal with real variables. If one tries to generalize by admitting complex coordinates, then the most fruitful generalization of the length of a vector is

$$\left\{ \sum_i x_i^* x_i \right\}^{1/2}$$

with x_i^* the complex conjugate of x_i . Transformations which leave this length invariant are unitary transformations, characterized in the notation of Appendix II by $\sum_i a_{ij} a_{ik}^* = \delta_{jk}$. The Lorentz transformation does not fall into this classification and consequently requires an entirely separate but parallel development for its complete elucidation. There is one crucial difference between Minkowski space and either unitary or orthogonal space: For the last two spaces the length of any component of a vector is less than or equal to the length of the vector itself, while the "length" of a component of a four-vector in Minkowski space is not so restricted. Similarly, in unitary and orthogonal transformations, the magnitudes of all coefficients are less than or equal to unity, but this is not true for Lorentz transformations. These points are important, but further discussion here would lead us too far afield.

primed system of Fig. 22-1 may be written as

$$\begin{aligned}
 x'_1 &= \frac{1}{\sqrt{1-\beta^2}}x_1 + 0 \cdot x_2 + 0 \cdot x_3 + \frac{i\beta}{\sqrt{1-\beta^2}}x_4 \\
 x'_2 &= 0 \cdot x_1 + x_2 + 0 \cdot x_3 + 0 \cdot x_4 \\
 x'_3 &= 0 \cdot x_1 + 0 \cdot x_2 + x_3 + 0 \cdot x_4 \\
 x'_4 &= -\frac{i\beta}{\sqrt{1-\beta^2}}x_1 + 0 \cdot x_2 + 0 \cdot x_3 + \frac{1}{\sqrt{1-\beta^2}}x_4
 \end{aligned} \tag{22-16}$$

with $\beta \equiv u/c$. The matrix of this transformation is

$$\mathbf{A} = \begin{bmatrix} \frac{1}{\sqrt{1-\beta^2}} & 0 & 0 & i\frac{\beta}{\sqrt{1-\beta^2}} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\frac{\beta}{\sqrt{1-\beta^2}} & 0 & 0 & \frac{1}{\sqrt{1-\beta^2}} \end{bmatrix} \tag{22-17}$$

We can easily verify that Eq. (22-17) is an orthogonal transformation*—that is, its components satisfy (II-6) in Appendix II.

Matrix \mathbf{A} is particularly simple (it has only six nonzero entries) in this case because the Lorentz transformation relates two systems that are in relative motion along one of the coordinate axes (namely, the x -axis). Thus, x and t are transformed into x' and t' but the y and z directions are unaffected. In the general case, when the direction of relative motion is not along a coordinate axis, the transformation is more complicated, but the matrix components still satisfy the orthogonality relationships. Since the coordinate directions can generally be chosen to fit the needs of a particular problem, we shall restrict ourselves in this book to Lorentz transformations of the type (22-17), or equivalently to transformations linking the coordinate systems in Fig. 22-1.

The Lorentz transformation (22-16) can be interpreted as a rotation in the x_1x_4 -plane. If this is the case, then the rotation angle θ is determined from

$$x'_1 = x_1 \cos \theta + x_4 \sin \theta$$

or

$$\tan \theta = i\beta = i(u/c) \tag{22-18}$$

* We use the term "orthogonal" instead of the more accurate term "complex orthogonal"; cf. the discussion of Section 22-3.

Thus, the rotation angle is not a real angle.* Mathematically, the Lorentz transformation acts like a rotation in our orthogonal four-dimensional space, but a rotation through an imaginary angle.

The inverse Lorentz transformation—that is, the transformation taking us from the primed system to the unprimed system—is given by the transpose matrix to (22-17):

$$\bar{\mathbf{A}} = \begin{bmatrix} \frac{1}{\sqrt{1-\beta^2}} & 0 & 0 & -i\frac{\beta}{\sqrt{1-\beta^2}} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ i\frac{\beta}{\sqrt{1-\beta^2}} & 0 & 0 & \frac{1}{\sqrt{1-\beta^2}} \end{bmatrix} \quad (22-19)$$

22-5

COVARIANT FORM OF THE ELECTROMAGNETIC EQUATIONS

The fundamental equations of electromagnetic theory, the Maxwell equations, which we discussed in Chapter 16, are written in terms of time and space derivatives of the \mathbf{E} - and \mathbf{B} -fields. In the usual three-dimensional system, time enters the equations as a scalar but the three space derivatives enter in certain symmetric combinations (divergence or curl). We may display the symmetry more directly by writing the divergence equation (e.g., Gauss's law) as

$$\sum_i \frac{\partial E_i}{\partial x_i} = \frac{1}{\epsilon_0} \rho \quad (22-20)$$

and the curl equation (e.g., Ampere's law) as

$$\frac{\partial B_j}{\partial x_i} - \frac{\partial B_i}{\partial x_j} = \mu_0 J_k + \mu_0 \epsilon_0 \frac{\partial E_k}{\partial t} \quad (22-21)$$

This last equation actually represents three equations (the three components of the vector curl equation) with i, j, k standing for a cyclic permutation of x, y, z .

In the preceding sections, however, we have noted that the Lorentz transformation mixes up the space coordinates and the time and may be regarded as a rotation in $x_1 x_2 x_3 x_4$ -space. Thus, x_1, x_2, x_3 , and x_4 should enter the Maxwell equations in a symmetrical way. We should, in fact, be able to write the Maxwell equations in terms of four-dimensional curls and divergences. A formulation of the equations of electricity and magnetism

* This reflects the unboundedness of the transformation alluded to earlier.

that treats space coordinates and time on an equal footing is called a *covariant* formulation. We must proceed with a certain amount of caution, however, because a vector quantity in three dimensions does not necessarily become part of a four-dimensional vector.

We start with the equation of continuity:

$$\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0 \quad (22-22)$$

Since J_x , J_y , and J_z are not independent of the charge density ρ , these four quantities form a natural four-vector. In fact, if we define the *four-vector current density* \mathfrak{J}_v by its components ($\mathfrak{J}_1, \mathfrak{J}_2, \mathfrak{J}_3, \mathfrak{J}_4 = ic\rho$), we may write the equation of continuity in covariant form:

$$\sum_v \frac{\partial \mathfrak{J}_v}{\partial x_v} = 0 \quad (22-23)$$

where the summation is understood to run from $v = 1$ to $v = 4$; or equivalently

$$\square \cdot \mathfrak{J} = 0 \quad (22-23a)$$

with $\square \cdot$ standing for a four-dimensional divergence.

Now the vector potential \mathbf{A} and the scalar potential φ satisfy inhomogeneous wave equations:

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_0 \mathbf{J} \quad (22-24)$$

$$\nabla^2 \varphi - \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} = -\frac{1}{\epsilon_0} \rho$$

Since \mathbf{J} and ρ are the components of a four-vector, Eq. (22-24) must represent the four components of a four-vector equation and \mathbf{A} and φ must also combine to form a four-vector. If we define the *four-potential* or *world-potential* \mathfrak{A}_λ by its components, $\mathfrak{A}_1 = A_1$, $\mathfrak{A}_2 = A_2$, $\mathfrak{A}_3 = A_3$, $\mathfrak{A}_4 = i\varphi/c$, then Eqs. (22-24) may be written as

$$\sum_v \frac{\partial^2 \mathfrak{A}_\lambda}{\partial x_v^2} = -\mu_0 \mathfrak{J}_\lambda \quad (22-25)$$

They may also be expressed as

$$\square^2 \mathfrak{A} = -\mu_0 \mathfrak{J} \quad (22-25a)$$

where $\square^2 \equiv \nabla^2 - (1/c^2)\partial^2/\partial t^2$ is the four-dimensional Laplacian operator, or the *d'Alembertian* operator. The Lorentz condition, Eq. (16-63), takes the form

$$\sum_v \frac{\partial \mathfrak{A}_v}{\partial x_v} = 0$$

or

$$\square \cdot \mathfrak{A} = 0 \quad (22-26)$$

We are now in a position to examine the electromagnetic field components. These may be obtained from the customary three-dimensional equations

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (16-57)$$

$$\mathbf{E} = -\nabla\varphi - \frac{\partial \mathbf{A}}{\partial t} \quad (16-60)$$

But \mathbf{A} and $i\varphi/c$ form a four-vector, so the last equation may be written (in component form) as

$$i \frac{1}{c} E_1 = \frac{\partial \mathfrak{A}_1}{\partial x_4} - \frac{\partial \mathfrak{A}_4}{\partial x_1}, \dots \quad (22-27)$$

Thus, \mathbf{B} and $i\mathbf{E}/c$ together make up the four-dimensional curl of \mathfrak{A} . The curl operation applied to a vector actually produces an antisymmetric tensor.* This fact is evident from the form of Eq. (22-27) since it is clear that a two-index quantity must be generated. We define the electromagnetic field tensor \mathbf{F} by the expression

$$F_{\mu\nu} = \frac{\partial \mathfrak{A}_\nu}{\partial x_\mu} - \frac{\partial \mathfrak{A}_\mu}{\partial x_\nu} \quad (22-28)$$

Here,

$$F_{11} = F_{22} = F_{33} = F_{44} = 0$$

$$F_{14} = -F_{41} = -iE_1/c$$

$$F_{24} = -F_{42} = -iE_2/c$$

$$F_{34} = -F_{43} = -iE_3/c$$

$$F_{12} = -F_{21} = B_3$$

$$F_{23} = -F_{32} = B_1$$

$$F_{31} = -F_{13} = B_2$$

* In three dimensions an antisymmetric tensor has three independent components, T_{12} , T_{23} , T_{31} , and these transform under space rotations like the components of a vector. It was therefore satisfactory to treat the curl of a vector as a vector. For an antisymmetric tensor, note that $T_{11} = 0$, $T_{21} = -T_{12}$, and so forth.

In four dimensions an antisymmetric tensor has six independent components, and the tensorial character of the quantity cannot be simplified.

In matrix form,

$$\mathbf{F} = \begin{bmatrix} 0 & B_3 & -B_2 & \frac{-iE_1}{c} \\ -B_3 & 0 & B_1 & \frac{-iE_2}{c} \\ B_2 & -B_1 & 0 & \frac{-iE_3}{c} \\ \frac{iE_1}{c} & \frac{iE_2}{c} & \frac{iE_3}{c} & 0 \end{bmatrix} \quad (22-29)$$

Suppose we now take the divergence of the field tensor. Because of its form, Eq. (22-28), we obtain

$$\sum_v \frac{\partial F_{\mu v}}{\partial x_v} = \frac{\partial}{\partial x_\mu} \sum_v \frac{\partial \mathcal{A}_v}{\partial x_v} - \sum_v \frac{\partial^2 \mathcal{A}_\mu}{\partial x_v^2} \quad (22-30)$$

In view of Eqs. (22-25) and (22-26), this equation becomes

$$\sum_v \frac{\partial F_{\mu v}}{\partial x_v} = \mu_0 \mathfrak{S}_\mu \quad (22-31)$$

or

$$\square \cdot \mathbf{F} = \mu_0 \mathfrak{S} \quad (22-31a)$$

This four-vector equation represents a covariant formulation of two of the Maxwell equations, namely, $\nabla \cdot \mathbf{E} = \rho/\epsilon_0$ and $\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + (1/c^2) \partial \mathbf{E}/\partial t$. Furthermore we have the identity

$$\frac{\partial F_{\mu\nu}}{\partial x_\lambda} + \frac{\partial F_{\nu\lambda}}{\partial x_\mu} + \frac{\partial F_{\lambda\mu}}{\partial x_\nu} = 0 \quad (22-32)$$

where μ , ν , and λ are all different and stand for any three of the subscripts 1, 2, 3, and 4. Equation (22-32) follows immediately from the form of $F_{\mu\nu}$ (Eq. 22-28). We can easily verify that Eq. (22-32) represents the other two Maxwell equations.

The majority of advanced books on relativistic electromagnetic theory achieve a certain compactness of notation by introducing what is known as the *summation convention*. In this formalism, all summation signs are suppressed, but summation is implied by means of a repeated index. Thus, for example, the equation of continuity becomes

$$\frac{\partial \mathfrak{S}_v}{\partial x_v} = 0$$

and the wave equation for the potential becomes

$$\frac{\partial^2 \mathcal{A}_\mu}{\partial x_v \partial x_v} = -\mu_0 \mathfrak{S}_\mu$$

Other equations may be written by analogy. We do not make explicit use of the summation convention in this book, but have introduced it here primarily as an aid to further reading.

22-6

TRANSFORMATION LAW FOR THE ELECTROMAGNETIC FIELD

Since the electromagnetic field is a tensor quantity in the four-dimensional formulation, its components will transform like those of a second rank tensor under a Lorentz transformation:

$$F'_{\mu\nu} = \sum_{\alpha} \sum_{\beta} a_{\mu\alpha} a_{\nu\beta} F_{\alpha\beta} \quad (22-33)$$

This equation is Eq. (II-16) of Appendix II, rewritten to include the fact that $(\tilde{a})_{\beta\nu} = a_{\nu\beta}$.

Again we take the primed system to be one moving with velocity u in the x -direction, relative to the unprimed system. The Lorentz transformation is given by Eq. (22-17). Thus,

$$B'_x = F'_{23} = \sum_{\alpha} \sum_{\beta} a_{2\alpha} a_{3\beta} F_{\alpha\beta} = F_{23} = B_x \quad (22-34)$$

and

$$\begin{aligned} B'_y = F'_{31} &= \sum_{\alpha} \sum_{\beta} a_{3\alpha} a_{1\beta} F_{\alpha\beta} = \frac{1}{\sqrt{1-\beta^2}} F_{31} + i \frac{\beta}{\sqrt{1-\beta^2}} F_{34} \\ &= \frac{1}{\sqrt{1-\beta^2}} [B_y + (\beta/c)E_z] \end{aligned} \quad (22-35)$$

Similarly, we find that

$$B'_z = \frac{1}{\sqrt{1-\beta^2}} [B_z - (\beta/c)E_y] \quad (22-36)$$

As far as the electric field is concerned,

$$\begin{aligned} E'_x = icF'_{14} &= ic \sum_{\alpha} \sum_{\beta} a_{1\alpha} a_{4\beta} F_{\alpha\beta} \\ &= ic \left[-\frac{i\beta}{1-\beta^2} F_{11} + \frac{1}{1-\beta^2} F_{14} + \frac{\beta^2}{1-\beta^2} F_{41} + \frac{i\beta}{1-\beta^2} F_{44} \right] \\ &= \frac{ic}{1-\beta^2} \left[-i \frac{E_x}{c} + \beta^2 i \frac{E_x}{c} \right] = E_x \end{aligned} \quad (22-37)$$

Finally, we verify that

$$E'_y = \frac{1}{\sqrt{1-\beta^2}} [E_y - c\beta B_z] \quad (22-38)$$

$$E'_z = \frac{1}{\sqrt{1 - \beta^2}} [E_z + c\beta B_y] \quad (22-39)$$

Thus the components of \mathbf{E} and \mathbf{B} in the direction of motion are unaffected, but the transverse components are modified.

The above results may be summarized by the following three-dimensional equations:

$$\begin{aligned} \mathbf{E}'_{\parallel} &= \mathbf{E}_{\parallel}, & \mathbf{E}'_{\perp} &= \frac{1}{\sqrt{1 - \beta^2}} [\mathbf{E}_{\perp} + \mathbf{u} \times \mathbf{B}] \\ \mathbf{B}'_{\parallel} &= \mathbf{B}_{\parallel}, & \mathbf{B}'_{\perp} &= \frac{1}{\sqrt{1 - \beta^2}} \left[\mathbf{B}_{\perp} - \frac{1}{c^2} \mathbf{u} \times \mathbf{E} \right] \end{aligned} \quad (22-40)$$

where \parallel and \perp mean components parallel to and perpendicular to the velocity \mathbf{u} of the Lorentz transformation.

The inverse transformation is obviously given by

$$\begin{aligned} \mathbf{E}_{\parallel} &= \mathbf{E}'_{\parallel}, & \mathbf{E}_{\perp} &= \frac{1}{\sqrt{1 - \beta^2}} [\mathbf{E}'_{\perp} - \mathbf{u} \times \mathbf{B}'] \\ \mathbf{B}_{\parallel} &= \mathbf{B}'_{\parallel}, & \mathbf{B}_{\perp} &= \frac{1}{\sqrt{1 - \beta^2}} \left[\mathbf{B}'_{\perp} + \frac{1}{c^2} \mathbf{u} \times \mathbf{E}' \right] \end{aligned} \quad (22-41)$$

Our discussion of the transformation law for the electromagnetic field components is now complete, and the results will be used in the following section.

22-7

THE FIELD OF A UNIFORMLY MOVING POINT CHARGE

To demonstrate the utility of the Lorentz transformation, we shall calculate the electric and magnetic fields of a point charge in uniform motion. Let us assume that the point charge q is moving with the velocity u along the x -axis. The geometry is shown in Fig. 22-2. We construct a second coordinate system (the primed system) that is moving along with the charge; for convenience we make the origin of this system, O' , coincide with the charge itself.

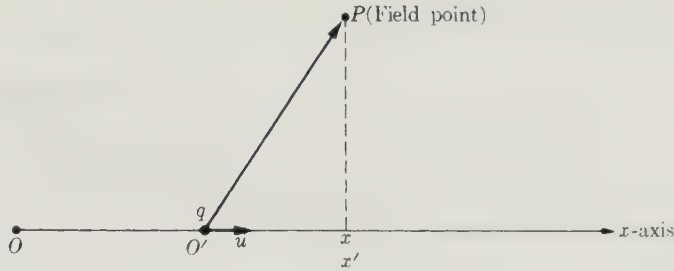
In the primed system the charge is at rest. Therefore, at the field point P ,

$$\mathbf{B}' = 0$$

$$\mathbf{E}' = \frac{q\mathbf{r}'}{4\pi\epsilon_0(r')^3} \quad (22-42)$$

FIGURE 22-2

Coordinate system for determining the \mathbf{E} - and \mathbf{B} -fields of a charge in uniform motion. The heavy line qP is the source-field vector (\mathbf{r}' in the primed system; \mathbf{R} in the laboratory system). Here x is the coordinate in the laboratory system and x' is the coordinate in the primed system.



The fields in the laboratory system may be obtained by using Eq. (22-41). Thus,

$$E_x = E_{\parallel} = E'_x = \frac{qx'}{4\pi\epsilon_0(r')^3}, \quad \mathbf{E}_{\perp} = \gamma\mathbf{E}'_{\perp} = \frac{\gamma q\mathbf{r}'_{\perp}}{4\pi\epsilon_0(r')^3} \quad (22-43)$$

where $\gamma \equiv 1/\sqrt{1 - \beta^2}$. Now from Eq. (22-3),

$$x' = \gamma(x - ut), \quad y' = y, \quad z' = z$$

where t is the time elapsed (in the laboratory system) from the moment at which the two origins coincide. The vector \mathbf{r}' is therefore given by the components:

$$\mathbf{r}' = \{\gamma(x - ut), y, z\} \quad (22-44)$$

It is convenient to define a quantity \mathbf{R}^* by

$$\gamma\mathbf{R}^* = \{\gamma(x - ut), y, z\} \quad (22-45)$$

Thus, Eq. (22-43) becomes

$$\begin{aligned} E_x &= \frac{q}{4\pi\epsilon_0} \frac{\gamma(x - ut)}{\gamma^3(R^*)^3} \\ E_y &= \frac{q}{4\pi\epsilon_0} \frac{\gamma y}{\gamma^3(R^*)^3} \\ E_z &= \frac{q}{4\pi\epsilon_0} \frac{\gamma z}{\gamma^3(R^*)^3} \end{aligned} \quad (22-46)$$

or

$$\mathbf{E} = \frac{q}{4\pi\epsilon_0} \frac{\mathbf{R}}{(R^*)^3} (1 - \beta^2) \quad (22-46a)$$

where \mathbf{R} is defined by

$$\mathbf{R} = \{x - ut, y, z\} \quad (22-47)$$

The electric field is directed radially away from the instantaneous position of the point charge, but in contrast to the static case it is no longer spherically symmetric. In fact, for a fast moving charge, the field is strongly concentrated in the plane at right angles to its motion.

The magnetic field is given by

$$B_x = B_{\parallel} = 0$$

$$\mathbf{B}_{\perp} = \gamma \frac{1}{c^2} \mathbf{u} \times \mathbf{E}' = \gamma \frac{1}{c^2} \mathbf{u} \times \mathbf{E}'_{\perp} = \frac{1}{c^2} \mathbf{u} \times \mathbf{E}_{\perp} \quad (22-48)$$

or

$$\mathbf{B} = \frac{1}{c^2} \mathbf{u} \times \mathbf{E} \quad (22-48a)$$

The magnetic field lines are circles with centers on the charge path.

22-8

SUMMARY

It is an experimental conclusion that electromagnetic radiation propagates with the constant velocity c in vacuum, in *every* uniformly moving coordinate system. Thus the wave equation must be unchanged by a transformation to a uniformly moving coordinate system. This end is accomplished by the **Lorentz transformation** of coordinates, which leaves the quadratic form $(x^2 + y^2 + z^2 - c^2t^2)$ invariant. The transformation is conveniently expressed in terms of the geometry of a four-dimensional, complex (non-Euclidean) **Minkowski space**, in which ict is the fourth coordinate. The Lorentz transformation (and also an ordinary spatial rotation) is represented by a complex orthogonal matrix, which operates on a four-vector, or world-vector. The Lorentz transformation to a system moving with velocity \mathbf{u} in the x -direction is

$$\mathbf{A} = \begin{pmatrix} \gamma & 0 & 0 & i\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\beta\gamma & 0 & 0 & \gamma \end{pmatrix}$$

where $\beta = u/c$, $\gamma = 1/\sqrt{1 - \beta^2}$. (This is a rotation through an imaginary angle.) The Maxwell equations are **covariant**—that is, they have the same form if the E - and B -field components are properly transformed by the Lorentz transformation.

• The principal electromagnetic four-vectors, the field tensor, and the relations among them are as follows:

Four-vectors:

$$\text{Space-time } \mathbf{x} = (x, y, z, ict)$$

$$\text{Current-charge } \mathfrak{J} = (J_x, J_y, J_z, ic\rho)$$

$$\text{Potential } \mathfrak{A} = (A_x, A_y, A_z, i\varphi/c)$$

Field tensor:

$$F_{\mu\nu} = \frac{\partial \mathfrak{A}_\nu}{\partial x_\mu} - \frac{\partial \mathfrak{A}_\mu}{\partial x_\nu}, \quad \mathbf{F} = \square \times \mathfrak{A}$$

Maxwell's equations:

$$\sum_\nu \frac{\partial F_{\mu\nu}}{\partial x_\nu} = \mu_0 \mathfrak{J}_\mu, \quad \square \cdot \mathbf{F} = \mu_0 \mathfrak{J}$$

$$\frac{\partial F_{\mu\nu}}{\partial x_\lambda} + \frac{\partial F_{\nu\lambda}}{\partial x_\mu} + \frac{\partial F_{\lambda\mu}}{\partial x_\nu} = 0$$

Wave equation for potential:

$$\sum_\nu \frac{\partial^2 \mathfrak{A}_\mu}{\partial x_\nu^2} = -\mu_0 \mathfrak{J}_\mu, \quad \square^2 \mathfrak{A} = -\mu_0 \mathfrak{J}$$

Lorentz condition:

$$\sum_\nu \frac{\partial \mathfrak{A}_\nu}{\partial x_\nu} = 0, \quad \square \cdot \mathfrak{A} = 0$$

Equation of continuity:

$$\sum_\nu \frac{\partial \mathfrak{J}_\nu}{\partial x_\nu} = 0, \quad \square \cdot \mathfrak{J} = 0$$

• The **field tensor** transforms according to

$$\mathbf{F}' = \mathbf{AFA}^{-1} = \mathbf{AF}\bar{\mathbf{A}}$$

In three dimensions,

$$\mathbf{E}'_{\parallel} = \mathbf{E}_{\parallel}, \quad \mathbf{E}'_{\perp} = \gamma(\mathbf{E}_{\perp} + \mathbf{u} \times \mathbf{B})$$

$$\mathbf{B}'_{\parallel} = \mathbf{B}_{\parallel}, \quad \mathbf{B}'_{\perp} = \gamma\left(\mathbf{B}_{\perp} - \frac{1}{c^2} \mathbf{u} \times \mathbf{E}\right)$$

• The field of a **uniformly moving point charge** is easily obtained by transforming from the rest system of the charge.

PROBLEMS

22-1 Transform the wave equation to the primed system of coordinates by using the Galilean transformation, Eq. (22-1). Show that

$$\varphi = F\{x - (c - u)t\} + G\{x + (c + u)t\}$$

where F and G are arbitrary functions of their arguments, is a solution to the transformed equation.

22-2 By making two consecutive Lorentz transformations, first to the Σ' -system moving with velocity u relative to the Σ -system and then to Σ'' moving with velocity u' relative to Σ' , prove the relativistic addition theorem for velocities:

$$u'' = \frac{u + u'}{1 + uu'/c^2}$$

22-3 An unstable particle has an intrinsic half-life of 6×10^{-19} s. If it is travelling at 0.9999 times the speed of light, what is its apparent half-life in the laboratory system?

22-4 Show that the Lorentz force equation is incorporated in the covariant form:

$$f_\mu = \sum_\nu F_{\mu\nu} \dot{x}_\nu$$

22-5 Given uniform electric and magnetic fields \mathbf{E} and \mathbf{B} , find a Lorentz transformation that will make \mathbf{E} and \mathbf{B} parallel to each other. [Hint: Choose the velocity \mathbf{u} of the Σ' -system in a direction perpendicular to both \mathbf{E} and \mathbf{B} , and determine the magnitude of $u/(1 + \beta^2)$ in terms of E^2 , B^2 , and $\mathbf{E} \times \mathbf{B}$.]

22-6 Equation (2-30) gives the electric field of a long straight wire containing λ units of charge per unit length. (a) Make a Lorentz transformation to a system moving with velocity u in a direction parallel to the wire. (b) Calculate the B -field in the new system and compare with the B -field of a current-carrying wire, Eq. (8-35). (c) Is there an E -field in the moving system? (d) What is the physical difference between a charged wire moving in the direction of its length and a current-carrying wire?

22-7 (a) Show that the scalar product $\mathbf{E} \cdot \mathbf{B}$ is unchanged by a Lorentz transformation. (b) Show the same for $E^2 - c^2 B^2$.

22-8 (a) Determine the Poynting vector for the uniformly moving point charge of Section 22-7, and (b) show that the total power radiated is zero.

22-9 A point magnetic dipole moment \mathbf{m} at rest produces a vector potential \mathbf{A} given by

$$\mathbf{A} = \frac{\mu_0 \mathbf{m} \times \mathbf{r}}{4\pi r^3}$$

If \mathbf{m} moves with a constant velocity \mathbf{v} (with $v \ll c$), show that there is also a scalar potential that appears to come from an electric dipole moment $\mathbf{p} = (\mathbf{v} \times \mathbf{m})/c^2$.

Computers and Electromagnetic Theory

In the course of the last fifty years, the digital computer has evolved from a rare item to one that is commonly available, not only in research establishments, but also to any individual who has a need for one. Science and engineering, as well as many other human activities, have been major beneficiaries of this increased availability. Very few electricity and magnetism texts, however, say anything about computers. As a result the transition from the neat problems that are treated analytically to the messy problems that occur in applications is abrupt if not traumatic. The objective of the computer material in this book is to ease the discontinuity and facilitate the use of computers in exploiting electromagnetic theory. The two important ingredients, in addition to electromagnetic theory, are numerical analysis and computer programming. However, while this appendix is relatively self-sufficient, it is not a text on numerical analysis, or computer programming, or the BASIC language. The bibliography at the end of this appendix contains a selection of material on each of these topics.

THE EVOLUTION OF COMPUTING

Computation has had a long history, but its most dramatic development phase began after World War II. Before 1940, the computers used by most physicists were arcane, hand-powered devices that required some skill to operate. During World War II a small number of large-scale (by 1940 standards) electronic digital computers were built, mainly to satisfy the needs of ballistics and cryptography. By the early 1950s, new general-purpose computers were beginning to appear, and one such device, WHIRLWIND, gained a place of honor in the Smithsonian Institution at the ripe old age of thirty. By the late fifties, standard, large electronic computers had become available from several vendors. These systems were

bulky and used vacuum tubes to implement the hardware requirements. But the astonishing development of semiconductor devices led to smaller computers and greater capacity (and incidentally lower cost computing). These trends continued, leading to mini-computers in the 1970s and micro-computers by the end of that decade. The introduction of the PC by IBM in mid-1981 was an important event because it led to a standardization of architecture and operating systems and so to the development of third-party software. At the present time, commercially available computers span the range from giant supercomputers to very capable pocket-sized micro-computers. Desktop computers, in particular, are capable of implementing important applications in electromagnetic theory and are very affordable.

The fact that computers must be told what to do requires a computer language. In the early days of computers, each one had its own language, typically built from very simple instructions that were used in enormous arrays. Learning a new language was painful and slow, and so impeded the adoption of new computers. In the late 1950s, this problem was solved with the invention of a standard language called FORTRAN (FORMula TRANslator) by John Backus and co-workers at IBM. Using FORTRAN made it possible to instruct the computer (write programs) in a language similar to ordinary English. This language was not understood by the computer, but it could be translated into machine language in a process now called compiling. Since compilers can be changed to adapt to new machine languages without affecting the language used by the programmer, early programmers could use FORTRAN to program any machine that was equipped with a FORTRAN compiler without learning a complex new language. FORTRAN remains a very popular high-level language among scientists and engineers. It has been standardized by the American National Standards Institute, and the current version is FORTRAN 77 (ANSI X3.9-1978). A revision, FORTRAN 90, should be available soon.

FORTRAN required too much memory to be useful on early micro-computers. Thus, another language, called BASIC (Beginners All-purpose Symbolic Instruction Code), was developed by John Kemeny and Thomas Kurtz to solve this problem and simplify the programming process. The BASIC language is also closely related to English and in many ways is similar to that of FORTRAN. The two most apparent differences are that standard BASIC identifies each instruction line with a number, and each line is converted to machine language and executed sequentially in a procedure called interpreting. Compiling, as is done with FORTRAN, converts the entire program and optimizes the machine language version before executing the program. BASIC was so useful that it was and still is supplied with most desktop computers. It too has evolved, with the most recent standard versions being BASIC 3.22 and 3.33. In 1990 a new version of BASIC called QBASIC appeared. It is incorporated into DOS 5.0, and currently it is being shipped with many desktop computers that use DOS.

QBASIC does not use line numbers and has several other advantages, but it is still interpreted. More advanced versions of BASIC avoid the line numbers and also compile rather than interpret. Other languages have been developed that have significant advantages for some purposes, but BASIC and FORTRAN are still the languages most commonly used by scientists and engineers.

What Computers Do

A computer system, regardless of size, typically consists of four units: *input*, *output*, *central processor*, and *memory*. The first two provide for communication with the computer and consist of keyboards, tape and disk drives, video screens, printers, plotters, and so on. The memory provides for the storage of data and instructions within the computer where they can be accessed readily as needed. It usually consists of two independent parts: the random access memory (RAM), which is used for storage while the central processor is carrying out its assignments, and disk or tape storage, which is used for long-term storage. There are also relatively small read-only memories (ROM), which are used to store frequently used material. BASIC, for example, is sometimes contained in ROM. The central processor unit (CPU) can do elementary arithmetic and logical operations in response to instructions provided by its controller. The controller in turn provides communication with the rest of the computer and conversion of instructions to a form suitable for use by the arithmetic and logical elements of the CPU. The basic functions may be thought of as *communications*, *storage*, *arithmetical operations*, and *logical operations*. From these basic functions, sequences that accomplish very complicated tasks can be constructed.

For the purposes of electromagnetic theory and mathematical physics in general, these capabilities translate into the ability to carry out complex computational instructions containing both arithmetic and logical elements, the ability to repeat calculations for different values of arguments or parameters, and the ability to do symbolic mathematics.

APPLICATIONS OF COMPUTERS TO ELECTROMAGNETIC THEORY

The computational capability facilitates many useful tasks, and we now illustrate a few with references to the main text where they are used. One of the most frequently used capabilities is the repeated evaluation of mathematical expressions. In Chapter 13, the *transfer function* of a simple network is obtained as an algebraic expression. A simple computer program evaluates this expression over a specified range of frequencies with a specified step size. Doing this evaluation manually for a single frequency is

not difficult, but for 100 different frequencies, the computer provides definite advantages. In Chapter 3, *Laplace's equation* is solved for a rectangular region using a series expansion in sines and hyperbolic sines. A computer program is used to evaluate the potential at specified points. This program takes advantage of the pre-programmed ability of BASIC to compute sines and exponentials. Here again the computation is not intrinsically complex, but summing 11 terms of the series for even one specific point is laborious. Also in Chapter 3, a set of finite difference equations is solved by an iterative technique. Each individual calculation is simple, but it must be repeated 50 times per iteration, and it may take 50 or more iterations to obtain a sufficiently accurate result. The computer does these calculations very rapidly. In Appendix VII, integrals are evaluated numerically in order to use the dispersion relations. This automation of a formerly laborious manual task is extremely important in science and engineering. In addition, circuits, boundary value problems, and many other applications involve the solution of sets of *simultaneous linear equations*. This process is sufficiently common that we consider it later in a separate section of this appendix. Chapters 3 and 13 contain examples requiring the solution of simultaneous linear equations.

Approach

Our approach to computing in connection with electromagnetic theory is epitomized by the imperative statement, "Use clever software written by other people." There is so much software now available that this valuable advice must be tempered with the observation that it is sometimes quicker to modify a good program that does almost what is needed than it is to locate a program that does exactly what is needed. The appropriate balance is a matter of taste and develops with experience. We make no claim of elegance for the listed programs. They are as simple and direct as we could make them while still serving the required purposes. To the best of our knowledge, none of the programs in the text appear elsewhere in the form we have used. Many of them, however, owe debts to long lists of people who have written and modified similar programs.

For each application in the text, we have tried to present several approaches of varying generality and varying sophistication. Where known to us and readily available to others, we have also cited more general and more sophisticated program listings. Finally, with the same caveats, we mention some commercially available software. Since commercial software changes rapidly we have not tried to be either complete or current, but only to give the flavor of what is available.

The problems require modification of programs listed in the text or, in some cases, the writing of new programs. They usually extend the text and often provide for more detailed exploration of idealized examples.

Specific Hardware and Software

The baseline hardware system for the material in the text is an IBM PC-XT, with two 360-K floppy drives, 640-K RAM, a graphics card, and an appropriate monitor (or any of the numerous clones of such a system). A printer, even of the inexpensive nine-pin dot matrix variety, is useful but probably not essential, except for submitting exercises. The essential software is a disk operating system, MS DOS version 2.11 or higher, and a compatible version of BASIC. Version 2.02 will work with DOS 2.11 or higher, but BASIC version 3.22 requires DOS version 3.2 or later. Alternatively, a Macintosh and an appropriate version of BASIC can be used, but this may require some modification of the programs listed in the text. A hard drive will save a great deal of time and represents a useful enhancement. In the realm of software for scientific and engineering applications, a graph plotting program is extremely useful;* but unfortunately, most inexpensive packages do not produce graphs of suitable quality for science and engineering use. Packages that produce good graphs tend to be expensive and often require a hard drive. A useful alternative is provided by BASIC routines that produce at least rough graphs. One such routine is listed by Miller.†

We decided to use standard BASIC (by *standard BASIC* we mean any of the versions using line numbers) to present programs in the text. The main reasons for this choice are simplicity, availability, adaptability to a variety of host computers, availability of compilers, and similarity to FORTRAN. BASIC programs will run on QBASIC, where the line numbers are optional, but the converse is subject to significant restrictions. The availability of compilers deserves comment. Because BASIC is interpreted, it executes relatively slowly, but compiled versions of BASIC have been developed that provide faster execution. These typically suppress the line numbers characteristic of ordinary BASIC and make some changes in the language, which generally make it even more like FORTRAN. The compiled versions of BASIC run faster but are not really needed for the problems in the text. One point that is often obscure is that programs written in ordinary BASIC or QBASIC (e.g., the programs in this text) can be compiled. To accomplish the compilation, the complete BASIC program, including all subroutines, is saved as an ASCII file (the SAVE "filename", A command in BASIC; automatic in QBASIC). This file is then loaded into the compiler, and the compile routine is executed. Alternatively, the program including the essential line numbers and all subroutines

* Some of the graphs in the text were produced using Tech*Graph*Pad (Copyright Binary Engineering Software, Inc.), which is a typical commercial program capable of producing publication-quality scientific and engineering graphs.

† See the bibliography listed at the end of this appendix.

can be typed directly into QuickBASIC. The compiled program will usually run several times faster than the original, depending on the program and the compiler. QuickBASIC (Microsoft Corporation) almost always gives a substantial (three- or four-fold) improvement. Some other compilers do not perform as well.

BASIC also has some disadvantages. The two that we note are that, first, the language does not deal with complex numbers as a standard type of numerical quantity. As a result, equations involving complex numbers must be explicitly decomposed into real and imaginary parts. This process is particularly annoying in dealing with ac circuits and wave propagation. The second disadvantage is that there seems to be no readily available, extensive compilation of useful scientific and engineering programs in BASIC. Books by Miller* and Walker* are useful, but for general utility they do not compare with *Numerical Recipes*,* which is now available for FORTRAN, PASCAL, and C. For QuickBASIC, the situation has improved with the appearance of *Numerical Recipes in BASIC*, which is a supplement to *Numerical Recipes*. The routines in this supplement run on QuickBASIC 4.5 and can often be adapted easily to QBASIC. Adapting them to BASIC is not something we recommend. BASIC has other shortcomings, but at this level they are clearly outweighed by the advantages.

Hardware and Software Summary

Hardware: Desktop computer using DOS operating system with at least two 360-K floppy drives

Graphics card

Monitor

Software: DOS operating system

BASIC, QBASIC, or QuickBASIC compatible with DOS version

It is almost impossible to buy BASIC alone. It is, however, usually supplied with DOS and packaged with IBM PC's and clones. QBASIC is packaged with DOS 5.0, and alternatively the disks are supplied with *Learn BASIC Now* by Halvorson and Rygmyr,* where it is called "Interpreted QuickBASIC." QuickBASIC, currently Version 4.5, is readily available at computer stores. QBASIC and QuickBASIC are substantial improvements on BASIC because they include additional statements and because they are more convenient to use. If you are new to BASIC, we would recommend starting with QBASIC rather than BASIC.

The programs and code in this book will run without change in almost any version of BASIC, QBASIC, or QuickBASIC. Line

* See the bibliography listed at the end of this appendix.

numbers are required in BASIC and are optional in the other two. The limitations of BASIC, however, require that the programs be written in such a way that some line numbers are essential. These essential line numbers are indicated in the heading box for each program. If the program is being typed into QBASIC or QuickBASIC, only the essential line numbers need be entered.

If a program is typed in BASIC (including line numbers), it can be used in QBASIC or QuickBASIC by saving it using SAVE "MYPROG.BAS",A and then opening it in the prescribed way in the other software. Programs originated in QBASIC or QuickBASIC *with line numbers* (note that line numbers are not entered automatically) will run in BASIC unless statements not supported by BASIC are used in the programs. In particular, subroutines are dealt with by more sophisticated methods in QBASIC and QuickBASIC and these will not work in BASIC. This procedure is not recommended.

COMPUTER IMPLEMENTATION OF SELECTED NUMERICAL METHODS

In this section, we discuss the computer implementation of two extremely common numerical procedures, *numerical integration* and *solving linear equations*.

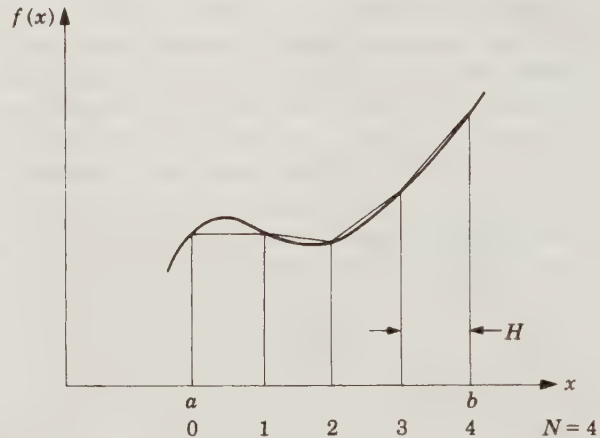
Numerical Integration

Our requirement for numerical integration can be satisfied by a simple trapezoidal formula and a program that implements it. The basic idea is illustrated in Fig. I-1. The integral of $f(x)$ is to be computed between the limits a and b . To approximate this integral the interval is divided into N equal intervals of width H . For each interval, the value of $f(x)$ at the beginning of the interval is added to the value at the end of the interval, and the sum is multiplied by $H/2$, where H is the width of the interval. These contributions from the N intervals are added to get the approximate value of the integral. The limit of this sum as the number of intervals becomes infinite is one definition of the Riemann integral so one would expect that a large number of small intervals could give a good approximation to the integral. In general, this expectation is correct; however, if the integrand is singular or otherwise pathological, considerable care may be needed to get the correct result, but we will not pursue this point further.

The short program listed in Fig. I-2 implements this procedure. The integrand is defined on line 200; as written, it is the integrand for evaluating the power radiated by a halfwave antenna as discussed in Section 20-2. Other integrands can be written in lines 201 to 209, and all but one may be "commented out" to provide flexibility for doing a variety of integrals as

FIGURE I-1

The area under the curve is approximated by trapezoids.



needed. The program first asks for the upper and lower limits and the tolerance as a fraction of the value of the integral. The approximation for $N = 1$ is defined. The range of integration is covered by means of a FOR...NEXT loop, and the integral is approximated by the sum of the areas of two trapezoids, I(2). If this approximation differs from I(1) by more than the tolerance, the program returns to line 270 and doubles the number of intervals. This process is repeated until the difference between

FIGURE I-2

BASIC program for integration by trapezoid rule (integrand is defined on line 200)

```

100 /*****
110 /*
120 /*           INTEG3:INTEGRATION BY TRAPEZOIDAL RULE
130 /*           NO ESSENTIAL LINE NUMBERS
140 /*
150 /*****
160 /
170 CLS
180 DIM S(1024): DIM I(1024): DIM D(1024)
190 PI = 4 * ATN(1)
200 DEF FNF (X) = ((COS(PI * X / 2)) ^ 2) / (1 - X ^ 2)
210 INPUT "Lower limit =", LL
220 INPUT "Upper limit =", UL
230 INPUT "Tolerance, T =", T
240 N = 1
250 D(1) = 2 * T
260   WHILE D(N) > ABS(T * I(N))
270     N = 2 * N
280     H = (UL - LL) / N
290     S(0) = (UL - LL) * (FNF(UL) + FNF(LL)) / (2 * N)
300     FOR J = 1 TO N - 1
310       S(N) = S(N) + H * (FNF(LL + J * H))
320     NEXT J
330     I(N) = S(0) + S(N)
340     D(N) = ABS(I(N) - I(N / 2))
350   WEND
360 PRINT "N =", N, "I(N) =", I(N), "D(N) =", D(N)
370 END

```

$I(N)$ and $I(N/2)$ is less than the tolerance. The number of intervals, the value of the integral, and the difference between $I(N)$ and $I(N/2)$ are printed.

If this program is simply loaded and run with the limits -1 and 1 , the computer announces that you are asking it to divide by zero. This problem occurs at the limits (in spite of the fact that the integrand actually vanishes at both limits), so it is easy to overcome by choosing limits slightly different from -1 and 1 ; for example, -0.999 and 0.999 instead. With these limits the program runs smoothly and gives a very satisfactory result for 512 intervals. However, the problem of dividing by zero is always a difficulty to be circumvented in one way or another.

Simultaneous Linear Equations

The solution of sets of simultaneous linear equations is one of the most frequently occurring tasks in mathematical physics. We limit ourselves to a nonsingular (i.e., $|A_{ij}| \neq 0$) set of equations of the form

$$\sum_j A_{ij} W_j = D_i \quad (\text{I-1})$$

with real coefficients. The obvious approach is to use Cramer's rule to find the inverse of the matrix \mathbf{A} and then multiply Eq. (I-1) by this inverse to obtain the W_j 's. It is, however, well known that this procedure is prohibitively time consuming even for fast computers if N , the number of equations, is even moderately large. Fortunately there are other faster techniques, and we describe here the *Gauss-Jordan elimination method* with full pivoting. In its simplest form, elimination involves subtracting multiples of the first equation from the other equations to eliminate W_1 from all equations except the first. The same procedure is followed using the second equation to eliminate W_2 from all equations except the second, and so on. The result is a series of equations of the form $W_i = C_i$ where the C_i 's are derived from the D_i 's in the process of elimination. A problem arises with this method when some of the diagonal elements of the coefficient matrix are small. In this case round-off errors can produce erroneous results. These are avoided by *pivoting*, which amounts to interchanging rows to put the largest element of the matrix on the diagonal and then eliminating the other elements in that column. The largest remaining element in the matrix is treated in the same way. After N repetitions, there remains a diagonal matrix that can be easily reduced to the identity matrix. In the process, the right-hand side is converted into the solution. It should be noted that full pivoting is not always necessary, and most sophisticated linear equation solvers pivot only if the diagonal element is smaller than some predetermined value. The full pivoting process is implemented by the program listed in Fig. I-3. The program first finds the largest element in the matrix and then interchanges rows to put this largest element on the diagonal. Each

FIGURE I-3

BASIC program for
Gauss-Jordan elimination

```

2000 /*****
2005 /*
2010 /*           G-J2KA: GAUSS-JORDAN ELIMINATION SUBROUTINE
2015 /*           LINE NUMBERS 2000, 2245, 2250, 2345 AND 2431 ARE ESSENTIAL
2020 /*
2025 /*****
2030 ' Required inputs: M1%, N1%, N2%, A(I,J), D(I); Dimensions of A, D, W
2035 ' should be specified by I-O program. Solution is generated as W(I) in
2040 ' in the same order as input, but not printed. Main program should order
2045 ' printing.
2050 '
2055 ' Identifiers
2060 ' A      coefficient matrix
2065 ' B1     biggest value
2070 ' D      constant vector
2075 ' H1     work variable
2080 ' IR%    row index
2085 ' IC%    column index
2090 ' M1%    maximum dimension of matrices
2095 ' N1%    number of rows
2100 ' N2%    number of columns
2105 ' P1     pivot element
2110 ' R%     work matrix
2115 ' W      solution matrix
2120 ' Note - I,J,K & L with or without suffixes are indices
2125 ' End of identifiers
2130 '
2135 DIM R%(M1%, 3)
2140 '
2145 ' Initialize R%(I%,3) and W(I%)
2150 FOR I% = 1 TO N2%
2155   W(I%) = D(I%)
2160   R%(I%, 3) = 0!
2165 NEXT I%
2170 '
2175 ' Diagonalize A(I%,J%)
2180 FOR I% = 1 TO N2%
2185 '
2190 ' Search for largest (pivot) element, B1
2195   B1 = 0!
2200   FOR J% = 1 TO N2%
2205     IF (R%(J%, 3) = 1) THEN 2250
2210     FOR K% = 1 TO N2%
2215       IF (R%(K%, 3) > 1) THEN PRINT "Error1 - Matrix is singular"
2220       IF (R%(K%, 3) = 1) THEN 2245
2225       IF (B1 >= ABS(A(J%, K%))) THEN 2245
2230       IR% = J%
2235       IC% = K%
2240       B1 = ABS(A(J%, K%))
2245     NEXT K%
2250   NEXT J%
2255 '
2260 ' Change R% to indicate biggest element in position IR%, IC%
2265   R%(IC%, 3) = R%(IC%, 3) + 1
2270   R%(IR%, 1) = IR%
2275   R%(IR%, 2) = IC%
2280 '
2285 ' Interchange rows to put pivot element on diagonal
2290 IF (IR% = IC%) THEN 2345
2295 FOR L% = 1 TO N2%
2300   H1 = A(IR%, L%)
2305   A(IR%, L%) = A(IC%, L%)
2310   A(IC%, L%) = H1
2315 NEXT L%
2320 H1 = W(IR%)
2325 W(IR%) = W(IC%)
2330 W(IC%) = H1
2335 '
2340 ' Divide pivot row by pivot element
2345 P1 = A(IC%, IC%)
2350 IF ABS(P1) < .0000001 THEN PRINT "Error2 - matrix is ill conditioned"
2355 A(IC%, IC%) = 1
2360 FOR L% = 1 TO N2%
2365   A(IC%, L%) = A(IC%, L%) / P1
2370 NEXT L%

```

FIGURE I-3
(continued)

```

2375     W(IC%) = W(IC%) / P1
2380 '
2385 'Reduce elements in column IC% in non-pivot rows
2390 FOR L1% = 1 TO N2%
2395     IF (L1% = IC%) THEN 2430
2400     T = A(L1%, IC%)
2405     A(L1%, IC%) = 0
2410     FOR L% = 1 TO N2%
2415         A(L1%, L%) = A(L1%, L%) - A(IC%, L%) * T
2420     NEXT L%
2425     W(L1%) = W(L1%) - W(IC%) * T
2430     NEXT L1%
2435 '
2440 NEXT I%
2445 '
2450 'Test for errors
2455 FOR K% = 1 TO N2%
2460     IF (R%(K%, 3) <> 1) THEN PRINT "Error3 - matrix is singular": STOP
2465 NEXT K%
2470 '
2475 RETURN: 'Return to main program from Gauss-Jordan subroutine
2480 END

```

element in that row is divided by the largest element to reduce the diagonal element to unity. Multiples of that row are subtracted from the other rows to make the nondiagonal elements in the pivot column zero. The process is repeated, excluding the first pivot row from the search for the largest matrix element. After a number of repetitions equal to the number of equations, the matrix **A** is converted into the identity matrix, and the vector **D** is converted to the solution vector. The only feature requiring further explanation is the register matrix **R**, which has three columns and as many rows as there are equations and keeps track of the pivoting. The third column of this matrix starts with all zero values. After the largest element is found, the first row is changed to give the row and column location of the largest element and the third column is changed to one. The row and column indices are changed as the largest element is moved to the diagonal. On the second iteration, this row is skipped and the largest remaining element is found. If, after all rows have been processed, any element in the third column of **R** is not one, it means the matrix is singular and a message to that effect is displayed. Many Gauss-Jordan routines also calculate the determinant of the coefficient matrix and explicitly calculate the inverse matrix. This program can be extended to do these calculations but, since we do not use them in the text, the shorter version has been listed. Under some circumstances the dimensions of **R** and **W** may be specified by the control program, in which case apostrophes should be added to the beginning of lines 2120 and 2125 to avoid error messages. It is convenient to test the program with a simple set of equations with a known solution. A test input is listed in Fig. I-4. If this short program is loaded and the Gauss-Jordan program (G-J2KA, Fig. I-3) merged with it, then running the program should produce the answers shown.

FIGURE I-4

BASIC input-output program to test Gauss-Jordan program. (Answers to test problem: $X_1 = 0.5$, $X_2 = 1$, $X_3 = 1.5$, and $X_4 = 2$.)

```

100 '*****
110 ' *
120 ' *           G-J_TEST: PROGRAM TO PROVIDE TEST INPUT AND PRINT OUTPUT FOR
130 ' *           GAUSS-JORDAN EQUATION SOLVING PROGRAM
140 ' *           NO ESSENTIAL LINE NUMBERS
145 ' *
150 ' *
160 '*****
170 INPUT "M1$ =", M1$
180 DIM A(M1$, M1$), D(M1$), W(M1$)
190 N1$ = 4: N2$ = 4
200 A(1, 1) = 1: A(1, 2) = 2: A(1, 3) = 3: A(1, 4) = 4: D(1) = 15
210 A(2, 1) = 2: A(2, 2) = 3: A(2, 3) = 4: A(2, 4) = 1: D(2) = 12
220 A(3, 1) = 3: A(3, 2) = 4: A(3, 3) = 1: A(3, 4) = 2: D(3) = 11
230 A(4, 1) = 4: A(4, 2) = 1: A(4, 3) = 2: A(4, 4) = 3: D(4) = 12
240 GOSUB 2000
250 PRINT "X1 =", W(1), "X2 =", W(2), "X3 =", W(3), "X4 =", W(4)
260 END

```

Comments on Numerical Analysis

We have said almost nothing about numerical analysis. Numerical analysis is too important to relegate to a few words in a text on another subject. It is important to note, however, that the very power of contemporary computers makes it extremely easy to get into trouble with convergence, stability, and many of the other traps of numerical analysis. There are a number of books that deal with numerical analysis in the context of modern computing; some of these are listed in the bibliography at the end of this appendix. It is useful to have one available for reference purposes when a problem arises.

I-4

USEFUL TIPS

Several useful tips that are hard to find in the usual references are listed here.

1. It is always a good idea to test programs by doing a problem with known solutions. Problems that can be done manually are often the best for testing. The process of solution by the computer can then be followed by judiciously inserted temporary PRINT and STOP statements. For similar reasons it is best to use data sets that are just large enough to exercise all of the features of the program. In Chapter 13, for example, three frequencies are enough for checking the program. Once it is running properly, the full data set can be inserted.

2. The line

```
PRINT "Press any key to continue.": K$ = INPUT $(1)
```

will stop the program until any keyboard key is pressed. K\$ can be any

otherwise unused string variable. It is useful for pausing to examine inputs on the screen after they have been entered.

3. To denote user choice, the lines

```
INPUT "Do you wish to (statement) (Y/N)"; A$
IF A$ = "Y" GOTO (line number) [or THEN GOSUB (line
number) . . . RETURN (line number)]
```

enable the user to choose whether or not to execute the statement. It can be used to quit, to print, or to execute a subprogram at the user's option. As always a little care is needed to insure that entries and exits are made at the proper points to implement the desired logic.

4. Summing series. The easiest way of evaluating the sum of the first N terms of a series $S_{N-1} = \sum_{j=0}^{N-1} T_j$ is to use a recursion relation $S_N = S_{N-1} + T_N$; $T_N = f(T_{N-1}, N - 1)$ in a FOR NEXT or WHILE WEND loop. For example,

```
INPUT "X = ", X: INPUT "N = ", N
S(0) = 1: T(0) = 1
FOR J = 1 TO N
T(J) = T(J - 1) * X / J
S(J) = S(J - 1) + T(J)
NEXT J
```

computes the sum, $S(N)$, of the first $N + 1$ terms of the series expansion for $\exp x$. This process is used several times in various programs in the text and appendices.

5. Variables must be initialized; otherwise they will start with whatever is left in the assigned storage. That is why the second line is included in the short program in item four above.

6. Many calculations require π . The line

```
PI = 4 * ATN(1)
```

makes it possible to type PI whenever π is needed.

7. Remember that DEF FN can be used to define functions in terms of the predefined functions of BASIC as well as algebraic expressions. Thus, for example,

```
DEF FN TANH(X) = (EXP(X) - EXP(-X)) / (EXP(X) + EXP(-X))
```

provides the hyperbolic tangent with the statement FN TANH(X).

8. Remember that arrays must be explicitly dimensioned if the dimension exceeds 10, and it is always a good idea to dimension them. Note, for example, that S(N), I(N), and D(N) are all dimensioned in line 160 of Fig. I-2. Dimensioning also sets all values to zero.

9. As mentioned earlier, PRINT and STOP are useful debugging tools.

10. It is sometimes useful to "comment out" lines of code by inserting

either REM or an apostrophe (') at the beginning of the line. Such lines are ignored in the execution of the program, but can be easily reactivated by deleting REM or the apostrophe.

11. Finally, we repeat the early advice: "Use clever software written by other people."

BIBLIOGRAPHY

The following bibliographic listing does not aim at completeness, but it does contain some useful references.

Numerical Analysis

1. Hanning, R. W., *Numerical Methods for Scientists and Engineers*, 2nd ed. (New York: McGraw-Hill, 1973; reprinted New York: Dover, 1986).
2. Stoer, J., and R. Bulirsch, *Introduction to Numerical Analysis* (New York: Springer, 1983).

Computer Programming

1. Knuth, Donald, *The Art of Computer Programming* (Reading, MA: Addison-Wesley, 1969).
2. Press, William H., Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge: Cambridge University Press, 1986, 1989). This book is available in FORTRAN, Pascal, and "C" versions, and a companion volume for use with QuickBASIC version 4.5 is available and listed below.

BASIC and Applications

1. "Microsoft GW-BASIC: User's Guide and User's Reference," Redmond, WA, Microsoft Corporation, 1987. This manual is essential for users of GW-BASIC; similar reference manuals exist for other versions of BASIC.
2. Halvorson, Michael, and David Rygmyr, *Learn BASIC Now* (Redmond, WA: Microsoft Press, 1989). This tutorial on QBASIC comes with disks containing QBASIC and exercises. Apparently there is no reference manual; the extensive on-line help files serve this purpose.
3. Rugg, T., and P. Feldman, *Using BASIC* (Carmel, IN: Que Corporation, 1990). This book is typical of many now on the market. It provides a very convenient way of learning the BASIC language. It does not contain much about solving mathematical problems.
4. Rugg, T., and P. Feldman, *Using QBASIC* (Carmel, IN: Que Corporation, 1991). The same comments apply as for *Using Basic*.

5. Miller, Alan R., *BASIC Programs for Scientists and Engineers* (San Francisco: SYBEX Corporation, 1981). This book contains about 50 general-purpose programs and illustrations of their use. It is a useful introduction to writing technical programs in BASIC.
6. Koonin, Steven E., *Computational Physics* (Reading, MA: Addison-Wesley, 1986). Koonin uses BASIC to deal with problems of real interest in physics. Some of the programs are very elaborate. The book is a good follow-on to Miller.
7. Walker, Robert D., *Numerical Methods for Engineers and Scientists* (Summit, PA: TAB Books, 1987). Walker lists 21 BASIC routines for doing useful mathematical tasks covering such things as interpolation, curve fitting, complex numbers, matrix algebra, solution of sets of simultaneous linear equations, and others. There is also a useful menu routine that provides convenient access to the individual programs. The programs are arranged to deal with mathematical problems, but they can be integrated into other programs.
8. Sprott, Julien C., *Numerical Recipes in BASIC* (Cambridge: Cambridge University Press, 1991). A companion volume to *Numerical Recipes*, this book does not contain all of the text from *Numerical Recipes*; both are needed. The routines and examples were prepared to work with QuickBASIC 4.5.

Coordinate Transformations, Vectors, and Tensors

In order to achieve a compactness in notation, we shall use x_1, x_2, x_3 for the Cartesian coordinates x, y, z . A coordinate transformation is *linear* if the new coordinates can be expressed as a linear combination of the old coordinates. Thus,

$$\begin{aligned}x'_1 &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\x'_2 &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\x'_3 &= a_{31}x_1 + a_{32}x_2 + a_{33}x_3\end{aligned}\tag{II-1}$$

or

$$x'_i = \sum_j a_{ij}x_j\tag{II-1a}$$

is a linear transformation. It is understood that the summation runs from $j = 1$ to $j = 3$. Furthermore i can take on any of the values 1, 2, or 3. The set of coefficients $\{a_{ij}\}$ describes the transformation.

Consider, for example, the coordinate transformation described by

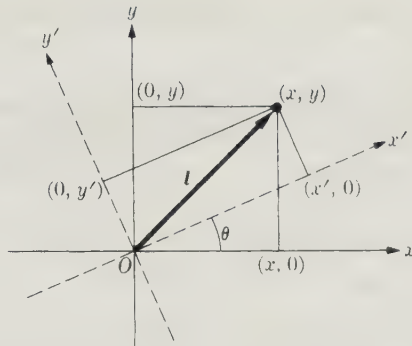
$$\begin{aligned}x' &= x \cos \theta + y \sin \theta \\y' &= -x \sin \theta + y \cos \theta \\z' &= z\end{aligned}\tag{II-2}$$

This linear transformation describes a rotation about the z -axis through the angle θ , in which the x - and y -axes are transformed into the x' - and y' -axes, respectively (see Fig. II-1). The length of the position vector, $l = \sqrt{x^2 + y^2 + z^2}$, is left invariant by the transformation since it is evident from (II-2) that

$$x^2 + y^2 + z^2 = x'^2 + y'^2 + z'^2\tag{II-3}$$

FIGURE II-1

Coordinate system rotation in two dimensions. The heavy line is the projection of vector l onto the xy -plane.



Equation (II-2) is an example of an *orthogonal transformation* in three dimensions, an orthogonal transformation being one which is real and leaves the length of a vector unchanged. Properties of orthogonal transformations will be discussed in more detail in the following paragraphs.

A transformation is orthogonal if it leaves the length of the displacement vector (or equivalently if it leaves $\sum x_i^2$) unchanged. Let us assume that (II-1a) is an orthogonal transformation. Then,

$$\sum_i (x'_i)^2 = \sum_k x_k^2 \quad (\text{II-4})$$

But

$$(x'_i)^2 = \sum_j \sum_k a_{ij} a_{ik} x_j x_k$$

and

$$\sum_i (x'_i)^2 = \sum_j \sum_k \sum_i a_{ij} a_{ik} x_j x_k \quad (\text{II-5})$$

Equations (II-4) and (II-5) agree only if the quantity

$$\sum_i a_{ij} a_{ik} = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases} \quad (\text{II-6})$$

This last equation may be written more compactly by introducing the Kronecker delta δ_{jk} , which was defined in Section 2-9.

$$\sum_i a_{ij} a_{ik} = \delta_{jk} \quad (\text{II-6a})$$

Equation (II-6a) is the condition which must be imposed to make the transformation $\{a_{ij}\}$ orthogonal. We easily verify that the rotation, (II-2), satisfies this criterion.

The transformation (II-1) may be written symbolically as

$$\mathbf{X}' = \mathbf{A}\mathbf{X} \quad (\text{II-7})$$

where \mathbf{X}' is the position vector with transformed components (x'_1, x'_2, x'_3) , \mathbf{X} is the position vector with original components (x_1, x_2, x_3) , and \mathbf{A} is regarded as a matrix operator. In fact \mathbf{A} is the matrix of the coefficients $\{a_{ij}\}$:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad (\text{II-8})$$

If we display the vectors \mathbf{X}' and \mathbf{X} as column matrices, we may write (II-7) as

$$\begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \quad (\text{II-7a})$$

The transformation (II-1) follows from this last equation through the usual laws of matrix multiplication. Thus the matrix \mathbf{A} given by (II-8) and Eq. (II-1a) are equivalent ways of describing the coordinate transformation.

The inverse of a transformation must bring us back again to the original set of coordinates. Thus, if $\{b_{ij}\}$ is the inverse transformation to $\{a_{ij}\}$, then

$$x_j = \sum_i b_{ji} x'_i \quad (\text{II-9})$$

Combining this with (II-1a), we find

$$x_j = \sum_k \sum_i b_{ji} a_{ik} x_k$$

which is an identity if

$$\sum_i b_{ji} a_{ik} = \delta_{jk} \quad (\text{II-10})$$

Equation (II-10) is the condition imposed in order for \mathbf{B} to be the inverse transformation to \mathbf{A} . If, in addition, \mathbf{A} is an orthogonal transformation, then (II-6a) holds. Comparison of (II-6a) and (II-10) shows that if matrix \mathbf{B} is constructed so that

$$b_{ji} = a_{ij} \quad (\text{II-11})$$

then \mathbf{B} will indeed describe the inverse transformation. According to (II-11), matrix \mathbf{B} is constructed from \mathbf{A} by interchanging its rows and columns. The new matrix is called the transpose of \mathbf{A} and is given the symbol $\tilde{\mathbf{A}}$. Thus *the inverse of an orthogonal transformation is the transpose of the original transformation.*

In Chapter 1, we defined a vector function as a quantity possessing both

direction and magnitude at each point in space. An alternative definition is:

A vector is a quantity whose components transform under an orthogonal transformation as the components of the position vector.

Thus, if \mathbf{F} is any vector function, its transform \mathbf{F}' , resulting from an orthogonal transform \mathbf{A} , is

$$\mathbf{F}' = \mathbf{A}\mathbf{F} \quad (\text{II-12})$$

Scalar functions of position, such as the length of a vector or the dot product of two vectors, are unchanged by an orthogonal transformation, or are *invariant*.

In addition to scalars and vectors, there are other more complicated quantities. One of these, whose transformation properties we need, is the *second-rank tensor* or simply *tensor*. A tensor is a quantity with two-index components; thus a component of tensor \mathbf{T} is T_{ij} , where i and j can each take on the values 1, 2, and 3. The reader will be exposed to the quadrupole moment tensor Q_{ij} (Section 2-9), the dielectric tensor ϵ_{ij} for anisotropic media (Section 16-3), and the four-dimensional field tensor $F_{\mu\nu}$ (Chapter 22). A familiar example from mechanics is the moment of inertia tensor. A linear relationship between two vector quantities can be expressed in terms of a second-rank tensor. Thus, the angular momentum \mathbf{L} of a rigid body can be related to its angular velocity $\boldsymbol{\omega}$ by means of the moment of inertia tensor \mathbf{I} :

$$\mathbf{L} = \mathbf{I}\boldsymbol{\omega}$$

or

$$L_i = \sum_j I_{ij}\omega_j$$

The tensor itself can be expressed in matrix form:

$$\mathbf{I} = \begin{bmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{bmatrix}$$

Suppose two vectors \mathbf{F} and \mathbf{X} are linearly related through the tensor relationship

$$\mathbf{F} = \mathbf{T}\mathbf{X} \quad (\text{II-13})$$

Now if an orthogonal transformation \mathbf{A} is performed, \mathbf{X} will be converted into \mathbf{X}' and \mathbf{F} into \mathbf{F}' . We should be able to express Eq. (II-13) in the transformed system as

$$\mathbf{F}' = \mathbf{T}'\mathbf{X}' \quad (\text{II-14})$$

where \mathbf{T}' is the transform of \mathbf{T} . But

$$\begin{aligned} F'_i &= \sum_j a_{ij} F_j = \sum_j \sum_k a_{ij} T_{jk} X_k \\ &= \sum_j \sum_k \sum_m a_{ij} T_{jk} (\bar{a})_{km} X'_m \\ &= \sum_m \left[\sum_j \sum_k a_{ij} T_{jk} (\bar{a})_{km} \right] X'_m \end{aligned} \quad (\text{II-15})$$

with $(\bar{a})_{km} \equiv a_{mk}$. For Eqs. (II-14) and (II-15) to be compatible,

$$T'_{im} = \sum_j \sum_k a_{ij} T_{jk} (\bar{a})_{km} \quad (\text{II-16})$$

This last equation expresses the transformation law of a second-rank tensor under an orthogonal transformation. Equation (II-16) also expresses the rule for multiplying three matrices together to find the i, m -component of the resulting matrix. Thus, (II-16) is written symbolically as

$$\mathbf{T}' = \mathbf{AT}\bar{\mathbf{A}} = \mathbf{ATA}^{-1} \quad (\text{II-16a})$$

Let us finally determine the transformation law for the vector differential operator ∇ . Consider the i th component of the gradient of a scalar invariant function φ ; we think of φ as a function of the coordinates x_j , which are in turn functions of the new coordinate x'_i . Using the chain rule of partial differentiation, we find

$$\frac{\partial \varphi}{\partial x'_i} = \sum_j \frac{\partial \varphi}{\partial x_j} \frac{\partial x_j}{\partial x'_i}$$

From Eq. (II-9),

$$\frac{\partial \varphi}{\partial x'_i} = \sum_j b_{ji} \frac{\partial \varphi}{\partial x_j} \quad (\text{II-17})$$

and from (II-11) for an orthogonal transformation

$$\frac{\partial}{\partial x'_i} = \sum_j a_{ij} \frac{\partial}{\partial x_j} \quad (\text{II-18})$$

Thus the components of ∇ transform like those of the position vector, (II-1a) under an orthogonal transformation.*

* Under a more general transformation, quantities that transform by (II-17) have to be distinguished from those that transform by (II-1a). They are called *covariant vectors* and *contravariant vectors*, respectively, but we will not need this distinction.

Systems of Units

In this book, the charge-rationalized mks system of units has been used. This system has the virtue of including the practical electrical units of potential difference (volt), current (ampere), resistance (ohm), and so on. As a result, the system rapidly gained favor with electrical engineers and is now the international standard. In other areas, notably atomic, solid-state, plasma, and nuclear physics, another system, known as the *Gaussian system*, has remained popular. Most other systems have faded from use, and hence only the Gaussian system will be discussed here at length.

In the mks system of units, the appearance of the numbers ϵ_0 and μ_0 in the formulation of Coulomb's law and the Biot law, respectively, causes an apparent difficulty. The difficulty is simply that Coulomb's law,

$$\mathbf{F}_2 = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2 \mathbf{r}_{12}}{r_{12}^3} \quad (\text{III-1})$$

cannot be used to define the coulomb unless ϵ_0 is known. By the same token, it cannot be used to define ϵ_0 unless the coulomb is previously defined. A technical point is that since ϵ_0 is taken as an experimentally determined number, using (III-1) to define the coulomb would result in a coulomb that would change every time ϵ_0 was redetermined. Thus it is clear that (III-1) should be used to define ϵ_0 , with the coulomb otherwise defined.

A corresponding difficulty does not arise in the magnetic case because one takes $\mu_0 = 4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}$, by definition. As a result, the expression

$$\frac{F}{l} = \frac{\mu_0}{2\pi} \frac{II'}{r} \quad (\text{III-2})$$

for the force per unit length between two parallel, current-carrying wires

can be used to define the ampere:

One ampere is that steady current which, when present in each of two long parallel conductors separated by a distance of one meter, results in a force per meter of length between them numerically equal to 2×10^{-7} N/m.

Of course, any other geometry could be used and would result in an equally unambiguous (and, in fact, numerically identical) definition of the ampere.

Having thus defined the ampere, the coulomb is defined as the charge transported by a steady current of one ampere flowing for one second. This in turn makes it feasible to use (III-1) to define ϵ_0 . There is thus no real problem, but only an artificial one arising from a desire to treat the mathematically simpler case of electrostatics before discussing the magnetic interaction of currents.

It is sometimes thought that this problem does not arise if the cgs Gaussian system of units is used. This idea is true only in the sense that the coefficient in Coulomb's law is chosen to be $1 \text{ dyne cm}^2/\text{esu}^2$, which places the burden of agreeing with experiment on the magnetic interactions. Thus, the velocity of light appears in the expression for the force between current-carrying conductors. Since in the usual treatment the conveniently defined quantity occurs first, this problem is less painfully obvious in the Gaussian system than in the mks system.

The Gaussian system is a combination of two earlier systems: the electrostatic system, esu, and the electromagnetic system, emu. The electrostatic system resulted from writing Coulomb's law in the form

$$\mathbf{F}_2 = \frac{q_1 q_2 \mathbf{r}_{12}}{r_{12}^3} \quad (\text{III-3})$$

and defining the esu of charge as that charge which when placed one centimeter from an exactly similar charge experiences a force of one dyne. It is obvious that the esu of charge is much smaller than the coulomb (in fact $1 \text{ coulomb} = 3 \times 10^9 \text{ esu}$). The electromagnetic system resulted from writing the Biot law, Eq. (8-25), without the factor $\mu_0/4\pi$ and defining the abampere as the current that, when present in a long straight wire, results in a force of 1 dyne/cm when the wire is placed 1 cm from a parallel conductor carrying the same current. From $|\mu_0/4\pi| = 10^{-7}$ and $1 \text{ newton} = 10^5 \text{ dynes}$, it is found that $1 \text{ abampere} = 10 \text{ ampere}$.

Either of the two starting points noted above could be used to initiate the development of a complete system of units. Historically, however, esu was used primarily for electrostatic problems and emu for electromagnetic problems. This being the case, it was natural that a hybrid system using esu for electrical quantities and emu for magnetic quantities should develop. The system that evolved in this way is known as the Gaussian system. The principal point of contact of esu and emu in the Gaussian system is in the

current density, where

$$\mathbf{J}_{\text{cmu}} = \frac{\mathbf{J}_{\text{esu}}}{c} \quad (\text{III-4})$$

In the Gaussian system we use \mathbf{J}_{esu} and explicitly exhibit the velocity of light in the magnetic equations. Thus, the Biot law has the form

$$d\mathbf{F}_2 = \frac{I_1 I_2 d\mathbf{l}_2 \times (d\mathbf{l}_1 \times \mathbf{r}_{12})}{c^2 r_{12}^3} \quad (\text{III-5})$$

with I in esu/s.

In Gaussian units, Maxwell's equations are

$$\begin{aligned} \nabla \times \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} &= 0 \\ \nabla \cdot \mathbf{D} &= 4\pi\rho \\ \nabla \times \mathbf{H} - \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t} &= \frac{4\pi\mathbf{J}}{c} \\ \nabla \cdot \mathbf{B} &= 0 \end{aligned} \quad (\text{III-6})$$

The fields are derived from scalar and vector potentials by means of

$$\mathbf{B} = \nabla \times \mathbf{A} \quad \text{and} \quad \mathbf{E} = -\nabla\varphi - \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} \quad (\text{III-7})$$

and the Lorentz force is

$$\mathbf{F} = q\left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B}\right) \quad (\text{III-8})$$

\mathbf{D} and \mathbf{B} are related to \mathbf{E} and \mathbf{H} by

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} \quad \text{and} \quad \mathbf{B} = \mathbf{H} + 4\pi\mathbf{M} \quad (\text{III-9})$$

where \mathbf{P} is the electric dipole moment ($\mathbf{p} = q\mathbf{l}$) per unit volume, and \mathbf{M} is the magnetic dipole moment ($\mathbf{m} = I\mathbf{An}/c$) per unit volume. These equations are substantially sufficient to define the Gaussian system of units. In addition, the energy density is

$$u = \frac{1}{8\pi} (\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H})$$

and the Poynting vector is

$$\mathbf{S} = \frac{c}{4\pi} \mathbf{E} \times \mathbf{H}$$

For convenience Table III-1 gives the numerical relationships of Gaussian units to mks units.

TABLE III-1

Quantity	Symbol	Gaussian units*	mks units
Capacitance	C	9×10^{11} cm (esu/statvolt)	= 1 F
Charge	Q	3×10^9 esu	= 1 C
Conductivity	g	9×10^9 s ⁻¹	= 1 ($\Omega \cdot \text{m}$) ⁻¹
Current	I	3×10^9 esu/s (=10 ⁻¹ abamp)	= 1 A
Electric displacement	D	$12\pi \times 10^5$ dyne/esu	= 1 C/m ²
Electric field	E	$\frac{1}{3} \times 10^{-4}$ dyne/esu	= 1 V/m
Energy	U	10^7 erg	= 1 J
Force	F	10^5 dyne	= 1 N
Inductance	L	$\frac{1}{9} \times 10^{-11}$ s ² /cm	= 1 H
Magnetic flux	Φ	10^8 maxwell	= 1 Wb
Magnetic induction	B	10^4 gauss	= 1 T
Magnetic intensity	H	$4\pi \times 10^{-3}$ oersted (gauss)	= 1 A/m
Potential	φ	$1/300$ erg/esu (statvolt)	= 1 V
Resistance	R	$\frac{1}{9} \times 10^{-11}$ s/cm (statvolt · s/esu)	= 1 Ω

* The factor 3 ($9 = 3^2$) in the conversions comes from the velocity of light, $c \cong 3 \times 10^8$ m/s. For greater accuracy, use 2.9979 instead of 3.

Some advantages of the Gaussian system, which account for its continued use in the physics literature, are that the factor $4\pi\epsilon_0$ does not occur in all the equations of atomic physics, where the Coulomb potential plays a central role, and that the velocity v always enters in the dimensionless form v/c prescribed by the Lorentz transformation. In relativity, it is natural for \mathbf{E} and \mathbf{B} to have the same units, since their components are simply different elements of the field tensor. There is no reason for the auxiliaries \mathbf{D} and \mathbf{H} to have other units. It is a convenience that in material media the permittivity and susceptibility are dimensionless and related by

$$\epsilon = 1 + 4\pi\chi_e$$

like the corresponding magnetic quantities,

$$\mu = 1 + 4\pi\chi_m$$

Vector Differential Operators

1. Rectangular Coordinates

$$\nabla\varphi = \mathbf{i} \frac{\partial\varphi}{\partial x} + \mathbf{j} \frac{\partial\varphi}{\partial y} + \mathbf{k} \frac{\partial\varphi}{\partial z}$$

$$\nabla \cdot \mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}$$

$$\nabla \times \mathbf{F} = \mathbf{i} \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right) + \mathbf{j} \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} \right) + \mathbf{k} \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right)$$

2. Cylindrical Coordinates

$$\nabla\varphi = \mathbf{a}_r \frac{\partial\varphi}{\partial r} + \mathbf{a}_\theta \frac{1}{r} \frac{\partial\varphi}{\partial\theta} + \mathbf{k} \frac{\partial\varphi}{\partial z}$$

$$\nabla \cdot \mathbf{F} = \frac{1}{r} \frac{\partial}{\partial r} (rF_r) + \frac{1}{r} \frac{\partial F_\theta}{\partial\theta} + \frac{\partial F_z}{\partial z}$$

$$\nabla \times \mathbf{F} = \mathbf{a}_r \left(\frac{1}{r} \frac{\partial F_z}{\partial\theta} - \frac{\partial F_\theta}{\partial z} \right) + \mathbf{a}_\theta \left(\frac{\partial F_r}{\partial z} - \frac{\partial F_z}{\partial r} \right) + \mathbf{k} \frac{1}{r} \left(\frac{\partial}{\partial r} (rF_\theta) - \frac{\partial F_r}{\partial\theta} \right)$$

3. Spherical Coordinates

$$\nabla\varphi = \mathbf{a}_r \frac{\partial\varphi}{\partial r} + \mathbf{a}_\theta \frac{1}{r} \frac{\partial\varphi}{\partial\theta} + \mathbf{a}_\phi \frac{1}{r \sin\theta} \frac{\partial\varphi}{\partial\phi}$$

$$\nabla \cdot \mathbf{F} = \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 F_r) + \frac{1}{r \sin\theta} \frac{\partial}{\partial\theta} (F_\theta \sin\theta) + \frac{1}{r \sin\theta} \frac{\partial F_\phi}{\partial\phi}$$

$$\begin{aligned} \nabla \times \mathbf{F} = & \mathbf{a}_r \frac{1}{r \sin \theta} \left[\frac{\partial}{\partial \theta} (F_\phi \sin \theta) - \frac{\partial F_\theta}{\partial \phi} \right] \\ & + \mathbf{a}_\theta \frac{1}{r} \left[\frac{1}{\sin \theta} \frac{\partial F_r}{\partial \phi} - \frac{\partial (r F_\phi)}{\partial r} \right] + \mathbf{a}_\phi \frac{1}{r} \left[\frac{\partial (r F_\theta)}{\partial r} - \frac{\partial F_r}{\partial \theta} \right] \end{aligned}$$

The Laplacian of a scalar quantity, $\nabla^2 \varphi$, is given in Section 3-1 for the three coordinate systems. Vector identities are given in Table 1-1.

The Dirac Delta Function

In one dimension, the delta function is defined by

$$\begin{aligned}\delta(x) &= 0 \quad \text{for } x \neq 0 \\ \int_{-\infty}^{+\infty} \delta(x) dx &= 1\end{aligned}\tag{V-1}$$

This function can be understood in an elementary way as the limit of an ordinary continuous function. For example, the Gaussian function

$$\frac{1}{\sqrt{\pi} \epsilon} e^{-x^2/\epsilon^2}$$

has a peak at $x = 0$, which is higher and narrower the smaller is ϵ , and its integral from $-\infty$ to $+\infty$ is 1 for any value of ϵ . Thus, we could define

$$\delta(x) = \lim_{\epsilon \rightarrow 0} \frac{1}{\sqrt{\pi} \epsilon} e^{-x^2/\epsilon^2}\tag{V-2}$$

Another useful representation is

$$\delta(x) = \lim_{\epsilon \rightarrow 0} \frac{\epsilon \sin^2 x/\epsilon}{\pi x^2}\tag{V-3}$$

As a third example, we may use the Lorentzian function introduced in Chapter 19:

$$\delta(x) = \lim_{\epsilon \rightarrow 0} \frac{\epsilon}{\pi x^2 + \epsilon^2}\tag{V-4}$$

An important property of the δ -function is

$$\delta(ax) = \frac{1}{a} \delta(x) \quad (\text{V-5})$$

where a is a constant. This expression follows directly from Eq. (V-1); first we change variable, x to u , which does not change the result. Then we let $u = ax$, which yields

$$\int_{-\infty}^{\infty} \delta(ax) dx = \frac{1}{a} \quad (\text{V-6})$$

The mean-value theorem of calculus gives

$$\int_{-a}^{+a} F(x)f(x) dx \cong F(0) \int_{-a}^{+a} f(x) dx$$

so that as $f(x) \rightarrow \delta(x)$

$$\int F(x) \delta(x) dx = F(0) \quad (\text{V-7})$$

In all of the above equations, we can replace the variable x by $(x - x_0)$; in particular,

$$\int F(x) \delta(x - x_0) dx = F(x_0) \quad (\text{V-8})$$

In any equation involving the delta function, the domain of integration can be reduced to any interval containing the point where the argument of the delta function vanishes.

An extension to three dimensions can be achieved by writing

$$\delta(\mathbf{r}) dv = \delta(x) \delta(y) \delta(z) dx dy dz \quad (\text{V-9})$$

In Chapter 2, we found that

$$\delta(\mathbf{r}) = \nabla^2 \left(\frac{-1}{4\pi r} \right) \quad (\text{V-10})$$

Fourier Integrals (Fourier Transforms)

A real function that is continuous except for a finite number of finite discontinuities in the interval $-l/2 < t < l/2$ can be expanded in a complex Fourier series:

$$f(t) = \sum_{n=-\infty}^{\infty} a_n e^{-2\pi i n t / l} \quad (\text{VI-1a})$$

$$a_n = \frac{1}{l} \int_{-l/2}^{l/2} f(t) e^{2\pi i n t / l} \quad (\text{VI-1b})$$

In order that $f(t)$ be real, a_n and a_{-n} must be complex conjugates. We wish to consider the case where $l \rightarrow \infty$. Let us define $\omega = 2\pi n / l$ and $l a_n = \sqrt{2\pi} \bar{f}(\omega)$; then we may convert the sum in (VI-1a) to an integral since $\sum F_n = \int F_n dn \rightarrow (l/2\pi) \int F(\omega) d\omega$. Equations (VI-1) become

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \bar{f}(\omega) e^{-i\omega t} d\omega \quad (\text{VI-2})$$

$$\bar{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{i\omega t} dt$$

In this form $f(t)$ and $\bar{f}(\omega)$ appear as a transform pair commonly called *Fourier transforms*. The factor $1/2\pi$ has been symmetrically apportioned, and we shall consistently use this convention, which although convenient, is not essential. In fact, changing both variables, ω and t , by dividing each by $\sqrt{2\pi}$ requires one to insert a factor of 2π in the exponents and remove the factors $1/\sqrt{2\pi}$ from in front of the integrals (this procedure is convenient for computer applications).

Equations (VI-2) may be combined to give

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} dt' f(t') e^{-i\omega(t-t')} \quad (\text{VI-3})$$

This equation is a statement of the *Fourier integral theorem*.

Fourier transforms can be formulated using only the physically defined positive frequencies, when $f(t)$ is nonzero only for positive values of t or when $f(t)$ is either even or odd. In the latter case, we may define new functions, $f_e(t)$ and $f_o(t)$, which are even or odd by

$$f_e(t) = f_e(-t), \quad f_o(t) = -f_o(-t) \quad (\text{VI-4})$$

(Every function can be expressed as the sum of an even and odd function.) Using Eqs. (VI-4) in Eqs. (VI-2) to transform the range of integration that extends over negative values of t , and noting that $f(t)$ is real we get

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} [\bar{f}(\omega) e^{-i\omega t} + \text{c.c.}] d\omega \quad (\text{VI-5})$$

$$f_e(t) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} \bar{f}_e(\omega) \cos \omega t d\omega \quad (\text{VI-6a})$$

$$\bar{f}_e(\omega) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} f_e(t) \cos \omega t dt$$

$$f_o(t) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} \bar{f}_o(\omega) \sin \omega t d\omega \quad (\text{VI-6b})$$

$$\bar{f}_o(\omega) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} f_o(t) \sin \omega t dt$$

We can also use the second equation in (VI-2) to find a useful representation of the Dirac delta function. We set $f(t) = \delta(t)$ so that

$$\bar{\delta}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \delta(t) e^{i\omega t} = \frac{1}{\sqrt{2\pi}} \quad (\text{VI-7})$$

The Fourier transform of the δ -function is a constant. Using this result in the first Eq. (VI-2) gives

$$\delta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega e^{-i\omega t} \quad (\text{VI-8})$$

The Fourier transform of a constant is the delta function. Changing variable t to $t - t_0$ gives

$$\delta(t - t_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega e^{-i\omega(t-t_0)} \quad (\text{VI-8a})$$

Taking the real part gives

$$\delta(t - t_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \cos \omega(t - t_0) \quad (\text{VI-8b})$$

and the imaginary part gives

$$\int_{-\infty}^{\infty} d\omega \sin \omega(t - t_0) = 0 \quad (\text{VI-8c})$$

In this calculation, we have violated an assumption about the continuity of $f(t)$, but the fundamental theorem can be generalized to accommodate this difficulty.

This representation of the Dirac delta function can be used to prove a very important theorem, the *convolution theorem*, which states that if

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(t'')h(t - t'') dt'' \quad (\text{VI-9})$$

then

$$\tilde{f}(\omega) = \tilde{g}(\omega)\tilde{h}(\omega) \quad (\text{VI-10})$$

There are many applications of Fourier integrals. One occurs in Section 19-6 and another in Appendix VII. Fourier integrals are one of the most valuable theoretical tools of the physicist and engineer.

Functions of a Complex Variable: Dispersion Relations

This appendix contains some notes on mathematical techniques involving complex variables. Since all physical quantities can be represented as real functions of real variables, such techniques are not essential to formulating or understanding electromagnetic theory. They may, however, facilitate and simplify the analysis.

VII-1

FUNCTIONS OF A COMPLEX VARIABLE

We assume a familiarity with the algebra of complex numbers: addition, subtraction, multiplication, division, exponentiation, complex conjugation, and so forth. The most useful results of analysis are stated in the following, usually without proof.

The definition of a complex function of a complex variable is a rule that associates with every pair of real numbers, x and y , representing the complex variable $z = x + iy$, another pair of real numbers, u and v , representing the complex quantity $w(z) = u(x, y) + iv(x, y)$. This statement is a simple extension of the definition of a function of a real variable, but it is so broad that no results useful in physics can be proved unless some restrictions on the functions are assumed. If the requirement that $w(z)$ be differentiable with respect to z is added, then $u(x, y)$ and $v(x, y)$ are related. The derivative of $w(z)$ with respect to z is defined formally as it is for real functions,

$$\frac{dw}{dz} = \lim_{z \rightarrow z_0} \left[\frac{w(z) - w(z_0)}{z - z_0} \right] \quad (\text{VII-1})$$

If this limit exists and is *independent* of the direction from which z approaches z_0 , the function $w(z)$ is said to be differentiable. Independence

of direction is the crucial difference between complex analysis and real analysis; there are now infinitely many directions from which z can tend to z_0 in the limit. A single-valued function of a complex variable that is continuous and differentiable in a region of the complex plane is called an *analytic* or *regular* or *holomorphic* function in that region. Because differentiability is a more severe restriction in complex analysis, many more results can be proved about analytic functions than about real functions.

The first important result is that for an analytic function $w(z) = u + iv$:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \quad (\text{VII-2})$$

The necessity of these conditions, which are known as the *Cauchy-Riemann differential equations*, can be proved most easily by taking the derivatives with $z - z_0$ first equal to δx and then equal to $i \delta y$. If the real and imaginary parts of the derivatives calculated in these two directions are set equal to one another, Eqs. (VII-2) result. These equations relate u and v , so they are not independent functions of x, y . We also note in passing that differentiating the first of these equations with respect to x and the second with respect to y and then adding gives

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \quad \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0 \quad (\text{VII-3})$$

so u (and v) satisfy Laplace's equation. This property of $w(z)$ forms the basis for a method of solving Laplace's equation in two dimensions (see Section 3-8).

Functions of a complex variable can also be integrated, according to a straightforward extension of the definition used for real functions:

$$\int_{aC}^b w(z) dz = \lim_{\delta z \rightarrow 0} \sum_i w(z_i) \delta z_i \quad (\text{VII-4})$$

where z_i is a point in δz_i and the limit means that the number of subintervals in the interval a to b becomes infinite while simultaneously every δz_i approaches zero. The C attached to the integral indicates the particular path of the many paths from a to b in the complex plane (as contrasted to just one in the real variable case), since they need not all yield the same value of the integral. The integral as defined by Eq. (VII-4) exists as long as $w(z)$ is continuous.

Several important theorems about integrals of a function of a complex variable can be proved. The first of these is *Cauchy's integral theorem* (sometimes called the fundamental theorem of the theory of functions), which states that for a function $w(z)$ that is analytic in a simply-connected region G and on a closed curve C inside that region,

$$\oint_C w(z) dz = 0 \quad (\text{VII-5})$$

FIGURE VII-1

Decomposition of contour C in the z' -plane into the sum of two contours. The \bullet indicates point z in all three cases.



This theorem is proved by considering a triangular-shaped curve and by (1) showing that the integral around a triangle is zero if the function is analytic, (2) showing that any closed polygon can be decomposed into triangular regions, (3) showing that the integral around the boundary of the polygon is the sum of the integrals around the triangles and consequently zero, and finally (4) showing that any simply-connected region can be approximated arbitrarily closely by a polygon.

From Cauchy's theorem there follows a series of formulas known as Cauchy's integral formulas. The first of these is

$$w(z) = \frac{1}{2\pi i} \oint_C \frac{w(z')}{z' - z} dz' \quad (\text{VII-6})$$

where $w(z')$ is analytic within a simply-connected region containing C , a closed curve enclosing the point z . This result is derived as follows (refer to Fig. VII-1): Replace the curve in (a) by the two curves in (b). The integral of $w(z') dz' / (z' - z)$ around C is equivalent to the integral of the function around C' plus an additional integral around the circle, since the two vertical legs in curve C' are traversed in opposite directions and the two circular traversals around point z are done oppositely. Now from the Cauchy integral theorem, Eq. (VII-5), the integral of $w(z') dz' / (z' - z)$ around curve C' is zero, since there are no enclosed singularities of the integrand. Thus, the entire contribution to the integral in Eq. (VII-6) comes from the small circle centered on z ; this integral can be evaluated directly, giving the result in (VII-6).

The other formulas are for the n th order derivatives of w :

$$\frac{d^n}{dz^n} w(z) = \frac{n!}{2\pi i} \oint \frac{w(z')}{(z' - z)^{n+1}} dz' \quad (\text{VII-7})$$

Informal derivations of these formulas can be made by successively differentiating Eq. (VII-6) with respect to z , but a rigorous proof is also possible. Thus, an analytic function possesses derivatives of all orders (not necessarily so for a differentiable function of a real variable).

PROOF OF THE DISPERSION RELATIONS

In Chapter 19, we demonstrated that the dispersion relations were valid for a system that approximates the classical damped harmonic oscillator (the prototype linear system) and then assumed that they applied to any linear causal system. We now provide a proof. We begin by considering the integral

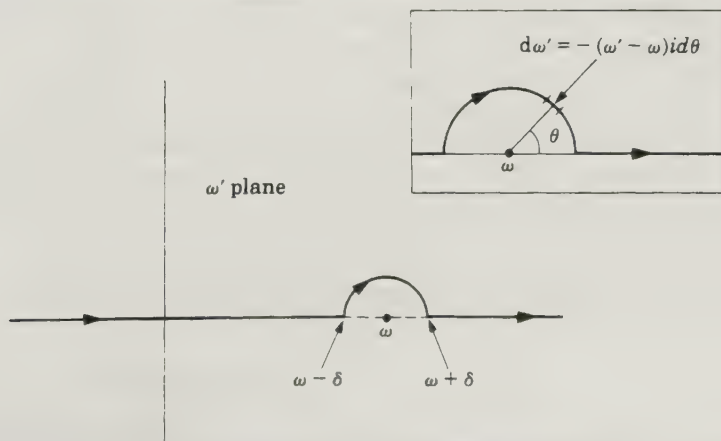
$$K(\omega) = \int_{-\infty}^{\infty} \frac{\tilde{R}(\omega')}{\omega' - \omega} d\omega' \quad (\text{VII-8})$$

where ω is real and the integral is taken along the real axis in the ω' -plane (treating ω' as a complex variable). The integrand at $\omega' = \omega$, and therefore the integral, are both undefined. This problem is dealt with by integrating up to a point $\omega - \delta$ and then going around a small semicircle above the real axis to the point $\omega + \delta$. From this point, the integration is continued to the upper limit as is shown in Fig. VII-2. The integral in Eq. (VII-8) is now defined as the limit as δ goes to zero of the integral calculated as just described. The integral around the small semicircle can be evaluated directly using the geometry shown in the inset of Fig. VII-2; it is just $-i\pi\tilde{R}(\omega)$, and $K(\omega)$ is found to be

$$K(\omega) = \lim_{\delta \rightarrow 0} \left\{ \int_{-\infty}^{\omega - \delta} \frac{\tilde{R}(\omega')}{\omega' - \omega} d\omega' + \int_{\omega + \delta}^{\infty} \frac{\tilde{R}(\omega')}{\omega' - \omega} d\omega' \right\} - i\pi\tilde{R}(\omega) \quad (\text{VII-9})$$

What we have done is to define a path of integration that leaves the singularity of the integrand below the path. The remaining two integrals over a real variable are well behaved and can be calculated by standard techniques if $\tilde{R}(\omega)$ is known. The limit of expressions like the sum of these

FIGURE VII-2

 Integration path for evaluating $K(\omega)$, Eq. (VII-8)


two remaining integrals occurs so frequently that it has been named the *Cauchy principal value* and defined by

$$\lim_{\delta \rightarrow 0} \left\{ \int_A^{\omega - \delta} f(\omega') d\omega' + \int_{\omega + \delta}^A f(\omega') d\omega' \right\} = P \int_A^A f(\omega') d\omega' \quad (\text{VII-10})$$

The Cauchy principal value, or frequently just the principal value, is designated by the P in front of the integral sign. Consolidating what has been done so far, Eq. (VII-9) becomes

$$K(\omega) = P \int_{-\infty}^{\infty} \frac{\tilde{R}(\omega')}{\omega' - \omega} d\omega' - i\pi\tilde{R}(\omega) \quad (\text{VII-11})$$

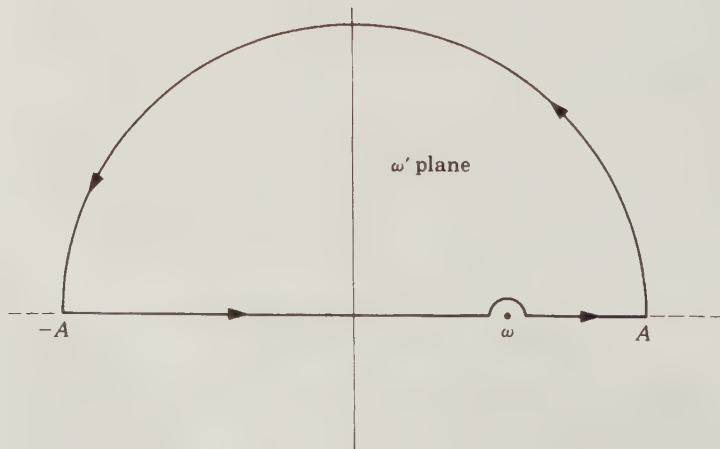
We now begin to see the desired relationship: If the real part of Eq. (VII-11) is taken, then the right-hand side involves an integral of the real part of \tilde{R} in the first term and the imaginary part of \tilde{R} in the second term. If $K(\omega)$ were known in some independent way, the dispersion relation problem would be solved.

Since ω' is considered as a complex variable then we can construct a contour as shown in Fig. VII-3. If $\tilde{R}(\omega')$ is analytic in the upper-half plane, then by Cauchy's theorem the integral of $\tilde{R}(\omega')$ around the entire contour shown in Fig. VII-3 is zero.

It remains to consider the integral around the large semicircle. The magnitude of this integral is less than $\pi A/(A - \omega)$ times the maximum value of $|\tilde{R}(\omega')|$ at any point on the contour. The integral will go to zero if $|\tilde{R}(\omega')|$ vanishes in any way whatsoever as A goes to infinity. Thus, the integral around the entire contour is zero, and the integral around the large

FIGURE VII-3

Closed contour for evaluating the function $K(\omega)$



semicircle is zero. Consequently, what is left, $K(\omega)$, must also be zero. Therefore,

$$K(\omega) = P \int_{-\infty}^{\infty} \frac{\tilde{R}(\omega')}{\omega' - \omega} d\omega' - i\pi\tilde{R}(\omega) = 0 \quad (\text{VII-12})$$

This expression may be written

$$\tilde{R}(\omega) = \frac{1}{\pi i} P \int_{-\infty}^{\infty} \frac{\tilde{R}(\omega')}{\omega' - \omega} d\omega' \quad (\text{VII-13})$$

or

$$R_r(\omega) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{R_i(\omega')}{\omega' - \omega} d\omega' \quad (\text{VII-14})$$

$$R_i(\omega) = -\frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{R_r(\omega')}{\omega' - \omega} d\omega'$$

which are the dispersion relations.

These dispersion relations are valid if $\tilde{R}(\omega')$, considered as a function of a complex variable, is analytic in the upper-half plane and vanishes as $|\omega'|$ goes to infinity, also in the upper-half plane. Writing $\tilde{R}(\omega)$ as the Fourier transform of $R(t)$, the response function described in Section 19-1,

$$\tilde{R}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} R(t)e^{i\omega t} dt \quad (\text{VII-15})$$

enables us to see that $\tilde{R}(\omega)$ approaches zero as $|\omega|$ goes to infinity if $R(t)$ satisfies the specified causality and boundedness conditions. If $R(t)$ is zero for $t < 0$, we need consider only $t > 0$. Then $i\omega t$ has a negative real part when ω has a positive imaginary part—that is, if ω is in the upper half of the ω -plane. Thus, as long as $R(t)$ is bounded, $\tilde{R}(\omega)$ will vanish as ω goes to infinity in the upper-half plane. That $\tilde{R}(\omega)$ is analytic can be seen by calculating the derivative with respect to ω of Eq. (VII-15). The derivative can be taken inside the integral since ω is a parameter as far as the integration is concerned. The function $\exp(i\omega t)$ is analytic everywhere; consequently, it has a unique derivative, and it follows that $\tilde{R}(\omega)$ has a unique derivative. Thus, all of the conditions required for the validity of the dispersion relations of $\tilde{R}(\omega)$ follow from the causality and boundedness conditions imposed on the response function $R(t)$ in Section 19-1.

By using the fact that $R(t)$ is real, the dispersion relations can be transformed to a form that depends only on values of the frequency response at positive values of the frequency. Since $R(t)$ is real, Eq. (VII-15) shows that

$$\tilde{R}(-\omega) = \tilde{R}^*(\omega) \quad (\text{VII-16})$$

or

$$R_i(-\omega) = -R_i(\omega), \quad R_r(-\omega) = R_r(\omega) \quad (\text{VII-17})$$

Using this result to convert the part of the integral in Eq. (VII-14) that extends over negative values of ω' to an integral over positive values of ω' gives

$$R_r(\omega) = \frac{1}{\pi} P \int_0^{\infty} R_i(\omega') \left[\frac{1}{\omega' - \omega} + \frac{1}{\omega' + \omega} \right] d\omega' \quad (\text{VII-18})$$

or

$$R_r(\omega) = \frac{2}{\pi} P \int_0^{\infty} \frac{\omega' R_i(\omega')}{\omega'^2 - \omega^2} d\omega' \quad (\text{VII-19a})$$

In an exactly parallel way,

$$R_i(\omega) = -\frac{2\omega}{\pi} P \int_0^{\infty} \frac{R_r(\omega')}{\omega'^2 - \omega^2} d\omega' \quad (\text{VII-19b})$$

The dispersion relations do not depend on what kind of linear physical system is being considered; it can be mechanical, electrical, quantum mechanical, or even an economic system, so long as a linear response function (with its Fourier transform) can be defined. The following simple examples demonstrate the use of the dispersion relations.

Calculation of $R_i(\omega)$ from the Dispersion Relation

To fix these ideas, we examine the simple circuit of Fig. 19-1. The frequency response function is just the voltage across the circuit resulting from a current with time dependence $\exp(-i\omega t)$ and unit amplitude as given by

$$\tilde{R}(\omega) = \frac{R}{1 + \omega^2 R^2 C^2} + i \frac{\omega R^2 C}{1 + \omega^2 R^2 C^2} \quad (\text{VII-20})$$

This equation is the complex conjugate of the impedance as usually defined because we have used the time dependence $\exp(-i\omega t)$. To show that the dispersion relation is satisfied we calculate $R_i(\omega)$ using Eq. (VII-14) with $R_r(\omega')$ from Eq. (VII-20). First add and subtract $R_r(\omega)/(\omega' - \omega)$ in the integrand. Then

$$R_i(\omega) = -\frac{1}{\pi} \left\{ P \int_{-\infty}^{\infty} \frac{R_r(\omega') - R_r(\omega)}{\omega' - \omega} d\omega' + R_r(\omega) P \int_{-\infty}^{\infty} \frac{d\omega'}{\omega' - \omega} \right\} \quad (\text{VII-21})$$

The second integral is just $\ln|\omega' - \omega|$ evaluated at the appropriate limits. Using $-A$ and A as the limits of integration we have

$$P \int_{-\infty}^{\infty} \frac{d\omega'}{\omega' - \omega} = \lim_{A \rightarrow \infty} [\ln(A - \omega) - \ln(A + \omega)] \\ + \lim_{\delta \rightarrow 0} [\ln|-\delta| - \ln|\delta|] = 0 \quad (\text{VII-22})$$

The limit at infinity must be taken symmetrically to be consistent with the

principal value. The remaining integral is evaluated by combining $R_r(\omega')$ and $R_r(\omega)$ as follows:

$$\begin{aligned} R_r(\omega') - R_r(\omega) &= R \left[\frac{1}{1 + \omega'^2 R^2 C^2} - \frac{1}{1 + \omega^2 R^2 C^2} \right] \\ &= - \frac{R^3 C^2 (\omega' - \omega)(\omega' + \omega)}{(1 + \omega'^2 R^2 C^2)(1 + \omega^2 R^2 C^2)} \end{aligned} \quad (\text{VII-23})$$

Using this result in Eq. (VII-21), we note, first, that $(\omega' - \omega)$ cancels in the numerator and denominator so that the integrand is well behaved everywhere and the principal value can be dropped and, second, that the integral of the term containing ω' in the numerator is an integral of an odd function over a symmetric interval and so vanishes. Combining these results gives

$$R_i(\omega) = \frac{1}{\pi} \frac{R\omega RC}{1 + \omega^2 R^2 C^2} \int_{-\infty}^{\infty} \frac{RC d\omega'}{1 + \omega'^2 R^2 C^2} = \frac{R\omega RC}{1 + \omega^2 R^2 C^2} \quad (\text{VII-24})$$

which, by comparison with Eq. (VII-20), is just $R_i(\omega)$.

The calculation done in this example is not particularly useful, and in fact, the computation of $R_i(\omega)$ from an analytic expression for $R_r(\omega)$ is almost never useful, since the two are usually obtained as the real and imaginary parts of a single calculation. What has been demonstrated, however, is that in this simple case the dispersion relations are exactly satisfied. In addition, the trick we used to eliminate the singularity at $\omega' = \omega$ is frequently useful in other cases—even (or especially) if the integral is to be evaluated numerically.

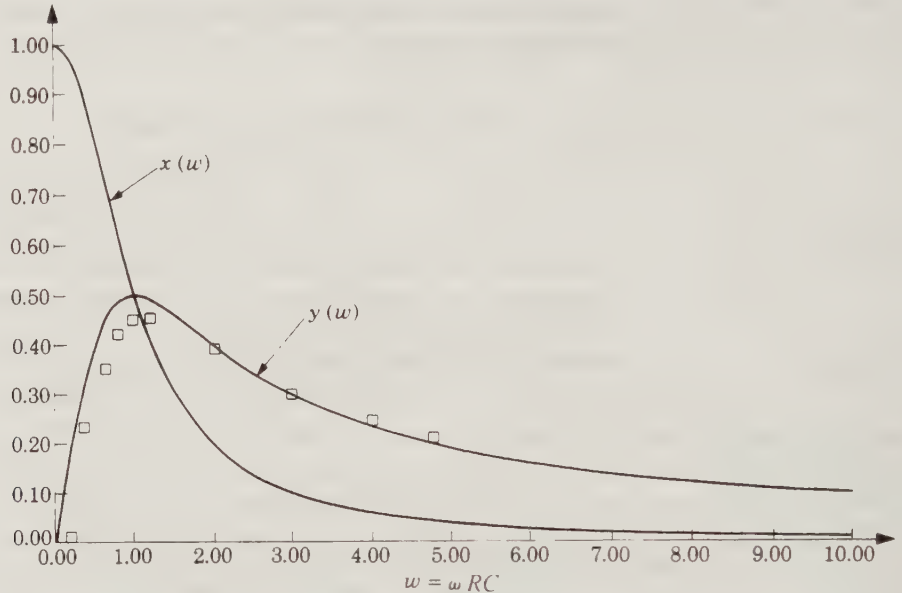
Calculation of $R_i(\omega)$ from Numerical Data for $R_r(\omega)$

As a second example we calculate the imaginary part of a response function from numerical data for the real part of the response function for the simple circuit we have been using as a demonstration. Suppose that the real part of the response function given in Eq. (VII-20) has been measured over some range of frequencies (all greater than zero), say from ω_1 to ω_2 at uniform intervals. (Practically speaking the measurements would most likely be made at constant frequency ratios, but uniform intervals simplify the numerical integrations we need to do.) We want to use these data to determine the imaginary part of the response function. First, Eq. (VII-20) is written in dimensionless form with $w = RC\omega$ and $z(w) = x(w) - iy(w) = \tilde{R}(w)/R$. Then, for $y(w)$, Eq. (VII-19b) becomes

$$y(w) = - \frac{2}{\pi} wP \int_0^{\infty} \frac{x(w')}{w'^2 - w^2} dw' \quad (\text{VII-25})$$

The numerical integrations can be done using the program in Fig. I-2 or a more sophisticated program. If $x(w)$ is known for $0 \leq w \leq 5$ at intervals of

FIGURE VII-4
 Calculation of $R_i(\omega)$
 from data for $R_r(\omega)$. See
 Eq. (VII-25). \square Exact
 values for $y(\omega)$



0.1 and x is essentially zero for $w > 5$, the numerical integration of Eq. (VII-25) reproduces $y(w)$ as shown in Fig. VII-4, where the exact values are shown for comparison. Note that in doing the numerical integration one must be careful to select points symmetrically located with respect to the value w at which $y(w)$ is being evaluated, at least for a region close to w . If this choice is not made, the cancellation that is implicit in the principal value does not work and erroneous results are obtained. Returning to Fig. VII-4, we see that the agreement is excellent for $w > 2.0$, but that there are significant deviations for smaller values of w . For small values of w , the integrand is very large for w' close to w . (The integrand is always large for w' close to w , but for $w = 1$, it is about ten times as large as it is for $w = 4$.) The discrepancy between the calculation and the exact values of $y(w)$ can be traced to this rather cavalier treatment of this part of the integral. Finally we note that in this example we have gone in the favorable direction—that is, from the real part to the imaginary part. Going the other way would be more difficult because of the long tail on the imaginary part. The data specified would not have been adequate to go the other way, and it would have been necessary to approximate $y(w)$ for large values of w by some reasonable function.

Answers to Odd-Numbered Problems

Chapter 1

- 1-1. $(\mathbf{A} - \mathbf{B}) \times (\mathbf{C} - \mathbf{D}) = 0, 1:3$
 1-3. $\mathbf{A} \cdot \mathbf{B} = 0, \mathbf{A} + \mathbf{B} = \mathbf{C}$
 1-7. The angle between \mathbf{R} and $\mathbf{R} - \mathbf{A}$, which is 90° , may be inscribed in a semicircle with \mathbf{A} as diameter. As \mathbf{R} varies, the various semicircles describe the surface of a sphere.
 1-13. $\mathbf{n} = (ax\mathbf{i} + by\mathbf{j} + cz\mathbf{k})/\sqrt{a^2x^2 + b^2y^2 + c^2z^2}$
 1-15. $\nabla \cdot \mathbf{F} = \frac{1}{r} \frac{\partial}{\partial r} (rF_r) + \frac{1}{r} \frac{\partial F_\theta}{\partial \theta} + \frac{\partial F_z}{\partial z}$
 1-17. No

Chapter 2

- 2-1. $\tan^3 \theta / (1 + \tan^2 \theta) = q^2 / 8\pi\epsilon_0 mg l^2$
 2-3. $\mathbf{E} = 3061 \text{ V/m}$ (along diagonal)
 2-5. (a) $E = (\sigma/2\epsilon_0)(1 - z/\sqrt{z^2 + R^2})$
 (b) $E = (\beta/2\epsilon_0) \left[\frac{L}{2} \left(\frac{L}{2} - \sqrt{\frac{L^2}{4} + R^2} \right) + R^2 \log \left(\frac{L}{2R} \sqrt{1 + \frac{L^2}{2R^2}} \right) \right]$
 2-7. (a) $x = a(3 + \sqrt{3})/2$, (b) saddlepoint
 2-9. $\varphi = (\rho/4\epsilon_0) \left[(z + \frac{1}{2}L) \{ (z + \frac{1}{2}L)^2 + R^2 \}^{1/2} - 2zL - (z - \frac{1}{2}L) \{ (z - \frac{1}{2}L)^2 + R^2 \}^{1/2} + R^2 \log \left\{ \frac{z + \frac{1}{2}L + \sqrt{(z + \frac{1}{2}L)^2 + R^2}}{z - \frac{1}{2}L + \sqrt{(z - \frac{1}{2}L)^2 + R^2}} \right\} \right]$
 2-11. (a) 240,000 V, (b) 4,730 km
 2-13. (a) $1.1 \times 10^{-12} \text{ C/m}^3$, (b) positive

- 2-15. (a) $\varphi = (A/\epsilon_0)(R - \frac{1}{2}r)$ for $r \leq R$
 $\varphi = AR^2/2\epsilon_0 r$ for $r \geq R$
 (b) $\varphi = (\rho_0/2\epsilon_0)(R^2 - \frac{1}{3}r^2)$ for $r \leq R$
 $\varphi = R^3\rho_0/3\epsilon_0 r$ for $r \geq R$
 2-17. (a) $\nabla \times \mathbf{r}/r^a = 0, \nabla \cdot \mathbf{r}/r^a = (3-a)/r^a$
 (b) $\rho = (3-a)q/4\pi r^a$
 (c) $\varphi = q/4\pi\epsilon_0(a-2)r^{a-2}(a \neq 2)$
 2-19. $\mathbf{E} = \frac{q}{4\pi\epsilon_0} (1/r + 1/\lambda)e^{-r/\lambda}\mathbf{r}/r^2$
 $\rho = q[\delta(\mathbf{r}) - 1/4\pi\lambda^2\mathbf{r}]e^{-r/\lambda}$
 2-21. (a) $\varphi = \frac{\lambda}{4\pi\epsilon_0} \ln \left\{ \frac{|z+a| + \sqrt{(z+a)^2 + r^2}}{|z-a| + \sqrt{(z-a)^2 + r^2}} \right\}$
 (b) $E_r = \frac{\lambda}{4\pi\epsilon_0 r} \times \left[\frac{a+z}{\sqrt{(a+z)^2 + r^2}} - \frac{a-z}{\sqrt{(a-z)^2 + r^2}} \right]$
 2-23. Treat dipole as two equal but oppositely charged point charges separated by a small distance.
 2-25. $Q_{11} = Q_{22} = -2ql^2; Q_{33} = 4ql^2$; other components zero
 2-27. (a) $\mathbf{p} = q\mathbf{l}_1 + q\mathbf{l}_2, p = 2ql \cos \theta/2$
 (b) $q = 5.26 \times 10^{-20} \text{ C} = 0.328 e$

Chapter 3

- 3-1. Between: $\varphi = \frac{r_b\varphi_b - r_a\varphi_a + (\varphi_a - \varphi_b)r_a r_b / r}{r_b - r_a}$
 For $r > r_b$: $\varphi = \varphi_b r_b / r$

3-9.
$$\varphi(r, \theta) = \frac{p}{4\pi\epsilon_0} \frac{P_1(\cos \theta)}{r^2} \left(1 - \frac{r^3}{a^3}\right), r \leq a$$

3-11.
$$\varphi = -(1 - a^3/r^3)E_0 r \cos \theta + Q/4\pi\epsilon_0 r$$

3-13.
$$\varphi = (rV_0/R) \sin \theta \text{ inside, } = (RV_0/r) \sin \theta \text{ outside}$$

 3-15. The mirror image of the charge distribution with ρ replaced by $-\rho$

3-17. (b)
$$F = -\frac{q^2}{16\pi\epsilon_0 d^2} \times \left[\frac{1}{u^2} + \sum_{n=1}^{\infty} \left(\frac{1}{(n+u)^2} - \frac{1}{(n-u)^2} \right) \right]$$

$$\cong -\frac{q^2}{16\pi\epsilon_0 d^2} \left[\frac{1}{u^2} - \frac{1}{(1-u)^2} + 1 - 2u \right],$$

 where $u = x/d$

3-21. (a) Image at $x_0(M^2 - 1)/(M^2 + 1)$

(b) $M = (x_0/a) + \sqrt{(x_0/a)^2 - 1}$

3-23. $3p^2/32\pi\epsilon_0 d^4$; attraction

3-25. Values at interior meshpoints (volts):

30		
18	34	
9	18	30

Chapter 4

4-1. $\rho_p = -2ax$; Q_p (on ends) = $A(aL^2 + b)$, $-bA$

4-3. Outside rod: $E_z = \frac{P}{2\epsilon_0} \times \left[\frac{\frac{1}{2}L - z}{\sqrt{(\frac{1}{2}L - z)^2 + R^2}} + \frac{\frac{1}{2}L + z}{\sqrt{(\frac{1}{2}L + z)^2 + R^2}} \right]$

4-5. $E = (1/\epsilon_0)P \cos \gamma$

4-7. $\frac{\tan \theta_1}{\tan \theta_2} = \frac{K_1}{K_2}$

4-9.
$$\varphi(\text{region 2}) = -E_0 \cos \theta \left\{ r - \frac{R_2^3}{r^2} + \frac{3(R_2^3 - R_1^3)}{r^2} \right.$$

$$\left. \times \frac{(K_2/K_1)}{[1 + 2(R_1/R_2)^3 + 2(K_2/K_1)(1 - R_1^3/R_2^3)]} \right\}$$
 electric field components by differentiation

4-11.
$$\varphi_1 = \frac{3}{K + 24\pi\epsilon_0} \frac{p \cos \theta}{r^2}, \quad r \geq a$$

$$\varphi_2 = \frac{p}{4\pi K \epsilon_0} \left[\frac{\cos \theta}{r^2} + \frac{K-1}{K+2a^3} r \cos \theta \right], \quad r \leq a$$

4-13. (a) $E_1 = \frac{K_2}{K_1} E_2$; (b) $P_1 = \frac{K_1 - 1}{K_1} K_2 \epsilon_0 E_2$;

(c) $\sigma_{P1} = \pm P_1$

4-15. $D = K\epsilon_0 \Delta\varphi/[Kd - (K-1)t]$

4-17. $E = Q/2\pi(\epsilon_1 + \epsilon_2)r^2$

4-19. (a) Inside: $\mathbf{E} = -\mathbf{P}/3\epsilon_0$

(b) Outside: $E_r = \frac{2R^3 P}{3\epsilon_0 r^3} \cos \theta$,

$$E_\theta = \frac{R^3 P}{3\epsilon_0 r^3} \sin \theta$$

Chapter 5

5-1. (a) $\alpha = 9.7 \times 10^{-41} \text{ C} \cdot \text{m}^2/\text{V}$

(b) $R_0 = 0.96 \times 10^{-10} \text{ m}$

5-3. (a) $2.6 \times 10^{-16} \text{ m}$

5-5. $2.94 \times 10^{-30} \text{ C} \cdot \text{m}$

Chapter 6

6-1. 39 cm

6-3.
$$U = -\frac{2q^2}{4\pi\epsilon_0} \left[\frac{1}{2l} - \frac{1}{2d+l} + \frac{1}{4d} + \frac{1}{4(d+l)} \right]$$

 (since $E = 0$ behind the plane)

6-5. $4\pi R^5 \rho_0^2 / 15\epsilon_0$

6-9. (a) $\Delta\varphi = 10.79 \text{ V}$, $U = 1.618 \mu\text{J}$

(b) $\Delta\varphi = 21.58 \text{ kV}$, $U = 3.24 \text{ mJ}$

6-11. $-(R/d)q$

6-13. $\epsilon_1 \epsilon_2 / (\epsilon_1 d_2 + \epsilon_2 d_1)$

6-15. 23.8 V

6-17. $C = 4\pi\epsilon / [(1/r_1) - (1/r_2)]$

6-19. (a) $Kl(\Delta\varphi)_0/[l + (K-1)x]$

(b) $(K-1)Q^2 d / 2\epsilon_0 w [l + (K-1)x]^2$

6-21. $[4mg/2\pi\epsilon_0(K-1)]^{1/2}$

6-25. $F_x = \frac{1}{2}(1 - 1/K)\epsilon_0 E^2 A$

6-27. (a) $Q_{\max} = \epsilon A E_{\max}$, with $A = \text{plate area}$

(b) Glass

Chapter 7

7-1. (b) $v = 0.739 \times 10^{-3} \text{ m/s}$

(c) $\tau = 2.5 \times 10^{-14} \text{ s}$

7-3. (a) $\varphi_{\text{int}} = \frac{\varphi_1 g_1 (d-a) + \varphi_2 g_2 a}{g_2 a + g_1 (d-a)}$

(b) $\sigma = \frac{(g_1 \epsilon_2 - g_2 \epsilon_1)(\varphi_1 - \varphi_2)}{g_2 a + g_1 (d-a)}$

7-5. (a) $\varphi_1 = -E_0 r \cos \theta - \frac{1}{2} E_0 a^3 \frac{\cos \theta}{r^2}$ (outside)

$$\varphi_2 = -\frac{3}{2} E_0 r \cos \theta$$
 (inside)

$$(b) \sigma = -\frac{3}{2}KE_0 \cos \theta$$

$$7-7. (a) I = 2\pi g \Delta\varphi / \ln(r_2/r_1)$$

$$7-11. I = \pi g s \Delta\varphi / \cosh^{-1}(b/2a)$$

$$7-13. (a) 675 \text{ W/m}^2, (b) 645 \text{ W/m}^2$$

$$7-15. 20 \Omega$$

$$7-17. 75.4 \text{ m}$$

$$7-19. (a) I = (\mathcal{V}_2 R_1 + \mathcal{V}_1 R_2) / (R_1 R + R_2 R + R_1 R_2)$$

$$(b) R_1 R_2 / (R_1 + R_2)$$

$$7-21. (a) 4R/5, (b) (11/20)R$$

$$7-23. (a) \frac{(R_4 R_5 - R_3 R_6) \mathcal{V}_1}{D + R_6 (R_3 + R_4)(R_5 + R_6)}$$

$$\text{where } D = R_3 R_4 R_5 + R_4 R_5 R_6 + R_5 R_6 R_3 + R_6 R_3 R_4$$

$$7-25. \text{One part in } 4 \times 10^6$$

Chapter 8

$$8-3. (a) 0.071 \text{ cm}, (b) 1.09 \times 10^{-7} \text{ s}$$

$$8-5. f = \frac{1}{2} \mu_0 \frac{N}{L} I^2 = 6.28 \text{ N/m}$$

$$8-7. B = \sqrt{3} \mu_0 I / \pi a$$

$$8-9. \mu_0 I N / 4a$$

$$8-11. \mu_0 N I$$

$$8-13. (a) \partial B_r / \partial z = \partial B_z / 2 \partial r$$

$$8-15. \nabla \times \nabla \times \mathbf{B} = \mu_0 \nabla \times \mathbf{J} = 0$$

$$8-17. B = \frac{1}{2} \mu_0 J s$$

$$8-19. (a) B = \mu_0 N I / 2\pi r, (b) b/a = 4/3$$

$$8-21. A_z = (\mu_0 I / 2\pi) \ln(r/b) \text{ between the conductors}$$

$$8-23. (b) \nabla \times (\mathbf{A} + \nabla\psi) = \nabla \times \mathbf{A}$$

$$(c) \nabla \cdot (\mathbf{A} + \nabla\psi) = f = \nabla^2 \psi, \psi = -\frac{1}{4\pi} \int \frac{f}{R} dv$$

$$8-25. \varphi^* \text{ is not single valued.}$$

$$8-27. (a) 2\sqrt{2} \mu_0 I / (\pi a), \perp xy\text{-plane}$$

$$(b) 0.631 \mu_0 I / (\pi a), \parallel xy\text{-plane}$$

$$0.647 \mu_0 I / (\pi a), \perp xy\text{-plane}$$

$$(c) 0.360 \mu_0 I / (\pi a), \perp xy\text{-plane}$$

$$(d) -0.631 \mu_0 I / (\pi a), \parallel xy\text{-plane}$$

$$-0.647 \mu_0 I / (\pi a), \perp xy\text{-plane}$$

$$8-29. (d) B_r = (\mu_0 I / 2a)$$

$$\times \left[\cos \theta - \frac{3r^2}{4a^2} (5 \cos^3 \theta - 3 \cos \theta) + \dots \right]$$

$$B_\theta = (\mu_0 I / 2a)$$

$$\times \left[-\sin \theta + \frac{3r^2}{4a^2} (5 \cos^2 \theta - 1) \sin \theta + \dots \right]$$

Chapter 9

$$9-1. (a) \mathbf{J}_M = \nabla \times \mathbf{M} = 0$$

$$\mathbf{j}_M = \mathbf{M} \times \mathbf{n}; j_M = M \text{ on cylindrical surfaces, } j_M = 0 \text{ on sides}$$

$$9-3. (b) (4/3)\pi R^3 \mathbf{M}_0$$

$$9-5. \sigma_M = M_0 x / [x^2 + (b^4/a^4)(y^2 + z^2)]^{1/2}; \rho_M = 0$$

$$9-7. (b) B_z =$$

$$\frac{1}{2} \mu_0 M \left[\frac{\frac{1}{2}L - z}{\sqrt{(\frac{1}{2}L - z)^2 + R^2}} + \frac{\frac{1}{2}L + z}{\sqrt{(\frac{1}{2}L + z)^2 + R^2}} \right]$$

$$9-9. (a) 0.14 \text{ A}, (b) 0.53 \text{ A}$$

$$9-11. \mathbf{B}_i = 2(\mu/\mu_0)\mathbf{B}_0/(1 + \mu/\mu_0)$$

$$9-13. J_M = 0, j_M = \frac{\chi I}{2\pi a} \text{ (inner surface),}$$

$$j_M = \frac{-\chi I}{2\pi b} \text{ (outer surface)}$$

$$I_M = \chi I \text{ (inner), } I_M = -\chi I \text{ (outer)}$$

$$9-15. (a) \theta_1 = 0, I \text{ parallel}$$

$$(b) \theta_1 = 90^\circ, I \text{ antiparallel}$$

$$9-17. 938$$

$$9-19. (a) \text{Sintered oxide } 0.4 \text{ T}$$

$$35 \text{ percent Co steel } 0.22 \text{ T}$$

$$(b) \text{Sintered oxide } 0.53 \text{ T}$$

$$35 \text{ percent Co steel } 0.96 \text{ T}$$

$$9-21. 0.64 \text{ T}$$

Chapter 10

$$10-3. 3.69 \times 10^{-4}$$

$$10-5. \gamma = 976$$

Chapter 11

$$11-5. (a) \mathcal{V} = \mathcal{V}_0 \cos \omega t; (b) \mathcal{V}_0 = 2\pi f N A B = 12.56 \text{ V}$$

$$11-7. (a) \text{From } b \text{ to } a; (b) \text{From } b \text{ to } a$$

$$(c) \text{From } a \text{ to } b$$

$$11-9. \frac{1}{2} B^2 a^2 r^2 \omega g t$$

$$11-11. (\mu_0 L / 2\pi) \ln(R_2/R_1)$$

$$11-13. \mu_0 \pi a^2 b^2 / 4r^3$$

$$11-15. M = (\mu_0 h / 2\pi) \ln(1 + d/r)$$

$$11-17. (\mu_0 / \pi) \ln(d/a)$$

$$11-21. (a) 73.4 \text{ Nm}, (b) 1.13 \text{ s}$$

$$11-25. \text{Let } \nabla \times \mathbf{B} = \alpha \mathbf{B}. \mathbf{J} = \mathbf{J}_0 e^{-(\alpha^2 r \mu_0)}$$

Chapter 12

$$12-3. U = (1/2s + 2)[L_1 I_1^2 + 2M_{12} I_1 I_2 + L_2 I_2^2]$$

$$12-5. (N^2 I^2 A \mu^2) / \mu_0 (l + d)^2$$

12-7. (a) $F = B_0^2 \chi_m A / 2\mu_0 (1 + \chi_m)$
 (b) $1.76 \times 10^{-4} \text{ N}$

12-11. (a) and (b) $\frac{1}{2} \mu_0 \frac{N}{l} I^2$

12-13. (a) $F = -\frac{1}{2} \mu_0 N^2 I^2 \left(\frac{b}{\sqrt{b^2 - a^2}} - 1 \right)$; (b) collapse

12-15. $I = I_0 \frac{L}{L + M}$

12-17. Commercial iron: 0.018 W/cm^3
 Tungsten steel: 0.395 W/cm^3

- 12-19. (a) Alnico 5, 45 joules
-
- (b) Alnico 12, 12 joules
-
- (c) Co steel, 12 joules
-
- (d) Sintered oxide, 8 joules

Chapter 13

- 13-1. (a)
- $I = 0.605 \text{ A}$
- ;
- $dI/dt = 1.59 \text{ A/s}$
-
- (b)
- $I = 1.295 \text{ A}$
- ;
- $dI/dt = 0.558 \text{ A/s}$
-
- (c)
- $I = 1.663 \text{ A}$
- ;
- $dI/dt = 0.0062 \text{ A/s}$

13-3. $Q = C \mathcal{V}_0 [1 - e^{-t/RC}]$

13-5. (a) $|Z| = \sqrt{\frac{R^2(1 - \omega^2 LC)^2 + (\omega L)^2}{1 + (\omega RC)^2}}$

- (b)
- $|Z| = R$
- for
- $\omega = 0$
- ;
-
- $|Z| \cong L/RC$
- for
- $\omega = 1/\sqrt{LC}$
- ;
-
- $|Z| \cong \omega L$
- for
- $\omega \rightarrow \infty$
- .

13-7. $Z =$

$$\frac{R\alpha - \omega C'R \frac{\omega L - 1}{\omega C} + i \left(\omega C'R^2 + \frac{\omega L - 1}{\omega C} \alpha \right)}{\alpha^2 + \omega^2 C'^2 R^2}$$

where $\alpha = 1 + \omega^2 C'L - C'/C$

- 13-9. (a)
- $3.2 \times 10^{-2} \text{ deg}$
-
- (b) zero to
- 1.8 MHz
-
- 13-13. (a)
- $f = 1.78 \times 10^3 \text{ Hz}$
-
- (b)
- $f = 1.78 \times 10^3 \text{ Hz}$
-
- (c)
- $f = 0.796 \times 10^3 \text{ Hz}$

- 13-15. (a)
- $L/C = 2R^2$
-
- (b)
- $L = \sqrt{2} R/\omega_c$
- ;
- $C = 1/\sqrt{2} R\omega_c$

13-17. $0.0713 - 0.0034i \text{ mA}$

13-19. $1/\sqrt{3LC}$

13-21. $G_T = \frac{3 + F^4 + 4iF}{(3 - F^2)^2 + 4F^2}$

13-23. $2W_1 - FW_2 - W_3 = 1$
 $FW_1 + 2W_2 - W_4 = 0$
 $-W_1 + \frac{3}{2}W_3 - 2FW_4 = 0$
 $-W_2 + 2FW_3 + \frac{3}{2}W_4 = F$
 where $V_2/\mathcal{V}_0 = W_1 + iW_2$
 $V_3/\mathcal{V}_0 = W_3 + iW_4$

Chapter 14

- 14-3. (a)
- $h = 0.62 \times 10^{-6} \text{ m}$
- ;
- $\omega_p = 9.7 \times 10^9 \text{ s}^{-1}$
-
- (b)
- $h = 0.84 \times 10^{-3} \text{ m}$
- ;
- $\omega_p = 1.8 \times 10^6 \text{ s}^{-1}$
-
- (c)
- $h = 3.1 \times 10^{-8} \text{ m}$
- ;
- $\omega_p = 1.8 \times 10^{11} \text{ s}^{-1}$
-
- 14-5.
- $u_d = 10^5 \text{ m/sec}$
-
- 14-7. (a)
- $v_p = 280 \text{ m/s}$
- ,
- $\delta = 0.11/f^2 \text{ m}$
- (
- f
- in Hz)
-
- (b)
- $v_p = 0.76 \text{ m/s}$
- ,
- $\delta = 0.03/f^2 \text{ m}$

Chapter 15

- 15-1. (a)
- $\mathbf{B} = B_0 \mathbf{i} - B_0(a/r)^2(\mathbf{a}_r \cos \theta + \mathbf{a}_\theta \sin \theta)$
-
- (outside)
-
- $\mathbf{B} = \mathbf{H} = 0$
- (inside)
-
- (b)
- $\mathbf{j}_s = -2B_0\mu_0^{-1} \sin \theta \mathbf{k}$

- 15-3.
- $\mathbf{B} = B_0 \mathbf{i} - b(a/r)^2(\mathbf{a}_r \cos \theta + \mathbf{a}_\theta \sin \theta)$
- (outside)
-
- $\mathbf{B} = c[\mathbf{a}_r(\lambda/r)I_1(r/\lambda) \cos \theta - \mathbf{a}_\theta I_1'(r/\lambda) \sin \theta]$
-
- (inside), with
- I_1
- the modified Bessel function of the first kind, and
- I_1'
- its first derivative

$$b = B_0 \frac{[I_1'(a/\lambda) - (\lambda/a)I_1(a/\lambda)]}{I_1'(a/\lambda) + (\lambda/a)I_1(a/\lambda)}$$

$$c = 2B_0/[I_1'(a/\lambda) + (\lambda/a)I_1(a/\lambda)]$$

- 15-5. (a)
- $B = 10^{-3} I$
- tesla where
- I
- is in A.
-
- (b) Pb:
- $I = 53 \text{ A}$
- ; Ta:
- $I = 9.7 \text{ A}$
- ; Nb:
- $I = 140 \text{ A}$

 15-7.

r/λ	B/B_0	B/B_0
	(1-D case)	(Fig. 15-3)
0	0.007	0
1	0.018	0.023
2	0.050	0.065
3	0.135	0.162
4	0.368	0.401
5	1.0	1.0

Chapter 16

- 16-1. (a)
- $Q = C(\Delta\varphi)e^{s/\epsilon}$
-
- (b)
- $-(g/\epsilon)C(\Delta\varphi)e^{s/\epsilon}$
- ; (c) Zero
-
- 16-3. (a)
- $H = \frac{1}{2}a \partial D/\partial t = I/2\pi a$
-
- (b)
- $S = \frac{1}{2}aE \partial D/\partial t$
- , inward
-
- (c)
- $\pi a^2 dE \partial D/\partial t$

$$16-7. \quad \mathbf{B} = -\mathbf{i}E_0\sqrt{\epsilon\mu}\sin\omega(\sqrt{\epsilon\mu}z-t) \\ + \mathbf{j}E_0\sqrt{\epsilon\mu}\cos\omega(\sqrt{\epsilon\mu}z-t)$$

$$\mathbf{S} = \mathbf{k}\sqrt{\frac{\epsilon}{\mu}}E_0^2$$

$$16-9. \quad \mathbf{A} = -\mathbf{i}(\lambda E_0/2\pi c)\cos[2\pi(z-ct)/\lambda]$$

$$16-11. \quad \nabla \cdot \mathbf{A} = -\epsilon\mu \frac{\partial \varphi}{\partial t} - \mu g \varphi$$

Chapter 17

$$17-5. \quad E = 700 \text{ V/m}; B = 2.33 \times 10^{-6} \text{ T, rms}$$

$$17-7. \quad \bar{S} = (1/c\mu_0)\bar{E}^2 \\ = (1/c\mu_0)^{1/2}(E_1^2 + E_2^2 + 2E_1E_2\cos\phi)$$

$$17-11. \quad n = 1.099\sqrt{K_r}, k = 0.455\sqrt{K_r}; \delta/\lambda = 0.384$$

$$17-13. \quad (b) \quad g = 0.92 \times 10^{-6} (\Omega\text{m})^{-1}$$

$$17-15. \quad l = 4.5 \text{ cm}$$

Chapter 18

$$18-1. \quad (a) \quad r_{12s} = \frac{1-n^2}{1+n^2}$$

$$(b) \quad \text{For } n = 1.5, R_s = 14.8 \text{ percent}$$

$$18-3. \quad R_p \cong 1 - 4K\sqrt{\frac{2\delta}{\sqrt{K}-1}}$$

$$18-5. \quad (a) \quad R_p \cong \frac{(n\cos\theta_1 - 1)^2 + k^2\cos^2\theta_1}{(n\cos\theta_1 + 1)^2 + k^2\cos^2\theta_1}$$

$$(b) \quad \cos\theta_1 = 1/\sqrt{n^2 + k^2}$$

$$(c) \quad \theta_1 = 80.5^\circ, R_p = 72 \text{ percent}$$

$$18-9. \quad K_i = 1.13; n \cong 3, k \cong 0.188; \alpha = 2.7^\circ$$

$$18-13. \quad \hat{r}_{23} = \pm 1, R = 1$$

$$18-15. \quad (a) \quad R = \frac{R_{12} + R_{23} - 2R_{12}R_{23}}{1 - R_{12}R_{23}}$$

$$(b) \quad \text{For } R_{12} = R_{23}, R = \frac{2R_{12}}{1 + R_{12}}$$

$$18-17. \quad B_x = B_1 \cos(\kappa y \cos\theta)e^{i(\kappa z \sin\theta - \omega t)} \\ E_y = -cB_1 \sin\theta \cos(\kappa y \cos\theta)e^{i(\kappa z \sin\theta - \omega t)} \\ E_z = icB_1 \cos\theta \sin(\kappa y \cos\theta)e^{i(\kappa z \sin\theta - \omega t)}$$

$$18-19. \quad \frac{\lambda}{2} < a < \frac{\lambda}{\sqrt{2}}$$

$$18-21. \quad (a) \quad R_s = \left[\frac{\cos\theta_1 - \sqrt{n_2^2 - \sin^2\theta_1}}{\cos\theta_1 + \sqrt{n_2^2 - \sin^2\theta_1}} \right]^2; \\ R_p = \left[\frac{n_2^2 \cos\theta_1 - \sqrt{n_2^2 - \sin^2\theta_1}}{n_2^2 \cos\theta_1 + \sqrt{n_2^2 - \sin^2\theta_1}} \right]^2$$

$$(c) \quad \text{Crown glass: } \theta_B = 56.7^\circ; \\ \text{flint glass: } \theta_B = 58.8^\circ; \\ \text{diamond: } \theta_B = 67.5^\circ$$

Chapter 19

$$19-1. \quad (a) \quad n_v - 1 = 1.63 \times 10^{-4}; (b) \quad 1.36 \times 10^{-3}$$

$$19-3. \quad \lambda_0 = 1560 \text{ \AA}$$

$$19-7. \quad (a) \quad k \cong K_i/2n_z, \alpha = 2k\omega_0/c$$

$$19-9. \quad n \cong k \cong 0.071$$

$$19-13. \quad \omega = \sqrt{K_0/\tau}; K_r \cong 1, K_i \cong \sqrt{K_0}$$

$$19-15. \quad P = \chi_0 E_0 \sin\omega_0 t$$

Chapter 20

$$20-3. \quad B_\theta = \frac{\mu_0 I_0}{4\pi} A \frac{\omega^2}{c^2 r} \sin\theta \cos\omega\left(t - \frac{r}{c}\right)$$

$$E_\phi = -\frac{I_0 A}{4\pi\epsilon_0 c^3 r} \omega^2 \sin\theta \cos\omega\left(t - \frac{r}{c}\right)$$

where A is the area of the circular loop

$$P = \frac{\mu_0}{6\pi} I_0^2 A^2 \frac{\omega^4}{c^3} \cos^2\omega\left(t - \frac{r}{c}\right)$$

$$20-5. \quad (a) \quad 5.23 \text{ A}, (b) \quad 2.45 \times 10^{-2} \text{ V/m}$$

$$20-7. \quad \frac{3}{2}$$

$$20-9. \quad (a) \quad 8\pi/3(v/c)^3; (b) \quad 4.07 \times 10^{-7}$$

$$20-11. \quad R_r = \frac{2\pi}{3} \sqrt{\frac{\mu_0}{\epsilon_0}} (2\pi)^2 \left(\frac{\pi a^2}{\lambda^2}\right)^2$$

$$20-13. \quad 52.1 \text{ m}^{-1}$$

Chapter 22

$$22-3. \quad 4.2 \times 10^{17} \text{ s}$$

$$22-5. \quad \frac{\mathbf{u}}{1 + u^2/c^2} = \frac{\mathbf{E} \times \mathbf{B}}{(E^2/c^2 + B^2)}$$

Index

- Abampere, 600
Absorptance, 456, 510
Absorption coefficient, 439, 544
Admittance, 327, 329
Air gap, 244–245, 246, 248, 249–250
Air–glass interface, 444, 452, 461, 466–469
Alfvén waves, 359
Alternating current. *See* Circuit, ac; Current, ac
Ampere (A), 162, 197, 600
Ampere’s circuital law, 204–208, 228, 236, 569
 and current in superconductors, 374; differential form of, 205; and displacement current, 386–389; and series magnetic circuit, 246–249; and superconductivity, 381
Angle of incidence, 446–447
 normal, 441–445
Angle of reflection, 447
Angle of refraction, real, 458, 459
Angular frequency, 305n
Anisotropic material, 104, 130, 229n, 391n
Anomalous dispersion, 506
Antenna, 535
 directivity of, 543; half-wave, radiation from, 529–531
Antiferromagnet, 268
Antisymmetric tensor, 571
Applied voltage, 175, 306, 307–308. *See also* Electromotive force
Atmospheric electricity, 164
Atomic current, 218, 219–220, 226. *See also* Magnetization current; Magnetization current density
Atomic polarizability, 132
Atomic radius, 132
Attenuation constant, 426–427, 461–462
 in conducting medium, 428
Average power, 317

Back cab rule, 5
BASIC, 82, 583–585
 programs in, 83, 87, 90, 331–333, 586, 588–589
Battery, 176
BCS theory, 369
Bessel functions, 382–383
 spherical, 435–436
Betatron, 285
B-value of plasma, 356–357
Biot-Savart law, 197–199, 218, 600, 601
Bohr magneton, 262, 263, 269, 270
Boltzmann factor, 133–134, 343, 515–516
Boltzmann transport equation, 342
Boundary conditions
 and coordinate system, 77–78; on field vectors, 112–115, 235–237; and Laplace’s equation, 67, 69; and Maxwell’s equations, 398–401; table, 602
Boundary-value problems, 63–77, 114–115, 398–401, 402
 involving dielectrics, 115–118; involving magnetic materials, 237–242; and Maxwell’s equations, 398–401, 441–483; for superconductors, 371–373
Bound charge, 97, 498. *See also* Polarization charge
 resonance absorption by, 499–506
Branch point, 178
Brewster’s angle, 451–454, 456–457
Brewster’s law, 445, 452
Bridged-T network, 329–331, 332

Capacitance, 152–153, 602
 coefficient of, 150–151
Capacitive reactance, 313
Capacitor, 53, 151–153, 173, 386–389
 and damping, 338; parallel and series connections, 153; in series circuits, 307–308, 310–320, 338
Cartesian coordinate system, 2
 del operator in, 17–18, 603
Cascading, 164
Cauchy principal value, 492, 614
Cauchy relation, 505
Cauchy-Riemann differential equations, 611
Cauchy’s integral theorem, 488n, 611–612
Causality, 487–490
Cavity field, 104–106, 125, 128, 257–258
Cavity resonator, 477–479
Charge, 26–27. *See also* Coulomb’s law; Image charge; Line charge; Point charge; Source charge; Test Charge
 bound, 97, 108n, 498, 499–506; on capacitor, 152; conservation of, 27, 153, 388, 390, 400, 525; external, 108; forces between, 27; free, 108, 273–277; total, in dielectric, 106–109; units of, 28–29, 602
Charge carrier, 36–37, 162, 163
Charge density, 30, 30–32, 39. *See also* Polarization charge density; Surface charge density
 on long, thin cylindrical conductor, 84–88; relation to current density, 166; relation to point charge, 48
Charged medium, 164
Circuit, 175–181. *See also* Ampere’s circuital law; Magnetic circuit; Parallel circuit; Series circuit
 inductively coupled, 279–283, 290–298; isolated, induced emf in, 277; magnetic field of, 209–211; magnetic forces on, 294–298; quality factor, 319–320; sign convention, 307–308; time constant of, 309–310
Circuit, ac, 315–320
 inductance in, 283; mesh analysis of, 325–327; mutual inductance in, 320–324; power factor, 316–317
Circuit, dc, 175
Circuit diagram, 179–180
Circular polarization, 420, 422
Clausius-Mossotti equation, 131, 132, 497
Coaxial cable, 124, 206
Coefficient of capacitance, 150–151
Coefficient of induction, 150–151
Coefficient of potential, 75–77, 148–150
Coercive force, 233
Coercivity, 232, 233
Coherent scattering, 540n
Collision time, 175n, 182–183
 and damping constant, 507; mean, for free electrons, 500
Complex conductivity, 499, 508
Complex constant vector amplitude of plane wave, 413, 418
Complex current, 313–314
Complex dielectric constant, 425–430, 459–460
 in bound electron system, 496–506; and Drude-Lorentz theory, 496–499, 501–505; in free electron system, 507
Complex frequency response function, 491–494
Complex Fresnel coefficients, 454–462
Complex impedance, 313–316, 320–324
Complex index of refraction, 425–428, 459–460
Complex orthogonal transformation, 566–567
Complex propagation vector, 417
Complex quantity, 313, 421–422, 454, 472, 566, 567n. *See also* Complex variable
 and linear media, 487
Complex variable, functions of, 610–617
Complex vector, 418, 420
Complex voltage, 312–314
Compton effect, 539
Computer, 579–590. *See also* Software
 and network equations, 329–336; and numerical integration, 585–587; solving linear equations on, 332, 585, 587–590; useful tips, 590–592
Computer language, 580
Computer program, 579. *See also specific programs*
 for electrostatic problems, 88–91
Conducting cylinder, 75, 475
Conducting medium, 162, 165, 170, 181
 complex dielectric constant of, 498; plane monochromatic waves in, 423–430; relaxation time of, 174–175; wave propagation in, 397–398, 412n
Conducting plane, 69–71, 75
Conducting sphere, 72–73, 94, 125
 in dielectric, 119–120; with infinite conductivity

- 370–371, 374–376; in a uniform electric field, 63–65
 onduction, 162–164
 microscopic theory of, 181–185
 onduction current, 164
 onductivity, 167–170, 498–499. *See also* Complex conductivity
 of dielectrics, 173; infinite, 352–353, 361, 370–373, 400–401, 402; of plasma, 352–353, 361; real, frequency dependence of, 517; relation to optical constants, 427–428; table, 168; units of, 602
 onductor, 31, 36–37, 162. *See also* Superconductor
 in boundary-value problems, 114–115; current-carrying, forces on, 193–197; electrostatic energy of, 145–151; plasma as, 341; self-energy, 146; system of, 75–77, 148–151
 onical surfaces, 62n
 onservaion of charge, 27, 153, 388, 390, 400, 525
 onservaion of energy, 391n, 393–394, 444
 onstitutive equations, 109, 167, 229, 415
 of medium, 390; of superconductivity, 384
 ontinuity equation, 165–167, 390
 ontour line, 7
 ontravariant vector, 598n
 onvection current, 164–165
 onvergence, 590
 onvolution theorem, 486, 487, 609
 oordinate systems. *See also* Cartesian coordinate system; Curvilinear coordinate system; Cylindrical coordinate system; Four-dimensional space; Orthonormal coordinate system; Rectangular coordinate system; Spherical coordinate system; Three-dimensional coordinate system
 and boundary conditions, 77–78; in motion, 559–566
 oordinate transformation, 594–597
 ouble (C), 29, 599, 600
 ouble gauge, 410
 ouble's law, 27–31, 190, 271, 389–390
 and electrostatic energy, 146; in mks system, 29, 602; and self-energy of point charge, 146; statement of, 27, 599
 ounter voltage
 and Kirchhoff's laws, 307; sign convention, 308n
 ouple circuit, and magnetic energy, 290–292
 ouple coefficient, 280, 282
 ovariant formulation, 569–570
 ovariant vector, 598n
 ritical angle, 451–454, 453
 ritical damping, 338
 ritical magnetic field, 366
 ritical temperature, of superconductor, 366, 367–368
 ross product. *See* Vector product
 urie temperature, 136, 265–266
 irl, 14–17
 econd-order operations with, 19–20
 urrent, 37, 162–189, 175, 218, 558. *See also* Atomic current; Complex current; Conduction current; Convection current; Displacement current; Eddy current; Image current; Magnetization current; Surface current; Transport current
 efinied, 162; effective value, 317; magnetic field of, 190–217; in metal, 162–163; periodic, 306–307; sense (direction) of, 164; slowly varying, 284, 290n, 305–340; in superconductors, 374; units of, 602
 urrent, ac, 315–320
 mesh analysis of, 325–327
 urrent density, 165–167. *See also* Displacement current density; Dissipative current density; Magnetization current density; Surface current density; Transport current density
 four-vector, 570; units of, 166, 601
 urrent gain, in four-terminal network, 329
 urrent generator, 276, 326–327
 urvilinear coordinate system, 18, 93
 utoff wavelength, 471, 477
 ylindrical coordinate system, 18n
 Laplace's equation in, 65–66; Maxwell's equations in, 603
 ylindrical harmonics, 65–66
 d'Alembertian operator, 570
 Damping. *See also* Radiation damping
 critical, 338; of wave equation, 395
 Damping coefficient, 491, 495, 500
 in free electron system, 507; and rate of energy loss, 537–541; and system of bound charges, 499–506
 Damping force, 537–538
 de Broglie wavelength, 269
 Debye equations, for dielectric constants, 516–517
 Debye length, and plasma, 344
 Debye relaxation time, 516
 Debye shielding distance, and plasma, 344
 Debye temperature, 184
 Decay rate, 500
 Decay time, mean, for oscillator, 500
 Decibel (dB), 439
 Deformation polarizability, 135
 del operator, 17–23, 415
 in cylindrical coordinates, 18n, 603; linear operator, 19, 22; in rectangular coordinates, 17–18, 603; in spherical coordinates, 18n, 603–604; transformation law, 598
 Delta function. *See* Dirac delta function; Kronecker delta
 Demagnetization factor, of sphere, 242
 Demagnetizing effect, 250–251
 Depolarizing field, 126, 137
 Diamagnetic material, 229, 259
 magnetic susceptibilities, 230
 Diamagnetic susceptibility, 230, 260–261
 Diamagnetism
 and Lenz's law, 260; origin of, 259–261; perfect, of superconductors, 370–371
 Dielectric coefficient. *See* Dielectric constant
 Dielectric constant, 109–110. *See also* Complex dielectric constant
 of conductor, 173n; Debye equations for, 516–517; and molecular polarizability, 131–132; relation to optical constants, 427–429; table, 111
 Dielectric loss, 513–514
 Dielectric medium, 36–37, 97–126, 127–140, 498. *See also* Insulator
 boundary conditions, 115–120, 400; conductivity of, 173; electrostatic field in, 103–106; and Gauss's law, 111, 173; and method of images, 120–122; molecular field in, 127–131; permittivity of, 109; point charge in, 110–112, 118–120; polarization, 37, 97–103; wave propagation in, 396–397
 Dielectric relaxation, 513–517
 Dielectric sphere, in uniform electric field, 116–118
 Dielectric strength, 110–111
 table, 111
 Dielectric tensor, 391n
 Dip angle, 217
 Dipole. *See* Electric dipole; Magnetic dipole
 Dipole field, static, 536
 Dipole moment. *See* Electric dipole moment; Magnetic dipole moment
 Dirac delta function, 48–50, 223, 486, 605–606
 and Ampere's law, 204–205; and approximation of dielectric constant, 501, 502, 509; and divergence theorem, 50; Fourier transform of, 608–609; and Gauss's law, 49
 Direct current. *See* Circuit, dc
 Directional derivative, 5–8
 Directivity of antenna, 543
 Disk storage (computer), 581
 Dispersion, 417
 anomalous, 506; in dispersive media, 484–524; normal, 502; and radiation pulse, 518–519
 Dispersion relations, 442, 447, 448, 491–494, 493, 610–617. *See also* Kramers-Kronig dispersion relations
 of conducting medium, 426; in linear medium, 424; longitudinal, 440; proof of, 613–616
 Dispersive medium
 dispersion and oscillating fields in, 484–524; and oscillating fields, 517–521
 Displacement, electric, 106–109
 flux in regions with no external charge, 115; at interface, 112–114, 117; relation to electric field, 484–490; units of, 107, 602
 Displacement current
 and Ampere's law, 386–389; and charge conservation, 390; and Maxwell's equations, 386–411; and wave propagation, 398
 Displacement current density, 377–378
 Dissipation factor, 338
 Dissipative current density, 377–378
 div. *See* Divergence
 Divergence, 12–14
 curl of, 19; four-dimensional, 570; gradient of, 20; notation for, 12, 17; second-order operations with, 19
 Divergence operator, 12
 Divergence theorem, 14, 59–60, 166
 applications, 81n, 101, 392, 398, 399; and Dirac delta function, 50; extensions of, 21–22; and Gauss's law, 40; and magnetic flux, 228
 Domain wall, 266, 267
 Dot product. *See* Scalar product
 Drift motion, 164, 165
 in plasma, 346–349
 Drift velocity, 165
 of electron, 183; steady-state solution for, 181–182
 Driving point impedance, 328–329
 Drude free-electron theory, 507–513
 Drude-Lorentz theory, 484, 494–499, 507–513
 Eddy current, 286
 Effective current, 317
 Effective voltage, 317
 Einstein's postulates of special relativity, 561–566
 Electric charge. *See* Charge
 Electric current. *See* Current
 Electric dipole, 43–45, 534. *See also* Electric dipole moment

- in external field, 45, 54; induced, 131–132; oscillating, 525–529; potential of, 45, 63, 526–527
- Electric dipole moment, 44, 209–210
of charge distribution, 46–48; of dielectric, *see* Polarization; of molecule, 130; and radiated power from moving charges, 532–535; units of, 601
- Electric displacement. *See* Displacement, electric
- Electric drift velocity, 346–347
- Electric field, 31–33, 34–36
of conductors, 37, 42–43; defined, 31, 105; in dielectric, 103–106; of earth, 53; of electric dipole, 43–45; and Gauss's law, 37–40; macroscopic, 103, 127, 129; motional, 352; multipole expansion of, 46–48; relation to electric displacement, 484–490; units of, 36, 602
- Electric flux, 40, 212
- Electric polarization. *See* Polarization
- Electric potential. *See* Electrostatic potential; Potential difference
- Electric susceptibility, 109–110, 494, 496, 602
- Electrodynamics, 545–557
- Electrolyte, 37
conduction in, 163–164
- Electrolytic tank, 171–172
- Electromagnetic energy, 390–394
- Electromagnetic equations
covariant form, 569–573
- Electromagnetic field tensor, 571
- Electromagnetic induction. *See* Faraday's Law
- Electromagnetic pulse, propagation of, 518–519
- Electromagnetic radiation. *See* Radiation
- Electromagnetic wave. *See* Wave, electromagnetic
- Electromotive force, 176n, 177. *See also* Open-circuit voltage; Work
defined, 271–272; direction, 273; induced, 271, 277, 308; motional, 275, 276; units of, 278
- Electron. *See also* Atomic current
charge of, 30, 183; collision time, 175n, 182–183, 500; and conduction, 36–37, 162, 163; mobility and drift velocity, 183; orbital motions of, 259–263; plasma oscillation, 358–363, 430; power radiated by accelerated electron, 539
- Electron pairing, in superconductors, 369
- Electron resonance frequency
of bound valence electron, 499–500; of free electron, 495; in nonconductors, 499–506
- Electrophorus, 159
- Electrostatic energy, 141–161
of charge distribution, 143–146
- Electrostatic energy density, 146
- Electrostatic equilibrium, approach to, 174–175
- Electrostatic field, 37
in dielectrics, 97–126; of earth, 53; total macroscopic, in a dielectric, 103
- Electrostatic force
conservative, potential energy of, 35–36; and steady current, 175–176
- Electrostatic image, 69–73
and Laplace's equation, 69–72
- Electrostatic potential, 33–36. *See also* Potential difference
calculation of, with Gauss's law, 41–42; distribution produced by point dipole, 44–45; and divergence theorem, 81n; due to arbitrary charge distribution, 46–47; relation to potential energy, 35–36
- Electrostatic problems, 56–96
numerical solutions of, 77–91
- Electrostatics, 26–55
- Elementary particles
charge on, 26–27; and Coulomb's law, 28
- Elliptical polarization, 420, 444
- Emf. *See* Electromotive force
- emu, 600
- Energy, units of, 36, 602
- Energy conservation, 391n, 393–394, 444
- Energy density, 148, 393, 421–423
of an electrostatic field, 136–148; and energy propagation in wave, 471–473; in magnetic field, 292–294; units of, 601
- Energy flow, 393
- Energy flux, 421–423, 471–473
- Energy methods, 156, 296
- Equation of continuity, 165–167, 170, 171, 178, 390, 570
and equilibrium, 174
- Equilibrium, 174–175
- Equilibrium theory in plasma, 343–344
- Equipotential surface, 7, 37, 63
and electrostatic images, 69–72; and Gauss's law, 42
- esu, 600
- Ether, 559, 560–561
- Euler equation, 351, 360
- Evanescent waves, 461
- External charge, 108
- Faraday's law of electromagnetic induction, 260, 272, 273–277
in diamagnetism, 260; differential form of, 273, 389
- Farad (F), 152, 602
- Ferrimagnet, 268
- Ferrie, 261, 268, 485
- Ferroelectricity, and permanent polarization, 136–138
- Ferroelectric material, 136–137
- Ferromagnetic domain, 265–267, 268
- Ferromagnetics, 231–234, 259
circuit problems, 242–248; magnetic energy changes in, 298–301
- Ferromagnetism, theory of, 263–265, 268
- Fiber optics, 454
- Field point, 32
- Field tensor, four-dimensional, 571
- Field vector, boundary conditions on, 112–115
- Finite difference method, 77–79
computer program for, 79, 82–84, 88–89; and Laplace's equation, 90
- Finite element method, 77, 79–82, 84, 89
- Flux. *See* Electric flux; Energy flux; Magnetic flux
- Flux exclusion, 367
perfect, 371–374; by surface transport currents, 374, 375
- Force, 175–176, 190–192. *See also* Coercive force; Damping force; Electromotive force; Electrostatic force; Lorentz force; Magnetic force; Mechanical force
on charge distribution, 141–142; between charges, 27; on current-carrying conductors, 193–197; in definition of ampere, 599–600; lines of, 69; on object in charge system, 154–156; on point charge in dielectric, 118–120; restoring, *see* Force constant; superposition principle, 29; units of, 28, 602
- Force constant, 491, 495
- Force-free field, 288
- FORTTRAN, 89, 580–581
- Four-dimensional space, 566–567
- Fourier integrals. *See* Fourier transforms
- Fourier integral theorem, 608
- Fourier series expansion, 67
- Fourier transforms, 89–91, 607–609
and plane waves, 417
- Four vector, 567
- Free charge, 97, 498, 499
and external charge, 108n; and ferroelectricity, 137–138; in wire, and Lorentz force, 273–277
- Frequency, 305–306, 396. *See also* Angular frequency; Damping frequency; Larmor frequency; Natural frequency; Plasma frequency; Resonance frequency
of electromagnetic radiation, 305–306, 413; half-power, 318
- Frequency response function. *See also* Complex frequency response function
and dispersion relations, 491–494; in linear media, 490
- Fresnel coefficient, 443, 444
for oblique incidence, 447, 449–451; and wave interference, 462–469
- Frustrated total reflection, 467–468
- f-sum rule, 497
- Fundamental theorem of the theory of functions, 611
- Fusion reactors, 355–358
- Galilean transformation, and Newton's laws, 559, 564
- Gamma rays, 412, 413
- Gas, ionized, *see* Plasma
- Gas discharge, and current, 164
- Gauge, 407, 410
- Gauge transformation, 407
- Gauss (G), 203
- Gaussian surface, 37–43, 42
in dielectric, 106–109
- Gaussian units, 107, 192n, 202, 416, 599, 600–602
conversion factor, 230; table, 602
- Gauss-Jordan elimination method, 587–589
- Gauss's law, 37–40, 92, 389, 569
in dielectric, 106–109, 111, 173; differential form, 106; and Dirac delta function, 49; and divergence theorem, 40; and Poisson's equation, 56; and potential, 41–42
- Generating function, 94
- Generator, 176, 276, 285, 326–327
- Geometric imperfections, 183
- Geometry of space-time, 566–567
- Giorgi system of units, 29
- grad. *See* Gradient
- Gradient, 5–9, 7
second-order operations with, 19–20
- Green's theorem, 21–22
- Grounded sphere, 73
- Group velocity, 518–519
- Guiding center, of particle in plasma, 345
- Gyromagnetic ratio, 270
- Hagen-Rubens relation, 456, 460, 509–510
- Half-power frequency, 318
- Hall resistivity, 363
- Harmonic oscillator, 494–499, 500
- Harmonic oscillator equation, 311, 396
- Heisenberg theory of ferromagnetism, 268

- Helmholtz coil, 201–203
 Helmholtz equation, 424
 and monochromatic waves, 396; scalar and vector, 379, 431–433
 Henry (H), 278, 279, 602
 Hertz vector, 411
 Homogeneous wave, 426n
 Homopolar generator, 285
 Hydrodynamic formulation in a plasma, 342, 350, 359
 Hysteresis, 229–235
 Hysteresis loop, 138, 233–235, 300
 Hysteresis loss, 298–301
 Ideal current generator, 326
 Ideal transformer, 323–324
 Ideal voltage generator, 326
 Image charge, 69–73, 121–122
 Image current, 254
 Imaginary unit, 68. *See also* Complex quantity;
 Complex variable
 Impedance, 313–314, 490. *See also* Complex
 impedance
 Driving point, 328–329; of free space, 438; load,
 322–326; parallel connection, 314–316; polar
 form, 314, 315; series connection, 314–316;
 transfer, 329
 Incidence. *See* Angle of incidence
 Incoherent X-ray scattering, 540n
 Index of refraction, 397, 404, 416, 425. *See also*
 Complex index of refraction
 In conducting medium, 428; for water, 444
 Induced dipole, model of, 131–132
 Inductance
 In ac circuits, 283, 320–324; incremental, 277;
 mutual, 279–282, 320–324; in parallel, 281; in
 series, 281; units of, 278, 602
 Induction. *See* Magnetic induction
 Induction drag on sphere, 364
 Inductively coupled circuits, 279–283, 290–298
 Inductive reactance, 313
 Inductor
 In parallel connection, 282–283; in series circuits,
 307–308; series connections, 281–283
 Infrared radiation, 413
 Inner product. *See* Scalar product
 Instantaneous power, 317
 Insulator. *See also* Dielectric medium
 Resistivity of, 168; resistivity table, 168;
 Resonance absorption by, 499–506
 Integral theorems. *See* Fourier integral theorem;
 Vector integral theorems
 Integration, numerical, 585–587
 Intensity. *See also* Magnetic intensity
 of wave, 443–444
 Interface between media, 112–115, 170–171
 Interference, 462–469
 Internal reflection, total, 454, 460, 467
 Internal resistance, 177
 Intrinsic magnetic moment, 261
 Inverse square law, 28
 I, 162, 163
 Ion confinement, 341
 Ionic absorption, 506
 Ionic vibrational frequency, 515
 Ionized gas, 341
 Ionosphere, reflectance of, 508
 Ion drop, 307
 Isolated magnetic pole, 198
 Isotope effect, and superconductivity, 369
 Isotropic medium, 109, 229
 Iteration
 computer programs for, 79, 82–84, 88–89; in
 finite difference method, 78–79; in magnetic
 circuit problems, 247
 Joule heating, 170, 465
 Joule heating rate, 392
 Joule heat loss, in magnetic circuit, 289
 Joule (J), 36
 Kinetic theory, and plasmas, 342, 343
 Kirchhoff's current law, and nodal analysis, 327–
 328
 Kirchhoff's laws, 175–181, 321–328
 and mutual inductances in circuits, 321–323; and
 slowly varying currents, 307–308
 Kirchhoff's loop law. *See* Kirchhoff's voltage law
 Kirchhoff's voltage law, 307, 308
 and mesh analysis, 324–327
 Kramers-Kronig dispersion relations, 494
 for low-frequency approximation, 512;
 modification of, 513
 Kronecker delta, 595
 Langevin-Debye formula, 134, 262
 and polar molecules, 132–136
 Laplace's equation, 58–60
 boundary conditions, 67, 69; and conducting
 sphere, 63–91; and dielectrics, 116; and finite
 difference method, 90; and magnetic field
 problems, 238; in one independent variable, 60;
 and plane waves, 423; solutions to, 58–60 (in
 cylindrical coordinates, 65–66; method of
 images, 69–71, 120–122; numerical, 77–91; in
 one-dimension, 77–82; in orthogonal,
 curvilinear coordinates, 93; in rectangular
 coordinates, 66–67; in spherical coordinates,
 60–63, 90, 430–431; in two dimensions, 67–68,
 82–84, 611); and steady-state conduction, 170;
 and superconductors, 372; in two variables, 60–
 63
 Laplacian operator, 57, 375–376, 431, 604
 and Dirac delta function, 50; notation for, 19; and
 scalar field, 19; and vector field, 20
 Larmor frequency, 259
 Larmor orbit, 345
 Larmor radius, 345
 Law of reflection, 445
 Lawson break-even condition, 356, 357
 Leakage flux, 248
 Legendre polynomials, 62, 94, 434
 Legendre's equation, 61–62, 93
 Lenz's law, 260, 273
 Level surface, 7
 Liénard-Wiechert potentials, 545–548
 and field of accelerated charge, 551–554; and
 field of uniformly moving charge, 548–551
 Light pipe, 454
 Linear dielectric medium, 110, 143, 148
 Linear equations, computer solutions, 332, 582,
 585, 587–590
 Linearity, 484–490
 Linear medium, 181, 229, 484–485
 isotropic, 167, 169; magnetic, 238, 290; polar
 molecules as, 132–136; wave propagation in,
 394–395, 412–417
 Linear polarization, 419, 422
 Linear transformation, 594
 Line charge
 and cylindrical conductors, 75; electric field of,
 41–42; and equipotential surface, 74–75
 Line image, 74–75
 Line integral, 10
 Lines of displacement, 115
 Lines of force, 33, 34
 and electrostatic images, 69
 Lines of magnetic induction, 236–237, 240, 242
 Liquid conducting medium, 171
 Load impedance, 322–326
 Local field, 257, 495
 Logarithmic frequency scale, 330–331
 London equations, 374–384
 London theory, of superconductivity, 369–370
 Longitudinal electromagnetic wave, 430, 440, 509
 Loop, 178
 Lorentz condition, 533
 and scalar potential of dipole, 526–527; and wave
 equation, 403, 407
 Lorentz contraction, 564, 565
 Lorentz force, 191, 272
 and conductor, current carrying, 193; on free
 charges in wire, 273–277; and plasma, 345, 348;
 units of, 601
 Lorentz force equation, 390
 Lorentz gauge, 407
 and wave equation, 403
 Lorentzian line shape, 503–504
 Lorentz local field, 495, 497
 Lorentz transformation, 561–566, 569
 covariant form, 569–570; for electromagnetic
 field, 573–574; and geometry of space-time, 566–
 567; and orthogonal transformation, 567–569;
 and particle velocity in plasma, 347; and
 uniformly moving charge, 574–576
 Loss angle, 517
 Lyddane-Sachs-Teller relation, 523
 Macroscopic electric field, 103, 127, 129
 Macroscopic magnetic field, 258
 Magnet
 poles of, 198, 225, 390; scalar potential, 210
 Magnetic circuit, 245–248
 concept of, applications, 247; containing
 permanent magnets, 248–251
 Magnetic confinement system, for controlled
 thermonuclear fusion, 355–358
 Magnetic cooling, 301
 Magnetic dipole, 210
 radiating, 436, 543
 Magnetic dipole field, 210
 Magnetic dipole moment, 196, 209–210. *See also*
 Magnetization
 in ferromagnetic material, 263; intrinsic, 261; and
 orbital motion, 261; in paramagnetic material,
 262; units of, 601
 Magnetic energy, 289–304
 of coupled circuits, 290–292
 Magnetic energy density, 292–294
 Magnetic field. *See also* Magnetic intensity
 of distant circuit, 208–210; inside magnetic
 material, 237–242; inside matter. *see* Molecular
 field; produced by magnetized material, 222–225;
 relation to magnetic intensity, 231–235; sources
 of, 218, 226–227; of steady currents, 190–217;
 uniform, 345–347, 359–362, 371–378
 Magnetic field equations, 227–228

- Magnetic field lines, 236
 Magnetic field strength. *See* Magnetic intensity
 Magnetic flux, 212, 228, 246–247
 leakage, 248; and superconductors, 370; units of, 602
 Magnetic force, 191, 276, 294
 on rigid circuits, 294–298
 Magnetic induction, 190–193. *See also* Faraday's law of electromagnetic induction
 inside toroidal coil, 278; lines of, 236–237, 240, 242; and moving point charge, 550–551, 554–555; units of, 602
 Magnetic intensity, 226–227
 in absence of transport current, 237–242; in boundary-value problems, 235–238, 241–242; four-dimensional, 571; at molecule in cavity, 258; relation to magnetic field, 231–235; relation to magnetization, 229–230; transformation law for, 573–574; and transport current density, 228; units of, 602
 Magnetic medium
 and boundary-value problems, 237–242; and current circuits, 242–245; linear, 238, 290; reversible, 290–298
 Magnetic mirror, 349–350
 and thermonuclear fusion, 355
 Magnetic mirror reactor, 357
 Magnetic moment, 196, 209–210, 229–230
 of gyrating particle, 346; intrinsic, 261
 Magnetic pole, single, 198, 390
 Magnetic pole density, 225–226
 Magnetic pressure, 351–352, 354–355
 Magnetic properties
 of ferromagnetic materials, 232; of matter, 218–255; of matter, table, 230, 232
 Magnetic scalar potential, 210–212, 225–226, 238
 Magnetic susceptibility, 229–230, 602
 table, 230
 Magnetic torque, 294, 295–298
 Magnetic vector potential, 207–208, 222, 402
 Magnetism
 macroscopic theory of, 218–255; microscopic theory of, 256–270
 Magnetite, 190, 268
 Magnetization, 218–221. *See also* Magnetic dipole moment
 and atomic current, 226; relation to magnetic intensity, 229–230; spontaneous, 264
 Magnetization current, 220
 of superconductors, 370–374
 Magnetization current density, 219, 221, 223, 228
 Magnetization curve, 231
 Magnetomotive force, 246
 Mass susceptibility, 229, 230
 table, 230
 Matrix formulation, 595, 596–597
 and network equations, 332, 335
 Matthiessen's rule, 183
 Maxwell-Boltzmann distribution, 134, 342
 Maxwell's equations, 51, 386–411
 and boundary conditions, 398–401; and boundary-value problems, 441–483; and complex quantities, 421–423; covariant form, 569–570; and displacement current, 386–411; and electromagnetic waves, 412–440; empirical basis of, 389–390; in Gaussian units, 601; and Lorentz transformations, 561; and monochromatic waves, 395–398; for plane waves, 415; and radiation problems, 441; and spherical waves, 433; and superconductivity, 366, 369–370, 378–381; and wave propagation, 425, 474–477, 559–560
 Mean free path, 182–183, 342
 for incident photons, 540
 Mechanical force, 175–176, 276
 Meissner effect, 367, 370, 377
 and London equations, 374
 Mesh analysis, 324–328
 Metal, 162–163, 182–183
 resistivity, 168, 183–184
 Metallic conduction, 161, 183
 Method of images, 69, 120–122
 Mho, 168, 429n
 Michelson-Morley experiment, 560–561, 562
 Microscopic field, 257
 Microscopic theory
 of conduction, 181–185; of dielectrics, 127–140
 Microwaves, 413, 429
 Minkowski space, 567
 mks units, 28, 36, 416
 table, 602
 Mmf. *See* Magnetomotive force
 Mobility, of electron, 183
 Mode, 476
 Molar susceptibility, 229
 Molecular field, 127, 257–259, 495
 in dielectric, 127–131
 Molecular polarizability, 130–131
 Momentum density of electromagnetic field, 310
 Monochromatic wave, 395–398. *See also* Plane wave; Radiation
 in bounded regions, 441–483; propagation of, 412–440
 Motional emf, 275, 276
 Mott-Zener formula, 510
 Multipole expansion, of electric fields, 46–48
 Mutual inductance, 279–282
 in ac circuits, 320–324; sign convention, 320–321
 Natural frequency, 312
 Natural line width, 539
 Needle-shaped cavity, 104–106
 Network equations
 algebraic approach to solutions, 331; computer solutions, 329–336; matrix approach to solutions, 332
 Network theory, 329
 Neumann function, spherical, 435–436
 Neumann's formula, 280–281, 296
 Neutral medium, 164
 Newtonian mechanics, 564
 Newton (N), 29, 600, 602
 Newton's laws of motion
 and Galilean transformation, 559
 Newton's third law, and Biot-Savart law, 198
 Nodal analysis, 324, 327–328
 Nonconducting media, plane monochromatic waves
 in, 412–417
 Nonperiodic behavior, 305, 306
 Nonpolar molecule, 131
 Nonreflecting coating, 467
 Normal dispersion, 502
 Nucleus, atomic, forces between, 28
 Numerical analysis, 579, 590
 Numerical integration, on computer, 585–587
 Numerical methods, 581–587
 Observation point. *See* Field point
 Oersted (Oe), 602
 Ohm (Ω), 168, 429n
 Ohmic medium, 167, 169
 Ohm's law, 167–170
 and equilibrium, 174; and Kirchhoff's laws, 179; and Laplace's equation, 170; and linear behavior, 181; and plasma, 352
 Open-circuit voltage, 177
 Optical constants, 426–429. *See also* Dielectric constant; Index of refraction
 Optical fiber, 454
 Orbit center, of particle in plasma, 345
 Orbit theory, 342, 343, 344–350
 Ordered spin structure, 268
 Orientational polarizability, 135
 Orthogonal transformation, 567–569, 595–598
 Orthogonal unit vectors, right-handed, 415, 418
 Orthonormal coordinate system, 18
 Oscillating electric dipole, radiation from, 525–529
 Oscillating field, in dispersive media, 484–524, 517–521
 Oscillator strength, 497
 Outer product. *See* Vector product
 Overdamped, 338
 Overrelaxation, 78
 Parallel connection
 of capacitors, 153, 489–490; of impedances, 314–316, 320; of inductors, 282–283; of resistors, 177, 489–490
 Parallel-plate capacitor, 152, 156
 Paramagnetic material, 229, 230, 259
 Paramagnetic resonance, 415n
 Paramagnetic susceptibility, 230, 262
 Paramagnetism, origin of, 261–262
 Particle accelerator, 345
 Particle motion in electric and magnetic fields, 294
 Particle orbits, in plasma, 344–349
 Penetration depth, of superconductors, 368, 377–378
 Periodic behavior, 306
 Permanent magnet, 233, 235, 240–242
 and magnetic energy changes, 298–301
 Permanent magnet circuit, 248–251
 Permeability, 230, 247, 278
 dimensions of, 278, 599, 600; incremental, 233; relative, 396n
 Permittivity, 143, 175, 599, 600, 602
 of dielectric material, 109; frequency dependence of, 489, 517; of vacuum, 29
 Phase retardation, 463–466
 Phase shift
 and propagation of radiation pulses, 519; on reflection, 455, 462–469
 Phase velocity, 459, 471, 473
 defined, 414; of plane waves, 416, 421–422, 426–427
 Pinch effect, 350, 353–355
 Plane of constant amplitude, 457, 458, 461
 Plane of constant phase, 457, 458, 461
 Plane of incidence, 446
 Plane wave
 in conducting media, 423–430; defined, 412; inhomogeneous, 481; in nonconducting media, 412–417
 Plane wave solutions, 412–417
 Plasma, 341–364
 electrical neutrality in, 343–344; electron density in, 343–344; particle orbits in, 344–349; resistivity of, 363; for thermonuclear fusion reactor, 355–358

- asma drift velocity, 346–347
 plasma dynamics, 342–343
 plasma-electron oscillation, 358–363, 430
 plasma frequency, 359, 507–511, 513
 table, 508
 plasma oscillation, 350
 plasma waves, 358–363
 point charge, 27–28
 accelerated, field of, 551–554; in dielectric, 110–112, 118–120; electric field of, 33; electrostatic potential due to, 35; and Liénard-Wiechert potentials, 545; near conducting objects, 69–73; near interface of two dielectrics, 121–122; potential energy of group, 142–143; potential of, 63; relation to charge distribution, 48; self-energy, 146, 148; and solution to wave equation, 404; uniformly moving, field of, 548–551, 574–576
 point dipole, 44
 potential distribution of, 44–45
 Poisson's equation, 56–58
 in cylindrical coordinates, 58; and dielectrics, 116; and Laplace's equation, 58; and magnetic vector potential, 207; and partial differential equation, 57; and potential in plasma, 344; in rectangular coordinates, 57–58; solutions of, 91–92
 polar coordinate system. *See* Spherical coordinate system
 polarizability
 atomic, 132; deformation, 135; molecular, 130–131, 135; orientational, 135
 polarization, 98, 495–496
 and boundary conditions, 114–115; circular, 420; of dielectric medium, 37, 97–103; elliptical, 420; linear, 419, 422; permanent, and ferroelectricity, 136–138; right-hand, 420n; s- and p-, 449–450; units of, 98, 107; of waves, 418–420
 polarization charge. *See also* Bound charge and free charge, 108n; relation to total charge, 106–109
 polarization charge density, 101–103
 polar molecule, 131, 454, 515
 and Langevin-Debye formula, 132–136
 poles, of magnet, 225
 potential, 33–36, 207, 210. *See also* Electrostatic potential; Scalar potential; Vector potential
 coefficient of, 75–77, 148–150; retarded, 545–557; screened Coulomb, 54; units of, 36, 602
 potential difference, 176–177
 of moving wire, 273–277; notation for, 177n
 potential energy, 141. *See also* Electrostatic energy of conservative electrostatic force, 35–36; of electric dipole, 45; of permanent dipole in electric field, 134; of point charges, 142–143; relation to electrostatic potential, 35–36
 power, 169–170
 average, 317; instantaneous, 317; measurement of, 316–317
 power density, 393
 power factor, 316–317, 323n
 power supply, electronic, 176
 Poynting vector, 393, 443, 450, 471–472
 of accelerated point charge, 555; and energy density and flux, 421; and radiation emission, 528, 530, 534, 539, 540; in radiation field of point dipole, 534; units of, 601
 principle of relativity, 560, 562
 sm, dispersion in, 519
 Propagation vector, 413, 445n
 Q-factor of circuit, 316–317, 319–320
 Q-factor of medium, 409
 Quadrupole moment tensor, 48
 Quality factor. *See* Q-factor
 Radiation
 from accelerated point charge, 535, 551–554; emission of, 525–544; frequencies and wavelengths, 305–306, 413; from group of moving charges, 531–535; from half-wave antenna, 529–531; from oscillating dipole, 525–529
 Radiation damping, 500, 537–541
 Radiation field, for small velocities, 554–556
 Radiation resistance, of dipole, 528–529
 Radiation zone field, 533–534, 535–537, 536
 Radio waves, 412, 413
 and skin depth, 430
 Radius of atom, 99, 495
 Radius of electron, 500, 539
 Radius of molecule, 516
 Reactance, and impedance, 313
 Real angle of refraction, 458
 Rectangular coordinate system, 603
 curl in, 16; divergence in, 12–13; gradient in, 8–9; Laplace's equation in, 66–67; Laplacian operator in, 19; Poisson's equations in, 57; verification of vector identities in, 20
 Reflectance, 444–450
 at air-metal interface, 454–462; from conducting plane, 454–462; dependence on angle of incidence, 451–454; and thin layer, 462–469
 Reflection
 angle of, 447; normal incidence for, 441–445; total internal, 454, 460, 467
 Refraction. *See also* Index of refraction
 oblique incidence for, 445–451; real angle of, 458, 459
 Refractive index. *See* Index of refraction
 Regular function, 611
 Relative permeability, 230, 396n
 of ferromagnetic materials, 231–234
 Relativity, 193, 276n, 558–578
 special theory of, 470, 561–566
 Relaxation time, 174, 181–183, 397, 514
 Reluctance, 247–248
 of leakage path, 248
 Remanence, 233
 Resistance, 169
 and impedance, 313; of inductors, 281–283; internal, 177; in superconductor, 365–366, 367; temperature coefficient of, 184–185; units of, 169, 602
 Resistance networks, and Kirchhoff's laws, 175–181
 Resistivity, 167, 177
 in alloys, 184–185; of ferrites, 268; in metals, 183–184; of plasma, 363; required for conductorlike behavior, 175; table, 168; units of, 167
 Resistor, 173, 177
 power to, 316–317; in series circuits, 307–308
 Resonance, 317–320
 Resonance absorption, by bound charges, 499–506
 Resonance frequency
 of electron, 495, 499–506; of parallel circuit, 320; of series circuit, 317–320
 Response function, 485–490
 Restoring force, 515n. *See also* Force constant
 Rest system, 565
 Retarded potential, 406, 545–557
 Retarded time, 406
 Retentivity, 233
 Reversible magnetic systems, 290–298
 Right-hand rule, 4, 11, 420n
 Rotation angle, in Lorentz transformation, 568–569
 Saturation magnetization, 231, 232
 Saturation polarization, 133
 Scalar, 1
 invariant, 595; notation for, 6; transformation, 595
 Scalar field, 1, 3, 12
 Scalar Helmholtz equation, 431–433
 Scalar potential, 402–407
 magnetic, 210–212, 225–226, 238; retarded, 406; units of, 601
 Scalar product, 3
 triple, 4
 Scalar wave equation, 559–562
 Scalar wave equation, inhomogeneous, solution of, 403–404
 Screened Coulomb potential, 54
 Second-order phase transition, in superconductors, 369
 Second-rank tensor, 597
 Self-energy, 146
 of conductor, 146; of point charge, 146, 148, 159
 Self-field, of point charge, 119
 Self-inductance, 277–278
 Semiconductor, 37
 conduction in, 182–183
 density and plasma frequency, 508–511
 Separation of variables, 61, 65, 423, 475
 Series circuit, 177, 312–316, 317–320
 magnetic, 246–247; resonance frequency of, 317–320; steady-state behavior, 312–314
 Series connection
 of capacitors, 153; of impedances, 314–316, 317–320; of inductors, 281–283; of resistor, 177
 Sheath, 341
 Shielding distance, 341, 343–344
 Siemens (S), 168, 429n
 Simultaneity, 564–565
 SI units, 29
 Skin depth, 509, 510, 540
 defined, 427; at oblique incidence, 460; in various media, 429–430
 Smakula's equation, 523
 Snell's law, 445, 453
 and complex Fresnel coefficients, 454, 458–459; and reflection at interface, 447, 450, 452
 Soft iron, 231n, 245, 249–250
 Software, 583–585
 modification of, for electromagnetic theory, 582; testing, 590
 Solenoid, 203, 215, 297–298, 377–378
 Space-time, geometry of, 566–567
 Space-time interval, 563
 Space-time point, 563
 Special theory of relativity, 558–578
 Spherical Bessel function, 435–436
 Spherical coordinate system, 18n, 603–604
 and divergence, 13; and electrostatic images, 72; Laplace's equation in, 60; Poisson's equations in, 57–58; polar coordinates in, 9; wave equation in,

- 430–437
 Spherical Neumann function, 435
 Spherical symmetry, 63
 Spherical wave, 405, 430–437
 SPICE, 335–336
 Spin, 261
 and ferromagnetism, 264, 268
 Spinel structure, 268
 Spontaneous magnetization, 264
 Steady-state behavior, 170–173, 306–307
 of series circuit, 312–314
 Steady-state conduction problems, 170–173
 Stokes's theorem, 16–17
 and Ampere's circuital law, 205; and boundary conditions, 399; extensions of, 21; and Neumann's formula, 281; and transport current, 228
 Summation convention, 572
 Sum rules, 524
 Superconducting cylinder, and perfect flux exclusion, 373–374
 Superconducting materials, properties of, 367–370
 Superconducting sphere
 and perfect flux exclusion, 370–373; in uniform field, 378–381
 Superconductivity, 365–385
 Superconductor, 365–385
 critical temperature, 366, 367–368; as magnetic material, 374, 375; perfect conductivity of, 367, 370–371; perfect diamagnetism of, 370–371; permeability and susceptibility, 367, 370, 376; theory of electromagnetic behavior of, 376–378
 Superposition of plane waves, 417
 Superposition theorem, 59
 Surface charge density, 30, 49, 101, 114, 236, 400
 and boundary conditions, 399–400; notation for, 167
 Surface current, 223
 Surface current density, 49, 236, 400
 Surface density of magnetic pole strength, 225
 Surface integral, 11
 Surface stress, 119
 Susceptibility
 diamagnetic, 230, 260–261; electric, 109–110, 494, 496, 602; magnetic, 229–230, 602; molar, 229; and molecular fields, 139; paramagnetic, 230, 262; of superconductors, 370, 376
 Temperature coefficient of resistance, 184–185
 table, 168
 Tensor, 2, 594–597
 antisymmetric, 571; defined, 597; dielectric, 391n; field, 567, 571n, 572–573; quadrupole moment, 48
 Tesla (T), 191, 202–203, 212n, 602
 Test charge, 31–32, 103
 TE wave. *See* Transverse electric wave
 Thermally induced imperfections, 183
 Thermal velocity, 182
 Thermionic emitting cathode, 162
 Thermoelectric effect, 164
 Thermonuclear fusion, and magnetic confinement, 355–358
 Thermonuclear reactors, 341
 Theta-pinch fusion reactor, 357
 Thomson scattering cross section, 537–541
 Time constant of circuit, 309–310
 Time constant of medium, 174, 183. *See also* Relaxation time
 Time dilation, 565
 TM wave. *See* Transverse magnetic wave
 Tokamak reactors, 356–357
 Toroid, 216
 filled with ferromagnetic material, 242–245; self-inductance of, 277–278
 Toroidal plasma, 356, 357n
 Torque, 154–156, 194
 magnetic, on rigid circuits, 294, 295–298
 Total internal reflection, 454, 460, 467
 Transfer functions, 328–329, 581
 Transformation
 linear, 594; orthogonal, 567–569, 595–598
 Transformation law
 for electromagnetic field, 573–574
 for vectors and tensors, 596–597, 598
 Transformer, 322–324
 Transient behavior, 306–307
 elementary, 308–312; and nonperiodic behavior, 305
 Transition temperature. *See* Critical temperature
 Transmittance, 444, 450
 relation to reflectance, 451; and thin layer, 462–469
 Transparent medium, 427, 453, 454, 505–511
 maximum thickness of, 466; normal dispersion behavior, 502
 Transport current, 218, 228
 Transport current density, 226–228
 Transpose matrix, 595
 Transverse dispersion relation, 416
 Transverse electric wave, 433, 436, 470, 474–476
 Transverse electromagnetic wave, 415, 427
 Transverse magnetic wave, 433, 436, 470
 Trapezoidal rule, 585–587
 Triple scalar product, 4
 Triple vector product, 4
 Turns ratio, 324
 Two-fluid model of superconductivity, 369
 Ultraviolet radiation, 413
 Underdamped, 338
 Uniformly magnetized material, 238–242
 Uniformly magnetized sphere, 240
 Uniformly moving charge, 548–551, 574–576
 Uniform wave, 426n
 Uniqueness theorem, 59
 Unitary transformation, 567n
 Units, 28–29, 599–602
 table, 602
 Unit vector, 4, 413
 Universal time, 564
 Vacuum
 current in, 162; electric properties of, 112; permittivity of, 29
 Van de Graaff generator, 176
 Vector, 1–3, 594–598
 addition, 3; curl of, 14–17; divergence of, 12–14; integration, 9; multiplication, 3; negative of, 2; transformation, 594–597
 Vector algebra, 2–5
 Vector analysis, 1–25
 Vector differential operator. *See* del operator
 Vector field, 2, 3, 10
 electric field as, 33
 Vector Helmholtz equation, 379, 431–433
 Vector identities, 20
 Vector integral theorems, 21–22, 23
 Vector potential, 402–407
 magnetic, 207–208, 222, 402; retarded, 406; units of, 601
 Vector product, 3–5
 del operations on, 20–21; triple, 4–5
 Vector wave equation, inhomogeneous, solution to, 403, 404–405
 Velocity of energy propagation, 471–473
 Velocity of light, 192, 396, 471, 600–602
 constancy of, 562–563
 Velocity of wave, 559
 Viscosity, 516
 Visible light, 412, 413
 Voltage. *See also* Applied voltage; Complex voltage; Counter voltage
 constant, 306n, 308–312; effective value, 317; open-circuit, 177; periodic, 306–307
 Voltage gain, in four-terminal network, 329
 Voltage generator, 326–327
 Volt (V), 36, 602
 Volume charge density, notation for, 167n
 Volume density of charge. *See* Charge density
 Volume integral, 11
 Wave. *See also* Monochromatic wave; Plane wave; Radiation
 homogeneous, 426n; intensity of, 443–444; longitudinal, 430, 440, 509; spherical, 431–433
 Wave, electromagnetic, 412–440, 509
 Wave equation, 394–395
 damped, 395; Lorentz transformation of, 561–562; one-dimensional, 405; with sources, 401–407
 Waveguide
 cylindrical, 474; parallel plate, 469–474; rectangular, 475–477
 Wavelength, 305–306, 396, 413
 Wave motion, in plasma, 358–363
 Wave propagation, 394–398, 412–440
 in conducting medium, 397–398, 412n; in moving reference system, 559–566; between parallel conducting plates, 469–474; in plasma, 360
 Wave velocity, group, 471
 Wave velocity, phase, 413, 471
 Weber (Wb), 212
 Weiss-Heisenberg theory, 264
 Weiss molecular field, 264
 Wheatstone bridge, 189
 Wire
 charged, 84–88; conduction in, 168–170; in magnetic field, 273–277; superconducting, 365, 374, 375, 381–384
 Work
 to assemble charge distribution, 141–142; to energize current circuits, 289–292
 World point, 567
 World scalars and vectors, 567
 X-ray, 412, 413
 X-ray scattering, 539, 540n
 Zero potential, 72
 Zonal harmonics, 60–63, 117, 239, 240

www.pearsonhighered.com



ISBN-13: 978-0-321-58174-7
ISBN-10: 0-321-58174-1



9 780321 581747