



CD-ROM
Included

1999 Edition

FAWWAZ T. ULABY



FUNDAMENTALS OF

APPLIED
ELECTROMAGNETICS

INTEL-1014
U.S. Patent 7,151,430

FUNDAMENTAL PHYSICAL CONSTANTS		
CONSTANT	SYMBOL	VALUE
speed of light in vacuum	c	$2.998 \times 10^8 \simeq 3 \times 10^8$ m/s
gravitational constant	G	6.67×10^{-11} N·m ² /kg ²
Boltzmann's constant	K	1.38×10^{-23} J/K
elementary charge	e	1.60×10^{-19} C
permittivity of free space	ϵ_0	$8.85 \times 10^{-12} \simeq \frac{1}{36\pi} \times 10^{-9}$ F/m
permeability of free space	μ_0	$4\pi \times 10^{-7}$ H/m
electron mass	m_e	9.11×10^{-31} kg
proton mass	m_p	1.67×10^{-27} kg
Planck's constant	h	6.63×10^{-34} J·s
intrinsic impedance of free space	η_0	$376.7 \simeq 120\pi$ Ω

FUNDAMENTAL SI UNITS		
DIMENSION	UNIT	SYMBOL
Length	meter	m
Mass	kilogram	kg
Time	second	s
Electric current	ampere	A
Temperature	kelvin	K
Amount of substance	mole	mol

MULTIPLE & SUBMULTIPLE PREFIXES					
PREFIX	SYMBOL	MAGNITUDE	PREFIX	SYMBOL	MAGNITUDE
exa	E	10^{18}	milli	m	10^{-3}
peta	P	10^{15}	micro	μ	10^{-6}
tera	T	10^{12}	nano	n	10^{-9}
giga	G	10^9	pico	p	10^{-12}
mega	M	10^6	femto	f	10^{-15}
kilo	k	10^3	atto	a	10^{-18}



When this book was first taught from, each student was asked to write a brief statement describing his or her understanding of what role electromagnetics plays in science, technology, and society. The following statement, submitted by Mr. Schaldenbrand, who has since graduated with a B.S.E. degree, was selected for inclusion here:

Electromagnetics has done more than just help science. Since we have such advanced communications, our understanding of other nations and nationalities has increased exponentially. This understanding has led and will lead the governments of the world to work towards global peace. The more knowledge we have about different cultures, the less foreign these cultures will seem. A global kinship will result, and the by-product will be harmony. Understanding is the first step, and communication is the means. Electromagnetics holds the key to this communication, and therefore is an important subject for not only science, but also the sake of humanity.

Mike Schaldenbrand

The University of Michigan

F U N D A M E N T A L S O F

**APPLIED
ELECTROMAGNETICS**
1999 Edition

F U N D A M E N T A L S O F

APPLIED ELECTROMAGNETICS

1999 Edition

Fawwaz T. Ulaby

The University of Michigan



PRENTICE HALL, Upper Saddle River, New Jersey 07458

Library of Congress Cataloging-in-Publication Data

Ulaby, Fawwaz T. (Fawwaz Tayssir)
Fundamentals of applied electromagnetics / Fawwaz T. Ulaby
p. cm.
Includes bibliographical references and index.
ISBN 0-13-011554-1
1. Electromagnetism. 2. Electromagnetism—Industrial applications
I. Title.
QC760.U49 1999
537--dc21 98-41930
CIP

Publisher: Tom Robbins
Acquisitions editor: Alice Dworkin
Editorial/production supervision: AnnMarie Kalajian
Cover photo: © Uniphoto Picture Agency
Cover designer: Joseph Sengotta/ Bruce Kensalaar
Creative director: Paula Maylahn
Managing editor: Eileen Clark
Editor-in-chief: Marcia Horton
Assistant vice president of production and manufacturing: David W. Riccardi
Manufacturing buyer: Pat Brown
Editorial assistant: Dan DePasquale



© 1997, 1999 Prentice-Hall, Inc.
Simon & Schuster / A Viacom Company
Upper Saddle River, New Jersey 07458

All rights reserved. No part of this book may be
reproduced, in any form or by any means,
without permission in writing from the publisher.

The author and publisher of this book have used their best efforts in preparing this book. These efforts
include the development, research, and testing of the theories and programs to determine their effective-
ness. The author and publisher make no warranty of any kind, expressed or implied, with regard to these
programs or the documentation contained in this book. The author and publisher shall not be liable in any
event for incidental or consequential damages in connection with, or arising out of, the furnishing, perfor-
mance, or use of these programs.

Printed in the United States of America
10 9 8 7 6 5 4 3 2

ISBN 0-13-011554-1

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan Inc., *Tokyo*
Simon & Schuster Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

Love is life's wonder!

I dedicate this book to Jean.

Children are life's joy.

*I also dedicate this book to
Aziza, Gabriel, Laith, and Neda.*

C O N T E N T S

PREFACE xiii

1 INTRODUCTION: WAVES AND PHASORS 3

- 1-1 Dimensions, Units, and Notation 4
- 1-2 The Nature of Electromagnetism 5
 - 1-2.1 The Gravitational Force: A Useful Analogue 6
 - 1-2.2 Electric Fields 7
 - 1-2.3 Magnetic Fields 9
 - 1-2.4 Static and Dynamic Fields 11
- 1-3 Traveling Waves 12
 - 1-3.1 Sinusoidal Wave in a Lossless Medium 14
 - 1-3.2 Sinusoidal Wave in a Lossy Medium 17
- 1-4 The Electromagnetic Spectrum 19
- 1-5 Review of Complex Numbers 22
- 1-6 Review of Phasors 24
- Problems 29

2 TRANSMISSION LINES 33

- 2-1 General Considerations 35
 - 2-1.1 The Role of Wavelength 36
 - 2-1.2 Propagation Modes 37
- 2-2 Lumped-Element Model 39
- 2-3 Transmission-Line Equations 43
- 2-4 Wave Propagation on a Transmission Line 44
- 2-5 The Lossless Transmission Line 47
 - 2-5.1 Voltage Reflection Coefficient 48
 - 2-5.2 Standing Waves 51
- 2-6 Input Impedance of the Lossless Line 55
- 2-7 Special Cases of the Lossless Line 58
 - 2-7.1 Short-Circuited Line 58
 - 2-7.2 Open-Circuited Line 60
 - 2-7.3 Application of Short-Circuit and Open-Circuit Measurements 60
 - 2-7.4 Lines of Length $l = n\lambda/2$ 62
 - 2-7.5 Quarter-Wave Transformer 62
 - 2-7.6 Matched Transmission Line: $Z_0 L = Z_0$ 62

- 2-8 Power Flow on a Lossless Transmission Line 64
 - 2-8.1 Instantaneous Power 64
 - 2-8.2 Time-Average Power 65
- 2-9 The Smith Chart 66
 - 2-9.1 Parametric Equations 66
 - 2-9.2 Input Impedance 71
 - 2-9.3 SWR, Voltage Maxima and Minima 73
 - 2-9.4 Impedance to Admittance Transformations 74
- 2-10 Impedance Matching 80
- 2-11 Transients on Transmission Lines 84
 - 2-11.1 Transient Response 84
 - 2-11.2 Bounce Diagrams 88
- Problems 92

3 VECTOR ANALYSIS 100

- 3-1 Basic Laws of Vector Algebra 101
 - 3-1.1 Equality of Two Vectors 102
 - 3-1.2 Vector Addition and Subtraction 103
 - 3-1.3 Position and Distance Vectors 103
 - 3-1.4 Vector Multiplication 104
 - 3-1.5 Scalar and Vector Triple Products 107
- 3-2 Orthogonal Coordinate Systems 108
 - 3-2.1 Cartesian Coordinates 109
 - 3-2.2 Cylindrical Coordinates 109
 - 3-2.3 Spherical Coordinates 113
- 3-3 Transformations between Coordinate Systems 115
 - 3-3.1 Cartesian to Cylindrical Transformations 115
 - 3-3.2 Cartesian to Spherical Transformations 117
 - 3-3.3 Cylindrical to Spherical Transformations 119
 - 3-3.4 Distance between Two Points 119
- 3-4 Gradient of a Scalar Field 120
 - 3-4.1 Gradient Operator in Cylindrical and Spherical Coordinates 121
 - 3-4.2 Properties of the Gradient Operator 122
- 3-5 Divergence of a Vector Field 123
 - 3-5.1 Divergence Theorem 125
 - 3-5.2 Remarks on Notation 125
- 3-6 Curl of a Vector Field 127
 - 3-6.1 Vector Identities Involving the Curl 128
 - 3-6.2 Stokes's Theorem 129
- 3-7 Laplacian Operator 130
 - Problems 132

4 ELECTROSTATICS 138

- 4-1 Maxwell's Equations 139
- 4-2 Charge and Current Distributions 140
 - 4-2.1 Charge Densities 140
 - 4-2.2 Current Density 142
- 4-3 Coulomb's Law 143
 - 4-3.1 Electric Field due to Multiple Point Charges 144
 - 4-3.2 Electric Field due to a Charge Distribution 145
- 4-4 Gauss's Law 148
- 4-5 Electric Scalar Potential 151
 - 4-5.1 Electric Potential as a Function of Electric Field 151
 - 4-5.2 Electric Potential due to Point Charges 153
 - 4-5.3 Electric Potential due to Continuous Distributions 153
 - 4-5.4 Electric Field as a Function of Electric Potential 147
 - 4-5.5 Poisson's Equation 155
- 4-6 Electrical Properties of Materials 156
- 4-7 Conductors 157
 - 4-7.1 Resistance 158
 - 4-7.2 Joule's Law 160
- 4-8 Dielectrics 161
- 4-9 Electric Boundary Conditions 163
 - 4-9.1 Dielectric—Conductor Boundary 166
 - 4-9.2 Conductor—Conductor Boundary 167
- 4-10 Capacitance 168
- 4-11 Electrostatic Potential Energy 172
- 4-12 Image Method 173
 - Problems 176

5 MAGNETOSTATICS 186

- 5-1 Magnetic Forces and Torques 187
 - 5-1.1 Magnetic Force on a Current-Carrying Conductor 189
 - 5-1.2 Magnetic Torque on a Current-Carrying Loop 192
- 5-2 The Biot—Savart Law 195
 - 5-2.1 Magnetic Field due to Surface and Volume Current Distributions 196
 - 5-2.2 Magnetic Field of a Magnetic Dipole 199
- 5-3 Magnetic Force between Two Parallel Conductors 200
- 5-4 Maxwell's Magnetostatic Equations 201
 - 5-4.1 Gauss's Law for Magnetism 201
 - 5-4.2 Ampere's Law 202
- 5-5 Vector Magnetic Potential 206
- 5-6 Magnetic Properties of Materials 208
 - 5-6.1 Orbital and Spin Magnetic Moments 208

5-6.2	Magnetic Permeability	209
5-6.3	Magnetic Hysteresis of Ferromagnetic Materials	210
5-7	Magnetic Boundary Conditions	213
5-8	Inductance	214
5-8.1	Magnetic Field in a Solenoid	215
5-8.2	Self-inductance	216
5-8.3	Mutual Inductance	218
5-9	Magnetic Energy	219
	Problems	221
6	MAXWELL'S EQUATIONS FOR TIME-VARYING FIELDS	229
6-1	Faraday's Law	231
6-2	Stationary Loop in a Time-Varying Magnetic Field	233
6-3	The Ideal Transformer	237
6-4	Moving Conductor in a Static Magnetic Field	238
6-5	The Electromagnetic Generator	241
6-6	Moving Conductor in a Time-Varying Magnetic Field	243
6-7	Displacement Current	244
6-8	Boundary Conditions for Electromagnetics	246
6-9	Charge-Current Continuity Relation	247
6-10	Free-Charge Dissipation in a Conductor	249
6-11	Electromagnetic Potentials	250
6-11.1	Retarded Potentials	250
6-11.2	Time-Harmonic Potentials	251
	Problems	255
7	PLANE-WAVE PROPAGATION	260
7-1	Time-Harmonic Fields	262
7-1.1	Complex Permittivity	263
7-1.2	Wave Equations for a Charge-Free Medium	263
7-2	Plane-Wave Propagation in Lossless Media	264
7-2.1	Uniform Plane Waves	264
7-2.2	General Relation between E and H	268
7-3	Wave Polarization	269
7-3.1	Linear Polarization	270
7-3.2	Circular Polarization	271
7-3.3	Elliptical Polarization	273
7-4	Plane-Wave Propagation in Lossy Media	276
7-4.1	Low-Loss Dielectric	278
7-4.2	Good Conductor	278
7-5	Current Flow in a Good Conductor	280
7-6	Electromagnetic Power Density	283
7-6.1	Plane Wave in a Lossless Medium	284

7-6.2	Plane Wave in a Lossy Medium	285
7-6.3	Decibel Scale for Power Ratios	286
	Problems	288
8	WAVE REFLECTION AND TRANSMISSION, AND GEOMETRIC OPTICS	292
8-1	Wave Reflection and Transmission at Normal Incidence	294
8-1.1	Boundary between Lossless Media	294
8-1.2	Transmission-Line Analogue	297
8-1.3	Power Flow in Lossless Media	298
8-1.4	Boundary between Lossy Media	301
8-2	Snell's Laws	303
8-3	Fiber Optics	306
8-4	Wave Reflection and Transmission at Oblique Incidence	308
8-4.1	Perpendicular Polarization	309
8-4.2	Parallel Polarization	313
8-4.3	Brewster Angle	315
8-5	Reflectivity and Transmissivity	316
8-6	Geometric Optics	319
8-7	Images Formed by Mirrors	320
8-7.1	Images Formed by Plane Mirrors	321
8-7.2	Images Formed by Spherical Mirrors	322
8-8	Images Formed by Spherical Lenses	324
	Problems	332
9	RADIATION AND ANTENNAS	340
9-1	The Short Dipole	343
9-1.1	Far-Field Approximation	345
9-1.2	Power Density	346
9-2	Antenna Radiation Characteristics	348
9-2.1	Antenna Pattern	349
9-2.2	Beam Dimensions	351
9-2.3	Antenna Directivity	351
9-2.4	Antenna Gain	354
9-2.5	Radiation Resistance	354
9-3	Half-Wave Dipole Antenna	355
9-3.1	Directivity of $\lambda/2$ Dipole	357
9-3.2	Radiation Resistance of $\lambda/2$ Dipole	357
9-3.3	Quarter-Wave Monopole Antenna	358
9-4	Dipole of Arbitrary Length	359
9-5	Effective Area of a Receiving Antenna	360
9-6	Friis Transmission Formula	363
9-7	Radiation by Large-Aperture Antennas	365

- 9-8 Rectangular Aperture with Uniform Aperture Distribution 368
 - 9-8.1 Beamwidth 369
 - 9-8.2 Directivity and Effective Area 370
- 9-9 Antenna Arrays 371
- 9-10 *N*-Element Array with Uniform Phase Distribution 378
- 9-11 Electronic Scanning of Arrays 380
 - 9-11.1 Uniform-Amplitude Excitation 382
 - 9-11.2 Array Feeding 382
- Problems 386

10 SATELLITE COMMUNICATION SYSTEMS AND RADAR SENSORS 390

- 10-1 Satellite Communication Systems 391
- 10-2 Satellite Transponders 393
- 10-3 Communication-Link Power Budget 396
- 10-4 Antenna Beams 398
- 10-5 Radar Sensors 399
 - 10-5.1 Basic Operation of a Radar System 399
 - 10-5.2 Unambiguous Range 400
 - 10-5.3 Range and Angular Resolutions 401
- 10-6 Target Detection 402
- 10-7 Doppler Radar 404
- 10-8 Monopulse Radar 406
- Problems 410

APPENDIX A SYMBOLS, QUANTITIES, AND UNITS 411

APPENDIX B MATERIAL CONSTANTS OF SOME COMMON MATERIALS 413

APPENDIX C MATHEMATICAL FORMULAS 415

APPENDIX D ANSWERS TO ODD-NUMBERED PROBLEMS 417

BIBLIOGRAPHY 423

INDEX 425

P R E F A C E

Why Another EM Book?

Several textbooks are currently available for teaching electromagnetics to students majoring in electrical engineering. So, why do we need another one? The answer is simple: (1) the curriculum for a bachelor's degree in electrical engineering is undergoing a radical change, perhaps more radical in both outlook and content than any of the curricular changes we have witnessed over the past several decades, and (2) available textbooks are incompatible with the philosophy of the new curriculum proposed for the 21st century (see, for example, the article by Director et al. in the *Proceedings of the IEEE*, September 1995).

The Changing Curriculum

For a bachelor's degree in electrical engineering, the curriculum has undergone about one major change per decade. In the 1960s, courses concerning solid-state devices were introduced and those on vacuum-tube electronics were slowly phased out. In the 1970s, courses on electric machinery nearly disappeared from the curricula of most universities and were replaced with computer-programming courses. More computer-related and digital processing courses were added in the 1980s, mainly by increasing the volume of required courses and reducing the number of technical and free electives. Also, there has been a sustained effort to incorporate new knowledge and assimilate the rapidly evolving role of technology in the undergraduate curriculum. More material was added to courses, greater effort was expected from students, and the number of elective credit hours rapidly approached zero. By the early 1990s, the average student at a U.S. university needed closer to five years to complete what was originally designed as a four-year program.

This scenario was not limited to electrical engineering; indeed, in almost any engineering discipline, the curriculum had become too demanding in terms of time to complete the B.S. degree and too inflexible to accommodate changes. From these pressures was born a new philosophy embracing the following tenets: (1) the B.S. degree program should be scaled back to four years, (2) the required part of the B.S. degree program should focus on the teaching of fundamentals, but with greater exposure to engineering applications, and (3) the electives portion of the program should be increased substantially to allow the student to explore both engineering and nonengineering areas of interest. While this new philosophy has been adopted, at least in principle, by many engineering schools, the task of revising the curriculum has been a major challenge, particularly at the course level. The new electrical engineering curriculum calls for allocating fewer credit hours in many of the traditionally core areas—electromagnetics (EM) among them. At many universities the required EM content of the program has been reduced from two courses down to one. For some, the intent is to continue to offer a two-course sequence, but with only the first being part of the required core and the second being for students interested in deepening their knowledge in electromagnetics.

Course Contents

Given these objectives and associated boundary conditions, what then should be the contents of a one-semester or a two-course sequence in electromagnetics and what texts might one use for this purpose? To answer these questions, we should briefly review the traditional approaches that have been in use over the past two decades. Most EM books share a common template. They begin with one or more chapters on vector calculus and coordinate systems, followed by two or

more chapters on statics (electrostatics and magnetostatics). These topics typically constitute half of the total material, and the second half usually covers dynamics (time-varying fields, wave propagation and reflection, waveguides and resonators, and antennas). Such an arrangement of topics presents two problems. First, starting a course with a heavy dose of mathematics tends to “turn the students off.” And second, whereas electrostatics and magnetostatics are interesting subjects in themselves and have many practical applications, their importance pales in comparison with the many applications of time-varying fields to communication systems, radar, optics, computers, and solid-state electronics, among others. Thus, teaching statics only in the required portion of the B.S. degree curriculum would leave an electrical engineering student seriously lacking in ability to deal with most electromagnetic phenomena.

Because of the heavy emphasis on mathematical manipulations using vector calculus and the relatively dry character of electrostatics and magnetostatics, the student who completes such a required course is unlikely to select to take the second course in the sequence, which is where many of the topics relevant to engineering practice are covered. Clearly, teaching from a book based on this traditional sequence of topics is no longer compatible with the format of the new curriculum proposed for the 21st century.

In past years, attempts have been made by educators at various universities to teach the dynamics of electromagnetics without preceding it with a study of statics. These attempts have not been very successful, primarily because in dynamics students must deal with several fundamental parameters simultaneously. These include spatial coordinates denoting location, vectors denoting direction, time, and the interconnection between the electric and magnetic fields. In statics, the electric and magnetic fields are independent of one another and time is not a variable, thereby reducing the number of parameters from four to two. By studying electrostatics and magnetostatics first, the student can much more easily tackle situations involving time-varying fields.

This Book

In view of this background, how then does one design

an appropriate one-semester or two-course sequence in electromagnetics? My answer to this question is characterized in this book by the following elements:

1. I wanted to begin by building a bridge between what was already familiar to a third-year electrical engineering student and the electromagnetics material. Prior to enrolling in an EM course, a typical student will have taken two or more courses in circuits. Thus, he or she is familiar with circuit analysis, Ohm's law, Kirchhoff's current and voltage laws, and related topics. Transmission lines constitute a *natural* bridge between electric circuits and electromagnetics. Without having to deal with vectors or fields, the student uses already familiar concepts to learn about wave motion, the reflection and transmission of power, phasors, impedance matching, and many of the properties of wave propagation in a guided structure. All of these newly learned concepts will prove invaluable later (in Chapters 7 through 9) and will facilitate the learning of how plane waves propagate in free space and in material media. Transmission lines are covered in Chapter 2, which is preceded in Chapter 1 with reviews of complex numbers and phasor analysis.

2. The next part of the book, contained in Chapters 3 through 5, covers vector analysis, electrostatics, and magnetostatics. Compared with most EM textbooks written for undergraduate instruction, the present book differs in terms of its presentation of these three topics in the following two ways: Of the total number of pages contained in the book, about 30% are allocated to these topics, compared with 50% or more in most EM textbooks. The electrostatics chapter begins with Maxwell's equations for the time-varying case, which are then specialized to electrostatics and magnetostatics, thereby providing the student with an overall framework for what is to come and showing him or her why electrostatics and magnetostatics are special cases of the more general time-varying case.

3. Unlike most EM textbooks, this book does not have a chapter on waveguides and resonators. Space limitations and the fact that waveguides are no longer used as widely as they had been prior to 1980 dictated this change.

4. Chapter 6 deals with time-varying fields and sets

Suggested Syllabi

	Two-semester Syllabus 6 credits (42 contact hours per semester)		One-semester Syllabus 4 credits (56 contact hours)	
Chapter	Sections	Hours	Sections	Hours
1 Introduction	All	4	All	4
2 Transmission Lines	All	12	2-1 to 2-8 and 2-11	8
3 Vector Analysis	All	8	All	8
4 Electrostatics	All	8	4-1 to 4-10	6
5 Magnetostatics	All	7	5-1 to 5-5 and 5-7 to 5-8	5
Exams		3		2
	Total for first semester	42		
6 Maxwell's Equations	All	6	6-1 to 6-3, and 6-6	3
7 Plane-wave Propagation	All	7	7-1 to 7-4, and 7-6	6
8 Wave Reflection and Transmission	All	9	8-1 to 8-3, and 8-6	7
9 Radiation and Antennas	All	10	9-1 to 9-6	6
10 Satellite Communication Systems and Radar Sensors	All	5	None	—
Exams		3	1	
	Total for second semester	40	Total	56
Extra Hours	2	0		

the stage for the material in Chapters 7 through 9. Chapter 7 covers plane-wave propagation in dielectric and conducting media, and Chapter 8 covers reflection and transmission at discontinuous boundaries and introduces the student to fiber optics and the imaging properties of mirrors and lenses.

In Chapter 9, the student is introduced to the principles of radiation by currents flowing in wires, such as dipoles, as well as to radiation by apertures, such as a horn antenna or an opening in an opaque screen illuminated by a light source.

5. To give the student a taste of the wide-ranging applications of electromagnetics in today's technological society, Chapter 10 concludes the book with overview presentations of two system examples: satellite communication systems and radar sensors.

6. The material in this book was written for a two-semester sequence of six credits, but it is possible to trim it down to generate a syllabus for a one-semester

four-credit course. The above table provides syllabi for each of these two options.

In writing this book, I avoided lengthy derivations of theorems, particularly those involving extensive use of vector calculus. My goal has been to help the student to develop competence in applying vector calculus to solve electromagnetic problems of practical interest. I view vector calculus and mathematics in general as useful tools and not as ends in themselves. Throughout the material, emphasis is placed on using the mathematics to explain and clarify the physics, followed with practical examples intended to demonstrate the engineering relevance of physical concepts. I believe the combination of the approach used in presenting the material, the arrangement of topics covered in the book, and the relative emphasis in favor of dynamics constitutes an effective algorithm for equipping our future graduates with a relevant foundation in applied electromagnetics.

Acknowledgments

In writing this book I have been fortunate to have had valuable help from many talented individuals. First and foremost, I would like to acknowledge my most sincere gratitude to Roger DeRoo, Richard Carnes and Jim Ryan. I am indebted to Roger DeRoo, not only for the numerous discussions we have had on how to best present electromagnetic concepts and applications, without getting lost in the mathematics, but also for his painstaking review of several drafts of the manuscript. Richard Carnes is unquestionably the best technical typist I have ever worked with; his mastery of LaTeX, coupled with his attention to detail, made it possible to arrange the material in a clear and smooth format. The artwork was done by Jim Ryan, who skillfully transformed my rough sketches into drawings that are both professional looking and esthetically pleasing.

I would like to thank my colleagues at the Radiation Laboratory of the University of Michigan for their advice and support and, in particular, I gratefully acknowledge Linda Katehi, Chen-To Tai and Kamal Sarabandi for their generous counsel and constant encouragement. Thanks also are due to my students,

who have suffered through two semesters of having a text in the form of notes and who have provided many suggestions for improving and clarifying the presentation. Moreover, I am grateful to the following graduate students for reading through parts or all of the manuscript and for helping me with the solutions manual: Bryan Hauck, Yanni Kouskoulas, and Paul Siqueira.

I am grateful to the reviewers for their valuable comments and suggestions. They include Constantine Balanis of Arizona State University, Harold Mott of the University of Alabama, David Pozar of the University of Massachusetts, S. N. Prasad of Bradley University, Robert Bond of New Mexico Institute of Technology, Mark Robinson of the University of Colorado at Colorado Springs, and Raj Mittra of the University of Illinois. I appreciate the dedicated efforts of the staff at Prentice Hall and I am grateful for their help in shepherding this project through the publication process in a very timely manner. I also would like to thank Mr. Ralph Pescatore for copyediting the manuscript.

Fawwaz T. Ulaby
Ann Arbor, Michigan

The 1999 Edition

This 1999 edition represents an intermediate step between the standard textbook format of the preceding edition and the CD-ROM interactive supplement to be introduced with the 2001 edition. The CD-ROM of the 2001 edition will include audio/video demonstrations, interactive “laboratory experiments,” drill exercises, and complete solutions of a select subset of the end-of-chapter problems. The material developed under this project will be added to a web site at the University of Michigan* on a continuing basis. Suggestions and ideas are very welcome.

The present edition differs from its predecessor in the following ways:

1. Typographical errors have been corrected.
2. The total number of end-of-chapter problems contained in the book has been increased from 238 problems to 373 problems, providing greater scope and a larger number of application examples.
3. Each copy of this book is accompanied with a CD-ROM, placed in the pocket attached to the inside back cover. The CD-ROM contains three folders:
 - (a) Figures: This folder, containing copies of all fig-

ures appearing in the book, is made available to facilitate the process of reproducing the figures practicably by instructors who may want to generate viewgraphs of the figures electronically.

(b) Smith Chart: The Smith-Chart file allows the student to generate a high quality print of a Smith chart for his or her use in Chapter 2.

(c) Sample Solutions: Exposure to examples of problems and their solutions is an indispensable learning tool for the student. The book contains many such examples, but in an effort to further enhance this component of the book, complete solutions are provided for 45 problems, selected from among the end-of-chapter problems appearing in the book. Those selected for inclusion in the CD-ROM are identified by the symbol  next to the problem statement.

I would like to take this opportunity to thank my Prentice Hall editor, Mr. Tom Robbins, for his encouragement and enthusiastic support of the CD-ROM project

Fawwaz T. Ulaby
Ann Arbor, Michigan

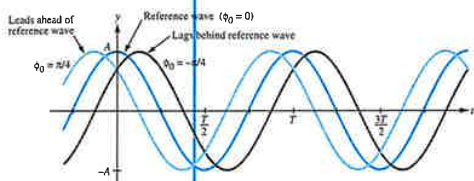
*<http://www.eecs.umich.edu/emag/>

CHAPTER 1

Introduction: Waves and Phasors

Overview

- 1-1 Dimensions, Units, and Notation
- 1-2 The Nature of Electromagnetism
- 1-3 Traveling Waves
- 1-4 The Electromagnetic Spectrum
- 1-5 Review of Complex Numbers
- 1-6 Review of Phasors



OVERVIEW

The subject of this book is applied electromagnetics, which encompasses the study of electric and magnetic phenomena and their engineering applications, under both static and dynamic conditions. Primary emphasis is placed on the fundamental properties of time-varying (dynamic) electromagnetic fields because of their greater relevance to practical problems in many engineering disciplines, including microwave and optical communications, radar systems, bioelectromagnetics, and high-speed microelectronics, among others. We shall study wave propagation in guided media, such as coaxial transmission lines and optical fibers, wave reflection and transmission at the interface between dissimilar media, imaging properties of mirrors and lenses, radiation by antennas, and several other related topics. The concluding chapter is intended to illustrate a few aspects of applied electromagnetics through an examination of design considerations associated with the use and operation of radar sensors and satellite communication systems.

By way of an example, let us consider the communication network depicted in Fig. 1-1. The satellite segment uses a *multiple-beam* antenna so as to communicate with several ground stations simultaneously. The signal received from ground station 1 at frequency $f_1 = 6$ GHz* is amplified and then downconverted to a frequency $f_2 = 4$ GHz for retransmission to ground station 2. The frequency conversion is facilitated by a microwave solid-state device called a *mixer*, which in this application converts the input frequency f_1 at port M_1 to a frequency f_2 at port M_2 such that $f_2 = f_1 - f_0$, where f_0 is the satellite

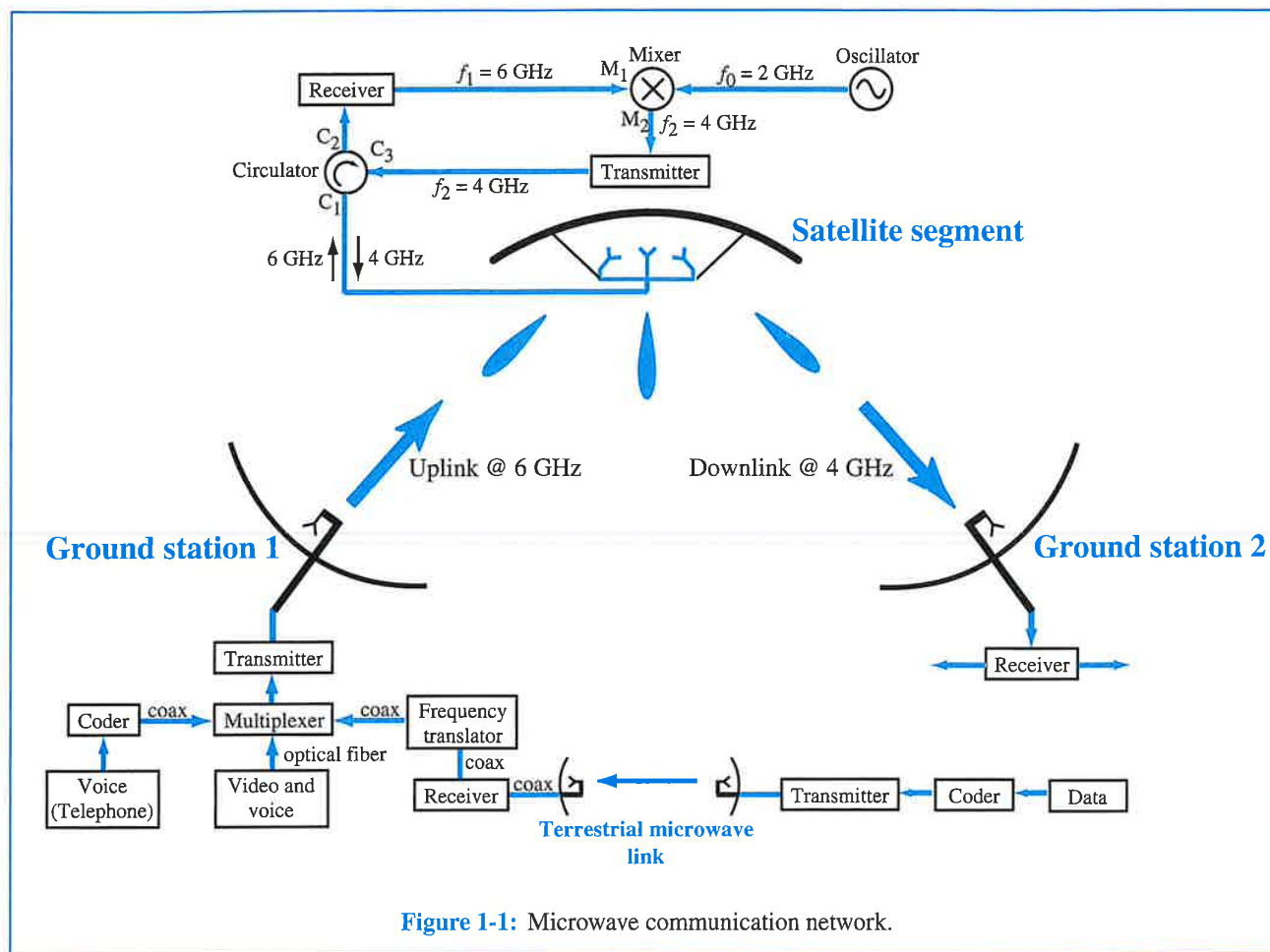
oscillator frequency. In this case $f_0 = 2$ GHz. Different frequencies are used for the uplink and downlink signals to avoid interference between them. The *circulator* connected to the satellite antenna allows the antenna to function as both a transmitter and a receiver. The circulator connects a signal received at port C_1 to the receiver at port C_2 , and it can also connect a signal at port C_3 to port C_1 for transmission by the antenna. Although in the configuration shown in Fig. 1-1 one antenna beam is used for reception and another for transmission, in reality both beams can be used for both functions.

Ground station 1 is connected to multiple terminals through various types of *transmission lines*, including *coaxial cables* and *optical fibers*. The information carried on these lines includes voice for telephone communication, video and voice for TV, and digital data. One of the terminals is connected to a terrestrial microwave link.

In this book we will examine several elements of this communication network, including wave transmission on coaxial lines, the fundamentals of optical fiber, free-space propagation between antennas, antenna design, and many other topics. Our goal is to help the reader to develop an understanding of the fundamental physical laws of electromagnetism, as well as to gain insight into their practical applications.

We begin this chapter by describing the system of units and the symbolic notation used in the book. Next, we introduce the fundamental electric and magnetic field quantities we use in electromagnetics, their relationships to each other and to electric charges and electric currents, and the basic parameters by which we characterize the electromagnetic properties of materials. The

*1 GHz = 10^9 Hz.



laws governing these relationships constitute the basic infrastructure we use in the study of electromagnetic phenomena. Then, in preparation for the material presented in Chapter 2, we provide short reviews of three topics: traveling waves, complex numbers, and phasor analysis. Although the reader most likely has encountered these topics in circuit analysis or other engineering disciplines, short reviews of the properties of traveling waves and of the convenience phasor notation

provides should prove useful in solving time-harmonic problems.

1-1 Dimensions, Units, and Notation

The *International System of Units*, abbreviated *SI* after its French name *Système Internationale*, is the standard system used in today's scientific literature for expressing

Table 1-1: Fundamental SI units.

Dimension	Unit	Symbol
Length	meter	m
Mass	kilogram	kg
Time	second	s
Electric Current	ampere	A
Temperature	kelvin	K
Amount of substance	mole	mol

the units of physical quantities. Length is a *dimension* and meter is the *unit* by which it is expressed relative to a reference standard. The SI system is based on the units for the six *fundamental dimensions* listed in Table 1-1. The units for all other dimensions are regarded as *secondary* because they are based on and can be expressed in terms of the six fundamental units. Appendix A contains a list of quantities used in this book, together with their symbols and units.

For quantities ranging in value between 10^{-18} and 10^{18} , a set of prefixes, arranged in steps of 10^3 , are commonly used to denote multiples and submultiples of units. These prefixes, all of which were derived from Greek, Latin, Spanish, and Danish terms, are listed in Table 1-2. A length of 5×10^{-9} m, for example, may be written as 5 nm.

In electromagnetics we work with scalar and vector quantities. In this book we use a medium-weight italic font for symbols (other than Greek letters) denoting scalar quantities, such as R for resistance, whereas we use a boldface roman font for symbols denoting vectors, such as $\mathbf{E}$ for the electric field vector. A vector consists of a magnitude (scalar) and a direction, with the direction usually denoted by a unit vector. For example,

$$\mathbf{E} = \hat{\mathbf{x}}E, \quad (1.1)$$

where E is the magnitude of $\mathbf{E}$ and $\hat{\mathbf{x}}$ is its direction. Unit vectors are printed in boldface with a circumflex ($\hat{}$) above the letter.

Table 1-2: Multiple and submultiple prefixes.

Prefix	Symbol	Magnitude
exa	E	10^{18}
peta	P	10^{15}
tera	T	10^{12}
giga	G	10^9
mega	M	10^6
kilo	k	10^3
milli	m	10^{-3}
micro	μ	10^{-6}
nano	n	10^{-9}
pico	p	10^{-12}
femto	f	10^{-15}
atto	a	10^{-18}

Throughout this book, we make extensive use of *phasor representation* in solving problems involving electromagnetic quantities that vary sinusoidally in time. Letters denoting phasor quantities are printed with a tilde ($\sim$) over the letter. Thus, $\tilde{\mathbf{E}}$ is the phasor electric field vector corresponding to the instantaneous electric field vector $\mathbf{E}(t)$. This notation is discussed in more detail in Section 1-6.

1-2 The Nature of Electromagnetism

Our physical universe is governed by four fundamental forces of nature:

- The *nuclear force*, which is the strongest of the four, but its range is limited to *submicroscopic* systems, such as nuclei.
- The *weak-interaction force*, whose strength is only 10^{-14} that of the nuclear force. Its primary role is in interactions involving certain radioactive elementary particles.

- The **electromagnetic force**, which exists between all charged particles. It is the dominant force in **microscopic** systems, such as atoms and molecules, and its strength is on the order of 10^{-2} that of the nuclear force.
- The **gravitational force**, which is the weakest of all four forces, having a strength on the order of 10^{-41} that of the nuclear force. However, it is the dominant force in **macroscopic** systems, such as the solar system.

Our interest in this book is with the electromagnetic force and its consequences. Even though the electromagnetic force operates at the atomic scale, its effects can be transmitted in the form of electromagnetic waves that can propagate through both free space and material media. The purpose of this section is to provide an overview of the basic **framework of electromagnetism**, which consists of certain fundamental laws governing the electric and magnetic fields induced by static and moving electric charges, respectively, the relations between the electric and magnetic fields, and how these fields interact with matter. As a precursor, however, we will take advantage of our familiarity with the gravitational force by describing some of its properties because they provide a useful analogue to those of the electromagnetic force.

1-2.1 The Gravitational Force: A Useful Analogue

According to Newton's law of gravity, the gravitational force $\mathbf{F}_{g_{21}}$ acting on mass m_2 due to a mass m_1 at a distance R_{12} from m_2 , as depicted in Fig. 1-2, is given by

$$\mathbf{F}_{g_{21}} = -\hat{\mathbf{R}}_{12} \frac{Gm_1m_2}{R_{12}^2} \quad (\text{N}), \quad (1.2)$$

where G is the universal gravitational constant, $\hat{\mathbf{R}}_{12}$ is a unit vector that points from m_1 to m_2 , and the unit for force is newton (N). The negative sign in Eq. (1.2) accounts for the fact that the gravitational force is attractive. Conversely, $\mathbf{F}_{g_{12}} = -\mathbf{F}_{g_{21}}$, where $\mathbf{F}_{g_{12}}$ is the force acting on mass m_1

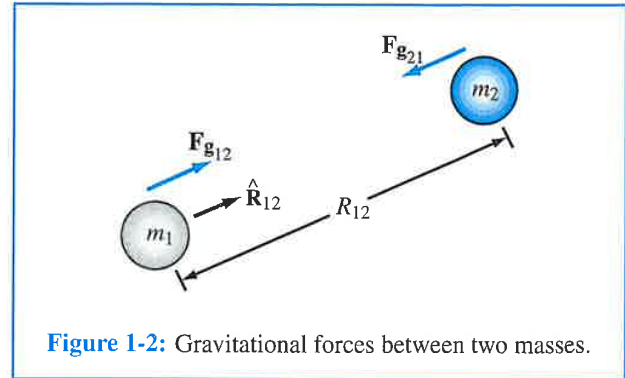


Figure 1-2: Gravitational forces between two masses.

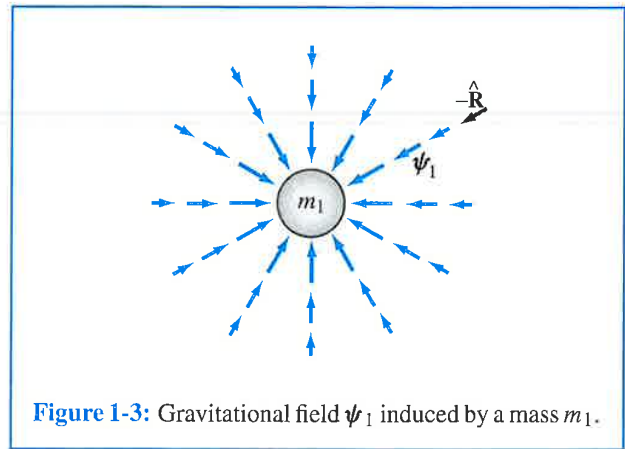


Figure 1-3: Gravitational field ψ_1 induced by a mass m_1 .

due to the gravitational pull of mass m_2 . Note that the first subscript of $\mathbf{F}_g$ denotes the mass experiencing the force and the second subscript denotes the source of the force.

The force of gravitation acts at a distance; that is, the two objects do not have to be in direct contact for each to experience the pull by the other. This phenomenon of direct action at a distance has led to the concept of **fields**. An object of mass m_1 induces a gravitational field ψ_1 (Fig. 1-3) that does not physically emanate from the object, but its influence exists at every point in space such that if another object of mass m_2 were to exist at a distance R_{12} from object m_1 then the second object m_2

would experience a force acting on it equal in strength to that given by Eq. (1.2). At a distance R from m_1 , the field ψ_1 is a vector defined as

$$\psi_1 = -\hat{\mathbf{R}} \frac{Gm_1}{R^2} \quad (\text{N/kg}), \quad (1.3)$$

where $\hat{\mathbf{R}}$ is a unit vector that points in the radial direction away from object m_1 , and therefore $-\hat{\mathbf{R}}$ points toward m_1 . The force due to ψ_1 acting on a mass m_2 at a distance $R = R_{12}$ along the direction $\hat{\mathbf{R}} = \hat{\mathbf{R}}_{12}$ is

$$\mathbf{F}_{g_{21}} = \psi_1 m_2 = -\hat{\mathbf{R}}_{12} \frac{Gm_1 m_2}{R_{12}^2}. \quad (1.4)$$

The field concept may be generalized by defining the gravitational field ψ at any point in space such that, when a test mass m is placed at that point, the force $\mathbf{F}_g$ acting on m is related to ψ by

$$\psi = \frac{\mathbf{F}_g}{m}. \quad (1.5)$$

The force $\mathbf{F}_g$ may be due to a single mass or a distribution of many masses.

1-2.2 Electric Fields

The electromagnetic force consists of an electrical force $\mathbf{F}_e$ and a magnetic force $\mathbf{F}_m$. The electrical force $\mathbf{F}_e$ is similar to the gravitational force, but with a major difference. *The source of the gravitational field is mass, and the source of the electrical field is electric charge*, and whereas both types of fields vary inversely as the square of the distance from their respective sources, electric charge may have positive or negative polarity, whereas mass does not exhibit such a property.

We know from atomic physics that all matter contains a mixture of neutrons, positively charged protons, and negatively charged electrons, with the fundamental quantity of charge being that of a single electron, usually denoted by the letter e . The unit by which electric charge is measured is the coulomb (C), named in honor of the

eighteenth-century French scientist Charles Augustin de Coulomb (1736–1806). The magnitude of e is

$$e = 1.6 \times 10^{-19} \quad (\text{C}). \quad (1.6)$$

The charge of a single electron is $q_e = -e$ and that of a proton is equal in magnitude but opposite in polarity: $q_p = e$. Coulomb's experiments demonstrated that:

- (1) *two like charges repel one another, whereas two charges of opposite polarity attract,*
- (2) *the force acts along the line joining the charges, and*
- (3) *its strength is proportional to the product of the magnitudes of the two charges and inversely proportional to the square of the distance between them.*

These properties constitute what today is called *Coulomb's law*, which can be expressed mathematically by the following equation:

$$\mathbf{F}_{e_{21}} = \hat{\mathbf{R}}_{12} \frac{q_1 q_2}{4\pi \epsilon_0 R_{12}^2} \quad (\text{N}) \quad (\text{in free space}), \quad (1.7)$$

where $\mathbf{F}_{e_{21}}$ is the electrical force acting on charge q_2 due to charge q_1 , R_{12} is the distance between the two charges, $\hat{\mathbf{R}}_{12}$ is a unit vector pointing from charge q_1 to charge q_2 (Fig. 1-4), and ϵ_0 is a universal constant called the *electrical permittivity of free space* [$\epsilon_0 = 8.854 \times 10^{-12}$ farad per meter (F/m)]. The two charges are assumed to be in *free space* (vacuum) and isolated from all other charges. The force $\mathbf{F}_{e_{12}}$ acting on charge q_1 due to charge q_2 is equal to force $\mathbf{F}_{e_{21}}$ in magnitude, but opposite in direction; $\mathbf{F}_{e_{12}} = -\mathbf{F}_{e_{21}}$.

The expression given by Eq. (1.7) for the electrical force is analogous to that given by Eq. (1.2) for the gravitational force, and we can extend the analogy further by defining

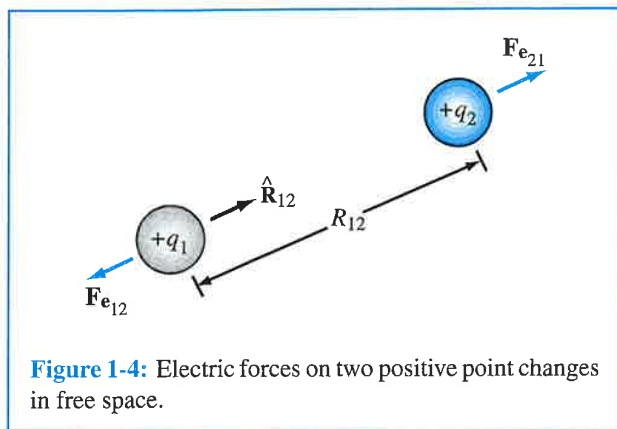


Figure 1-4: Electric forces on two positive point charges in free space.

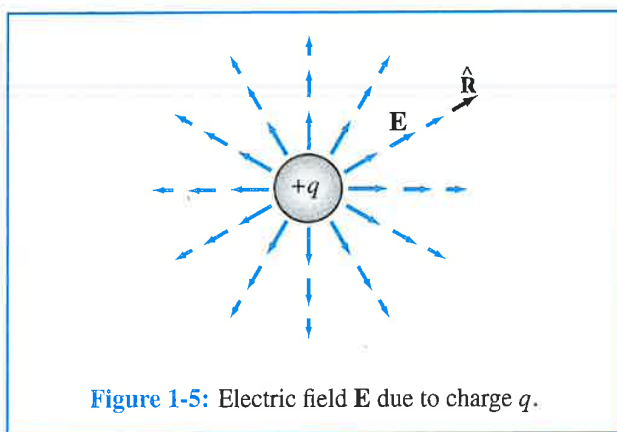


Figure 1-5: Electric field $\mathbf{E}$ due to charge q .

the existence of an *electric field intensity* $\mathbf{E}$ due to any charge q as follows:

$$\mathbf{E} = \hat{\mathbf{R}} \frac{q}{4\pi\epsilon_0 R^2} \quad (\text{V/m}) \quad (\text{in free space}), \quad (1.8)$$

where R is the distance between the charge and the observation point, and $\hat{\mathbf{R}}$ is the radial unit vector pointing away from the charge. Figure 1-5 depicts the electric-field lines due to a positive charge. For reasons that will become apparent in later chapters, the unit for $\mathbf{E}$ is volt per meter (V/m).

Electric charge exhibits two important properties. The first is the *law of conservation of electric charge*, which states that *the (net) electric charge can neither be created nor destroyed*. If a volume contains n_p protons and n_e electrons, then the total charge is

$$q = n_p e - n_e e = (n_p - n_e) e \quad (\text{C}). \quad (1.9)$$

Even if some of the protons were to combine with an equal number of electrons to produce neutrons or other elementary particles, the net charge q remains unchanged. In matter, the quantum mechanical laws governing the behavior of the protons inside the atom's nucleus and the electrons outside it do not allow them to combine.

The second important property of electric charge is the *principle of linear superposition*, which states that *the total vector electric field at a point in space due to a system of point charges is equal to the vector sum of the electric fields at that point due to the individual charges*. This seemingly simple concept will allow us in future chapters to compute the electric field due to complex distributions of charge without having to be concerned with the forces acting on each individual charge due to the fields by all of the other charges.

The expression given by Eq. (1.8) describes the field induced by an electric charge when in free space. Let us now consider what happens when we place a positive point charge in a material composed of atoms. In the absence of the point charge, the material is electrically neutral, with each atom having a positively charged nucleus surrounded by a cloud of electrons of equal but opposite polarity. Hence, at any point in the material not occupied by an atom the electric field $\mathbf{E}$ is zero. Upon placing a point charge in the material, as shown in Fig. 1-6, the atoms experience forces that cause them to become distorted. The center of symmetry of the electron cloud is altered with respect to the nucleus, with one pole of the atom becoming more positively charged and the other pole becoming more negatively charged. Such a polarized atom is called an *electric dipole*, and

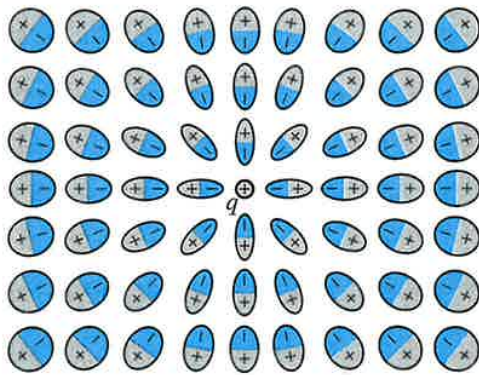


Figure 1-6: Polarization of the atoms of a dielectric material by a positive charge q .

the distortion process is called **polarization**. The degree of polarization depends on the distance between the atom and the isolated point charge, and the orientation of the dipole is such that the dipole axis connecting its two poles is directed toward the point charge, as illustrated schematically in Fig. 1-6. The net result of this polarization process is that the electric dipoles of the atoms (or molecules) tend to counteract the field due to the point charge. Consequently, the electric field at any point in the material would be different from the field that would have been induced by the point charge in the absence of the material. To extend Eq. (1.8) from the free-space case to any medium, we replace the permittivity of free space ϵ_0 with ϵ , where ϵ is now the permittivity of the material in which the electric field is measured and is therefore characteristic of that particular material. Thus,

$$\mathbf{E} = \hat{\mathbf{R}} \frac{q}{4\pi\epsilon R^2} \quad (\text{V/m}). \quad (1.10)$$

Often, ϵ is expressed in the form

$$\epsilon = \epsilon_r \epsilon_0 \quad (\text{F/m}), \quad (1.11)$$

where ϵ_r is a dimensionless quantity called the **relative permittivity** or **dielectric constant** of the material. For vacuum, $\epsilon_r = 1$; for air near Earth's surface, $\epsilon_r = 1.0006$; and for materials that we will have occasion to use in this book, their values of ϵ_r are tabulated in Appendix B.

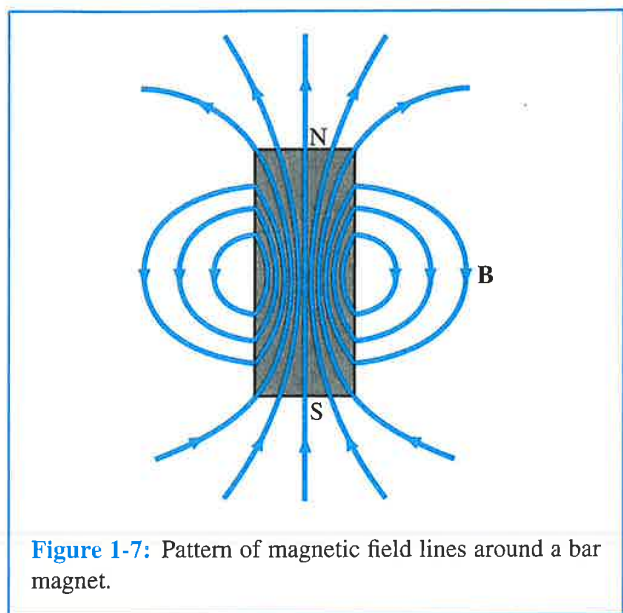
In addition to the electric field intensity $\mathbf{E}$, we will often find it convenient to also use a related quantity called the **electric flux density** $\mathbf{D}$, given by

$$\mathbf{D} = \epsilon \mathbf{E} \quad (\text{C/m}^2), \quad (1.12)$$

and its unit is coulomb per square meter (C/m^2). These two electrical quantities, $\mathbf{E}$ and $\mathbf{D}$, constitute one of two fundamental pairs of electromagnetic fields. The second pair consists of the magnetic fields discussed next.

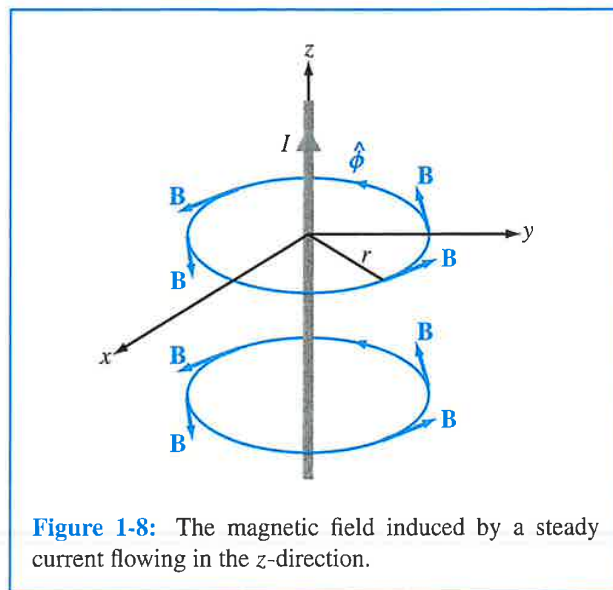
1-2.3 Magnetic Fields

As early as 800 B.C., the Greeks discovered that certain kinds of stones exhibit a force that attracts pieces of iron. These stones are now called **magnetite** (Fe_3O_4) and the phenomenon they exhibit is **magnetism**. In the thirteenth century, French scientists discovered that, when a needle was placed on the surface of a spherical natural magnet, the needle oriented itself along different directions for different locations on the magnet. By mapping the directions taken by the needle, it was determined that the magnetic force formed magnetic-field lines that encircled the sphere and appeared to pass through two points diametrically opposite each other. These points, called the **north and south poles** of the magnet, were found to exist for every magnet, regardless of its shape. The magnetic-field pattern of a bar magnet is displayed in Fig. 1-7. It was also observed that like poles of different magnets repel each other and unlike poles attract each other. This attraction–repulsion property is similar to the electric force between electric charges, except for one important difference: **electric charges can be isolated, but magnetic poles always exist in pairs**. If a permanent



magnet is cut into small pieces, no matter how small each piece is, it will always have a north and a south pole.

The magnetic lines encircling a magnet are called *magnetic-field lines* and represent the existence of a magnetic field called the *magnetic flux density* $\mathbf{B}$. A magnetic field not only exists around permanent magnets but can also be created by electric current. This connection between electricity and magnetism was discovered in 1819 by the Danish scientist Hans Oersted (1777–1851), who found that an electric current in a wire caused a compass needle placed in its vicinity to deflect and that the needle turned so that its direction was always perpendicular to the wire and to the radial line connecting the wire to the needle. From these observations, it was deduced that the current-carrying wire induced a magnetic field that formed closed circular loops around the wire, as illustrated in Fig. 1-8. Shortly after Oersted's discovery, French scientists Jean Baptiste Biot and Felix Savart developed an expression that relates the magnetic flux density $\mathbf{B}$ at a point in space to the current I in the conductor. Application



of their formulation, known today as the *Biot–Savart law*, to the situation depicted in Fig. 1-8 for a very long wire leads to the result that the *magnetic flux density* $\mathbf{B}$ induced by a constant current I flowing in the z -direction is given by

$$\mathbf{B} = \hat{\phi} \frac{\mu_0 I}{2\pi r} \quad (\text{T}), \quad (1.13)$$

where r is the radial distance from the current and $\hat{\phi}$ is an azimuthal unit vector denoting the fact that the magnetic field direction is tangential to the circle surrounding the current, as shown in Fig. 1-8. The magnetic field is measured in tesla (T), named in honor of Nikola Tesla (1856–1943), a Croatian-American electrical engineer whose work on transformers made it possible to transport electricity over long wires without too much loss. The quantity μ_0 is called the *magnetic permeability of free space* [$\mu_0 = 4\pi \times 10^{-7}$ henry per meter (H/m)], and it is analogous to the electric permittivity ϵ_0 . In fact, as we will see in Chapter 2, the product of ϵ_0 and μ_0 specifies c ,

the *velocity of light in free space*, as follows:

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}} = 3 \times 10^8 \quad (\text{m/s}). \quad (1.14)$$

The majority of natural materials are *nonmagnetic*, meaning that they exhibit a magnetic permeability $\mu = \mu_0$. For ferromagnetic materials, such as iron and nickel, μ can be much larger than μ_0 . The magnetic permeability μ accounts for *magnetization* properties of a material. In analogy with Eq. (1.11), μ of a particular material can be defined as

$$\mu = \mu_r \mu_0 \quad (\text{H/m}), \quad (1.15)$$

where μ_r is a dimensionless quantity called the *relative magnetic permeability* of the material. The values of μ_r for commonly used ferromagnetic materials are given in Appendix B.

We stated earlier that **E** and **D** constitute one of two pairs of electromagnetic field quantities. The second pair is **B** and the *magnetic field intensity* **H**, which are related to each other through μ :

$$\mathbf{B} = \mu \mathbf{H}. \quad (1.16)$$

1-2.4 Static and Dynamic Fields

Because the electric field **E** is governed by the charge q and the magnetic field **H** is governed by $I = dq/dt$, and since q and dq/dt are independent variables, the induced electric and magnetic fields are independent of one another as long as I remains constant. To demonstrate the validity of this statement, consider for example a small section of a beam of charged particles that are moving at a constant velocity. The moving charges constitute a d-c current. The electric field due to that section of the beam is determined by the total charge q contained in that

section. The magnetic field does not depend on q , but rather on the rate of charge (current) flowing through that section. Few charges moving very fast can constitute the same current as many charges moving slowly. In these two cases the induced magnetic field will be the same because the current I is the same, but the induced electric field will be quite different because the numbers of charges are not the same.

Electrostatics and *magnetostatics*, corresponding to stationary charges and steady currents, respectively, are special cases of electromagnetics. They represent two *independent* branches, so characterized because the induced electric and magnetic fields are uncoupled to each other. *Dynamics*, the third and more general branch of electromagnetics, involves *time-varying fields* induced by time-varying sources, that is, currents and charge densities. If the current associated with the beam of moving charged particles varies with time, then the amount of charge present in a given section of the beam also varies with time, and vice versa. As we will see in Chapter 6, the electric and magnetic fields become coupled to each other in that case. In fact, *a time-varying electric field will generate a time-varying magnetic field, and vice versa*. Table 1-3 provides a summary of the three branches of electromagnetics.

The electric and magnetic properties of materials are characterized by the two parameters ϵ and μ , respectively. A third fundamental parameter is also needed, the *conductivity* of a material σ , which is measured in siemens per meter (S/m). The conductivity characterizes the ease with which charges (electrons) can move freely in a material. If $\sigma = 0$, the charges do not move more than atomic distances and the material is said to be a *perfect dielectric*, and if $\sigma = \infty$, the charges can move very freely throughout the material, which is then called a *perfect conductor*. The material parameters ϵ , μ , and σ are often referred to as the *constitutive parameters* of a material (Table 1-4). A medium is said to be *homogeneous* if its constitutive parameters are constant throughout the medium.

Table 1-3: The three branches of electromagnetics.

Branch	Condition	Field Quantities (Units)
Electrostatics	Stationary charges ($\partial q/\partial t = 0$)	Electric field intensity $\mathbf{E}$ (V/m) Electric flux density $\mathbf{D}$ (C/m ²) $\mathbf{D} = \epsilon \mathbf{E}$
Magnetostatics	Steady currents ($\partial I/\partial t = 0$)	Magnetic flux density $\mathbf{B}$ (T) Magnetic field intensity $\mathbf{H}$ (A/m) $\mathbf{B} = \mu \mathbf{H}$
Dynamics (Time-varying fields)	Time-varying currents ($\partial I/\partial t \neq 0$)	$\mathbf{E}$, $\mathbf{D}$, $\mathbf{B}$, and $\mathbf{H}$ ($\mathbf{E}$, $\mathbf{D}$) coupled to ($\mathbf{B}$, $\mathbf{H}$)

Table 1-4: Constitutive parameters of materials.

Parameter	Units	Free-space Value
Electrical permittivity ϵ	F/m	$\epsilon_0 = 8.854 \times 10^{-12}$ (F/m) $\simeq \frac{1}{36\pi} \times 10^{-9}$ (F/m)
Magnetic permeability μ	H/m	$\mu_0 = 4\pi \times 10^{-7}$ (H/m)
Conductivity σ	S/m	0

REVIEW QUESTIONS

Q1.1 What are the four fundamental forces of nature and what are their relative strengths?

Q1.2 What is Coulomb's law? State its properties.

Q1.3 What are the two important properties of electric charge?

Q1.4 What do the electrical permittivity and magnetic permeability of a material account for?

Q1.5 What are the three branches and associated conditions of electromagnetics?

1-3 Traveling Waves

Waves are a natural consequence of many physical processes: waves and ripples on oceans and lakes; sound waves that travel through air; mechanical waves on stretched strings; electromagnetic waves that constitute light; earthquake waves; and many others. All these various types of waves exhibit a number of common properties, including the following:

- *Moving waves carry energy from one point to another.*
- *Waves have velocity;* it takes time for a wave to travel from one point to another. In vacuum, light waves

travel at a speed of 3×10^8 m/s and sound waves in air travel at a speed approximately a million times slower, specifically 330 m/s.

- *Some waves exhibit a property called linearity.* Waves that do not affect the passage of other waves are called **linear** because they pass right through each other, and the total of two linear waves is simply the sum of the two waves as they would exist separately. Electromagnetic waves are linear, as are sound waves. When two people speak to one another, their sound waves do not reflect from one another, but simply pass through independently of each other. Water waves are approximately linear; the expanding circles of ripples caused by two pebbles thrown into two locations on a lake surface do not affect each other. Although the interaction of the two circles may exhibit a complicated pattern, it is simply the linear superposition of two independent expanding circles.

Waves are of two types: **transient waves** caused by a short-duration disturbance and **continuous harmonic waves** generated by an oscillating source. We will encounter both types of waves in this book, but most of our discussion will deal with the propagation of continuous waves that vary sinusoidally with time.

An essential feature of a propagating wave is that it is a self-sustaining disturbance of the medium through which it travels. If this disturbance varies as a function of one space variable, such as the vertical displacement of the string shown in Fig. 1-9, we call the wave a **one-dimensional wave**. The vertical displacement varies with time and with the location along the length of the string. Even though the string rises up into a second dimension, the wave is only one-dimensional because the disturbance varies with only one space variable. A **two-dimensional wave** propagates out across a surface, like the ripples on a pond [Fig. 1-10(a)], and its disturbance can be described by two space variables. And by extension, a **three-dimensional wave** propagates

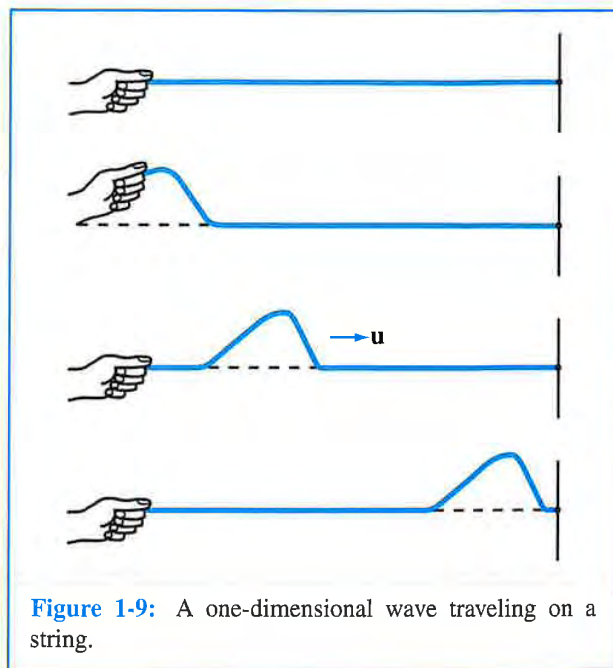


Figure 1-9: A one-dimensional wave traveling on a string.

through a volume and its disturbance may be a function of all three space variables. Three-dimensional waves may take on many different shapes; they include **plane waves**, **cylindrical waves**, and **spherical waves**. A plane wave is characterized by a disturbance that at a given point in time has uniform properties across an infinite plane perpendicular to the direction of wave propagation [Fig. 1-10(b)] and, similarly, for cylindrical and spherical waves the disturbances are uniform across cylindrical and spherical surfaces, as shown in Figs. 1-10(b) and (c).

In the material that follows, we will examine some of the basic properties of waves by developing mathematical formulations that describe their functional dependence on time and space variables. To keep the presentation simple, we will limit our present discussion to sinusoidally varying waves whose disturbances are functions of only one space variable, and we will defer discussion of more complicated waves to later chapters.

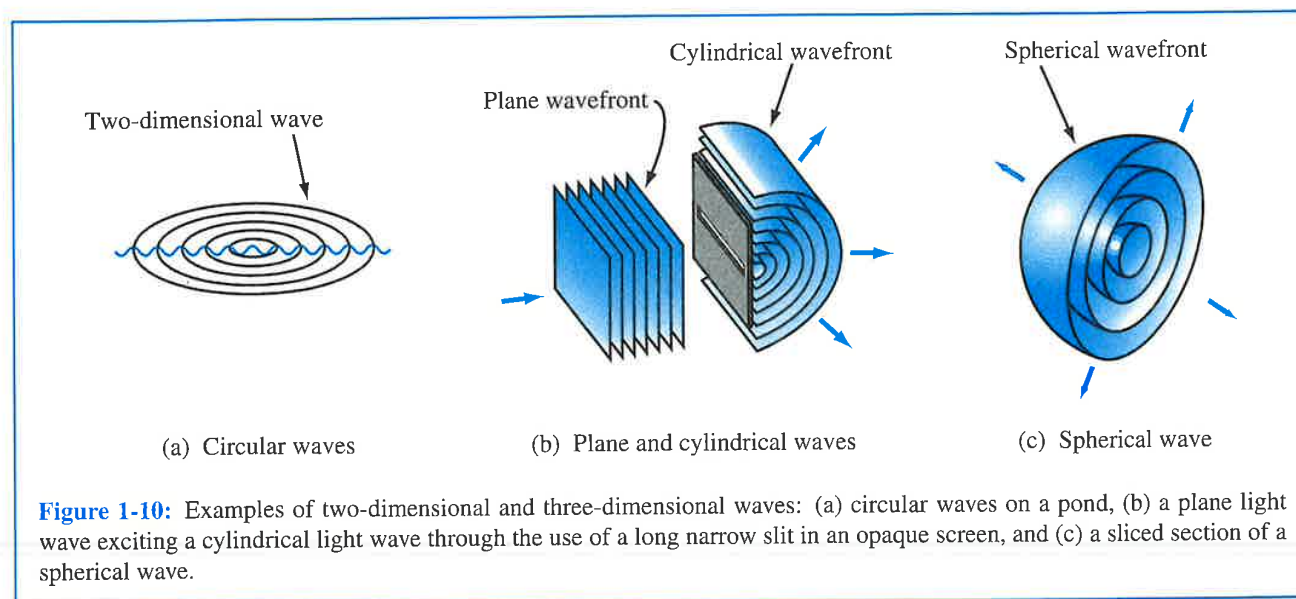


Figure 1-10: Examples of two-dimensional and three-dimensional waves: (a) circular waves on a pond, (b) a plane light wave exciting a cylindrical light wave through the use of a long narrow slit in an opaque screen, and (c) a sliced section of a spherical wave.

1-3.1 Sinusoidal Wave in a Lossless Medium

Regardless of the mechanism responsible for generating them, all waves can be described mathematically in common terms. By way of an example, let us consider a wave traveling on a lake surface. *A medium is said to be lossless if it does not attenuate the amplitude of the wave traveling within it or on its surface.* Let us assume for the time being that frictional forces can be ignored, thereby allowing a wave generated on the water surface to travel indefinitely with no loss in energy. If y denotes the height of the water surface relative to the mean height (undisturbed condition) and x denotes the distance of wave travel, the functional dependence of y on time t and the spatial coordinate x has the general form

$$y(x, t) = A \cos \left(\frac{2\pi t}{T} - \frac{2\pi x}{\lambda} + \phi_0 \right) \text{ (m)}, \quad (1.17)$$

where A is the **amplitude** of the wave, T is its **time period**, λ is its **spatial wavelength**, and ϕ_0 is a **reference phase**.

The quantity $y(x, t)$ can also be expressed in the form

$$y(x, t) = A \cos \phi(x, t), \quad (1.18)$$

where

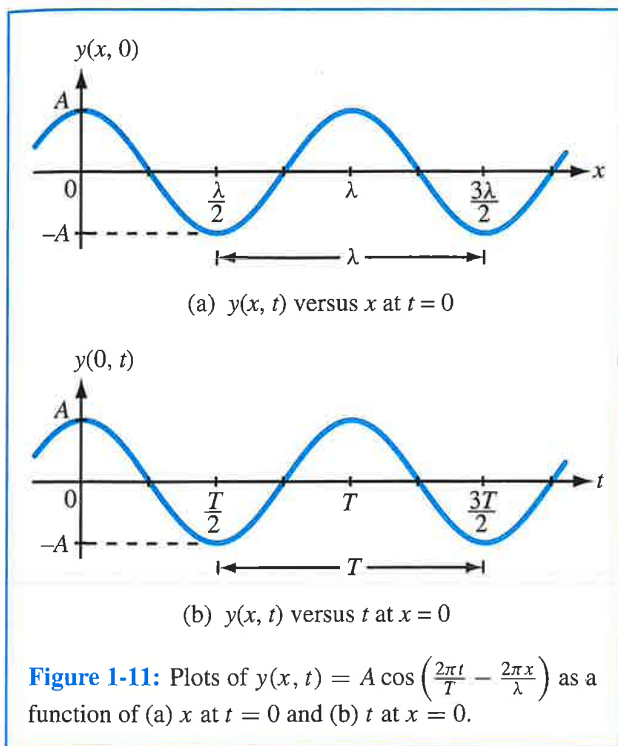
$$\phi(x, t) = \left(\frac{2\pi t}{T} - \frac{2\pi x}{\lambda} + \phi_0 \right) \text{ (rad)}. \quad (1.19)$$

The angle $\phi(x, t)$ is called the **phase** of the wave, and it should not be confused with the reference phase ϕ_0 , which is constant with respect to both time and space. Phase is measured by the same units as angles, that is, radians (rad) or degrees, with 2π radians = 360° .

Let us first analyze the simple case when $\phi_0 = 0$:

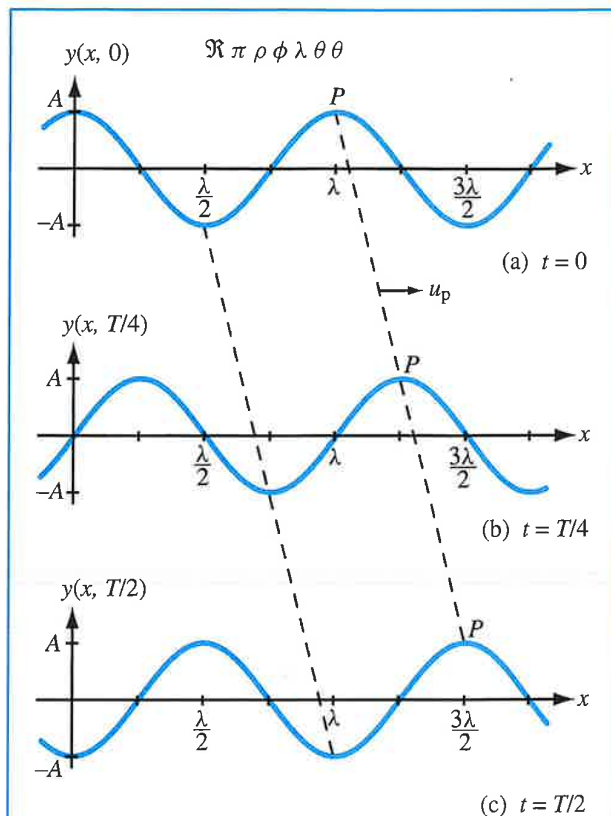
$$y(x, t) = A \cos \left(\frac{2\pi t}{T} - \frac{2\pi x}{\lambda} \right) \text{ (m)}. \quad (1.20)$$

The plots in Fig. 1-11 show the variation of $y(x, t)$ with x at $t = 0$ and with t at $x = 0$. The wave pattern repeats itself at a spatial period λ along x and at a temporal period T along t .



If we take time snapshots of the water surface, the height profile $y(x)$ would exhibit the sinusoidal patterns shown in Fig. 1-12. For each plot, corresponding to a specific value of t , the spacing between peaks is equal to the wavelength λ , but the patterns are shifted relative to one another because they correspond to different observation times. Because the pattern advances along the $+x$ -direction at progressively increasing values of t , the height profile behaves like a wave traveling in that direction. If we choose any height level, such as the peak P , and follow it in time, we can measure the **phase velocity** of the wave. The peak corresponds to when the phase $\phi(x, t)$ of the wave is equal to zero or multiples of 2π radians. Thus,

$$\phi(x, t) = \frac{2\pi t}{T} - \frac{2\pi x}{\lambda} = 2n\pi, \quad n = 0, 1, 2, \dots \quad (1.21)$$



Had we chosen any other fixed height of the wave, say y_0 , and monitored its movement as a function of t and x , this again is equivalent to setting the phase $\phi(x, t)$ constant such that

$$y(x, t) = y_0 = A \cos\left(\frac{2\pi t}{T} - \frac{2\pi x}{\lambda}\right), \quad (1.22)$$

or

$$\frac{2\pi t}{T} - \frac{2\pi x}{\lambda} = \cos^{-1}\left(\frac{y_0}{A}\right) = \text{constant}. \quad (1.23)$$

The apparent velocity of that fixed height is obtained by taking the time derivative of Eq. (1.23),

$$\frac{2\pi}{T} - \frac{2\pi}{\lambda} \frac{dx}{dt} = 0, \quad (1.24)$$

which gives the **phase velocity** u_p as

$$u_p = \frac{dx}{dt} = \frac{\lambda}{T} \quad (\text{m/s}). \quad (1.25)$$

The phase velocity, also called the **propagation velocity**, is **the velocity of the wave pattern** as it moves across the water surface. The water itself mostly moves up and down; when the wave moves from one point to another, the water does not move physically along with it.

The direction of wave propagation is easily determined by inspecting the signs of the t and x terms in the expression for the phase $\phi(x, t)$ given by Eq. (1.19): *if one of the signs is positive and the other is negative, then the wave is traveling in the positive x -direction, and if both signs are positive or both are negative, then the wave is traveling in the negative x -direction.* The constant phase reference ϕ_0 has no influence on either the speed or the direction of wave propagation.

The **frequency** of a sinusoidal wave, f , is the reciprocal of its time period T :

$$f = \frac{1}{T} \quad (\text{Hz}). \quad (1.26)$$

Combining the preceding two equations gives the relation

$$u_p = f\lambda \quad (\text{m/s}). \quad (1.27)$$

The wave frequency f , which is measured in cycles per second, has been assigned the unit (Hz) (pronounced

“hertz”), named in honor of the German physicist Heinrich Hertz (1857–1894), who pioneered the development of radio waves.

Using Eq. (1.26), Eq. (1.20) can be rewritten in the shortened form as

$$\begin{aligned} y(x, t) &= A \cos\left(2\pi f t - \frac{2\pi}{\lambda} x\right) \\ &= A \cos(\omega t - \beta x), \end{aligned} \quad (1.28)$$

where ω is the **angular velocity** of the wave and β is its **phase constant** (or **wavenumber**), defined as

$$\omega = 2\pi f \quad (\text{rad/s}), \quad (1.29a)$$

$$\beta = \frac{2\pi}{\lambda} \quad (\text{rad/m}). \quad (1.29b)$$

In terms of these two quantities,

$$u_p = f\lambda = \frac{\omega}{\beta}. \quad (1.30)$$

So far, we have examined the behavior of a wave traveling in the $+x$ -direction. To describe a wave traveling in the $-x$ -direction, we reverse the sign of x in Eq. (1.28):

$$y(x, t) = A \cos(\omega t + \beta x). \quad (1.31)$$

We now examine the role of the phase reference ϕ_0 given previously in Eq. (1.17). If ϕ_0 is not zero, then Eq. (1.28) should be written as

$$y(x, t) = A \cos(\omega t - \beta x + \phi_0). \quad (1.32)$$

A plot of $y(x, t)$ as a function of x at a specified t or as a function of t at a specified x will be shifted in space or time, respectively, relative to a plot with $\phi_0 = 0$ by an amount ϕ_0 . This is illustrated by the plots shown in Fig. 1-13. We observe that when ϕ_0 is positive, $y(t)$ reaches its peak value, or any other specified value, sooner than when $\phi_0 = 0$. Thus, the wave with $\phi_0 = \pi/4$ is said to **lead** the wave with $\phi_0 = 0$ by a **phase lead** of $\pi/4$; and similarly, the wave with $\phi_0 = -\pi/4$ is said to **lag** the wave with $\phi_0 = 0$ by a **phase lag** of $\pi/4$. A wave function with a negative ϕ_0 takes longer to

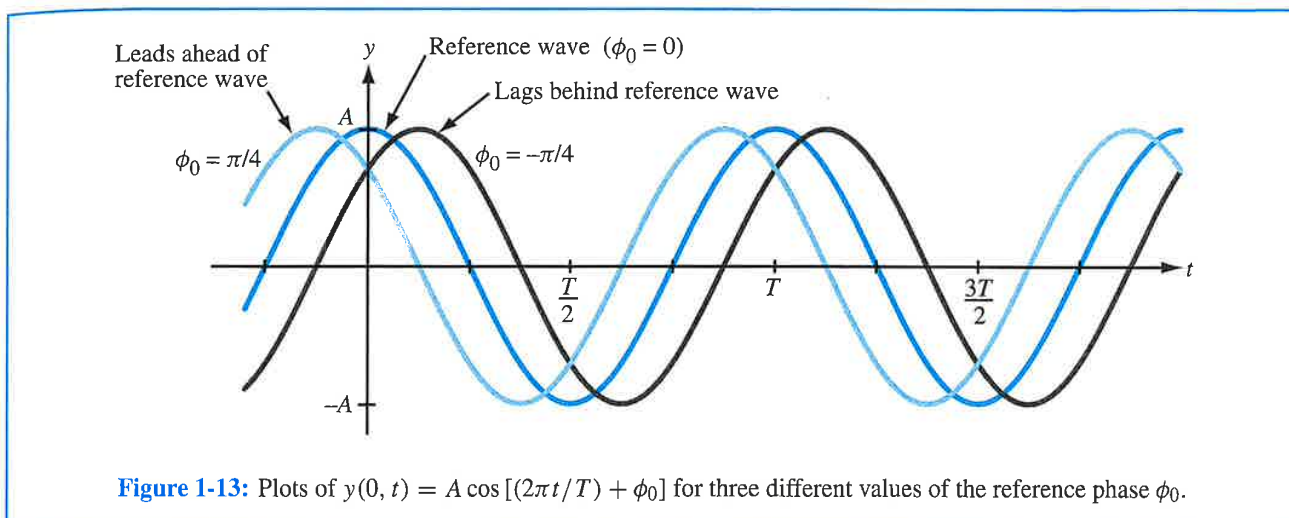


Figure 1-13: Plots of $y(0, t) = A \cos[(2\pi t/T) + \phi_0]$ for three different values of the reference phase ϕ_0 .

reach a given value of $y(t)$ than the zero-phase reference function. When its value is positive, ϕ_0 signifies a phase lead in time, and when it is negative, it signifies a phase lag.

1-3.2 Sinusoidal Wave in a Lossy Medium

If a wave is traveling in the x -direction in a *lossy medium*, its amplitude will decrease as $e^{-\alpha x}$. This factor is called the *attenuation factor*, and α is called the *attenuation constant* of the medium and its unit is neper per meter (Np/m). Thus, in general,

$$y(x, t) = Ae^{-\alpha x} \cos(\omega t - \beta x + \phi_0). \quad (1.33)$$

The wave amplitude is now $Ae^{-\alpha x}$, and not just A . Figure 1-14 shows a plot of $y(x, t)$ as a function of x at $t = 0$ for $A = 10$ m, $\lambda = 2$ m, $\alpha = 0.2$ Np/m, and $\phi_0 = 0$. Note that the envelope of the wave pattern decreases as $e^{-\alpha x}$.

The real unit of α is (1/m); the neper (Np) part is a dimensionless, artificial adjective traditionally used as a reminder that the unit (Np/m) refers to the attenuation constant of the medium, α . A similar practice is applied

to the phase constant β by assigning it the unit (rad/m) instead of just (1/m).

REVIEW QUESTIONS

Q1.6 How can you tell if a wave is traveling in the positive x -direction or the negative x -direction?

Q1.7 How does the envelope of the wave pattern vary with distance in (a) a lossless medium and (b) a lossy medium?

Q1.8 Why does a negative value of ϕ_0 signify a phase lag?

Example 1-1 Sound Wave in Water

An acoustic wave traveling in the x -direction in a fluid (liquid or gas) is characterized by a differential pressure $p(x, t)$. The unit for pressure is newton per square meter (N/m^2). Find an expression for $p(x, t)$ for a sinusoidal sound wave traveling in the positive x -direction in water,

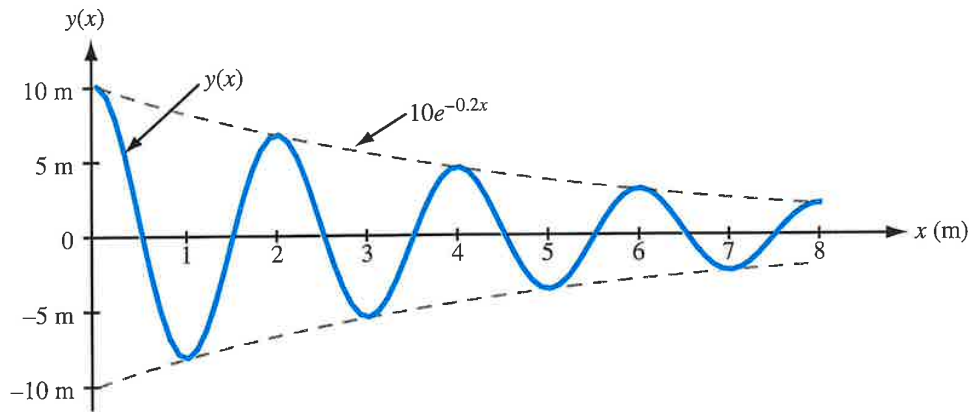


Figure 1-14: Plot of $y(x) = (10e^{-0.2x} \cos \pi x)$ meters. Note that the envelope is bounded between the curve given by $10e^{-0.2x}$ and its mirror image.

given that the wave frequency is 1 kHz, the velocity of sound in water is 1.5 km/s, the wave amplitude is 10 N/m^2 , and $p(x, t)$ was observed to be at its maximum value at $t = 0$ and $x = 0.25 \text{ m}$. Treat water as a lossless medium.

Solution: According to the general form given by Eq. (1.17) for a wave traveling in the positive x -direction,

$$p(x, t) = A \cos\left(\frac{2\pi}{T}t - \frac{2\pi}{\lambda}x + \phi_0\right) \quad (\text{N/m}^2).$$

The amplitude $A = 10 \text{ N/m}^2$, $T = 1/f = 10^{-3} \text{ s}$, and from $u_p = f\lambda$,

$$\lambda = \frac{u_p}{f} = \frac{1.5 \times 10^3}{10^3} = 1.5 \text{ m}.$$

Hence,

$$p(x, t) = 10 \cos\left(2\pi \times 10^3 t - \frac{4\pi}{3}x + \phi_0\right) \quad (\text{N/m}^2).$$

Since at $t = 0$ and $x = 0.25 \text{ m}$, $p(0.25, 0) = 10 \text{ N/m}^2$, we have

$$\begin{aligned} 10 &= 10 \cos\left(\frac{-4\pi}{3} 0.25 + \phi_0\right) \\ &= 10 \cos\left(\frac{-\pi}{3} + \phi_0\right), \end{aligned}$$

which yields the result $(\phi_0 - \pi/3) = \cos^{-1}(1)$, or $\phi_0 = \pi/3$. Hence,

$$p(x, t) = 10 \cos\left(2\pi \times 10^3 t - \frac{4\pi}{3}x + \frac{\pi}{3}\right) \quad (\text{N/m}^2).$$

Example 1-2 Power Loss

A laser beam of light propagating through the atmosphere is characterized by an electric field intensity given by

$$\begin{aligned} E(x, t) &= \\ 150e^{-0.03x} \cos(3 \times 10^{15}t - 10^7x) \quad (\text{V/m}), \end{aligned}$$

where x is the distance from the source in meters. The attenuation is due to absorption by atmospheric gases. Determine (a) the direction of wave travel, (b) the wave velocity, and (c) the wave amplitude at a distance of 200 m.

Solution: (a) Since the coefficients of t and x in the argument of the cosine function have opposite signs, the wave must be traveling in the $+x$ -direction.

(b)

$$u_p = \frac{\omega}{\beta} = \frac{3 \times 10^{15}}{10^7} = 3 \times 10^8 \text{ m/s,}$$

which is equal to c , the velocity of light in free space.

(c) At $x = 200$ m, the amplitude of $E(x, t)$ is

$$150e^{-0.03 \times 200} = 0.37 \quad (\text{V/m}). \quad \blacksquare$$

EXERCISE 1.1 The electric field of a traveling electromagnetic wave is given by

$$E(z, t) = 10 \cos(\pi \times 10^7 t + \pi z/15 + \pi/6) \quad (\text{V/m}).$$

Determine (a) the direction of wave propagation, (b) the wave frequency f , (c) its wavelength λ , and (d) its phase velocity u_p .

Ans. (a) $-z$ -direction, (b) $f = 5$ MHz, (c) $\lambda = 30$ m, (d) $u_p = 1.5 \times 10^8$ m/s.

EXERCISE 1.2 An electromagnetic wave is propagating in the z -direction in a lossy medium with attenuation constant $\alpha = 0.5$ Np/m. If the wave's electric-field amplitude is 100 V/m at $z = 0$, how far can the wave travel before its amplitude will have been reduced to (a) 10 V/m, (b) 1 V/m, (c) 1 μ V/m?

Ans. (a) 4.6 m, (b) 9.2 m, (c) 37 m.

1-4 The Electromagnetic Spectrum

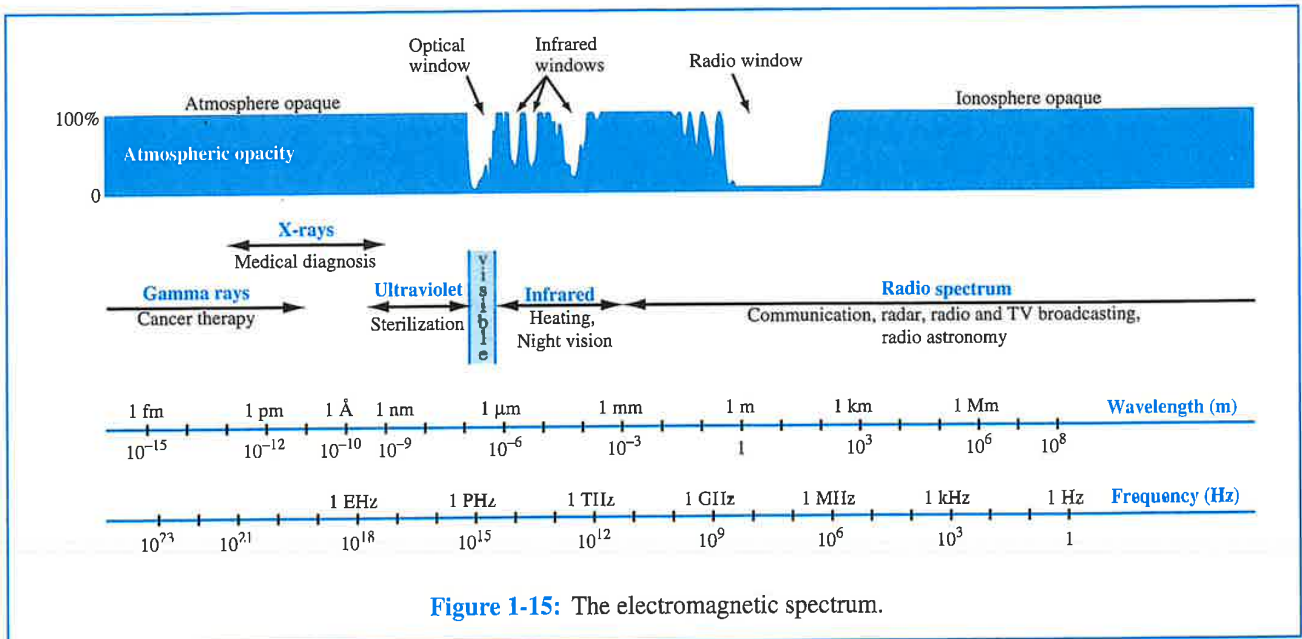
Visible light belongs to a family of waves called the **electromagnetic spectrum** (Fig. 1-15). Other members of this family include gamma rays, x rays, infrared waves, and radio waves. Generically, they all are called electromagnetic (EM) waves because they share the following fundamental properties:

- An EM wave consists of electric and magnetic field intensities that oscillate at the same frequency f .
- The phase velocity of an EM wave propagating in vacuum is a universal constant given by the velocity of light c , defined earlier by Eq. (1.14).
- In vacuum, the wavelength λ of an EM wave is related to its oscillation frequency f by

$$\lambda = \frac{c}{f}. \quad (1.34)$$

Whereas all EM waves share these properties, each is distinguished by its own wavelength λ , or equivalently by its own oscillation frequency f .

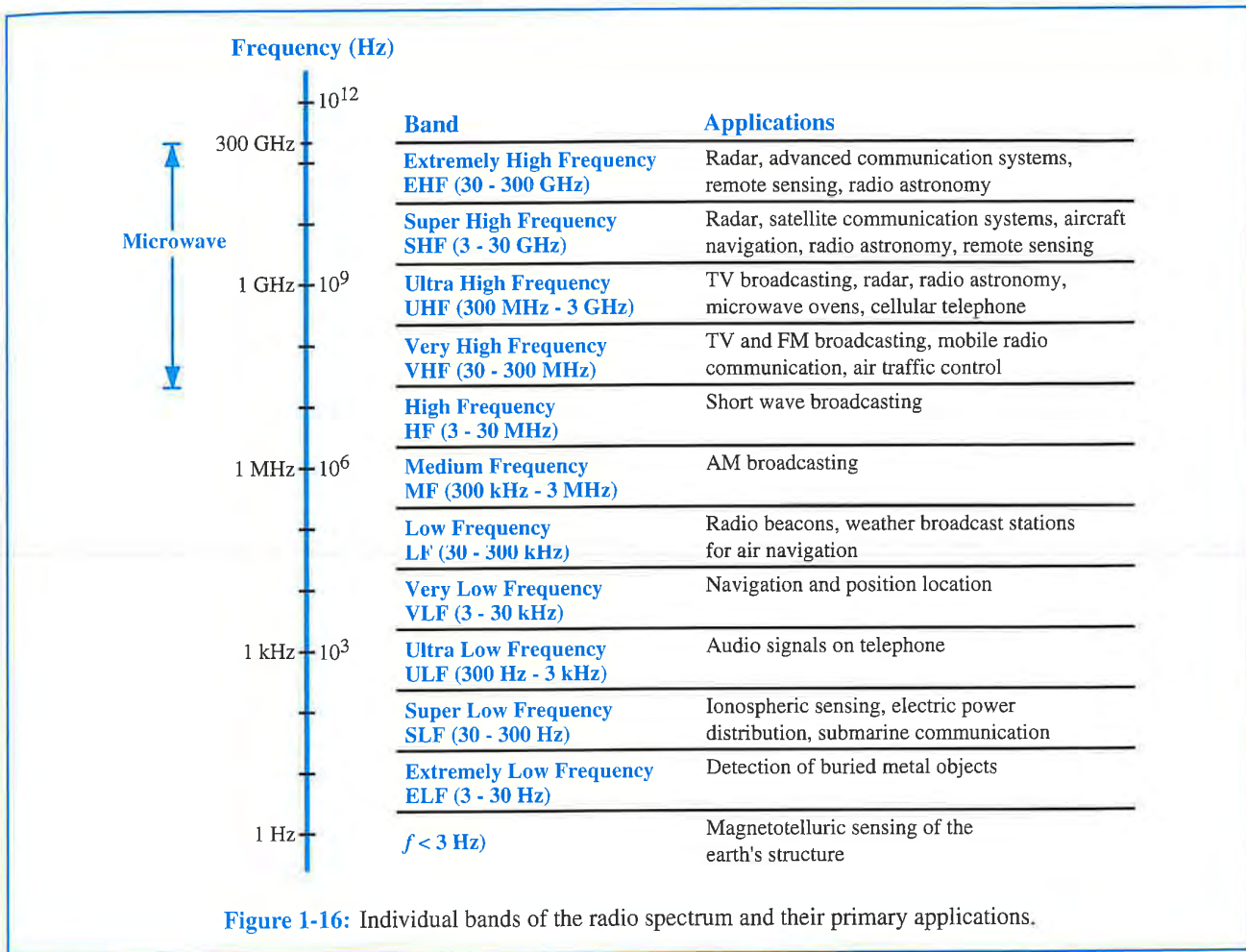
The visible part of the EM spectrum shown in Fig. 1-15 covers a very narrow wavelength range extending between $\lambda = 0.4 \mu\text{m}$ (violet) and $\lambda = 0.7 \mu\text{m}$ (red). As we move progressively toward shorter wavelengths, we encounter the ultraviolet, x-ray, and gamma-ray bands, each so named because of historical reasons associated with the discovery of waves with those wavelengths. On the other side of the visible spectrum lie the infrared band and then the radio region. Because of the link between λ and f given by Eq. (1.34), each of these spectral ranges may be specified in terms of its wavelength range or alternatively in terms of its frequency range. In practice, however, a wave is specified in terms of its wavelength λ if $\lambda < 1$ mm, which encompasses all parts of the EM spectrum except for the radio region, and the wave is specified in terms of its frequency f if $\lambda > 1$ mm (i.e., in the radio region). A wavelength of 1 mm corresponds to a frequency of 3×10^{11} Hz = 300 GHz in free space.



The radio spectrum consists of several individual bands, as shown in the chart of Fig. 1-16. Each band covers one decade of the radio spectrum and has a letter designation based on a nomenclature defined by the International Telecommunication Union. Different frequencies have different applications because they are excited by different mechanisms, and the properties of an EM wave propagating in a material may vary considerably from one band to another. The extremely low frequency (ELF) band from 3 to 30 Hz is used primarily for the detection of buried metal objects. Lower frequencies down to 0.1 Hz are used in magnetotelluric sensing of the structure of the earth, and frequencies in the range from 1 Hz to 1 kHz sometimes are used for communications with submerged submarines and for certain kinds of sensing of Earth's ionosphere. The very low frequency (VLF) region from 3 to 30 kHz is used both for submarine communications and for position location by the Omega navigation system. The low-frequency

(LF) band, from 30 to 300 kHz, is used for some forms of communication and for the Loran C position-location system. Some radio beacons and weather broadcast stations used in air navigation operate at frequencies in the higher end of the LF band. The medium-frequency (MF) region from 300 kHz to 3 MHz contains the standard AM broadcast band from 0.5 to 1.5 MHz.

Long-distance communications and short-wave broadcasting over long distances use frequencies in the high-frequency (HF) band from 3 to 30 MHz because waves in this band are strongly affected by reflections by the ionosphere and least affected by absorption in the ionosphere. The next frequency region, the very high frequency (VHF) band from 30 to 300 MHz, is used primarily for television and FM broadcasting over line-of-sight distances and also for communicating with aircraft and other vehicles. Some early radio-astronomy research was also conducted in this range. The ultrahigh frequency (UHF) region from 300 MHz to 3 GHz is



extensively populated with radars, although part of this band also is used for television broadcasting and mobile communications with aircraft and surface vehicles. The radars in this region of the spectrum are normally used for aircraft detection and tracking. Some parts of this region have been reserved for radio astronomical observation.

Many point-to-point radio communication systems and various kinds of ground-based radars and ship radars

operate at frequencies in the superhigh frequency (SHF) range from 3 to 30 GHz. Some aircraft navigation systems operate in this range as well.

Most of the extremely high frequency (EHF) band from 30 to 300 GHz is used less extensively, primarily because the technology is not as well developed and because of excessive absorption by the atmosphere in some parts of this band. Some advanced communication

systems are being developed for operation at frequencies in the “atmospheric windows,” where atmospheric absorption is not a serious problem, as are automobile collision-avoidance radars and some military imaging radar systems. These atmospheric windows include the ranges from 30 to 35 GHz, 70 to 75 GHz, 90 to 95 GHz, and 135 to 145 GHz.

Although no precise definition exists for the extent of the **microwave band**, it is conventionally regarded to cover the full ranges of the UHF, SHF, and EHF bands, with the EHF band sometimes referred to as the **millimeter-wave band**, because the wavelength range covered by this band extends from 1 mm (300 GHz) to 1 cm (30 GHz).

REVIEW QUESTIONS

Q1.9 What are the three fundamental properties of EM waves?

Q1.10 What is the range of frequencies covered by the microwave band?

Q1.11 What is the wavelength range of the visible spectrum? What are some of the applications of the infrared band?

1-5 Review of Complex Numbers

A **complex number** z is written in the form

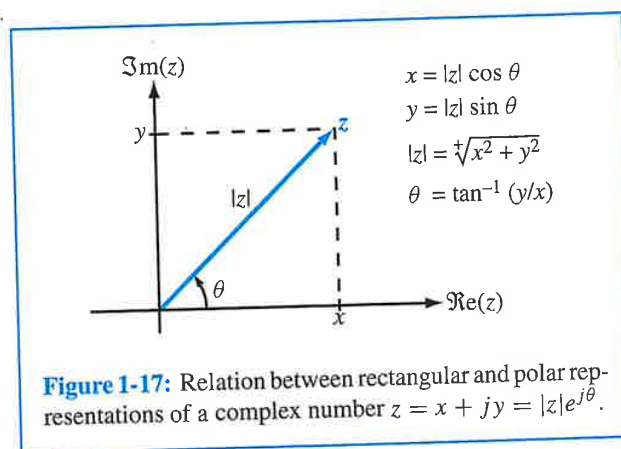
$$z = x + jy, \quad (1.35)$$

where x and y are the **real** ($\Re$) and **imaginary** ($\Im$) parts of z , respectively, and $j = \sqrt{-1}$. That is,

$$x = \Re(z), \quad y = \Im(z). \quad (1.36)$$

Alternatively, z may be written in **polar form** as

$$z = |z|e^{j\theta} = |z|\angle\theta \quad (1.37)$$



where $|z|$ is the magnitude of z , θ is its phase angle, and the form $\angle\theta$ is a useful shorthand representation commonly used in numerical calculations. Applying **Euler's identity**,

$$e^{j\theta} = \cos \theta + j \sin \theta, \quad (1.38)$$

we can convert z from polar form, as in Eq. (1.37), into rectangular form, as in Eq. (1.35),

$$z = |z|e^{j\theta} = |z| \cos \theta + j|z| \sin \theta, \quad (1.39)$$

which leads to the relations

$$x = |z| \cos \theta, \quad y = |z| \sin \theta, \quad (1.40)$$

$$|z| = \sqrt{x^2 + y^2}, \quad \theta = \tan^{-1}(y/x). \quad (1.41)$$

The two forms are illustrated graphically in Fig. 1-17. When using Eq. (1.41), care should be taken to ensure that θ is in the proper quadrant. Also note that, since $|z|$ is a positive quantity, only the positive root in Eq. (1.41) is applicable. This is denoted by the $+$ sign above the square-root sign.

The **complex conjugate** of z , denoted with a star superscript (or asterisk), is obtained by replacing j (wherever it appears) with $-j$, so that

$$z^* = (x + jy)^* = x - jy = |z|e^{-j\theta} = |z|\angle -\theta. \quad (1.42)$$

The magnitude $|z|$ is equal to the positive square root of the product of z and its complex conjugate:

$$|z| = \sqrt[4]{z z^*}. \quad (1.43)$$

We now highlight some of the properties of complex algebra that we will likely encounter in future chapters.

Equality: If two complex numbers z_1 and z_2 are given by

$$z_1 = x_1 + jy_1 = |z_1|e^{j\theta_1}, \quad (1.44)$$

$$z_2 = x_2 + jy_2 = |z_2|e^{j\theta_2}, \quad (1.45)$$

then $z_1 = z_2$ if and only if $x_1 = x_2$ and $y_1 = y_2$ or, equivalently, $|z_1| = |z_2|$ and $\theta_1 = \theta_2$.

Addition:

$$z_1 + z_2 = (x_1 + x_2) + j(y_1 + y_2). \quad (1.46)$$

Multiplication:

$$\begin{aligned} z_1 z_2 &= (x_1 + jy_1)(x_2 + jy_2) \\ &= (x_1 x_2 - y_1 y_2) + j(x_1 y_2 + x_2 y_1), \end{aligned} \quad (1.47a)$$

or

$$\begin{aligned} z_1 z_2 &= |z_1|e^{j\theta_1} \cdot |z_2|e^{j\theta_2} \\ &= |z_1||z_2|e^{j(\theta_1 + \theta_2)} \\ &= |z_1||z_2|[\cos(\theta_1 + \theta_2) + j \sin(\theta_1 + \theta_2)]. \end{aligned} \quad (1.47b)$$

Division: For $z_2 \neq 0$,

$$\begin{aligned} \frac{z_1}{z_2} &= \frac{x_1 + jy_1}{x_2 + jy_2} \\ &= \frac{(x_1 + jy_1)}{(x_2 + jy_2)} \cdot \frac{(x_2 - jy_2)}{(x_2 - jy_2)} \\ &= \frac{(x_1 x_2 + y_1 y_2) + j(x_2 y_1 - x_1 y_2)}{x_2^2 + y_2^2}, \end{aligned} \quad (1.48a)$$

or

$$\begin{aligned} \frac{z_1}{z_2} &= \frac{|z_1|e^{j\theta_1}}{|z_2|e^{j\theta_2}} \\ &= \frac{|z_1|}{|z_2|}e^{j(\theta_1 - \theta_2)} \\ &= \frac{|z_1|}{|z_2|}[\cos(\theta_1 - \theta_2) + j \sin(\theta_1 - \theta_2)]. \end{aligned} \quad (1.48b)$$

Powers: For any positive integer n ,

$$\begin{aligned} z^n &= (|z|e^{j\theta})^n \\ &= |z|^n e^{jn\theta} = |z|^n (\cos n\theta + j \sin n\theta), \end{aligned} \quad (1.49)$$

$$\begin{aligned} z^{1/2} &= \pm |z|^{1/2} e^{j\theta/2} \\ &= \pm |z|^{1/2} [\cos(\theta/2) + j \sin(\theta/2)]. \end{aligned} \quad (1.50)$$

Useful Relations:

$$\begin{aligned} -1 &= e^{j\pi} = e^{-j\pi} = 1\angle 180^\circ, \\ j &= e^{j\pi/2} = 1\angle 90^\circ, \end{aligned} \quad (1.51)$$

$$-j = -e^{j\pi/2} = e^{-j\pi/2} = 1\angle -90^\circ, \quad (1.52)$$

$$\sqrt{j} = (e^{j\pi/2})^{1/2} = \pm e^{j\pi/4} = \frac{\pm(1 + j)}{\sqrt{2}}, \quad (1.53)$$

$$\sqrt{-j} = \pm e^{-j\pi/4} = \frac{\pm(1 - j)}{\sqrt{2}}. \quad (1.54)$$

Example 1-3 Working with Complex Numbers

Given two complex numbers

$$V = 3 - j4,$$

$$I = -(2 + j3).$$

(a) Express V and I in polar form, and find (b) VI , (c) VI^* , (d) V/I , and (e) $\sqrt{I}$.**Solution**

$$(a) |V| = \sqrt[4]{VV^*}$$

$$= \sqrt[4]{(3 - j4)(3 + j4)} = \sqrt[4]{9 + 16} = 5,$$

$$\theta_V = \tan^{-1}(-4/3) = -53.1^\circ,$$

$$V = |V|e^{j\theta_V} = 5e^{-j53.1^\circ} = 5\angle -53.1^\circ,$$

$$|I| = \sqrt[4]{2^2 + 3^2} = \sqrt[4]{13} = 3.61.$$

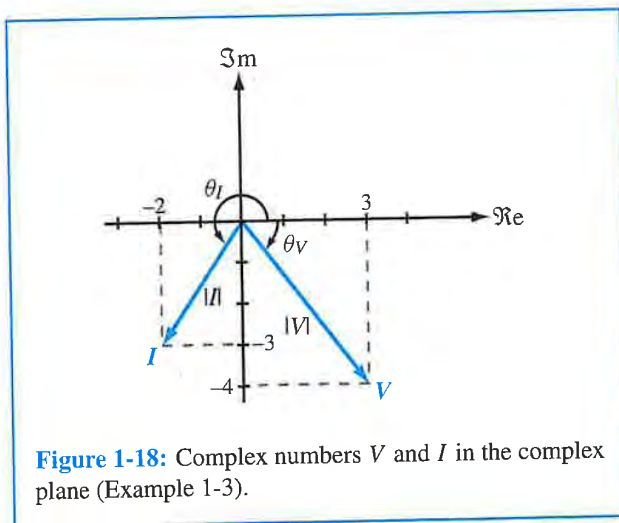
Since $I = (-2 - j3)$ is in the third quadrant in the complex plane [Fig. 1-18],

$$\theta_I = 180^\circ + \tan^{-1}\left(\frac{3}{2}\right) = 236.3^\circ,$$

$$I = 3.61\angle 236.3^\circ.$$

$$(b) VI = 5e^{-j53.1^\circ} \times 3.61e^{j236.3^\circ}$$

$$= 18.05e^{j(236.3^\circ - 53.1^\circ)} = 18.05e^{j183.2^\circ}.$$

**Figure 1-18:** Complex numbers V and I in the complex plane (Example 1-3).

$$(c) VI^* = 5e^{-j53.1^\circ} \times 3.61e^{-j236.3^\circ}$$

$$= 18.05e^{-j289.4^\circ} = 18.05e^{j70.6^\circ}.$$

$$(d) \frac{V}{I} = \frac{5e^{-j53.1^\circ}}{3.61e^{j236.3^\circ}}$$

$$= 1.39e^{-j289.4^\circ} = 1.39e^{j70.6^\circ}.$$

$$(e) \sqrt{I} = \sqrt{3.61e^{j236.3^\circ}}$$

$$= \pm\sqrt{3.61}e^{j236.3^\circ/2} = \pm 1.90e^{j118.15^\circ}. \quad \blacksquare$$

EXERCISE 1.3 Express the following complex functions in polar form:

$$z_1 = (4 - j3)^2,$$

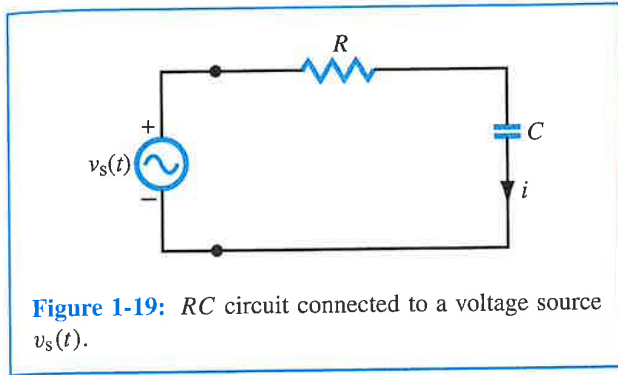
$$z_2 = (4 - j3)^{1/7}.$$

$$\text{Ans. } z_1 = 25\angle -73.7^\circ, z_2 = \pm\sqrt[7]{5}\angle -18.4^\circ.$$

EXERCISE 1.4 Show that $\sqrt{2j} = \pm(1 + j)$.**1-6 Review of Phasors**

Phasor analysis is a useful mathematical tool for solving problems involving linear systems in which the excitation is a periodic time function. Many engineering problems are cast in the form of linear integro-differential equations. If the excitation, more commonly known as the **forcing function**, varies sinusoidally with time, the use of phasor notation to represent time-dependent variables allows us to convert the integro-differential equation into a linear equation with no sinusoidal functions, thereby simplifying the method of solution. After solving for the desired variable, such as the voltage or current in a circuit, conversion from the phasor domain back to the time domain provides the desired result.

The phasor technique can also be used for analyzing linear systems when the forcing function is any arbitrary (nonsinusoidal) periodic time function, such as a square wave or a sequence of pulses. By expanding the forcing



function into a Fourier series of sinusoidal components, we can solve for the desired variable using phasor analysis for each Fourier component of the forcing function separately. According to the principle of superposition, the sum of the solutions due to all of the Fourier components gives the same result as one would obtain had the problem been solved entirely in the time domain without the aid of Fourier representation. The obvious advantage of the phasor–Fourier approach is **simplicity**. Moreover, in the case of nonperiodic source functions, such as a single pulse, the functions can be expressed as Fourier integrals, and a similar application of the principle of superposition can be used as well.

The simple RC circuit shown in Fig. 1-19 contains a sinusoidally time-varying voltage source given by

$$v_s(t) = V_0 \sin(\omega t + \phi_0), \quad (1.55)$$

where V_0 is the amplitude, ω is the angular frequency, and ϕ_0 is a reference phase. Application of Kirchhoff's voltage law gives the following loop equation:

$$R i(t) + \frac{1}{C} \int i(t) dt = v_s(t) \quad (\text{time domain}). \quad (1.56)$$

Our objective is to obtain an expression for the current $i(t)$. We can do this by solving Eq. (1.56) in the time domain, which is somewhat cumbersome because the

forcing function $v_s(t)$ is a sinusoid. Alternatively, we can take advantage of the phasor technique as follows.

Step 1: Adopt a cosine reference

This means that we should express the forcing function as a cosine, if not already in that form, and hence all time-varying functions, such as the current in the circuit and the voltages across R and C , will also have a cosine reference. Thus,

$$\begin{aligned} v_s(t) &= V_0 \sin(\omega t + \phi_0) \\ &= V_0 \cos\left(\frac{\pi}{2} - \omega t - \phi_0\right) \\ &= V_0 \cos\left(\omega t + \phi_0 - \frac{\pi}{2}\right), \end{aligned} \quad (1.57)$$

where we used the properties $\sin x = \cos(\pi/2 - x)$ and $\cos(-x) = \cos x$.

Step 2: Express time-dependent variables as phasors

Any cosinusoidally time-varying function $z(t)$ can be expressed in the form

$$z(t) = \Re \left[\tilde{Z} e^{j\omega t} \right], \quad (1.58)$$

where $\tilde{Z}$ is a time-independent function called the **phasor** of the **instantaneous** function $z(t)$. To distinguish instantaneous quantities from their phasor counterparts, a letter denoting a phasor is given a tilde ($\sim$) over the letter. The voltage $v_s(t)$ given by Eq. (1.57) can be cast in the form

$$\begin{aligned} v_s(t) &= \Re \left[V_0 e^{j(\omega t + \phi_0 - \pi/2)} \right] \\ &= \Re \left[V_0 e^{j(\phi_0 - \pi/2)} e^{j\omega t} \right] \\ &= \Re \left[\tilde{V}_s e^{j\omega t} \right], \end{aligned} \quad (1.59)$$

where

$$\tilde{V}_s = V_0 e^{j(\phi_0 - \pi/2)}. \quad (1.60)$$

The phasor $\tilde{V}_s$, corresponding to the time function $v_s(t)$, contains amplitude and phase information but is

independent of the time variable t . Next we define the unknown variable $i(t)$ in terms of a phasor $\tilde{I}$,

$$i(t) = \Re(\tilde{I}e^{j\omega t}), \quad (1.61)$$

and if the equation we are trying to solve contains derivatives or integrals, we use the following two properties:

$$\begin{aligned} \frac{di}{dt} &= \frac{d}{dt} [\Re(\tilde{I}e^{j\omega t})] \\ &= \Re \left[\frac{d}{dt} (\tilde{I}e^{j\omega t}) \right] \\ &= \Re[j\omega \tilde{I}e^{j\omega t}], \end{aligned} \quad (1.62)$$

and

$$\begin{aligned} \int i dt &= \int \Re(\tilde{I}e^{j\omega t}) dt \\ &= \Re \left(\int \tilde{I}e^{j\omega t} dt \right) \\ &= \Re \left(\frac{\tilde{I}}{j\omega} e^{j\omega t} \right). \end{aligned} \quad (1.63)$$

Thus, differentiation of the time function $i(t)$ is equivalent to multiplication of its phasor $\tilde{I}$ by $j\omega$, and integration is equivalent to division by $j\omega$.

Step 3: Recast the differential/integral equation in phasor form

Upon using Eqs. (1.59), (1.61), and (1.63) in Eq. (1.56), we have

$$R \Re(\tilde{I}e^{j\omega t}) + \frac{1}{C} \Re \left(\frac{\tilde{I}}{j\omega} e^{j\omega t} \right) = \Re(\tilde{V}_s e^{j\omega t}). \quad (1.64)$$

Since both R and C are real quantities and the $\Re(\cdot)$ operation is distributive, Eq. (1.64) simplifies to

$$\tilde{I} \left(R + \frac{1}{j\omega C} \right) = \tilde{V}_s \quad (\text{phasor domain}). \quad (1.65)$$

The time factor $e^{j\omega t}$ has disappeared because it was contained in all three terms. Equation (1.65) is the phasor-domain equivalent of Eq. (1.56).

Step 4: Solve the phasor-domain equation

From Eq. (1.65) the phasor current $\tilde{I}$ is given by

$$\tilde{I} = \frac{\tilde{V}_s}{R + 1/(j\omega C)}. \quad (1.66)$$

Before we apply the next step, we need to convert the right-hand side of Eq. (1.66) into the form $I_0 e^{j\theta}$ with I_0 being a real quantity. Thus,

$$\begin{aligned} \tilde{I} &= V_0 e^{j(\phi_0 - \pi/2)} \left[\frac{j\omega C}{1 + j\omega RC} \right] \\ &= V_0 e^{j(\phi_0 - \pi/2)} \left[\frac{\omega C e^{j\pi/2}}{\sqrt{1 + \omega^2 R^2 C^2} e^{j\phi_1}} \right] \\ &= \frac{V_0 \omega C}{\sqrt{1 + \omega^2 R^2 C^2}} e^{j(\phi_0 - \phi_1)}, \end{aligned} \quad (1.67)$$

where we have used the identity $j = e^{j\pi/2}$. The phase angle $\phi_1 = \tan^{-1}(\omega RC)$ and lies in the first quadrant of the complex plane.

Step 5: Find the instantaneous value

To find $i(t)$, we simply apply Eq. (1.61). That is, we multiply the phasor $\tilde{I}$ given by Eq. (1.67) by $e^{j\omega t}$ and then take the real part:

$$\begin{aligned} i(t) &= \Re \left[\tilde{I} e^{j\omega t} \right] \\ &= \Re \left[\frac{V_0 \omega C}{\sqrt{1 + \omega^2 R^2 C^2}} e^{j(\phi_0 - \phi_1)} e^{j\omega t} \right] \\ &= \frac{V_0 \omega C}{\sqrt{1 + \omega^2 R^2 C^2}} \cos(\omega t + \phi_0 - \phi_1). \end{aligned} \quad (1.68)$$

In summary, we converted all time-varying quantities into the phasor domain, solved for the phasor $\tilde{I}$ of the desired instantaneous current $i(t)$, and then converted back to the time domain to obtain an expression for $i(t)$. Table 1-5 provides a summary of some time-domain functions and their phasor-domain equivalents.

Example 1-4 RL Circuit

The voltage source of the circuit shown in Fig. 1-20 is given by

$$v_s(t) = 5 \sin(4 \times 10^4 t - 30^\circ) \quad (\text{V}). \quad (1.69)$$

Obtain an expression for the voltage across the inductor.

Solution: The voltage loop equation of the RL circuit is

$$Ri + L \frac{di}{dt} = v_s(t). \quad (1.70)$$

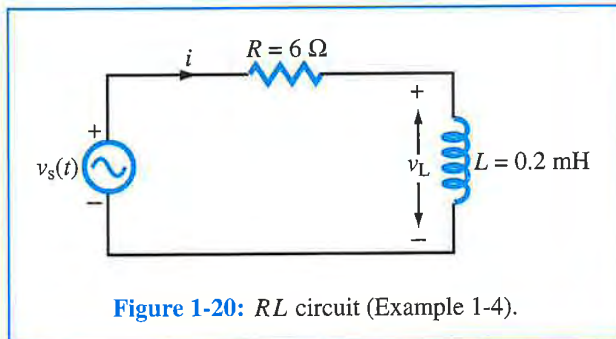


Figure 1-20: RL circuit (Example 1-4).

Table 1-5: Time-domain sinusoidal functions $z(t)$ and their cosine-reference phasor-domain equivalents $\tilde{Z}$, where $z(t) = \Re[\tilde{Z}e^{j\omega t}]$.

$z(t)$		$\tilde{Z}$
$A \cos \omega t$	$\leftrightarrow$	A
$A \cos(\omega t + \phi_0)$	$\leftrightarrow$	$Ae^{j\phi_0}$
$A \cos(\omega t + \beta x + \phi_0)$	$\leftrightarrow$	$Ae^{j(\beta x + \phi_0)}$
$Ae^{-\alpha x} \cos(\omega t + \beta x + \phi_0)$	$\leftrightarrow$	$Ae^{-\alpha x} e^{j(\beta x + \phi_0)}$
$A \sin \omega t$	$\leftrightarrow$	$Ae^{-j\pi/2}$
$A \sin(\omega t + \phi_0)$	$\leftrightarrow$	$Ae^{j(\phi_0 - \pi/2)}$
$\frac{d}{dt}(z_1(t))$	$\leftrightarrow$	$j\omega \tilde{Z}_1$
$\frac{d}{dt}[A \cos(\omega t + \phi_0)]$	$\leftrightarrow$	$j\omega A e^{j\phi_0}$
$\int z_1(t) dt$	$\leftrightarrow$	$\frac{1}{j\omega} \tilde{Z}_1$
$\int A \sin(\omega t + \phi_0) dt$	$\leftrightarrow$	$\frac{1}{j\omega} A e^{j(\phi_0 - \pi/2)}$

Before converting Eq. (1.70) into the phasor domain, we need to express Eq. (1.69) in terms of a cosine reference:

$$\begin{aligned} v_s(t) &= 5 \sin(4 \times 10^4 t - 30^\circ) \\ &= 5 \cos(4 \times 10^4 t - 120^\circ) \quad (\text{V}). \end{aligned} \quad (1.71)$$

The coefficient of t specifies the angular frequency as $\omega = 4 \times 10^4$ (rad/s). The voltage phasor corresponding to $v_s(t)$ is

$$\tilde{V}_s = 5e^{-j120^\circ} \quad (\text{V}),$$

and the phasor equation corresponding to Eq. (1.70) is

$$R\tilde{I} + j\omega L\tilde{I} = \tilde{V}_s. \quad (1.72)$$

Solving for the current phasor $\tilde{I}$, we have

$$\begin{aligned}\tilde{I} &= \frac{\tilde{V}_s}{R + j\omega L} \\ &= \frac{5e^{-j120^\circ}}{6 + j4 \times 10^4 \times 2 \times 10^{-4}} \\ &= \frac{5e^{-j120^\circ}}{6 + j8} = \frac{5e^{-j120^\circ}}{10e^{j53.1^\circ}} = 0.5e^{-j173.1^\circ} \quad (\text{A}).\end{aligned}$$

The voltage phasor across the inductor is related to $\tilde{I}$ by

$$\begin{aligned}\tilde{V}_L &= j\omega L \tilde{I} \\ &= j4 \times 10^4 \times 2 \times 10^{-4} \times 0.5e^{-j173.1^\circ} \\ &= 4e^{j(90^\circ - 173.1^\circ)} = 4e^{-j83.1^\circ} \quad (\text{V}),\end{aligned}$$

and the corresponding instantaneous voltage $v_L(t)$ is therefore

$$\begin{aligned}v_L(t) &= \Re[\tilde{V}_L e^{j\omega t}] \\ &= \Re[4e^{-j83.1^\circ} e^{j4 \times 10^4 t}] \\ &= 4 \cos(4 \times 10^4 t - 83.1^\circ) \quad (\text{V}). \quad \blacksquare\end{aligned}$$

REVIEW QUESTIONS

Q1.12 Why is the phasor technique useful? When is it used? Describe the process.

Q1.13 How is the phasor technique used when the forcing function is a non-sinusoidal periodic waveform, such as a train of pulses?

EXERCISE 1.5 A series RL circuit is connected to a voltage source given by $v_s(t) = 150 \cos \omega t$ (V). Find (a) the phasor current $\tilde{I}$ and (b) the instantaneous current $i(t)$ for $R = 400 \, \Omega$, $L = 3 \, \text{mH}$, and $\omega = 10^5 \, \text{rad/s}$.

Ans. (a) $\tilde{I} = 150/(R + j\omega L) = 0.3 \angle -36.9^\circ$ (A), (b) $i(t) = 0.3 \cos(\omega t - 36.9^\circ)$ (A).

EXERCISE 1.6 A phasor voltage is given by $\tilde{V} = j5$ V. Find $v(t)$.

Ans. $v(t) = 5 \cos(\omega t + \pi/2) = -5 \sin \omega t$ (V).

CHAPTER HIGHLIGHTS

- Electromagnetics is the study of electric and magnetic phenomena and their engineering applications.
- The International System of Units consists of the six fundamental dimensions listed in Table 1-1. The units of all other physical quantities can be expressed in terms of the six fundamental units.
- The four fundamental forces of nature are the nuclear, weak-interaction, electromagnetic and gravitational forces.
- The source of the electric field quantities $\mathbf{E}$ and $\mathbf{D}$ is the electric charge q . In a material, $\mathbf{E}$ and $\mathbf{D}$ are related by $\mathbf{D} = \epsilon \mathbf{E}$, where ϵ is the electrical permittivity of the material. In free space, $\epsilon = \epsilon_0 \simeq (1/36\pi) \times 10^{-9}$ (F/m).
- The source of the magnetic field quantities $\mathbf{B}$ and $\mathbf{H}$ is the electric current I . In a material, $\mathbf{B}$ and $\mathbf{H}$ are related by $\mathbf{B} = \mu \mathbf{H}$, where μ is the magnetic permeability of the medium. In free space, $\mu = \mu_0 = 4\pi \times 10^{-7}$ (H/m).
- Electromagnetics consists of three branches: (1) electrostatics, which pertains to stationary charges, (2) magnetostatics, which pertains to steady currents, and (3) electrodynamics, which pertains to time-varying currents.
- A traveling wave is characterized by a spatial wavelength λ , a time period T , and a phase velocity $u_p = \lambda/T$.

- An electromagnetic (EM) wave consists of oscillating electric and magnetic field intensities and travels in free space at the velocity of light $c = 1/\sqrt{\epsilon_0\mu_0}$. The EM spectrum encompasses gamma rays, x rays, visible light, infrared waves, and radio waves.
- Phasor analysis is a useful mathematical tool for solving problems involving time-periodic sources.

GLOSSARY OF IMPORTANT TERMS

Provide definitions or explain the meaning of the following terms:

SI system of units
 fundamental dimensions
 Coulomb's law
 electric field intensity **E**
 law of conservation of electric charge
 principle of linear superposition
 electric dipole
 electric polarization
 electrical permittivity ϵ
 relative permittivity or dielectric constant ϵ_r
 electric flux density **D**
 electrostatics
 magnetostatics
 magnetic flux density **B**
 magnetic permeability μ
 velocity of light c
 nonmagnetic materials
 magnetic field intensity **H**
 electrodynamics
 conductivity σ
 perfect dielectric
 perfect conductor
 constitutive parameters
 transient wave
 continuous harmonic wave
 wave amplitude
 wave period T

wavelength λ
 reference phase ϕ_0
 phase velocity (propagation velocity) u_p
 wave frequency f
 angular velocity ω
 phase constant (wave number) β
 phase lag and lead
 attenuation factor
 attenuation constant α
 EM spectrum
 microwave band
 complex number
 Euler's identity
 complex conjugate
 forcing function
 phasor
 instantaneous function

PROBLEMS

Section 1-3: Traveling Waves

1.1* A 4-kHz sound wave traveling in the x -direction in air was observed to have a differential pressure $p(x, t) = 5 \text{ N/m}^2$ at $x = 0$ and $t = 25 \mu\text{s}$. If the reference phase of $p(x, t)$ is 42° , find a complete expression for $p(x, t)$. The velocity of sound in air is 330 m/s.

1.2 For the pressure wave described in Example 1-1, plot the following:

(a) $p(x, t)$ versus x at $t = 0$

(b) $p(x, t)$ versus t at $x = 0$

Be sure to use appropriate scales for x and t so that each of your plots covers at least two cycles.

1.3* A harmonic wave traveling along a string is generated by an oscillator that completes 120 vibrations per minute. If it is observed that a given crest, or maximum, travels 250 cm in 10 s, what is the wavelength?

*Answer(s) available in Appendix D.

 Solution available in CD-ROM.

- 1.4** Two waves, $y_1(t)$ and $y_2(t)$, have identical amplitudes and oscillate at the same frequency, but $y_2(t)$ leads $y_1(t)$ by a phase angle of 60° . If

$$y_1(t) = 4 \cos(2\pi \times 10^3 t)$$

write the expression appropriate for $y_2(t)$ and plot both functions over the time span from 0 to 2 ms.

- 1.5*** The height of an ocean wave is described by the function

$$y(x, t) = 1.5 \sin(0.5t - 0.6x) \quad (\text{m})$$

Determine the phase velocity and wavelength, and then sketch $y(x, t)$ at $t = 2\text{ s}$ over the range from $x = 0$ to $x = 2\lambda$.

- 1.6** A wave traveling along a string in the $+x$ -direction is given by

$$y_1(x, t) = A \cos(\omega t - \beta x)$$

where $x = 0$ is the end of the string, which is tied rigidly to a wall, as shown in Fig. 1-21. When wave $y_1(x, t)$ arrives at the wall, a reflected wave $y_2(x, t)$ is generated. Hence, at any location on the string, the vertical displacement y_s is the sum of the incident and reflected waves:

$$y_s(x, t) = y_1(x, t) + y_2(x, t)$$

- Write an expression for $y_2(x, t)$, keeping in mind its direction of travel and the fact that the end of the string cannot move.
- Generate plots of $y_1(x, t)$, $y_2(x, t)$ and $y_s(x, t)$ versus x over the range $-2\lambda \leq x \leq 0$ at $\omega t = \pi/4$ and at $\omega t = \pi/2$.

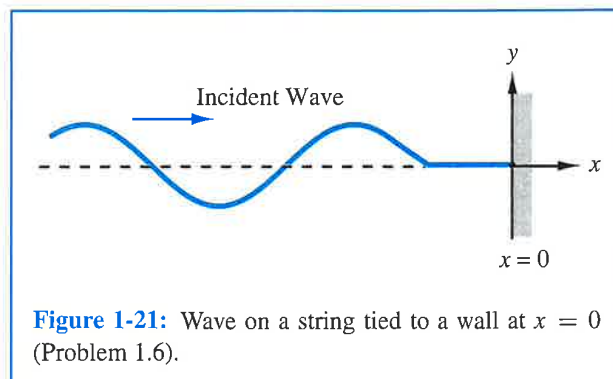


Figure 1-21: Wave on a string tied to a wall at $x = 0$ (Problem 1.6).

- 1.7*** Two waves on a string are given by the following functions:

$$y_1(x, t) = 3 \cos(20t - 30x) \quad (\text{cm})$$

$$y_2(x, t) = -3 \cos(20t + 30x) \quad (\text{cm})$$

where x is in centimeters. The waves are said to interfere constructively when their superposition $|y_s| = |y_1 + y_2|$ is a maximum, and they interfere destructively when $|y_s|$ is a minimum.

- What are the directions of propagation of waves $y_1(x, t)$ and $y_2(x, t)$?
- At $t = (\pi/50)\text{ s}$, at what location x do the two waves interfere constructively, and what is the corresponding value of $|y_s|$?
- At $t = (\pi/50)\text{ s}$, at what location x do the two waves interfere destructively, and what is the corresponding value of $|y_s|$?

- 1.8** Give expressions for $y(x, t)$ for a sinusoidal wave traveling along a string in the negative x -direction, given that $y_{\text{max}} = 20\text{ cm}$, $\lambda = 30\text{ cm}$, $f = 5\text{ Hz}$, and

- $y(x, 0) = 0$ at $x = 0$
- $y(x, 0) = 0$ at $x = 7.5\text{ cm}$

1.9* An oscillator that generates a sinusoidal wave on a string completes 20 vibrations in 30 s. The wave peak is observed to travel a distance of 2.8 m along the string in 5 s. What is the wavelength?

1.10 The vertical displacement of a string is given by the harmonic function:

$$y(x, t) = 5 \cos(12\pi t - 20\pi x) \quad (\text{m})$$

where x is the horizontal distance along the string in meters. Suppose a tiny particle were attached to the string at $x = 5$ cm. Obtain an expression for the vertical velocity of the particle as a function of time.

1.11* Given two waves characterized by the following:

$$y_1(t) = 6 \cos \omega t$$

$$y_2(t) = 6 \sin(\omega t + 30^\circ)$$

does $y_2(t)$ lead or lag $y_1(t)$ and by what phase angle?

1.12 The voltage of an electromagnetic wave traveling on a transmission line is given by

$$v(z, t) = 3e^{-\alpha z} \sin(2\pi \times 10^9 t - 10\pi z) \quad (\text{V})$$

where z is the distance in meters from the generator.

- (a) Find the frequency, wavelength, and phase velocity of the wave.
- (b) At $z = 2$ m, the amplitude of the wave was measured to be 1 V. Find α .

1.13* A certain electromagnetic wave traveling in seawater was observed to have an amplitude of 19.025 (V/m) at a depth of 10 m, and an amplitude of 12.13 (V/m) at a depth of 100 m. What is the attenuation constant of seawater?

Section 1-5: Complex Numbers

1.14 Evaluate each of the following complex numbers and express the result in rectangular form:

- (a) $z_1 = 3e^{j\pi/4}$
- (b) $z_2 = \sqrt{3}e^{j3\pi/4}$
- (c) $z_3 = 2e^{-j\pi/2}$
- (d) $z_4 = j^3$
- (e) $z_5 = j^{-4}$
- (f) $z_6 = (1 - j)^3$
- (g) $z_7 = (1 - j)^{1/2}$

1.15* Complex numbers z_1 and z_2 are given by the following:

$$z_1 = 3 - j2$$

$$z_2 = -4 + j2$$

- (a) Express z_1 and z_2 in polar form.
- (b) Find $|z_1|$ by first applying Eq. (1.41) and then by applying Eq. (1.43).
- (c) Determine the product $z_1 z_2$ in polar form.
- (d) Determine the ratio z_1/z_2 in polar form.
- (e) Determine z_1^3 in polar form.

1.16 If $z = -2 + j3$, determine the following quantities in polar form:

- (a) $1/z$
- (b) z^3
- (c) $|z|^2$
- (d) $\Im\{z\}$
- (e) $\Im\{z^*\}$

1.17* Find complex numbers $t = z_1 + z_2$ and $s = z_1 - z_2$, both in polar form, for each of the following pairs:

- (a) $z_1 = 2 + j3$ and $z_2 = 1 - j2$
- (b) $z_1 = 2$ and $z_2 = -j2$

(c) $z_1 = 3\angle 30^\circ$ and $z_2 = 3\angle -30^\circ$

(d) $z_1 = 3\angle 30^\circ$ and $z_2 = 3\angle -150^\circ$

1.18 Complex numbers z_1 and z_2 are given by the following:

$$z_1 = 5\angle -60^\circ$$

$$z_2 = 2\angle 45^\circ$$

- (a) Determine the product $z_1 z_2$ in polar form.
- (b) Determine the product $z_1 z_2^*$ in polar form.
- (c) Determine the ratio z_1/z_2 in polar form.
- (d) Determine the ratio z_1^*/z_2^* in polar form.
- (e) Determine $\sqrt{z_1}$ in polar form.

1.19* If $z = 3 - j4$, find the value of $\ln(z)$.

1.20 If $z = 3 - j4$, find the value of e^z .

Section 1-6: Phasors

1.21* A voltage source given by

$$v_s(t) = 10 \cos(2\pi \times 10^3 t - 30^\circ) \quad (\text{V})$$

is connected to a series RC load as shown in Fig. 1-19. If $R = 1 \text{ M}\Omega$ and $C = 100 \text{ pF}$, obtain an expression for $v_c(t)$, the voltage across the capacitor.

1.22 Find the phasors of the following time functions:

- (a) $v(t) = 3 \cos(\omega t - \pi/4) \text{ (V)}$
- (b) $v(t) = 12 \sin(\omega t + \pi/4) \text{ (V)}$
- (c) $i(x, t) = 4e^{-3x} \sin(\omega t - \pi/6) \text{ (A)}$
- (d) $i(t) = -2 \cos(\omega t + 3\pi/4) \text{ (A)}$
- (e) $i(t) = 2 \sin(\omega t + \pi/3) + 3 \cos(\omega t - \pi/6) \text{ (A)}$

1.23* Find the instantaneous time sinusoidal functions corresponding to the following phasors:

- (a) $\tilde{V} = -3e^{j\pi/3} \text{ (V)}$
- (b) $\tilde{V} = j6e^{j\pi/4} \text{ (V)}$
- (c) $\tilde{I} = (3 + j4) \text{ (A)}$

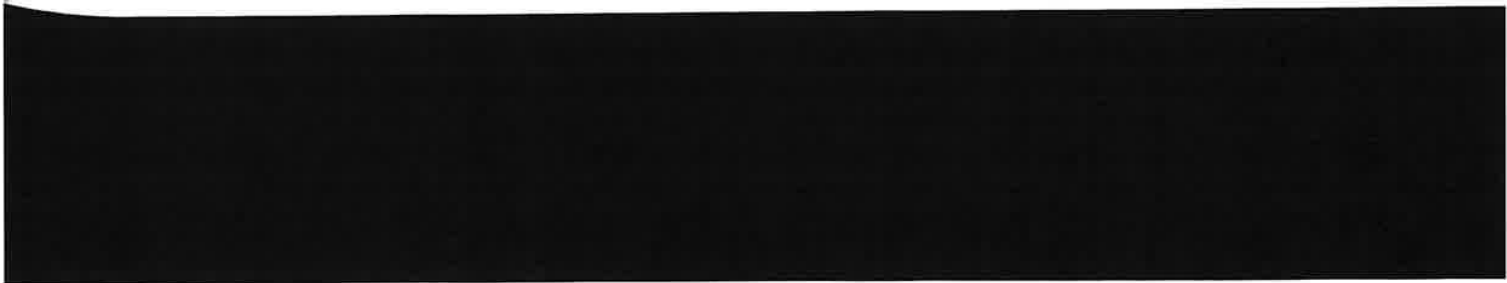
(d) $\tilde{I} = -3 + j2 \text{ (A)}$

(e) $\tilde{I} = j \text{ (A)}$

(f) $\tilde{I} = 2e^{j3\pi/4} \text{ (A)}$

1.24 A series RLC circuit is connected to a generator with a voltage $v_s(t) = V_0 \cos(\omega t + \pi/3) \text{ (V)}$.

- (a) Write the voltage loop equation in terms of the current $i(t)$, R , L , C , and $v_s(t)$.
- (b) Obtain the corresponding phasor-domain equation.
- (c) Solve the equation to obtain an expression for the phasor current $\tilde{I}$.



CHAPTER

2-1	General Considerations
2-2	Lumped-Element Model
2-3	Transmission-Line Equations
2-4	Wave Propagation on a Transmission Line
2-5	The Lossless Transmission Line
2-6	Input Impedance of the Lossless Line
2-7	Special Cases of the Lossless Line
2-8	Power Flow on a Lossless Transmission Line
2-9	The Smith Chart
2-10	Impedance Matching
2-11	Transients on Transmission Lines

2-1 General Considerations

In most electrical engineering curricula, the study of electromagnetics is preceded by one or more courses on electrical circuits. In this book, we use this background to build a bridge between circuit theory and electromagnetic theory. The bridge is provided by transmission lines, the topic of this chapter. By modeling the transmission line in the form of an equivalent circuit, we can use Kirchhoff's voltage and current laws to develop wave equations whose solutions provide an understanding of wave propagation, standing waves, and power transfer. Familiarity with these concepts facilitates the presentation of material in later chapters.

Although the family of *transmission lines* may encompass all structures and media that serve to transfer energy or information between two points, including nerve fibers in the human body, acoustic waves in fluids, and mechanical pressure waves in solids, we shall focus our treatment in this chapter on transmission lines used for

guiding electromagnetic signals. Such transmission lines include telephone wires, coaxial cables carrying audio and video information to TV sets or digital data to computer monitors, and optical fibers carrying light waves for the transmission of data at very high rates. Fundamentally, a transmission line is a two-port network, with each port consisting of two terminals, as illustrated in Fig. 2-1. One of the ports is the sending end and the other is the receiving end. The source connected to its sending end may be any circuit with an output voltage, such as a radar transmitter, an amplifier, or a computer terminal operating in the transmission mode. From circuit theory, any such source can be represented by a Thévenin-equivalent *generator circuit* consisting of a generator voltage V_g in series with a generator resistance R_g , as shown in Fig. 2-1. The generator voltage may consist of digital pulses, a modulated time-varying sinusoidal signal, or any other signal waveform. In the case of a-c signals, the generator circuit is represented by a voltage phasor $\tilde{V}_g$ and an impedance Z_g .

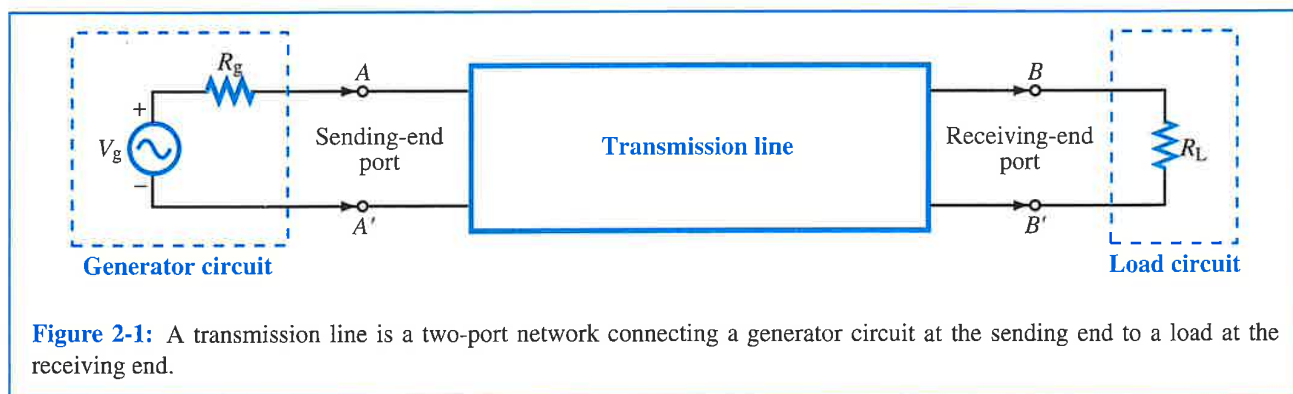
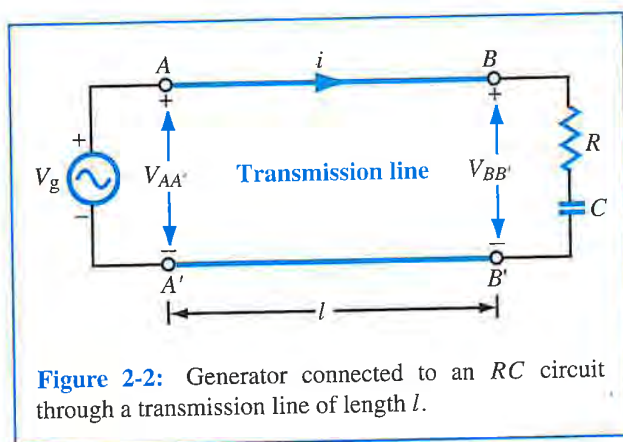


Figure 2-1: A transmission line is a two-port network connecting a generator circuit at the sending end to a load at the receiving end.

The circuit connected to the receiving end of the transmission line is called the load circuit, or simply the **load**. This may be an antenna in the case of a radar, a computer terminal operating in the receiving mode, the input terminals of an amplifier, or any output circuit whose input terminals can be represented by an equivalent load resistance R_L , or a load impedance Z_L in the a-c case.

2-1.1 The Role of Wavelength

In low-frequency electrical circuits, we usually use wires to connect the elements of the circuit in the desired configuration. In the circuit shown in Fig. 2-2, for example, the generator is connected to a simple RC load via a pair of wires. In view of our definition in the preceding paragraphs of what constitutes a transmission line, we pose the following question: Is the pair of wires between terminals AA' and terminals BB' a transmission line? If so, why is it important? After all, we usually solve for the current in the circuit and the voltage across its elements without regard for the wires connecting the elements. The answer to this question is yes; indeed the pair of wires constitutes a transmission line, but the impact of the line on the current and voltages in the circuit depends on the length of the line l and the frequency f of



the signal provided by the generator. (As we will see later, the determining factor is the ratio of the length l to the wavelength λ of the wave propagating on the transmission line between AA' and BB' .) If the generator voltage is cosinusoidal in time, then the voltage across the input terminals AA' is

$$V_{AA'} = V_g(t) = V_0 \cos \omega t \quad (\text{V}), \quad (2.1)$$

where $\omega = 2\pi f$ is the angular frequency, and if we assume that the current flowing through the wires travels at the speed of light, $c = 3 \times 10^8$ m/s, then the voltage across the output terminals BB' will have to be delayed in time relative to that across AA' by the travel delay time l/c . Thus, assuming no significant ohmic losses in the transmission line,

$$\begin{aligned} V_{BB'}(t) &= V_{AA'}(t - l/c) \\ &= V_0 \cos [\omega(t - l/c)] \quad (\text{V}). \end{aligned} \quad (2.2)$$

Let us compare $V_{BB'}$ to $V_{AA'}$ at $t = 0$ for an ultralow-frequency electronic circuit operating at a frequency $f = 1$ kHz. For a typical wire length $l = 5$ cm, Eqs. (2.1) and (2.2) give $V_{AA'} = V_0$ and $V_{BB'} = V_0 \cos(2\pi f l/c) = 0.99999999998 V_0$. Thus, for all practical purposes, the length of the transmission line may be ignored and terminal AA' may be treated as identical with BB' . On the other hand, had the line been a 20-km long telephone cable carrying a 1-kHz voice signal, then the same calculation would have led to $V_{BB'} = 0.91 V_0$. The determining factor is the magnitude of $\omega l/c$. From Eq. (1.27), the velocity of propagation u_p of a traveling wave is related to the oscillation frequency f and the wavelength λ by

$$u_p = f\lambda \quad (\text{m/s}). \quad (2.3)$$

In the present case, $u_p = c$. Hence, the phase factor

$$\frac{\omega l}{c} = \frac{2\pi f l}{c} = 2\pi \frac{l}{\lambda} \quad \text{radians}. \quad (2.4)$$

When l/λ is very small, transmission-line effects may be ignored, but when $l/\lambda \gtrsim 0.01$, it may be necessary to

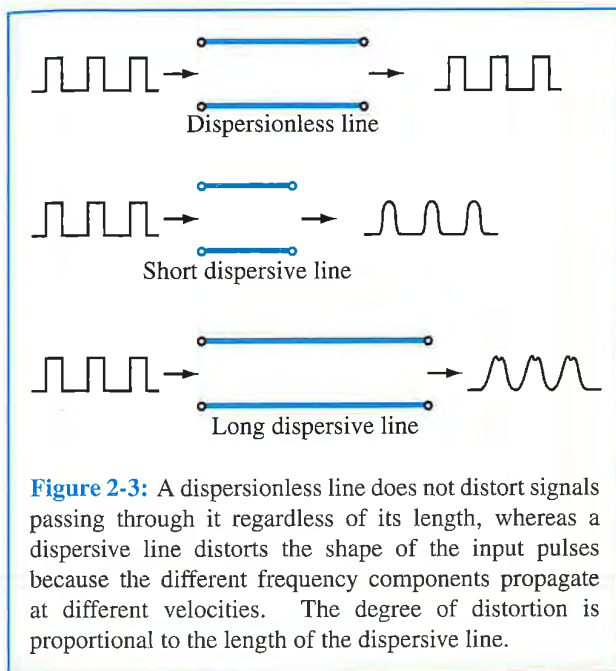


Figure 2-3: A dispersionless line does not distort signals passing through it regardless of its length, whereas a dispersive line distorts the shape of the input pulses because the different frequency components propagate at different velocities. The degree of distortion is proportional to the length of the dispersive line.

account not only for the phase shift associated with the time delay, but also for the presence of *reflected* signals that may have been bounced back by the load toward the generator. *Power loss* on the line and *dispersive* effects may need to be considered as well. A dispersive transmission line is one on which the wave velocity is not constant as a function of the frequency f . This means that the shape of a rectangular pulse, which through Fourier analysis is composed of many waves of different frequencies, will be distorted as it travels down the line because its different frequency components will not propagate at the same velocity (Fig. 2-3). Preservation of pulse shape is very important in high-speed data transmission, both between terminals as well as in high-speed integrated circuits in which transmission-line design and fabrication processes are an integral part of the IC design process. At 10 GHz, for example, the wavelength $\lambda = 3$ cm in air and is on the order of 1 cm in a semiconductor material. Hence, even connection lengths

between devices on the order of millimeters become significant, and their presence has to be incorporated in the overall design of the circuit.

2-1.2 Propagation Modes

A few examples of common types of transmission lines are shown in Fig. 2-4. Transmission lines may be classified into two basic types:

- **Transverse electromagnetic (TEM) transmission lines:** Waves propagating along these lines are characterized by electric and magnetic fields that are entirely *transverse* to the direction of propagation. This is called a TEM mode. A good example is the coaxial line shown in Fig. 2-5; the electric field lines are in the radial direction between the inner and outer conductors, the magnetic field forms circles around the inner conductor, and hence neither has any components along the length of the line (the direction of wave propagation). Other TEM transmission lines include the two-wire line and the parallel-plate line, both shown in Fig. 2-4. Although the fields present on a microstrip line do not adhere to the exact definition of a TEM mode, the nontransverse field components are sufficiently small in comparison to the transverse components to be ignored, thereby allowing the inclusion of microstrip lines in the TEM class. A common feature among TEM lines is that they consist of two parallel conducting surfaces.
- **Higher-order transmission lines:** Waves propagating along these lines have at least one significant field component in the direction of propagation. Hollow conducting waveguides, dielectric rods, and optical fibers belong to this class of lines.

Only TEM-mode transmission lines will be treated in this chapter. This is because less mathematical rigor is required for treating this class of lines than that required for treating waves characterized by higher-order modes

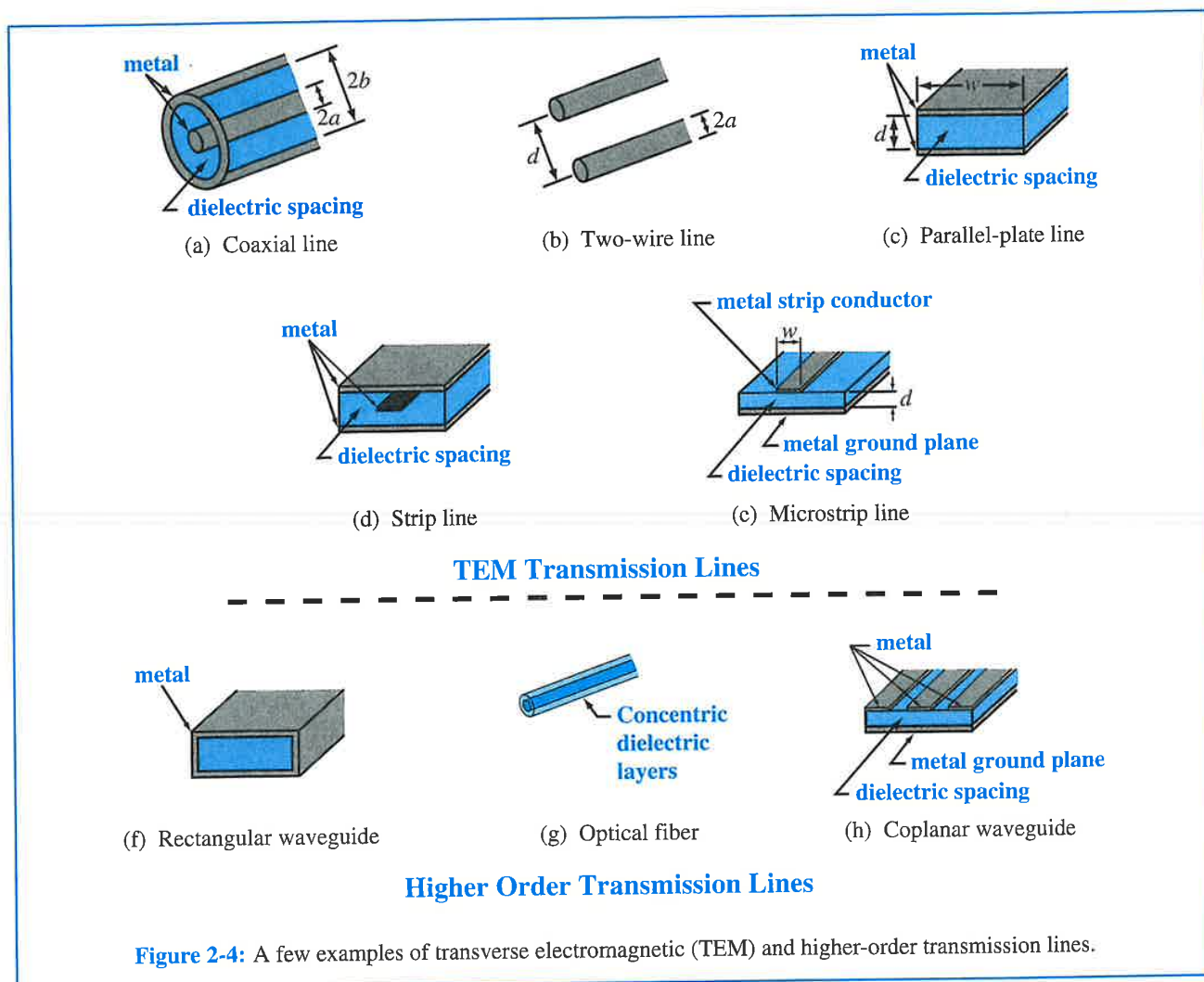


Figure 2-4: A few examples of transverse electromagnetic (TEM) and higher-order transmission lines.

and, in addition, TEM lines are more commonly used in practice. We start our treatment by representing the transmission line in terms of a lumped-element circuit model, and then we apply Kirchhoff's voltage and current laws to derive a set of two governing equations known as the *telegrapher's equations*. By combining these

equations, we obtain wave equations for the voltage and current at any point on the line. Solution of the wave equations for the sinusoidal steady-state case leads to a set of formulas that can be used for solving a wide range of practical problems. In the latter part of this chapter we introduce a graphical technique known as

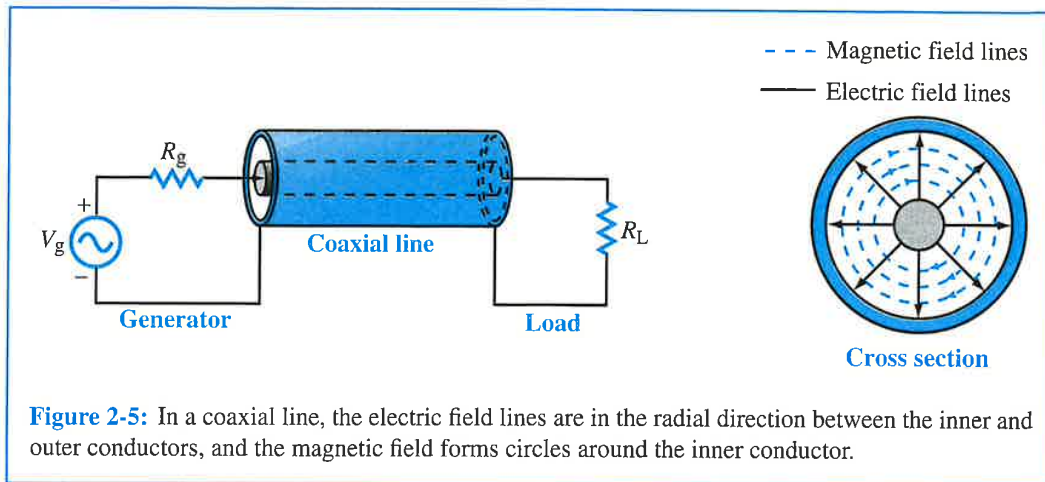


Figure 2-5: In a coaxial line, the electric field lines are in the radial direction between the inner and outer conductors, and the magnetic field forms circles around the inner conductor.

the *Smith chart*, which facilitates the solution of many transmission-line problems without having to perform laborious calculations involving complex numbers.

2-2 Lumped-Element Model

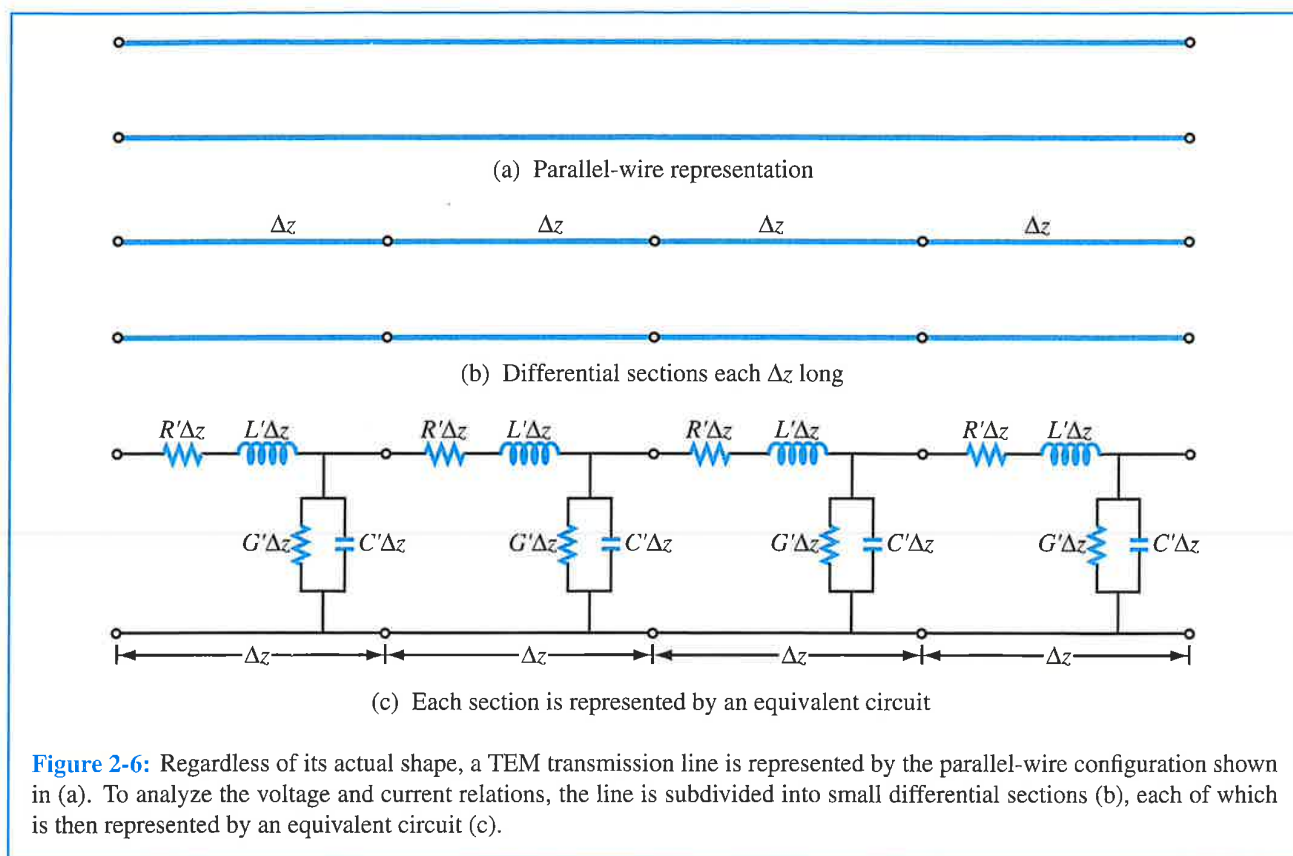
When we draw a schematic of an electronic circuit, we use specific symbols to represent resistors, capacitors, inductors, diodes, and the like. In each case, the symbol represents the functionality of the device, rather than its shape, size or other attributes. We shall do the same with regard to transmission lines; *we shall represent a transmission line by a parallel-wire configuration*, as shown in Fig. 2-6(a), *regardless of the specific shape of the line under consideration*. Thus, Fig. 2-6(a) may represent a coaxial line, a two-wire line, or any other TEM line.

Drawing again on our familiarity with electronic circuits, when we analyze a circuit containing a transistor, we represent the functionality of the transistor by an equivalent circuit composed of sources, resistors, and capacitors. We will apply the same approach to the transmission line by orienting the line along

the z -direction, subdividing it into differential sections each of length Δz [Fig. 2-6(b)] and then representing each section by an equivalent circuit, as illustrated in Fig. 2-6(c). This representation, which is called the *lumped-element circuit* model, consists of four basic elements, which henceforth will be called the *transmission line parameters*. These are

- R' : The combined *resistance* of both conductors per unit length, in Ω/m ,
- L' : The combined *inductance* of both conductors per unit length, in H/m ,
- G' : The *conductance* of the insulation medium per unit length, in S/m , and
- C' : The *capacitance* of the two conductors per unit length, in F/m .

Whereas the four line parameters have different expressions for different types and dimensions of transmission lines, the equivalent model represented by Fig. 2-6(c) is equally applicable to all transmission lines characterized by TEM-mode wave propagation. *The prime superscript is used as a reminder that the line parameters are differential quantities whose units are per unit length.*



Expressions for the line parameters R' , L' , G' , and C' are given in Table 2-1 for the three types of TEM transmission lines diagrammed in parts (a) through (c) of Fig. 2-4. For each of these lines, the expressions are functions of two sets of parameters: (1) geometric parameters defining the cross-sectional dimensions of the given line and (2) electromagnetic constitutive parameters characteristic of the materials of which the conductors and the insulating material between them are made. The pertinent geometric parameters are as follows:

Coaxial line [Fig. 2-4(a)]:

- a = outer radius of inner conductor, m
- b = inner radius of outer conductor, m

Two-wire line [Fig. 2-4(b)]:

- a = radius of each wire, m
- d = spacing between wires' centers, m

Parallel-plate line [Fig. 2-4(c)]:

- w = width of each plate, m
- d = thickness of insulation between plates, m

Table 2-1: Transmission-line parameters R' , L' , G' , and C' for three types of lines.

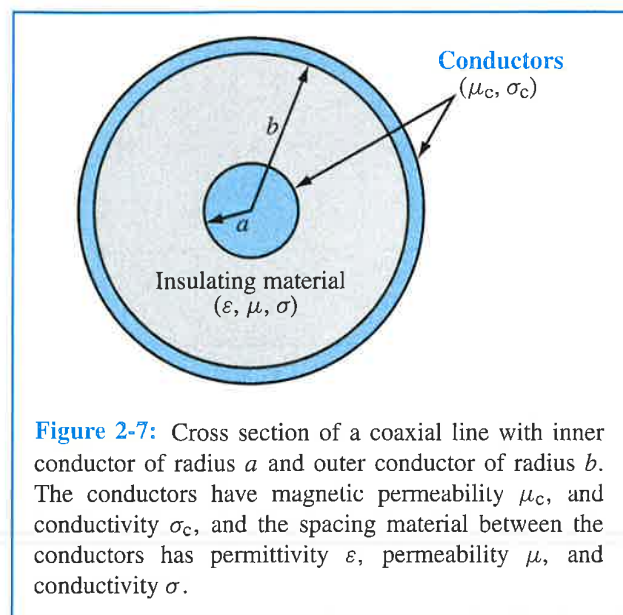
Parameter	Coaxial	Two Wire	Parallel Plate	Unit
R'	$\frac{R_s}{2\pi} \left(\frac{1}{a} + \frac{1}{b} \right)$	$\frac{R_s}{\pi a}$	$\frac{2R_s}{w}$	Ω/m
L'	$\frac{\mu}{2\pi} \ln(b/a)$	$\frac{\mu}{\pi} \ln \left[(d/2a) + \sqrt{(d/2a)^2 - 1} \right]$	$\frac{\mu d}{w}$	H/m
G'	$\frac{2\pi\sigma}{\ln(b/a)}$	$\frac{\pi\sigma}{\ln \left[(d/2a) + \sqrt{(d/2a)^2 - 1} \right]}$	$\frac{\sigma w}{d}$	S/m
C'	$\frac{2\pi\epsilon}{\ln(b/a)}$	$\frac{\pi\epsilon}{\ln \left[(d/2a) + \sqrt{(d/2a)^2 - 1} \right]}$	$\frac{\epsilon w}{d}$	F/m

Notes: (1) Refer to Fig. 2-4 for definitions of dimensions. (2) μ , ϵ , and σ pertain to the insulating material between the conductors. (3) $R_s = \sqrt{\pi f \mu_c / \sigma_c}$. (4) μ_c and σ_c pertain to the conductors. (5) If $(d/2a)^2 \gg 1$, then $\ln \left[(d/2a) + \sqrt{(d/2a)^2 - 1} \right] \simeq \ln(d/a)$.

The constitutive parameters apply to all three lines and consist of two groups: μ_c and σ_c are the magnetic permeability and electrical conductivity of the conductors, and ϵ , μ , and σ are the electrical permittivity, magnetic permeability, and electrical conductivity of the insulation material separating the conductors. Appendix B contains tabulated values for these constitutive parameters for various types of materials. For the purposes of the present chapter, we need not concern ourselves with the derivations responsible for the expressions given in Table 2-1. The formulations necessary for computing R' , L' , G' , and C' will be made available in later chapters for the general case of any two-conductor configuration.

The lumped-element model shown in Fig. 2-6(c) represents the physical processes associated with the currents and voltages on any TEM transmission line. Other equivalent models are available also and are

equally applicable as well. All these models, however, lead to exactly the same set of telegrapher's equations, from which all our future results will be derived. Hence, only the model described in Fig. 2-6(c) will be examined in the present treatment. It consists of two series elements, R' and L' , and two shunt elements, G' and C' . By way of providing a physical explanation for the lumped-element model, let us consider a small section of a coaxial line, as shown in Fig. 2-7. The line consists of an inner conductor of radius a separated from an outer conducting cylinder of radius b by a material with permittivity ϵ , permeability μ , and conductivity σ . The two metal conductors are made of a material with conductivity σ_c and permeability μ_c . When a voltage source is connected across the two conductors at the sending end of the line, currents will flow through the conductors, primarily along the outer surface of the inner conductor and the inner surface of



the outer conductor. The line resistance R' accounts for the combined resistance per unit length of the inner and outer conductors. The expression for R' is derived in Chapter 7 and is given by Eq. (7.96) as

$$R' = \frac{R_s}{2\pi} \left(\frac{1}{a} + \frac{1}{b} \right) \quad (\Omega/\text{m}), \quad (2.5)$$

where R_s , which represents the surface resistance of the conductors, is called the **intrinsic resistance** and is given by Eq. (7.92a) as

$$R_s = \sqrt{\frac{\pi f \mu_c}{\sigma_c}} \quad (\Omega). \quad (2.6)$$

The intrinsic resistance depends not only on the material properties of the conductors (σ_c and μ_c), but on the frequency f of the wave traveling on the line as well. For a **perfect conductor** with $\sigma_c = \infty$ or a high-conductivity

material such that $(f \mu_c / \sigma_c) \ll 1$, R_s approaches zero, and so does R' .

Next, let us examine the inductance per unit length L' . Application of Ampère's law in Chapter 5 to the definition of inductance leads to the following expression [Eq. (5.99)] for the inductance per unit length of a coaxial line:

$$L' = \frac{\mu}{2\pi} \ln \left(\frac{b}{a} \right) \quad (\text{H/m}). \quad (2.7)$$

The shunt conductance per unit length G' accounts for current flow between the outer and inner conductors, made possible by the material conductivity σ of the insulator. It is precisely because the current flow is from one conductor to the other that G' is a shunt element in the lumped-element model. Its expression is given by Eq. (4.76) as

$$G' = \frac{2\pi\sigma}{\ln(b/a)} \quad (\text{S/m}). \quad (2.8)$$

If the material separating the inner and outer conductors is a **perfect dielectric** with $\sigma = 0$, then $G' = 0$.

The last line parameter on our list is the capacitance per unit length C' . When equal and opposite charges are placed on any two noncontacting conductors, a voltage difference will be induced between them. Capacitance is defined as the ratio of charge to voltage difference. For the coaxial line, C' is given by Eq. (4.117) as

$$C' = \frac{2\pi\epsilon}{\ln(b/a)} \quad (\text{F/m}). \quad (2.9)$$

All TEM transmission lines share the following useful relations:

$$L' C' = \mu \epsilon, \quad (2.10)$$

and

$$\frac{G'}{C'} = \frac{\sigma}{\epsilon}. \quad (2.11)$$

If the insulating medium between the conductors is air, the transmission line is called an **air line** (e.g., coaxial air line or two-wire air line). For an air line, $\epsilon = \epsilon_0 = 8.854 \times 10^{-12}$ F/m, $\mu = \mu_0 = 4\pi \times 10^{-7}$ H/m, $\sigma = 0$, and $G' = 0$.

REVIEW QUESTIONS

Q2.1 What is a transmission line? When should transmission-line effects be considered?

Q2.2 What is the difference between dispersive and nondispersive transmission lines? What is the practical significance?

Q2.3 What constitutes a TEM transmission line?

Q2.4 What purpose does the lumped-element circuit model serve? How are the line parameters R' , L' , G' , and C' related to the physical and electromagnetic constitutive properties of the transmission line?

EXERCISE 2.1 Use Table 2-1 to compute the line parameters of a two-wire air line whose wires are separated by a distance of 2 cm, and each is 1 mm in radius. The wires may be treated as perfect conductors with $\sigma_c = \infty$.

Ans. $R' = 0$, $L' = 1.20$ ($\mu\text{H/m}$), $G' = 0$, $C' = 9.29$ (pF/m).

EXERCISE 2.2 Calculate the transmission line parameters at 1 MHz for a rigid coaxial air line with an inner conductor diameter of 0.6 cm and an outer conductor diameter of 1.2 cm. The conductors are made of copper [see Appendix B for μ_c and σ_c of copper].

Ans. $R' = 2.08 \times 10^{-2}$ (Ω/m), $L' = 0.14$ ($\mu\text{H/m}$), $G' = 0$, $C' = 80.3$ (pF/m).

2-3 Transmission-Line Equations

A transmission line usually connects a source on one end to a load on the other end. Before we consider the complete circuit, however, we need to develop equations that describe the voltage across the transmission line and the current carried by the line as a function of time t and spatial position z . Using the lumped-element model described in Fig. 2-6(c), we begin by considering a differential length Δz as shown in Fig. 2-8. The quantities $v(z, t)$ and $i(z, t)$ denote the instantaneous voltage and current at the left end of the differential section (node N), and similarly $v(z + \Delta z, t)$ and $i(z + \Delta z, t)$ denote the same quantities at the right end (node $N + 1$). Application of Kirchhoff's voltage law accounts for the voltage drop across the series resistance $R'\Delta z$ and inductance $L'\Delta z$:

$$v(z, t) - R'\Delta z i(z, t) - L'\Delta z \frac{\partial i(z, t)}{\partial t} - v(z + \Delta z, t) = 0. \quad (2.12)$$

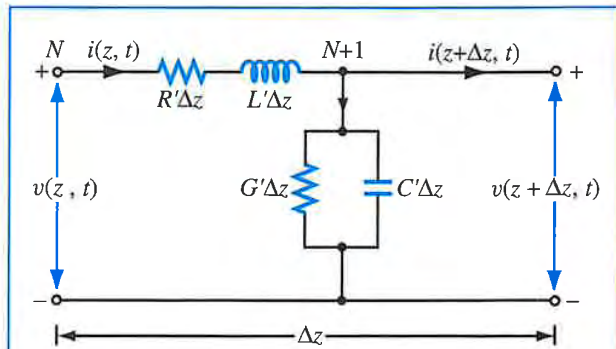


Figure 2-8: Equivalent circuit of a differential length Δz of a two-conductor transmission line.

Upon dividing all terms by Δz and rearranging terms, we obtain

$$-\left[\frac{v(z + \Delta z, t) - v(z, t)}{\Delta z}\right] = R' i(z, t) + L' \frac{\partial i(z, t)}{\partial t}. \quad (2.13)$$

In the limit as $\Delta z \rightarrow 0$, Eq. (2.13) becomes a differential equation:

$$-\frac{\partial v(z, t)}{\partial z} = R' i(z, t) + L' \frac{\partial i(z, t)}{\partial t}. \quad (2.14)$$

Similarly, application of Kirchhoff's current law at node $N + 1$ in Fig. 2-8 leads to

$$i(z, t) - G' \Delta z v(z + \Delta z, t) - C' \Delta z \frac{\partial v(z + \Delta z, t)}{\partial t} - i(z + \Delta z, t) = 0. \quad (2.15)$$

Upon dividing all terms by Δz and taking the limit as $\Delta z \rightarrow 0$, Eq. (2.15) provides a second differential equation,

$$-\frac{\partial i(z, t)}{\partial z} = G' v(z, t) + C' \frac{\partial v(z, t)}{\partial t}. \quad (2.16)$$

The first-order differential equations given by Eqs. (2.14) and (2.16) are the time-domain form of the **transmission line equations**, otherwise called the **telegrapher's equations**.

Except for the last section, our primary interest in this chapter is in sinusoidal steady-state conditions. To this end, we shall make use of phasors with the cosine reference notation as outlined in Section 1-6. Thus, we define

$$v(z, t) = \Re[\tilde{V}(z) e^{j\omega t}], \quad (2.17a)$$

$$i(z, t) = \Re[\tilde{I}(z) e^{j\omega t}], \quad (2.17b)$$

where $\tilde{V}(z)$ and $\tilde{I}(z)$ are phasor quantities, each of which may be real or complex. Upon substituting Eqs. (2.17a) and (2.17b) into Eqs. (2.14) and (2.16) and utilizing the property given by Eq. (1.62) that $\partial/\partial t$ in the time domain

becomes equivalent to multiplication by $j\omega$ in the phasor domain, we obtain the following pair of equations:

$$-\frac{d\tilde{V}(z)}{dz} = (R' + j\omega L') \tilde{I}(z), \quad (2.18a)$$

$$-\frac{d\tilde{I}(z)}{dz} = (G' + j\omega C') \tilde{V}(z). \quad (2.18b)$$

These are the telegrapher's equations in phasor form.

2-4 Wave Propagation on a Transmission Line

The two first-order coupled equations given by Eqs. (2.18a) and (2.18b) can be combined to give two second-order uncoupled wave equations, one for $\tilde{V}(z)$ and another for $\tilde{I}(z)$. The wave equation for $\tilde{V}(z)$ is derived by differentiating both sides of Eq. (2.18a) with respect to z , giving

$$-\frac{d^2\tilde{V}(z)}{dz^2} = (R' + j\omega L') \frac{d\tilde{I}(z)}{dz}, \quad (2.19)$$

and upon substituting Eq. (2.18b) into Eq. (2.19) for $d\tilde{I}(z)/dz$, Eq. (2.19) becomes

$$\frac{d^2\tilde{V}(z)}{dz^2} - (R' + j\omega L')(G' + j\omega C') \tilde{V}(z) = 0, \quad (2.20)$$

or

$$\frac{d^2\tilde{V}(z)}{dz^2} - \gamma^2 \tilde{V}(z) = 0, \quad (2.21)$$

where

$$\gamma = \sqrt{(R' + j\omega L')(G' + j\omega C')}. \quad (2.22)$$

Application of the same steps to Eqs. (2.18a) and (2.18b) but in reverse order, leads to

$$\frac{d^2 \tilde{I}(z)}{dz^2} - \gamma^2 \tilde{I}(z) = 0. \quad (2.23)$$

Equations (2.21) and (2.23) are called **wave equations** for $\tilde{V}(z)$ and $\tilde{I}(z)$, respectively, and γ is called the **complex propagation constant** of the transmission line. As such, γ consists of a real part α , called the **attenuation constant** of the line with units of Np/m, and an imaginary part β , called the **phase constant** of the line with units of rad/m. Thus,

$$\gamma = \alpha + j\beta \quad (2.24)$$

with

$$\begin{aligned} \alpha &= \Re(\gamma) \\ &= \Re\left(\sqrt{(R' + j\omega L')(G' + j\omega C')}\right) \quad (\text{Np/m}), \end{aligned} \quad (2.25a)$$

$$\begin{aligned} \beta &= \Im(\gamma) \\ &= \Im\left(\sqrt{(R' + j\omega L')(G' + j\omega C')}\right) \quad (\text{rad/m}). \end{aligned} \quad (2.25b)$$

In Eqs. (2.25a) and (2.25b), we choose the square-root values that give positive values for α and β . For passive transmission lines, α is either zero or positive. Most transmission lines, and all those considered in this chapter, are of the passive type. The active region of a laser is an example of an active transmission line with a negative α .

The wave equations given by Eqs. (2.21) and (2.23) have traveling wave solutions of the following form:

$$\tilde{V}(z) = V_0^+ e^{-\gamma z} + V_0^- e^{\gamma z} \quad (\text{V}), \quad (2.26a)$$

$$\tilde{I}(z) = I_0^+ e^{-\gamma z} + I_0^- e^{\gamma z} \quad (\text{A}), \quad (2.26b)$$

where, as will be shown later, the $e^{-\gamma z}$ term represents wave propagation in the $+z$ -direction and the $e^{\gamma z}$ term represents wave propagation in the $-z$ -direction. Verification that these are indeed valid solutions is easily accomplished by substituting the expressions given by Eqs. (2.26a) and (2.26b), as well as their second derivatives, into Eqs. (2.21) and (2.23). In their present form, the solutions given by Eqs. (2.26a) and (2.26b) contain four unknowns, the wave amplitudes (V_0^+ , I_0^+) of the $+z$ propagating wave and (V_0^- , I_0^-) of the $-z$ propagating wave. We can easily relate the current wave amplitudes, I_0^+ and I_0^- , to the voltage wave amplitudes, V_0^+ and V_0^- , respectively, by using Eq. (2.26a) in Eq. (2.18a) and then solving for the current $\tilde{I}(z)$ to get the result

$$\tilde{I}(z) = \frac{\gamma}{R' + j\omega L'} [V_0^+ e^{-\gamma z} - V_0^- e^{\gamma z}]. \quad (2.27)$$

Comparison of each term with the corresponding term in the expression given by Eq. (2.26b) leads to the conclusion that

$$\frac{V_0^+}{I_0^+} = Z_0 = \frac{-V_0^-}{I_0^-}, \quad (2.28)$$

where

$$Z_0 = \frac{R' + j\omega L'}{\gamma} = \sqrt{\frac{R' + j\omega L'}{G' + j\omega C'}} \quad (\Omega), \quad (2.29)$$

is defined as the **characteristic impedance** of the line. It should be noted that Z_0 is equal to the ratio of the voltage amplitude to the current amplitude for each of the traveling waves individually (with an additional minus sign in the case of the $-z$ propagating wave), but it is not equal to the ratio of the total voltage $\tilde{V}(z)$ to the total current $\tilde{I}(z)$, unless one of the two waves is absent. In terms of Z_0 , Eq. (2.27) can be rewritten in the form

$$\tilde{I}(z) = \frac{V_0^+}{Z_0} e^{-\gamma z} - \frac{V_0^-}{Z_0} e^{\gamma z}. \quad (2.30)$$

In later sections, we will apply boundary conditions at the load and at the sending end of the transmission line to obtain expressions for the remaining wave amplitudes V_0^+ and V_0^- . In general, each will be a complex quantity composed of a magnitude and a phase angle. Thus,

$$V_0^+ = |V_0^+|e^{j\phi^+}, \quad (2.31a)$$

$$V_0^- = |V_0^-|e^{j\phi^-}. \quad (2.31b)$$

Upon substituting these definitions in Eq. (2.26a) and replacing γ with Eq. (2.24), we can convert back to the time domain to obtain an expression for $v(z, t)$, the instantaneous voltage on the line:

$$\begin{aligned} v(z, t) &= \Re\{\tilde{V}(z)e^{j\omega t}\} \\ &= \Re\{(V_0^+e^{-\gamma z} + V_0^-e^{\gamma z})e^{j\omega t}\} \\ &= \Re[|V_0^+|e^{j\phi^+}e^{j\omega t}e^{-(\alpha+j\beta)z} \\ &\quad + |V_0^-|e^{j\phi^-}e^{j\omega t}e^{(\alpha+j\beta)z}] \\ &= |V_0^+|e^{-\alpha z}\cos(\omega t - \beta z + \phi^+) \\ &\quad + |V_0^-|e^{\alpha z}\cos(\omega t + \beta z + \phi^-). \end{aligned} \quad (2.32)$$

From our review of traveling waves in Section 1-3, we recognize the first term in Eq. (2.32) as a wave traveling in the $+z$ -direction (the coefficients of t and z have opposite signs) and the second term as a wave traveling in the $-z$ -direction (the coefficients of t and z are both positive), both propagating with a phase velocity u_p given by Eq. (1.30):

$$u_p = f\lambda = \frac{\omega}{\beta}. \quad (2.33)$$

The factor $e^{-\alpha z}$ accounts for the attenuation of the $+z$ propagating wave, and the $e^{\alpha z}$ accounts for the attenuation of the $-z$ propagating wave. The presence of two waves on the line propagating in opposite directions produces a **standing wave**. To gain a physical understanding of what this means, we shall first examine the relatively simple but important case of a **lossless line** ($\alpha = 0$) and then extend the results to the more general

case of **lossy transmission lines** ($\alpha \neq 0$). In fact, we shall devote the next several sections to the study of lossless transmission lines because in practice many lines can be designed to exhibit very low-loss characteristics.

Example 2-1 Air Line

An **air line** is a transmission line for which air is the dielectric material present between the two conductors, which renders $G' = 0$. In addition, the conductors are made of a material with high conductivity so that $R' \simeq 0$. For an air line with characteristic impedance of 50Ω and phase constant of 20 rad/m at 700 MHz , find the inductance per meter and the capacitance per meter of the line.

Solution: The following quantities are given:

$$\begin{aligned} Z_0 &= 50 \Omega, & \beta &= 20 \text{ rad/m}, \\ f &= 700 \text{ MHz} = 7 \times 10^8 \text{ Hz}. \end{aligned}$$

With $R' = G' = 0$, Eqs. (2.25b) and (2.29) reduce to

$$\begin{aligned} \beta &= \Im\left(\sqrt{(j\omega L')(j\omega C')}\right) \\ &= \Im\left(j\omega\sqrt{L'C'}\right) = \omega\sqrt{L'C'}, \\ Z_0 &= \sqrt{\frac{j\omega L'}{j\omega C'}} = \sqrt{\frac{L'}{C'}}. \end{aligned}$$

The ratio is given by

$$\frac{\beta}{Z_0} = \omega C',$$

or

$$\begin{aligned} C' &= \frac{\beta}{\omega Z_0} \\ &= \frac{20}{2\pi \times 7 \times 10^8 \times 50} \\ &= 9.09 \times 10^{-11} \text{ (F/m)} = 90.9 \text{ (pF/m)}. \end{aligned}$$

From $Z_0 = \sqrt{L'/C'}$,

$$\begin{aligned} L' &= Z_0^2 C' \\ &= (50)^2 \times 90.9 \times 10^{-12} \\ &= 2.27 \times 10^{-7} \text{ (H/m)} = 227 \text{ (nH/m)}. \quad \blacksquare \end{aligned}$$

EXERCISE 2.3 Verify that Eq. (2.26a) is indeed a solution of the wave equation given by Eq. (2.21).

EXERCISE 2.4 A two-wire air line has the following line parameters: $R' = 0.404 \text{ (m}\Omega/\text{m)}$, $L' = 2.0 \text{ (}\mu\text{H/m)}$, $G' = 0$, and $C' = 5.56 \text{ (pF/m)}$. For operation at 5 kHz, determine (a) the attenuation constant α , (b) the phase constant β , (c) the phase velocity u_p , and (d) the characteristic impedance Z_0 .

Ans. (a) $\alpha = 3.37 \times 10^{-7} \text{ (Np/m)}$, (b) $\beta = 1.05 \times 10^{-4} \text{ (rad/m)}$, (c) $u_p = 3.0 \times 10^8 \text{ (m/s)}$, (d) $Z_0 = (600 - j2.0) \Omega = 600 \angle -0.19^\circ \Omega$.

2-5 The Lossless Transmission Line

According to the preceding section, a transmission line is characterized by two fundamental properties, its propagation constant γ and characteristic impedance Z_0 , both of which are specified by the angular frequency ω and the line parameters R' , L' , G' , and C' . In many practical situations, the transmission line can be designed to minimize ohmic losses by selecting conductors with very high conductivities and dielectric materials (separating the conductors) with negligible conductivities. As a result, R' and G' assume very small values such that $R' \ll \omega L'$ and $G' \ll \omega C'$. These lossless-line conditions allow us to set $R' = G' = 0$ in Eq. (2.22), which then gives the result

$$\gamma = \alpha + j\beta = j\omega\sqrt{L'C'}, \quad (2.34)$$

which means that

$$\begin{aligned} \alpha &= 0 && \text{(lossless line),} \\ \beta &= \omega\sqrt{L'C'} && \text{(lossless line).} \end{aligned} \quad (2.35)$$

Application of the lossless-line conditions to Eq. (2.29) gives the characteristic impedance as

$$Z_0 = \sqrt{\frac{L'}{C'}} \quad \text{(lossless line),} \quad (2.36)$$

which is now a real number. Using the lossless-line expression for β given by Eq. (2.35), we obtain the following relations for the wavelength λ and the phase velocity u_p :

$$\lambda = \frac{2\pi}{\beta} = \frac{2\pi}{\omega\sqrt{L'C'}}, \quad (2.37)$$

$$u_p = \frac{\omega}{\beta} = \frac{1}{\sqrt{L'C'}}. \quad (2.38)$$

Upon using the relation given by Eq. (2.10), which is shared by all TEM transmission lines, Eqs. (2.35) and (2.38) may be rewritten as

$$\beta = \omega\sqrt{\mu\epsilon} \quad \text{(rad/m),} \quad (2.39)$$

$$u_p = \frac{1}{\sqrt{\mu\epsilon}} \quad \text{(m/s),} \quad (2.40)$$

where μ and ϵ are, respectively, the magnetic permeability and electrical permittivity of the insulating material separating the conductors. Materials used for this purpose are usually characterized by a permeability $\mu = \mu_0$, where $\mu_0 = 4\pi \times 10^{-7} \text{ H/m}$ is the permeability of free space, and the permittivity is usually specified in terms of the relative permittivity ϵ_r defined as

$$\epsilon_r = \epsilon/\epsilon_0, \quad (2.41)$$

where $\epsilon_0 = 8.854 \times 10^{-12} \text{ F/m} \simeq (1/36\pi) \times 10^{-9} \text{ F/m}$ is the permittivity of free space. Hence, Eq. (2.40) becomes

$$u_p = \frac{1}{\sqrt{\mu_0 \epsilon_r \epsilon_0}} = \frac{1}{\sqrt{\mu_0 \epsilon_0}} \cdot \frac{1}{\sqrt{\epsilon_r}} = \frac{c}{\sqrt{\epsilon_r}}, \quad (2.42)$$

where $c = 1/\sqrt{\mu_0 \epsilon_0} = 3 \times 10^8 \text{ m/s}$ is the velocity of light in a vacuum. If the insulating material between the conductors is air, then $\epsilon_r = 1$ and $u_p = c$. In view of Eq. (2.41) and the relationship between λ and u_p given by Eq. (2.33), the wavelength is given by

$$\lambda = \frac{u_p}{f} = \frac{c}{f} \frac{1}{\sqrt{\epsilon_r}} = \frac{\lambda_0}{\sqrt{\epsilon_r}}, \quad (2.43)$$

where $\lambda_0 = c/f$ is the wavelength in air corresponding to a frequency f . Note that, because both u_p and λ depend on ϵ_r , the choice of the type of insulating material used in a transmission line is dictated not only by its mechanical properties, but by its electrical properties as well.

When the phase velocity of a medium is independent of frequency, the medium is called **nondispersive**, which clearly is the case for a lossless TEM transmission line. This is an important feature for the transmission of digital data in the form of pulses. A rectangular pulse or a series of pulses is composed of many Fourier components with different frequencies. If the phase velocity is the same for all frequency components (or at least for the dominant ones), the pulse shape will remain the same as the pulse travels on the line. In contrast, the shape of a pulse propagating in a dispersive medium gets progressively distorted, and the pulse length increases (stretches out) as a function of distance in the medium, thereby imposing a limitation on the maximum data rate (which is related to the length of the individual pulses and the spacing between adjacent pulses) that can be transmitted through the medium without loss of information.

Table 2-2 provides a list of the expressions for γ , Z_0 , and u_p for the general case of a lossy line and for several types of lossless lines. The expressions for the lossless

lines are based on the equations for L' and C' given in Table 2-1.

EXERCISE 2.5 For a lossless transmission line, $\lambda = 20.7 \text{ cm}$ at 1 GHz . Find ϵ_r of the insulating material.

Ans. $\epsilon_r = 2.1$.

EXERCISE 2.6 A lossless transmission line uses a dielectric insulating material with $\epsilon_r = 4$. If its line capacitance is $C' = 10 \text{ (pF/m)}$, find (a) the phase velocity u_p , (b) the line inductance L' , and (c) the characteristic impedance Z_0 .

Ans. (a) $u_p = 1.5 \times 10^8 \text{ (m/s)}$, (b) $L' = 4.45 \text{ (}\mu\text{H/m)}$, (c) $Z_0 = 667.1 \text{ }\Omega$.

2-5.1 Voltage Reflection Coefficient

With $\gamma = j\beta$ for the lossless line, the expressions given by Eqs. (2.26a) and (2.30) for the total voltage and current on the line become

$$\tilde{V}(z) = V_0^+ e^{-j\beta z} + V_0^- e^{j\beta z}, \quad (2.44a)$$

$$\tilde{I}(z) = \frac{V_0^+}{Z_0} e^{-j\beta z} - \frac{V_0^-}{Z_0} e^{j\beta z}. \quad (2.44b)$$

These expressions contain two unknowns, V_0^+ and V_0^- , the voltage amplitudes of the incident and reflected waves, respectively. To determine V_0^+ and V_0^- , we need to consider the lossless transmission line in the context of the complete circuit, including a generator circuit at its input terminals and a load at its output terminals, as shown in Fig. 2-9. The line, of length l , is terminated in an arbitrary **load impedance** Z_L . *For convenience, the reference of the spatial coordinate z is chosen such that $z = 0$ corresponds to the location of the load.* At the sending end at $z = -l$, the line is connected to a sinusoidal voltage source with phasor $\tilde{V}_g$ and an internal

Table 2-2: Characteristic parameters of transmission lines.

	Propagation Constant $\gamma = \alpha + j\beta$	Phase Velocity u_p	Characteristic Impedance Z_0
General case	$\gamma = \sqrt{(R' + j\omega L')(G' + j\omega C')}$	$u_p = \omega/\beta$	$Z_0 = \sqrt{\frac{(R' + j\omega L')}{(G' + j\omega C')}}$
Lossless ($R' = G' = 0$)	$\alpha = 0, \beta = \omega\sqrt{\epsilon_r}/c$	$u_p = c/\sqrt{\epsilon_r}$	$Z_0 = \sqrt{L'/C'}$
Lossless coaxial	$\alpha = 0, \beta = \omega\sqrt{\epsilon_r}/c$	$u_p = c/\sqrt{\epsilon_r}$	$Z_0 = (60/\sqrt{\epsilon_r}) \ln(b/a)$
Lossless two wire	$\alpha = 0, \beta = \omega\sqrt{\epsilon_r}/c$	$u_p = c/\sqrt{\epsilon_r}$	$Z_0 = (120/\sqrt{\epsilon_r}) \ln[(d/2a) + \sqrt{(d/2a)^2 - 1}]$ $Z_0 \simeq (120/\sqrt{\epsilon_r}) \ln(d/a),$ if $d \gg a$
Lossless parallel plate	$\alpha = 0, \beta = \omega\sqrt{\epsilon_r}/c$	$u_p = c/\sqrt{\epsilon_r}$	$Z_0 = (120\pi/\sqrt{\epsilon_r}) (d/w)$

Notes: (1) $\mu = \mu_0$, $\epsilon = \epsilon_r \epsilon_0$, $c = 1/\sqrt{\mu_0 \epsilon_0}$, and $\sqrt{\mu_0/\epsilon_0} \simeq (120\pi) \Omega$, where ϵ_r is the relative permittivity of insulating material. (2) For coaxial line, a and b are radii of inner and outer conductors. (3) For two-wire line, a = wire radius and d = separation between wire centers. (4) For parallel-plate line, w = width of plate and d = separation between the plates.

impedance Z_g . At the load, the phasor voltage across it, $\tilde{V}_L$, and the phasor current through it, $\tilde{I}_L$, are related by the load impedance Z_L as follows:

$$Z_L = \frac{\tilde{V}_L}{\tilde{I}_L}. \quad (2.45)$$

The voltage $\tilde{V}_L$ is equal to the total voltage on the line $\tilde{V}(z)$ given by Eq. (2.44a), and $\tilde{I}_L$ is equal to $\tilde{I}(z)$ given by Eq. (2.44b), both evaluated at $z = 0$:

$$\tilde{V}_L = \tilde{V}(z=0) = V_0^+ + V_0^-, \quad (2.46a)$$

$$\tilde{I}_L = \tilde{I}(z=0) = \frac{V_0^+}{Z_0} - \frac{V_0^-}{Z_0}. \quad (2.46b)$$

Upon using these expressions in Eq. (2.45), we obtain the result:

$$Z_L = \left(\frac{V_0^+ + V_0^-}{V_0^+ - V_0^-} \right) Z_0. \quad (2.47)$$

Solving for V_0^- gives

$$V_0^- = \left(\frac{Z_L - Z_0}{Z_L + Z_0} \right) V_0^+. \quad (2.48)$$

The ratio of the amplitude of the reflected voltage wave to the amplitude of the incident voltage wave at the load

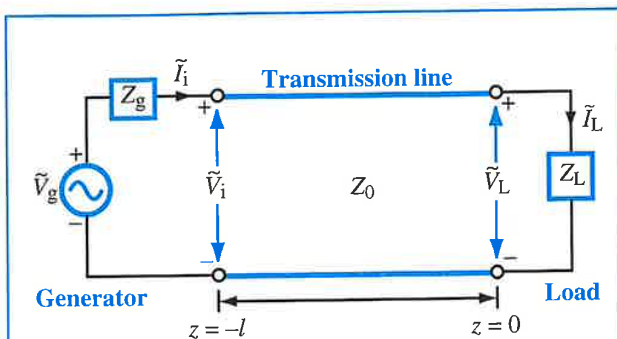


Figure 2-9: Transmission line of length l connected on one end to a generator circuit and on the other end to a load Z_L . The load is located at $z = 0$ and the generator terminals are at $z = -l$.

is known as the **voltage reflection coefficient** Γ . From Eq. (2.48), this definition gives the result

$$\Gamma = \frac{V_0^-}{V_0^+} = \frac{Z_L - Z_0}{Z_L + Z_0} = \frac{Z_L/Z_0 - 1}{Z_L/Z_0 + 1} \quad (\text{dimensionless}), \quad (2.49a)$$

and in view of Eq. (2.28), the ratio of the current amplitudes is

$$\frac{I_0^-}{I_0^+} = -\frac{V_0^-}{V_0^+} = -\Gamma. \quad (2.49b)$$

We note that Γ is governed by a single parameter, the load impedance Z_L , normalized to the characteristic impedance of the line, Z_0 . As indicated by Eq. (2.36), Z_0 of a lossless line is a real number. However, Z_L is in general a complex quantity, as in the case of a series RL

circuit, for example, for which $Z_L = R + j\omega L$. Hence, in general Γ may be complex also:

$$\Gamma = |\Gamma|e^{j\theta_r}, \quad (2.50)$$

where $|\Gamma|$ is the magnitude of Γ and θ_r is its phase angle. Note that $|\Gamma| \leq 1$.

A load is said to be **matched to the line** if $Z_L = Z_0$ because then there will be no reflection by the load ($\Gamma = 0$ and $V_0^- = 0$). On the other hand, when the load is an open circuit ($Z_L = \infty$), $\Gamma = 1$ and $V_0^- = V_0^+$, and when it is a short circuit ($Z_L = 0$), $\Gamma = -1$ and $V_0^- = -V_0^+$.

Example 2-2 Reflection Coefficient of a Series RC Load

A 100- Ω transmission line is connected to a load consisting of a 50- Ω resistor in series with a 10-pF capacitor. Find the reflection coefficient at the load for a 100-MHz signal.

Solution: The following quantities are given [Fig. 2-10]:

$$R_L = 50 \, \Omega, \quad C_L = 10 \, \text{pF} = 10^{-11} \, \text{F},$$

$$Z_0 = 100 \, \Omega, \quad f = 100 \, \text{MHz} = 10^8 \, \text{Hz}.$$

The load impedance is

$$Z_L = R_L - j/\omega C_L = 50 - j \frac{1}{2\pi \times 10^8 \times 10^{-11}} = (50 - j159) \, \Omega.$$

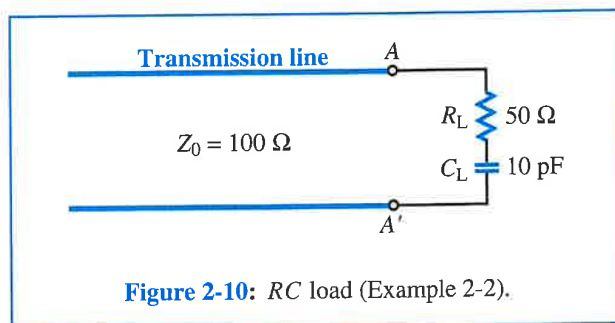


Figure 2-10: RC load (Example 2-2).

From Eq. (2.49a), the voltage reflection coefficient is given by

$$\begin{aligned}\Gamma &= \frac{Z_L/Z_0 - 1}{Z_L/Z_0 + 1} \\ &= \frac{0.5 - j1.59 - 1}{0.5 - j1.59 + 1} \\ &= \frac{-0.5 - j1.59}{1.5 - j1.59} = \frac{-1.67e^{j72.6^\circ}}{2.19e^{-j46.7^\circ}} = -0.76e^{j119.3^\circ}.\end{aligned}$$

This result may be converted into a form with positive magnitude for Γ by replacing the minus sign with e^{-j180° . Thus,

$$\Gamma = 0.76e^{j119.3^\circ}e^{-j180^\circ} = 0.76e^{-j60.7^\circ} = 0.76\angle-60.7^\circ,$$

or

$$|\Gamma| = 0.76, \quad \theta_r = -60.7^\circ. \quad \blacksquare$$

Example 2-3 $|\Gamma|$ for Purely Reactive Load

Show that $|\Gamma| = 1$ for a purely reactive load.

Solution: The load impedance of a purely reactive load is given by

$$Z_L = jX_L.$$

From Eq. (2.49a), the reflection coefficient is

$$\begin{aligned}\Gamma &= \frac{Z_L - Z_0}{Z_L + Z_0} \\ &= \frac{jX_L - Z_0}{jX_L + Z_0} \\ &= \frac{-(Z_0 - jX_L)}{(Z_0 + jX_L)} = \frac{-\sqrt{Z_0^2 + X_L^2}e^{-j\theta}}{\sqrt{Z_0^2 + X_L^2}e^{j\theta}} = -e^{-j2\theta},\end{aligned}$$

where $\theta = \tan^{-1} X_L/Z_0$. Hence

$$|\Gamma| = |-e^{-j2\theta}| = [(e^{-j2\theta})(e^{-j2\theta})^*]^{1/2} = 1. \quad \blacksquare$$

EXERCISE 2.7 A $50\text{-}\Omega$ lossless transmission line is terminated in a load impedance $Z_L = (30 - j200)\text{ }\Omega$. Calculate the voltage reflection coefficient at the load.

Ans. $\Gamma = 0.93\angle-27.5^\circ$.

EXERCISE 2.8 A $150\text{-}\Omega$ lossless line is terminated in a capacitor whose impedance is $Z_L = -j30\text{ }\Omega$. Calculate Γ .

Ans. $\Gamma = 1\angle-157.4^\circ$.

2-5.2 Standing Waves

Using the relation $V_0^- = \Gamma V_0^+$ in Eqs. (2.44a) and (2.44b) gives the expressions

$$\tilde{V}(z) = V_0^+(e^{-j\beta z} + \Gamma e^{j\beta z}), \quad (2.51a)$$

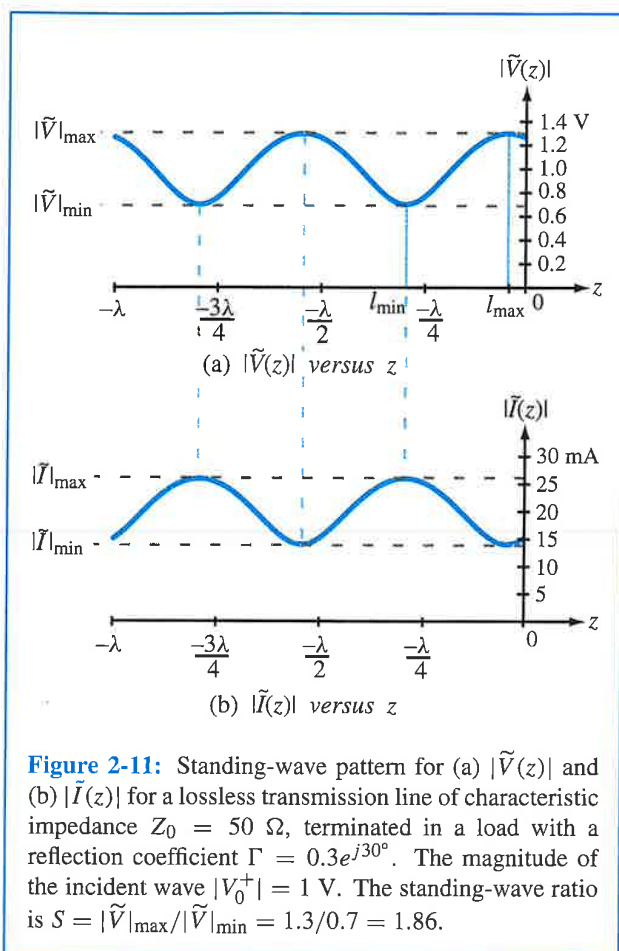
$$\tilde{I}(z) = \frac{V_0^+}{Z_0}(e^{-j\beta z} - \Gamma e^{j\beta z}), \quad (2.51b)$$

which now contain only one, yet to be determined, unknown, V_0^+ . Before we proceed toward that goal, however, let us examine the physical meaning represented by these expressions. We begin by deriving an expression for $|\tilde{V}(z)|$, the magnitude of $\tilde{V}(z)$. Upon using Eq. (2.50) in (2.51a) and applying the relation $|\tilde{V}(z)| = [\tilde{V}(z)\tilde{V}^*(z)]^{1/2}$, where $\tilde{V}^*(z)$ is the complex conjugate of $\tilde{V}(z)$, we have

$$\begin{aligned}|\tilde{V}(z)| &= \{[V_0^+(e^{-j\beta z} + |\Gamma|e^{j\theta_r}e^{j\beta z})] \\ &\quad \cdot [(V_0^+)^*(e^{j\beta z} + |\Gamma|e^{-j\theta_r}e^{-j\beta z})]\}^{1/2} \\ &= |V_0^+| \left[1 + |\Gamma|^2 \right. \\ &\quad \left. + |\Gamma|(e^{j(2\beta z + \theta_r)} + e^{-j(2\beta z + \theta_r)}) \right]^{1/2} \\ &= |V_0^+| [1 + |\Gamma|^2 + 2|\Gamma|\cos(2\beta z + \theta_r)]^{1/2} \quad (2.52)\end{aligned}$$

where we have used the identity

$$e^{jx} + e^{-jx} = 2\cos x \quad (2.53)$$



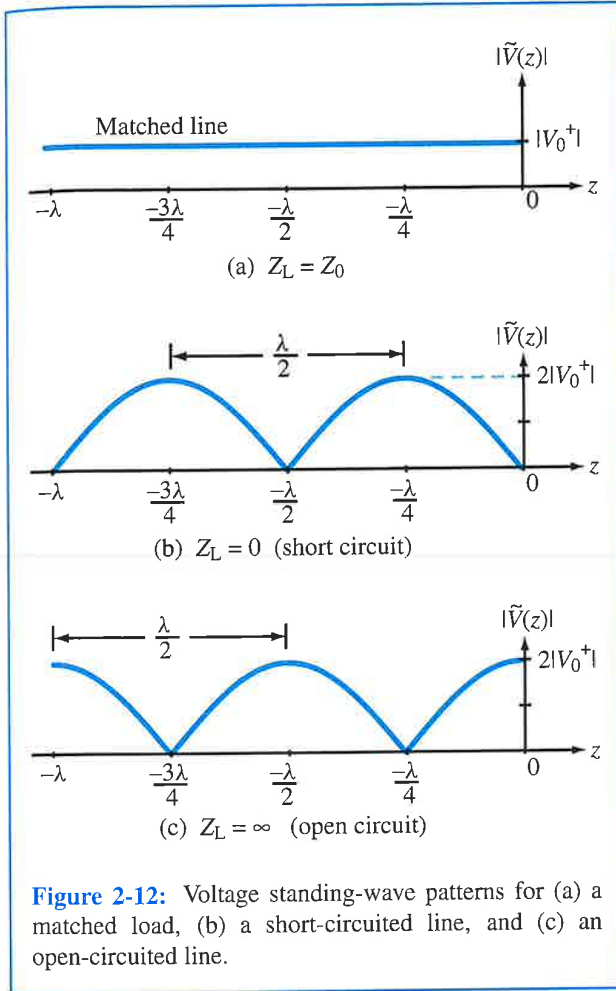
for any real quantity x . By applying the same steps to Eq. (2.51b), a similar expression can be derived for $|\tilde{I}(z)|$, the magnitude of the current $\tilde{I}(z)$.

The variations of $|\tilde{V}(z)|$ and $|\tilde{I}(z)|$ as a function of z , the position on the line relative to the load at $z = 0$, are illustrated in Fig. 2-11 for a line with $|V_0^+| = 1$ V, $|\Gamma| = 0.3$, $\theta_r = 30^\circ$, and $Z_0 = 50 \Omega$. The sinusoidal pattern is called a **standing wave**, and it is caused by the **interference** of the two waves. The maximum value of the **standing-wave pattern** of $|\tilde{V}(z)|$ corresponds

to the position on the line at which the incident and reflected waves are **in phase** [$2\beta z + \theta_r = -2n\pi$ in Eq. (2.52)] and therefore add **constructively** to give a value equal to $(1 + |\Gamma|)|V_0^+| = 1.3$ V. The minimum value of $|\tilde{V}(z)|$ corresponds to destructive interference, which occurs when the incident and reflected waves are in **phase opposition** [$2\beta z + \theta_r = -(2n + 1)\pi$]. In this case, $|\tilde{V}(z)| = (1 - |\Gamma|)|V_0^+| = 0.7$ V. Whereas the repetition period is λ for the incident and reflected waves individually, **the repetition period of the standing-wave pattern is $\lambda/2$** . The standing-wave pattern describes the spatial variation of the magnitude of $\tilde{V}(z)$ as a function of z . If one were to observe the variations of the instantaneous voltage as a function of time at any location z , corresponding to one of the maxima in the standing-wave pattern, for example, that variation would be as $\cos \omega t$ and would have an amplitude equal to 1.3 V [i.e., $v(t)$ would oscillate between -1.3 V and $+1.3$ V]. Similarly, the time oscillation of $v(z, t)$ at any location z will have an amplitude equal to $|\tilde{V}(z)|$ at that z .

Close inspection of the voltage and current standing-wave patterns shown in Fig. 2-11 reveals that the two patterns are in phase opposition (when one is at a maximum, the other is at a minimum, and vice versa). This is a consequence of the fact that the second term in Eq. (2.51a) is preceded by a plus sign, whereas the second term in Eq. (2.51b) is preceded by a negative sign.

The standing-wave patterns shown in Fig. 2-11 are for a typical situation with $\Gamma = 0.3e^{j30^\circ}$. The peak-to-peak variation of the pattern depends on $|\Gamma|$, which can vary between 0 and 1. For the special case of a matched line with $Z_L = Z_0$, we have $|\Gamma| = 0$ and $|\tilde{V}(z)| = |V_0^+|$ for all values of z , as shown in Fig. 2-12(a). **With no reflected wave present, there will be no interference and no standing waves.** The other end of the $|\Gamma|$ scale, at $|\Gamma| = 1$, corresponds to when the load is a short circuit ($\Gamma = -1$) or an open circuit ($\Gamma = 1$). The standing-wave patterns for these two cases are shown in Figs. 2-12(b) and (c), both of which have maxima equal to $2|V_0^+|$ and



minima equal to zero, but the two patterns are shifted in z relative to each other by a distance of $\lambda/4$.

Now let us examine the maximum and minimum values of the voltage magnitude. From Eq. (2.52), $|\tilde{V}(z)|$ is a maximum when the argument of the cosine function is equal to zero or multiples of 2π . Noting that the location on the line always corresponds to negative values of z (since the load is at $z = 0$), if we denote $l_{\max} = -z$ as the distance from the load at which $|\tilde{V}(z)|$ is a maximum,

then

$$|\tilde{V}(z)| = |\tilde{V}|_{\max} = |V_0^+|[1 + |\Gamma|], \quad (2.54)$$

and this occurs when

$$2\beta z + \theta_r = -2\beta l_{\max} + \theta_r = -2n\pi, \quad (2.55)$$

with $n = 0$ or a positive integer. Solving Eq. (2.55) for $l_{\max}$, we have

$$-z = l_{\max} = \frac{\theta_r + 2n\pi}{2\beta} = \frac{\theta_r \lambda}{4\pi} + \frac{n\lambda}{2}, \quad (2.56)$$

$$\begin{cases} n = 1, 2, \dots & \text{if } \theta_r < 0, \\ n = 0, 1, 2, \dots & \text{if } \theta_r \geq 0, \end{cases}$$

where we have used the relation $\beta = 2\pi/\lambda$. The phase angle of the voltage reflection coefficient, θ_r , is bounded between $-\pi$ and π radians. If $\theta_r \geq 0$, the **first voltage maximum** occurs at $l_{\max} = \theta_r \lambda / 4\pi$, corresponding to $n = 0$, but if $\theta_r < 0$, the first physically meaningful maximum occurs at $l_{\max} = (\theta_r \lambda / 4\pi) + \lambda/2$, corresponding to $n = 1$. Negative values of $l_{\max}$ correspond to locations "beyond" the load at the end of the line and therefore have no physical significance. As was mentioned earlier, **the locations on the line corresponding to voltage maxima also correspond to current minima, and vice versa.**

Similarly, the minimum values of $|\tilde{V}(z)|$ occur at distances $l_{\min} = -z$ corresponding to when the argument of the cosine function in Eq. (2.52) is equal to $-(2n+1)\pi$, which gives the result

$$|\tilde{V}|_{\min} = |V_0^+|[1 - |\Gamma|],$$

$$\text{when } (\theta_r - 2\beta l_{\min}) = -(2n+1)\pi, \quad (2.57)$$

with $-\pi \leq \theta_r \leq \pi$. The first minimum corresponds to $n = 0$. The spacing between a maximum $l_{\max}$ and the

adjacent minimum $l_{\min}$ is $\lambda/4$. Hence, the *first minimum* occurs at

$$l_{\min} = \begin{cases} l_{\max} + \lambda/4, & \text{if } l_{\max} < \lambda/4, \\ l_{\max} - \lambda/4, & \text{if } l_{\max} \geq \lambda/4. \end{cases} \quad (2.58)$$

The ratio of $|\tilde{V}|_{\max}$ to $|\tilde{V}|_{\min}$ is called the *voltage standing-wave ratio* S , which from Eqs. (2.54) and (2.57) is given by

$$S = \frac{|\tilde{V}|_{\max}}{|\tilde{V}|_{\min}} = \frac{1 + |\Gamma|}{1 - |\Gamma|} \quad (\text{dimensionless}). \quad (2.59)$$

This quantity, which often is referred to by its acronym, VSWR, or the shorter acronym SWR, provides a measure of the mismatch between the load and the transmission line; for a matched load with $\Gamma = 0$, we get $S = 1$, and for a line with $|\Gamma| = 1$, $S = \infty$.

REVIEW QUESTIONS

Q2.5 The attenuation constant α represents ohmic losses. In view of the model given in Fig. 2-6(c), what should R' and G' be in order to have no losses? Verify your expectation through the expression for α given by Eq. (2.25a).

Q2.6 How is the wavelength λ of the wave traveling on the transmission line related to the free-space wavelength λ_0 ?

Q2.7 When is a load matched to the line? Why is it important?

Q2.8 What is a standing-wave pattern? Why is its period $\lambda/2$ and not λ ?

Q2.9 What is the separation between the location of a voltage maximum and the adjacent current maximum on the line?

Example 2-4 Standing-wave Ratio

A $50\text{-}\Omega$ transmission line is terminated in a load with $Z_L = (100 + j50)\text{ }\Omega$. Find the voltage reflection coefficient and the voltage standing-wave ratio (SWR).

Solution: From Eq. (2.49a), Γ is given by

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0} = \frac{(100 + j50) - 50}{(100 + j50) + 50} = \frac{50 + j50}{150 + j50}.$$

Converting the numerator and denominator to polar form and then simplifying yields

$$\Gamma = \frac{70.7e^{j45^\circ}}{158.1e^{j18.4^\circ}} = 0.45e^{j26.6^\circ}.$$

Using the definition for S given by Eq. (2.59), we have

$$S = \frac{1 + |\Gamma|}{1 - |\Gamma|} = \frac{1 + 0.45}{1 - 0.45} = 2.6. \quad \blacksquare$$

Example 2-5 Measuring Z_L

A *slotted-line* probe is an instrument used to measure the unknown impedance of a load, Z_L . A coaxial slotted line contains a narrow longitudinal slit in the outer conductor. A small probe inserted in the slit can be used to sample the magnitude of the electric field and, hence, the magnitude $|\tilde{V}|$ of the voltage on the line (Fig. 2-13). By moving the probe along the length of the slotted line, it is possible to measure $|\tilde{V}|_{\max}$ and $|\tilde{V}|_{\min}$

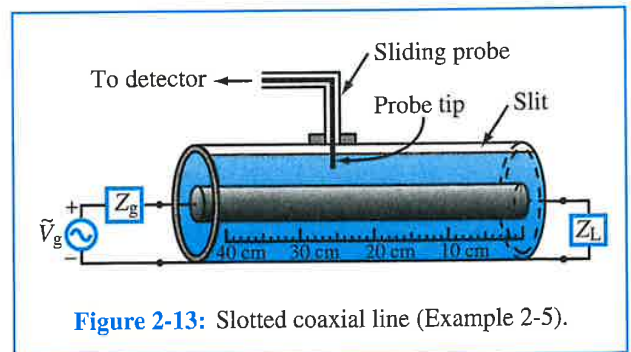


Figure 2-13: Slotted coaxial line (Example 2-5).

and the distances from the load at which they occur. Use of Eq. (2.59) then provides the voltage standing-wave ratio S . Measurements with a 50- Ω slotted line connected to an unknown load impedance determined that $S = 3$. The distance between successive voltage minima was found to be 30 cm, and the first voltage minimum was located at 12 cm from the load. Determine the load impedance Z_L .

Solution: The following quantities are given:

$$Z_0 = 50 \, \Omega, \quad S = 3, \quad l_{\min} = 12 \, \text{cm}.$$

Since the distance between successive voltage minima is equal to $\lambda/2$,

$$\lambda = 2 \times 0.3 = 0.6 \, \text{m},$$

and

$$\beta = \frac{2\pi}{\lambda} = \frac{2\pi}{0.6} = \frac{10\pi}{3} \quad (\text{rad/m}).$$

From Eq. (2.59), solving for $|\Gamma|$ in terms of S gives

$$|\Gamma| = \frac{S - 1}{S + 1} = \frac{3 - 1}{3 + 1} = 0.5.$$

Next, we use the condition given by Eq. (2.57) for the location of a voltage minimum to find θ_r :

$$\theta_r - 2\beta l_{\min} = -\pi, \quad \text{for } n = 0 \text{ (first minimum),}$$

which gives

$$\begin{aligned} \theta_r &= 2\beta l_{\min} - \pi \\ &= 2 \times \frac{10\pi}{3} \times 0.12 - \pi = -0.2\pi \text{ (rad)} = -36^\circ. \end{aligned}$$

Hence,

$$\Gamma = |\Gamma|e^{j\theta_r} = 0.5e^{-j36^\circ} = 0.405 - j0.294.$$

Solving Eq. (2.49a) for Z_L , we have

$$\begin{aligned} Z_L &= Z_0 \left[\frac{1 + \Gamma}{1 - \Gamma} \right] \\ &= 50 \left[\frac{1 + 0.405 - j0.294}{1 - 0.405 + j0.294} \right] = (85 - j67) \, \Omega. \quad \blacksquare \end{aligned}$$

EXERCISE 2.9 If $\Gamma = 0.5\angle -60^\circ$ and $\lambda = 24 \, \text{cm}$, find the locations of the voltage maximum and minimum nearest to the load.

Ans. $l_{\max} = 10 \, \text{cm}$, $l_{\min} = 4 \, \text{cm}$.

EXERCISE 2.10 A 140- Ω lossless line is terminated in a load impedance $Z_L = (280 + j182) \, \Omega$. If $\lambda = 72 \, \text{cm}$, find (a) the reflection coefficient Γ , (b) the voltage standing-wave ratio S , (c) the locations of voltage maxima, and (d) the locations of voltage minima.

Ans. (a) $\Gamma = 0.5\angle 29^\circ$, (b) $S = 3.0$, (c) $l_{\max} = 2.9 \, \text{cm} + n\lambda/2$, (d) $l_{\min} = 20.9 \, \text{cm} + n\lambda/2$, where $n = 0, 1, 2, \dots$

2-6 Input Impedance of the Lossless Line

The standing-wave patterns indicate that for a mismatched line the voltage and current magnitudes are oscillatory with position on the line and in phase opposition with each other. Hence, the voltage to current ratio, called the **input impedance** Z_{in} , must vary with position also. Using Eqs. (2.51a) and (2.51b), Z_{in} is given by

$$\begin{aligned} Z_{\text{in}}(z) &= \frac{\tilde{V}(z)}{\tilde{I}(z)} \\ &= \frac{V_0^+[e^{-j\beta z} + \Gamma e^{j\beta z}]}{V_0^+[e^{-j\beta z} - \Gamma e^{j\beta z}]} Z_0 \\ &= Z_0 \left[\frac{1 + \Gamma e^{j2\beta z}}{1 - \Gamma e^{j2\beta z}} \right] \quad (\Omega). \end{aligned} \quad (2.60)$$

Note that $Z_{in}(z)$ is the ratio of the total voltage (incident- and reflected-wave voltages) to the total current at any point z on the line, in contrast with the characteristic impedance of the line Z_0 , which relates the voltage and current of each of the two waves individually ($Z_0 = V_0^+/I_0^+ = -V_0^-/I_0^-$).

Of particular interest in many transmission-line problems is the input impedance at the input of the line at $z = -l$, which is given by

$$\begin{aligned} Z_{in}(-l) &= Z_0 \left[\frac{e^{j\beta l} + \Gamma e^{-j\beta l}}{e^{j\beta l} - \Gamma e^{-j\beta l}} \right] \\ &= Z_0 \left[\frac{1 + \Gamma e^{-j2\beta l}}{1 - \Gamma e^{-j2\beta l}} \right]. \end{aligned} \quad (2.61)$$

By replacing Γ with Eq. (2.49a) and using the relations

$$e^{j\beta l} = \cos \beta l + j \sin \beta l, \quad (2.62a)$$

$$e^{-j\beta l} = \cos \beta l - j \sin \beta l, \quad (2.62b)$$

Eq. (2.61) can be rewritten in terms of Z_L as

$$\begin{aligned} Z_{in}(-l) &= Z_0 \left(\frac{Z_L \cos \beta l + j Z_0 \sin \beta l}{Z_0 \cos \beta l + j Z_L \sin \beta l} \right) \\ &= Z_0 \left(\frac{Z_L + j Z_0 \tan \beta l}{Z_0 + j Z_L \tan \beta l} \right). \end{aligned} \quad (2.63)$$

From the standpoint of the generator circuit, the transmission line can be replaced with an impedance Z_{in} , as shown in Fig. 2-14. The phasor voltage across Z_{in} is given by

$$\tilde{V}_i = \tilde{I}_i Z_{in} = \frac{\tilde{V}_g Z_{in}}{Z_g + Z_{in}}, \quad (2.64)$$

but from the standpoint of the transmission line, the voltage across it at the input of the line is given by Eq. (2.51a) with $z = -l$:

$$\tilde{V}_i = \tilde{V}(-l) = V_0^+ [e^{j\beta l} + \Gamma e^{-j\beta l}]. \quad (2.65)$$

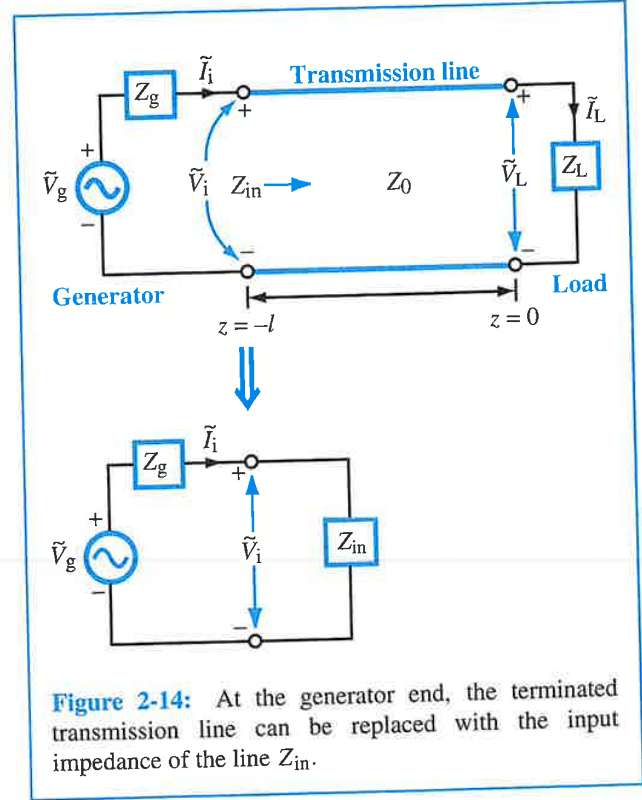


Figure 2-14: At the generator end, the terminated transmission line can be replaced with the input impedance of the line Z_{in} .

Equating Eq. (2.64) to Eq. (2.65) and then solving for V_0^+ leads to the result

$$V_0^+ = \left(\frac{\tilde{V}_g Z_{in}}{Z_g + Z_{in}} \right) \left(\frac{1}{e^{j\beta l} + \Gamma e^{-j\beta l}} \right). \quad (2.66)$$

This completes the solution of the transmission-line wave equations, given by Eqs. (2.21) and (2.23), for the special case of a lossless transmission line. We started out with the general solutions given by Eqs. (2.26a) and (2.26b), which included four unknown amplitudes, V_0^+ , V_0^- , I_0^+ , and I_0^- . We then found out that $Z_0 = V_0^+/I_0^+ = -V_0^-/I_0^-$, thereby reducing the unknowns to the two voltage amplitudes only. Upon

applying the boundary condition at the load, we were able to relate V_0^- to V_0^+ through Γ , and, finally, by applying the boundary condition at the sending end of the line, we obtained an expression for V_0^+ .

Example 2-6 Complete Solution for $v(z, t)$ and $i(z, t)$

A 1.05-GHz generator circuit with series impedance $Z_g = 10 \Omega$ and voltage source given by

$$v_g(t) = 10 \sin(\omega t + 30^\circ) \quad (\text{V})$$

is connected to a load $Z_L = (100 + j50) \Omega$ through a 50- Ω , 67-cm-long lossless transmission line. The phase velocity of the line is $0.7c$, where c is the velocity of light in a vacuum. Find $v(z, t)$ and $i(z, t)$ on the line.

Solution: From the relationship $u_p = \lambda f$, we find the wavelength:

$$\lambda = \frac{u_p}{f} = \frac{0.7 \times 3 \times 10^8}{1.05 \times 10^9} = 0.2 \text{ m},$$

and

$$\begin{aligned} \tan(\beta l) &= \tan\left(\frac{2\pi}{\lambda} l\right) \\ &= \tan\left(\frac{2\pi}{0.2} \times 0.67\right) \\ &= \tan 6.7\pi = \tan 0.7\pi = \tan 126^\circ, \end{aligned}$$

where we have subtracted multiples of 2π . The voltage reflection coefficient at the load is

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0} = \frac{(100 + j50) - 50}{(100 + j50) + 50} = 0.45e^{j26.6^\circ}.$$

With reference to Fig. 2-14, the input impedance of the line, given by Eq. (2.63), is

$$\begin{aligned} Z_{\text{in}} &= Z_0 \left[\frac{Z_L + jZ_0 \tan \beta l}{Z_0 + jZ_L \tan \beta l} \right] \\ &= Z_0 \left[\frac{Z_L/Z_0 + j \tan \beta l}{1 + j(Z_L/Z_0) \tan \beta l} \right] \\ &= 50 \left[\frac{(2 + j1) + j \tan 126^\circ}{1 + j(2 + j1) \tan 126^\circ} \right] = (21.9 + j17.4) \Omega. \end{aligned}$$

Rewriting the expression for the generator voltage with the cosine reference, we have

$$\begin{aligned} v_g(t) &= 10 \sin(\omega t + 30^\circ) \\ &= 10 \cos(\pi/2 - \omega t - 30^\circ) \\ &= 10 \cos(\omega t - 60^\circ) \\ &= \Re[10e^{-j60^\circ} e^{j\omega t}] = \Re[\tilde{V}_g e^{j\omega t}] \quad (\text{V}). \end{aligned}$$

Hence, the phasor voltage $\tilde{V}_g$ is given by

$$\tilde{V}_g = 10 e^{-j60^\circ} \text{ (V)} = 10 \angle -60^\circ \quad (\text{V}).$$

Application of Eq. (2.66) gives

$$\begin{aligned} V_0^+ &= \left(\frac{\tilde{V}_g Z_{\text{in}}}{Z_g + Z_{\text{in}}} \right) \left(\frac{1}{e^{j\beta l} + \Gamma e^{-j\beta l}} \right) \\ &= \left[\frac{10e^{-j60^\circ} (21.9 + j17.4)}{10 + 21.9 + j17.4} \right] \\ &\quad \cdot [e^{j126^\circ} + 0.45e^{j26.6^\circ} e^{-j126^\circ}]^{-1} \\ &= 10.2e^{j159^\circ} \text{ (V)} = 10.2 \angle 159^\circ \quad (\text{V}). \end{aligned}$$

The phasor voltage on the line is then

$$\begin{aligned} \tilde{V}(z) &= V_0^+ (e^{-j\beta z} + \Gamma e^{j\beta z}) \\ &= 10.2e^{j159^\circ} (e^{-j\beta z} + 0.45e^{j26.6^\circ} e^{j\beta z}), \end{aligned}$$

and the instantaneous voltage $v(z, t)$ is

$$\begin{aligned} v(z, t) &= \Re[\tilde{V}(z)e^{j\omega t}] \\ &= 10.2 \cos(\omega t - \beta z + 159^\circ) \\ &\quad + 4.55 \cos(\omega t + \beta z + 185.6^\circ) \quad (\text{V}). \end{aligned}$$

Similarly, use of V_0^+ in Eq. (2.51b) leads to

$$\begin{aligned} \tilde{I}(z) &= 0.20e^{j159^\circ}(e^{-j\beta z} - 0.45e^{j26.6^\circ}e^{j\beta z}), \\ i(z, t) &= 0.20 \cos(\omega t - \beta z + 159^\circ) \\ &\quad + 0.091 \cos(\omega t + \beta z + 5.6^\circ) \quad (\text{A}). \quad \blacksquare \end{aligned}$$

2-7 Special Cases of the Lossless Line

We often encounter situations involving lossless transmission lines with particular terminations or lines whose lengths exhibit particularly useful properties. We shall now consider some of these special cases.

2-7.1 Short-Circuited Line

The transmission line shown in Fig. 2-15(a) is terminated in a short circuit, $Z_L = 0$. Consequently, the voltage reflection coefficient defined by Eq. (2.49a) is $\Gamma = -1$, and the voltage standing-wave ratio given by Eq. (2.59) is $S = \infty$. From Eqs. (2.51a) and (2.51b), the voltage and current on a short-circuited lossless transmission line are given by

$$\tilde{V}_{sc}(z) = V_0^+[e^{-j\beta z} - e^{j\beta z}] = -2jV_0^+ \sin \beta z, \quad (2.67a)$$

$$\tilde{I}_{sc}(z) = \frac{V_0^+}{Z_0}[e^{-j\beta z} + e^{j\beta z}] = \frac{2V_0^+}{Z_0} \cos \beta z. \quad (2.67b)$$

The voltage $\tilde{V}_{sc}(z)$ is zero at the load ($z = 0$), as it should be for a short circuit, and its amplitude varies as $\sin \beta z$, whereas the current $\tilde{I}_{sc}(z)$ is a maximum at the load and it varies as $\cos \beta z$. Both quantities are displayed in Fig. 2-15 as a function of negative z .

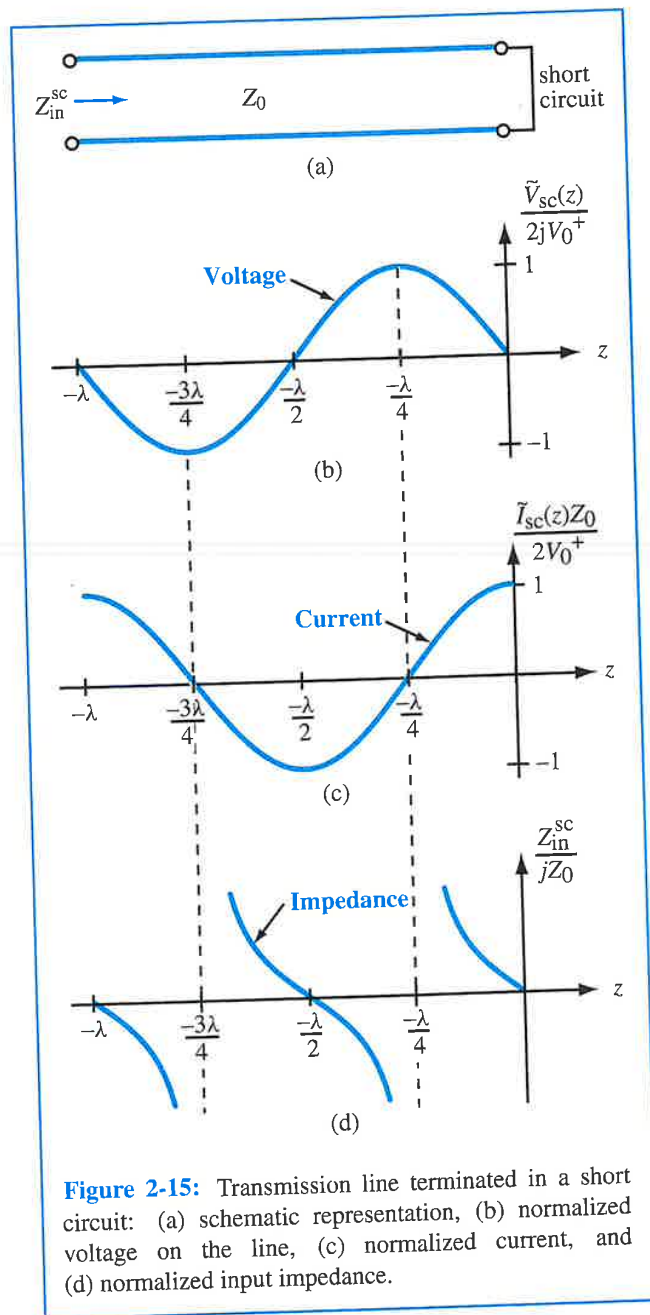


Figure 2-15: Transmission line terminated in a short circuit: (a) schematic representation, (b) normalized voltage on the line, (c) normalized current, and (d) normalized input impedance.

The input impedance of the line at $z = -l$ is given by the ratio of $\tilde{V}_{sc}(-l)$ to $\tilde{I}_{sc}(-l)$. Denoting Z_{in}^{sc} as the input impedance for a short-circuited line, we have

$$Z_{in}^{sc} = \frac{\tilde{V}_{sc}(-l)}{\tilde{I}_{sc}(-l)} = jZ_0 \tan \beta l. \quad (2.68)$$

A plot of Z_{in}^{sc}/jZ_0 versus negative z is shown in Fig. 2-15(d).

In general, the input impedance Z_{in} may consist of a real part, or input resistance R_{in} , and an imaginary part, or input reactance X_{in} :

$$Z_{in} = R_{in} + jX_{in}. \quad (2.69)$$

In the case of the short-circuited lossless line, the input impedance is purely reactive ($R_{in} = 0$). If $\tan \beta l \geq 0$, the line appears inductive, acting like an equivalent inductor L_{eq} whose impedance is equal to Z_{in}^{sc} . Thus,

$$j\omega L_{eq} = jZ_0 \tan \beta l, \quad \text{if } \tan \beta l \geq 0, \quad (2.70a)$$

or

$$L_{eq} = \frac{Z_0 \tan \beta l}{\omega} \quad (\text{H}). \quad (2.70b)$$

The minimum line length l that would result in an input impedance Z_{in}^{sc} equivalent to that of an inductor of inductance L_{eq} is

$$l = \frac{1}{\beta} \tan^{-1} \left(\frac{\omega L_{eq}}{Z_0} \right) \quad (\text{m}). \quad (2.70c)$$

Similarly, if $\tan \beta l \leq 0$, the input impedance is capacitive, in which case the line acts like an equivalent capacitor C_{eq} such that

$$\frac{1}{j\omega C_{eq}} = jZ_0 \tan \beta l, \quad \text{if } \tan \beta l \leq 0, \quad (2.71a)$$

or

$$C_{eq} = -\frac{1}{Z_0 \omega \tan \beta l} \quad (\text{F}). \quad (2.71b)$$

Since l is a positive number, the shortest length l for which $\tan \beta l \leq 0$ corresponds to the range $\pi/2 \leq \beta l \leq \pi$. Hence, the minimum line length l that would result in an input impedance Z_{in}^{sc} equivalent to that of a capacitor of capacitance C_{eq} is

$$l = \frac{1}{\beta} \left[\pi - \tan^{-1} \left(\frac{1}{\omega C_{eq} Z_0} \right) \right] \quad (\text{m}). \quad (2.71c)$$

These results mean that, through proper choice of the length of a short-circuited line, we can make substitutes for capacitors and inductors with any desired reactance. Such a practice is indeed common in the design of microwave circuits and high-speed integrated circuits, because making an actual capacitor or inductor is often more difficult than making a shorted transmission line.

Example 2-7 Equivalent Reactive Elements

Choose the length of a shorted $50\text{-}\Omega$ lossless transmission line (Fig. 2-16) such that its input impedance at 2.25 GHz is equivalent to the reactance of a capacitor with capacitance $C_{eq} = 4\text{ pF}$. The wave velocity on the line is $0.75c$.

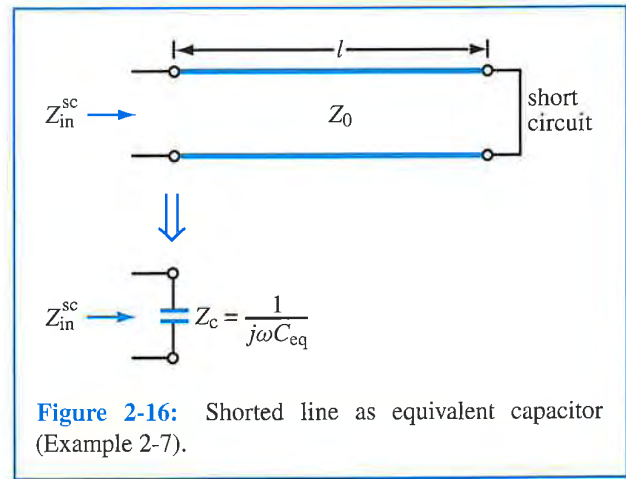


Figure 2-16: Shorted line as equivalent capacitor (Example 2-7).

Solution: We are given

$$u_p = 0.75c = 0.75 \times 3 \times 10^8 = 2.25 \times 10^8 \text{ m/s,}$$

$$Z_0 = 50 \Omega,$$

$$f = 2.25 \text{ GHz} = 2.25 \times 10^9 \text{ Hz,}$$

$$C_{eq} = 4 \text{ pF} = 4 \times 10^{-12} \text{ F.}$$

The phase constant is

$$\beta = \frac{2\pi}{\lambda} = \frac{2\pi f}{u_p} = \frac{2\pi \times 2.25 \times 10^9}{2.25 \times 10^8} = 62.8 \text{ (rad/m).}$$

From Eq. (2.71a),

$$\begin{aligned} \tan \beta l &= -\frac{1}{Z_0 \omega C_{eq}} \\ &= -\frac{1}{50 \times 2\pi \times 2.25 \times 10^9 \times 4 \times 10^{-12}} = -0.354. \end{aligned}$$

The tangent function is negative when its argument is in the second or fourth quadrants. The solution for the second quadrant is

$$\beta l_1 = 2.8 \text{ rad} \quad \text{or} \quad l_1 = \frac{2.8}{\beta} = \frac{2.8}{62.8} = 4.46 \text{ cm,}$$

and the solution for the fourth quadrant is

$$\beta l_2 = 5.94 \text{ rad} \quad \text{or} \quad l_2 = \frac{5.94}{62.8} = 9.46 \text{ cm.}$$

We also could have obtained the value of l_1 by applying Eq. (2.71c). The length l_2 is greater than l_1 by exactly $\lambda/2$. In fact, any length $l = 4.46 \text{ cm} + n\lambda/2$, where n is a positive integer, is also a solution. ■

2-7.2 Open-Circuited Line

With $Z_L = \infty$, as illustrated in Fig. 2-17(a), we have $\Gamma = 1$, $S = \infty$, and the voltage, current, and input impedance are given by

$$\tilde{V}_{oc}(z) = V_0^+[e^{-j\beta z} + e^{j\beta z}] = 2V_0^+ \cos \beta z, \quad (2.72a)$$

$$\tilde{I}_{oc}(z) = \frac{V_0^+}{Z_0}[e^{-j\beta z} - e^{j\beta z}] = \frac{-2jV_0^+}{Z_0} \sin \beta z, \quad (2.72b)$$

$$Z_{in}^{oc} = \frac{\tilde{V}_{oc}(-l)}{\tilde{I}_{oc}(-l)} = -jZ_0 \cot \beta l. \quad (2.73)$$

Plots of these quantities are displayed in Fig. 2-17 as a function of negative z .

2-7.3 Application of Short-Circuit and Open-Circuit Measurements

A network analyzer is a radio-frequency (RF) instrument capable of measuring the impedance of any load connected to its input terminal. When used to measure Z_{in}^{sc} , the input impedance of a lossless line terminated in a short circuit, and again Z_{in}^{oc} , the input impedance of the line when terminated in an open circuit, the combination of the two measurements can be used to determine the characteristic impedance of the line Z_0 and its phase constant β . The product of Eqs. (2.68) and (2.73) gives the result

$$Z_0 = \sqrt{Z_{in}^{sc} Z_{in}^{oc}}, \quad (2.74)$$

and the ratio of the same equations leads to

$$\tan \beta l = \sqrt{\frac{-Z_{in}^{sc}}{Z_{in}^{oc}}}. \quad (2.75)$$

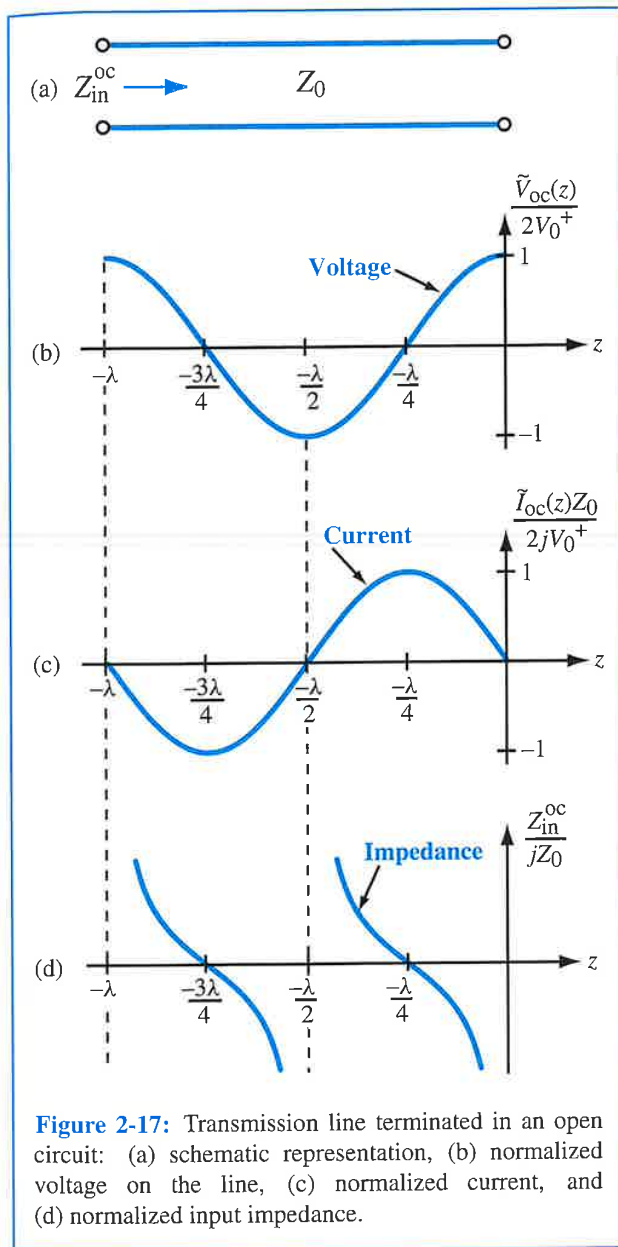


Figure 2-17: Transmission line terminated in an open circuit: (a) schematic representation, (b) normalized voltage on the line, (c) normalized current, and (d) normalized input impedance.

Because of the π phase ambiguity associated with the tangent function, the length l should be less than or equal to $\lambda/2$ to provide an unambiguous result.

Example 2-8 Measuring Z_0 and β

Find Z_0 and β of a 57-cm-long lossless transmission line whose input impedance was measured as $Z_{in}^{sc} = j40.42 \, \Omega$ when terminated in a short circuit and as $Z_{in}^{oc} = -j121.24 \, \Omega$ when terminated in an open circuit. From other measurements, we know that the line is between 3 and 3.25 wavelengths long.

Solution: From Eqs. (2.74) and (2.75),

$$Z_0 = \sqrt{Z_{in}^{sc} Z_{in}^{oc}} = \sqrt{(j40.42)(-j121.24)} = 70 \, \Omega,$$

$$\tan \beta l = \sqrt{\frac{-Z_{in}^{sc}}{Z_{in}^{oc}}} = \sqrt{\frac{1}{3}}.$$

Since l is between 3λ and 3.25λ , $\beta l = (2\pi l/\lambda)$ is between 6π radians and $(13\pi/2)$ radians. This places βl in the first quadrant (0 to $\pi/2$) in a polar coordinate system. Hence, the only acceptable solution for the above equation is $\beta l = \pi/6$ radians. This value, however, does not include the 2π multiples associated with the integer λ multiples of l . Hence, the true value of βl is

$$\beta l = 6\pi + \frac{\pi}{6} = 19.4 \quad (\text{rad}),$$

in which case

$$\beta = \frac{19.4}{0.57} = 34 \quad (\text{rad/m}). \quad \blacksquare$$

2-7.4 Lines of Length $l = n\lambda/2$

If $l = n\lambda/2$, where n is an integer,

$$\tan \beta l = \tan [(2\pi/\lambda)(n\lambda/2)] = \tan n\pi = 0.$$

Consequently, Eq. (2.63) reduces to

$$Z_{in} = Z_L, \quad \text{for } l = n\lambda/2, \quad (2.76)$$

which means that a half-wavelength line (or any integer multiple of $\lambda/2$) does not modify the load impedance. Thus, *a generator connected to a load through a half-wavelength lossless line would induce the same voltage across the load and current through it as when the line is not there.*

2-7.5 Quarter-Wave Transformer

Another case of interest is when the length of the line is a quarter-wavelength (or $\lambda/4 + n\lambda/2$, where $n = 0$ or a positive integer), corresponding to $\beta l = (2\pi/\lambda)(\lambda/4) = \pi/2$. From Eq. (2.63), the input impedance becomes

$$Z_{in} = \frac{Z_0^2}{Z_L}, \quad \text{for } l = \lambda/4 + n\lambda/2. \quad (2.77)$$

The utility of such a quarter-wave transformer is illustrated by the problem in Example 2-9.

Example 2-9 Quarter-Wave Transformer

A 50- Ω lossless transmission line is to be matched to a resistive load impedance with $Z_L = 100 \Omega$ via a quarter-wave section as shown in Fig. 2-18, thereby eliminating reflections along the feedline. Find the characteristic impedance of the quarter-wave transformer.

Solution: To eliminate reflections at terminal AA', the input impedance Z_{in} looking into the quarter-wave line

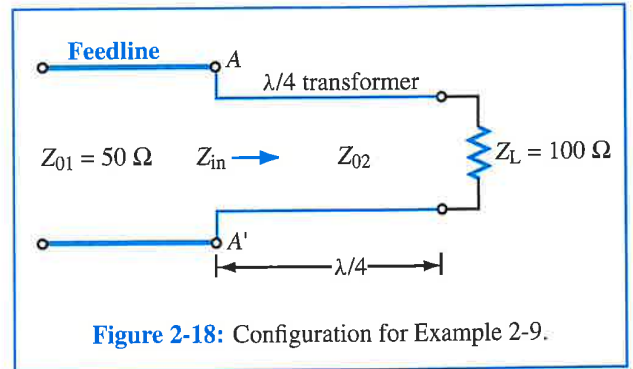


Figure 2-18: Configuration for Example 2-9.

should be equal to Z_{01} , the characteristic impedance of the feedline. Thus, $Z_{in} = 50 \Omega$. From Eq. (2.77),

$$Z_{in} = \frac{Z_{02}^2}{Z_L},$$

or

$$Z_{02} = \sqrt{Z_{in} Z_L} = \sqrt{50 \times 100} = 70.7 \Omega.$$

Whereas this eliminates reflections on the feedline, it does not eliminate them on the $\lambda/4$ line. However, since the lines are lossless, all the incident power will end up getting transferred into the load Z_L . ■

2-7.6 Matched Transmission Line: $Z_L = Z_0$

For a matched lossless transmission line with $Z_L = Z_0$, (1) the input impedance $Z_{in} = Z_0$ for all locations z on the line, (2) $\Gamma = 0$, and (3) all the incident power is delivered to the load, regardless of the line length l . A summary of the properties of standing waves is given in Table 2-3.

REVIEW QUESTIONS

Q2.10 What is the difference between the characteristic impedance Z_0 and the input impedance Z_{in} ? When are they the same?

Table 2-3: Properties of standing waves on a lossless transmission line.

Voltage maximum	$ \tilde{V} _{\max} = V_0^+ [1 + \Gamma]$
Voltage minimum	$ \tilde{V} _{\min} = V_0^+ [1 - \Gamma]$
Positions of voltage maxima (also positions of current minima)	$l_{\max} = \frac{\theta_r \lambda}{4\pi} + \frac{n\lambda}{2}, \quad n = 0, 1, 2, \dots$
Position of first maximum (also position of first current minimum)	$l_{\max} = \begin{cases} \frac{\theta_r \lambda}{4\pi}, & \text{if } 0 \leq \theta_r \leq \pi \\ \frac{\theta_r \lambda}{4\pi} + \frac{\lambda}{2}, & \text{if } -\pi \leq \theta_r \leq 0 \end{cases}$
Positions of voltage minima (also positions of first current maxima)	$l_{\min} = \frac{\theta_r \lambda}{4\pi} + \frac{(2n+1)\lambda}{4}, \quad n = 0, 1, 2, \dots$
Position of first minimum (also position of first current maximum)	$l_{\min} = \frac{\lambda}{4} \left(1 + \frac{\theta_r}{\pi} \right)$
Input impedance	$Z_{\text{in}} = Z_0 \left(\frac{Z_L + jZ_0 \tan \beta l}{Z_0 + jZ_L \tan \beta l} \right)$
Positions at which Z_{in} is real	at voltage maxima and minima
Z_{in} at voltage maxima	$Z_{\text{in}} = Z_0 \left(\frac{1 + \Gamma }{1 - \Gamma } \right)$
Z_{in} at voltage minima	$Z_{\text{in}} = Z_0 \left(\frac{1 - \Gamma }{1 + \Gamma } \right)$
Z_{in} of short-circuited line	$Z_{\text{in}}^{\text{sc}} = jZ_0 \tan \beta l$
Z_{in} of open-circuited line	$Z_{\text{in}}^{\text{oc}} = -jZ_0 \cot \beta l$
Z_{in} of line of length $l = n\lambda/2$	$Z_{\text{in}} = Z_L, \quad n = 0, 1, 2, \dots$
Z_{in} of line of length $l = \lambda/4 + n\lambda/2$	$Z_{\text{in}} = Z_0^2/Z_L, \quad n = 0, 1, 2, \dots$
Z_{in} of matched line	$Z_{\text{in}} = Z_0$
$ V_0^+ $ = amplitude of incident wave, $\Gamma = \Gamma e^{j\theta_r}$ with $-\pi < \theta_r < \pi$; θ_r in radians.	

Q2.11 What is a quarter-wave transformer? How is it used?

Q2.12 A lossless transmission line of length l is terminated in a short circuit. If $l < \lambda/4$, is the input impedance inductive or capacitive?

Q2.13 What is the input impedance of an infinitely long line?

Q2.14 If the input impedance of a lossless line is inductive when terminated in a short circuit, will it be inductive or capacitive when the line is terminated in an open circuit?

EXERCISE 2.11 A $50\text{-}\Omega$ lossless transmission line uses an insulating material with $\epsilon_r = 2.25$. When terminated in an open circuit, how long should the line be for its input impedance to be equivalent to a 10-pF capacitor at 50 MHz ?

Ans. $l = 9.92\text{ cm}$.

EXERCISE 2.12 A $300\text{-}\Omega$ feedline is to be connected to a 3-m long, $150\text{-}\Omega$ line terminated in a $150\text{-}\Omega$ resistor. Both lines are lossless and use air as the insulating material, and the operating frequency is 50 MHz . Determine (a) the input impedance of the 3-m long line, (b) the voltage standing-wave ratio on the feedline, and (c) the characteristic impedance of a quarter-wave transformer were it to be used between the two lines in order to achieve $S = 1$ on the feedline.

Ans. (a) $Z_{\text{in}} = 150\text{ }\Omega$, (b) $S = 2$, (c) $Z_0 = 212.1\text{ }\Omega$.

2-8 Power Flow on a Lossless Transmission Line

Our discussion thus far has focused on the voltage and current aspects of wave propagation on a transmission line. Now we shall examine the flow of power carried by the incident and reflected waves. We begin by

reintroducing Eqs. (2.51a) and (2.51b), the general expressions for the voltage and current phasors on a lossless transmission line:

$$\tilde{V}(z) = V_0^+(e^{-j\beta z} + \Gamma e^{j\beta z}), \quad (2.78a)$$

$$\tilde{I}(z) = \frac{V_0^+}{Z_0}(e^{-j\beta z} - \Gamma e^{j\beta z}), \quad (2.78b)$$

In these expressions, the first terms represent the incident-wave voltage and current, and the terms involving Γ represent the reflected-wave voltage and current. At the load ($z = 0$), the incident and reflected voltages and currents are

$$\tilde{V}^i = V_0^+, \quad \tilde{I}^i = \frac{V_0^+}{Z_0}, \quad (\text{at } z = 0), \quad (2.79)$$

$$\tilde{V}^r = \Gamma V_0^+, \quad \tilde{I}^r = -\Gamma \frac{V_0^+}{Z_0}, \quad (\text{at } z = 0). \quad (2.80)$$

2-8.1 Instantaneous Power

The instantaneous power carried by the incident wave, *as it arrives at the load*, is equal to the product of the instantaneous voltage $v^i(t)$ and the instantaneous current $i^i(t)$,

$$\begin{aligned} P^i(t) &= v^i(t) \cdot i^i(t) \\ &= \Re[\tilde{V}^i e^{j\omega t}] \cdot \Re[\tilde{I}^i e^{j\omega t}] \\ &= \Re[|V_0^+| e^{j\phi^+} e^{j\omega t}] \cdot \Re\left[\frac{|V_0^+|}{Z_0} e^{j\phi^+} e^{j\omega t}\right] \\ &= |V_0^+| \cos(\omega t + \phi^+) \cdot \frac{|V_0^+|}{Z_0} \cos(\omega t + \phi^+) \\ &= \frac{|V_0^+|^2}{Z_0} \cos^2(\omega t + \phi^+) \quad (\text{W}), \end{aligned} \quad (2.81)$$

where use was made of Eq. (2.31a) to express V_0^+ in terms of its magnitude $|V_0^+|$ and phase angle ϕ^+ .

Similarly, upon replacing Γ in Eq. (2.80) with $|\Gamma|e^{j\theta_r}$ and then following the same steps, we obtain

the following expression for the instantaneous power reflected by the load:

$$P^r(t) = v^r(t) \cdot i^r(t) \\ = -|\Gamma|^2 \frac{|V_0^+|^2}{Z_0} \cos^2(\omega t + \phi^+ + \theta_r) \quad (\text{W}). \quad (2.82)$$

The negative sign in Eq. (2.82) signifies the fact that the reflected power flows in the $-z$ -direction.

2-8.2 Time-Average Power

From a practical standpoint, we usually are more interested in the time-average power flow along the transmission line, P_{av} , than in the instantaneous power $P(t)$. To compute P_{av} , we can use a time-domain approach or a computationally simpler phasor-domain approach. For completeness, we will consider both.

Time-Domain Approach

The time-average power flow is equal to the instantaneous power averaged over one time period $T = 1/f = 2\pi/\omega$. For the incident wave, the time-average power is

$$P_{av}^i = \frac{1}{T} \int_0^T P^i(t) dt = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} P^i(t) dt. \quad (2.83)$$

Upon inserting Eq. (2.81) for the incident power $P^i(t)$ into Eq. (2.83) and performing the integration, we obtain the result

$$P_{av}^i = \frac{|V_0^+|^2}{2Z_0} \quad (\text{W}). \quad (2.84)$$

The factor of $1/2$ is a consequence of the integration of $\cos^2(\omega t + \phi^+)$ over one period. A similar treatment for the reflected wave gives

$$P_{av}^r = -|\Gamma|^2 \frac{|V_0^+|^2}{2Z_0} = -|\Gamma|^2 P_{av}^i \quad (2.85)$$

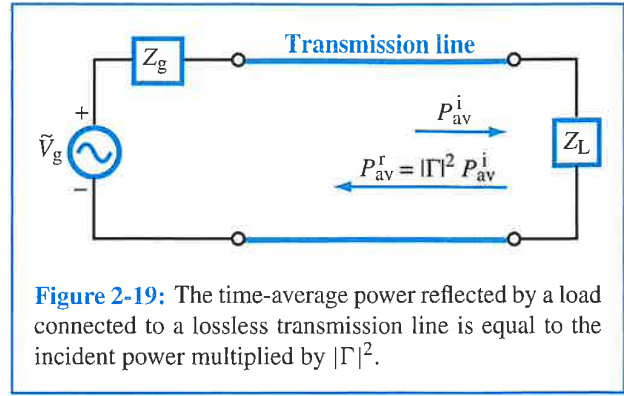


Figure 2-19: The time-average power reflected by a load connected to a lossless transmission line is equal to the incident power multiplied by $|\Gamma|^2$.

Thus, the magnitude of the average reflected power is equal to the average incident power, reduced by a multiplicative factor of $|\Gamma|^2$.

The **net average power delivered to the load** shown in Fig. 2-19 is

$$P_{av} = P_{av}^i + P_{av}^r = \frac{|V_0^+|^2}{2Z_0} [1 - |\Gamma|^2] \quad (\text{W}). \quad (2.86)$$

Phasor-Domain Approach

For any propagating wave with voltage and current phasors $\tilde{V}$ and $\tilde{I}$, a useful formula for computing the time-average power is

$$P_{av} = \frac{1}{2} \Re[\tilde{V} \cdot \tilde{I}^*], \quad (2.87)$$

where $\tilde{I}^*$ is the complex conjugate of $\tilde{I}$. Application of this formula to Eqs. (2.79) and (2.80) gives

$$P_{av}^i = \frac{1}{2} \Re \left[V_0^+ \cdot \frac{V_0^{+*}}{Z_0} \right] = \frac{|V_0^+|^2}{2Z_0}, \quad (2.88)$$

$$P_{av}^r = \frac{1}{2} \Re \left[\Gamma V_0^+ \cdot \left(\frac{-\Gamma^* V_0^{+*}}{Z_0} \right) \right]$$

$$= -|\Gamma|^2 \frac{|V_0^+|^2}{2Z_0}, \quad (2.89)$$

which are, respectively, identical to the expressions given by Eqs. (2.84) and (2.85).

EXERCISE 2.13 For a 50- Ω lossless transmission line terminated in a load impedance $Z_L = (100 + j50) \Omega$, determine the fraction of the average incident power reflected by the load.

Ans. 20%.

EXERCISE 2.14 For the line of Exercise 2.13, what is the magnitude of the average reflected power if $|V_0^+| = 1 \text{ V}$?

Ans. $P_{\text{av}}^r = 2 \text{ (mW)}$.

REVIEW QUESTIONS

Q2.15 According to Eq. (2.82), the instantaneous value of the reflected power depends on the phase of the reflection coefficient θ_r , but the average reflected power given by Eq. (2.85) does not. Explain.

Q2.16 What is the average power delivered by a lossless transmission line to a reactive load?

Q2.17 What fraction of the incident power is delivered to a matched load?

Q2.18 Verify that

$$\frac{1}{T} \int_0^T \cos^2 \left(\frac{2\pi t}{T} + \phi \right) dt = \frac{1}{2},$$

regardless of the value of ϕ .

2-9 The Smith Chart

Prior to the age of computers and programmable calculators, several types of charts were developed to assist in the solution of transmission-line problems. The **Smith chart**, which was developed by P. H. Smith in 1939, has been and continues to be the most widely used graphical technique for analyzing and designing transmission-line circuits. Even though the original intent of its inventor was to provide a useful graphical tool for performing calculations involving complex impedances, the Smith chart has become a principal presentation medium in computer-aided design (CAD) software for displaying the performance of microwave circuits. As the material in this and the next section will demonstrate, use of the Smith chart not only avoids tedious manipulations of complex numbers, but it also allows an engineer to design impedance-matching circuits with relative ease. The Smith chart can be used for both lossy and lossless transmission lines. In the present treatment, however, we will confine our discussion to the lossless case.

2-9.1 Parametric Equations

The reflection coefficient Γ is, in general, a complex quantity composed of a magnitude $|\Gamma|$ and a phase angle θ_r or, equivalently, a real part Γ_r and an imaginary part Γ_i ,

$$\Gamma = |\Gamma|e^{j\theta_r} = \Gamma_r + j\Gamma_i, \quad (2.90)$$

where

$$\Gamma_r = |\Gamma| \cos \theta_r, \quad (2.91a)$$

$$\Gamma_i = |\Gamma| \sin \theta_r. \quad (2.91b)$$

The Smith chart lies in the complex plane of Γ . In Fig. 2-20, point A represents a reflection coefficient $\Gamma_A = 0.3 + j0.4$ or, equivalently,

$$|\Gamma_A| = [(0.3)^2 + (0.4)^2]^{1/2} = 0.5$$

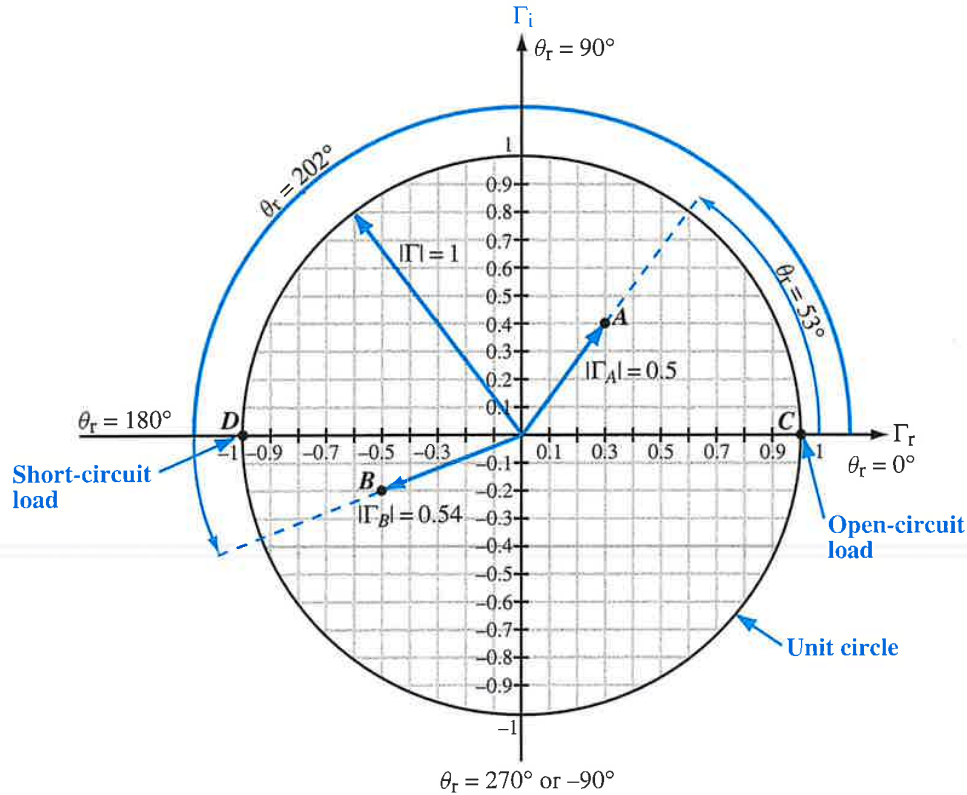


Figure 2-20: The complex Γ plane. Point A is at $\Gamma_A = 0.3 + j0.4 = 0.5e^{j53^\circ}$, and point B is at $\Gamma_B = -0.5 - j0.2 = |0.54|e^{j202^\circ}$. The unit circle corresponds to $|\Gamma| = 1$. At point C, $\Gamma = 1$, corresponding to an open-circuit load, and at point D, $\Gamma = -1$, corresponding to a short circuit.

and

$$\theta_r = \tan^{-1}(0.4/0.3) = 53^\circ.$$

Similarly, point B represents $\Gamma_B = -0.5 - j0.2$, or $|\Gamma_B| = 0.54$ and $\theta_r = 202^\circ$ [or, equivalently, $\theta_r = (360^\circ - 202^\circ) = -158^\circ$]. Note that when both Γ_r and Γ_i are negative numbers θ_r is in the third quadrant in the Γ_r - Γ_i plane. Thus, when using $\theta = \tan^{-1}(\Gamma_i/\Gamma_r)$ to

compute θ_r , it may be necessary to add or subtract 180° to obtain the correct value of θ_r .

The **unit circle** shown in Fig. 2-20 corresponds to $|\Gamma| = 1$. Because $|\Gamma| \leq 1$ for a transmission line, only that part of the Γ_r - Γ_i plane that lies within the unit circle has physical meaning; hence, future drawings will be limited to the domain contained within the unit circle.

Impedances on a Smith chart are represented by

normalized values, with Z_0 , the characteristic impedance of the line, serving as the normalization constant. *Normalized impedances are denoted by lowercase letters*, as in $z = Z/Z_0$. The *normalized load impedance* is then given by

$$z_L = Z_L/Z_0 \quad (\text{dimensionless}), \quad (2.92)$$

and the reflection coefficient Γ , defined by Eq. (2.49a), can be written as

$$\Gamma = \frac{Z_L/Z_0 - 1}{Z_L/Z_0 + 1} = \frac{z_L - 1}{z_L + 1}. \quad (2.93)$$

The inverse relation of Eq. (2.93) is

$$z_L = \frac{1 + \Gamma}{1 - \Gamma}. \quad (2.94)$$

The normalized load impedance z_L is, in general, a complex quantity composed of a *normalized load resistance* r_L and a *normalized load reactance* x_L :

$$z_L = r_L + jx_L. \quad (2.95)$$

Using Eqs. (2.90) and (2.95) in Eq. (2.94), we have

$$r_L + jx_L = \frac{(1 + \Gamma_r) + j\Gamma_i}{(1 - \Gamma_r) - j\Gamma_i}, \quad (2.96)$$

which can be solved to obtain explicit expressions for r_L and x_L in terms of Γ_r and Γ_i . This is accomplished by multiplying the numerator and denominator of the right-hand side of Eq. (2.96) by the complex conjugate of the denominator and then separating the result into real and imaginary parts. These steps lead to

$$r_L = \frac{1 - \Gamma_r^2 - \Gamma_i^2}{(1 - \Gamma_r)^2 + \Gamma_i^2}, \quad (2.97a)$$

$$x_L = \frac{2\Gamma_i}{(1 - \Gamma_r)^2 + \Gamma_i^2}. \quad (2.97b)$$

These expressions state that for a given set of values for Γ_r and Γ_i there corresponds a unique set of values for r_L

and x_L . However, if we fix the value of r_L , say at 2, many possible combinations of values can be assigned to Γ_r and Γ_i , each of which can give the same value of r_L . For example, $(\Gamma_r, \Gamma_i) = (0.33, 0)$ gives $r_L = 2$, as does the combination $(\Gamma_r, \Gamma_i) = (0.5, 0.29)$, as well as an infinite number of other combinations. In fact, if we were to plot in the Γ_r - Γ_i plane all the possible combinations of Γ_r and Γ_i corresponding to $r_L = 2$, we would obtain the circle denoted by $r_L = 2$ in Fig. 2-21. Similar circles apply to other values of r_L , and within the $|\Gamma| = 1$ domain all these circles pass through the point $(\Gamma_r, \Gamma_i) = (1, 0)$. After some algebraic manipulations, Eq. (2.97a) can be rearranged to give the following *parametric equation* for the circle in the Γ_r - Γ_i plane corresponding to a given value of r_L :

$$\left(\Gamma_r - \frac{r_L}{1 + r_L}\right)^2 + \Gamma_i^2 = \left(\frac{1}{1 + r_L}\right)^2. \quad (2.98)$$

The standard equation for a circle in the x - y plane with center at (x_0, y_0) and radius a is given by

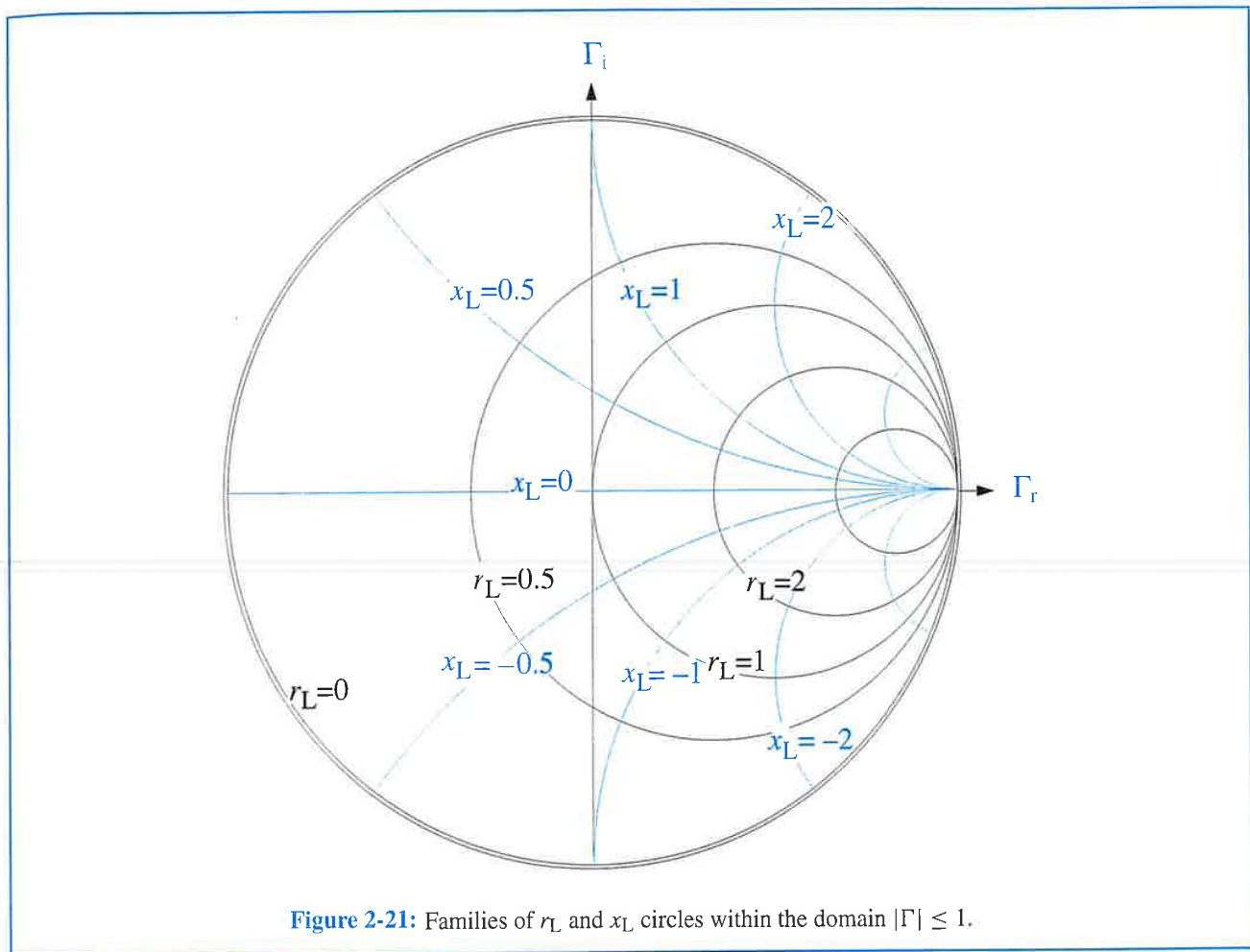
$$(x - x_0)^2 + (y - y_0)^2 = a^2. \quad (2.99)$$

Comparison of Eq. (2.98) with Eq. (2.99) shows that the r_L circle is centered at $\Gamma_r = r_L/(1 + r_L)$ and $\Gamma_i = 0$, and its radius is $1/(1 + r_L)$. The largest circle shown in Fig. 2-21 corresponds to $r_L = 0$, which is also the unit circle corresponding to $|\Gamma| = 1$. This is to be expected, because when $r_L = 0$, $|\Gamma| = 1$ regardless of the magnitude of x_L .

A similar examination of the expression for x_L given by Eq. (2.97b) also leads to an equation for a circle given by

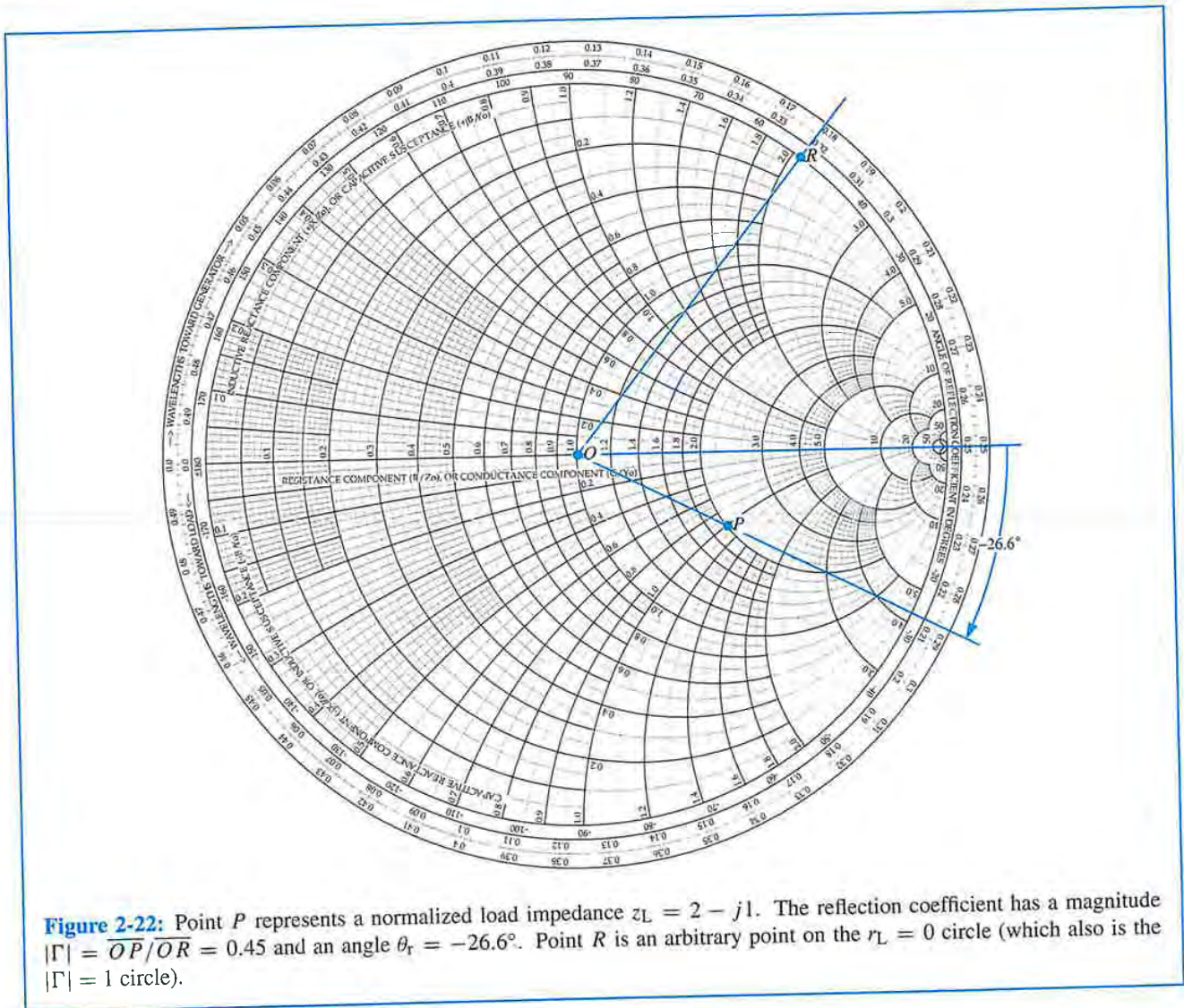
$$(\Gamma_r - 1)^2 + \left(\Gamma_i - \frac{1}{x_L}\right)^2 = \left(\frac{1}{x_L}\right)^2, \quad (2.100)$$

but the x_L circles in the Γ_r - Γ_i plane exhibit a different character from that for r_L . To start with, the normalized reactance x_L may assume both positive and negative values, whereas the normalized resistance cannot be negative (negative resistances are physically



meaningless). Hence, Eq. (2.100) can generate two families of circles, one family corresponding to positive values of x_L and another corresponding to negative values of x_L . Furthermore, as shown in Fig. 2-21, only part of a given circle falls within the bounds of the unit circle. The families of circles of the two parametric equations given by Eqs. (2.98) and (2.100) plotted for selected

values of r_L and x_L constitute the Smith chart shown in Fig. 2-22. A given point on the Smith chart, such as point P in Fig. 2-22, represents a normalized load impedance $z_L = 2 - j1$, with a corresponding voltage reflection coefficient $\Gamma = 0.45 \exp(-j26.6^\circ)$. The magnitude $|\Gamma| = 0.45$ is obtained by dividing the length of the line between the center of the Smith chart and the point P by



the length of the line between the center of the Smith chart and the edge of the unit circle (the radius of the unit circle corresponds to $|\Gamma| = 1$). The perimeter of the Smith chart contains three concentric scales. The innermost scale is

labeled *angle of reflection coefficient in degrees*. This is the scale for θ_r . As indicated in Fig. 2-22, $\theta_r = -26.6^\circ$ for point P . The meanings and uses of the other two scales are discussed next.

EXERCISE 2.15 Use the Smith chart to find the values of Γ corresponding to the following normalized load impedances: (a) $z_L = 2 + j0$, (b) $z_L = 1 - j1$, (c) $z_L = 0.5 - j2$, (d) $z_L = -j3$, (e) $z_L = 0$ (short circuit), (f) $z_L = \infty$ (open circuit), (g) $z_L = 1$ (matched load).

Ans. (a) $\Gamma = 0.33$, (b) $\Gamma = 0.45 \angle -63.4^\circ$, (c) $\Gamma = 0.83 \angle -50.9^\circ$, (d) $\Gamma = 1 \angle -36.9^\circ$, (e) $\Gamma = -1$, (f) $\Gamma = 1$, (g) $\Gamma = 0$.

2-9.2 Input Impedance

From Eq. (2.61), the input impedance looking toward the load at a distance l from the load is given by

$$Z_{in} = Z_0 \left[\frac{1 + \Gamma e^{-j2\beta l}}{1 - \Gamma e^{-j2\beta l}} \right] \quad (\Omega). \quad (2.101)$$

To use the Smith chart, we always normalize impedances to the characteristic impedance Z_0 . Hence, the **normalized input impedance** is

$$z_{in} = \frac{Z_{in}}{Z_0} = \frac{1 + \Gamma e^{-j2\beta l}}{1 - \Gamma e^{-j2\beta l}} \quad (\text{dimensionless}). \quad (2.102)$$

The quantity $\Gamma = |\Gamma|e^{j\theta_r}$ is the voltage reflection coefficient at the load. Let us define

$$\Gamma_l = \Gamma e^{-j2\beta l} = |\Gamma|e^{j\theta_r}e^{-j2\beta l} = |\Gamma|e^{j(\theta_r - 2\beta l)} \quad (2.103)$$

as the **phase-shifted** voltage reflection coefficient, meaning that Γ_l has the same magnitude as Γ , but the phase of Γ_l is shifted by $2\beta l$ relative to that of Γ . In terms of Γ_l , Eq. (2.102) can be rewritten as

$$z_{in} = \frac{1 + \Gamma_l}{1 - \Gamma_l}. \quad (2.104)$$

The form of Eq. (2.104) is identical with that for z_L given by Eq. (2.94):

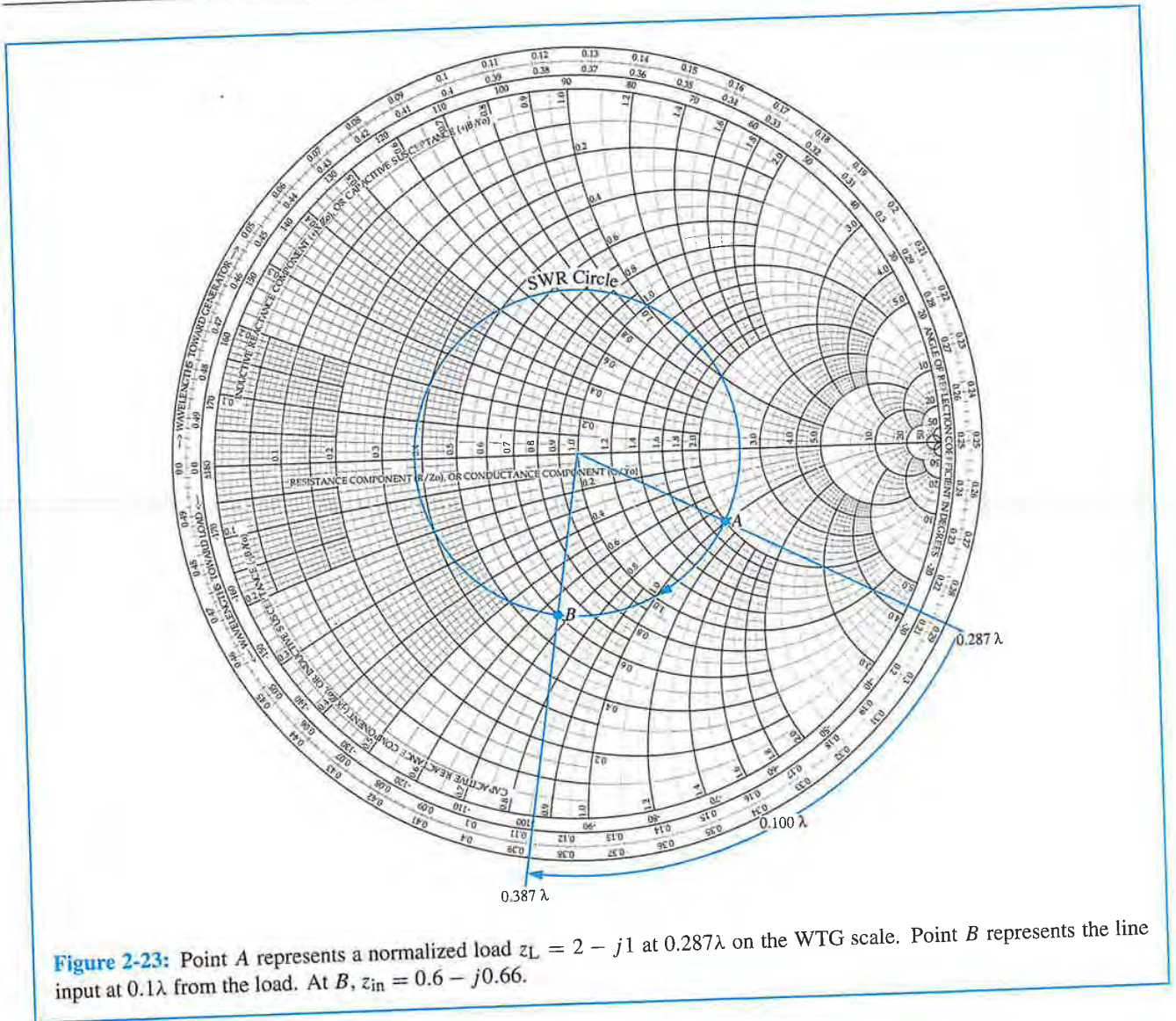
$$z_L = \frac{1 + \Gamma}{1 - \Gamma}. \quad (2.105)$$

The similarity in form suggests that, if Γ is transformed into Γ_l , z_L gets transformed into z_{in} . On the Smith chart, transforming Γ into Γ_l means maintaining $|\Gamma|$ constant and decreasing the phase θ_r by $2\beta l$, which corresponds to rotation in a clockwise direction on the Smith chart. Noting that a complete rotation around the Smith chart is equal to a phase change of 2π , the length l corresponding to such a change is obtained from

$$2\beta l = 2 \frac{2\pi}{\lambda} l = 2\pi, \quad (2.106)$$

or $l = \lambda/2$. The outermost scale around the perimeter of the Smith chart (Fig. 2-22), called the **wavelengths toward generator** (WTG) scale, has been constructed to denote movement on the transmission line toward the generator, in units of the wavelength λ . That is, l is measured in wavelengths, and one complete rotation corresponds to $l = \lambda/2$. In some transmission-line problems, it may be necessary to move from some point on the transmission line toward another point closer to the load, in which case the phase is increased, which corresponds to rotation in a counterclockwise direction. For convenience, the Smith chart contains a third scale around its perimeter (in between the θ_r scale and the WTG scale) for accommodating such a need. It is called the **wavelengths toward load** (WTL) scale.

To illustrate how the Smith chart is used to find Z_{in} , let us consider a $50\text{-}\Omega$ lossless transmission line terminated in a load impedance $Z_L = (100 - j50)\text{ }\Omega$. Our objective is to find Z_{in} at a distance $l = 0.1\lambda$ from the load. The normalized load impedance is $z_L = Z_L/Z_0 = 2 - j1$, and it is denoted by point A on the Smith chart shown in Fig. 2-23. On the WTG scale, the location of point A is at 0.287λ . Using a compass, a circle is drawn through point A, with the center of the circle being at the center of the Smith chart. Since the center of the Smith chart is the intersection point of the Γ_r and Γ_i axes, all points on the drawn circle have the same value of $|\Gamma|$. This is called the **constant- $|\Gamma|$ circle**, or more commonly the **SWR circle**. The reason for this second name is that the



voltage standing-wave ratio (SWR) is related to $|\Gamma|$ by Eq. (2.59) as

$$S = \frac{1 + |\Gamma|}{1 - |\Gamma|}. \quad (2.107)$$

Thus, a constant value of $|\Gamma|$ corresponds to a specific value for S . As was stated earlier, to transform z_L to z_{in} , we need to maintain $|\Gamma|$ constant, which means staying on the SWR circle, and to decrease the phase of Γ by $2\beta l$. This is equivalent to moving a distance $l = 0.1\lambda$ toward

the generator on the WTG scale. Since the location of point A is at 0.287λ , we need to move to location $0.287\lambda + 0.1\lambda = 0.387\lambda$ on the WTG scale. A radial line through this new position on the WTG scale intersects the SWR circle at point B. This point represents z_{in} , and its value is $z_{in} = 0.6 - j0.66$. Finally, we unnormalize z_{in} by multiplying it by $Z_0 = 50 \Omega$ to get $Z_{in} = (30 - j33) \Omega$. This result can be verified analytically using Eq. (2.101). The points between points A and B on the SWR circle represent different points along the transmission line.

EXERCISE 2.16 Use the Smith chart to find the normalized input impedance of a lossless line of length l terminated in a normalized load impedance z_L for each of the following combinations: (a) $l = 0.25\lambda$, $z_L = 1 + j0$, (b) $l = 0.5\lambda$, $z_L = 1 + j1$, (c) $l = 0.3\lambda$, $z_L = 1 - j1$, (d) $l = 1.2\lambda$, $z_L = 0.5 - j0.5$, (e) $l = 0.1\lambda$, $z_L = 0$ (short circuit), (f) $l = 0.4\lambda$, $z_L = j3$, (g) $l = 0.2\lambda$, $z_L = \infty$ (open circuit).

Ans. (a) $z_{in} = 1 + j0$, (b) $z_{in} = 1 + j1$, (c) $z_{in} = 0.76 + j0.84$, (d) $z_{in} = 0.59 + j0.66$, (e) $z_{in} = 0 + j0.73$, (f) $z_{in} = 0 + j0.72$, (g) $z_{in} = 0 - j0.32$.

2-9.3 SWR, Voltage Maxima and Minima

Consider a load with $z_L = 2 + j1$. Figure 2-24 shows a Smith chart with a SWR circle drawn through z_L (point A). The SWR circle intersects the real axis (Γ_r) at two points, designated P_{max} and P_{min} . Thus, at both points $\Gamma_i = 0$ and $\Gamma = \Gamma_r$. Also, on the real axis, the imaginary part of the load impedance $x_L = 0$. From the definition of Γ ,

$$\Gamma = \frac{z_L - 1}{z_L + 1}, \quad (2.108)$$

points P_{max} and P_{min} correspond to the special case

$$\Gamma = \Gamma_r = \frac{r_L - 1}{r_L + 1} \quad (\text{for } \Gamma_i = 0), \quad (2.109)$$

with P_{min} corresponding to the condition when $r_L < 1$ and P_{max} corresponding to the condition when $r_L > 1$. Rewriting Eq. (2.107) for $|\Gamma|$ in terms of S , we have

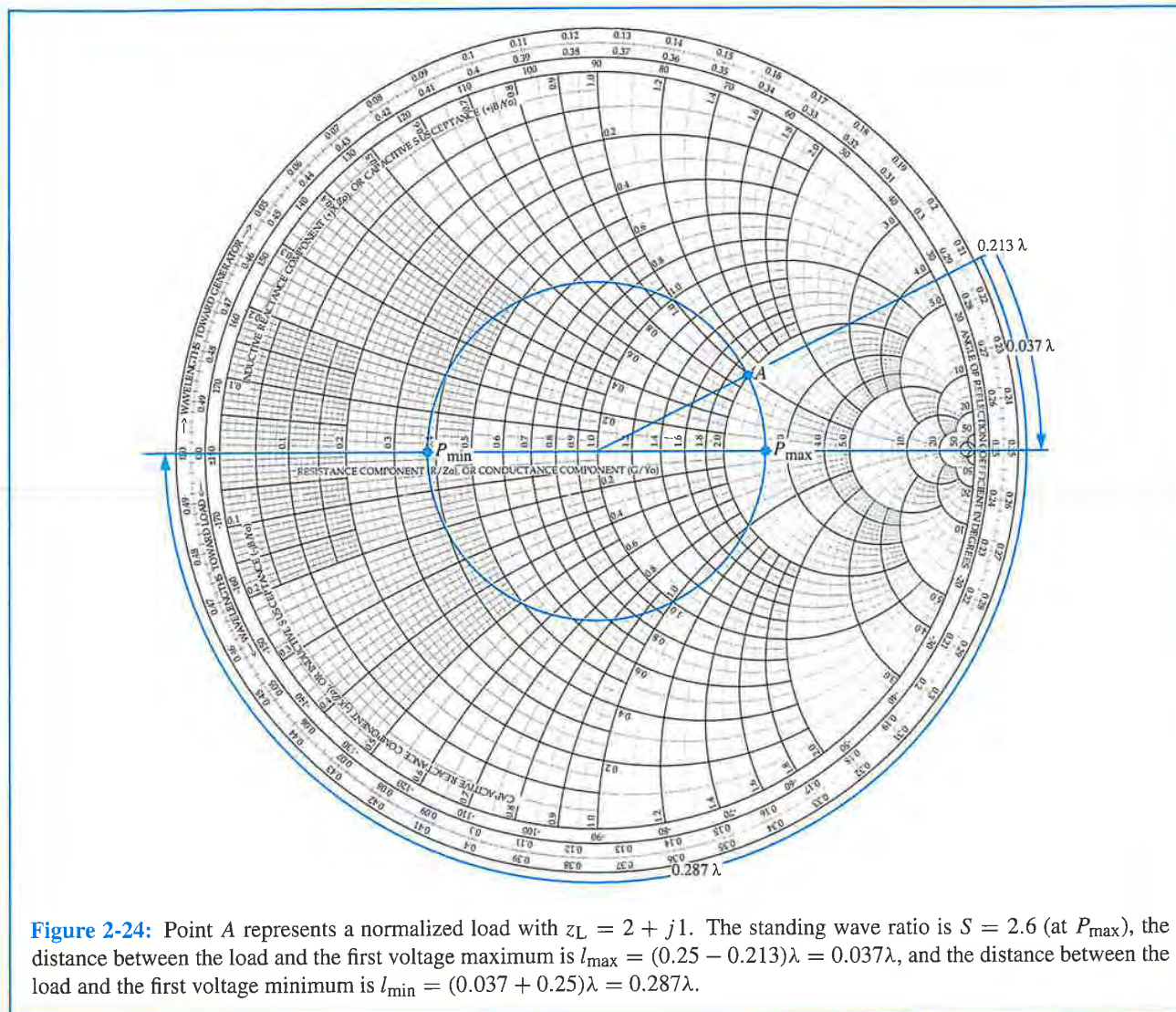
$$|\Gamma| = \frac{S - 1}{S + 1}. \quad (2.110)$$

For points P_{max} and P_{min} , $|\Gamma| = \Gamma_r$; hence

$$\Gamma_r = \frac{S - 1}{S + 1}. \quad (2.111)$$

The similarity in form of Eqs. (2.109) and (2.111) suggests that $S = r_L$. However, since by definition $S \geq 1$, only point P_{max} (for which $r_L > 1$) satisfies the similarity condition. In Fig. 2-24, $r_L = 2.6$ at P_{max} ; hence $S = 2.6$. In other words, *S is numerically equal to the value of r_L at P_{max} , the point at which the SWR circle intersects the real Γ axis on the right-hand side of the chart's center.*

The points P_{min} and P_{max} also represent the distances from the load at which the magnitude of the voltage on the line, $|\tilde{V}|$, is a minimum and a maximum, respectively. This statement is easily demonstrated by considering the definition of Γ_l given by Eq. (2.103). At point P_{max} , the total phase of Γ_l , that is, $(\theta_r - 2\beta l)$, is equal to zero (if $\theta_r > 0$) or 2π (if $\theta_r < 0$), which is the condition corresponding to $|\tilde{V}|_{max}$, as indicated by Eq. (2.55). Similarly, at P_{min} the total phase of Γ_l is equal to π , which is the condition for $|\tilde{V}|_{min}$. Thus, for the transmission line represented by the SWR circle shown in Fig. 2-24, the distance between the load and the nearest voltage maximum is $l_{max} = 0.037\lambda$, obtained by moving clockwise from the load at point A to point P_{max} , and the distance to the nearest voltage minimum is $l_{min} = 0.287\lambda$, corresponding to the clockwise rotation from A to P_{min} . Since the location of $|\tilde{V}|_{max}$ is also the location of $|\tilde{I}|_{min}$ and the location of $|\tilde{V}|_{min}$ is also the location of $|\tilde{I}|_{max}$, the Smith chart provides a convenient way to determine the distances to all maxima and minima on the line (the standing-wave pattern has a repetition period of $\lambda/2$).



2-9.4 Impedance to Admittance Transformations

In solving certain types of transmission line problems, it is often more convenient to work with admittances than with impedances. Any impedance Z is in general

a complex quantity consisting of a resistance R and a reactance X :

$$Z = R + jX \quad (\Omega), \quad (2.112)$$

The **admittance** Y corresponding to Z is the reciprocal of Z :

$$Y = \frac{1}{Z} = \frac{1}{R + jX} = \frac{R - jX}{R^2 + X^2} \quad (\text{S}). \quad (2.113)$$

The real part of Y is called the **conductance** G , and the imaginary part of Y is called the **susceptance** B . That is,

$$Y = G + jB \quad (\text{S}). \quad (2.114)$$

Comparison of Eq. (2.114) with Eq. (2.113) leads to

$$G = \frac{R}{R^2 + X^2} \quad (\text{S}), \quad (2.115a)$$

$$B = \frac{-X}{R^2 + X^2} \quad (\text{S}). \quad (2.115b)$$

A normalized impedance z is defined as the ratio of Z to Z_0 , the characteristic impedance of the line. The same concept applies to the definition of the **normalized admittance** y ; that is,

$$y = \frac{Y}{Y_0} = \frac{G}{Y_0} + j \frac{B}{Y_0} = g + jb \quad (\text{dimensionless}), \quad (2.116)$$

where $Y_0 = 1/Z_0$ is the **characteristic admittance of the line** and

$$g = \frac{G}{Y_0} = GZ_0 \quad (\text{dimensionless}), \quad (2.117a)$$

$$b = \frac{B}{Y_0} = BZ_0 \quad (\text{dimensionless}). \quad (2.117b)$$

The lowercase quantities g and b represent the **normalized conductance** and **normalized susceptance** of y , respectively. Of course, the normalized admittance y is the reciprocal of the normalized impedance z ,

$$y = \frac{Y}{Y_0} = \frac{Z_0}{Z} = \frac{1}{z}. \quad (2.118)$$

Accordingly, using Eq. (2.105), the normalized load admittance y_L is given by

$$y_L = \frac{1}{z_L} = \frac{1 - \Gamma}{1 + \Gamma} \quad (\text{dimensionless}). \quad (2.119)$$

Now let us consider the normalized input impedance z_{in} at a distance $l = \lambda/4$ from the load. Using Eq. (2.102) with $2\beta l = 4\pi l/\lambda = 4\pi\lambda/4\lambda = \pi$ gives

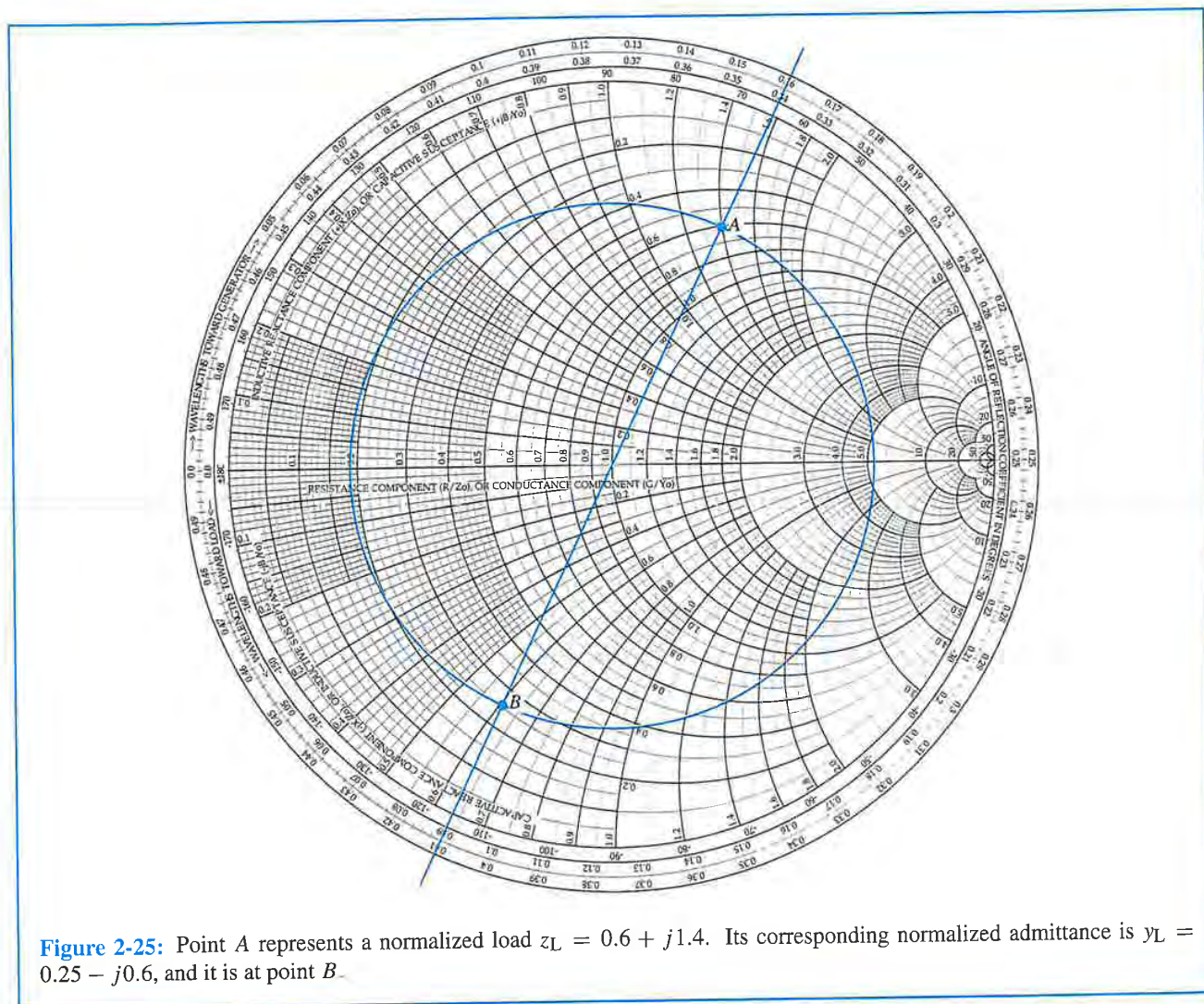
$$z_{in}(l = \lambda/4) = \frac{1 + \Gamma e^{-j\pi}}{1 - \Gamma e^{-j\pi}} = \frac{1 - \Gamma}{1 + \Gamma} = y_L. \quad (2.120)$$

Thus, **rotation by $\lambda/4$ on the Smith chart transforms z_L into y_L** . In Fig. 2-25, the points representing z_L and y_L are diametrically opposite to each other on the SWR circle. In fact, such a transformation on the Smith chart can be used to determine any normalized admittance from its corresponding normalized impedance, and vice versa.

The Smith chart can be used with normalized impedances or with normalized admittances. As an impedance chart, the Smith chart consists of r_L and x_L circles, representing the normalized resistance and reactance of a normalized load impedance z_L . When used as an admittance chart, the r_L circles become g_L circles and the x_L circles become b_L circles, where g_L and b_L are the normalized conductance and susceptance of the normalized load admittance y_L , respectively.

Example 2-10 Smith Chart Calculations

A 50- Ω lossless transmission line is terminated in a load impedance $Z_L = (25 + j50) \Omega$. Use the Smith chart to find (a) the voltage reflection coefficient, (b) the voltage standing-wave ratio, (c) the distances of the first voltage maximum and first voltage minimum from the load, (d) the input impedance of the line, given that the line is 3.3λ long, and (e) the input admittance of the line.



Solution: (a) The normalized load impedance is

$$z_L = \frac{Z_L}{Z_0} = \frac{25 + j50}{50} = 0.5 + j1,$$

which is marked as point A on the Smith chart in

Fig. 2-26. Using a ruler, a radial line is drawn from the center of the chart at point O through point A, outward to the outer perimeter of the chart. The line crosses the scale labeled “angle of reflection coefficient in degrees” at $\theta_r = 83^\circ$. Next, a ruler is used to measure the length d_A

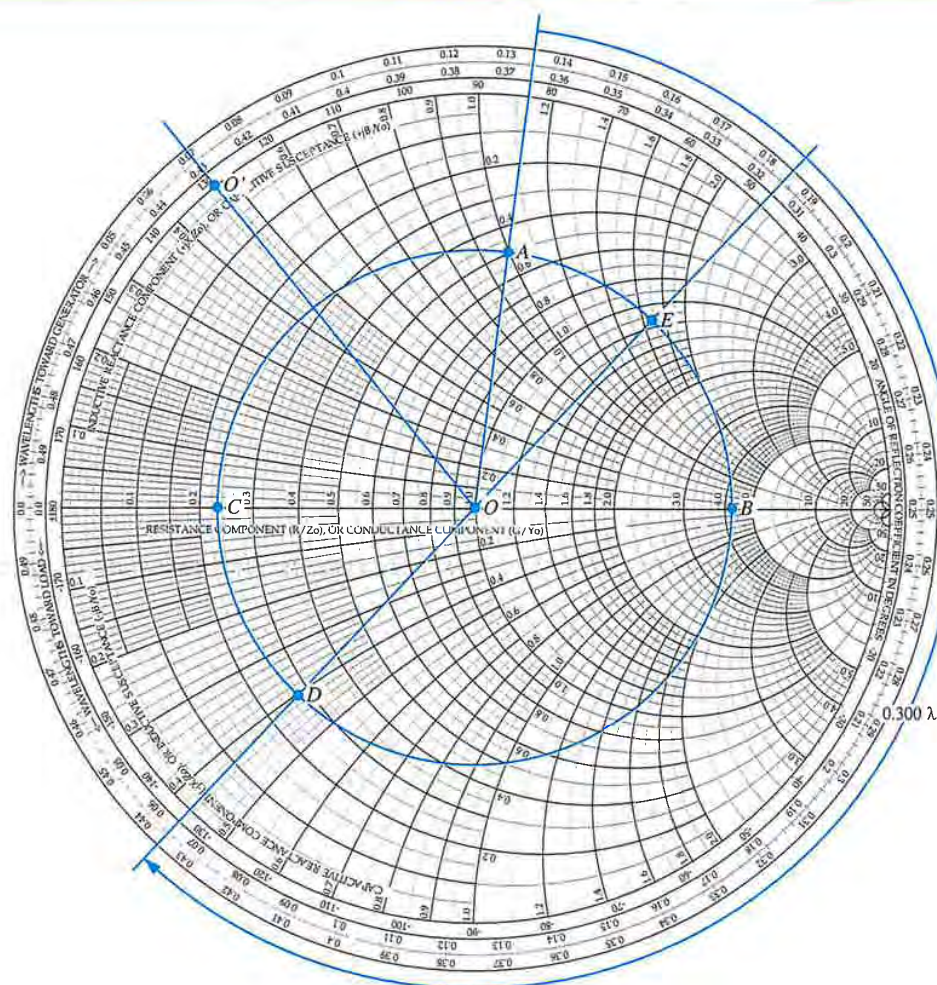


Figure 2-26: Solution for Example 2-10. Point A represents a normalized load $z_L = 0.5 + j1$ at 0.135λ on the WTG scale. At A, $\theta_r = 83^\circ$ and $|\Gamma| = d_A/d_{O'} = \overline{OA}/\overline{OO'} = 0.62$. At B, the standing-wave ratio is $S = 4.26$. The distance from A to B gives $l_{\max} = 0.115\lambda$ and from A to C gives $l_{\min} = 0.365\lambda$. Point D represents the normalized input impedance $z_{\text{in}} = 0.28 - j0.40$, and point E represents the normalized input admittance $y_{\text{in}} = 1.15 + j1.7$.

of the line between points O and A and the length $d_{O'}$ of the line between points O and O', where O' is an arbitrary point on the $r_L = 0$ circle. The length $d_{O'}$ is equal to the radius of the $|\Gamma| = 1$ circle. The magnitude

of Γ is then obtained from $|\Gamma| = d_A/d_{O'} = 0.62$. Hence,

$$\Gamma = 0.62e^{j83^\circ} = 0.62\angle 83^\circ. \quad (2.121)$$

(b) Using a compass, the SWR circle with center at

point O is drawn through point A . The circle crosses the Γ_r axis at points B and C . The value of r_L at point B is 4.26, which also is equal to S . Thus,

$$S = 4.26.$$

(c) The first voltage maximum is at point B on the SWR circle, which is at location 0.25λ on the WTG scale. The load, represented by point A , is at 0.135λ on the WTG scale. Hence, the distance between the load and the first voltage maximum is

$$l_{\max} = (0.25 - 0.135)\lambda = 0.115\lambda.$$

The first voltage minimum is at point C . Moving on the WTG scale between A and C gives

$$l_{\min} = (0.5 - 0.135)\lambda = 0.365\lambda,$$

which is 0.25λ past $l_{\max}$.

(d) The line is 3.3λ long; subtracting multiples of 0.5λ leaves 0.3λ . From the load at 0.135λ on the WTG scale, the input of the line is at $(0.135 + 0.3)\lambda = 0.435\lambda$. This is labeled as point D on the SWR circle, and the normalized impedance is read as

$$z_{\text{in}} = 0.28 - j0.40,$$

and therefore

$$Z_{\text{in}} = z_{\text{in}}Z_0 = (0.28 - j0.40)50 = (14 - j20) \Omega.$$

(e) The normalized input admittance y_{in} is found by moving 0.25λ on the Smith chart to the image point of z_{in} across the circle, labeled point E on the SWR circle. The coordinates of point E give

$$y_{\text{in}} = 1.15 + j1.7,$$

and the corresponding input admittance is

$$Y_{\text{in}} = y_{\text{in}}Y_0 = \frac{y_{\text{in}}}{Z_0} = \frac{1.15 + j1.7}{50} = (0.023 + j0.034) \text{ S}.$$

Example 2-11 Determining Z_L Using the Smith Chart

This problem is similar to Example 2-5, except now we demonstrate the solution using the Smith chart.

Given that the voltage standing-wave ratio is $S = 3$ on a $50\text{-}\Omega$ line, that the first voltage minimum occurs at 5 cm from the load, and that the distance between successive minima is 20 cm, find the load impedance.

Solution: The distance between successive minima is equal to $\lambda/2$. Hence, $\lambda = 40$ cm. In wavelength units, the first voltage minimum is at

$$l_{\min} = \frac{5}{40} = 0.125\lambda.$$

Point A on the Smith chart in Fig. 2-27 corresponds to $r_L = S = 3$. Using a compass, the constant S circle is drawn through point A . Point B corresponds to the locations of voltage minima. Upon moving 0.125λ from point B toward the load on the WTL scale (counterclockwise), we arrive at point C , which represents the location of the load. The normalized load impedance at point C is

$$z_L = 0.6 - j0.8.$$

Multiplying by $Z_0 = 50 \Omega$, we obtain

$$Z_L = 50(0.6 - j0.8) = (30 - j40) \Omega. \quad \blacksquare$$

REVIEW QUESTIONS

Q2.19 The outer perimeter of the Smith chart represents what value of $|\Gamma|$? What point on the Smith chart represents a matched load?

Q2.20 What is the SWR circle? What quantity is constant at all points on the SWR circle?

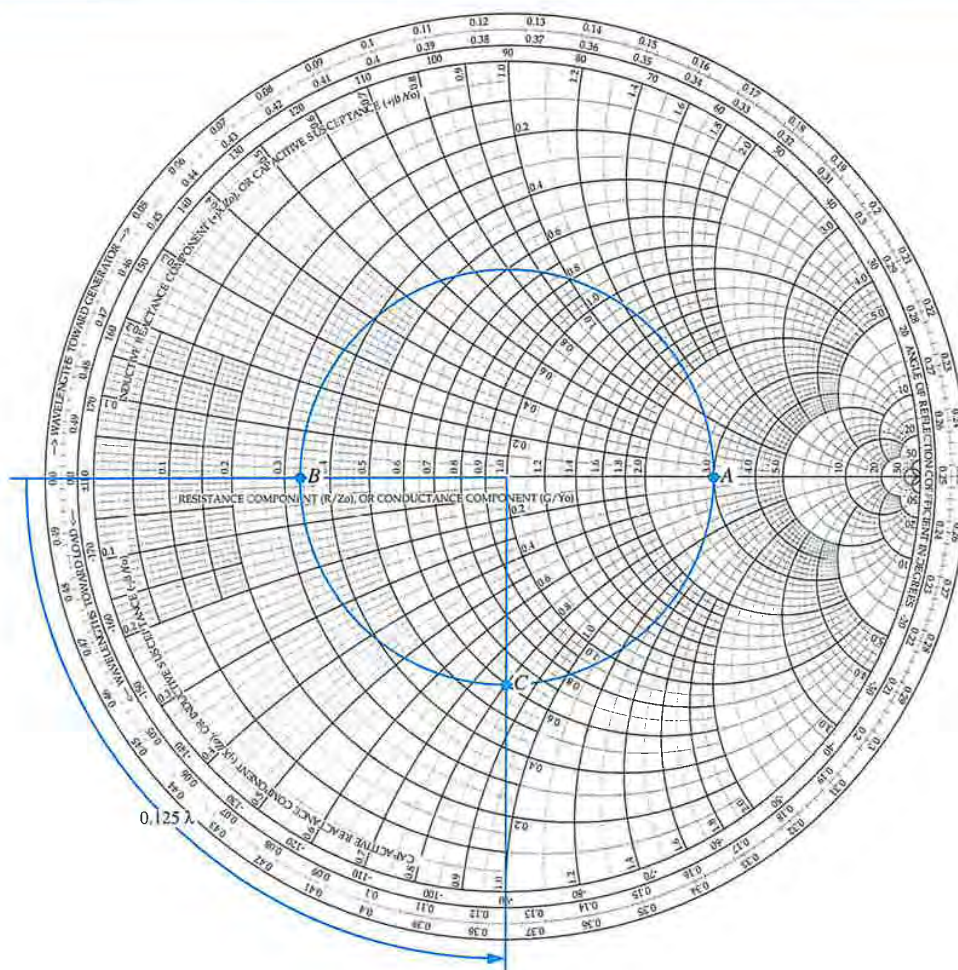


Figure 2-27: Solution for Example 2-11. Point A denotes that $S = 3$, point B represents the location of the voltage minimum, and point C represents the load at 0.125λ on the WTL scale from point B. At C, $z_L = 0.6 - j0.8$.

Q2.21 What line length corresponds to one complete rotation around the Smith chart? Why?

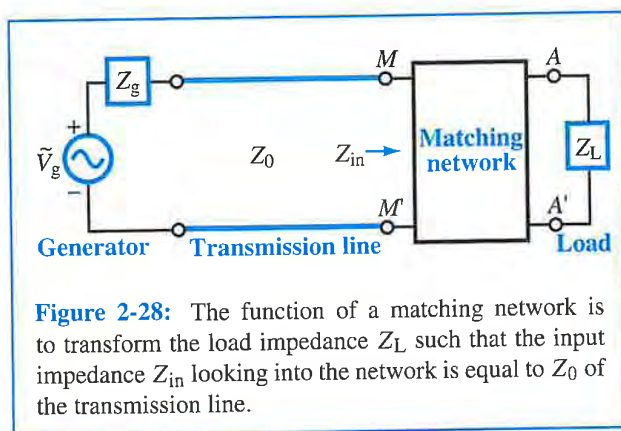
Q2.22 What points on the SWR circle correspond to the locations of the voltage maxima and minima on the line and why?

Q2.23 Given a normalized impedance z_L , how do you use the Smith chart to find the corresponding normalized admittance $y_L = 1/z_L$?

2-10 Impedance Matching

A transmission line usually connects a generator circuit at one end to a load at the other end. The load may be an antenna or any circuit with an equivalent input impedance Z_L . *The transmission line is said to be matched to the load when its characteristic impedance $Z_0 = Z_L$,* in which case no reflection occurs at the load end of the line. Since the primary uses of transmission lines are to transfer power and to transmit coded signals (such as digital data), a matched load ensures that the power delivered to the load is a maximum.

The simplest solution to matching a load to a transmission line is to design the load circuit such that its impedance $Z_L = Z_0$. Unfortunately, this may not be possible in practice because the load circuit may have to satisfy other requirements. An alternative solution is to place an *impedance-matching network* between the load and the transmission line as illustrated in Fig. 2-28. The purpose of the matching network is to eliminate reflections at the juncture MM' between the transmission line and the network. This is achieved by designing the matching network to exhibit an impedance equal to Z_0 at MM' when looking into the network from the transmission line side. If the network is lossless, then all the power going into it will end up in the load. Matching networks may consist of lumped elements (and in order



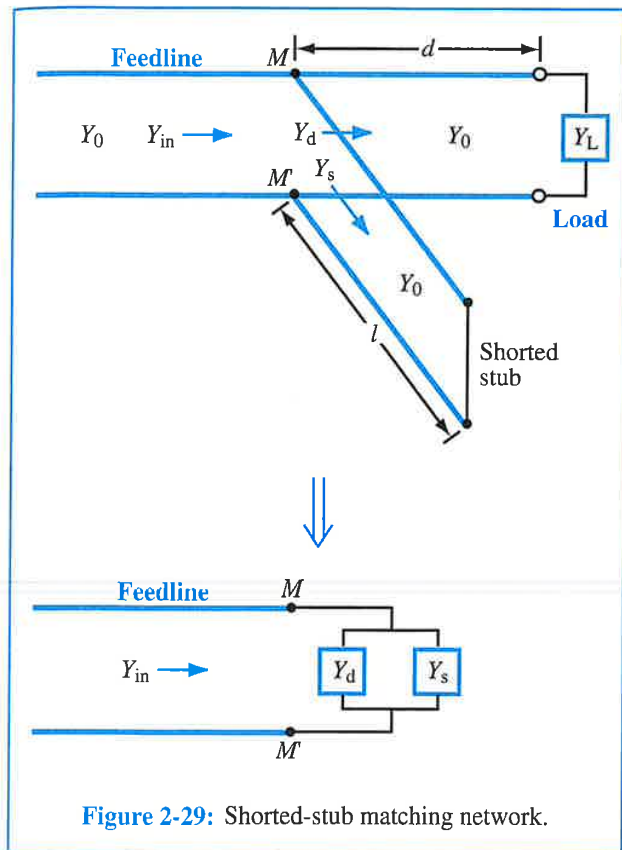
to avoid ohmic losses only capacitors and inductors are used) or of sections of transmission lines with appropriate lengths and terminations. We will demonstrate the latter approach using a *single-stub matching network*.

The matching network is intended to match a load impedance $Z_L = R_L + jX_L$ to a lossless transmission line with characteristic impedance Z_0 . This means that the network has to transform the real part of the load impedance from R_L at the load to Z_0 at MM' in Fig. 2-28 and to transform the reactive part from X_L at the load to zero at MM' . To achieve these two transformations, the matching network has to have at least two degrees of freedom; that is, at least two adjustable parameters. The single-stub matching network shown in Fig. 2-29 consists of two transmission line sections, one of length d connecting the load to the feedline at MM' and another of length l connected in parallel with the other two lines at MM' . This second line is called a *stub*, and it is usually terminated in either a short circuit or open circuit. The stub shown in Fig. 2-29 has a short-circuit termination.

The required two degrees of freedom are provided by the length l of the stub and the distance d from the load to the stub position. Because at MM' the stub is added in "parallel" to the line (and hence called a *shunt* stub), it is easier to work with admittances than with impedances. The matching procedure consists of two basic steps. In the first step, the distance d is selected so as to transform the load admittance $Y_L = 1/Z_L$ into an admittance of the form $Y_d = Y_0 + jB$, when looking toward the load at MM' . Then, in the second step, the length l of the stub line is selected so that its input admittance Y_s at MM' is equal to $-jB$. The parallel sum of the two admittances at MM' yields Y_0 , the characteristic admittance of the line. The procedure is illustrated by Example 2-12.

Example 2-12 Single-Stub Matching

A $50\text{-}\Omega$ transmission line is connected to an antenna with load impedance $Z_L = (25 - j50)\text{ }\Omega$. Find the



position and length of the short-circuited stub required to match the line.

Solution:

The normalized load impedance is

$$z_L = \frac{Z_L}{Z_0} = \frac{25 - j50}{50} = 0.5 - j1,$$

which is located at point A in the Smith chart of Fig. 2-30. Next, we draw the constant S circle through point A. To perform the matching task, it is easier to work with admittances than with impedances. The normalized load

admittance y_L is at point B, obtained by rotating 0.25λ , or equivalently by drawing a line from point A through the center to the image point on the S circle. The value of y_L at B is

$$y_L = 0.4 + j0.8,$$

and it is located at position 0.115λ on the WTG scale. In the admittance domain, the r_L circles become g_L circles, and the x_L circles become b_L circles. To achieve matching, we need to move from the load toward the generator a distance d such that the normalized input admittance y_d of the line terminated in the load (Fig. 2-29) has a real part equal to 1. This condition is satisfied by either of the two **matching points** C and D on the Smith charts of Figs. 2-30 and 2-31, respectively, corresponding to the intersections of the S circle with the $g_L = 1$ circle. Points C and D represent two possible solutions for the distance d in Fig. 2-29.

Solution for Point C [Fig. 2-30]: At C,

$$y_d = 1 + j1.58,$$

and it is located at 0.178λ on the WTG scale. The distance between points B and C is

$$d_1 = (0.178 - 0.115)\lambda = 0.063\lambda.$$

Looking from the generator toward the parallel combination of the line connected to the load and the short-circuited stub, the normalized input admittance at the juncture is

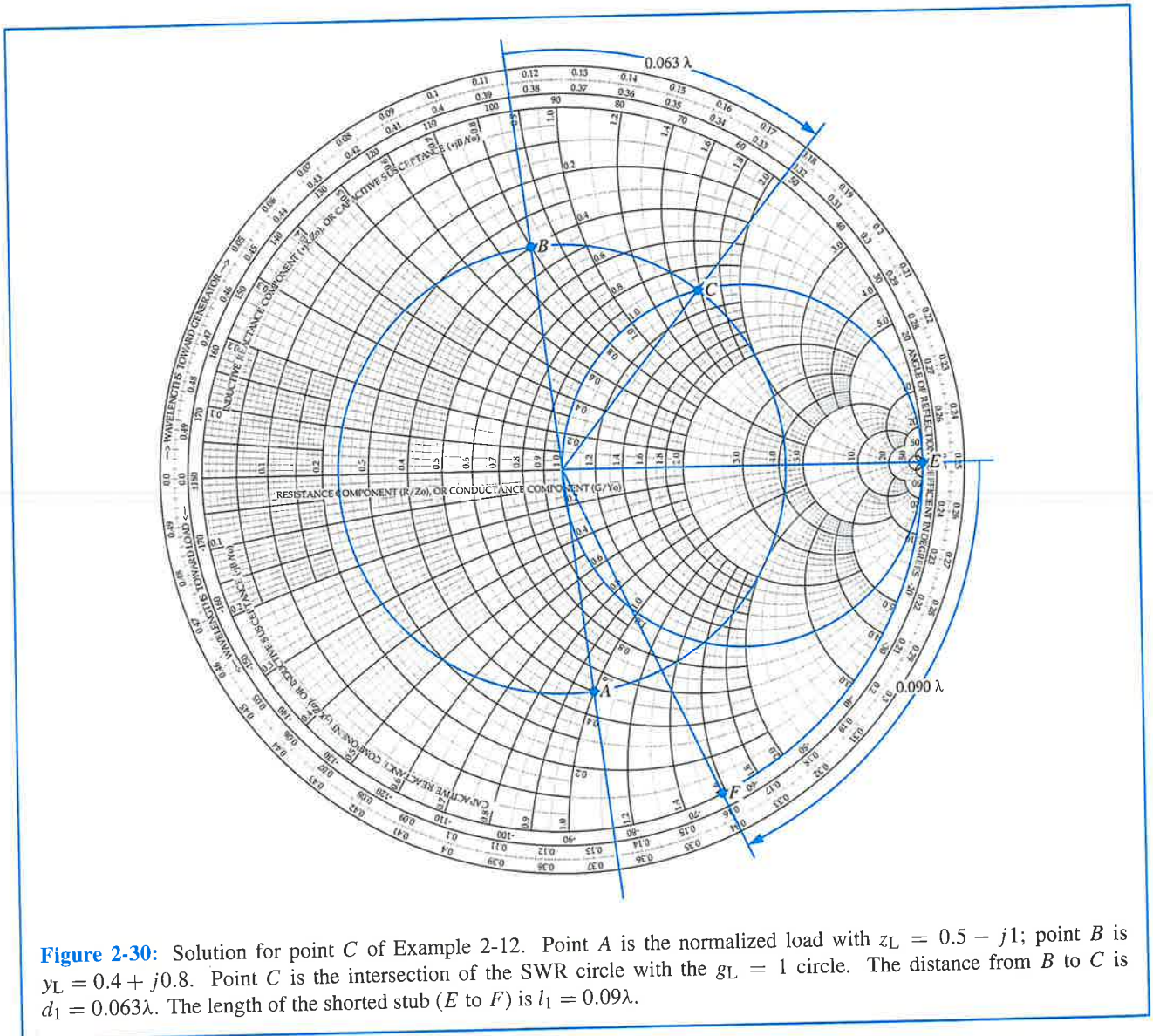
$$y_{in} = y_s + y_d,$$

where y_s is the normalized input admittance of the stub line. To match the feed line to the parallel combinations, we need to have $y_{in} = 1 + j0$. Thus,

$$1 + j0 = y_s + 1 + j1.58,$$

or

$$y_s = -j1.58.$$



The normalized admittance of a short circuit is $-j\infty$ and is located at point E on the Smith chart, whose position is 0.25λ on the WTG scale. An input normalized

admittance of $-j1.58$ is located at point F and is at position 0.34λ on the WTG scale. Hence,

$$l_1 = (0.34 - 0.25)\lambda = 0.09\lambda.$$

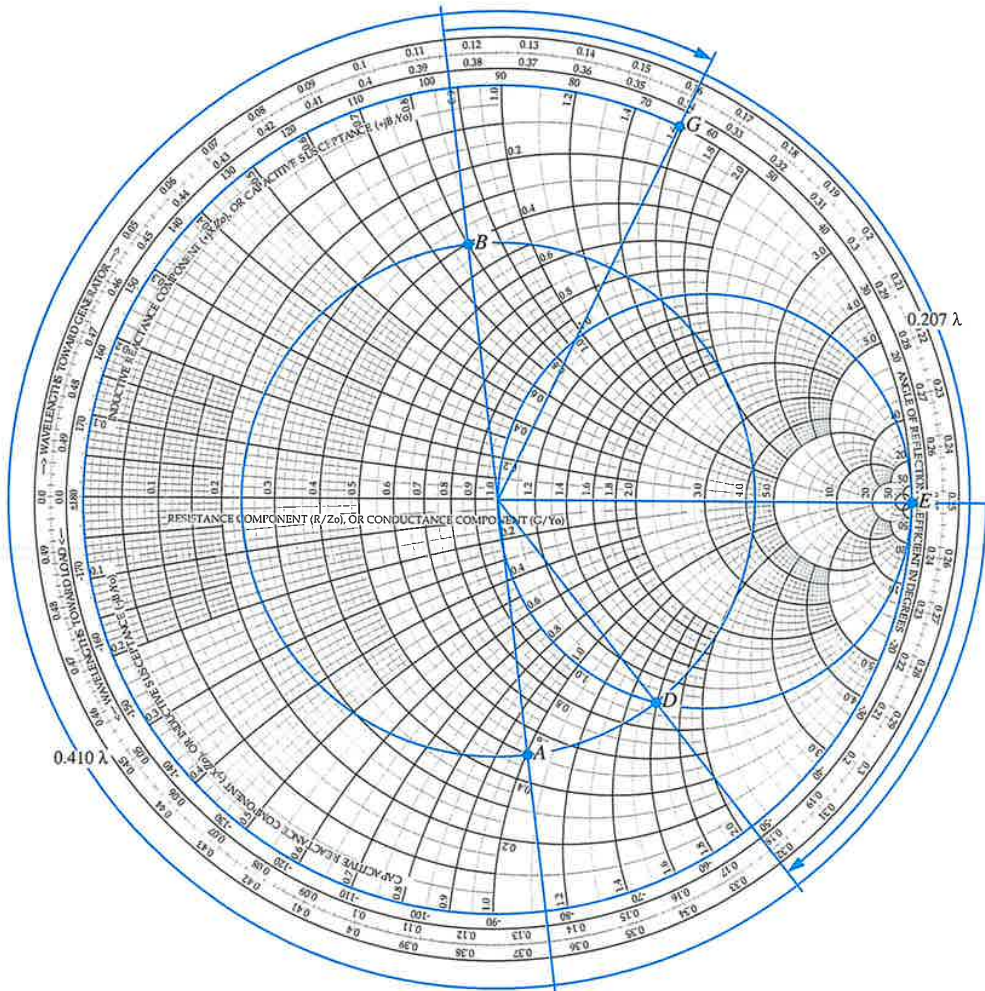


Figure 2-31: Solution for point D of Example 2-12. Point D is the second point of intersection of the SWR circle and the $y_L = 1$ circle. The distance B to D gives $d_2 = 0.207\lambda$, and the distance E to G gives $l_2 = 0.410\lambda$.

Solution for Point D (Fig. 2-31): At point D ,

$$y_d = 1 - j1.58,$$

and the distance between points B and D is

$$d_2 = (0.322 - 0.115)\lambda = 0.207\lambda.$$

The needed normalized input admittance of the stub is $y_s = +j1.58$, which is located at point G at position

0.16λ on the WTG scale. Rotating from point E to point G involves a rotation of 0.25λ plus an additional rotation of 0.16λ , or

$$l_2 = (0.25 + 0.16)\lambda = 0.41\lambda. \quad \blacksquare$$

REVIEW QUESTIONS

Q2.24 To match an arbitrary load impedance to a lossless transmission line through a matching network, what is the required minimum number of degrees of freedom that the network should provide?

Q2.25 In the case of the single-stub matching network, what are the two degrees of freedom?

Q2.26 When a transmission line is matched to a load through a single-stub matching network, no waves will be reflected toward the generator. What happens to the waves reflected by the load and by the shorted stub when they arrive at terminals MM' in Fig. 2-29?

2-11 Transients on Transmission Lines

Thus far, our treatment of wave propagation on transmission lines has focused on the analysis of single-frequency, time-harmonic signals under steady-state conditions. The tools we developed—including the impedance-matching techniques and the use of the Smith chart—are useful for a wide range of applications, but they are inappropriate for dealing with digital or wideband signals on telephone lines or in a computer network. For such signals, we need to examine their transient behavior as a function of time. The **transient response** of a voltage pulse on a transmission line is a time record of its back and forth travel between the sending and receiving ends of the line, taking into account all the multiple reflections (echoes) at both ends.

Let us start by considering the simple case of a single rectangular pulse of amplitude V_0 and duration τ , as shown in Fig. 2-32(a). The amplitude of the pulse is zero prior to $t = 0$, V_0 over the duration $0 \leq t \leq \tau$, and again zero afterward. The pulse can be described mathematically as the sum of two unit step functions:

$$\begin{aligned} V(t) &= V_1(t) + V_2(t) \\ &= V_0 U(t) - V_0 U(t - \tau), \end{aligned} \quad (2.122)$$

where the unit step function $U(x)$ is defined in terms of its argument x as

$$U(x) = \begin{cases} 1 & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases} \quad (2.123)$$

The first component, $V_1(t) = V_0 U(t)$, represents a d-c voltage of amplitude V_0 that gets switched on at $t = 0$ and remains that way indefinitely, and the second component, $V_2(t) = -V_0 U(t - \tau)$, represents a d-c voltage of amplitude $-V_0$ that gets switched on at $t = \tau$ and then remains that way indefinitely. As can be seen from Fig. 2-32(b), the sum of the two components is equal to V_0 for $0 \leq t \leq \tau$ and equal to zero for $t > \tau$. This representation of a pulse in terms of two step functions allows us to analyze the transient behavior of the pulse on a transmission line as the superposition of two d-c signals. Hence, if we can develop basic tools for describing the transient behavior of a single step function, we can apply the same tools for each of the two components of the pulse and then add the results appropriately.

2-11.1 Transient Response

The circuit shown in Fig. 2-33(a) consists of a generator composed of a d-c voltage source V_g and a series resistance R_g connected to a lossless transmission line of length l and characteristic impedance Z_0 . The line is terminated in a purely resistive load Z_L at $z = l$. Hence, all impedances in the circuit are real.

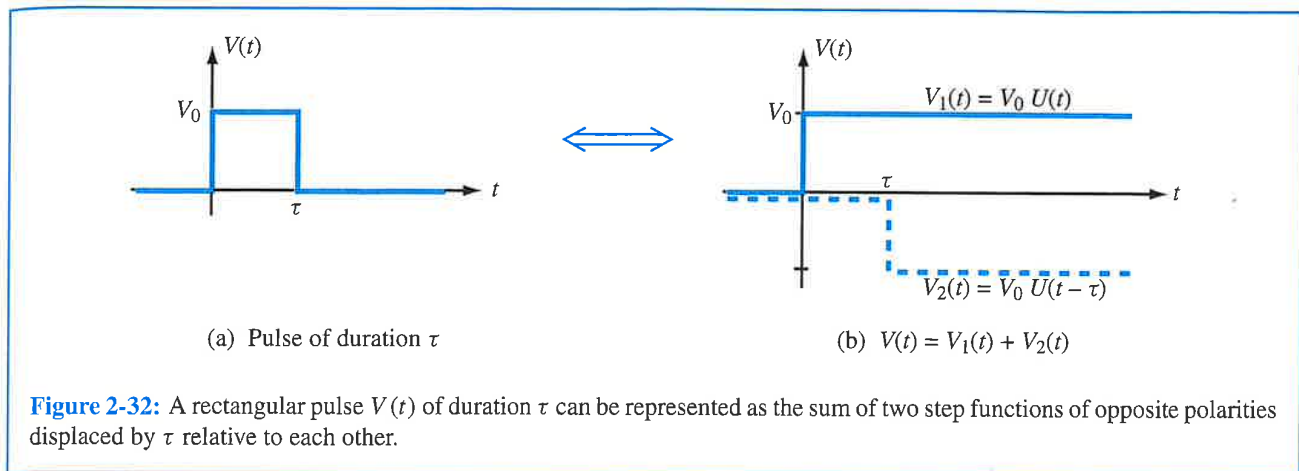


Figure 2-32: A rectangular pulse $V(t)$ of duration τ can be represented as the sum of two step functions of opposite polarities displaced by τ relative to each other.

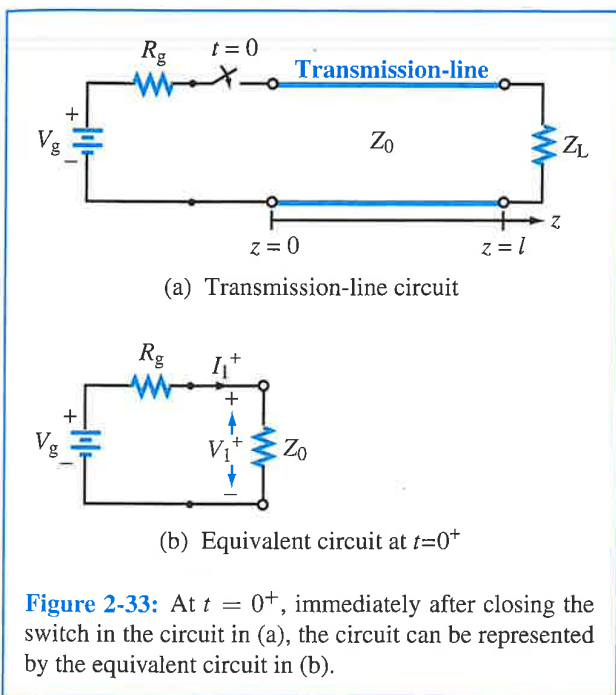


Figure 2-33: At $t = 0^+$, immediately after closing the switch in the circuit in (a), the circuit can be represented by the equivalent circuit in (b).

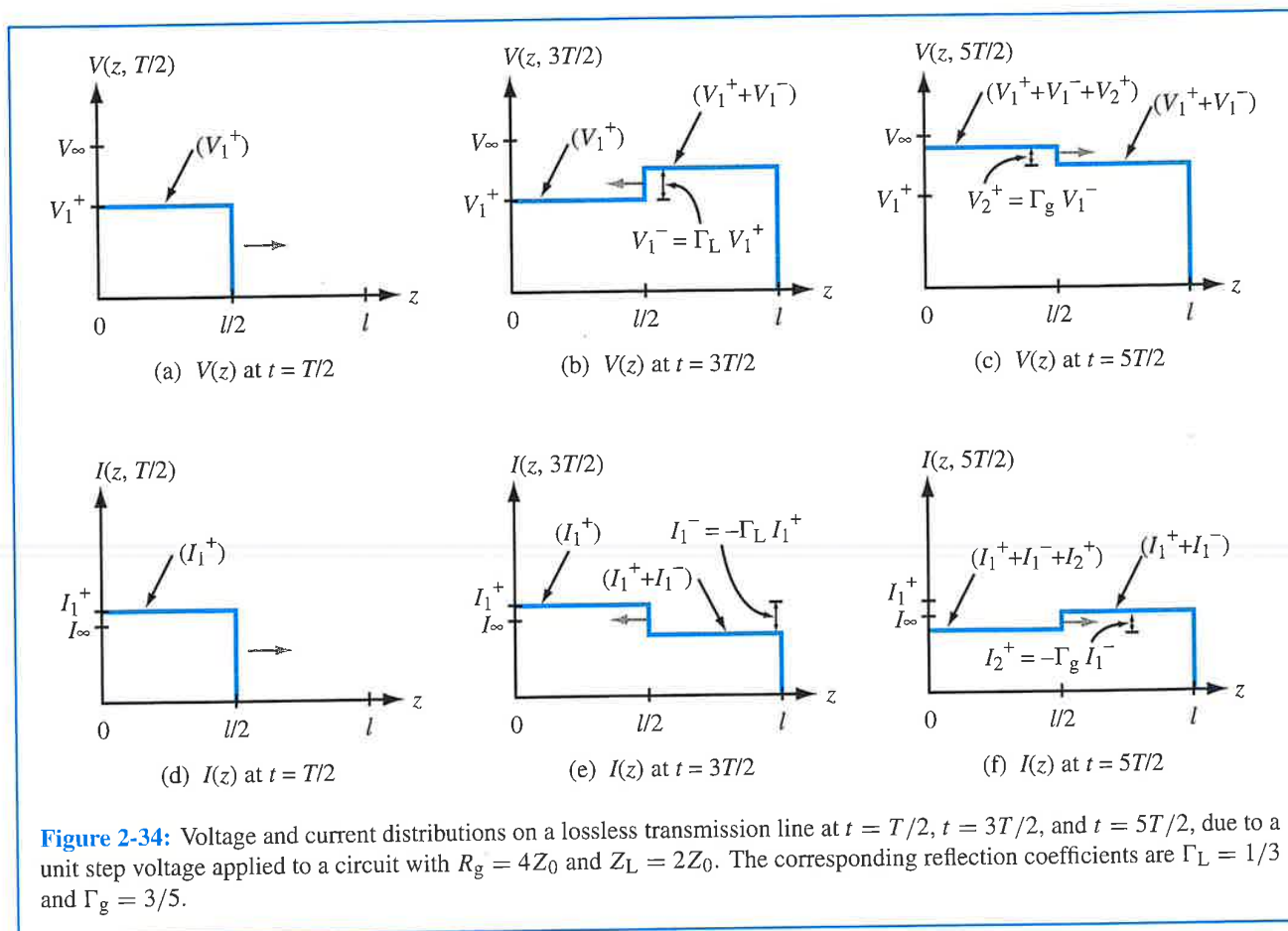
The switch between the generator circuit and the transmission line is switched on at $t = 0$. At the instant

the switch is closed, the transmission line appears to the generator circuit as a load with impedance Z_0 , the characteristic impedance of the line. This is because, in the absence of a signal on the line, the input impedance of the line is unaffected by the load impedance Z_L . The circuit representing the *initial condition* is shown in Fig. 2-33(b). The *initial current* I_1^+ and corresponding *initial voltage* V_1^+ at the input end of the transmission line are given by

$$I_1^+ = \frac{V_g}{R_g + Z_0}, \quad (2.124a)$$

$$V_1^+ = I_1^+ Z_0 = \frac{V_g Z_0}{R_g + Z_0}. \quad (2.124b)$$

The combination of V_1^+ and I_1^+ constitutes a wave that starts to travel along the line with a velocity $u_p = 1/\sqrt{\mu\epsilon}$, immediately after the instant at which the switch is closed. The plus-sign superscript denotes the fact that the wave is traveling in the $+z$ -direction. The transient response of the wave is shown in Fig. 2-34 at each of three instances in time for a circuit with



$R_g = 4Z_0$ and $Z_L = 2Z_0$. The first response is at time $t_1 = T/2$, where $T = l/u_p$ is the time it takes the wave to travel the full length of the line. By time t_1 , the wave has traveled half-way down the line; consequently, the voltage on the first half of the line is equal to V_1^+ , and the voltage on the second half is still zero [Fig. 2-34(a)]. At $t = T$, the wave reaches the load at $z = l$, and because $Z_L \neq Z_0$, the mismatch generates a reflected wave with amplitude

$$V_1^- = \Gamma_L V_1^+, \quad (2.125)$$

where

$$\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0} \quad (2.126)$$

is the reflection coefficient of the load. For the specific case illustrated in Fig. 2-34, $Z_L = 2Z_0$, which results in $\Gamma_L = 1/3$. After this first reflection, the voltage on the line consists of the sum of two waves, the initial wave V_1^+ and the reflected wave V_1^- . The voltage on the transmission line at $t_2 = 3T/2$ is shown in Fig. 2-34(b); $V(z, 3T/2)$ is equal to V_1^+ on the first half of the line

($0 \leq z < l/2$), and it is equal to $(V_1^+ + V_1^-)$ on the second half ($l/2 \leq z \leq l$).

At $t = 2T$, the reflected wave V_1^- arrives at the sending end of the line. If $R_g \neq Z_0$, the mismatch at the sending end generates a reflection at $z = 0$ in the form of a wave with voltage amplitude V_2^+ given by

$$V_2^+ = \Gamma_g V_1^- = \Gamma_g \Gamma_L V_1^+, \quad (2.127)$$

where

$$\Gamma_g = \frac{R_g - Z_0}{R_g + Z_0} \quad (2.128)$$

is the reflection coefficient of the generator resistance R_g . For $R_g = 4Z_0$, we have $\Gamma_g = 0.6$. As time progresses after $t = 2T$, the wave V_2^+ travels down the line toward the load and, as it does that, it adds to the previous voltage condition on the line. Hence, at $t = 5T/2$, the total voltage on the first half of the line is

$$\begin{aligned} V(z, 5T/2) &= V_1^+ + V_1^- + V_2^+ \\ &= (1 + \Gamma_L + \Gamma_L \Gamma_g) V_1^+ \\ &\quad (0 \leq z < l/2), \end{aligned} \quad (2.129a)$$

and on the second half of the line the voltage is

$$\begin{aligned} V(z, 5T/2) &= V_1^+ + V_1^- \\ &= (1 + \Gamma_L) V_1^+ \quad (l/2 \leq z \leq l). \end{aligned} \quad (2.129b)$$

The voltage distribution is shown in Fig. 2-34(c).

So far, we have examined the transient response of only the voltage wave $V(z, t)$. The associated transient response of the current $I(z, t)$ is shown in Figs. 2-34(d)–(f). The process is similar to that we described for the voltage $V(z, t)$, except for one important difference. Whereas at either end of the line the reflected voltage is related to the incident voltage by the reflection coefficient at that end, the reflected current is related to the incident current by the negative of the

reflection coefficient. This property of wave reflection is expressed by Eq. (2.49b). Accordingly,

$$I_1^- = -\Gamma_L I_1^+, \quad (2.130a)$$

$$I_2^+ = -\Gamma_g I_1^- = \Gamma_g \Gamma_L I_1^+, \quad (2.130b)$$

and so on.

The multiple-reflection process continues indefinitely, and the ultimate value that $V(z, t)$ reaches as t approaches ∞ is the same at all locations on the transmission line and is given by

$$\begin{aligned} V_\infty &= V_1^+ + V_1^- + V_2^+ + V_2^- + V_3^+ + V_3^- + \cdots \\ &= V_1^+ [1 + \Gamma_L + \Gamma_L \Gamma_g + \Gamma_L^2 \Gamma_g + \Gamma_L^2 \Gamma_g^2 + \Gamma_L^3 \Gamma_g^2 + \cdots] \\ &= V_1^+ [(1 + \Gamma_L)(1 + \Gamma_L \Gamma_g + \Gamma_L^2 \Gamma_g^2 + \cdots)] \\ &= V_1^+ (1 + \Gamma_L) [1 + x + x^2 + \cdots], \end{aligned} \quad (2.131)$$

where $x = \Gamma_L \Gamma_g$. The series inside the square bracket is the binomial series of the function

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots \quad \text{for } |x| < 1. \quad (2.132)$$

Hence, Eq. (2.131) can be rewritten in the compact form

$$V_\infty = V_1^+ \frac{1 + \Gamma_L}{1 - \Gamma_L \Gamma_g}. \quad (2.133)$$

Upon replacing V_1^+ , Γ_L , and Γ_g with the expressions given by Eqs. (2.124b), (2.126), and (2.128), respectively, and then simplifying the resulting expression, we obtain

$$V_\infty = \frac{V_g Z_L}{R_g + Z_L}. \quad (2.134)$$

The voltage V_∞ is called the **steady-state voltage** on the line, and its expression is exactly what we should expect on the basis of d-c analysis of the circuit in Fig. 2-33(a) if we were to treat the transmission line as simply a

connecting wire between the generator circuit and the load. The corresponding *steady-state current* is

$$I_{\infty} = \frac{V_{\infty}}{Z_L} = \frac{V_g}{R_g + Z_L} \quad (2.135)$$

2-11.2 Bounce Diagrams

Keeping track of the voltage and current waves as they bounce back and forth on the line is a rather tedious process. The *bounce diagram* is a graphical presentation that allows us to accomplish the same goal, but with relative ease. The horizontal axis in Figs. 2-35(a) and (b) represents position along the transmission line, and the vertical axis denotes time. Figure 2-35(a) pertains to $V(z, t)$ and part (b) pertains to $I(z, t)$. The bounce diagram in Fig. 2-35(a) consists of a zigzag line indicating the progress of the voltage wave on the line. The incident wave V_1^+ starts at $z = t = 0$ and travels in the $+z$ -direction until it reaches the load at $z = l$ at time $t = T$. At the very top of the bounce diagram, the reflection coefficients are indicated by $\Gamma = \Gamma_g$ at the generator end and by $\Gamma = \Gamma_L$ at the load end. At the end of the first straight-line segment of the zigzag line, a second line is drawn to represent the reflected voltage wave $V_1^- = \Gamma_L V_1^+$. The amplitude of each new straight-line segment is equal to the product of the amplitude of the preceding straight-line segment and the reflection coefficient at that end of the line. The bounce diagram for the current $I(z, t)$ in Fig. 2-35(b) follows the same procedure as for the voltage except for the reversal of the signs of Γ_L and Γ_g at the top of the bounce diagram.

Using the bounce diagram, the total voltage (or current) at any point z_1 and time t_1 can be determined by drawing a vertical line through the point z_1 , then adding the voltages (or currents) of all the zigzag segments intersected by that line between $t = 0$ and $t = t_1$. To find the voltage

at $z = l/4$ and $T = 4T$, for example, we draw a dashed vertical line in Fig. 2-35(a) through $z = l/4$ and we extend it from $t = 0$ to $t = 4T$. The dashed line intersects four line segments. The total voltage at $z = l/4$ and $t = 4T$ is therefore given by

$$\begin{aligned} V(l/4, 4T) &= V_1^+ + \Gamma_L V_1^+ + \Gamma_g \Gamma_L V_1^+ + \Gamma_g \Gamma_L^2 V_1^+ \\ &= V_1^+ (1 + \Gamma_L + \Gamma_g \Gamma_L + \Gamma_g \Gamma_L^2). \end{aligned}$$

The time variation of V at a specific location z can be obtained by plotting the values of V along the (dashed) vertical line passing through z . Figure 2-35(c) shows the variation of V as a function of time at $z = l/4$ for a circuit with $\Gamma_g = 3/5$ and $\Gamma_L = 1/3$.

Example 2-13 Time-Domain Reflectometer

A time-domain reflectometer (TDR) is an instrument used to locate faults on a transmission line. Consider, for example, a long underground or undersea cable that gets damaged at some distance d from the sending end of the line. The damage may alter the electrical properties or the shape of the cable, causing it to exhibit an impedance Z_L at the fault location that is different from its characteristic impedance, Z_0 . A TDR sends a step voltage down the line, and by observing the voltage at the sending end as a function of time, it is possible to determine the location of the fault and its severity.

If the voltage waveform shown in Fig. 2-36(a) is seen on an oscilloscope connected to the input of a $75\text{-}\Omega$ matched transmission line, determine (a) the generator voltage, (b) the location of the fault, and (c) the fault shunt resistance. The line's insulating material is Teflon with $\epsilon_r = 2.1$.

Solution: (a) Since the line is properly matched, it means that $R_g = Z_L = Z_0$. In Fig. 2-36(b), the fault located at distance d from the sending end is represented by a shunt resistance R_f . For a matched line, Eq. (2.124b) gives

$$V_1^+ = \frac{V_g Z_0}{R_g + Z_0} = \frac{V_g Z_0}{2Z_0} = \frac{V_g}{2}.$$

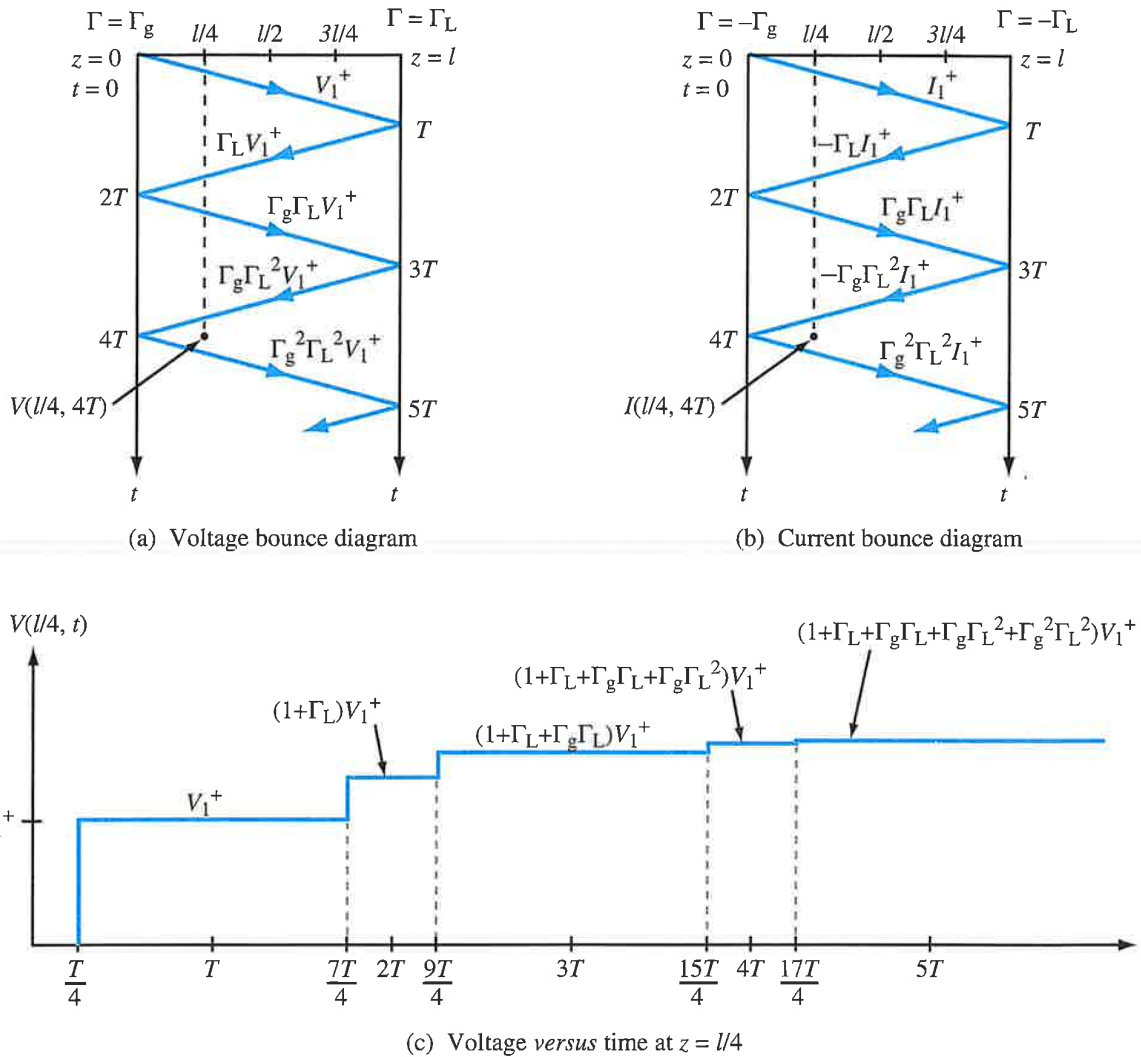


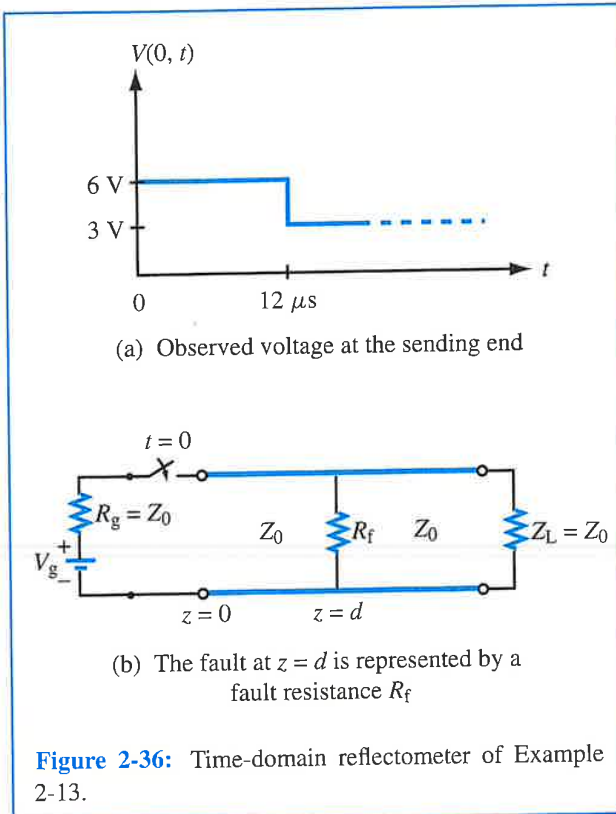
Figure 2-35: Bounce diagrams for (a) voltage and (b) current. In (c), the voltage variation with time at $z = l/4$ for a circuit with $\Gamma_g = 3/5$ and $\Gamma_L = 1/3$ is deduced from the vertical dashed line at $l/4$ in (a).

According to Fig. 2-36(a), $V_1^+ = 6$ V. Hence,

$$V_g = 2V_1^+ = 12 \text{ V.}$$

(b) The propagation velocity on the line is

$$u_p = \frac{c}{\sqrt{\epsilon_r}} = \frac{3 \times 10^8}{\sqrt{2.1}} = 2.07 \times 10^8 \text{ m/s.}$$



For a fault at a distance d , the round-trip time delay of the echo is

$$\Delta t = \frac{2d}{u_p}.$$

From Fig. 2-36(a), $\Delta t = 12 \mu\text{s}$. Hence,

$$d = \frac{\Delta t}{2} u_p = \frac{12 \times 10^{-6}}{2} \times 2.07 \times 10^8 = 1,242 \text{ m}.$$

(c) The change in level of $V(0, t)$ shown in Fig. 2-36(a) represents V_1^- . Thus,

$$V_1^- = \Gamma_f V_1^+ = -3 \text{ V},$$

or

$$\Gamma_f = \frac{-3}{6} = -0.5,$$

where Γ_f is the reflection coefficient due to the fault load Z_{Lf} that appears at $z = d$.

From Eq. (2.49a),

$$\Gamma_f = \frac{Z_{Lf} - Z_0}{Z_{Lf} + Z_0},$$

and it follows that $Z_{Lf} = 25 \Omega$. This fault load is composed of the fault shunt resistance R_f and the characteristic impedance Z_0 of the line to the right of the fault:

$$\frac{1}{Z_{Lf}} = \frac{1}{R_f} + \frac{1}{Z_0},$$

so that the shunt resistance is 37.5Ω . ■

REVIEW QUESTIONS

Q2.27 What is transient analysis used for?

Q2.28 The transient analysis presented in this section was for a step voltage. How does one use it for analyzing the response to a pulse?

Q2.29 What is the difference between the bounce diagram for voltage and the bounce diagram for current?

CHAPTER HIGHLIGHTS

- A transmission line is a two-port network connecting a generator to a load. EM waves traveling on the line may experience ohmic power losses, dispersive effects, and reflections at the generator and load ends of the line. These transmission-line effects may be ignored if the line length is much shorter than λ .

- TEM transmission lines consist of two conductors that can support the propagation of transverse electromagnetic waves characterized by electric and magnetic fields that are transverse to the direction of propagation. TEM lines may be represented by a lumped-element model consisting of four line parameters (R' , L' , G' , and C') whose values are specified by the specific line geometry, the constitutive parameters of the conductors and of the insulating material between them, and the angular frequency ω .
- Wave propagation on a transmission line, which is represented by the phasor voltage $\tilde{V}(z)$ and associated current $\tilde{I}(z)$, is governed by the propagation constant of the line, $\gamma = \alpha + j\beta$, and its characteristic impedance Z_0 . Both γ and Z_0 are specified by ω and the four line parameters.
- If $R' = G' = 0$, the line becomes lossless ($\alpha = 0$). A lossless line is nondispersive, meaning that the phase velocity of a wave is independent of its oscillation frequency.
- In general, a line supports two waves, an incident wave supplied by the generator and another wave reflected by the load. The sum of the two waves generates a standing-wave pattern with a period of $\lambda/2$. The voltage standing-wave ratio S , which is equal to the ratio of the maximum to minimum voltage magnitude on the line, varies between 1 for a matched load ($Z_L = Z_0$) to ∞ for a line terminated in an open circuit, a short circuit, or a purely reactive load.
- The input impedance of a line terminated in a short circuit or open circuit is purely reactive. This property can be used to design equivalent inductors and capacitors.
- The fraction of the incident power delivered to the load by a lossless line is equal to $(1 - |\Gamma|^2)$.
- The Smith chart is a useful graphical technique for analyzing transmission-line problems and for designing impedance-matching networks.
- Matching networks are placed between the load and the feed transmission line for the purpose of eliminating reflections toward the generator. A matching network may consist of lumped elements in the form of capacitors or inductors, or it may consist of sections of transmission lines with appropriate lengths and terminations.
- Transient analysis of pulses can be performed using a bounce-diagram graphical technique that tracks reflections at both the load and generator ends of the transmission line.

GLOSSARY OF IMPORTANT TERMS

Provide definitions or explain the meaning of the following terms:

dispersive transmission line
 TEM transmission lines
 higher-order transmission lines
 lumped-element model
 transmission-line parameters
 intrinsic resistance R_s
 perfect conductor
 perfect dielectric
 air line
 telegrapher's equations
 complex propagation constant γ
 attenuation constant α
 phase constant β
 characteristic impedance Z_0
 lossless line
 standing wave
 distortionless line
 voltage reflection coefficient Γ
 matched transmission line
 standing-wave pattern
 in-phase
 phase opposition
 voltage maxima and minima

current maxima and minima
 voltage standing-wave ratio (VSWR or SWR) S
 slotted line
 input impedance Z_{in}
 short-circuited line
 open-circuited line
 quarter-wave transformer
 Smith chart
 unit circle
 normalized impedance
 normalized load resistance r_L
 normalized load reactance x_L
 WTG and WTL
 SWR circle
 admittance Y
 conductance G and susceptance B
 impedance matching
 matching network
 single-stub matching
 transient response
 bounce diagram

PROBLEMS

Sections 2-1 to 2-4: Transmission-Line Model

2.1* A transmission line of length l connects a load to a sinusoidal voltage source with an oscillation frequency f . Assuming that the velocity of wave propagation on the line is c , for which of the following situations is it reasonable to ignore the presence of the transmission line in the solution of the circuit:

- (a) $l = 20$ cm, $f = 10$ kHz

* Answer(s) available in Appendix D.

 Solution available in CD-ROM.

- (b) $l = 50$ km, $f = 60$ Hz

- (c) $l = 20$ cm, $f = 300$ MHz

- (d) $l = 1$ mm, $f = 100$ GHz

2.2 Calculate the line parameters R' , L' , G' , and C' for a coaxial line with an inner conductor diameter of 0.5 cm and an outer conductor diameter of 1 cm, filled with an insulating material where $\mu = \mu_0$, $\epsilon_r = 2.25$, and $\sigma = 10^{-3}$ S/m. The conductors are made of copper with $\mu_c = \mu_0$ and $\sigma_c = 5.8 \times 10^7$ S/m. The operating frequency is 1 GHz.

2.3* A 1-GHz parallel-plate transmission line consists of 1.5-cm-wide copper strips separated by a 0.2-cm-thick layer of polystyrene. Appendix B gives $\mu_c = \mu_0 = 4\pi \times 10^{-7}$ (H/m) and $\sigma_c = 5.8 \times 10^7$ (S/m) for copper, and $\epsilon_r = 2.6$ for polystyrene. Use Table 2-1 to determine the line parameters of the transmission line. Assume that $\mu = \mu_0$ and $\sigma \simeq 0$ for polystyrene.

2.4 Show that the transmission-line model shown in Fig. 2-37 yields the same telegrapher's equations given by Eqs. (2.14) and (2.16).

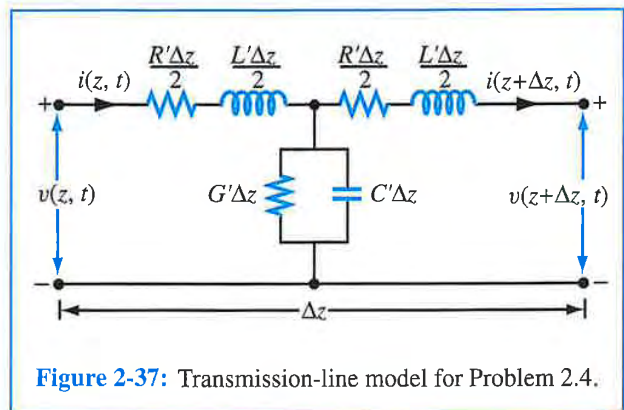


Figure 2-37: Transmission-line model for Problem 2.4.

2.5* Find α , β , u_p , and Z_0 for the coaxial line of Problem 2.2.

Section 2-5: The Lossless Line

2.6 In addition to not dissipating power, a lossless line has two important features: (1) it is dispersionless (u_p is independent of frequency); and (2) its characteristic impedance Z_0 is purely real. Sometimes, it is not possible to design a transmission line such that $R' \ll \omega L'$ and $G' \ll \omega C'$, but it is possible to choose the dimensions of the line and its material properties so as to satisfy the condition

$$R'C' = L'G' \quad (\text{distortionless line})$$

Such a line is called a **distortionless** line, because despite the fact that it is not lossless, it nonetheless possesses the previously mentioned features of the lossless line. Show that for a distortionless line,

$$\begin{aligned}\alpha &= R' \sqrt{\frac{C'}{L'}} = \sqrt{R'G'} \\ \beta &= \omega \sqrt{L'C'} \\ Z_0 &= \sqrt{\frac{L'}{C'}}\end{aligned}$$

2.7* For a distortionless line [see Problem 2.6] with $Z_0 = 50 \, \Omega$, $\alpha = 40 \, (\text{mNp/m})$, and $u_p = 2.5 \times 10^8 \, (\text{m/s})$, find the line parameters and λ at 250 MHz.

2.8 Find α and Z_0 of a distortionless line whose $R' = 4 \, \Omega/\text{m}$ and $G' = 4 \times 10^{-4} \, \text{S/m}$.

 **2.9*** A transmission line operating at 125 MHz has $Z_0 = 40 \, \Omega$, $\alpha = 0.02 \, (\text{Np/m})$, and $\beta = 0.75 \, \text{rad/m}$. Find the line parameters R' , L' , G' , and C' .

2.10 Using a slotted line, the voltage on a lossless transmission line was found to have a maximum magnitude of 1.5 V and a minimum magnitude of 0.8 V. Find the magnitude of the load's reflection coefficient.

2.11* Polyethylene with $\epsilon_r = 2.25$ is used as the insulating material in a lossless coaxial line with a characteristic impedance of $50 \, \Omega$. The radius of the inner conductor is 1 mm.

- (a) What is the radius of the outer conductor?
- (b) What is the phase velocity of the line?

2.12 A $50\text{-}\Omega$ lossless transmission line is terminated in a load with impedance $Z_L = (30 - j60) \, \Omega$. The wavelength is 5 cm. Find the following:

- (a) The reflection coefficient at the load.
- (b) The standing-wave ratio on the line.
- (c) The position of the voltage maximum nearest the load.
- (d) The position of the current maximum nearest the load.

2.13* On a $150\text{-}\Omega$ lossless transmission line, the following observations were noted: distance of first voltage minimum from the load = 3 cm; distance of first voltage maximum from the load = 9 cm; $S = 3$. Find Z_L .

2.14 Using a slotted line, the following results were obtained: distance of first minimum from the load = 4 cm; distance of second minimum from the load = 14 cm; voltage standing-wave ratio = 2.5. If the line is lossless and $Z_0 = 50 \, \Omega$, find the load impedance.

 **2.15*** A load with impedance $Z_L = (25 - j50) \, \Omega$ is to be connected to a lossless transmission line with characteristic impedance Z_0 , with Z_0 chosen such that the standing-wave ratio is the smallest possible. What should Z_0 be?

2.16 A $50\text{-}\Omega$ lossless line terminated in a purely resistive load has a voltage standing-wave ratio of 4. Find all possible values of Z_L .

Section 2-6: Input Impedance

2.17* At an operating frequency of 300 MHz, a lossless 50- Ω air-spaced transmission line 2.5 m in length is terminated with an impedance $Z_L = (60 + j20) \Omega$. Find the input impedance.

2.18 A lossless transmission line of electrical length $l = 0.35\lambda$ is terminated in a load impedance as shown in Fig. 2-38. Find Γ , S , and Z_{in} .

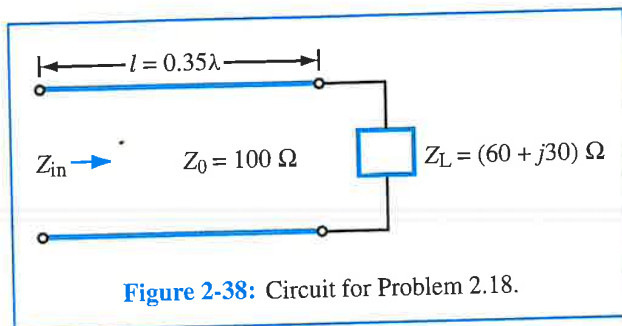


Figure 2-38: Circuit for Problem 2.18.

2.19 Show that the input impedance of a quarter-wavelength-long lossless line terminated in a short circuit appears as an open circuit.

2.20 Show that at the position where the magnitude of the voltage on the line is a maximum, the input impedance is purely real.

2.21* A voltage generator with

$$v_g(t) = 5 \cos(2\pi \times 10^9 t) \text{ V}$$

and internal impedance $Z_g = 50 \Omega$ is connected to a 50- Ω lossless air-spaced transmission line. The line length is 5 cm and it is terminated in a load with impedance $Z_L = (100 - j100) \Omega$. Find the following:

- (a) Γ at the load.
- (b) Z_{in} at the input to the transmission line.

(c) The input voltage $\tilde{V}_i$ and input current $\tilde{I}_i$.

2.22 A 6-m section of 150- Ω lossless line is driven by a source with

$$v_g(t) = 5 \cos(8\pi \times 10^7 t - 30^\circ) \text{ (V)}$$

and $Z_g = 150 \Omega$. If the line, which has a relative permittivity $\epsilon_r = 2.25$, is terminated in a load $Z_L = (150 - j50) \Omega$, find the following:

- (a) λ on the line.
- (b) The reflection coefficient at the load.
- (c) The input impedance.
- (d) The input voltage $\tilde{V}_i$.
- (e) The time-domain input voltage $v_i(t)$.

2.23* Two half-wave dipole antennas, each with an impedance of 75 Ω , are connected in parallel through a pair of transmission lines, and the combination is connected to a feed transmission line, as shown in Fig. 2-39. All lines are 50 Ω and lossless.

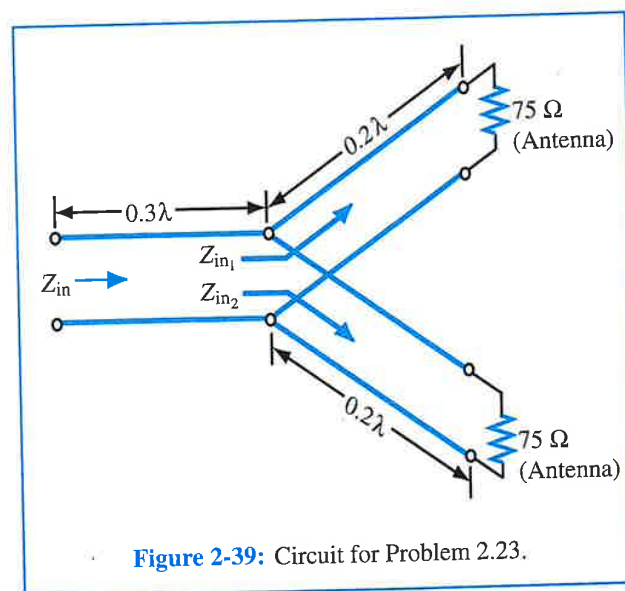


Figure 2-39: Circuit for Problem 2.23.

- (a) Calculate Z_{in_1} , the input impedance of the antenna-terminated line, at the parallel juncture.
- (b) Combine Z_{in_1} and Z_{in_2} in parallel to obtain Z'_L , the effective load impedance of the feedline.
- (c) Calculate Z_{in} of the feedline.

Section 2-7: Special Cases

2.24 At an operating frequency of 200 MHz, it is desired to use a section of a lossless 50- Ω transmission line terminated in a short circuit to construct an equivalent load with reactance $X = 25 \Omega$. If the phase velocity of the line is $0.75c$, what is the shortest possible line length that would exhibit the desired reactance at its input?

2.25* A lossless transmission line is terminated in a short circuit. How long (in wavelengths) should the line be for it to appear as an open circuit at its input terminals?

2.26 The input impedance of a 31-cm-long lossless transmission line of unknown characteristic impedance was measured at 1 MHz. With the line terminated in a short circuit, the measurement yielded an input impedance equivalent to an inductor with inductance of $0.128 \mu\text{H}$, and when the line was open-circuited, the measurement yielded an input impedance equivalent to a capacitor with capacitance of 20 pF. Find Z_0 of the line, the phase velocity, and the relative permittivity of the insulating material.

2.27* A 60- Ω resistive load is preceded by a $\lambda/4$ section of a 50- Ω lossless line, which itself is preceded by another $\lambda/4$ section of a 100- Ω line. What is the input impedance?

2.28 A 100-MHz FM broadcast station uses a 300- Ω transmission line between the transmitter and a tower-mounted half-wave dipole antenna. The antenna impedance is 73Ω . You are asked to design a quarter-wave transformer to match the antenna to the line.

- (a) Determine the electrical length and characteristic impedance of the quarter-wave section.
- (b) If the quarter-wave section is a two-wire line with $d = 2.5 \text{ cm}$, and the spacing between the wires is made of polystyrene with $\epsilon_r = 2.6$, determine the physical length of the quarter-wave section and the radius of the two wire conductors.



2.29* A 50-MHz generator with $Z_g = 50 \Omega$ is connected to a load $Z_L = (50 - j25) \Omega$. The time-average power transferred from the generator into the load is maximum when $Z_g = Z_L^*$, where Z_L^* is the complex conjugate of Z_L . To achieve this condition without changing Z_g , the effective load impedance can be modified by adding an open-circuited line in series with Z_L , as shown in Fig. 2-40. If the line's $Z_0 = 100 \Omega$,

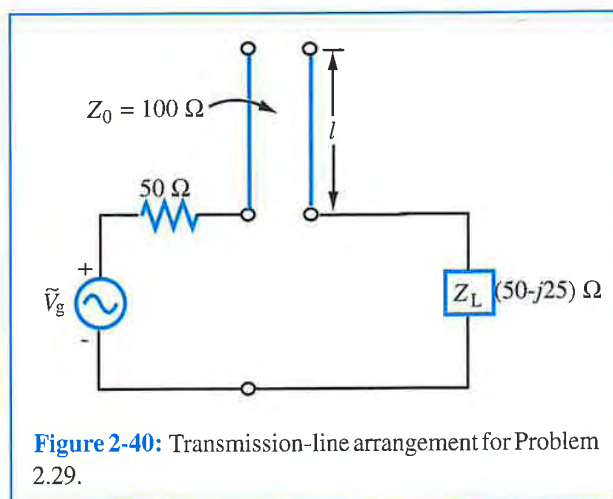


Figure 2-40: Transmission-line arrangement for Problem 2.29.

determine the shortest length of line (in wavelengths) necessary for satisfying the maximum-power-transfer condition.

2.30 A 50- Ω lossless line of length $l = 0.375\lambda$ connects a 200-MHz generator with $\tilde{V}_g = 150 \text{ V}$ and $Z_g = 50 \Omega$ to a load Z_L . Determine the time-domain current through the load for:

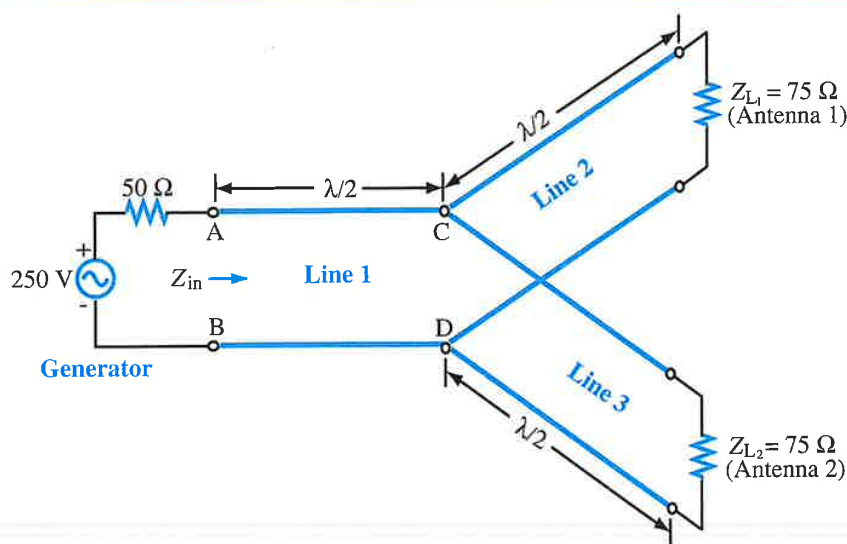


Figure 2-41: Antenna configuration for Problem 2.32.

- (a) $Z_L = (50 - j50) \Omega$
- (b) $Z_L = 50 \Omega$
- (c) $Z_L = 0$ (short circuit)

Section 2-8: Power Flow on Lossless Line

2.31* A generator with $\tilde{V}_g = 100$ V and $Z_g = 50 \Omega$ is connected to a load $Z_L = 75 \Omega$ through a $50\text{-}\Omega$ lossless line of length $l = 0.15\lambda$.

- (a) Compute Z_{in} , the input impedance of the line at the generator end.
- (b) Compute $\tilde{I}_i$ and $\tilde{V}_i$.
- (c) Compute the time-average power delivered to the line, $P_{in} = \frac{1}{2} \Re[\tilde{V}_i \tilde{I}_i^*]$.
- (d) Compute $\tilde{V}_L$, $\tilde{I}_L$, and the time-average power delivered to the load, $P_L = \frac{1}{2} \Re[\tilde{V}_L \tilde{I}_L^*]$. How does P_{in} compare to P_L ? Explain.

- (e) Compute the time-average power delivered by the generator, P_g , and the time-average power dissipated in Z_g . Is conservation of power satisfied?

2.32 If the two-antenna configuration shown in Fig. 2-41 is connected to a generator with $\tilde{V}_g = 250$ V and $Z_g = 50 \Omega$, how much average power is delivered to each antenna?

2.33* For the circuit shown in Fig. 2-42, calculate the average incident power, the average reflected power, and the average power transmitted into the infinite $100\text{-}\Omega$ line. The $\lambda/2$ line is lossless and the infinitely long line is slightly lossy. (Hint: The input impedance of an infinitely long line is equal to its characteristic impedance so long as $\alpha \neq 0$.)

2.34 An antenna with a load impedance

$$Z_L = (75 + j25) \Omega$$

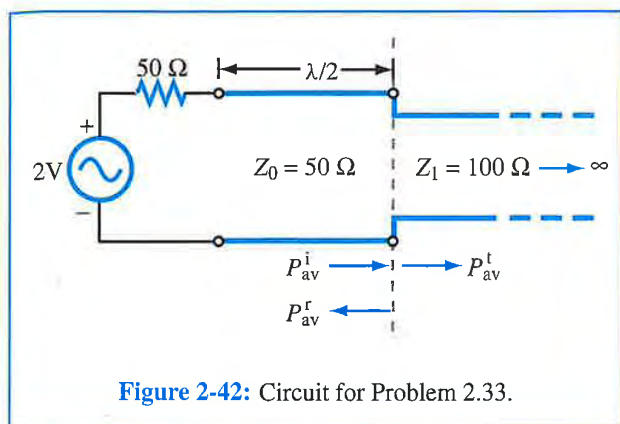


Figure 2-42: Circuit for Problem 2.33.

is connected to a transmitter through a 50-Ω lossless transmission line. If under matched conditions (50-Ω load) the transmitter can deliver 10 W to the load, how much power can it deliver to the antenna? Assume that $Z_g = Z_0$.

Section 2-9: Smith Chart

2.35* Use the Smith chart to find the reflection coefficient corresponding to a load impedance of

- (a) $Z_L = 3Z_0$
- (b) $Z_L = (2 - 2j)Z_0$
- (c) $Z_L = -2jZ_0$
- (d) $Z_L = 0$ (short circuit)

2.36 Use the Smith chart to find the normalized load impedance corresponding to a reflection coefficient of

- (a) $\Gamma = 0.5$
- (b) $\Gamma = 0.5 \angle 60^\circ$
- (c) $\Gamma = -1$
- (d) $\Gamma = 0.3 \angle -30^\circ$
- (e) $\Gamma = 0$
- (f) $\Gamma = j$

2.37* On a lossless transmission line terminated in a load $Z_L = 100 \Omega$, the standing-wave ratio was measured to be 2.5. Use the Smith chart to find the two possible values of Z_0 .

2.38 A lossless 50-Ω transmission line is terminated in a load with $Z_L = (50 + j25) \Omega$. Use the Smith chart to find the following:

- (a) The reflection coefficient Γ .
- (b) The standing-wave ratio.
- (c) The input impedance at 0.35λ from the load.
- (d) The input admittance at 0.35λ from the load.
- (e) The shortest line length for which the input impedance is purely resistive.
- (f) The position of the first voltage maximum from the load.

2.39* A lossless 50-Ω transmission line is terminated in a short circuit. Use the Smith chart to find the following:

- (a) The input impedance at a distance 2.3λ from the load.
- (b) The distance from the load at which the input admittance is $Y_{in} = -j0.04 \text{ S}$.

2.40 Use the Smith chart to find y_L if $z_L = 1.5 - j0.7$.

2.41* A lossless 100-Ω transmission line $3\lambda/8$ in length is terminated in an unknown impedance. If the input impedance is $Z_{in} = -j2.5 \Omega$,

- (a) Use the Smith chart to find Z_L .
- (b) What length of open-circuit line could be used to replace Z_L ?

2.42 A 75-Ω lossless line is 0.6λ long. If $S = 1.8$ and $\theta_r = -60^\circ$, use the Smith chart to find $|\Gamma|$, Z_L , and Z_{in} .

2.43* Using a slotted line on a 50-Ω air-spaced lossless line, the following measurements were obtained:

$S = 1.6$ and $|\tilde{V}|_{\max}$ occurred only at 10 cm and 24 cm from the load. Use the Smith chart to find Z_L .

2.44 At an operating frequency of 5 GHz, a 50- Ω lossless coaxial line with insulating material having a relative permittivity $\epsilon_r = 2.25$ is terminated in an antenna with an impedance $Z_L = 75 \Omega$. Use the Smith chart to find Z_{in} . The line length is 30 cm.

Section 2-10: Impedance Matching

2.45* A 50- Ω lossless line 0.6λ long is terminated in a load with $Z_L = (50 + j25) \Omega$. At 0.3λ from the load, a resistor with resistance $R = 30 \Omega$ is connected as shown in Fig. 2-43. Use the Smith chart to find Z_{in} .

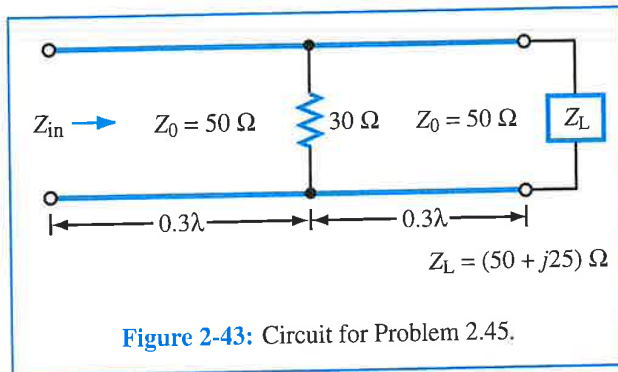


Figure 2-43: Circuit for Problem 2.45.

2.46 A 50- Ω lossless line is to be matched to an antenna with $Z_L = (75 - j20) \Omega$ using a shorted stub. Use the Smith chart to determine the stub length and distance between the antenna and stub.

2.47* Repeat Problem 2.46 for a load with $Z_L = (100 + j50) \Omega$.

2.48 Use the Smith chart to find Z_{in} of the feed line shown in Fig. 2-44. All lines are lossless with $Z_0 = 50 \Omega$.

2.49* Repeat Problem 2.48 for the case where all three transmission lines are $\lambda/4$ in length.

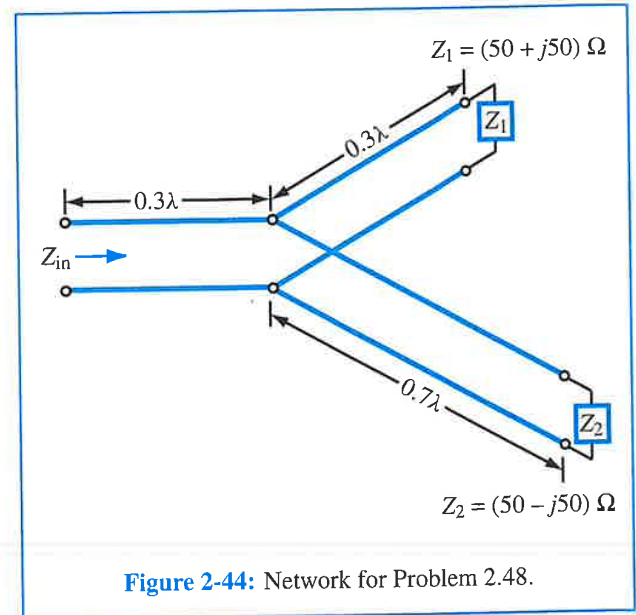


Figure 2-44: Network for Problem 2.48.

Section 2-11: Transients on Transmission Lines

2.50 Generate a bounce diagram for the voltage $V(z, t)$ for a 1-m-long lossless line characterized by $Z_0 = 50 \Omega$ and $u_p = 2c/3$ (where c is the velocity of light) if the line is fed by a step voltage applied at $t = 0$ by a generator circuit with $V_g = 60 \text{ V}$ and $R_g = 100 \Omega$. The line is terminated in a load $Z_L = 25 \Omega$. Use the bounce diagram to plot $V(t)$ at a point midway along the length of the line from $t = 0$ to $t = 25 \text{ ns}$.

2.51 Repeat Problem 2.50 for the current I on the line.

2.52 In response to a step voltage, the voltage waveform shown in Fig. 2-45 was observed at the sending end of a lossless transmission line with $R_g = 50 \Omega$, $Z_0 = 50 \Omega$, and $\epsilon_r = 2.25$. Determine the following:

- The generator voltage.
- The length of the line.
- The load impedance.

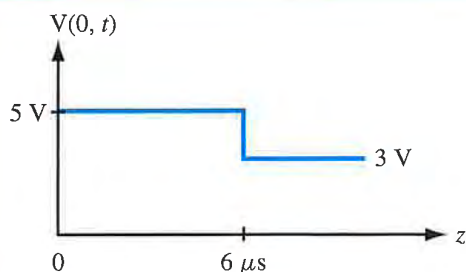


Figure 2-45: Voltage waveform for Problems 2.52 and 2.54.

2.53* In response to a step voltage, the voltage waveform shown in Fig. 2-46 was observed at the sending end of a shorted line with $Z_0 = 50 \Omega$ and $\epsilon_r = 4$. Determine V_g , R_g , and the line length.

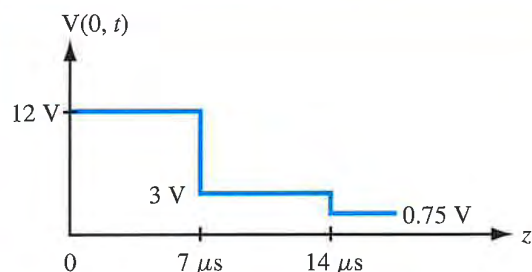


Figure 2-46: Voltage waveform of Problem 2.53.

2.54 Suppose the voltage waveform shown in Fig. 2-45 was observed at the sending end of a $50\text{-}\Omega$ transmission line in response to a step voltage introduced by a generator with $V_g = 15\text{ V}$ and an unknown series resistance R_g . The line is 1 km in length, its velocity of propagation is $1 \times 10^8\text{ m/s}$, and it is terminated in a load $Z_L = 100 \Omega$.

- Determine R_g .
- Explain why the drop in level of $V(0, t)$ at $t = 6 \mu s$ cannot be due to reflection from the load.
- Determine the shunt resistance R_f and location of the fault responsible for the observed waveform.

2.55 A generator circuit with $V_g = 200\text{ V}$ and $R_g = 25 \Omega$ was used to excite a $75\text{-}\Omega$ lossless line with a rectangular pulse of duration $\tau = 0.4 \mu s$. The line is 200 m long, its $u_p = 2 \times 10^8\text{ m/s}$, and it is terminated in a load $Z_L = 125 \Omega$.

- Synthesize the voltage pulse exciting the line as the sum of two step functions, $V_{g1}(t)$ and $V_{g2}(t)$.
- For each voltage step function, generate a bounce diagram for the voltage on the line.
- Use the bounce diagrams to plot the total voltage at the sending end of the line.

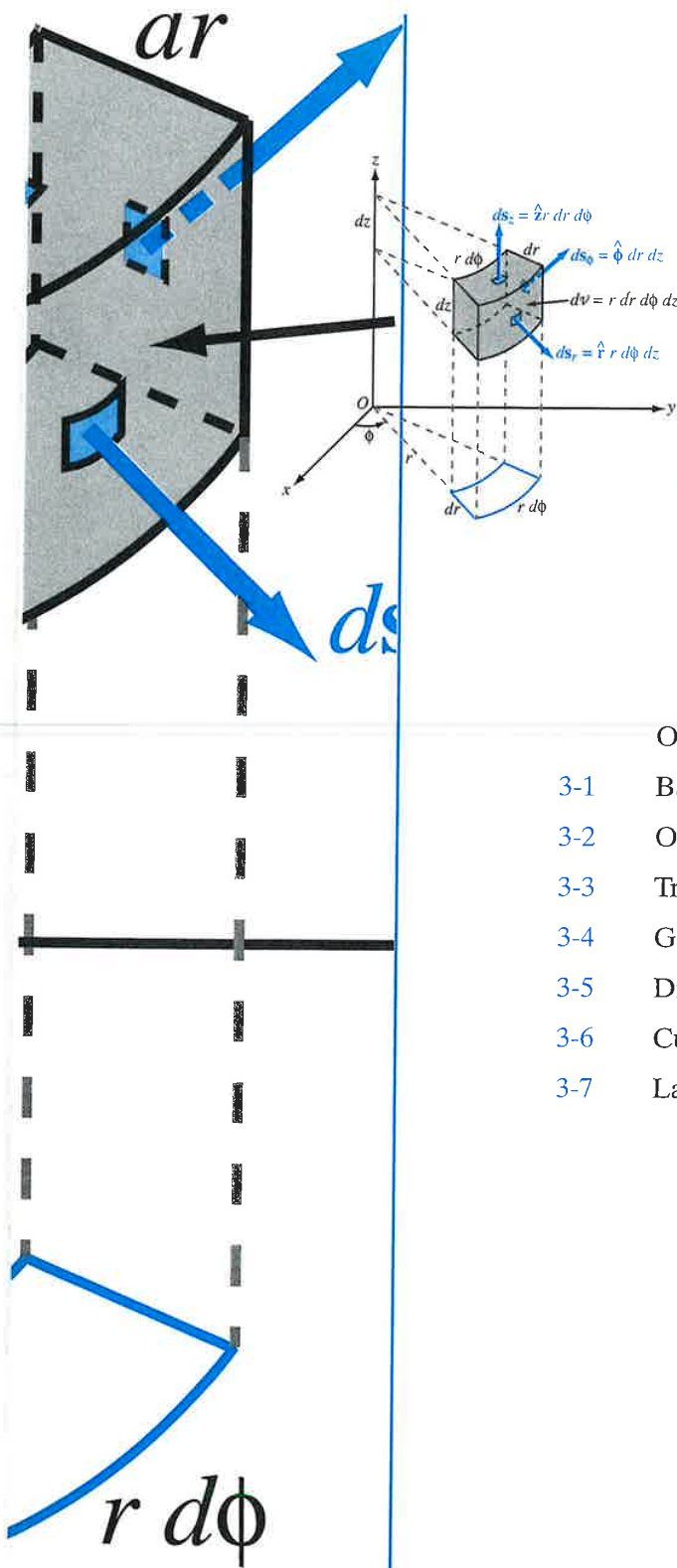
2.56 For the circuit of Problem 2.55, generate a bounce diagram for the current and plot its time history at the middle of the line.

CHAPTER

Vector Analysis

Overview

- 3-1 Basic Laws of Vector Algebra
- 3-2 Orthogonal Coordinate Systems
- 3-3 Transformations between Coordinate Systems
- 3-4 Gradient of a Scalar Field
- 3-5 Divergence of a Vector Field
- 3-6 Curl of a Vector Field
- 3-7 Laplacian Operator



OVERVIEW

In our examination of wave propagation on a transmission line in Chapter 2, the primary quantities we worked with were voltage, current, impedance, and power. Each of these is a **scalar** quantity, meaning that it can be completely specified by its magnitude, if it is a positive real number, or by its magnitude and phase angle if it is a negative or a complex number (a negative number has a positive magnitude and a phase angle of π (rad)). A **vector** specifies both the magnitude and direction of a quantity. The speed of an object is a scalar, whereas its velocity is a vector.

Starting in the next chapter and throughout the succeeding chapters in the book, the electromagnetic quantities we will deal with the most are the electric and magnetic fields, $\mathbf{E}$ and $\mathbf{H}$. These, and many other related quantities, are vectors. Vector analysis provides the mathematical tools necessary for expressing and manipulating vector quantities in an efficient and convenient manner. To specify a vector in three-dimensional space, it is necessary to specify its components along each of the three dimensions. Several types of coordinate systems are used in the study of vector quantities, the most common being the Cartesian (or rectangular), cylindrical, and spherical systems. A particular coordinate system is usually chosen to best suit the geometry of the particular problem under consideration.

Vector algebra governs the laws of addition, subtraction, and multiplication of vectors in any given coordinate system. The rules of vector algebra and vector representation in each of the aforementioned orthogonal coordinate systems (including vector transformation between them) are two of the three major topics treated in this chapter. The third topic is **vector calculus**, which

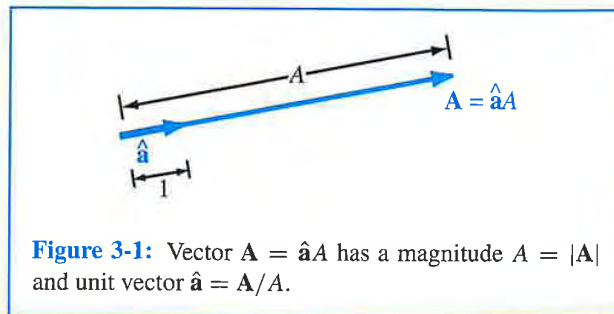


Figure 3-1: Vector $\mathbf{A} = \hat{\mathbf{a}}A$ has a magnitude $A = |\mathbf{A}|$ and unit vector $\hat{\mathbf{a}} = \mathbf{A}/A$.

encompasses the laws of differentiation and integration of vectors, the use of special vector operators (gradient, divergence, and curl), and the application of certain particularly useful theorems, most notably the divergence and Stokes's theorems.

3-1 Basic Laws of Vector Algebra

A vector $\mathbf{A}$ has a **magnitude** $A = |\mathbf{A}|$ and a direction specified by a unit vector $\hat{\mathbf{a}}$:

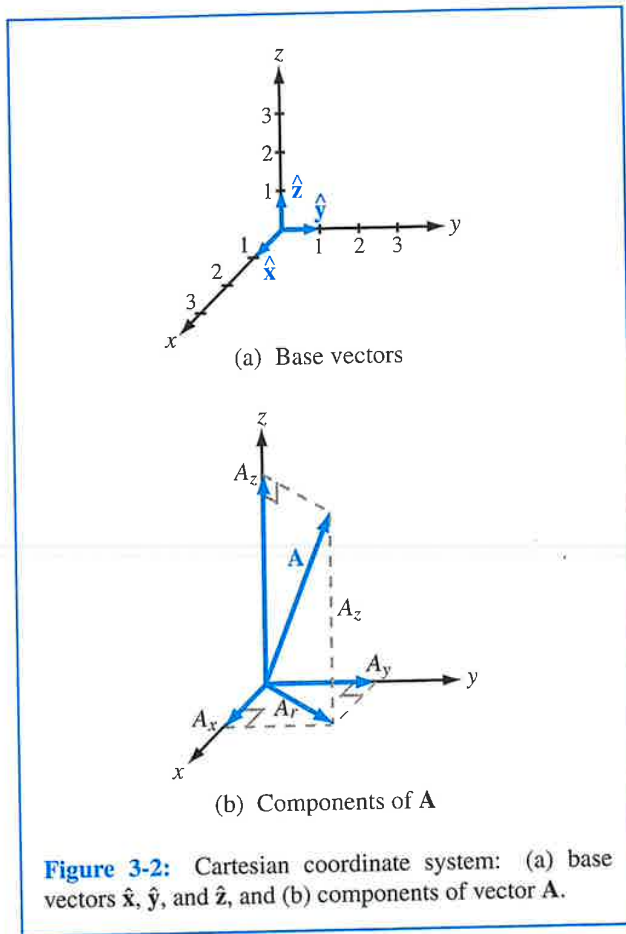
$$\mathbf{A} = \hat{\mathbf{a}}|\mathbf{A}| = \hat{\mathbf{a}}A. \quad (3.1)$$

The **unit vector** $\hat{\mathbf{a}}$ has a magnitude of unity ($|\hat{\mathbf{a}}| = 1$), and its direction is given by

$$\hat{\mathbf{a}} = \frac{\mathbf{A}}{|\mathbf{A}|} = \frac{\mathbf{A}}{A}. \quad (3.2)$$

Figure 3-1 shows a graphical representation of the vector $\mathbf{A}$ as a straight line of length A with its tip pointing in the direction of $\hat{\mathbf{a}}$.

In the Cartesian (or rectangular) coordinate system shown in Fig. 3-2(a), the directions of the x , y ,



and z coordinates are denoted by the three mutually perpendicular unit vectors $\hat{x}$, $\hat{y}$, and $\hat{z}$, which are called the **base vectors**. The vector $\mathbf{A}$ in Fig. 3-2(b) may be represented as

$$\mathbf{A} = \hat{x}A_x + \hat{y}A_y + \hat{z}A_z, \quad (3.3)$$

where A_x , A_y , and A_z are the components of $\mathbf{A}$ along the x -, y -, and z -directions, respectively. The component A_z is equal to the perpendicular projection of $\mathbf{A}$ onto the

z -axis, and similar definitions apply to A_x and A_y . Application of the Pythagorean theorem, first to the right triangle in the x - y plane to express the hypotenuse A_r in terms of A_x and A_y , and then again to the vertical right triangle with sides A_r and A_z and hypotenuse A , gives the following expression for the magnitude of $\mathbf{A}$:

$$A = |\mathbf{A}| = \sqrt{A_x^2 + A_y^2 + A_z^2}. \quad (3.4)$$

Since A is a nonnegative scalar, only the positive root applies. From Eq. (3.2), the unit vector $\hat{\mathbf{a}}$ is given by

$$\hat{\mathbf{a}} = \frac{\mathbf{A}}{A} = \frac{\hat{x}A_x + \hat{y}A_y + \hat{z}A_z}{\sqrt{A_x^2 + A_y^2 + A_z^2}}. \quad (3.5)$$

Occasionally, we shall use the shorthand notation $\mathbf{A} = (A_x, A_y, A_z)$ to denote a vector with components A_x , A_y , and A_z in a Cartesian coordinate system.

3-1.1 Equality of Two Vectors

Two vectors $\mathbf{A}$ and $\mathbf{B}$ are said to be equal if they have equal magnitudes and identical unit vectors. Thus, if

$$\mathbf{A} = \hat{\mathbf{a}}A = \hat{x}A_x + \hat{y}A_y + \hat{z}A_z, \quad (3.6a)$$

$$\mathbf{B} = \hat{\mathbf{b}}B = \hat{x}B_x + \hat{y}B_y + \hat{z}B_z, \quad (3.6b)$$

then $\mathbf{A} = \mathbf{B}$ if and only if $A = B$ and $\hat{\mathbf{a}} = \hat{\mathbf{b}}$, which requires that $A_x = B_x$, $A_y = B_y$, and $A_z = B_z$. *Equality of two vectors does not necessarily imply that they are identical*; in Cartesian coordinates, two displaced parallel vectors of equal magnitude and pointing in the same direction are equal, but they are identical only if they lie on top of one another.

3-1.2 Vector Addition and Subtraction

The sum of two vectors **A** and **B** is a vector **C** given by

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}. \quad (3.7)$$

Graphically, vector addition is obtained by either the parallelogram rule or the head-to-tail rule, as illustrated in Fig. 3-3. With **A** and **B** drawn from the same point, while keeping their magnitudes and directions unchanged, the vector **C** is the diagonal of the parallelogram found by **A** and **B**. With the head-to-tail rule, we may either add **A** to **B** or **B** to **A**. When **A** is added to **B**, it is positioned so that its tail starts at the tip of **B**, again keeping its length and direction unchanged. The sum vector **C** starts at the tail of **B** and ends at the tip of **A**.

If **A** and **B** are given in a rectangular coordinate system by Eqs. (3.6a) and (3.6b), vector addition gives

$$\begin{aligned} \mathbf{C} &= \mathbf{A} + \mathbf{B} \\ &= (\hat{x}A_x + \hat{y}A_y + \hat{z}A_z) + (\hat{x}B_x + \hat{y}B_y + \hat{z}B_z) \\ &= \hat{x}(A_x + B_x) + \hat{y}(A_y + B_y) + \hat{z}(A_z + B_z). \end{aligned} \quad (3.8)$$

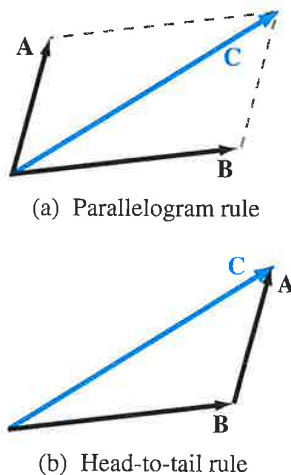


Figure 3-3: Vector addition by (a) the parallelogram rule and (b) the head-to-tail rule.

Subtraction of vector **B** from vector **A** is equivalent to the addition of **A** to negative **B**. Thus,

$$\begin{aligned} \mathbf{D} &= \mathbf{A} - \mathbf{B} \\ &= \mathbf{A} + (-\mathbf{B}) \\ &= \hat{x}(A_x - B_x) + \hat{y}(A_y - B_y) + \hat{z}(A_z - B_z). \end{aligned} \quad (3.9)$$

Graphically, the same rules used for vector addition are also applicable to vector subtraction; the only difference is that the arrowhead of $(-\mathbf{B})$ is drawn on the opposite end of the line segment representing the vector **B** (i.e., the tail and head are interchanged).

3-1.3 Position and Distance Vectors

In a given coordinate system, the *position vector* of a point *P* in space is the vector from the origin to *P*. Points *P*₁ and *P*₂ in Fig. 3-4 are located at (*x*₁, *y*₁, *z*₁)

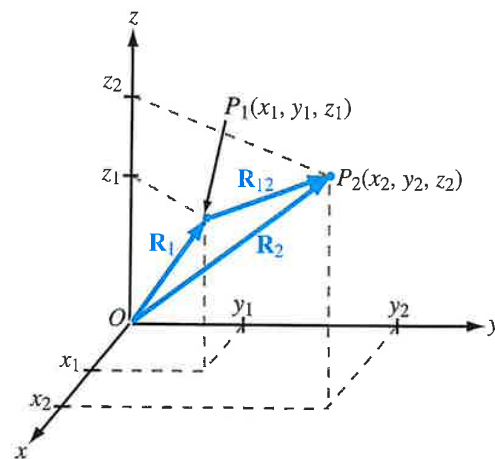


Figure 3-4: Position vector $\mathbf{R}_{12} = \overrightarrow{P_1 P_2} = \mathbf{R}_2 - \mathbf{R}_1$.

and (x_2, y_2, z_2) , respectively. Their position vectors are

$$\mathbf{R}_1 = \overrightarrow{OP_1} = \hat{x}x_1 + \hat{y}y_1 + \hat{z}z_1, \quad (3.10a)$$

$$\mathbf{R}_2 = \overrightarrow{OP_2} = \hat{x}x_2 + \hat{y}y_2 + \hat{z}z_2, \quad (3.10b)$$

where point O is the origin. The **distance vector** from P_1 to P_2 is defined as

$$\begin{aligned} \mathbf{R}_{12} &= \overrightarrow{P_1P_2} \\ &= \mathbf{R}_2 - \mathbf{R}_1 \\ &= \hat{x}(x_2 - x_1) + \hat{y}(y_2 - y_1) + \hat{z}(z_2 - z_1), \end{aligned} \quad (3.11)$$

and the distance d between P_1 and P_2 is equal to the magnitude of $\mathbf{R}_{12}$:

$$\begin{aligned} d &= |\mathbf{R}_{12}| \\ &= [(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2]^{1/2}. \end{aligned} \quad (3.12)$$

Note that the first subscript of $\mathbf{R}_{12}$ denotes the location of the tail of vector $\mathbf{R}_{12}$ and the second subscript denotes the location of its head, as shown in Fig. 3-4.

3-1.4 Vector Multiplication

Three types of products can occur in vector calculus. These are the simple, scalar (or dot), and vector (or cross) products.

Simple Product

Multiplication of a vector by a scalar is called a **simple product**. The product of the vector $\mathbf{A} = \hat{a}A$ by a scalar k results in a vector $\mathbf{B}$ whose magnitude is kA and whose direction is the same as that of $\mathbf{A}$. That is,

$$\mathbf{B} = k\mathbf{A} = \hat{a}kA = \hat{x}(kA_x) + \hat{y}(kA_y) + \hat{z}(kA_z). \quad (3.13)$$

Scalar or Dot Product

The **scalar** (or **dot**) **product** of two vectors $\mathbf{A}$ and $\mathbf{B}$, denoted by $\mathbf{A} \cdot \mathbf{B}$ and pronounced "A dot B," is defined geometrically as the product of the magnitude of one of the vectors and the projection of the other vector onto the first one, or vice versa. Thus,

$$\mathbf{A} \cdot \mathbf{B} = AB \cos \theta_{AB}, \quad (3.14)$$

where θ_{AB} is the angle between $\mathbf{A}$ and $\mathbf{B}$, as shown in Fig. 3-5. The scalar product of two vectors yields a scalar whose magnitude is less than or equal to the products of the magnitudes of the two vectors (equality holds when $\theta_{AB} = 0$) and whose sign is positive if $0 < \theta_{AB} < 90^\circ$ and negative if $90^\circ < \theta_{AB} < 180^\circ$. When $\theta_{AB} = 90^\circ$, the two vectors are orthogonal. The dot product of two orthogonal vectors is zero. The quantity $A \cos \theta_{AB}$ is the component of $\mathbf{A}$ along $\mathbf{B}$ and

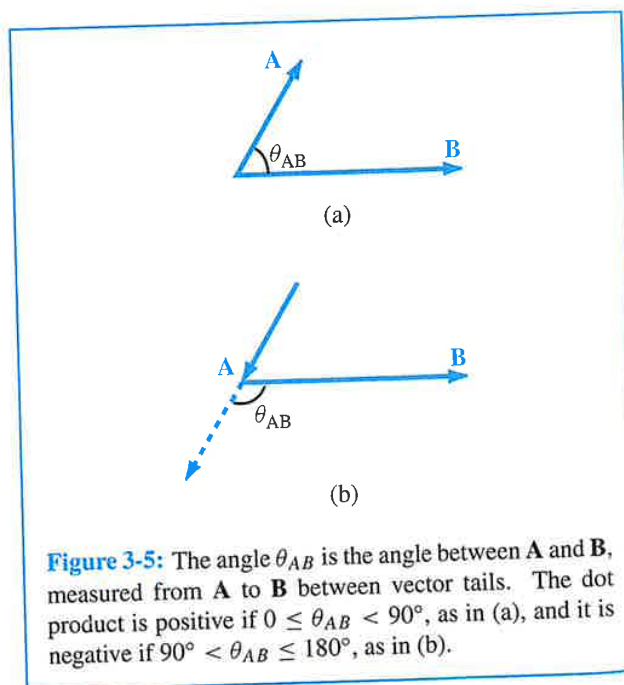


Figure 3-5: The angle θ_{AB} is the angle between $\mathbf{A}$ and $\mathbf{B}$, measured from $\mathbf{A}$ to $\mathbf{B}$ between vector tails. The dot product is positive if $0 \leq \theta_{AB} < 90^\circ$, as in (a), and it is negative if $90^\circ < \theta_{AB} \leq 180^\circ$, as in (b).

is equal to the *projection* of vector **A** along the direction of vector **B**, and similarly $B \cos \theta_{AB}$ is the component of **B** along **A**.

If $\mathbf{A} = (A_x, A_y, A_z)$ and $\mathbf{B} = (B_x, B_y, B_z)$, then

$$\mathbf{A} \cdot \mathbf{B} = (\hat{x}A_x + \hat{y}A_y + \hat{z}A_z) \cdot (\hat{x}B_x + \hat{y}B_y + \hat{z}B_z). \quad (3.15)$$

Since the base vectors $\hat{x}$, $\hat{y}$, and $\hat{z}$ are each orthogonal to the other two, it follows that

$$\hat{x} \cdot \hat{x} = \hat{y} \cdot \hat{y} = \hat{z} \cdot \hat{z} = 1, \quad (3.16a)$$

$$\hat{x} \cdot \hat{y} = \hat{y} \cdot \hat{z} = \hat{z} \cdot \hat{x} = 0. \quad (3.16b)$$

Use of Eqs. (3.16a) and (3.16b) in Eq. (3.15) leads to

$$\mathbf{A} \cdot \mathbf{B} = A_x B_x + A_y B_y + A_z B_z. \quad (3.17)$$

The dot product obeys both the commutative and distributive properties of multiplication; that is,

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A} \quad (\text{commutative property}), \quad (3.18a)$$

$$\mathbf{A} \cdot (\mathbf{B} + \mathbf{C}) = \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \cdot \mathbf{C} \quad (\text{distributive property}). \quad (3.18b)$$

The dot product of a vector with itself gives

$$\mathbf{A} \cdot \mathbf{A} = |\mathbf{A}|^2 = A^2. \quad (3.19)$$

If the vector **A** is defined in a given coordinate system, its magnitude A can be determined from

$$A = |\mathbf{A}| = \sqrt{\mathbf{A} \cdot \mathbf{A}}. \quad (3.20)$$

Also, if vectors **A** and **B** are specified in a given coordinate system, then the smaller angle between them, θ_{AB} , can be determined from

$$\theta_{AB} = \cos^{-1} \left[\frac{\mathbf{A} \cdot \mathbf{B}}{\sqrt{\mathbf{A} \cdot \mathbf{A}} \sqrt{\mathbf{B} \cdot \mathbf{B}}} \right]. \quad (3.21)$$

Vector or Cross Product

The *vector* (or *cross*) *product* of two vectors **A** and **B**, denoted by $\mathbf{A} \times \mathbf{B}$ and pronounced “A cross B,” yields a vector defined as

$$\mathbf{A} \times \mathbf{B} = \hat{n} AB \sin \theta_{AB}, \quad (3.22)$$

where θ_{AB} is the angle between **A** and **B**, measured *from* the tail of **A** *to* the tail of **B**, and $\hat{n}$ is a unit vector normal to the plane containing **A** and **B**. The magnitude of the cross product is equal to the area of the parallelogram defined by the two vectors, as illustrated in Fig. 3-6(a), and its direction is specified by $\hat{n}$ in accordance with the

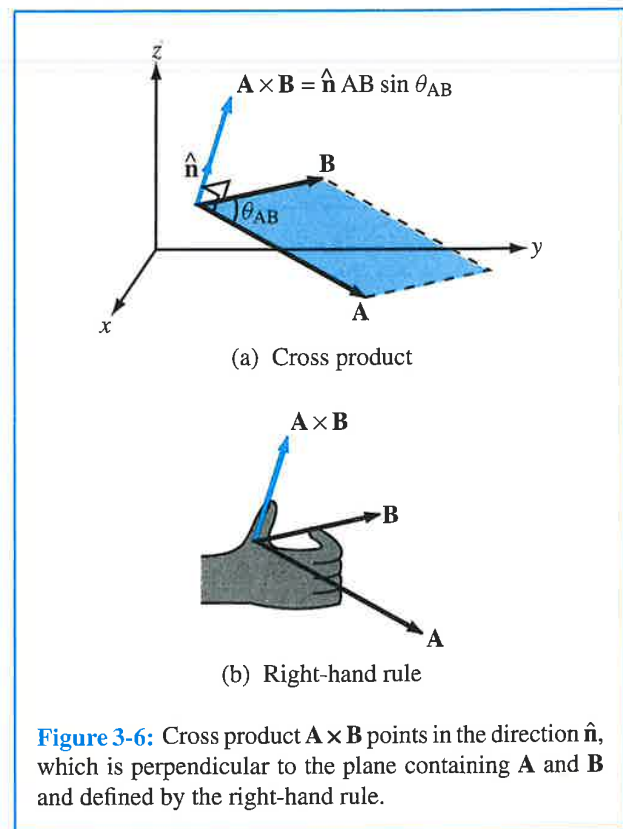


Figure 3-6: Cross product $\mathbf{A} \times \mathbf{B}$ points in the direction $\hat{n}$, which is perpendicular to the plane containing **A** and **B** and defined by the right-hand rule.

following **right-hand rule** [Fig. 3-6(b)]: the direction of $\hat{n}$ points along the right thumb when the fingers rotate from $\mathbf{A}$ to $\mathbf{B}$ through the angle θ_{AB} . We note that, since $\hat{n}$ is perpendicular to the plane containing $\mathbf{A}$ and $\mathbf{B}$, it is also perpendicular to the vectors $\mathbf{A}$ and $\mathbf{B}$.

The cross product is anticommutative, meaning that

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A} \quad (\text{anticommutative}). \quad (3.23)$$

This property can be verified by rotating the fingers of the right hand from $\mathbf{B}$ to $\mathbf{A}$ through the angle θ_{AB} . Other properties of the cross product include

$$\mathbf{A} \times (\mathbf{B} + \mathbf{C}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C} \quad (\text{distributive}), \quad (3.24a)$$

$$\mathbf{A} \times \mathbf{A} = 0. \quad (3.24b)$$

From the definition of the cross product given by Eq. (3.22), it is easy to verify that the base vectors $\hat{x}$, $\hat{y}$, and $\hat{z}$ of the Cartesian coordinate system obey the following right-hand cyclic relations:

$$\hat{x} \times \hat{y} = \hat{z}, \quad \hat{y} \times \hat{z} = \hat{x}, \quad \hat{z} \times \hat{x} = \hat{y}. \quad (3.25)$$

Note the cyclic order ($xyzxyz \dots$). Also,

$$\hat{x} \times \hat{x} = \hat{y} \times \hat{y} = \hat{z} \times \hat{z} = 0. \quad (3.26)$$

If $\mathbf{A} = (A_x, A_y, A_z)$ and $\mathbf{B} = (B_x, B_y, B_z)$, use of Eqs. (3.25) and (3.26) leads to

$$\begin{aligned} \mathbf{A} \times \mathbf{B} &= (\hat{x}A_x + \hat{y}A_y + \hat{z}A_z) \times (\hat{x}B_x + \hat{y}B_y + \hat{z}B_z) \\ &= \hat{x}(A_yB_z - A_zB_y) + \hat{y}(A_zB_x - A_xB_z) \\ &\quad + \hat{z}(A_xB_y - A_yB_x). \end{aligned} \quad (3.27)$$

The cyclical form of the result given by Eq. (3.27) allows us to express the cross product in the form of a determinant:

$$\mathbf{A} \times \mathbf{B} = \begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix}. \quad (3.28)$$

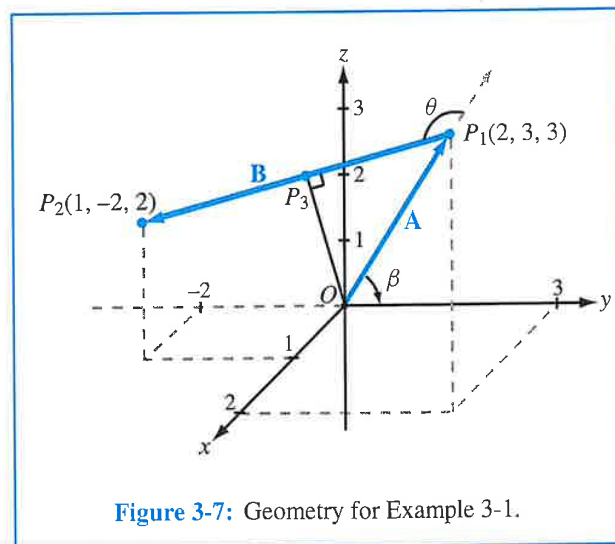


Figure 3-7: Geometry for Example 3-1.

Example 3-1 Vectors and Angles

In Cartesian coordinates, vector $\mathbf{A}$ is directed from the origin to point $P_1(2, 3, 3)$, and vector $\mathbf{B}$ is directed from P_1 to point $P_2(1, -2, 2)$. Find

- vector $\mathbf{A}$, its magnitude A , and unit vector $\hat{a}$,
- the angle that $\mathbf{A}$ makes with the y -axis,
- vector $\mathbf{B}$,
- the angle between $\mathbf{A}$ and $\mathbf{B}$, and
- the perpendicular distance from the origin to vector $\mathbf{B}$.

Solution: (a) Vector $\mathbf{A}$ is given by the distance vector from $O(0, 0, 0)$ to $P_1(2, 3, 3)$ as shown in Fig. 3-7. Thus,

$$\mathbf{A} = \hat{x}2 + \hat{y}3 + \hat{z}3,$$

$$A = |\mathbf{A}| = \sqrt{2^2 + 3^2 + 3^2} = \sqrt{22},$$

$$\hat{a} = \frac{\mathbf{A}}{A} = (\hat{x}2 + \hat{y}3 + \hat{z}3)/\sqrt{22}.$$

- (b) The angle β between $\mathbf{A}$ and the y -axis is obtained from

$$\mathbf{A} \cdot \hat{y} = |\mathbf{A}| \cos \beta,$$

or

$$\beta = \cos^{-1} \left(\frac{\mathbf{A} \cdot \hat{\mathbf{y}}}{A} \right) = \cos^{-1} \left(\frac{3}{\sqrt{22}} \right) = 50.2^\circ.$$

(c)

$$\mathbf{B} = \hat{\mathbf{x}}(1 - 2) + \hat{\mathbf{y}}(-2 - 3) + \hat{\mathbf{z}}(2 - 3) = -\hat{\mathbf{x}} - \hat{\mathbf{y}}5 - \hat{\mathbf{z}}.$$

(d)

$$\theta = \cos^{-1} \left[\frac{\mathbf{A} \cdot \mathbf{B}}{|\mathbf{A}||\mathbf{B}|} \right] = \cos^{-1} \left[\frac{(-2 - 15 - 3)}{\sqrt{22} \sqrt{27}} \right] = 145.1^\circ.$$

(e) The perpendicular distance between the origin and vector $\mathbf{B}$ is the distance $|\overrightarrow{OP_3}|$ shown in Fig. 3-7. From right triangle OP_1P_3 ,

$$\begin{aligned} |\overrightarrow{OP_3}| &= |\mathbf{A}| \sin(180^\circ - \theta) \\ &= \sqrt{22} \sin(180^\circ - 145.1^\circ) = 2.68. \quad \blacksquare \end{aligned}$$

EXERCISE 3.1 Find the distance vector between $P_1(1, 2, 3)$ and $P_2(-1, -2, 3)$ in Cartesian coordinates.

Ans. $\overrightarrow{P_1P_2} = -\hat{\mathbf{x}}2 - \hat{\mathbf{y}}4.$

EXERCISE 3.2 Find the angle θ between vectors $\mathbf{A}$ and $\mathbf{B}$ of Example 3-1 using the cross product between them.

Ans. $\theta = 145.1^\circ.$

EXERCISE 3.3 Find the angle that vector $\mathbf{B}$ of Example 3-1 makes with the z -axis.

Ans. $101.1^\circ.$

3-1.5 Scalar and Vector Triple Products

When three vectors are multiplied, not all combinations of dot and cross products are meaningful. For example, the product

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

does not make sense because $\mathbf{B} \cdot \mathbf{C}$ gives a scalar, and the cross product of the vector $\mathbf{A}$ with a scalar is not defined under the rules of vector algebra. Other than the product of the form $\mathbf{A}(\mathbf{B} \cdot \mathbf{C})$, the only two meaningful products of three vectors are the scalar triple product and the vector triple product.

Scalar Triple Product

The dot product of a vector with the cross product of two other vectors is called a scalar triple product, so named because the result is a scalar. A scalar triple product obeys the following cyclic order:

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{B} \cdot (\mathbf{C} \times \mathbf{A}) = \mathbf{C} \cdot (\mathbf{A} \times \mathbf{B}). \quad (3.29)$$

The equalities hold as long as the cyclic order ($ABCABC \dots$) is preserved. The scalar triple product of vectors $\mathbf{A} = (A_x, A_y, A_z)$, $\mathbf{B} = (B_x, B_y, B_z)$, and $\mathbf{C} = (C_x, C_y, C_z)$ can be written in the form of a 3×3 determinant:

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \begin{vmatrix} A_x & A_y & A_z \\ B_x & B_y & B_z \\ C_x & C_y & C_z \end{vmatrix}. \quad (3.30)$$

The validity of Eqs. (3.29) and (3.30) can be verified by expanding $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ in component form and carrying out the multiplications.

Vector Triple Product

The vector triple product involves the cross product of a vector with the cross product of two others, such as

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}). \quad (3.31)$$

Since each cross product yields a vector, the result of a vector triple product is also a vector. The vector triple product does not, in general, obey the associative law. That is,

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) \neq (\mathbf{A} \times \mathbf{B}) \times \mathbf{C}, \quad (3.32)$$

which means that it is important to specify which cross multiplication is to be performed first. By expanding the vectors $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ in component form, it can be shown that

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B}), \quad (3.33)$$

which sometimes is known as the “bac-cab” rule.

Example 3-2 Vector Triple Product

Given $\mathbf{A} = \hat{\mathbf{x}} - \hat{\mathbf{y}} + \hat{\mathbf{z}}$, $\mathbf{B} = \hat{\mathbf{y}} + \hat{\mathbf{z}}$, and $\mathbf{C} = -\hat{\mathbf{x}} + \hat{\mathbf{z}}$, find $(\mathbf{A} \times \mathbf{B}) \times \mathbf{C}$ and compare it with $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$.

Solution

$$\mathbf{A} \times \mathbf{B} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ 1 & -1 & 2 \\ 0 & 1 & 1 \end{vmatrix} = -\hat{\mathbf{x}}\mathbf{3} - \hat{\mathbf{y}} + \hat{\mathbf{z}}$$

and

$$(\mathbf{A} \times \mathbf{B}) \times \mathbf{C} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ -3 & -1 & 1 \\ -2 & 0 & 3 \end{vmatrix} = -\hat{\mathbf{x}}\mathbf{3} + \hat{\mathbf{y}}\mathbf{7} - \hat{\mathbf{z}}\mathbf{2}.$$

A similar procedure gives $\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \hat{\mathbf{x}}\mathbf{2} + \hat{\mathbf{y}}\mathbf{4} + \hat{\mathbf{z}}$. The fact that the results of two vector triple products are different is a demonstration of the inequality stated in Eq. (3.32). ■

REVIEW QUESTIONS

Q3.1 When are two vectors equal and when are they identical?

Q3.2 When is the position vector of a point identical to the distance vector between two points?

Q3.3 If $\mathbf{A} \cdot \mathbf{B} = 0$, what is θ_{AB} ?

Q3.4 If $\mathbf{A} \times \mathbf{B} = 0$, what is θ_{AB} ?

Q3.5 Is $\mathbf{A}(\mathbf{B} \cdot \mathbf{C})$ a vector triple product?

Q3.6 If $\mathbf{A} \cdot \mathbf{B} = \mathbf{A} \cdot \mathbf{C}$, does it follow that $\mathbf{B} = \mathbf{C}$?

3-2 Orthogonal Coordinate Systems

In electromagnetics, the physical quantities we deal with are, in general, functions of space and time. A three-dimensional coordinate system allows us to uniquely specify the location of a point in space or the direction of a vector quantity. Coordinate systems may be orthogonal or nonorthogonal. An *orthogonal coordinate system* is one whose coordinates are mutually perpendicular, whereas in a nonorthogonal system not all three coordinates are mutually perpendicular. Nonorthogonal systems are very specialized and seldom used in solving practical problems. Many orthogonal coordinate systems have been devised, but the most standard and commonly used are

- the Cartesian (otherwise called the rectangular) coordinate system,
- the cylindrical coordinate system, and
- the spherical coordinate system.

Why do we need more than one coordinate system? Whereas a point in space has the same location and an object has the same shape regardless of which specific coordinate system is used to describe them, the solution of a given practical problem can be greatly facilitated by

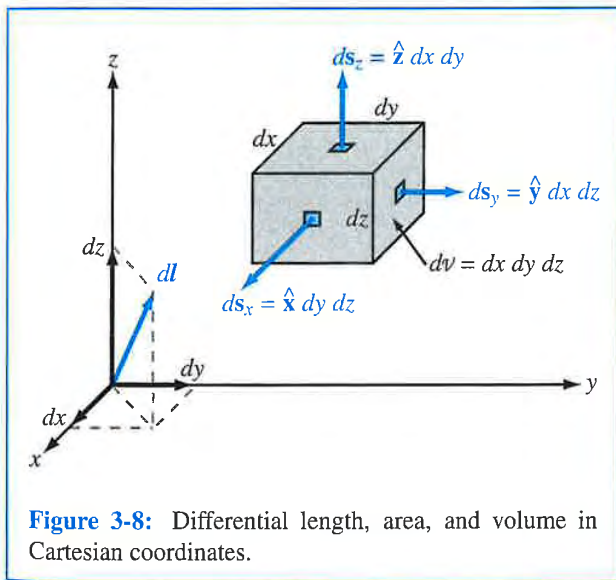
the proper choice of a coordinate system that best fits the geometry of the problem. Hence, in the following subsections we shall examine the properties of each of the aforementioned orthogonal systems, and in Section 3-3 we shall describe how a point or vector may be transformed from one coordinate system to another.

3-2.1 Cartesian Coordinates

The Cartesian coordinate system was introduced in Section 3-1, where we used it to illustrate many of the laws of vector algebra. In lieu of repeating these laws for the Cartesian system, we have summarized them for easy access in Table 3-1. In differential calculus, we often work with differential quantities. Differential length in Cartesian coordinates is a vector (Fig. 3-8) defined as

$$d\mathbf{l} = \hat{\mathbf{x}} dl_x + \hat{\mathbf{y}} dl_y + \hat{\mathbf{z}} dl_z = \hat{\mathbf{x}} dx + \hat{\mathbf{y}} dy + \hat{\mathbf{z}} dz, \quad (3.34)$$

where $dl_x = dx$ is a differential length along $\hat{\mathbf{x}}$, and similar definitions apply to $dl_y = dy$ and $dl_z = dz$.



A differential surface area ds is a vector quantity with a magnitude ds equal to the product of two differential lengths (such as dl_y and dl_z), and its direction is denoted by a unit vector along the third direction (such as $\hat{\mathbf{x}}$). Thus, for a differential area in the y - z plane,

$$ds_x = \hat{\mathbf{x}} dl_y dl_z = \hat{\mathbf{x}} dy dz \quad (y\text{-}z \text{ plane}), \quad (3.35a)$$

with the subscript of ds denoting its direction. Similarly,

$$ds_y = \hat{\mathbf{y}} dx dz \quad (x\text{-}z \text{ plane}), \quad (3.35b)$$

$$ds_z = \hat{\mathbf{z}} dx dy \quad (x\text{-}y \text{ plane}). \quad (3.35c)$$

A differential volume is equal to the product of all three differential lengths:

$$dv = dx dy dz. \quad (3.36)$$

3-2.2 Cylindrical Coordinates

A cylindrical coordinate system is useful for solving problems having cylindrical symmetry, such as calculating the capacitance per unit length of a coaxial transmission line. The location of a point in space is uniquely defined by three variables, r , ϕ , and z , as shown in Fig. 3-9. The coordinate r is the **radial distance** in the x - y plane, ϕ is the **azimuth angle** measured from the positive x -axis, and z is as previously defined in the Cartesian coordinate system. Their ranges are $0 \leq r < \infty$, $0 \leq \phi < 2\pi$, and $-\infty < z < \infty$. Point $P(r_1, \phi_1, z_1)$ in Fig. 3-9 is located at the intersection of three surfaces. These are the cylindrical surface defined by $r = r_1$, the vertical half-plane defined by $\phi = \phi_1$ (which extends outwardly from the z -axis), and the horizontal plane defined by $z = z_1$. The mutually perpendicular base vectors are $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\phi}}$, and $\hat{\mathbf{z}}$, with $\hat{\mathbf{r}}$ pointing away from the origin along r , $\hat{\boldsymbol{\phi}}$ pointing in a direction tangential to the cylindrical surface, and $\hat{\mathbf{z}}$ pointing along the vertical. Unlike the Cartesian system, in which the base vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are independent of the location of P , in the cylindrical system both $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\phi}}$ are functions of ϕ .

Table 3-1: Summary of vector relations.

	Cartesian Coordinates	Cylindrical Coordinates	Spherical Coordinates
Coordinate variables	x, y, z	r, ϕ, z	R, θ, ϕ
Vector representation, $\mathbf{A} =$	$\hat{\mathbf{x}}A_x + \hat{\mathbf{y}}A_y + \hat{\mathbf{z}}A_z$	$\hat{\mathbf{r}}A_r + \hat{\boldsymbol{\phi}}A_\phi + \hat{\mathbf{z}}A_z$	$\hat{\mathbf{R}}A_R + \hat{\boldsymbol{\theta}}A_\theta + \hat{\boldsymbol{\phi}}A_\phi$
Magnitude of $\mathbf{A}$, $ \mathbf{A} =$	$\sqrt{A_x^2 + A_y^2 + A_z^2}$	$\sqrt{A_r^2 + A_\phi^2 + A_z^2}$	$\sqrt{A_R^2 + A_\theta^2 + A_\phi^2}$
Position vector $\overrightarrow{OP_1} =$	$\hat{\mathbf{x}}x_1 + \hat{\mathbf{y}}y_1 + \hat{\mathbf{z}}z_1$, for $P(x_1, y_1, z_1)$	$\hat{\mathbf{r}}r_1 + \hat{\mathbf{z}}z_1$, for $P(r_1, \phi_1, z_1)$	$\hat{\mathbf{R}}R_1$, for $P(R_1, \theta_1, \phi_1)$
Base vectors properties	$\hat{\mathbf{x}} \cdot \hat{\mathbf{x}} = \hat{\mathbf{y}} \cdot \hat{\mathbf{y}} = \hat{\mathbf{z}} \cdot \hat{\mathbf{z}} = 1$ $\hat{\mathbf{x}} \cdot \hat{\mathbf{y}} = \hat{\mathbf{y}} \cdot \hat{\mathbf{z}} = \hat{\mathbf{z}} \cdot \hat{\mathbf{x}} = 0$ $\hat{\mathbf{x}} \times \hat{\mathbf{y}} = \hat{\mathbf{z}}$ $\hat{\mathbf{y}} \times \hat{\mathbf{z}} = \hat{\mathbf{x}}$ $\hat{\mathbf{z}} \times \hat{\mathbf{x}} = \hat{\mathbf{y}}$	$\hat{\mathbf{r}} \cdot \hat{\mathbf{r}} = \hat{\boldsymbol{\phi}} \cdot \hat{\boldsymbol{\phi}} = \hat{\mathbf{z}} \cdot \hat{\mathbf{z}} = 1$ $\hat{\mathbf{r}} \cdot \hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\phi}} \cdot \hat{\mathbf{z}} = \hat{\mathbf{z}} \cdot \hat{\mathbf{r}} = 0$ $\hat{\mathbf{r}} \times \hat{\boldsymbol{\phi}} = \hat{\mathbf{z}}$ $\hat{\boldsymbol{\phi}} \times \hat{\mathbf{z}} = \hat{\mathbf{r}}$ $\hat{\mathbf{z}} \times \hat{\mathbf{r}} = \hat{\boldsymbol{\phi}}$	$\hat{\mathbf{R}} \cdot \hat{\mathbf{R}} = \hat{\boldsymbol{\theta}} \cdot \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\phi}} \cdot \hat{\boldsymbol{\phi}} = 1$ $\hat{\mathbf{R}} \cdot \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} \cdot \hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\phi}} \cdot \hat{\mathbf{R}} = 0$ $\hat{\mathbf{R}} \times \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\phi}}$ $\hat{\boldsymbol{\theta}} \times \hat{\boldsymbol{\phi}} = \hat{\mathbf{R}}$ $\hat{\boldsymbol{\phi}} \times \hat{\mathbf{R}} = \hat{\boldsymbol{\theta}}$
Dot product, $\mathbf{A} \cdot \mathbf{B} =$	$A_x B_x + A_y B_y + A_z B_z$	$A_r B_r + A_\phi B_\phi + A_z B_z$	$A_R B_R + A_\theta B_\theta + A_\phi B_\phi$
Cross product, $\mathbf{A} \times \mathbf{B} =$	$\begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix}$	$\begin{vmatrix} \hat{\mathbf{r}} & \hat{\boldsymbol{\phi}} & \hat{\mathbf{z}} \\ A_r & A_\phi & A_z \\ B_r & B_\phi & B_z \end{vmatrix}$	$\begin{vmatrix} \hat{\mathbf{R}} & \hat{\boldsymbol{\theta}} & \hat{\boldsymbol{\phi}} \\ A_R & A_\theta & A_\phi \\ B_R & B_\theta & B_\phi \end{vmatrix}$
Differential length, $d\mathbf{l} =$	$\hat{\mathbf{x}} dx + \hat{\mathbf{y}} dy + \hat{\mathbf{z}} dz$	$\hat{\mathbf{r}} dr + \hat{\boldsymbol{\phi}} r d\phi + \hat{\mathbf{z}} dz$	$\hat{\mathbf{R}} dR + \hat{\boldsymbol{\theta}} R d\theta + \hat{\boldsymbol{\phi}} R \sin \theta d\phi$
Differential surface areas	$ds_x = \hat{\mathbf{x}} dy dz$ $ds_y = \hat{\mathbf{y}} dx dz$ $ds_z = \hat{\mathbf{z}} dx dy$	$ds_r = \hat{\mathbf{r}} r d\phi dz$ $ds_\phi = \hat{\boldsymbol{\phi}} dr dz$ $ds_z = \hat{\mathbf{z}} r dr d\phi$	$ds_R = \hat{\mathbf{R}} R^2 \sin \theta d\theta d\phi$ $ds_\theta = \hat{\boldsymbol{\theta}} R \sin \theta dR d\phi$ $ds_\phi = \hat{\boldsymbol{\phi}} R dR d\theta$
Differential volume, $d\nu =$	$dx dy dz$	$r dr d\phi dz$	$R^2 \sin \theta dR d\theta d\phi$

The base unit vectors obey the following right-hand cyclic relations:

$$\hat{\mathbf{r}} \times \hat{\boldsymbol{\phi}} = \hat{\mathbf{z}}, \quad \hat{\boldsymbol{\phi}} \times \hat{\mathbf{z}} = \hat{\mathbf{r}}, \quad \hat{\mathbf{z}} \times \hat{\mathbf{r}} = \hat{\boldsymbol{\phi}}, \quad (3.37)$$

and like all unit vectors, $\hat{\mathbf{r}} \cdot \hat{\mathbf{r}} = \hat{\boldsymbol{\phi}} \cdot \hat{\boldsymbol{\phi}} = \hat{\mathbf{z}} \cdot \hat{\mathbf{z}} = 1$, and $\hat{\mathbf{r}} \times \hat{\mathbf{r}} = \hat{\boldsymbol{\phi}} \times \hat{\boldsymbol{\phi}} = \hat{\mathbf{z}} \times \hat{\mathbf{z}} = 0$.

In cylindrical coordinates, a vector is expressed as

$$\mathbf{A} = \hat{\mathbf{a}}|\mathbf{A}| = \hat{\mathbf{r}}A_r + \hat{\boldsymbol{\phi}}A_\phi + \hat{\mathbf{z}}A_z, \quad (3.38)$$

where A_r , A_ϕ , and A_z are the components of $\mathbf{A}$ along the $\hat{\mathbf{r}}$ -, $\hat{\boldsymbol{\phi}}$ -, and $\hat{\mathbf{z}}$ -directions. The magnitude of $\mathbf{A}$ is obtained by applying Eq. (3.20), which gives

$$|\mathbf{A}| = \sqrt{\mathbf{A} \cdot \mathbf{A}} = \sqrt{A_r^2 + A_\phi^2 + A_z^2}. \quad (3.39)$$

The position vector $\overrightarrow{OP}$ shown in Fig. 3-9 has components along r and z only. Thus,

$$\mathbf{R}_1 = \overrightarrow{OP} = \hat{\mathbf{r}}r_1 + \hat{\mathbf{z}}z_1. \quad (3.40)$$

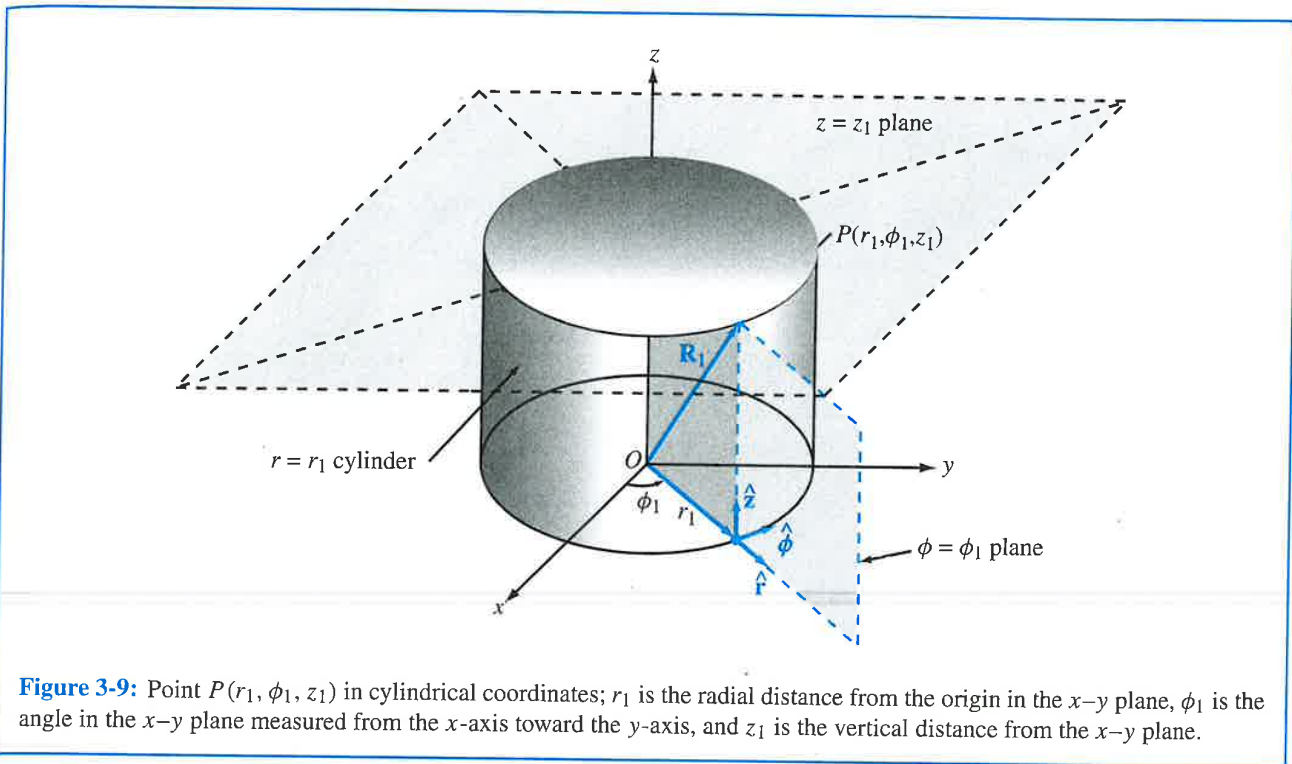


Figure 3-9: Point $P(r_1, \phi_1, z_1)$ in cylindrical coordinates; r_1 is the radial distance from the origin in the x - y plane, ϕ_1 is the angle in the x - y plane measured from the x -axis toward the y -axis, and z_1 is the vertical distance from the x - y plane.

The dependence of $\mathbf{R}_1$ on ϕ_1 is implicit through the dependence of $\hat{\mathbf{r}}$ on ϕ_1 . Hence, when using Eq. (3.40) to denote the position vector of point $P(r_1, \phi_1, z_1)$, it is necessary to specify that $\hat{\mathbf{r}}$ is at ϕ_1 .

Figure 3-10 shows a differential volume element in cylindrical coordinates. The differential lengths along $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\phi}}$, and $\hat{\mathbf{z}}$ are

$$dl_r = dr, \quad dl_\phi = r d\phi, \quad dl_z = dz. \quad (3.41)$$

Note that the differential length along $\hat{\boldsymbol{\phi}}$ is $r d\phi$, not just $d\phi$. The differential length $d\mathbf{l}$ in cylindrical coordinates is given by

$$d\mathbf{l} = \hat{\mathbf{r}} dl_r + \hat{\boldsymbol{\phi}} dl_\phi + \hat{\mathbf{z}} dl_z = \hat{\mathbf{r}} dr + \hat{\boldsymbol{\phi}} r d\phi + \hat{\mathbf{z}} dz. \quad (3.42)$$

As was stated previously for the Cartesian coordinate system, the product of any pair of differential lengths is

equal to the magnitude of a vector differential surface area with a surface normal pointing along the direction of the third coordinate. Thus,

$$ds_r = \hat{\mathbf{r}} dl_\phi dl_z = \hat{\mathbf{r}} r d\phi dz \quad (\phi\text{-}z \text{ cylindrical surface}), \quad (3.43a)$$

$$ds_\phi = \hat{\boldsymbol{\phi}} dl_r dl_z = \hat{\boldsymbol{\phi}} dr dz \quad (r\text{-}z \text{ plane}), \quad (3.43b)$$

$$ds_z = \hat{\mathbf{z}} dl_r dl_\phi = \hat{\mathbf{z}} r dr d\phi \quad (r\text{-}\phi \text{ plane}). \quad (3.43c)$$

The differential volume is the product of the three differential lengths,

$$dv = dl_r dl_\phi dl_z = r dr d\phi dz. \quad (3.44)$$

The preceding properties of the cylindrical coordinate system are summarized in Table 3-1.

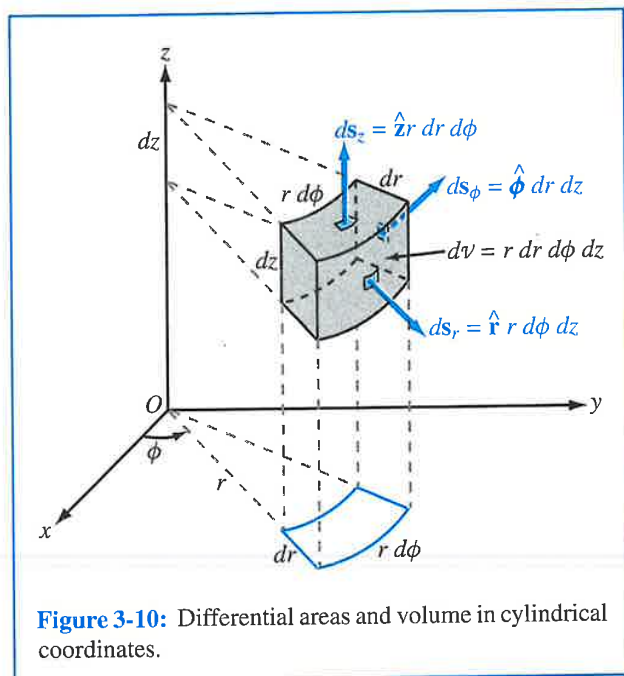


Figure 3-10: Differential areas and volume in cylindrical coordinates.

Example 3-3 Distance Vector in Cylindrical Coordinates

Find an expression for the unit vector of vector **A** shown in Fig. 3-11 in cylindrical coordinates.

Solution: In triangle OP_1P_2 ,

$$\overrightarrow{OP_2} = \overrightarrow{OP_1} + \mathbf{A}.$$

Hence,

$$\mathbf{A} = \overrightarrow{OP_2} - \overrightarrow{OP_1} = \hat{\mathbf{r}}r_0 - \hat{\mathbf{z}}h,$$

and

$$\hat{\mathbf{a}} = \frac{\mathbf{A}}{|\mathbf{A}|} = \frac{\hat{\mathbf{r}}r_0 - \hat{\mathbf{z}}h}{\sqrt{r_0^2 + h^2}}.$$

We note that the expression for **A** is independent of ϕ_0 . That is, all vectors from point P_1 to any point on the

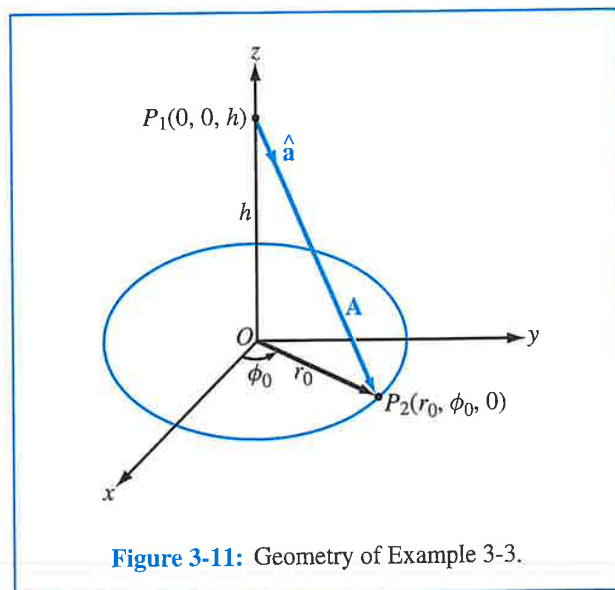


Figure 3-11: Geometry of Example 3-3.

circle defined by $r = r_0$ in the x - y plane are equal in the cylindrical coordinate system. The ambiguity can be eliminated by specifying that **A** passes through a point whose $\phi = \phi_0$. ■

Example 3-4 Cylindrical Area

Find the area of a cylindrical surface described by $r = 5$, $30^\circ \leq \phi \leq 60^\circ$, and $0 \leq z \leq 3$ (Fig. 3-12).

Solution: The prescribed surface is shown in Fig. 3-12. Use of Eq. (3.43a) for a surface element with constant r gives

$$S = r \int_{\phi=30^\circ}^{60^\circ} d\phi \int_{z=0}^3 dz = 5\phi \Big|_{\pi/6}^{\pi/3} z \Big|_0^3 = \frac{5\pi}{2}.$$

Note that ϕ had to be converted to radians before evaluating the integration limits. ■

EXERCISE 3.4 A circular cylinder of radius $r = 5$ cm is concentric with the z -axis and extends between

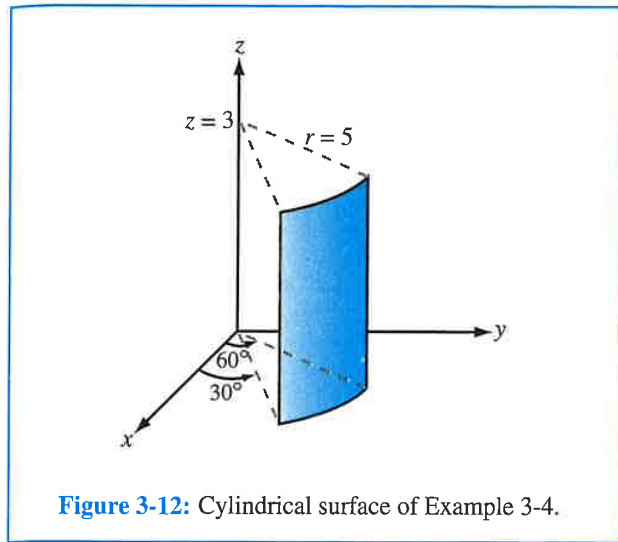


Figure 3-12: Cylindrical surface of Example 3-4.

$z = -3$ cm and $z = 3$ cm. Use Eq. (3.44) to find the cylinder's volume.

Ans. 471.2 cm³.

3-2.3 Spherical Coordinates

In the spherical coordinate system, the location of a point in space is uniquely specified by the variables R , θ , and ϕ , as shown in Fig. 3-13. The coordinate R , which sometimes is called the *range* coordinate, describes a sphere of radius R centered at the origin. The *zenith* angle θ is measured from the positive z -axis and it describes a conical surface with its apex at the origin, and the azimuth angle ϕ is the same as in the cylindrical coordinate system. The ranges of R , θ , and ϕ are $0 \leq R < \infty$, $0 \leq \theta \leq \pi$, and $0 \leq \phi < 2\pi$. The base vectors $\hat{\mathbf{R}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\phi}}$ obey the following right-hand cyclic relations:

$$\hat{\mathbf{R}} \times \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\phi}}, \quad \hat{\boldsymbol{\theta}} \times \hat{\boldsymbol{\phi}} = \hat{\mathbf{R}}, \quad \hat{\boldsymbol{\phi}} \times \hat{\mathbf{R}} = \hat{\boldsymbol{\theta}}. \quad (3.45)$$

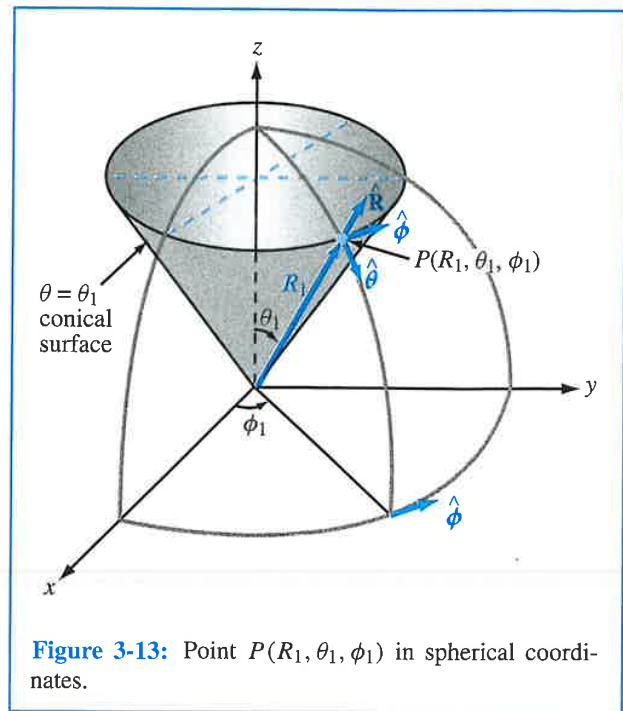


Figure 3-13: Point $P(R_1, \theta_1, \phi_1)$ in spherical coordinates.

A vector with components A_R , A_θ , and A_ϕ is written as

$$\mathbf{A} = \hat{\mathbf{a}}|\mathbf{A}| = \hat{\mathbf{R}}A_R + \hat{\boldsymbol{\theta}}A_\theta + \hat{\boldsymbol{\phi}}A_\phi, \quad (3.46)$$

and its magnitude is given by

$$|\mathbf{A}| = \sqrt{\mathbf{A} \cdot \mathbf{A}} = \sqrt{A_R^2 + A_\theta^2 + A_\phi^2}. \quad (3.47)$$

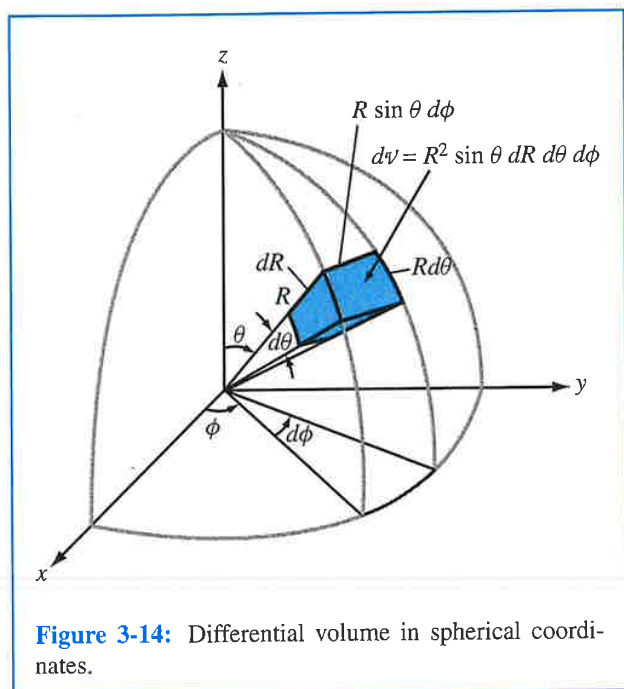
The position vector of point $P(R_1, \theta_1, \phi_1)$ is simply

$$\mathbf{R}_1 = \overrightarrow{OP} = \hat{\mathbf{R}}R_1, \quad (3.48)$$

while keeping in mind that $\hat{\mathbf{R}}$ is implicitly dependent on θ_1 and ϕ_1 .

As shown in Fig. 3-14, the differential lengths along $\hat{\mathbf{R}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\phi}}$ are

$$dl_R = dR, \quad dl_\theta = R d\theta, \quad dl_\phi = R \sin \theta d\phi. \quad (3.49)$$



Hence, the expressions for the vector differential length $d\mathbf{l}$, the vector differential surface $d\mathbf{s}$, and the differential volume dV are

$$\begin{aligned} d\mathbf{l} &= \hat{\mathbf{R}} dl_R + \hat{\boldsymbol{\theta}} dl_\theta + \hat{\boldsymbol{\phi}} dl_\phi \\ &= \hat{\mathbf{R}} dR + \hat{\boldsymbol{\theta}} R d\theta + \hat{\boldsymbol{\phi}} R \sin \theta d\phi, \end{aligned} \quad (3.50a)$$

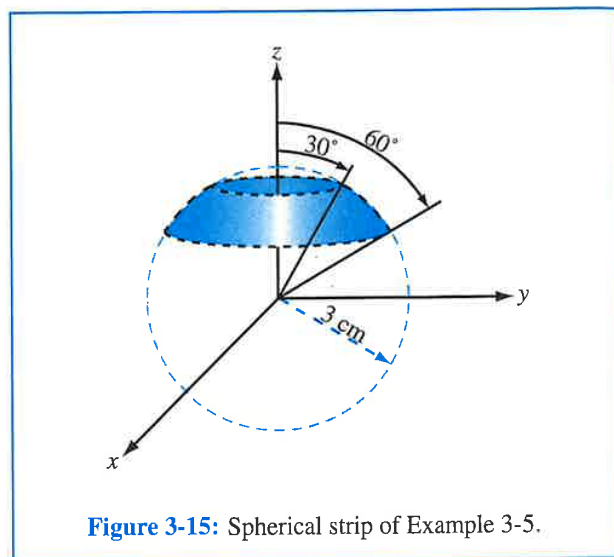
$$\begin{aligned} d\mathbf{s}_R &= \hat{\mathbf{R}} dl_\theta dl_\phi = \hat{\mathbf{R}} R^2 \sin \theta d\theta d\phi \\ &\quad (\theta\text{-}\phi \text{ spherical surface}), \end{aligned} \quad (3.50b)$$

$$\begin{aligned} d\mathbf{s}_\theta &= \hat{\boldsymbol{\theta}} dl_R dl_\phi = \hat{\boldsymbol{\theta}} R \sin \theta dR d\phi \\ &\quad (R\text{-}\phi \text{ conical surface}), \end{aligned} \quad (3.50c)$$

$$d\mathbf{s}_\phi = \hat{\boldsymbol{\phi}} dl_R dl_\theta = \hat{\boldsymbol{\phi}} R dR d\theta \quad (R\text{-}\theta \text{ plane}), \quad (3.50d)$$

$$dV = dl_R dl_\theta dl_\phi = R^2 \sin \theta dR d\theta d\phi. \quad (3.50e)$$

These relations are summarized in Table 3-1.



Example 3-5 Surface Area in Spherical Coordinates

The spherical strip shown in Fig. 3-15 is a section of a sphere of radius 3 cm. Find the area of the strip.

Solution: Use of Eq. (3.50b) for the area of an elemental spherical area with constant radius R gives

$$\begin{aligned} S &= R^2 \int_{\theta=30^\circ}^{60^\circ} \sin \theta d\theta \int_{\phi=0}^{2\pi} d\phi \\ &= 9(-\cos \theta) \Big|_{30^\circ}^{60^\circ} \phi \Big|_0^{2\pi} \quad (\text{cm}^2) \\ &= 18\pi(\cos 30^\circ - \cos 60^\circ) = 20.7 \text{ cm}^2. \quad \blacksquare \end{aligned}$$

Example 3-6 Charge in a Sphere

A sphere of radius 2 cm contains a volume charge density ρ_v given by

$$\rho_v = 4 \cos^2 \theta \quad (\text{C/m}^3).$$

Find the total charge Q contained in the sphere.

Solution

$$\begin{aligned}
Q &= \int_V \rho_v dV \\
&= \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{R=0}^{2 \times 10^{-2}} (4 \cos^2 \theta) R^2 \sin \theta dR d\theta d\phi \\
&= 4 \int_0^{2\pi} \int_0^{\pi} \left(\frac{R^3}{3} \right) \Big|_0^{2 \times 10^{-2}} \sin \theta \cos^2 \theta d\theta d\phi \\
&= \frac{32}{3} \times 10^{-6} \int_0^{2\pi} \left(-\frac{\cos^3 \theta}{3} \right) \Big|_0^{\pi} d\phi \\
&= \frac{64}{9} \times 10^{-6} \int_0^{2\pi} d\phi \\
&= \frac{128\pi}{9} \times 10^{-6} = 44.68 \quad (\mu\text{C}).
\end{aligned}$$

Note that the limits on R were converted to meters prior to evaluating the integral on R . ■

3-3 Transformations between Coordinate Systems

The position of a given point in space is invariant with respect to the choice of coordinate system. That is, its location is the same irrespective of which specific coordinate system is used to represent it. The same is true for vectors. In this section, we shall establish the relations between the variables (x, y, z) of the Cartesian system, (r, ϕ, z) of the cylindrical system, and (R, θ, ϕ) of the spherical system. These relations will then be used to transform vectors expressed in any one of the three systems into vectors expressed in any of the other two.

3-3.1 Cartesian to Cylindrical Transformations

Point P in Fig. 3-16 has Cartesian coordinates (x, y, z) and cylindrical coordinates (r, ϕ, z) . Both systems share the coordinate z , and the relations between the other two

pairs of coordinates can be obtained from the geometry in Fig. 3-16. They are

$$r = \sqrt{x^2 + y^2}, \quad \phi = \tan^{-1} \left(\frac{y}{x} \right), \quad (3.51)$$

and the inverse relations are

$$x = r \cos \phi, \quad y = r \sin \phi. \quad (3.52)$$

Next, with the help of Fig. 3-17, which shows the

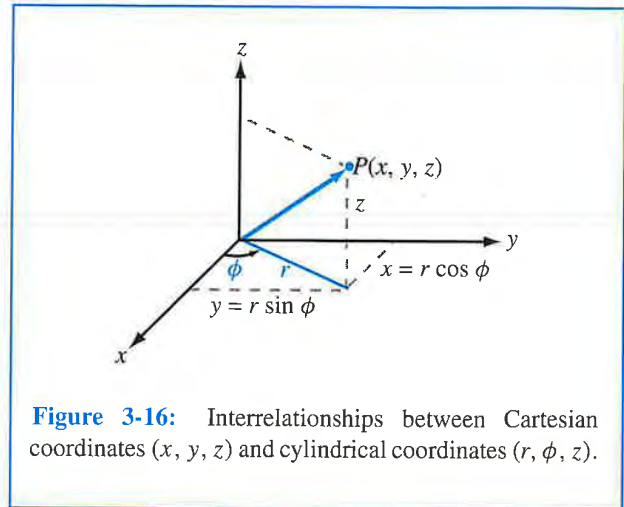


Figure 3-16: Interrelationships between Cartesian coordinates (x, y, z) and cylindrical coordinates (r, ϕ, z) .

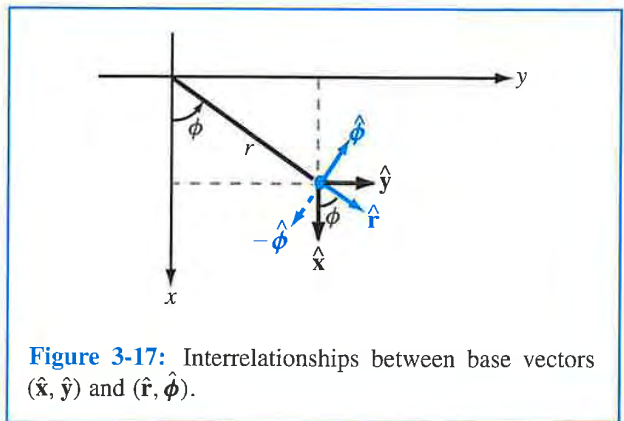


Figure 3-17: Interrelationships between base vectors $(\hat{x}, \hat{y})$ and $(\hat{r}, \hat{\phi})$.

directions of the unit vectors $\hat{x}$, $\hat{y}$, $\hat{r}$, and $\hat{\phi}$ in the x - y plane, we obtain the following relations:

$$\hat{r} \cdot \hat{x} = \cos \phi, \quad \hat{r} \cdot \hat{y} = \sin \phi, \quad (3.53a)$$

$$\hat{\phi} \cdot \hat{x} = -\sin \phi, \quad \hat{\phi} \cdot \hat{y} = \cos \phi, \quad (3.53b)$$

To express $\hat{r}$ in terms of $\hat{x}$ and $\hat{y}$, let us write $\hat{r}$ as

$$\hat{r} = \hat{x}a + \hat{y}b, \quad (3.54)$$

where a and b are unknown transformation coefficients. The dot product $\hat{r} \cdot \hat{x}$ gives

$$\hat{r} \cdot \hat{x} = \hat{x} \cdot \hat{x}a + \hat{y} \cdot \hat{x}b = a. \quad (3.55)$$

Comparison of Eq. (3.55) with Eq. (3.53a) leads to the conclusion that $a = \cos \phi$. Similarly, application of the dot product $\hat{r} \cdot \hat{y}$ to Eq. (3.54) gives $b = \sin \phi$. Hence,

$$\hat{r} = \hat{x} \cos \phi + \hat{y} \sin \phi. \quad (3.56a)$$

Repetition of the procedure for $\hat{\phi}$ leads to

$$\hat{\phi} = -\hat{x} \sin \phi + \hat{y} \cos \phi. \quad (3.56b)$$

The third base vector $\hat{z}$ is the same in both coordinate systems. By solving Eqs. (3.56a) and (3.56b) simultaneously for $\hat{x}$ and $\hat{y}$, we obtain the following inverse relations:

$$\hat{x} = \hat{r} \cos \phi - \hat{\phi} \sin \phi, \quad (3.57a)$$

$$\hat{y} = \hat{r} \sin \phi + \hat{\phi} \cos \phi. \quad (3.57b)$$

The relations given by Eqs. (3.56a) to (3.57b) are not only useful for transforming the base vectors ($\hat{x}$, $\hat{y}$) into ($\hat{r}$, $\hat{\phi}$), and vice versa, they can also be used to transform the components of a vector expressed in either coordinate system into its corresponding components expressed in the other system. For example, a vector $\mathbf{A} = \hat{x}A_x + \hat{y}A_y + \hat{z}A_z$ in Cartesian coordinates can be

transformed into $\mathbf{A} = \hat{r}A_r + \hat{\phi}A_\phi + \hat{z}A_z$ in cylindrical coordinates by applying Eqs. (3.56a) and (3.56b). That is,

$$A_r = A_x \cos \phi + A_y \sin \phi, \quad (3.58a)$$

$$A_\phi = -A_x \sin \phi + A_y \cos \phi, \quad (3.58b)$$

and, conversely,

$$A_x = A_r \cos \phi - A_\phi \sin \phi, \quad (3.59a)$$

$$A_y = A_r \sin \phi + A_\phi \cos \phi. \quad (3.59b)$$

The transformation relations given in this and the following two subsections are summarized in Table 3-2.

Example 3-7 Cartesian to Cylindrical Transformations

Given point $P_1(3, -4, 3)$ and vector $\mathbf{A} = \hat{x}2 - \hat{y}3 + \hat{z}4$, defined in Cartesian coordinates, express P_1 and $\mathbf{A}$ in cylindrical coordinates and evaluate $\mathbf{A}$ at P_1 .

Solution: For point P_1 , $x = 3$, $y = -4$, and $z = 3$. Using Eq. (3.51), we have

$$r = \sqrt{x^2 + y^2} = 5, \quad \phi = \tan^{-1} \frac{y}{x} = -53.1^\circ = 306.9^\circ,$$

and z remains unchanged. Hence, $P_1 = P_1(5, 306.9^\circ, 3)$ in cylindrical coordinates.

For vector $\mathbf{A} = \hat{r}A_r + \hat{\phi}A_\phi + \hat{z}A_z$ in cylindrical coordinates, its components can be determined by applying Eqs. (3.58a) and (3.58b):

$$A_r = A_x \cos \phi + A_y \sin \phi = 2 \cos \phi - 3 \sin \phi,$$

$$A_\phi = -A_x \sin \phi + A_y \cos \phi = -2 \sin \phi - 3 \cos \phi,$$

$$A_z = 4.$$

Hence,

$$\mathbf{A} = \hat{r}(2 \cos \phi - 3 \sin \phi) - \hat{\phi}(2 \sin \phi + 3 \cos \phi) + \hat{z}4.$$

At point P , $\phi = 306.9^\circ$, which gives

$$\mathbf{A} = \hat{r}3.60 - \hat{\phi}0.20 + \hat{z}4. \quad \blacksquare$$

Table 3-2: Coordinate transformation relations.

Transformation	Coordinate Variables	Unit Vectors	Vector Components
Cartesian to cylindrical	$r = \sqrt{x^2 + y^2}$ $\phi = \tan^{-1}(y/x)$ $z = z$	$\hat{\mathbf{r}} = \hat{\mathbf{x}} \cos \phi + \hat{\mathbf{y}} \sin \phi$ $\hat{\boldsymbol{\phi}} = -\hat{\mathbf{x}} \sin \phi + \hat{\mathbf{y}} \cos \phi$ $\hat{\mathbf{z}} = \hat{\mathbf{z}}$	$A_r = A_x \cos \phi + A_y \sin \phi$ $A_\phi = -A_x \sin \phi + A_y \cos \phi$ $A_z = A_z$
Cylindrical to Cartesian	$x = r \cos \phi$ $y = r \sin \phi$ $z = z$	$\hat{\mathbf{x}} = \hat{\mathbf{r}} \cos \phi - \hat{\boldsymbol{\phi}} \sin \phi$ $\hat{\mathbf{y}} = \hat{\mathbf{r}} \sin \phi + \hat{\boldsymbol{\phi}} \cos \phi$ $\hat{\mathbf{z}} = \hat{\mathbf{z}}$	$A_x = A_r \cos \phi - A_\phi \sin \phi$ $A_y = A_r \sin \phi + A_\phi \cos \phi$ $A_z = A_z$
Cartesian to spherical	$R = \sqrt{x^2 + y^2 + z^2}$ $\theta = \tan^{-1}[\sqrt{x^2 + y^2}/z]$ $\phi = \tan^{-1}(y/x)$	$\hat{\mathbf{R}} = \hat{\mathbf{x}} \sin \theta \cos \phi + \hat{\mathbf{y}} \sin \theta \sin \phi + \hat{\mathbf{z}} \cos \theta$ $\hat{\boldsymbol{\theta}} = \hat{\mathbf{x}} \cos \theta \cos \phi + \hat{\mathbf{y}} \cos \theta \sin \phi - \hat{\mathbf{z}} \sin \theta$ $\hat{\boldsymbol{\phi}} = -\hat{\mathbf{x}} \sin \phi + \hat{\mathbf{y}} \cos \phi$	$A_R = A_x \sin \theta \cos \phi + A_y \sin \theta \sin \phi + A_z \cos \theta$ $A_\theta = A_x \cos \theta \cos \phi + A_y \cos \theta \sin \phi - A_z \sin \theta$ $A_\phi = -A_x \sin \phi + A_y \cos \phi$
Spherical to Cartesian	$x = R \sin \theta \cos \phi$ $y = R \sin \theta \sin \phi$ $z = R \cos \theta$	$\hat{\mathbf{x}} = \hat{\mathbf{R}} \sin \theta \cos \phi + \hat{\boldsymbol{\theta}} \cos \theta \cos \phi - \hat{\boldsymbol{\phi}} \sin \phi$ $\hat{\mathbf{y}} = \hat{\mathbf{R}} \sin \theta \sin \phi + \hat{\boldsymbol{\theta}} \cos \theta \sin \phi + \hat{\boldsymbol{\phi}} \cos \phi$ $\hat{\mathbf{z}} = \hat{\mathbf{R}} \cos \theta - \hat{\boldsymbol{\theta}} \sin \theta$	$A_x = A_R \sin \theta \cos \phi + A_\theta \cos \theta \cos \phi - A_\phi \sin \phi$ $A_y = A_R \sin \theta \sin \phi + A_\theta \cos \theta \sin \phi + A_\phi \cos \phi$ $A_z = A_R \cos \theta - A_\theta \sin \theta$
Cylindrical to spherical	$R = \sqrt{r^2 + z^2}$ $\theta = \tan^{-1}(r/z)$ $\phi = \phi$	$\hat{\mathbf{R}} = \hat{\mathbf{r}} \sin \theta + \hat{\mathbf{z}} \cos \theta$ $\hat{\boldsymbol{\theta}} = \hat{\mathbf{r}} \cos \theta - \hat{\mathbf{z}} \sin \theta$ $\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\phi}}$	$A_R = A_r \sin \theta + A_z \cos \theta$ $A_\theta = A_r \cos \theta - A_z \sin \theta$ $A_\phi = A_\phi$
Spherical to cylindrical	$r = R \sin \theta$ $\phi = \phi$ $z = R \cos \theta$	$\hat{\mathbf{r}} = \hat{\mathbf{R}} \sin \theta + \hat{\mathbf{z}} \cos \theta$ $\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\phi}}$ $\hat{\mathbf{z}} = \hat{\mathbf{R}} \cos \theta - \hat{\boldsymbol{\theta}} \sin \theta$	$A_r = A_R \sin \theta + A_\theta \cos \theta$ $A_\phi = A_\phi$ $A_z = A_R \cos \theta - A_\theta \sin \theta$

3-3.2 Cartesian to Spherical Transformations

From Fig. 3-18, we obtain the following relations between the Cartesian coordinates (x, y, z) and the spherical coordinates (R, θ, ϕ) :

$$R = \sqrt{x^2 + y^2 + z^2}, \quad (3.60a)$$

$$\theta = \tan^{-1} \left[\frac{\sqrt{x^2 + y^2}}{z} \right], \quad (3.60b)$$

$$\phi = \tan^{-1} \left(\frac{y}{x} \right), \quad (3.60c)$$

and the converse relations are

$$x = R \sin \theta \cos \phi, \quad (3.61a)$$

$$y = R \sin \theta \sin \phi, \quad (3.61b)$$

$$z = R \cos \theta. \quad (3.61c)$$

The unit vector $\hat{\mathbf{R}}$ lies in the $\hat{\mathbf{r}}\text{-}\hat{\mathbf{z}}$ plane. Hence, it can be expressed as a linear combination of $\hat{\mathbf{r}}$ and $\hat{\mathbf{z}}$ as follows:

$$\hat{\mathbf{R}} = \hat{\mathbf{r}}a + \hat{\mathbf{z}}b, \quad (3.62)$$

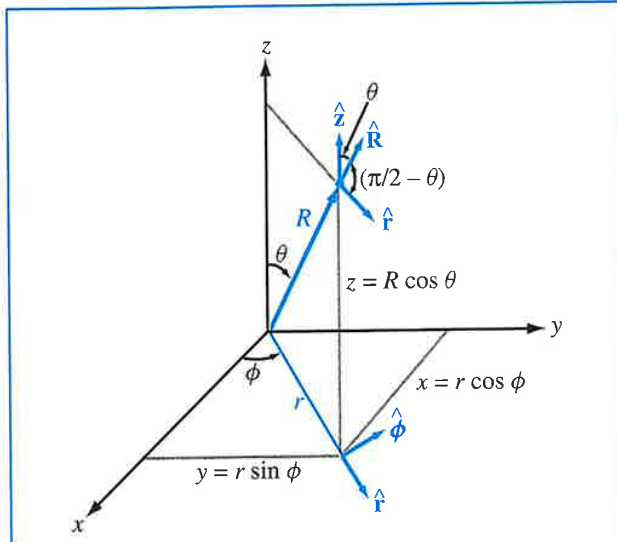


Figure 3-18: Interrelationships between (x, y, z) and (R, θ, ϕ) .

where a and b are transformation coefficients. Since $\hat{\mathbf{r}}$ and $\hat{\mathbf{z}}$ are mutually perpendicular,

$$\hat{\mathbf{R}} \cdot \hat{\mathbf{r}} = a, \quad (3.63a)$$

$$\hat{\mathbf{R}} \cdot \hat{\mathbf{z}} = b. \quad (3.63b)$$

From Fig. 3-18, the angle between $\hat{\mathbf{R}}$ and $\hat{\mathbf{r}}$ is the complement of θ and that between $\hat{\mathbf{R}}$ and $\hat{\mathbf{z}}$ is θ . Hence, $a = \hat{\mathbf{R}} \cdot \hat{\mathbf{r}} = \sin \theta$ and $b = \hat{\mathbf{R}} \cdot \hat{\mathbf{z}} = \cos \theta$. Upon inserting these expressions for a and b in Eq. (3.62) and replacing $\hat{\mathbf{r}}$ with Eq. (3.56a), we have

$$\hat{\mathbf{R}} = \hat{\mathbf{x}} \sin \theta \cos \phi + \hat{\mathbf{y}} \sin \theta \sin \phi + \hat{\mathbf{z}} \cos \theta. \quad (3.64a)$$

A similar procedure can be followed to obtain the following expression for $\hat{\theta}$:

$$\hat{\theta} = \hat{\mathbf{x}} \cos \theta \cos \phi + \hat{\mathbf{y}} \cos \theta \sin \phi - \hat{\mathbf{z}} \sin \theta, \quad (3.64b)$$

and the expression for $\hat{\phi}$ is given by Eq. (3.56b) as

$$\hat{\phi} = -\hat{\mathbf{x}} \sin \phi + \hat{\mathbf{y}} \cos \phi. \quad (3.64c)$$

Equations (3.64a) through (3.64c) can be solved simultaneously to give the following expressions for $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$ in terms of $(\hat{\mathbf{R}}, \hat{\theta}, \hat{\phi})$:

$$\hat{\mathbf{x}} = \hat{\mathbf{R}} \sin \theta \cos \phi + \hat{\theta} \cos \theta \cos \phi - \hat{\phi} \sin \phi, \quad (3.65a)$$

$$\hat{\mathbf{y}} = \hat{\mathbf{R}} \sin \theta \sin \phi + \hat{\theta} \cos \theta \sin \phi + \hat{\phi} \cos \phi, \quad (3.65b)$$

$$\hat{\mathbf{z}} = \hat{\mathbf{R}} \cos \theta - \hat{\theta} \sin \theta. \quad (3.65c)$$

Equations (3.64a) to (3.65c) can also be used to transform (A_x, A_y, A_z) of vector $\mathbf{A}$ into its spherical components (A_R, A_θ, A_ϕ) , and vice versa, by replacing $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}, \hat{\mathbf{R}}, \hat{\theta}, \hat{\phi})$ with $(A_x, A_y, A_z, A_R, A_\theta, A_\phi)$, respectively.

Example 3-8 Cartesian to Spherical Transformation

Express vector $\mathbf{A} = \hat{\mathbf{x}}(x+y) + \hat{\mathbf{y}}(y-x) + \hat{\mathbf{z}}z$ in spherical coordinates.

Solution: Using the transformation relation for A_R given in Table 3-2, we have

$$\begin{aligned} A_R &= A_x \sin \theta \cos \phi + A_y \sin \theta \sin \phi + A_z \cos \theta \\ &= (x+y) \sin \theta \cos \phi + (y-x) \sin \theta \sin \phi + z \cos \theta. \end{aligned}$$

Using the expressions for x, y , and z given by Eq. (3.61c), we have

$$\begin{aligned} A_R &= (R \sin \theta \cos \phi + R \sin \theta \sin \phi) \sin \theta \cos \phi \\ &\quad + (R \sin \theta \sin \phi - R \sin \theta \cos \phi) \sin \theta \sin \phi + R \cos^2 \theta \\ &= R \sin^2 \theta (\cos^2 \phi + \sin^2 \phi) + R \cos^2 \theta \\ &= R \sin^2 \theta + R \cos^2 \theta = R. \end{aligned}$$

Similarly,

$$A_\theta = (x + y) \cos \theta \cos \phi + (y - x) \cos \theta \sin \phi - z \sin \theta,$$

$$A_\phi = -(x + y) \sin \phi + (y - x) \cos \phi,$$

and following the procedure used with A_R , we obtain the results

$$A_\theta = 0,$$

$$A_\phi = -R \sin \theta.$$

Hence,

$$\mathbf{A} = \hat{\mathbf{R}}A_R + \hat{\boldsymbol{\theta}}A_\theta + \hat{\boldsymbol{\phi}}A_\phi = \hat{\mathbf{R}}R - \hat{\boldsymbol{\phi}}R \sin \theta. \quad \blacksquare$$

3-3.3 Cylindrical to Spherical Transformations

Transformations between cylindrical and spherical coordinates can be realized by combining the transformation relations of the preceding two subsections. The results are given in Table 3-2.

3-3.4 Distance between Two Points

In Cartesian coordinates, the distance d between two points $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$ is given by Eq. (3.12) as

$$\begin{aligned} d &= |\mathbf{R}_{12}| \\ &= [(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2]^{1/2}. \end{aligned} \quad (3.66)$$

Upon using Eq. (3.52) to convert the Cartesian coordinates of P_1 and P_2 into their cylindrical equivalents, we have

$$\begin{aligned} d &= [(r_2 \cos \phi_2 - r_1 \cos \phi_1)^2 \\ &\quad + (r_2 \sin \phi_2 - r_1 \sin \phi_1)^2 + (z_2 - z_1)^2]^{1/2} \\ &= [r_2^2 + r_1^2 - 2r_1r_2 \cos(\phi_2 - \phi_1) + (z_2 - z_1)^2]^{1/2} \\ &\quad \text{(cylindrical)}. \end{aligned} \quad (3.67)$$

A similar transformation using Eqs. (3.61a-c) leads to an expression for d in terms of the spherical coordinates of P_1 and P_2 :

$$\begin{aligned} d &= \{R_2^2 + R_1^2 - 2R_1R_2[\cos \theta_2 \cos \theta_1 \\ &\quad + \sin \theta_1 \sin \theta_2 \cos(\phi_2 - \phi_1)]\}^{1/2} \\ &\quad \text{(spherical)}. \end{aligned} \quad (3.68)$$

REVIEW QUESTIONS

Q3.7 Why do we use more than one coordinate system?

Q3.8 Why is it that the base vectors $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$ are independent of the location of a point, but $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\phi}}$ are not?

Q3.9 What are the cyclic relations for the base vectors in (a) Cartesian coordinates, (b) cylindrical coordinates, and (c) spherical coordinates?

Q3.10 How is the position vector of a point in cylindrical coordinates related to its position vector in spherical coordinates?

EXERCISE 3.5 Point $P(2\sqrt{3}, \pi/3, -2)$ is given in cylindrical coordinates. Express P in spherical coordinates.

Ans. $P(4, 2\pi/3, \pi/3)$.

EXERCISE 3.6 Transform vector

$$\mathbf{A} = \hat{\mathbf{x}}(x + y) + \hat{\mathbf{y}}(y - x) + \hat{\mathbf{z}}z$$

from Cartesian to cylindrical coordinates.

Ans. $\mathbf{A} = \hat{\mathbf{r}}r - \hat{\boldsymbol{\phi}}r + \hat{\mathbf{z}}z$.

3-4 Gradient of a Scalar Field

When dealing with a scalar physical quantity whose magnitude depends on a single variable, such as the temperature T as a function of height z , the rate of change of T with height can be described by the derivative dT/dz . However, if T is also a function of x and y in a Cartesian coordinate system, its space rate of change becomes more difficult to describe because we now have to deal not only with three separate variables, but also do it in a unified arrangement. The differential change in T along x , y , and z can be described in terms of the partial derivatives of T with respect to the three coordinate variables, but it is not immediately obvious as to how we should combine the three partial derivatives so as to describe the space rate of change of T along a specified direction. Furthermore, many of the quantities we deal with in electromagnetics are vectors, and therefore both their magnitudes and directions may vary with spatial position. In vector calculus, we use three fundamental operators to describe the differential spatial variations of scalars and vectors; these are the **gradient**, **divergence**, and **curl** operators. The gradient operator applies to scalar fields and is the subject of the present section. The other two operators, which apply to vector fields, are discussed in succeeding sections.

Suppose that $T_1(x, y, z)$ is the temperature at point $P_1(x, y, z)$ in some region of space, and $T_2(x + dx, y + dy, z + dz)$ is the temperature at a nearby point P_2 , as shown in Fig. 3-19. The differential distances dx , dy , and dz are the components of the differential distance vector $d\mathbf{l}$. That is,

$$d\mathbf{l} = \hat{\mathbf{x}} dx + \hat{\mathbf{y}} dy + \hat{\mathbf{z}} dz. \quad (3.69)$$

From differential calculus, the differential temperature $dT = T_2 - T_1$ is given by

$$dT = \frac{\partial T}{\partial x} dx + \frac{\partial T}{\partial y} dy + \frac{\partial T}{\partial z} dz, \quad (3.70)$$

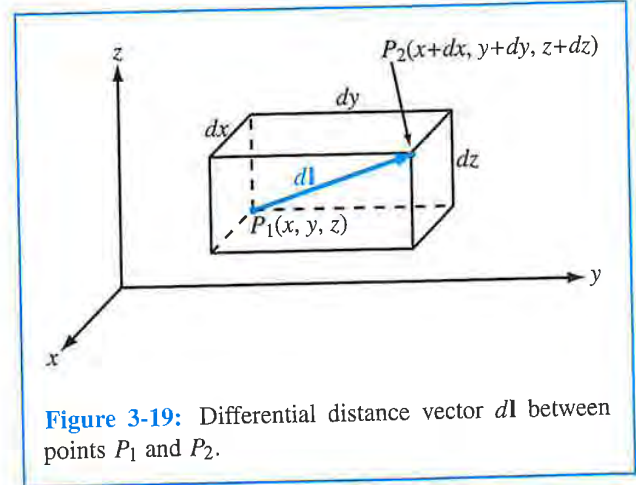


Figure 3-19: Differential distance vector $d\mathbf{l}$ between points P_1 and P_2 .

and since by definition $dx = \hat{\mathbf{x}} \cdot d\mathbf{l}$, $dy = \hat{\mathbf{y}} \cdot d\mathbf{l}$, and $dz = \hat{\mathbf{z}} \cdot d\mathbf{l}$, Eq. (3.70) can be rewritten as

$$\begin{aligned} dT &= \hat{\mathbf{x}} \frac{\partial T}{\partial x} \cdot d\mathbf{l} + \hat{\mathbf{y}} \frac{\partial T}{\partial y} \cdot d\mathbf{l} + \hat{\mathbf{z}} \frac{\partial T}{\partial z} \cdot d\mathbf{l} \\ &= \left[\hat{\mathbf{x}} \frac{\partial T}{\partial x} + \hat{\mathbf{y}} \frac{\partial T}{\partial y} + \hat{\mathbf{z}} \frac{\partial T}{\partial z} \right] \cdot d\mathbf{l}. \end{aligned} \quad (3.71)$$

The vector inside the square brackets in Eq. (3.71) defines the change in temperature dT corresponding to a vector change in position $d\mathbf{l}$. This vector is called the **gradient** of T , or **grad** T for short, and it is usually written symbolically as ∇T . That is,

$$\nabla T = \text{grad } T \triangleq \hat{\mathbf{x}} \frac{\partial T}{\partial x} + \hat{\mathbf{y}} \frac{\partial T}{\partial y} + \hat{\mathbf{z}} \frac{\partial T}{\partial z}, \quad (3.72)$$

and Eq. (3.71) can be expressed in the form

$$dT = \nabla T \cdot d\mathbf{l}. \quad (3.73)$$

The symbol ∇ is called the **del** or **gradient operator** and is defined as

$$\nabla \triangleq \hat{\mathbf{x}} \frac{\partial}{\partial x} + \hat{\mathbf{y}} \frac{\partial}{\partial y} + \hat{\mathbf{z}} \frac{\partial}{\partial z} \quad (\text{Cartesian}). \quad (3.74)$$

We note that, *whereas the gradient operator has no physical meaning by itself, it attains a physical meaning once it operates on a scalar physical quantity, and the result of the operation is a vector whose magnitude is equal to the maximum rate of change of the physical quantity per unit distance and whose direction is along the direction of maximum increase.* With $d\mathbf{l} = \hat{\mathbf{a}}_l dl$, where $\hat{\mathbf{a}}_l$ is the unit vector of $d\mathbf{l}$, the *directional derivative* of T along the direction $\hat{\mathbf{a}}_l$ is given by

$$\frac{dT}{dl} = \nabla T \cdot \hat{\mathbf{a}}_l. \quad (3.75)$$

If ∇T is a known function of the coordinate variables of a given coordinate system, we can find the difference $(T_2 - T_1)$, where T_1 and T_2 are the values of T at points P_1 and P_2 , respectively, by integrating both sides of Eq. (3.73). Thus,

$$T_2 - T_1 = \int_{P_1}^{P_2} \nabla T \cdot d\mathbf{l}. \quad (3.76)$$

Example 3-9 Directional Derivative

Find the directional derivative of $T = x^2 + y^2z$ along the direction $\hat{\mathbf{x}}2 + \hat{\mathbf{y}}3 - \hat{\mathbf{z}}2$ and evaluate it at $(1, -1, 2)$.

Solution: The directional derivative dT/dl is given by Eq. (3.75). First, we find the gradient of T :

$$\begin{aligned} \nabla T &= \left(\hat{\mathbf{x}} \frac{\partial}{\partial x} + \hat{\mathbf{y}} \frac{\partial}{\partial y} + \hat{\mathbf{z}} \frac{\partial}{\partial z} \right) (x^2 + y^2z) \\ &= \hat{\mathbf{x}}2x + \hat{\mathbf{y}}2yz + \hat{\mathbf{z}}y^2. \end{aligned}$$

We denote $\mathbf{l}$ as the given direction,

$$\mathbf{l} = \hat{\mathbf{x}}2 + \hat{\mathbf{y}}3 - \hat{\mathbf{z}}2.$$

Its unit vector is

$$\hat{\mathbf{a}}_l = \frac{\mathbf{l}}{|\mathbf{l}|} = \frac{\hat{\mathbf{x}}2 + \hat{\mathbf{y}}3 - \hat{\mathbf{z}}2}{\sqrt{2^2 + 3^2 + 2^2}} = \frac{\hat{\mathbf{x}}2 + \hat{\mathbf{y}}3 - \hat{\mathbf{z}}2}{\sqrt{17}}.$$

Application of Eq. (3.75) gives

$$\begin{aligned} \frac{dT}{dl} &= \nabla T \cdot \hat{\mathbf{a}}_l = (\hat{\mathbf{x}}2x + \hat{\mathbf{y}}2yz + \hat{\mathbf{z}}y^2) \cdot \left(\frac{\hat{\mathbf{x}}2 + \hat{\mathbf{y}}3 - \hat{\mathbf{z}}2}{\sqrt{17}} \right) \\ &= \frac{4x + 6yz - 2y^2}{\sqrt{17}}. \end{aligned}$$

At $(1, -1, 2)$,

$$\left. \frac{dT}{dl} \right|_{(1, -1, 2)} = \frac{4 - 12 - 2}{\sqrt{17}} = \frac{-10}{\sqrt{17}}. \quad \blacksquare$$

3-4.1 Gradient Operator in Cylindrical and Spherical Coordinates

Even though Eq. (3.73) was derived using Cartesian coordinates, it should be equally valid in any orthogonal coordinate system. In order to apply the gradient operator to a scalar quantity expressed in cylindrical or spherical coordinates, we need expressions for ∇ in those coordinate systems. To convert Eq. (3.72) into the cylindrical coordinates (r, ϕ, z) , we start by restating the coordinate relations

$$r = \sqrt{x^2 + y^2}, \quad \tan \phi = \frac{y}{x}. \quad (3.77)$$

From differential calculus,

$$\frac{\partial T}{\partial x} = \frac{\partial T}{\partial r} \frac{\partial r}{\partial x} + \frac{\partial T}{\partial \phi} \frac{\partial \phi}{\partial x} + \frac{\partial T}{\partial z} \frac{\partial z}{\partial x}. \quad (3.78)$$

Since z is orthogonal to x , the last term is equal to zero because $\partial z / \partial x = 0$. Using the coordinate relations given by Eq. (3.77), it is easy to show that

$$\frac{\partial r}{\partial x} = \frac{x}{\sqrt{x^2 + y^2}} = \cos \phi, \quad (3.79a)$$

$$\frac{\partial \phi}{\partial x} = -\frac{1}{r} \sin \phi. \quad (3.79b)$$

Hence,

$$\frac{\partial T}{\partial x} = \cos \phi \frac{\partial T}{\partial r} - \frac{\sin \phi}{r} \frac{\partial T}{\partial \phi}. \quad (3.80)$$

This expression can be used to replace the coefficient of $\hat{x}$ in Eq. (3.72), and a similar procedure can be followed to obtain an expression for $\partial T/\partial y$ in terms of r and ϕ . If, in addition, we use the relations $\hat{x} = \hat{r} \cos \phi - \hat{\phi} \sin \phi$ and $\hat{y} = \hat{r} \sin \phi + \hat{\phi} \cos \phi$ [from Eqs. (3.57a) and (3.57b)], Eq. (3.72) becomes

$$\nabla T = \hat{r} \frac{\partial T}{\partial r} + \hat{\phi} \frac{1}{r} \frac{\partial T}{\partial \phi} + \hat{z} \frac{\partial T}{\partial z}, \quad (3.81)$$

and therefore the gradient operator in cylindrical coordinates can be defined as

$$\nabla = \hat{r} \frac{\partial}{\partial r} + \hat{\phi} \frac{1}{r} \frac{\partial}{\partial \phi} + \hat{z} \frac{\partial}{\partial z} \quad (\text{cylindrical}). \quad (3.82)$$

A similar procedure leads to the following expression for the gradient in spherical coordinates:

$$\nabla = \hat{R} \frac{\partial}{\partial R} + \hat{\theta} \frac{1}{R} \frac{\partial}{\partial \theta} + \hat{\phi} \frac{1}{R \sin \theta} \frac{\partial}{\partial \phi} \quad (\text{spherical}). \quad (3.83)$$

3-4.2 Properties of the Gradient Operator

For any two scalar functions U and V , the following relations apply:

$$(1) \quad \nabla(U + V) = \nabla U + \nabla V, \quad (3.84a)$$

$$(2) \quad \nabla(UV) = U \nabla V + V \nabla U, \quad (3.84b)$$

$$(3) \quad \nabla V^n = n V^{n-1} \nabla V, \quad \text{for any } n. \quad (3.84c)$$

The gradient of a vector is meaningless under the rules of vector calculus.

Example 3-10 Calculating the Gradient

Find the gradient of each of the following scalar functions and then evaluate it at the given point.

- (a) $V_1 = 24V_0 \cos(\pi y/3) \sin(2\pi z/3)$ at $(3, 2, 1)$ in Cartesian coordinates,

- (b) $V_2 = V_0 e^{-2r} \sin 3\phi$ at $(1, \pi/2, 3)$ in cylindrical coordinates,

- (c) $V_3 = V_0 (a/R) \cos 2\theta$ at $(2a, 0, \pi)$ in spherical coordinates.

Solution: (a) In Cartesian coordinates,

$$\begin{aligned} \nabla V_1 &= \hat{x} \frac{\partial V_1}{\partial x} + \hat{y} \frac{\partial V_1}{\partial y} + \hat{z} \frac{\partial V_1}{\partial z} \\ &= -\hat{y} 8\pi V_0 \sin \frac{\pi y}{3} \sin \frac{2\pi z}{3} + \hat{z} 16\pi V_0 \cos \frac{\pi y}{3} \cos \frac{2\pi z}{3} \\ &= 8\pi V_0 \left[-\hat{y} \sin \frac{\pi y}{3} \sin \frac{2\pi z}{3} + \hat{z} 2 \cos \frac{\pi y}{3} \cos \frac{2\pi z}{3} \right]. \end{aligned}$$

At $(3, 2, 1)$,

$$\begin{aligned} \nabla V_1 &= 8\pi V_0 \left[-\hat{y} \sin^2 \frac{2\pi}{3} + \hat{z} 2 \cos^2 \frac{2\pi}{3} \right] \\ &= \pi V_0 [-\hat{y} 6 + \hat{z} 4]. \end{aligned}$$

- (b) The function V_2 is expressed in terms of cylindrical variables. Hence, we need to use Eq. (3.82) for ∇ :

$$\begin{aligned} \nabla V_2 &= \left(\hat{r} \frac{\partial}{\partial r} + \hat{\phi} \frac{1}{r} \frac{\partial}{\partial \phi} + \hat{z} \frac{\partial}{\partial z} \right) V_0 e^{-2r} \sin 3\phi \\ &= -\hat{r} 2 V_0 e^{-2r} \sin 3\phi + \hat{\phi} (3 V_0 e^{-2r} \cos 3\phi) / r \\ &= \left[-\hat{r} 2 \sin 3\phi + \hat{\phi} \frac{3 \cos 3\phi}{r} \right] V_0 e^{-2r}. \end{aligned}$$

At $(1, \pi/2, 3)$, $r = 1$ and $\phi = \pi/2$. Hence,

$$\begin{aligned} \nabla V_2 &= \left[-\hat{r} 2 \sin \frac{3\pi}{2} + \hat{\phi} 3 \cos \frac{3\pi}{2} \right] V_0 e^{-2} \\ &= \hat{r} 2 V_0 e^{-2} = \hat{r} 0.27 V_0. \end{aligned}$$

- (c) As V_3 is expressed in spherical coordinates, we apply Eq. (3.83) to V_3 :

$$\begin{aligned} \nabla V_3 &= \left(\hat{R} \frac{\partial}{\partial R} + \hat{\theta} \frac{1}{R} \frac{\partial}{\partial \theta} + \hat{\phi} \frac{1}{R \sin \theta} \frac{\partial}{\partial \phi} \right) V_0 \left(\frac{a}{R} \right) \cos 2\theta \\ &= -\hat{R} \frac{V_0 a}{R^2} \cos 2\theta - \hat{\theta} \frac{2 V_0 a}{R^2} \sin 2\theta \\ &= -[\hat{R} \cos 2\theta + \hat{\theta} 2 \sin 2\theta] \frac{V_0 a}{R^2}. \end{aligned}$$

At $(2a, 0, \pi)$, $R = 2a$ and $\theta = 0$, which gives

$$\nabla V_3 = -\hat{\mathbf{R}} \frac{V_0}{4a}. \quad \blacksquare$$

EXERCISE 3.7 Given $V = x^2y + xy^2 + xz^2$, (a) find the gradient of V , and (b) evaluate it at $(1, -1, 2)$.

Ans. (a) $\nabla V = \hat{\mathbf{x}}(2xy + y^2 + z^2) + \hat{\mathbf{y}}(x^2 + 2xy) + \hat{\mathbf{z}}2xz$,
(b) $\nabla V|_{(1,-1,2)} = \hat{\mathbf{x}}3 - \hat{\mathbf{y}} + \hat{\mathbf{z}}4$.

EXERCISE 3.8 Find the directional derivative of $V = rz^2 \cos 2\phi$ along the direction $\mathbf{A} = \hat{\mathbf{r}}2 - \hat{\mathbf{z}}$ and evaluate it at $(1, \pi/2, 2)$.

Ans. $(dV/dl)|_{(1,\pi/2,2)} = -4/\sqrt{5}$.

3-5 Divergence of a Vector Field

From our brief introduction of Coulomb's law in Chapter 1, we know that an isolated, positive point charge q induces an electric field $\mathbf{E}$ in the space around it, with the direction of $\mathbf{E}$ being along the outward direction away from the charge. Also, the strength (magnitude) of $\mathbf{E}$ is proportional to q and decreases with distance R from the charge as $1/R^2$. In a graphical presentation, a vector field is usually represented by **field lines**, as shown in Fig. 3-20. The arrowhead denotes the direction of the field at the point where the field line is drawn, and the length of the line provides a qualitative depiction of the field's magnitude.

Even though the electric field vector does not actually move, we regard its presence as a flux flowing through space, and we refer to its field lines as **flux lines**. At a surface boundary, **flux density** is defined as the amount of outward flux crossing a unit surface ds :

$$\text{Flux density of } \mathbf{E} = \frac{\mathbf{E} \cdot d\mathbf{s}}{|d\mathbf{s}|} = \frac{\mathbf{E} \cdot \hat{\mathbf{n}} ds}{ds}, \quad (3.85)$$

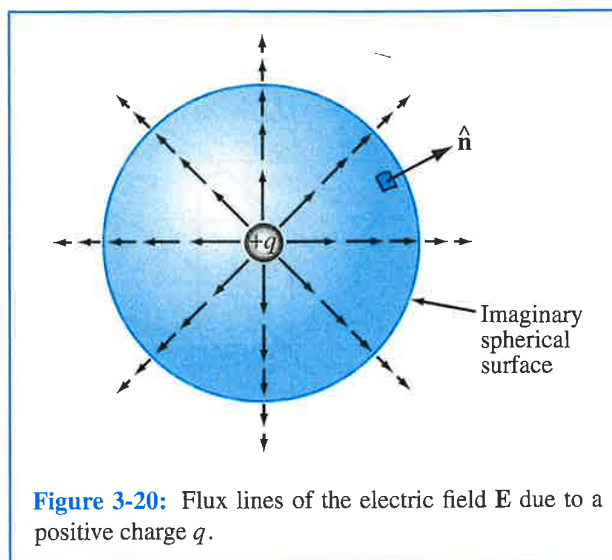


Figure 3-20: Flux lines of the electric field $\mathbf{E}$ due to a positive charge q .

where $\hat{\mathbf{n}}$ is the outward surface normal of ds . The **total flux** crossing a closed surface S , such as the enclosed surface of the imaginary sphere outlined in Fig. 3-20, is

$$\text{Total flux} = \oint_S \mathbf{E} \cdot d\mathbf{s}. \quad (3.86)$$

Let us now consider the case of a differential rectangular parallelepiped, such as a cube, whose edges are lined up with the axes of a Cartesian coordinate system as shown in Fig. 3-21. The lengths of the edges are Δx along x , Δy along y , and Δz along z . A vector field $\mathbf{E}(x, y, z)$ exists in the region of space containing the parallelepiped, and we wish to determine the flux of $\mathbf{E}$ through its total surface S . Since S includes six faces, we need to sum up the fluxes through all of them, and by definition the flux through any face is the **outward** flux from the volume Δv through that face.

Let $\mathbf{E}$ be defined as

$$\mathbf{E} = \hat{\mathbf{x}}E_x + \hat{\mathbf{y}}E_y + \hat{\mathbf{z}}E_z. \quad (3.87)$$

The area of the face marked 1 in Fig. 3-21 is $\Delta y \Delta z$, and its unit vector $\hat{\mathbf{n}}_1 = -\hat{\mathbf{x}}$. Hence, the outward flux F_1

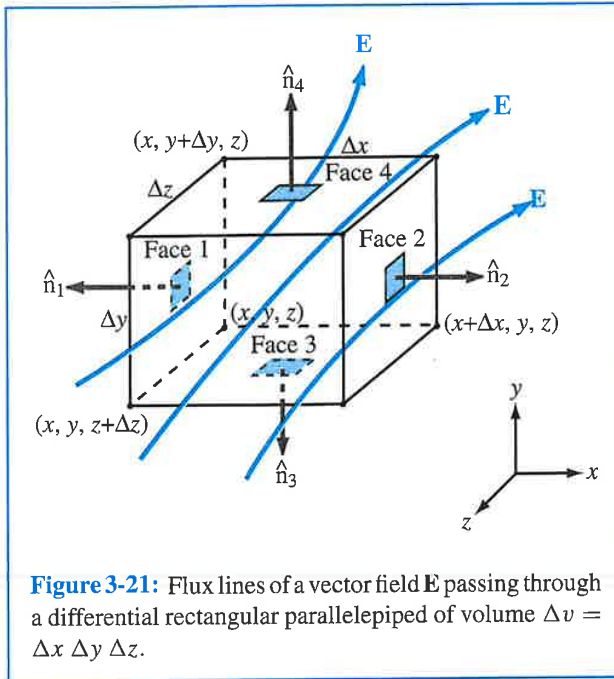


Figure 3-21: Flux lines of a vector field $\mathbf{E}$ passing through a differential rectangular parallelepiped of volume $\Delta v = \Delta x \Delta y \Delta z$.

through face 1 is

$$\begin{aligned} F_1 &= \int_{\text{Face 1}} \mathbf{E} \cdot \hat{\mathbf{n}}_1 \, ds \\ &= \int_{\text{Face 1}} (\hat{\mathbf{x}}E_x + \hat{\mathbf{y}}E_y + \hat{\mathbf{z}}E_z) \cdot (-\hat{\mathbf{x}}) \, dy \, dz \\ &= -E_x(1) \Delta y \Delta z, \end{aligned} \quad (3.88)$$

where $E_x(1)$ is the value of E_x at the center of face 1. Approximating E_x over face 1 by its value at the center is justified by the assumption that the differential volume under consideration is very small.

Similarly, the flux out of face 2 (with $\hat{\mathbf{n}}_2 = \hat{\mathbf{x}}$) is

$$F_2 = E_x(2) \Delta y \Delta z, \quad (3.89)$$

where $E_x(2)$ is the value of E_x at the center of face 2. Over a differential separation Δx , $E_x(2)$ is related to

$E_x(1)$ by

$$E_x(2) = E_x(1) + \frac{\partial E_x}{\partial x} \Delta x, \quad (3.90)$$

where we have ignored higher-order terms involving $(\Delta x)^2$ and higher powers because their contributions are negligibly small when Δx is very small. Substituting Eq. (3.90) into Eq. (3.89) gives

$$F_2 = \left[E_x(1) + \frac{\partial E_x}{\partial x} \Delta x \right] \Delta y \Delta z. \quad (3.91)$$

The sum of the fluxes out of faces 1 and 2 is obtained by adding Eqs. (3.88) and (3.91),

$$F_1 + F_2 = \frac{\partial E_x}{\partial x} \Delta x \Delta y \Delta z. \quad (3.92a)$$

Repeating the same procedure to each of the other pairs of faces leads to

$$F_3 + F_4 = \frac{\partial E_y}{\partial y} \Delta x \Delta y \Delta z, \quad (3.92b)$$

$$F_5 + F_6 = \frac{\partial E_z}{\partial z} \Delta x \Delta y \Delta z. \quad (3.92c)$$

The sum of fluxes F_1 through F_6 gives the total flux through surface S of the parallelepiped:

$$\begin{aligned} \oint_S \mathbf{E} \cdot d\mathbf{s} &= \left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} \right) \Delta x \Delta y \Delta z \\ &= (\text{div } \mathbf{E}) \Delta v, \end{aligned} \quad (3.93)$$

where $\Delta v = \Delta x \Delta y \Delta z$ and $\text{div } \mathbf{E}$ is a differential function called the **divergence** of $\mathbf{E}$ and is defined in Cartesian coordinates as

$$\text{div } \mathbf{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}. \quad (3.94)$$

By shrinking the volume Δv to zero, we define the divergence of $\mathbf{E}$ at a point as the net outward flux per

unit volume over a closed incremental surface. Thus, from Eq. (3.93), we have

$$\text{div } \mathbf{E} \triangleq \lim_{\Delta v \rightarrow 0} \frac{\oint_S \mathbf{E} \cdot d\mathbf{s}}{\Delta v}, \quad (3.95)$$

where S encloses the elemental volume Δv . Instead of denoting the divergence of $\mathbf{E}$ by $\text{div } \mathbf{E}$, it is common practice to denote it as $\nabla \cdot \mathbf{E}$. That is,

$$\nabla \cdot \mathbf{E} \triangleq \text{div } \mathbf{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} \quad (3.96)$$

for a vector $\mathbf{E}$ in Cartesian coordinates. Further discussion of these two forms of notation is given in Section 3-5.2.

From the definition of the divergence of $\mathbf{E}$ given by Eq. (3.95), the field $\mathbf{E}$ has positive divergence if the net flux out of surface S is positive, which may be viewed as if the volume Δv contains a *source* of flux. If the divergence is negative, Δv may be viewed as a *sink* because the net flux is inward into Δv . For a uniform field $\mathbf{E}$, the same amount of flux enters Δv as leaves from it; hence, its divergence is zero and the field is said to be *divergenceless*. The divergence is a differential operator, it can operate only on vectors, and the result of its operation is a scalar. This is in contrast with the gradient operator, which can operate only on scalars and the result is a vector. Expressions for the divergence of a vector in cylindrical and spherical coordinates are given on the inside back cover of this book.

The divergence operator is distributive. That is, for any pair of vectors $\mathbf{E}_1$ and $\mathbf{E}_2$,

$$\nabla \cdot (\mathbf{E}_1 + \mathbf{E}_2) = \nabla \cdot \mathbf{E}_1 + \nabla \cdot \mathbf{E}_2. \quad (3.97)$$

If $\nabla \cdot \mathbf{E} = 0$, the vector field $\mathbf{E}$ is called *solenoidal*.

3-5.1 Divergence Theorem

The result given by Eq. (3.93) for a differential volume Δv can be extended to relate the volume integral of $\nabla \cdot \mathbf{E}$ over any volume v to the flux of $\mathbf{E}$ through the closed surface S that bounds v . That is,

$$\int_v \nabla \cdot \mathbf{E} dv = \oint_S \mathbf{E} \cdot d\mathbf{s} \quad (\text{divergence theorem}). \quad (3.98)$$

This relationship, known as the *divergence theorem*, is used extensively in electromagnetics.

3-5.2 Remarks on Notation

Books on vector analysis use one or both of the following two notations for denoting the gradient, divergence, and curl operators. In the linguistic notation, the gradient of a scalar T is written as “grad T ,” and the divergence and curl of a vector $\mathbf{E}$ are written as “div $\mathbf{E}$ ” and “curl $\mathbf{E}$.” A few years after James Clerk Maxwell published his classic treatise on electricity and magnetism in 1873, at which time vector analysis was not yet available, the American mathematician J. Willard Gibbs published a pair of pamphlets entitled *Elements of Vector Analysis*, in which he established the field of vector analysis as we know it today. In his presentation, he introduced a new notation for the divergence and curl operations: “ $\nabla \cdot \mathbf{E}$ ” for div $\mathbf{E}$ and “ $\nabla \times \mathbf{E}$ ” for curl $\mathbf{E}$. Because of its symbolic simplicity, Gibbs’s notation was quickly adopted and is now used in the overwhelming majority of books on electromagnetics. The problem with Gibbs’s notation, however, is that it is prone to misinterpretation. In Gibbs’s notation, $\nabla \cdot \mathbf{E}$ stands for the divergence of vector $\mathbf{E}$. This is commonly misinterpreted as the dot product of the operator ∇ with the vector $\mathbf{E}$, although Gibbs himself did not call it that. The reason for the misinterpretation is that algebraically the dot product of ∇ , as given by Eq. (3.74), with $\mathbf{E}$, as given by

Eq. (3.87), gives the same result as the expression for the divergence of $\mathbf{E}$ given by Eq. (3.96). However, the “dot product” interpretation is incorrect on two grounds. First, the operator ∇ is treated as if it were a vector quantity, which it is not. As we mentioned earlier, a differential operator has no physical meaning in itself and only attains a meaning after it operates on a physical quantity. This may be a subtle point, but it is nonetheless fundamental. The second criticism of the dot-product interpretation is that it can lead to invalid results under coordinate transformations, which is unacceptable because the divergence of a physical quantity should be the same regardless of the choice of coordinate system. Similar criticisms apply to the interpretation of $\nabla \times \mathbf{E}$ as the cross product of ∇ and $\mathbf{E}$.

To avoid the misinterpretation problems associated with Gibbs's notation, Tai* introduced the notation $\nabla \mathbf{E}$ for the divergence of $\mathbf{E}$ and $\nabla \mathbf{E}$ for the curl of $\mathbf{E}$. Tai's proposed notation, however, is relatively new, and it will take some time before its logical foundation is properly appreciated. Hence, in keeping with the established tradition, we will still adopt Gibbs's notation in this book, but we reiterate that “ $\nabla \cdot$ ” and “ $\nabla \times$ ” should not be regarded as the del operator followed with a dot or cross product.

Example 3-11 Calculating the Divergence

Determine the divergence of each of the following vector fields and then evaluate it at the indicated point:

(a) $\mathbf{E} = \hat{x}3x^2 + \hat{y}2z + \hat{z}x^2z$ at $(2, -2, 0)$;

(b) $\mathbf{E} = \hat{R}(a^3 \cos \theta / R^2) - \hat{\theta}(a^3 \sin \theta / R^2)$ at $(a/2, 0, \pi)$.

Solution

$$\begin{aligned} \text{(a)} \quad \nabla \cdot \mathbf{E} &= \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} \\ &= \frac{\partial}{\partial x}(3x^2) + \frac{\partial}{\partial y}(2z) + \frac{\partial}{\partial z}(x^2z) \\ &= 6x + 0 + x^2 = x^2 + 6x. \end{aligned}$$

*C. T. Tai, *Generalized Vector and Dyadic Analysis*, IEEE Press, New York, 1992, pp. 41–44.

At $(2, -2, 0)$, $\nabla \cdot \mathbf{E} \Big|_{(2,-2,0)} = 16$.

(b) Using the expression given on the inside of the back cover of the book for the divergence of a vector in spherical coordinates, we have

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \frac{1}{R^2} \frac{\partial}{\partial R}(R^2 E_R) + \frac{1}{R \sin \theta} \frac{\partial}{\partial \theta}(E_\theta \sin \theta) \\ &\quad + \frac{1}{R \sin \theta} \frac{\partial E_\phi}{\partial \phi} \\ &= \frac{1}{R^2} \frac{\partial}{\partial R}(a^3 \cos \theta) + \frac{1}{R \sin \theta} \frac{\partial}{\partial \theta} \left(-\frac{a^3 \sin^2 \theta}{R^2} \right) \\ &= 0 - \frac{2a^3 \cos \theta}{R^3} = -\frac{2a^3 \cos \theta}{R^3}. \end{aligned}$$

At $R = a/2$ and $\theta = 0$, $\nabla \cdot \mathbf{E} \Big|_{(a/2,0,\pi)} = -16$. ■

EXERCISE 3.9 Given $\mathbf{A} = e^{-2y}(\hat{x} \sin 2x + \hat{y} \cos 2x)$, find $\nabla \cdot \mathbf{A}$.

Ans. $\nabla \cdot \mathbf{A} = 0$.

EXERCISE 3.10 Given $\mathbf{A} = \hat{r}r \cos \phi + \hat{\phi}r \sin \phi + \hat{z}3z$, find $\nabla \cdot \mathbf{A}$ at $(2, 0, 3)$.

Ans. $\nabla \cdot \mathbf{A} = 6$.

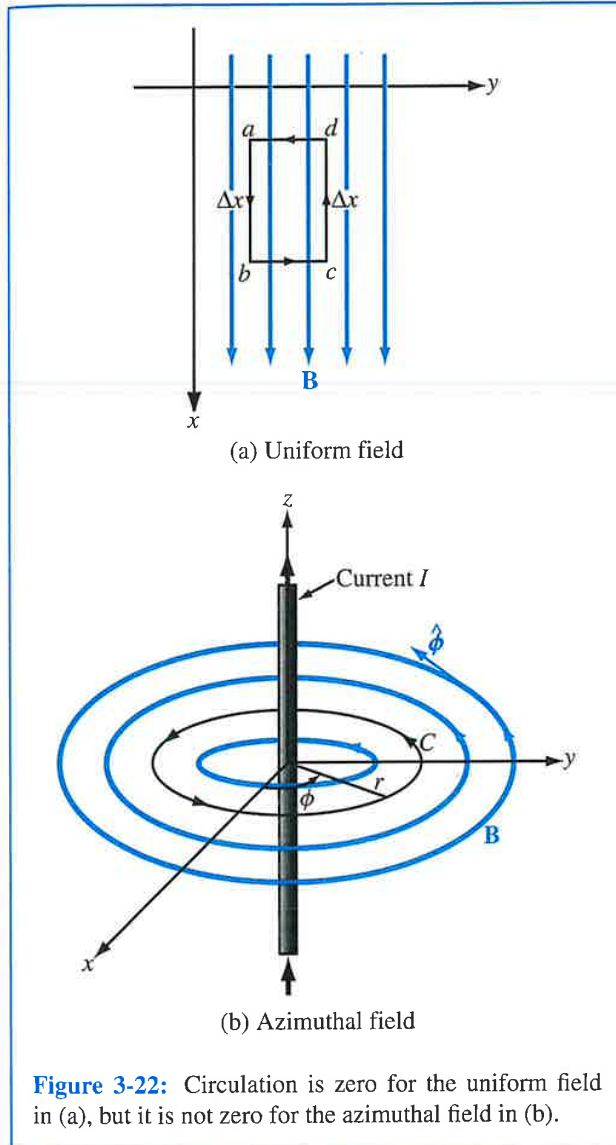
EXERCISE 3.11 If $\mathbf{E} = \hat{R}AR$ in spherical coordinates, calculate the flux of $\mathbf{E}$ through a spherical surface of radius a , centered at the origin.

Ans. $\oint_S \mathbf{E} \cdot d\mathbf{s} = 4\pi Aa^3$.

EXERCISE 3.12 Verify the divergence theorem by calculating the volume integral of the divergence of the field $\mathbf{E}$ of Exercise 3.11 over the volume bounded by the surface of radius a .

3-6 Curl of a Vector Field

So far we have defined and discussed two of the three fundamental operators used in vector analysis, the gradient of a scalar and the divergence of a vector.



Now we introduce the *curl operator*. The curl of a vector field $\mathbf{B}$ describes the rotational property, or the *circulation*, of $\mathbf{B}$. For a closed contour C , the circulation of $\mathbf{B}$ is defined as the line integral of $\mathbf{B}$ around C . That is,

$$\text{Circulation} = \oint_C \mathbf{B} \cdot d\mathbf{l} \quad (3.99)$$

To gain some physical understanding for this definition, we consider two examples. The first example is for a uniform field $\mathbf{B} = \hat{\mathbf{x}}B_0$, whose field lines are as depicted in Fig. 3-22(a). For the rectangular contour $abcd$ shown in the figure, we have

$$\begin{aligned} \text{Circulation} &= \int_a^b \hat{\mathbf{x}}B_0 \cdot \hat{\mathbf{x}} dx + \int_b^c \hat{\mathbf{x}}B_0 \cdot \hat{\mathbf{y}} dy \\ &\quad + \int_c^d \hat{\mathbf{x}}B_0 \cdot \hat{\mathbf{x}} dx + \int_d^a \hat{\mathbf{x}}B_0 \cdot \hat{\mathbf{y}} dy \\ &= B_0 \Delta x - B_0 \Delta x = 0, \end{aligned} \quad (3.100)$$

where $\Delta x = b - a = c - d$ and, because $\hat{\mathbf{x}} \cdot \hat{\mathbf{y}} = 0$, the second and fourth integrals are zero. According to Eq. (3.100), *the circulation of a uniform field is zero*.

Next, we consider the magnetic field $\mathbf{B}$ induced by an infinite wire carrying a d-c current I . If the current is in free space and it is oriented along the z -direction, then, from Eq. (1.13),

$$\mathbf{B} = \hat{\phi} \frac{\mu_0 I}{2\pi r}, \quad (3.101)$$

where μ_0 is the permeability of free space and r is the radial distance from the current in the x - y plane. The direction of $\mathbf{B}$ is along the azimuth direction $\hat{\phi}$. The field lines of $\mathbf{B}$ are concentric circles around the current source, as shown in Fig. 3-22(b). For a circular contour of radius r , the differential length vector $d\mathbf{l} = \hat{\phi} r d\phi$, and the circulation of $\mathbf{B}$ around C is

$$\begin{aligned} \text{Circulation} &= \oint_C \mathbf{B} \cdot d\mathbf{l} \\ &= \int_0^{2\pi} \hat{\phi} \frac{\mu_0 I}{2\pi r} \cdot \hat{\phi} r d\phi = \mu_0 I. \end{aligned} \quad (3.102)$$

In this case, the circulation is not zero, but had the contour C been in any plane perpendicular to the x - y plane, $d\mathbf{l}$ would not have had a $\hat{\phi}$ component, and the integral would have yielded a net circulation of zero. In other words, the magnitude of the circulation of $\mathbf{B}$ depends on the choice of contour. Also, the direction of the contour determines whether the circulation is positive or negative. If we are trying to describe the circulation of a tornado, for example, we would like to choose our contour such that the circulation of the wind field is maximum, and we would like the circulation to have both a magnitude and a direction, with the direction being toward the tornado's vortex. The **curl operator** has been defined to accommodate those properties. The curl of a vector field $\mathbf{B}$, denoted $\text{curl } \mathbf{B}$ or $\nabla \times \mathbf{B}$, is defined as

$$\nabla \times \mathbf{B} = \text{curl } \mathbf{B} \triangleq \lim_{\Delta s \rightarrow 0} \frac{1}{\Delta s} \left[\hat{\mathbf{n}} \oint_C \mathbf{B} \cdot d\mathbf{l} \right]_{\max} \quad (3.103)$$

Thus, $\text{curl } \mathbf{B}$ is the circulation of $\mathbf{B}$ per unit area, with the area Δs of the contour C being oriented such that the circulation is maximum. The direction of $\text{curl } \mathbf{B}$ is $\hat{\mathbf{n}}$, the unit normal of Δs , defined according to the right-hand rule: with the four fingers of the right hand following the contour direction $d\mathbf{l}$, the thumb points along $\hat{\mathbf{n}}$ (Fig. 3-23). The curl of a vector is defined at a point; this follows from the definition given by Eq. (3.103) in which the circulation is normalized to the area Δs , and Δs is made to approach zero. As was mentioned earlier in Section 3-5.2, when we use the notation $\nabla \times \mathbf{B}$ to denote $\text{curl } \mathbf{B}$, it should not be interpreted as the cross product of ∇ and $\mathbf{B}$.

For a vector $\mathbf{B}$ given in Cartesian coordinates as

$$\mathbf{B} = \hat{\mathbf{x}}B_x + \hat{\mathbf{y}}B_y + \hat{\mathbf{z}}B_z, \quad (3.104)$$

it can be shown, through a rather lengthy and involved derivation, that the definition given by Eq. (3.103) leads to

$$\nabla \times \mathbf{B} = \hat{\mathbf{x}} \left(\frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} \right) + \hat{\mathbf{y}} \left(\frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x} \right)$$

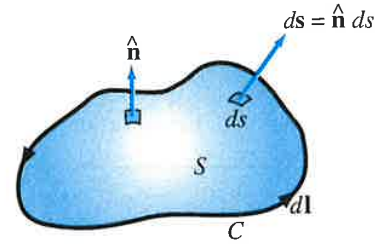


Figure 3-23: The direction of the unit vector $\hat{\mathbf{n}}$ is along the thumb when the other four fingers of the right hand follow $d\mathbf{l}$.

$$+ \hat{\mathbf{z}} \left(\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} \right) \\ = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ B_x & B_y & B_z \end{vmatrix} \quad (3.105)$$

Expressions for $\nabla \times \mathbf{B}$ are given on the inside back cover of the book for the three orthogonal coordinate systems considered in this chapter.

3-6.1 Vector Identities Involving the Curl

For any two vectors $\mathbf{A}$ and $\mathbf{B}$,

$$(1) \quad \nabla \times (\mathbf{A} + \mathbf{B}) = \nabla \times \mathbf{A} + \nabla \times \mathbf{B}, \quad (3.106a)$$

$$(2) \quad \nabla \cdot (\nabla \times \mathbf{A}) = 0 \quad \text{for any vector } \mathbf{A}, \quad (3.106b)$$

$$(3) \quad \nabla \times (\nabla V) = 0 \quad \text{for any scalar function } V. \quad (3.106c)$$

Identities (2) and (3) are important properties that we will use in succeeding chapters.

3-6.2 Stokes's Theorem

Stokes's theorem converts the surface integral of the curl of a vector over an open surface S into a line integral of the vector along the contour C bounding the surface S . The geometry is shown in Fig. 3-23. Mathematically, **Stokes's theorem** is given by

$$\int_S (\nabla \times \mathbf{B}) \cdot d\mathbf{s} = \oint_C \mathbf{B} \cdot d\mathbf{l} \quad (\text{Stokes's theorem}), \quad (3.107)$$

and its derivation follows from the definition of $\nabla \times \mathbf{B}$ given by Eq. (3.103). The conversion process represented by Eq. (3.107) is used extensively in solutions of electromagnetic problems. If $\nabla \times \mathbf{B} = 0$, the field $\mathbf{B}$ is said to be **conservative** or **irrotational** because its circulation, represented by the right-hand side of Eq. (3.107), is zero.

Example 3-12 Verification of Stokes's Theorem

A vector field is given by $\mathbf{B} = \hat{\mathbf{z}} \cos \phi / r$. Verify Stokes's theorem for a segment of a cylindrical surface defined by $r = 2$, $\pi/3 \leq \phi \leq \pi/2$, and $0 \leq z \leq 3$, as depicted in Fig. 3-24.

Solution: Stokes's theorem states that

$$\int_S (\nabla \times \mathbf{B}) \cdot d\mathbf{s} = \oint_C \mathbf{B} \cdot d\mathbf{l}.$$

Left-hand side: With $\mathbf{B}$ having only a component $B_z = \cos \phi / r$, use of the expression for $\nabla \times \mathbf{B}$ in cylindrical coordinates from the inside back cover of the

book gives

$$\begin{aligned} \nabla \times \mathbf{B} &= \hat{\mathbf{r}} \left(\frac{1}{r} \frac{\partial B_z}{\partial \phi} - \frac{\partial B_\phi}{\partial z} \right) + \hat{\phi} \left(\frac{\partial B_r}{\partial z} - \frac{\partial B_z}{\partial r} \right) \\ &\quad + \hat{\mathbf{z}} \frac{1}{r} \left(\frac{\partial}{\partial r} (r B_\phi) - \frac{\partial B_r}{\partial \phi} \right) \\ &= \hat{\mathbf{r}} \frac{1}{r} \frac{\partial}{\partial \phi} \left(\frac{\cos \phi}{r} \right) - \hat{\phi} \frac{\partial}{\partial r} \left(\frac{\cos \phi}{r} \right) \\ &= -\hat{\mathbf{r}} \frac{\sin \phi}{r^2} + \hat{\phi} \frac{\cos \phi}{r^2}. \end{aligned}$$

The integral of $\nabla \times \mathbf{B}$ over the specified surface S is

$$\begin{aligned} \int_S (\nabla \times \mathbf{B}) \cdot d\mathbf{s} &= \int_{z=0}^3 \int_{\phi=\pi/3}^{\pi/2} \left(-\hat{\mathbf{r}} \frac{\sin \phi}{r^2} + \hat{\phi} \frac{\cos \phi}{r^2} \right) \cdot \hat{\mathbf{r}} r d\phi dz \\ &= \int_0^3 \int_{\pi/3}^{\pi/2} -\frac{\sin \phi}{r} d\phi dz = -\frac{3}{2r} = -\frac{3}{4}, \end{aligned}$$

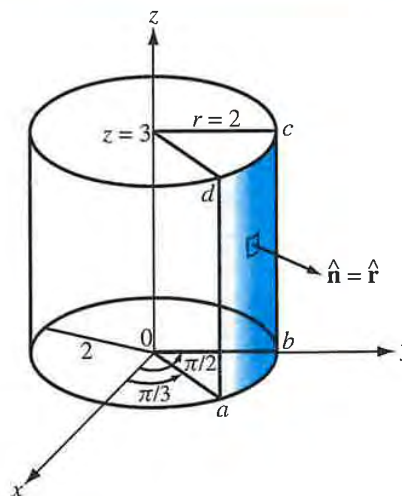


Figure 3-24: Geometry for Example 3-12.

where we used the fact that $r = 2$.

Right-hand side: The surface S is bounded by contour $C = abcd$ in Fig. 3-24. The direction of C is chosen so that it is compatible with the surface normal $\hat{\mathbf{r}}$ by the right-hand rule. Hence,

$$\oint_C \mathbf{B} \cdot d\mathbf{l} = \int_a^b \mathbf{B}_{ab} \cdot d\mathbf{l} + \int_b^c \mathbf{B}_{bc} \cdot d\mathbf{l} + \int_c^d \mathbf{B}_{cd} \cdot d\mathbf{l} + \int_d^a \mathbf{B}_{da} \cdot d\mathbf{l},$$

where $\mathbf{B}_{ab}$, $\mathbf{B}_{bc}$, $\mathbf{B}_{cd}$, and $\mathbf{B}_{da}$ are the expressions for the field $\mathbf{B}$ evaluated for segments ab , bc , cd , and da , respectively. Over segment ab , the dot product of $\mathbf{B}_{ab} = \hat{\mathbf{z}}(\cos \phi)/2$ and $d\mathbf{l} = \hat{\phi}r d\phi$ is zero, and the same is true for segment cd . Over segment bc , $\phi = \pi/2$; hence, $\mathbf{B}_{bc} = \hat{\mathbf{z}}(\cos \pi/2)/2 = 0$. For the last segment, $\mathbf{B}_{da} = \hat{\mathbf{z}}(\cos \pi/3)/2 = \hat{\mathbf{z}}/4$ and $d\mathbf{l} = \hat{\mathbf{z}} dz$. Hence,

$$\oint_C \mathbf{B} \cdot d\mathbf{l} = \int_d^a \left(\hat{\mathbf{z}} \frac{1}{4} \right) \cdot \hat{\mathbf{z}} dz = \int_3^0 \frac{1}{4} dz = -\frac{3}{4},$$

which is the same result obtained by evaluating the left-hand side of Stokes's equation. ■

EXERCISE 3.13 Find $\nabla \times \mathbf{A}$ at $(2, 0, 3)$ in cylindrical coordinates for the vector field

$$\mathbf{A} = \hat{\mathbf{r}}10e^{-2r} \cos \phi + \hat{\mathbf{z}}10 \sin \phi.$$

Ans.

$$\nabla \times \mathbf{A} = \left(\hat{\mathbf{r}} \frac{10 \cos \phi}{r} + \frac{\hat{\mathbf{z}} 10 e^{-2r}}{r} \sin \phi \right) \bigg|_{(2,0,3)} = \hat{\mathbf{r}}5.$$

EXERCISE 3.14 Find $\nabla \times \mathbf{A}$ at $(3, \pi/6, 0)$ in spherical coordinates for the vector field $\mathbf{A} = \hat{\theta}12 \sin \theta$.

Ans.

$$\nabla \times \mathbf{A} = \hat{\phi} \frac{12 \sin \theta}{R} \bigg|_{(3,\pi/6,0)} = \hat{\phi}2.$$

3-7 Laplacian Operator

In later chapters, we will sometimes deal with problems involving multiple combinations of operations on scalars and vectors. A frequently encountered combination is the divergence of the gradient of a scalar. For a scalar function V defined in Cartesian coordinates, its gradient is

$$\begin{aligned} \nabla V &= \hat{\mathbf{x}} \frac{\partial V}{\partial x} + \hat{\mathbf{y}} \frac{\partial V}{\partial y} + \hat{\mathbf{z}} \frac{\partial V}{\partial z} \\ &= \hat{\mathbf{x}} A_x + \hat{\mathbf{y}} A_y + \hat{\mathbf{z}} A_z = \mathbf{A}, \end{aligned} \quad (3.108)$$

where we defined a vector $\mathbf{A}$ with components $A_x = \partial V / \partial x$, $A_y = \partial V / \partial y$, and $A_z = \partial V / \partial z$. The divergence of ∇V is

$$\begin{aligned} \nabla \cdot (\nabla V) &= \nabla \cdot \mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \\ &= \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}. \end{aligned} \quad (3.109)$$

For convenience, $\nabla \cdot (\nabla V)$ is called the **Laplacian** of V and is denoted by $\nabla^2 V$ (the symbol ∇^2 is pronounced "del square"). That is,

$$\nabla^2 V \triangleq \nabla \cdot (\nabla V) = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}. \quad (3.110)$$

As we can see from Eq. (3.110), the Laplacian of a scalar function is a scalar. Expressions for $\nabla^2 V$ in cylindrical and spherical coordinates are given on the inside back cover of the book.

The Laplacian of a scalar can be used to define the Laplacian of a vector. For a vector $\mathbf{E}$ given in Cartesian coordinates by

$$\mathbf{E} = \hat{\mathbf{x}} E_x + \hat{\mathbf{y}} E_y + \hat{\mathbf{z}} E_z, \quad (3.111)$$

the Laplacian of $\mathbf{E}$ is defined as

$$\begin{aligned}\nabla^2 \mathbf{E} &= \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \mathbf{E} \\ &= \hat{\mathbf{x}} \nabla^2 E_x + \hat{\mathbf{y}} \nabla^2 E_y + \hat{\mathbf{z}} \nabla^2 E_z. \quad (3.112)\end{aligned}$$

Thus, in Cartesian coordinates the Laplacian of a vector is a vector whose components are equal to the Laplacians of the vector components. Through direct substitution, it can be shown that

$$\nabla^2 \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \nabla \times (\nabla \times \mathbf{E}). \quad (3.113)$$

This identity will prove useful in succeeding chapters.

REVIEW QUESTIONS

Q3.11 What do the magnitude and direction of the gradient of a scalar quantity represent?

Q3.12 Prove the validity of Eq. (3.84c) in Cartesian coordinates.

Q3.13 What is the physical meaning of the divergence of a vector field?

Q3.14 If a vector field is solenoidal at a given point in space, does it necessarily follow that the vector field is zero at that point? Explain.

Q3.15 What is the meaning of the transformation provided by the divergence theorem?

Q3.16 How is the curl of a vector field at a point related to the circulation of the vector field?

Q3.17 What is the meaning of the transformation provided by Stokes's theorem?

Q3.18 When is a vector field "conservative"?

CHAPTER HIGHLIGHTS

- Vector algebra governs the laws of addition, subtraction, and multiplication of vectors, and vector calculus encompasses the laws of differentiation and integration of vectors.
- In a right-handed orthogonal coordinate system, the three base vectors are mutually perpendicular to each other at any point in space, and the cyclic relations governing the cross products of the base vectors obey the right-hand rule.
- The dot product of two vectors produces a scalar, whereas the cross product of two vectors produces another vector.
- A vector expressed in a given coordinate system can be expressed in another coordinate system through the use of transformation relations linking the two coordinate systems.
- The fundamental differential functions in vector calculus are the gradient, the divergence, and the curl.
- The gradient of a scalar function is a vector whose magnitude is equal to the maximum rate of increasing change of the scalar function per unit distance, and its direction is along the direction of maximum increase.
- The divergence of a vector field is a measure of the net outward flux per unit volume through a closed surface surrounding the unit volume.
- The divergence theorem transforms the volume integral of the divergence of a vector field into a surface integral of the field's flux through a closed surface surrounding the volume.
- The curl of a vector field is a measure of the circulation of the vector field per unit area Δs , with the orientation of Δs chosen such that the circulation is maximum.
- Stokes's theorem transforms the surface integral of the curl of a vector field into a line integral of the field over a contour that bounds the surface.

- The Laplacian of a scalar function is defined as the divergence of the gradient of that function.

GLOSSARY OF IMPORTANT TERMS

Provide definitions or explain the meaning of the following terms:

scalar quantity
vector quantity
magnitude
unit vector
base vectors
position vector
distance vector
simple product
scalar (or dot) product
vector (or cross) product
orthogonal coordinate system
Cartesian coordinate system
cylindrical coordinate system
spherical coordinate system
radial distance r
azimuth angle ϕ
zenith angle θ
range R
gradient operator
directional derivative
flux lines
flux density
divergence operator
solenoidal field
divergence theorem
circulation of a vector
curl operator
conservative field
Stokes's theorem
Laplacian operator

PROBLEMS

Section 3-1: Vector Algebra

3.1* Vector $\mathbf{A}$ starts at point $(1, -1, -2)$ and ends at point $(2, -1, 0)$. Find a unit vector in the direction of $\mathbf{A}$.

3.2 Given vectors $\mathbf{A} = \hat{x}2 - \hat{y}3 + \hat{z}$, $\mathbf{B} = \hat{x}2 - \hat{y} + \hat{z}3$, and $\mathbf{C} = \hat{x}4 + \hat{y}2 - \hat{z}2$, show that $\mathbf{C}$ is perpendicular to both $\mathbf{A}$ and $\mathbf{B}$.

3.3* In Cartesian coordinates, the three corners of a triangle are $P_1(0, 2, 2)$, $P_2(2, -2, 2)$, and $P_3(1, 1, -2)$. Find the area of the triangle.

3.4 Given $\mathbf{A} = \hat{x}2 - \hat{y}3 + \hat{z}1$ and $\mathbf{B} = \hat{x}B_x + \hat{y}2 + \hat{z}B_z$:

- Find B_x and B_z if $\mathbf{A}$ is parallel to $\mathbf{B}$.
- Find a relation between B_x and B_z if $\mathbf{A}$ is perpendicular to $\mathbf{B}$.

3.5* Given vectors $\mathbf{A} = \hat{x} + \hat{y}2 - \hat{z}3$, $\mathbf{B} = \hat{x}3 - \hat{y}4$, and $\mathbf{C} = \hat{y}3 - \hat{z}4$, find the following

- A and $\hat{a}$
- The component of $\mathbf{B}$ along $\mathbf{C}$
- θ_{AC}
- $\mathbf{A} \times \mathbf{C}$
- $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})$
- $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$
- $\hat{x} \times \mathbf{B}$
- $(\mathbf{A} \times \hat{y}) \cdot \hat{z}$

3.6 Given vectors $\mathbf{A} = \hat{x}2 - \hat{y} + \hat{z}3$ and $\mathbf{B} = \hat{x}3 - \hat{z}2$, find a vector $\mathbf{C}$ whose magnitude is 6 and whose direction is perpendicular to both $\mathbf{A}$ and $\mathbf{B}$.

3.7* Given $\mathbf{A} = \hat{x}(2x + 3y) - \hat{y}(2y + 3z) + \hat{z}(3x - y)$, determine a unit vector parallel to $\mathbf{A}$ at point $P(1, -1, 2)$.

3.8 By expansion in Cartesian coordinates, prove:

* Answer(s) available in Appendix D.

• Solution available in CD-ROM.

- (a) The relation for the scalar triple product given by Eq. (3.29).
 (b) The relation for the vector triple product given by Eq. (3.33).

3.9* Find an expression for the unit vector directed toward the origin from an arbitrary point on the line described by $x = 1$ and $z = 2$.

3.10 Find an expression for the unit vector directed toward the point P located on the z -axis at a height h above the x - y plane from an arbitrary point $Q(x, y, 2)$ in the plane $z = 2$.

3.11* Find a unit vector parallel to either direction of the line described by

$$2x - z = 4$$

3.12 Two lines in the x - y plane are described by the expressions:

$$\begin{array}{ll} \text{Line 1} & x + 2y = -6 \\ \text{Line 2} & 3x + 4y = 8 \end{array}$$

Use vector algebra to find the smaller angle between the lines at their intersection point.

3.13* A given line is described by

$$x + 2y = 4$$

Vector $\mathbf{A}$ starts at the origin and ends at point P on the line such that $\mathbf{A}$ is orthogonal to the line. Find an expression for $\mathbf{A}$.

3.14 Show that, given two vectors $\mathbf{A}$ and $\mathbf{B}$,

- (a) The vector $\mathbf{C}$ defined as the vector component of $\mathbf{B}$ in the direction of $\mathbf{A}$ is given by

$$\mathbf{C} = \hat{\mathbf{a}}(\mathbf{B} \cdot \hat{\mathbf{a}}) = \frac{\mathbf{A}(\mathbf{B} \cdot \mathbf{A})}{|\mathbf{A}|^2},$$

where $\hat{\mathbf{a}}$ is the unit vector of $\mathbf{A}$.

- (b) The vector $\mathbf{D}$ defined as the vector component of $\mathbf{B}$ perpendicular to $\mathbf{A}$ is given by

$$\mathbf{D} = \mathbf{B} - \frac{\mathbf{A}(\mathbf{B} \cdot \mathbf{A})}{|\mathbf{A}|^2}$$

3.15* A certain plane is described by

$$2x + 3y + 4z = 16$$

Find the unit vector normal to the surface in the direction away from the origin.

3.16 Given $\mathbf{B} = \hat{\mathbf{x}}(2z - 3y) + \hat{\mathbf{y}}(2x - 3z) - \hat{\mathbf{z}}(x + y)$, find a unit vector parallel to $\mathbf{B}$ at point $P(1, 0, -1)$.

3.17 When sketching or demonstrating the spatial variation of a vector field, we often use arrows, as in Fig. 3-25, wherein the length of the arrow is made to be

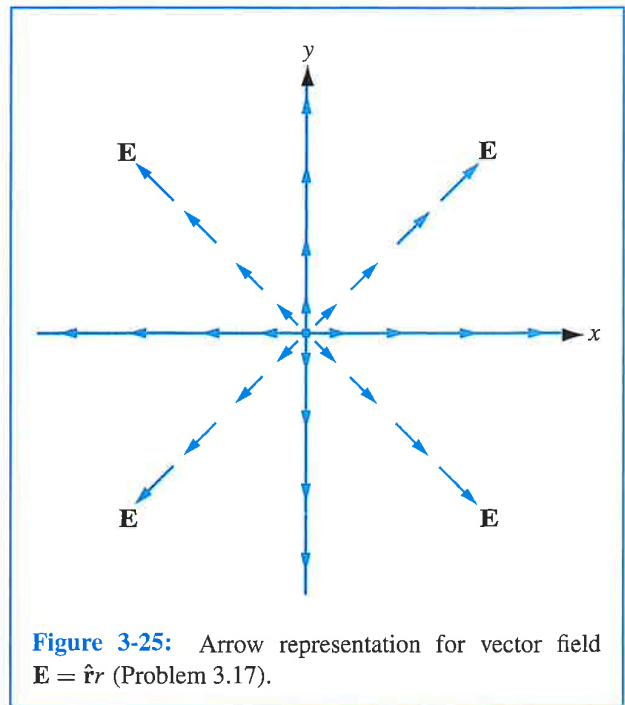


Figure 3-25: Arrow representation for vector field $\mathbf{E} = \hat{\mathbf{r}}$ (Problem 3.17).

proportional to the strength of the field and the direction of the arrow is the same as that of the field's. The sketch shown in Fig. 3-25, which represents the vector field $\mathbf{E} = \hat{\mathbf{r}}r$, consists of arrows pointing radially away from the origin and their lengths increasing linearly in proportion to their distance away from the origin. Using this arrow representation, sketch each of the following vector fields:

- (a) $\mathbf{E}_1 = -\hat{\mathbf{x}}y$
- ⊗ (b) $\mathbf{E}_2 = \hat{\mathbf{y}}x$
- (c) $\mathbf{E}_3 = \hat{\mathbf{x}}x + \hat{\mathbf{y}}y$
- (d) $\mathbf{E}_4 = \hat{\mathbf{x}}x + \hat{\mathbf{y}}2y$
- ⊗ (e) $\mathbf{E}_5 = \hat{\phi}r$
- (f) $\mathbf{E}_6 = \hat{\mathbf{r}} \sin \phi$

3.18 Use arrows to sketch each of the following vector fields:

- (a) $\mathbf{E}_1 = \hat{\mathbf{x}}x - \hat{\mathbf{y}}y$
- (b) $\mathbf{E}_2 = -\hat{\phi}$
- (c) $\mathbf{E}_3 = \hat{\mathbf{y}}(1/x)$
- (d) $\mathbf{E}_4 = \hat{\mathbf{r}} \cos \phi$

Sections 3-2 and 3-3: Coordinate Systems

3.19* Convert the coordinates of the following points from Cartesian to cylindrical and spherical coordinates:

- ⊗ (a) $P_1(1, 2, 0)$
- (b) $P_2(0, 0, 3)$
- (c) $P_3(1, 1, 2)$
- (d) $P_4(-3, 3, -3)$

3.20 Convert the coordinates of the following points from cylindrical to Cartesian coordinates:

- (a) $P_1(2, \pi/4, -3)$
- (b) $P_2(3, 0, 0)$
- (c) $P_3(4, \pi, 2)$

3.21* Convert the coordinates of the following points from spherical to cylindrical coordinates:

- (a) $P_1(5, 0, 0)$
- ⊗ (b) $P_2(5, 0, \pi)$
- (c) $P_3(3, \pi/2, \pi)$

3.22 Use the appropriate expression for the differential surface area ds to determine the area of each of the following surfaces:

- (a) $r = 3; 0 \leq \phi \leq \pi/3; -2 \leq z \leq 2$
 - (b) $2 \leq r \leq 5; \pi/2 \leq \phi \leq \pi; z = 0$
 - (c) $2 \leq r \leq 5; \phi = \pi/4; -2 \leq z \leq 2$
 - ⊗ (d) $R = 2; 0 \leq \theta \leq \pi/3; 0 \leq \phi \leq \pi$
 - (e) $0 \leq R \leq 5; \theta = \pi/3; 0 \leq \phi \leq 2\pi$
- Also sketch the outline of each surface.

3.23* Find the volumes described by the following:

- (a) $2 \leq r \leq 5; \pi/2 \leq \phi \leq \pi; 0 \leq z \leq 2$
 - (b) $0 \leq R \leq 5; 0 \leq \theta \leq \pi/3; 0 \leq \phi \leq 2\pi$
- Also sketch the outline of each volume.

3.24 A section of a sphere is described by $0 \leq R \leq 2$, $0 \leq \theta \leq 90^\circ$, and $30^\circ \leq \phi \leq 90^\circ$. Find the following:

- (a) The surface area of the spherical section.
 - (b) The enclosed volume.
- Also sketch the outline of the section.

3.25* A vector field is given in cylindrical coordinates by

$$\mathbf{E} = \hat{\mathbf{r}}r \cos \phi + \hat{\phi}r \sin \phi + \hat{\mathbf{z}}z^2$$

Point $P(4, \pi, 2)$ is located on the surface of the cylinder described by $r = 4$. At point P , find:

- (a) The vector component of $\mathbf{E}$ perpendicular to the cylinder.
- (b) The vector component of $\mathbf{E}$ tangential to the cylinder.

3.26 At a given point in space, vectors **A** and **B** are given in spherical coordinates by

$$\mathbf{A} = \hat{\mathbf{R}}4 + \hat{\boldsymbol{\theta}}2 - \hat{\boldsymbol{\phi}}$$

$$\mathbf{B} = -\hat{\mathbf{R}}2 + \hat{\boldsymbol{\phi}}3$$

Find:

- (a) The scalar component, or projection, of **B** in the direction of **A**.
- (b) The vector component of **B** in the direction of **A**.
- (c) The vector component of **B** perpendicular to **A**.

3.27* Given vectors

$$\mathbf{A} = \hat{\mathbf{r}}(\cos \phi + 3z) - \hat{\boldsymbol{\phi}}(2r + 4 \sin \phi) + \hat{\mathbf{z}}(r - 2z)$$

$$\mathbf{B} = -\hat{\mathbf{r}} \sin \phi + \hat{\mathbf{z}} \cos \phi$$

find

- (a) θ_{AB} at $(2, \pi/2, 0)$
- (b) A unit vector perpendicular to both **A** and **B** at $(2, \pi/3, 1)$

3.28 Find the distance between the following pairs of points:

- (a) $P_1(1, 2, 3)$ and $P_2(-2, -3, 2)$ in Cartesian coordinates
- (b) $P_3(1, \pi/4, 2)$ and $P_4(3, \pi/4, 4)$ in cylindrical coordinates
- (c) $P_5(2, \pi/2, 0)$ and $P_6(3, \pi, 0)$ in spherical coordinates

3.29* Determine the distance between the following pairs of points:

- (a) $P_1(1, 1, 2)$ and $P_2(0, 2, 2)$
- (b) $P_3(2, \pi/3, 1)$ and $P_4(4, \pi/2, 0)$
- (c) $P_5(3, \pi, \pi/2)$ and $P_6(4, \pi/2, \pi)$

3.30 Transform the following vectors into cylindrical coordinates and then evaluate them at the indicated points:

(a) $\mathbf{A} = \hat{\mathbf{x}}(x + y)$ at $P_1(1, 2, 3)$

(b) $\mathbf{B} = \hat{\mathbf{x}}(y - x) + \hat{\mathbf{y}}(x - y)$ at $P_2(1, 0, 2)$

(c) $\mathbf{C} = \hat{\mathbf{x}}y^2/(x^2 + y^2) - \hat{\mathbf{y}}x^2/(x^2 + y^2) + \hat{\mathbf{z}}4$ at $P_3(1, -1, 2)$

(d) $\mathbf{D} = \hat{\mathbf{R}} \sin \theta + \hat{\boldsymbol{\theta}} \cos \theta + \hat{\boldsymbol{\phi}} \cos^2 \phi$ at $P_4(2, \pi/2, \pi/4)$

(e) $\mathbf{E} = \hat{\mathbf{R}} \cos \phi + \hat{\boldsymbol{\theta}} \sin \phi + \hat{\boldsymbol{\phi}} \sin^2 \theta$ at $P_5(3, \pi/2, \pi)$

3.31* Transform the following vectors into spherical coordinates and then evaluate them at the indicated points:

(a) $\mathbf{A} = \hat{\mathbf{x}}y^2 + \hat{\mathbf{y}}xz + \hat{\mathbf{z}}4$ at $P_1(1, -1, 2)$

(b) $\mathbf{B} = \hat{\mathbf{y}}(x^2 + y^2 + z^2) - \hat{\mathbf{z}}(x^2 + y^2)$ at $P_2(-1, 0, 2)$

(c) $\mathbf{C} = \hat{\mathbf{r}} \cos \phi - \hat{\boldsymbol{\phi}} \sin \phi + \hat{\mathbf{z}} \cos \phi \sin \phi$ at $P_3(2, \pi/4, 2)$

(d) $\mathbf{D} = \hat{\mathbf{x}}y^2/(x^2 + y^2) - \hat{\mathbf{y}}x^2/(x^2 + y^2) + \hat{\mathbf{z}}4$ at $P_4(1, -1, 2)$

Sections 3-4 to 3-7: Gradient, Divergence, and Curl Operators

3.32 Find the gradient of the following scalar functions:

(a) $T = 2/(x^2 + z^2)$

(b) $V = xy^2z^3$

(c) $U = z \cos \phi / (1 + r^2)$

(d) $W = e^{-R} \sin \theta$

(e) $S = x^2 e^{-z} + y^2$

(f) $N = r^2 \cos \phi$

(g) $M = R \cos \theta \sin \phi$

3.33* The gradient of a scalar function T is given by

$$\nabla T = \hat{\mathbf{z}}e^{-2z}$$

If $T = 10$ at $z = 0$, find $T(z)$.

3.34 Follow a procedure similar to that leading to Eq. (3.82) to derive the expression given by Eq. (3.83) for ∇ in spherical coordinates.

3.35* For the scalar function $V = xy - z^2$, determine its directional derivative along the direction of vector $\mathbf{A} = (\hat{x} - \hat{y}z)$ and then evaluate it at $P(1, -1, 2)$.

3.36 For the scalar function $T = e^{-r/5} \cos \phi$, determine its directional derivative along the radial direction $\hat{r}$ and then evaluate it at $P(2, \pi/4, 3)$.

3.37* For the scalar function $U = \frac{1}{R} \sin^2 \theta$, determine its directional derivative along the range direction $\hat{R}$ and then evaluate it at $P(4, \pi/4, \pi/2)$.

3.38 Vector field $\mathbf{E}$ is characterized by the following properties: (a) $\mathbf{E}$ points along $\hat{R}$; (b) the magnitude of $\mathbf{E}$ is a function of only the distance from the origin; (c) $\mathbf{E}$ vanishes at the origin, and (d) $\nabla \cdot \mathbf{E} = 6$, everywhere. Find an expression for $\mathbf{E}$ that satisfies these properties.

3.39* For the vector field $\mathbf{E} = \hat{x}xz - \hat{y}yz^2 - \hat{z}xy$, verify the divergence theorem by computing

- (a) The total outward flux flowing through the surface of a cube centered at the origin and with sides equal to 2 units each and parallel to the Cartesian axes.
- (b) The integral of $\nabla \cdot \mathbf{E}$ over the cube's volume.

3.40 For the vector field $\mathbf{E} = \hat{r}10e^{-r} - \hat{z}3z$, verify the divergence theorem for the cylindrical region enclosed by $r = 2$, $z = 0$, and $z = 4$.

3.41* A vector field $\mathbf{D} = \hat{r}r^3$ exists in the region between two concentric cylindrical surfaces defined by $r = 1$ and $r = 2$, with both cylinders extending between $z = 0$ and $z = 5$. Verify the divergence theorem by evaluating the following

- (a) $\oint_S \mathbf{D} \cdot d\mathbf{s}$
- (b) $\int_V \nabla \cdot \mathbf{D} dv$

3.42 For the vector field $\mathbf{D} = \hat{R}3R^2$, evaluate both sides of the divergence theorem for the region enclosed between the spherical shells defined by $R = 1$ and $R = 2$.

3.43* For the vector field $\mathbf{E} = \hat{x}xy - \hat{y}(x^2 + 2y^2)$, calculate the following:

- (a) $\oint_C \mathbf{E} \cdot d\mathbf{l}$ around the triangular contour shown in Fig. 3-26(a).
- (b) $\int_S (\nabla \times \mathbf{E}) \cdot d\mathbf{s}$ over the area of the triangle.

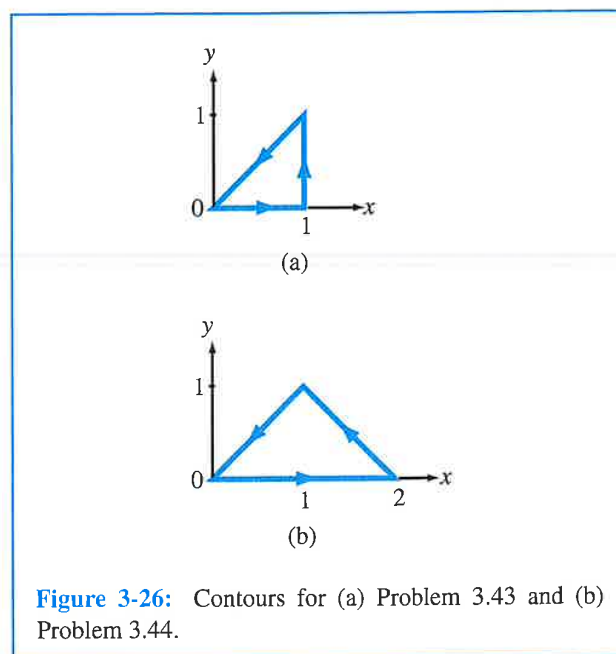


Figure 3-26: Contours for (a) Problem 3.43 and (b) Problem 3.44.

3.44 Repeat Problem 3.43 for the contour shown in Fig. 3-26(b).

3.45* Verify Stokes's theorem for the vector field

$$\mathbf{B} = (\hat{r}r \cos \phi + \hat{\phi} \sin \phi)$$

by evaluating the following:

- (a) $\oint_C \mathbf{B} \cdot d\mathbf{l}$ over the semicircular contour shown in Fig. 3-27(a).
- (b) $\int_S (\nabla \times \mathbf{B}) \cdot d\mathbf{s}$ over the surface of the semicircle.

3.46 Repeat Problem 3.45 for the contour shown in Fig. 3-27(b).

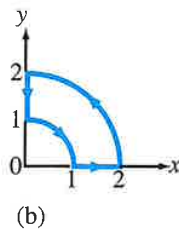
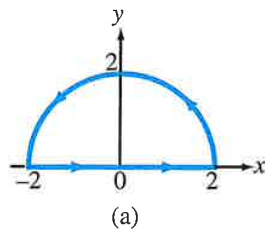


Figure 3-27: Contour paths for (a) Problem 3.45 and (b) Problem 3.46.

3.47 Verify Stokes's theorem for the vector field $\mathbf{A} = \hat{\mathbf{R}} \cos \theta + \hat{\boldsymbol{\phi}} \sin \theta$ by evaluating it on the hemisphere of unit radius.

3.48* Determine if each of the following vector fields is solenoidal, conservative, or both:

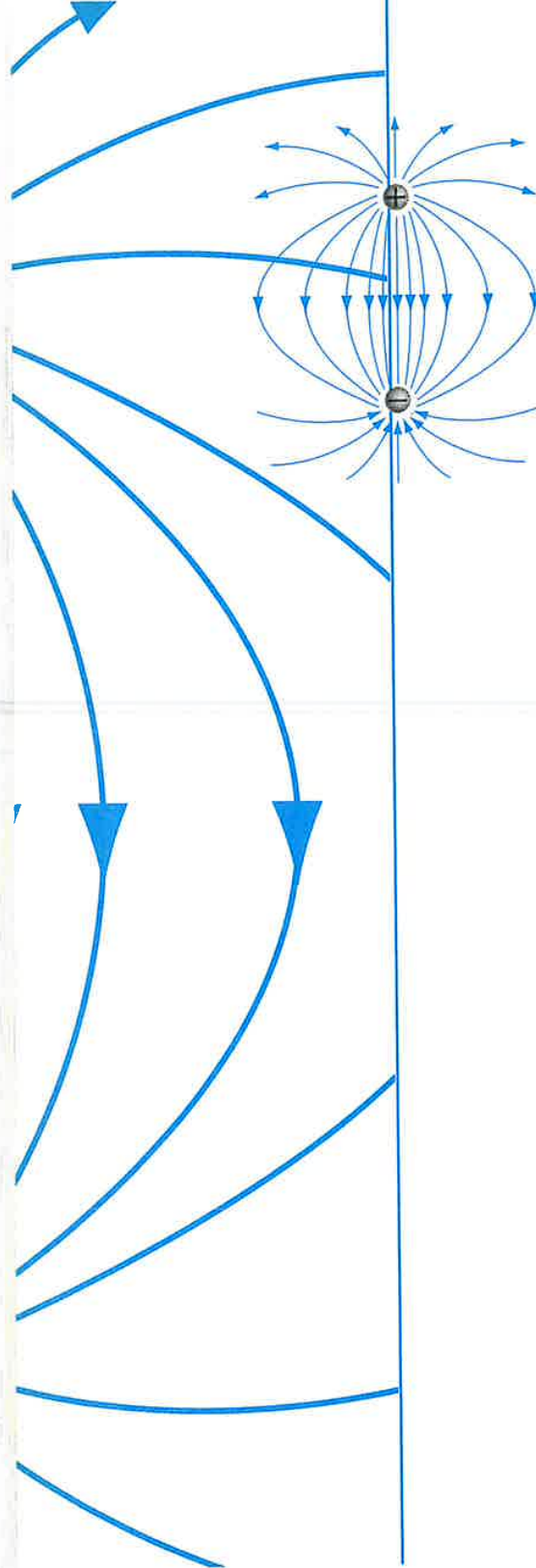
- (a) $\mathbf{A} = \hat{\mathbf{x}}2xy - \hat{\mathbf{y}}y^2$
- (b) $\mathbf{B} = \hat{\mathbf{x}}x^2 - \hat{\mathbf{y}}y^2 + \hat{\mathbf{z}}2z$
- (c) $\mathbf{C} = \hat{\mathbf{r}}(\sin \phi)/r^2 + \hat{\boldsymbol{\phi}}(\cos \phi)/r^2$
- (d) $\mathbf{D} = \hat{\mathbf{R}}/R$
- (e) $\mathbf{E} = \hat{\mathbf{r}}\left(3 - \frac{r}{1+r}\right) + \hat{\mathbf{z}}z$
- (f) $\mathbf{F} = (\hat{\mathbf{x}}y - \hat{\mathbf{y}}x)/(x^2 + y^2)$
- (g) $\mathbf{G} = \hat{\mathbf{x}}(x^2 + z^2) + \hat{\mathbf{y}}(y^2 + x^2) + \hat{\mathbf{z}}(y^2 + z^2)$
- (h) $\mathbf{H} = \hat{\mathbf{R}}(Re^{-R})$

3.49 Find the Laplacian of the following scalar functions:

- (a) $V = xy^2z^3$
- (b) $V = xy + yz + zx$
- (c) $V = 1/(x^2 + y^2)$
- (d) $V = 5e^{-r} \cos \phi$
- (e) $V = 10e^{-R} \sin \theta$

CHAPTER 4

Electrostatics

- 
- The diagram on the left side of the page illustrates electrostatic concepts. It features two positive point charges at the top, with blue field lines radiating outwards. Below them, a vertical line represents a distribution of charges, with field lines curving around it. At the bottom, two large blue arrows point downwards, indicating the direction of the electric field in that region.
- 4-1 Maxwell's Equations
 - 4-2 Charge and Current Distributions
 - 4-3 Coulomb's Law
 - 4-4 Gauss's Law
 - 4-5 Electric Scalar Potential
 - 4-6 Electrical Properties of Materials
 - 4-7 Conductors
 - 4-8 Dielectrics
 - 4-9 Electric Boundary Conditions
 - 4-10 Capacitance
 - 4-11 Electrostatic Potential Energy
 - 4-12 Image Method

4-1 Maxwell's Equations

Modern electromagnetism is based on a set of four fundamental relations known as *Maxwell's equations*:

$$\nabla \cdot \mathbf{D} = \rho_v, \quad (4.1a)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (4.1b)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (4.1c)$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \quad (4.1d)$$

where $\mathbf{E}$ and $\mathbf{D}$ are electric field quantities interrelated by $\mathbf{D} = \epsilon \mathbf{E}$, with ϵ being the electrical permittivity of the material; $\mathbf{B}$ and $\mathbf{H}$ are magnetic field quantities interrelated by $\mathbf{B} = \mu \mathbf{H}$, with μ being the magnetic permeability of the material; ρ_v is the electric charge density per unit volume; and $\mathbf{J}$ is the current density per unit area. The field quantities $\mathbf{E}$, $\mathbf{D}$, $\mathbf{B}$, and $\mathbf{H}$ were introduced in Section 1-2, and ρ_v and $\mathbf{J}$ will be defined in Section 4-2. These equations hold in any material, including free space (vacuum), and at any spatial location (x, y, z) . In general, all the quantities in Maxwell's equations may be a function of time t . By his formulation of these equations, published in a classic treatise in 1873, James Clerk Maxwell established the first unified theory of electricity and magnetism. His equations, which he deduced from experimental observations reported by Gauss, Ampère, Faraday, and others, not only encapsulate the connection between the electric field and electric charge and between the magnetic field and electric current, but they also define

the bilateral coupling between the electric and magnetic field quantities. Together with some auxiliary relations, Maxwell's equations form the fundamental tenets of electromagnetic theory.

In the *static* case, none of the quantities appearing in Maxwell's equations are a function of time (i.e., $\partial/\partial t = 0$). *This happens when all charges are permanently fixed in space, or, if they move, they do so at a steady rate so that ρ_v and $\mathbf{J}$ are constant in time.* Under these circumstances, the time derivatives of $\mathbf{B}$ and $\mathbf{D}$ in Eqs. (4.1b) and (4.1d) are zero, and Maxwell's equations reduce to

Electrostatics

$$\nabla \cdot \mathbf{D} = \rho_v, \quad (4.2a)$$

$$\nabla \times \mathbf{E} = 0. \quad (4.2b)$$

Magnetostatics

$$\nabla \cdot \mathbf{B} = 0, \quad (4.3a)$$

$$\nabla \times \mathbf{H} = \mathbf{J}. \quad (4.3b)$$

Maxwell's four equations separate into two uncoupled pairs, with the first pair involving only the electric field quantities $\mathbf{E}$ and $\mathbf{D}$ and the second pair involving only the magnetic field quantities $\mathbf{B}$ and $\mathbf{H}$. *The electric and magnetic fields are no longer interconnected in the static case.* This allows us to study electricity and magnetism as two distinct and separate phenomena, as long as the spatial distributions of charge and current flow remain constant in time. We refer to the study of electric and magnetic phenomena under

static conditions as *electrostatics* and *magnetostatics*, respectively. Electrostatics is the subject of the present chapter, and in Chapter 5 we learn about magnetostatics. The experience gained through handling the relatively simpler situations in electrostatics and magnetostatics will prove valuable in tackling the more involved material in subsequent chapters, which deals with time-varying fields, charge densities, and currents.

We study electrostatics not only as a prelude to the study of time-varying fields, but also because it is an important field of study in its own right. Many electronic devices and systems are based on the principles of electrostatics. They include x-ray machines, oscilloscopes, ink-jet electrostatic printers, liquid crystal displays, copying machines, capacitance keyboards, and many solid-state control devices. Electrostatics is also used in the design of medical diagnostic sensors, such as the electrocardiogram (for recording the heart's pumping pattern) and the electroencephalogram (for recording brain activity), as well as in numerous industrial applications.

4-2 Charge and Current Distributions

In electromagnetics, we encounter various forms of electric charge distributions, and if the charges are in motion, they constitute current distributions. Charge may be distributed over a volume of space, across a surface, or along a line.

4-2.1 Charge Densities

At the atomic scale, the charge distribution in a material is discrete, meaning that charge exists only where electrons and nuclei are and nowhere else. In electromagnetics, we usually are interested in studying phenomena at a much larger scale, typically three or more orders of magnitudes greater than the spacing between adjacent atoms. At such a macroscopic scale, we can disregard the discontinuous nature of the charge distribution and treat

the net charge contained in an elemental volume Δv as if it were uniformly distributed within it. Accordingly, we define the *volume charge density* ρ_v as

$$\rho_v = \lim_{\Delta v \rightarrow 0} \frac{\Delta q}{\Delta v} = \frac{dq}{dv} \quad (\text{C/m}^3), \quad (4.4)$$

where Δq is the charge contained in Δv . In general, ρ_v is defined at a given point in space, specified by (x, y, z) in a Cartesian coordinate system, and at a given time t . Thus, $\rho_v = \rho_v(x, y, z, t)$. Physically, ρ_v represents the average charge per unit volume for a volume Δv centered at (x, y, z) , with Δv being large enough to contain a large number of atoms and yet small enough to be regarded as a point at the macroscopic scale under consideration. The variation of ρ_v with spatial location is called its *spatial distribution*, or simply its *distribution*. The total charge contained in a given volume v is given by

$$Q = \int_v \rho_v dv \quad (\text{C}). \quad (4.5)$$

In some cases, particularly when dealing with conductors, electric charge may be distributed across the surface of a material, in which case the relevant quantity of interest is the *surface charge density* ρ_s , defined as

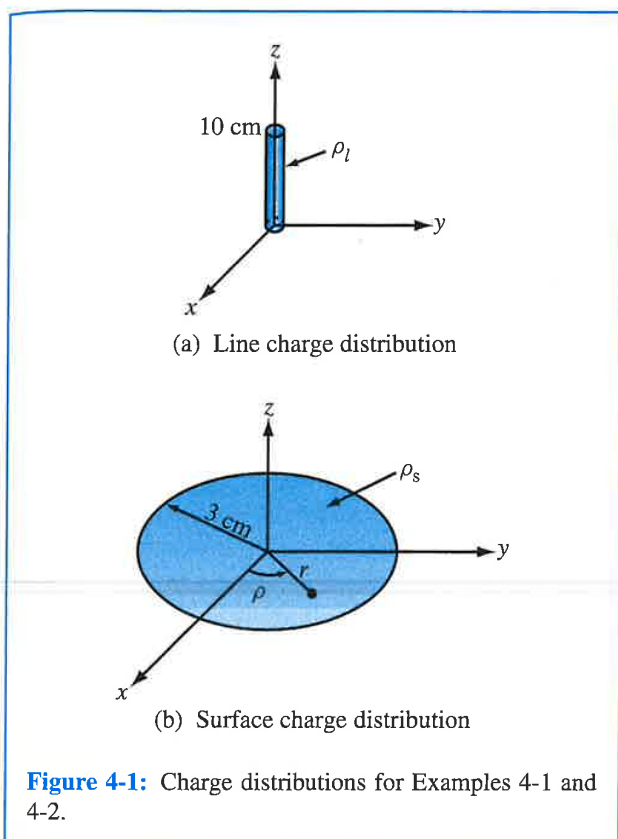
$$\rho_s = \lim_{\Delta s \rightarrow 0} \frac{\Delta q}{\Delta s} = \frac{dq}{ds} \quad (\text{C/m}^2), \quad (4.6)$$

where Δq is the charge present across an elemental surface area Δs . Similarly, if the charge is distributed along a line, which need not be straight, we characterize the distribution in terms of the *line charge density* ρ_l , defined as

$$\rho_l = \lim_{\Delta l \rightarrow 0} \frac{\Delta q}{\Delta l} = \frac{dq}{dl} \quad (\text{C/m}). \quad (4.7)$$

Example 4-1 Line Charge Distribution

Calculate the total charge Q contained in a cylindrical tube of charge oriented along the z -axis as shown in Fig. 4-1(a). The line charge density is $\rho_l = 2z$, where z



is the distance in meters from the bottom end of the tube. The tube length is 10 cm.

Solution: The total charge Q is

$$Q = \int_0^{0.1} \rho_l dz = \int_0^{0.1} 2z dz = z^2 \Big|_0^{0.1} = 10^{-2} \text{ C.} \quad \blacksquare$$

Example 4-2 Surface Charge Distribution

The circular disk of electric charge shown in Fig. 4-1(b) is characterized by an azimuthally symmetric surface charge density that increases linearly with r from zero at the center to 6 C/m^2 at $r = 3 \text{ cm}$. Find the total charge present on the disk surface.

Solution: Since ρ_s is symmetrical with respect to the azimuth angle ϕ , its functional form depends only on r and is given by

$$\rho_s = \frac{6r}{3 \times 10^{-2}} = 2 \times 10^2 r \quad (\text{C/m}^2),$$

where r is in meters. In polar coordinates, an element of area is $ds = r dr d\phi$, and for the disk shown in Fig. 4-1(b), the limits of integration are from 0 to 2π (rad) for ϕ and from 0 to $3 \times 10^{-2} \text{ m}$ for r . Hence,

$$\begin{aligned} Q &= \int_S \rho_s ds \\ &= \int_{\phi=0}^{2\pi} \int_{r=0}^{3 \times 10^{-2}} (2 \times 10^2 r) r dr d\phi \\ &= 2\pi \times 2 \times 10^2 \frac{r^3}{3} \Big|_0^{3 \times 10^{-2}} = 11.31 \text{ (nC)}. \quad \blacksquare \end{aligned}$$

EXERCISE 4.1 A square plate in the x - y plane is situated in the space defined by $-3 \text{ m} \leq x \leq 3 \text{ m}$ and $-3 \text{ m} \leq y \leq 3 \text{ m}$. Find the total charge on the plate if the surface charge density is given by $\rho_s = 4y^2 (\mu\text{C/m}^2)$.

Ans. $Q = 0.432 \text{ (nC)}$.

EXERCISE 4.2 A spherical shell centered at the origin extends between $R = 2 \text{ cm}$ and $R = 3 \text{ cm}$. If the volume charge density is given by $\rho_v = 3R \times 10^{-4} (\text{C/m}^3)$, find the total charge contained in the shell.

Ans. $Q = 0.61 \text{ (nC)}$.

4-2.2 Current Density

Consider a tube of charge with volume charge density ρ_v , as shown in Fig. 4-2(a). The charges are moving with a mean velocity $\mathbf{u}$ along the axis of the tube. Over a period Δt , the charges move a distance $\Delta l = u \Delta t$. The amount of charge that crosses the tube's cross-sectional surface $\Delta s'$ in time Δt is therefore

$$\Delta q' = \rho_v \Delta v = \rho_v \Delta l \Delta s' = \rho_v u \Delta s' \Delta t. \quad (4.8)$$

Now consider the more general case where the charges are flowing through a surface Δs whose surface normal $\hat{\mathbf{n}}$ is not necessarily parallel to $\mathbf{u}$, as shown in Fig. 4-2(b). In this case, the amount of charge Δq flowing through Δs is

$$\Delta q = \rho_v \mathbf{u} \cdot \Delta \mathbf{s} \Delta t, \quad (4.9)$$

and the corresponding current is

$$\Delta I = \frac{\Delta q}{\Delta t} = \rho_v \mathbf{u} \cdot \Delta \mathbf{s} = \mathbf{J} \cdot \Delta \mathbf{s}, \quad (4.10)$$

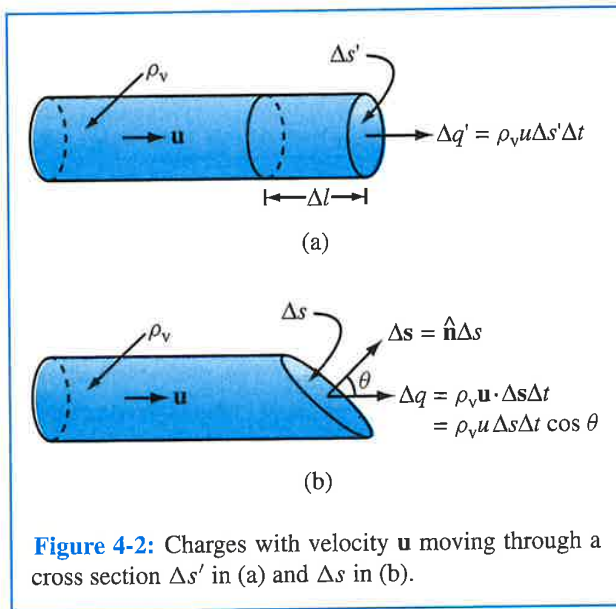


Figure 4-2: Charges with velocity $\mathbf{u}$ moving through a cross section $\Delta s'$ in (a) and Δs in (b).

where

$$\mathbf{J} = \rho_v \mathbf{u} \quad (\text{A/m}^2) \quad (4.11)$$

is defined as the **current density** in ampere per square meter. For an arbitrary surface S , the total current flowing through it is then given by

$$I = \int_S \mathbf{J} \cdot d\mathbf{s} \quad (\text{A}). \quad (4.12)$$

When the current is generated by the actual movement of electrically charged matter, it is called a **convection current**, and $\mathbf{J}$ is called the **convection current density**. A wind-driven charged cloud, for example, gives rise to a convection current. In some cases, the charged matter constituting the convection current consists solely of charged particles, such as the electrons of an electron beam in a cathode ray tube (the picture tube of televisions and computer monitors).

This is distinct from a **conduction current**, where atoms of the conducting medium do not move. In a metal wire, for example, there are equal amounts of positive charges (in atomic nuclei) and negative charges (in the electron shells of the atom). All the positive charges and most of the negative charges cannot move; only those electrons in the outermost electronic shells of the atoms can be easily pushed from one atom to the next if a voltage is applied across the ends of the wire. This movement of electrons from atom to atom gives rise to conduction current. The electrons that emerge from the wire are not necessarily the same electrons that entered the wire at the other end.

Because the two types of current are generated by different physical mechanisms, conduction current obeys Ohm's law, whereas convection current does not. Conduction current is discussed in more detail in Section 4-7.

REVIEW QUESTIONS

Q4.1 What happens to Maxwell's equations under static conditions? Why is that significant?

Q4.2 How is the current density $\mathbf{J}$ related to the volume charge density ρ_v ?

Q4.3 What is the difference between convection current and conduction current?

4-3 Coulomb's Law

One of the major goals of this chapter is to develop expressions relating the *electric field intensity* $\mathbf{E}$ and the associated *electric flux density* $\mathbf{D}$ to any specified distribution of charge. Our discussions, however, will be limited to electrostatic fields induced by static charge distributions.

We begin by reviewing how the electric field was introduced and defined in Section 1-2.2 on the basis of the results of Coulomb's experiments on the electrical force between charged bodies. *Coulomb's law*, which was first introduced for electrical charges in air and later generalized to material media, states that

- (1) an isolated charge q induces an electric field $\mathbf{E}$ at every point in space, and at any specific point P , $\mathbf{E}$ is given by

$$\mathbf{E} = \hat{\mathbf{R}} \frac{q}{4\pi\epsilon R^2} \quad (\text{V/m}), \quad (4.13)$$

where $\hat{\mathbf{R}}$ is a unit vector pointing from q to P (Fig. 4-3), R is the distance between them, and ϵ is the electrical permittivity of the medium containing the observation point P ; and

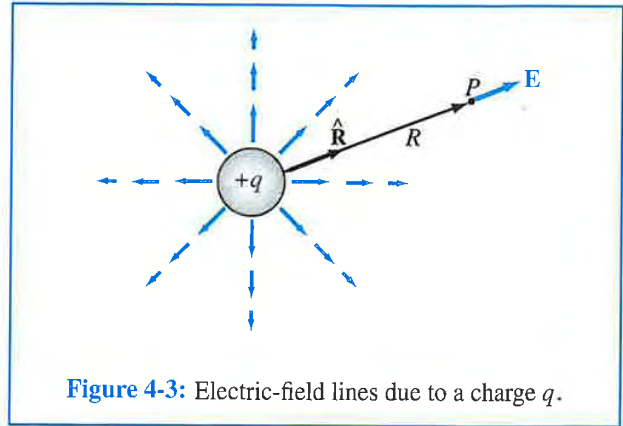


Figure 4-3: Electric-field lines due to a charge q .

- (2) in the presence of an electric field $\mathbf{E}$ at a given point in space, which may be due to a single charge or a distribution of many charges, the force acting on a test charge q' , when the charge is placed at that point, is given by

$$\mathbf{F} = q'\mathbf{E} \quad (\text{N}). \quad (4.14)$$

With $\mathbf{F}$ measured in newtons (N) and q' in coulombs (C), the unit of $\mathbf{E}$ is (N/C), which is shown later in Section 4-5 to be the same as volt per meter (V/m).

For a material with electrical permittivity ϵ , the electrical field quantities $\mathbf{D}$ and $\mathbf{E}$ are related by

$$\mathbf{D} = \epsilon\mathbf{E} \quad (4.15)$$

with

$$\epsilon = \epsilon_r\epsilon_0, \quad (4.16)$$

where

$$\epsilon_0 = 8.85 \times 10^{-12} \simeq (1/36\pi) \times 10^{-9} \quad (\text{F/m})$$

is the electrical permittivity of free space, and $\epsilon_r = \epsilon/\epsilon_0$ is called the *relative permittivity* (or *dielectric constant*) of the material. For most materials and under most conditions, ϵ of the material has a constant value independent of both the magnitude and direction of $\mathbf{E}$.

If ϵ is independent of the magnitude of $\mathbf{E}$, then the material is said to be **linear** because $\mathbf{D}$ and $\mathbf{E}$ are related linearly, and if it is independent of the direction of $\mathbf{E}$, the material is said to be **isotropic**. Materials do not usually exhibit nonlinear permittivity behavior except when the amplitude of $\mathbf{E}$ is very high (at levels approaching the dielectric breakdown conditions discussed later in Section 4-8), and anisotropy is peculiar only to certain materials with particular crystalline structures. Hence, except for materials under these very special circumstances, the quantities $\mathbf{D}$ and $\mathbf{E}$ are effectively redundant; for a material with known ϵ , knowledge of either $\mathbf{D}$ or $\mathbf{E}$ is sufficient to specify the other in that material.

4-3.1 Electric Field due to Multiple Point Charges

The expression given by Eq. (4.13) for the field $\mathbf{E}$ due to a single charge can be extended to find the field due to multiple point charges. We begin by considering two point charges, q_1 and q_2 , located at position vectors $\mathbf{R}_1$ and $\mathbf{R}_2$ from the origin of a given coordinate system, as shown in Fig. 4-4. The electric field $\mathbf{E}$ is to be evaluated at a point P with position vector $\mathbf{R}$. At P , the electric field $\mathbf{E}_1$ due to q_1 alone is given by Eq. (4.13) with R , the distance between q_1 and P , replaced with $|\mathbf{R} - \mathbf{R}_1|$ and the unit vector $\hat{\mathbf{R}}$ replaced with $(\mathbf{R} - \mathbf{R}_1)/|\mathbf{R} - \mathbf{R}_1|$. Thus,

$$\mathbf{E}_1 = \frac{q_1(\mathbf{R} - \mathbf{R}_1)}{4\pi\epsilon|\mathbf{R} - \mathbf{R}_1|^3} \quad (\text{V/m}). \quad (4.17a)$$

Similarly, the electric field due to q_2 alone is

$$\mathbf{E}_2 = \frac{q_2(\mathbf{R} - \mathbf{R}_2)}{4\pi\epsilon|\mathbf{R} - \mathbf{R}_2|^3} \quad (\text{V/m}). \quad (4.17b)$$

The electric field obeys the principle of linear superposition. Consequently, the total electric field $\mathbf{E}$ at any point

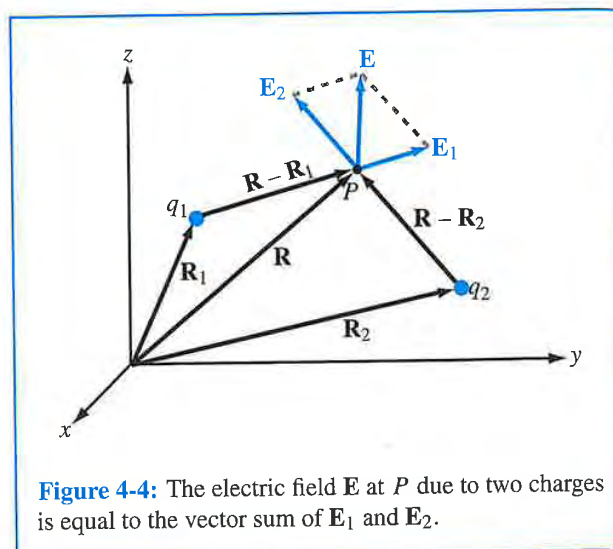


Figure 4-4: The electric field $\mathbf{E}$ at P due to two charges is equal to the vector sum of $\mathbf{E}_1$ and $\mathbf{E}_2$.

in space is equal to the vector sum of the electric fields induced by all the individual charges. In the present case,

$$\begin{aligned} \mathbf{E} &= \mathbf{E}_1 + \mathbf{E}_2 \\ &= \frac{1}{4\pi\epsilon} \left[\frac{q_1(\mathbf{R} - \mathbf{R}_1)}{|\mathbf{R} - \mathbf{R}_1|^3} + \frac{q_2(\mathbf{R} - \mathbf{R}_2)}{|\mathbf{R} - \mathbf{R}_2|^3} \right] \end{aligned} \quad (4.18)$$

Generalizing the preceding result to the case of N point charges, the electric field $\mathbf{E}$ at position vector $\mathbf{R}$ caused by charges $q_1, q_2, \dots, q_N$ located at points with position vectors $\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N$, is given by

$$\mathbf{E} = \frac{1}{4\pi\epsilon} \sum_{i=1}^N \frac{q_i(\mathbf{R} - \mathbf{R}_i)}{|\mathbf{R} - \mathbf{R}_i|^3} \quad (\text{V/m}). \quad (4.19)$$

Example 4-3 Electric Field due to Two Point Charges

Two point charges with $q_1 = 2 \times 10^{-5} \text{ C}$ and $q_2 = -4 \times 10^{-5} \text{ C}$ are located in free space at $(1, 3, -1)$ and $(-3, 1, -2)$, respectively, in a Cartesian coordinate system. Find (a) the electric field $\mathbf{E}$ at $(3, 1, -2)$ and (b)

the force on a 8×10^{-5} C charge located at that point. All distances are in meters.

Solution: (a) From Eq. (4.18), the electric field $\mathbf{E}$ with $\epsilon = \epsilon_0$ (free space) is given by

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \left[q_1 \frac{(\mathbf{R} - \mathbf{R}_1)}{|\mathbf{R} - \mathbf{R}_1|^3} + q_2 \frac{(\mathbf{R} - \mathbf{R}_2)}{|\mathbf{R} - \mathbf{R}_2|^3} \right] \quad (\text{V/m}).$$

The vectors $\mathbf{R}_1$, $\mathbf{R}_2$, and $\mathbf{R}$ are given by

$$\mathbf{R}_1 = \hat{x} + \hat{y}3 - \hat{z},$$

$$\mathbf{R}_2 = -\hat{x}3 + \hat{y} - \hat{z}2,$$

$$\mathbf{R} = \hat{x}3 + \hat{y} - \hat{z}2.$$

Hence,

$$\begin{aligned} \mathbf{E} &= \frac{1}{4\pi\epsilon_0} \left[\frac{2(\hat{x}2 - \hat{y}2 - \hat{z})}{27} - \frac{4(\hat{x}6)}{216} \right] \times 10^{-5} \\ &= \frac{\hat{x} - \hat{y}4 - \hat{z}2}{108\pi\epsilon_0} \times 10^{-5} \quad (\text{V/m}). \end{aligned}$$

(b)

$$\begin{aligned} \mathbf{F} &= q_3 \mathbf{E} = 8 \times 10^{-5} \times \frac{\hat{x} - \hat{y}4 - \hat{z}2}{108\pi\epsilon_0} \times 10^{-5} \\ &= \frac{\hat{x}2 - \hat{y}8 - \hat{z}4}{27\pi\epsilon_0} \times 10^{-10} \quad (\text{N}). \quad \blacksquare \end{aligned}$$

EXERCISE 4.3 Four charges of $10 \mu\text{C}$ each are located in free space at $(-3, 0, 0)$, $(3, 0, 0)$, $(0, -3, 0)$, and $(0, 3, 0)$ in a Cartesian coordinate system. Find the force on a $20\text{-}\mu\text{C}$ charge located at $(0, 0, 4)$. All distances are in meters.

Ans. $\mathbf{F} = \hat{z}0.23$ (N).

EXERCISE 4.4 Two identical charges are located on the x -axis at $x = 3$ and $x = 7$. At what point in space is the net electric field zero?

Ans. At point $(5, 0, 0)$.

EXERCISE 4.5 In a hydrogen atom the electron and proton are separated by an average distance of 5.3×10^{-11} m. Find the magnitude of the electrical force F_e between the two particles, and compare it with the gravitational force F_g between them.

Ans. $F_e = 8.2 \times 10^{-8}$ N, and $F_g = 3.6 \times 10^{-47}$ N.

4-3.2 Electric Field due to a Charge Distribution

We now extend the results we obtained for the field caused by discrete point charges to the case of a continuous charge distribution. Consider volume v' shown in Fig. 4-5. It contains a distribution of electric charge characterized by a volume charge density ρ_v , whose magnitude may vary with spatial location within v' . The differential electric field at a point P due to a differential amount of charge $dq = \rho_v dv'$ contained in a differential volume dv' is

$$d\mathbf{E} = \hat{\mathbf{R}}' \frac{dq}{4\pi\epsilon R'^2} = \hat{\mathbf{R}}' \frac{\rho_v dv'}{4\pi\epsilon R'^2}, \quad (4.20)$$

where $\mathbf{R}'$ is the vector from the differential volume dv' to point P . Applying the principle of linear superposition, the total electric field $\mathbf{E}$ can be obtained by integrating

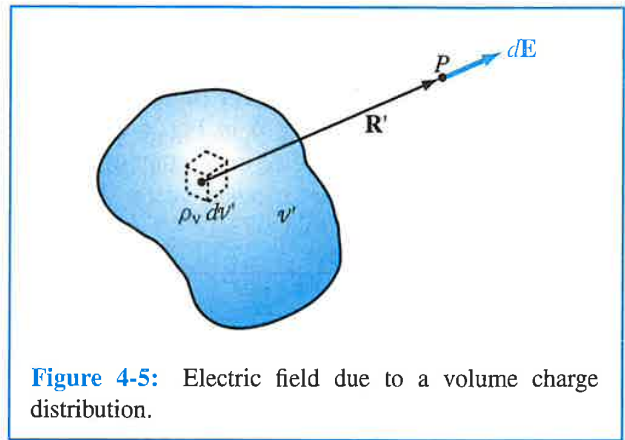


Figure 4-5: Electric field due to a volume charge distribution.

the fields contributed by all the charges making up the charge distribution. Thus,

$$\mathbf{E} = \int_{\nu'} d\mathbf{E} = \frac{1}{4\pi\epsilon} \int_{\nu'} \hat{\mathbf{R}}' \frac{\rho_v d\nu'}{R'^2} \quad (\text{volume distribution}). \quad (4.21a)$$

It is important to note that, in general, both R' and $\hat{\mathbf{R}}'$ vary as a function of position over the integration volume ν' .

If the charge is distributed across a surface S' with surface charge density ρ_s , then $dq = \rho_s ds'$, and if it is distributed along a line l' with a line charge density ρ_l , then $dq = \rho_l dl'$. Accordingly,

$$\mathbf{E} = \frac{1}{4\pi\epsilon} \int_{S'} \hat{\mathbf{R}}' \frac{\rho_s ds'}{R'^2} \quad (\text{surface distribution}), \quad (4.21b)$$

$$\mathbf{E} = \frac{1}{4\pi\epsilon} \int_{l'} \hat{\mathbf{R}}' \frac{\rho_l dl'}{R'^2} \quad (\text{line distribution}). \quad (4.21c)$$

Example 4-4 Electric Field of a Ring of Charge

A ring of charge of radius b is characterized by a uniform line charge density of positive polarity ρ_l . With the ring in free space and positioned in the x - y plane as shown in Fig. 4-6, determine the electric field intensity $\mathbf{E}$ at a point $P(0, 0, h)$ along the axis of the ring at a distance h from its center.

Solution: We start by considering the electric field generated by a differential segment of the ring, such as segment 1 located at $(b, \phi, 0)$ in Fig. 4-6(a). The segment has length $dl = b d\phi$ and contains charge $dq = \rho_l dl = \rho_l b d\phi$. The distance vector $\mathbf{R}'_1$ from segment 1 to point $P(0, 0, h)$ is

$$\mathbf{R}'_1 = -\hat{\mathbf{r}}b + \hat{\mathbf{z}}h,$$

from which we have

$$R'_1 = |\mathbf{R}'_1| = \sqrt{b^2 + h^2}, \quad \hat{\mathbf{R}}'_1 = \frac{\mathbf{R}'_1}{|\mathbf{R}'_1|} = \frac{-\hat{\mathbf{r}}b + \hat{\mathbf{z}}h}{\sqrt{b^2 + h^2}}.$$

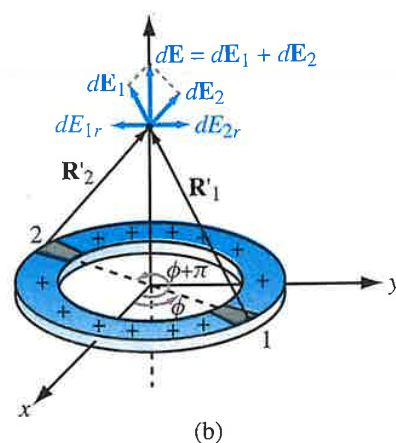
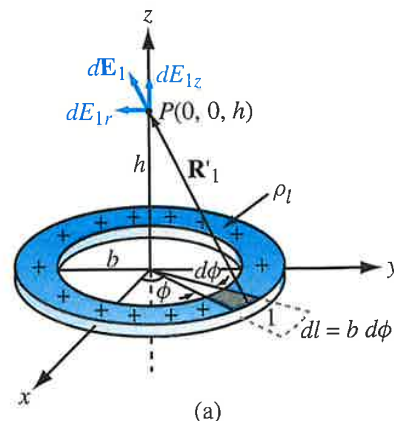


Figure 4-6: Ring of charge with line density ρ_l . (a) The field $d\mathbf{E}_1$ due to infinitesimal segment 1 and (b) the fields $d\mathbf{E}_1$ and $d\mathbf{E}_2$ due to segments at diametrically opposite locations (Example 4-4).

The electric field at $P(0, 0, h)$ due to the charge of segment 1 is

$$d\mathbf{E}_1 = \frac{1}{4\pi\epsilon_0} \hat{\mathbf{R}}'_1 \frac{\rho_l dl}{R'^2_1} = \frac{\rho_l b}{4\pi\epsilon_0} \frac{(-\hat{\mathbf{r}}b + \hat{\mathbf{z}}h)}{(b^2 + h^2)^{3/2}} d\phi.$$

The field $d\mathbf{E}_1$ has component dE_{1r} along $-\hat{\mathbf{r}}$ and component dE_{1z} along $\hat{\mathbf{z}}$. From symmetry considerations, the field $d\mathbf{E}_2$ generated by segment 2 in Fig. 4-6(b), which is located diametrically opposite the location of segment 1, is identical with $d\mathbf{E}_1$ except that the $\hat{\mathbf{r}}$ -component of $d\mathbf{E}_2$ is opposite that of $d\mathbf{E}_1$. Hence, the $\hat{\mathbf{r}}$ -components of the sum cancel and the $\hat{\mathbf{z}}$ -contributions add. The sum of the two contributions is

$$d\mathbf{E} = d\mathbf{E}_1 + d\mathbf{E}_2 = \hat{\mathbf{z}} \frac{\rho_l b h}{2\pi\epsilon_0 (b^2 + h^2)^{3/2}} d\phi \quad (4.22)$$

Since for every ring segment in the semicircle defined over the range $0 \leq \phi \leq \pi$ (the right-hand half of the circular ring) there is a corresponding segment located diametrically opposite at $(\phi + \pi)$, we can obtain the total field generated by the ring by integrating Eq. (4.22) over a semicircle as follows:

$$\begin{aligned} \mathbf{E} &= \hat{\mathbf{z}} \frac{\rho_l b h}{2\pi\epsilon_0 (b^2 + h^2)^{3/2}} \int_0^\pi d\phi \\ &= \hat{\mathbf{z}} \frac{\rho_l b h}{2\epsilon_0 (b^2 + h^2)^{3/2}} \\ &= \hat{\mathbf{z}} \frac{h}{4\pi\epsilon_0 (b^2 + h^2)^{3/2}} Q, \end{aligned} \quad (4.23)$$

where $Q = 2\pi b \rho_l$ is the total charge contained in the ring. ■

Example 4-5 Electric Field of a Circular Disk of Charge

Find the electric field at a point $P(0, 0, h)$ in free space at a height h on the z -axis due to a circular disk of charge in the x - y plane with uniform charge density ρ_s , as shown in Fig. 4-7, and then evaluate $\mathbf{E}$ for the infinite-sheet case by letting $a \rightarrow \infty$.

Solution: Building on the expression obtained in Example 4-4 for the on-axis electric field due to a circular ring of charge, we can determine the field due to the circular disk by treating the disk as a set of concentric rings. A ring of radius r and width dr has an area $ds = 2\pi r dr$ and contains charge

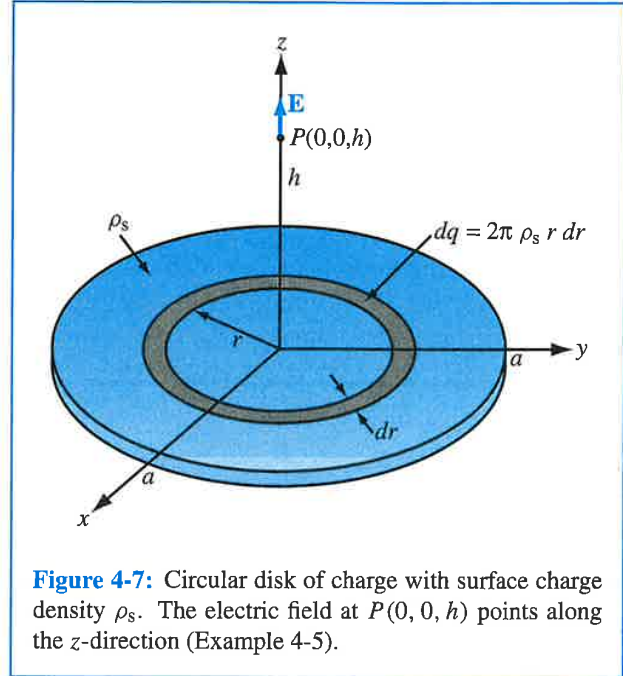


Figure 4-7: Circular disk of charge with surface charge density ρ_s . The electric field at $P(0, 0, h)$ points along the z -direction (Example 4-5).

$dq = \rho_s ds = 2\pi\rho_s r dr$. Upon using this expression in Eq. (4.23) and also replacing b with r , we obtain the following expression for the field due to the ring:

$$d\mathbf{E} = \hat{\mathbf{z}} \frac{h}{4\pi\epsilon_0 (r^2 + h^2)^{3/2}} (2\pi\rho_s r dr).$$

The total field at P is obtained by integrating the expression over the limits $r = 0$ to $r = a$:

$$\begin{aligned} \mathbf{E} &= \hat{\mathbf{z}} \frac{\rho_s h}{2\epsilon_0} \int_0^a \frac{r dr}{(r^2 + h^2)^{3/2}} \\ &= \pm \hat{\mathbf{z}} \frac{\rho_s}{2\epsilon_0} \left[1 - \frac{|h|}{\sqrt{a^2 + h^2}} \right], \end{aligned} \quad (4.24)$$

with the plus sign corresponding to when $h > 0$ and the minus sign to when $h < 0$ (below the disk).

For an infinite sheet of charge with $a = \infty$,

$$\mathbf{E} = \hat{\mathbf{z}} \frac{\rho_s}{2\epsilon_0} \quad (\text{infinite sheet of charge}). \quad (4.25)$$

We note that for an infinite sheet of charge $\mathbf{E}$ has the same value at any point above the x - y plane. For points located below the x - y plane, the unit vector $\hat{\mathbf{z}}$ in Eq. (4.25) should be replaced with $-\hat{\mathbf{z}}$. ■

REVIEW QUESTIONS

Q4.4 When we say that an electric charge induces an electric field at every point in space, does that mean that the charge “radiates” the electric field? Explain.

Q4.5 If the electric field is zero at a given point in space, does this imply the absence of electric charges?

Q4.6 State the principle of linear superposition as it applies to the electric field due to a distribution of electric charge.

EXERCISE 4.6 An infinite sheet of charge with uniform surface charge density ρ_s is located at $z = 0$ (x - y plane), and another infinite sheet with density $-\rho_s$ is located at $z = 2$ m, both in free space. Determine $\mathbf{E}$ in all regions.

Ans. $\mathbf{E} = 0$ for $z < 0$; $\mathbf{E} = \hat{\mathbf{z}}\rho_s/\epsilon_0$ for $0 < z < 2$ m; and $\mathbf{E} = 0$ for $z > 2$ m.

4-4 Gauss's Law

We now return to Eq. (4.1a):

$$\nabla \cdot \mathbf{D} = \rho_v \quad (\text{Gauss's law}), \quad (4.26)$$

which is called the differential form of **Gauss's law**. The adjective “differential” refers to the fact that the divergence operation involves spatial derivatives. As we will see shortly, Eq. (4.26) can be converted and expressed in integral form. When solving electromagnetic problems, we often convert back and forth between the differential and integral forms of equations, depending on which form happens to be the more applicable or convenient to use in each step of a solution. To convert Eq. (4.26) into integral form, we multiply both sides by dV and take the volume integral over an arbitrary volume V . Hence,

$$\int_V \nabla \cdot \mathbf{D} dV = \int_V \rho_v dV = Q, \quad (4.27)$$

where Q is the total charge enclosed in V . The divergence theorem, given by Eq. (3.98), states that the volume integral of the divergence of any vector over a volume V is equal to the total outward flux of that vector through the surface S enclosing V . Thus, for the vector $\mathbf{D}$,

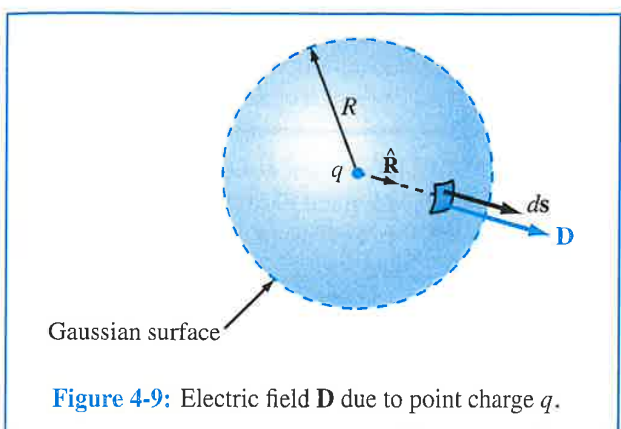
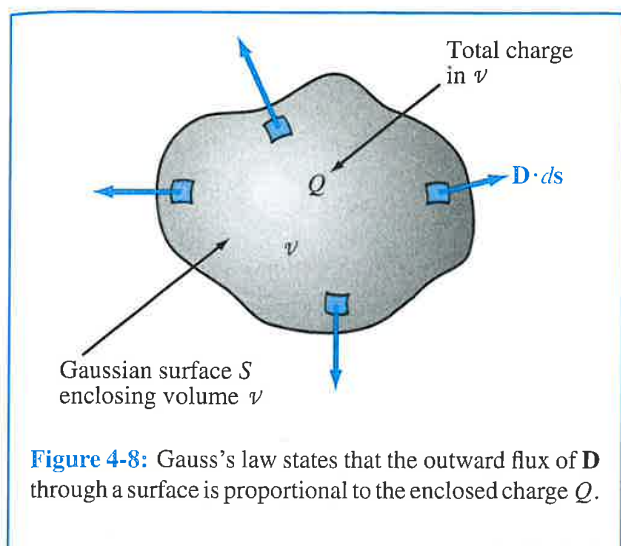
$$\int_V \nabla \cdot \mathbf{D} dV = \oint_S \mathbf{D} \cdot d\mathbf{s}. \quad (4.28)$$

Comparison of Eq. (4.27) with Eq. (4.28) leads to

$$\oint_S \mathbf{D} \cdot d\mathbf{s} = Q \quad (\text{Gauss's law}). \quad (4.29)$$

The integral form of Gauss's law is illustrated diagrammatically in Fig. 4-8; for each differential surface element $d\mathbf{s}$, $\mathbf{D} \cdot d\mathbf{s}$ is the electric field flux flowing outwardly through $d\mathbf{s}$, and the total flux through the surface S is equal to the enclosed charge Q . The surface S is called a **Gaussian surface**.

When the dimensions of a very small volume ΔV containing a total charge q are much smaller than the distance from ΔV to the point at which the electric flux density $\mathbf{D}$ is to be evaluated, then q may be regarded as a **point charge**. The integral form of Gauss's law can be applied to determine $\mathbf{D}$ due to a single isolated charge q by constructing a closed, spherical, Gaussian surface S of an



arbitrary radius R centered at q , as shown in Fig. 4-9. From symmetry considerations, assuming that q is positive, the direction of $\mathbf{D}$ must be radially outward along the unit vector $\hat{\mathbf{R}}$, and D_R , the magnitude of $\mathbf{D}$, must be the same at all points on the Gaussian surface S . Thus, at any point on the surface, defined by position vector $\mathbf{R}$,

$$\mathbf{D}(\mathbf{R}) = \hat{\mathbf{R}} D_R, \quad (4.30)$$

and $d\mathbf{s} = \hat{\mathbf{R}} ds$. Applying Gauss's law gives

$$\begin{aligned} \oint_S \mathbf{D} \cdot d\mathbf{s} &= \oint_S \hat{\mathbf{R}} D_R \cdot \hat{\mathbf{R}} ds \\ &= \oint_S D_R ds = D_R (4\pi R^2) = q. \end{aligned} \quad (4.31)$$

Solving for D_R and then inserting the result in Eq. (4.30) gives the following expression for the electric field $\mathbf{E}$ induced by an isolated point charge in a medium with permittivity ϵ :

$$\mathbf{E}(\mathbf{R}) = \frac{\mathbf{D}(\mathbf{R})}{\epsilon} = \hat{\mathbf{R}} \frac{q}{4\pi \epsilon R^2} \quad (\text{V/m}). \quad (4.32)$$

This is identical with Eq. (4.13) obtained from Coulomb's law. For this simple case of an isolated point charge, it does not much matter whether Coulomb's law or Gauss's law is used to obtain the expression for $\mathbf{E}$. However, it does matter as to which approach we follow when we deal with multiple point charges or continuous charge distributions. Even though Coulomb's law can be used to find $\mathbf{E}$ for any specified distribution of charge, Gauss's law is easier to apply than Coulomb's law, but its utility is limited to symmetrical charge distributions.

Gauss's law, as given by Eq. (4.29), provides a convenient method for determining the electrostatic flux density $\mathbf{D}$ when the charge distribution possesses symmetry properties that allow us to make valid assumptions about the variations of the magnitude and direction of $\mathbf{D}$ as a function of spatial location. Because at every point on the surface the direction of $d\mathbf{s}$ is the outward normal to the surface, only the normal component of $\mathbf{D}$ at the surface contributes to the integral in Eq. (4.29). To successfully apply Gauss's law, the surface S should be chosen such that, from symmetry considerations, the magnitude of $\mathbf{D}$ is constant and its direction is normal or tangential at every point of each subsurface of S (the surface of a cube, for example, has six subsurfaces). These aspects are illustrated in Example 4-6.

Example 4-6 Electric Field of an Infinite Line of Charge

Use Gauss's law to obtain an expression for $\mathbf{E}$ in free space due to an infinitely long line of charge with uniform charge density ρ_l along the z -axis.

Solution: Since the line of charge is infinite in extent and is along the z -axis, symmetry considerations dictate that $\mathbf{D}$ must be in the radial $\hat{\mathbf{r}}$ -direction and must not depend on ϕ or z . Thus, $\mathbf{D} = \hat{\mathbf{r}}D_r$. In Fig. 4-10, we construct a cylindrical Gaussian surface of radius r , concentric around the line of charge. The total charge contained within the cylinder is $Q = \rho_l h$, where h is the height of the cylinder. Since $\mathbf{D}$ is along $\hat{\mathbf{r}}$, the top and bottom surfaces of the cylinder do not contribute to the surface integral on the left-hand side of Eq. (4.29), and only the side surface contributes to the integral. Hence,

$$\int_{z=0}^h \int_{\phi=0}^{2\pi} \hat{\mathbf{r}}D_r \cdot \hat{\mathbf{r}} d\phi dz = \rho_l h$$

or

$$2\pi h D_r r = \rho_l h,$$

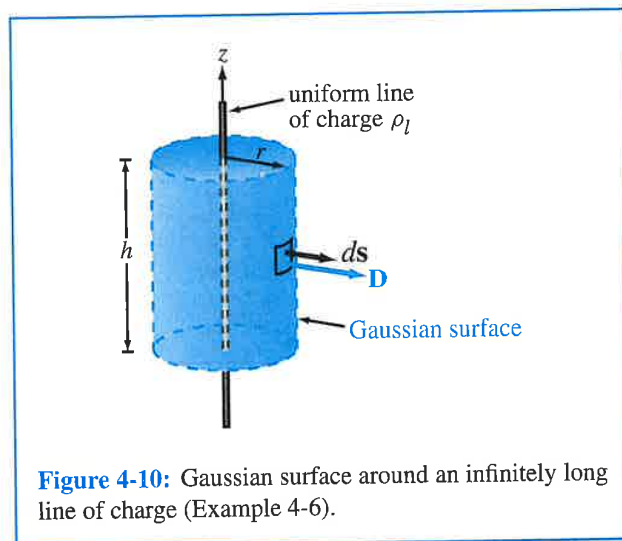


Figure 4-10: Gaussian surface around an infinitely long line of charge (Example 4-6).

which gives the result

$$\mathbf{E} = \frac{\mathbf{D}}{\epsilon_0} = \hat{\mathbf{r}} \frac{D_r}{\epsilon_0} = \hat{\mathbf{r}} \frac{\rho_l}{2\pi \epsilon_0 r} \quad (\text{infinite line of charge}). \quad (4.33)$$

Note that Eq. (4.33) is applicable for any infinite line of charge, regardless of its location and direction, as long as $\hat{\mathbf{r}}$ is properly defined as the radial distance vector from the line charge to the observation point (i.e., $\hat{\mathbf{r}}$ is perpendicular to the line of charge). ■

REVIEW QUESTIONS

Q4.7 Explain Gauss's law. Under what circumstances is it useful?

Q4.8 How should one choose a Gaussian surface?

Q4.9 When is it reasonable to treat a charge distribution as a point charge?

EXERCISE 4.7 Two infinite lines of charge, each carrying a charge density ρ_l , are parallel to the z -axis and located at $x = 1$ and $x = -1$. Determine $\mathbf{E}$ at an arbitrary point in free space along the y -axis.

Ans. $\mathbf{E} = \hat{\mathbf{y}}\rho_l y / [\pi \epsilon_0 (y^2 + 1)]$.

EXERCISE 4.8 A thin spherical shell of radius a carries a uniform surface charge density ρ_s . Use Gauss's law to determine $\mathbf{E}$.

Ans. $\mathbf{E} = 0$ for $R < a$;
 $\mathbf{E} = \hat{\mathbf{R}}\rho_s a^2 / (\epsilon R^2)$ for $R > a$.

EXERCISE 4.9 A spherical volume of radius a contains a uniform volume charge density ρ_v . Use Gauss's law to determine $\mathbf{D}$ for (a) $R \leq a$ and (b) $R \geq a$.

Ans. (a) $\mathbf{D} = \hat{\mathbf{R}}\rho_v R/3$,
 (b) $\mathbf{D} = \hat{\mathbf{R}}\rho_v a^3 / (3R^2)$.

4-5 Electric Scalar Potential

In electric circuits, we work with voltages and currents. The voltage V between two points in the circuit represents the amount of work, or *potential* energy, required to move a unit charge between the two points. In fact, the term “voltage” is a shortened version of “voltage potential” and is the same as *electric potential*. Even though when we solve a circuit problem we usually do not consider the electric fields present in the circuit, in fact it is the existence of an electric field between two points that gives rise to the voltage difference between them, such as across a resistor or a capacitor. The relationship between the electric field $\mathbf{E}$ and the electric potential V is the subject of this section.

4-5.1 Electric Potential as a Function of Electric Field

We begin by considering the simple case of a positive charge q in a uniform electric field $\mathbf{E} = -\hat{\mathbf{y}}E$, parallel to the $-y$ -direction, as shown in Fig. 4-11. The presence of the field $\mathbf{E}$ exerts a force $\mathbf{F}_e = q\mathbf{E}$ on the charge in the negative y -direction. If we attempt to move the charge along the positive y -direction (against the force $\mathbf{F}_e$), we will need to provide an external force $\mathbf{F}_{\text{ext}}$ to counteract $\mathbf{F}_e$, which requires the expenditure of energy. To move q without any acceleration (at a constant speed),

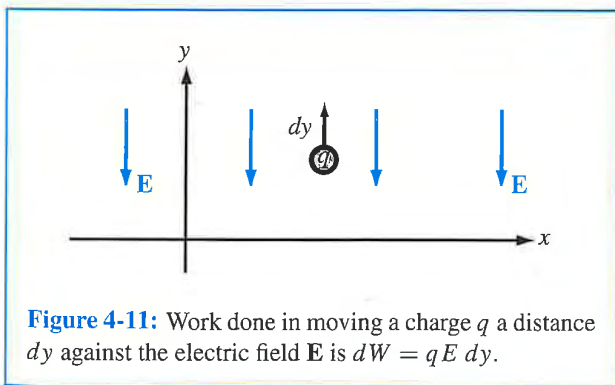


Figure 4-11: Work done in moving a charge q a distance dy against the electric field $\mathbf{E}$ is $dW = qE dy$.

it is necessary that the net force acting on the charge be zero, which means that $\mathbf{F}_{\text{ext}} + \mathbf{F}_e = 0$, or

$$\mathbf{F}_{\text{ext}} = -\mathbf{F}_e = -q\mathbf{E}. \quad (4.34)$$

The work done, or energy expended, in moving any object a vector differential distance $d\mathbf{l}$ under the influence of a force $\mathbf{F}_{\text{ext}}$ is

$$dW = \mathbf{F}_{\text{ext}} \cdot d\mathbf{l} = -q\mathbf{E} \cdot d\mathbf{l} \quad (\text{J}). \quad (4.35)$$

Work, or energy, is measured in joules (J). In the present case, if the charge is moved a distance dy along $\hat{\mathbf{y}}$, then

$$dW = -q(-\hat{\mathbf{y}}E) \cdot \hat{\mathbf{y}} dy = qE dy. \quad (4.36)$$

The differential electric potential energy dW per unit charge is called the *differential electric potential* (or differential voltage) dV . That is,

$$dV = \frac{dW}{q} = -\mathbf{E} \cdot d\mathbf{l} \quad (\text{J/C or V}). \quad (4.37)$$

The unit of V is the volt (V), with $1 \text{ V} \triangleq 1 \text{ J/C}$, and since V is measured in volts, the electric field is expressed in volts per meter (V/m).

The potential difference between any two points P_2 and P_1 (Fig. 4-12) is obtained by integrating Eq. (4.37) *along any path* between them. That is,

$$\int_{P_1}^{P_2} dV = - \int_{P_1}^{P_2} \mathbf{E} \cdot d\mathbf{l}, \quad (4.38)$$

or

$$V_{21} = V_2 - V_1 = - \int_{P_1}^{P_2} \mathbf{E} \cdot d\mathbf{l}, \quad (4.39)$$

where V_1 and V_2 are the electric potentials at points P_1 and P_2 , respectively. The result of the line integral on the right-hand side of Eq. (4.39) should be independent of the specific integration path taken between points P_1 and P_2 . This requirement is mandated by the law of conservation of energy. To illustrate with an example,

let us consider a particle in Earth's gravitational field. If the particle is moved from a height h_1 above Earth's surface to a higher height h_2 , the particle gains potential energy in an amount proportional to $(h_2 - h_1)$. If, instead, we were to first raise the particle from height h_1 to a height h_3 greater than h_2 , thereby giving it potential energy proportional to $(h_3 - h_1)$, and then we let it drop back to height h_2 by expending an energy amount proportional to $(h_3 - h_2)$, its net gain in potential energy will again be proportional to $(h_2 - h_1)$. The same principle applies to the electric potential energy W and to the potential difference $(V_2 - V_1)$. The voltage difference between two nodes in an electric circuit has the same value regardless of which path in the circuit we follow between the nodes. Moreover, **Kirchhoff's voltage law** states that the net voltage drop around a closed loop is zero. If we go from P_1 to P_2 by path 1 in Fig. 4-12 and then return from P_2 to P_1 by path 2, the right-hand side of Eq. (4.39) becomes a closed contour and the left-hand side becomes zero. In fact, the line integral of the electrostatic field $\mathbf{E}$ around any closed contour C

is zero:

$$\oint_C \mathbf{E} \cdot d\mathbf{l} = 0 \quad (\text{Electrostatics}). \quad (4.40)$$

A vector field whose line integral along any closed path is zero is called a **conservative** or an **irrotational** field. Hence, the electrostatic field $\mathbf{E}$ is conservative. As we will see later in Chapter 6, if $\mathbf{E}$ is a time-varying function, it is no longer conservative, and its line integral along a closed path is not necessarily equal to zero.

The conservative property of the electrostatic field can also be deduced from Maxwell's second equation, Eq. (4.1b). If $\partial/\partial t = 0$, then

$$\nabla \times \mathbf{E} = 0. \quad (4.41)$$

If we take the surface integral of $\nabla \times \mathbf{E}$ over an open surface S and then apply Stokes's theorem, given by Eq. (3.107), to convert the surface integral into a line integral, we have

$$\int_S (\nabla \times \mathbf{E}) \cdot d\mathbf{s} = \oint_C \mathbf{E} \cdot d\mathbf{l} = 0, \quad (4.42)$$

where C is a closed contour surrounding S . Thus, Eq. (4.41) is the differential-form equivalent of Eq. (4.40).

We now define what we mean by the electric potential V at a point in space. Before we do so, however, let us revisit our electric-circuit analogue. Absolute voltage at a point in a circuit has no defined meaning, nor does absolute electric potential at a point in space. When we talk of the voltage V of a point in a circuit, we do so in reference to the voltage of some conveniently chosen point to which we have assigned a reference voltage of zero, which we call **ground**. The same principle applies to the electric potential V . Usually, the reference-potential point is chosen to be at infinity. That is, in Eq. (4.39) we

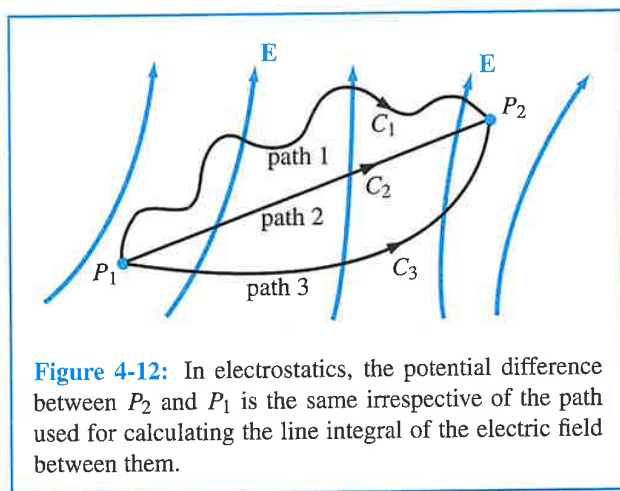


Figure 4-12: In electrostatics, the potential difference between P_2 and P_1 is the same irrespective of the path used for calculating the line integral of the electric field between them.

assume that $V_1 = 0$ when P_1 is at infinity, and therefore the electric potential V at any point P is given by

$$V = - \int_{\infty}^P \mathbf{E} \cdot d\mathbf{l} \quad (\text{V}). \quad (4.43)$$

4-5.2 Electric Potential due to Point Charges

For a point charge q located at the origin of a spherical coordinate system, the electric field at a distance R is given by Eq. (4.32) as

$$\mathbf{E} = \hat{\mathbf{R}} \frac{q}{4\pi\epsilon R^2} \quad (\text{V/m}). \quad (4.44)$$

As was stated earlier, the choice of integration path between the two end points in Eq. (4.43) is arbitrary. Hence, we will conveniently choose the path to be along the radial direction $\hat{\mathbf{R}}$, in which case $d\mathbf{l} = \hat{\mathbf{R}} dR$ and

$$V = - \int_{\infty}^R \left(\hat{\mathbf{R}} \frac{q}{4\pi\epsilon R^2} \right) \cdot \hat{\mathbf{R}} dR = \frac{q}{4\pi\epsilon R} \quad (\text{V}). \quad (4.45)$$

If the charge q is at a location other than the origin, specified by a source position vector $\mathbf{R}_1$, then V at observation position vector $\mathbf{R}$ becomes

$$V(\mathbf{R}) = \frac{q}{4\pi\epsilon |\mathbf{R} - \mathbf{R}_1|} \quad (\text{V}), \quad (4.46)$$

where $|\mathbf{R} - \mathbf{R}_1|$ is the distance between the observation point and the location of the charge q . The principle of superposition that we applied previously to the electric field $\mathbf{E}$ also applies to the electric potential V . Hence, for N discrete point charges $q_1, q_2, \dots, q_N$ having position vectors $\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N$, the electric potential is

$$V(\mathbf{R}) = \frac{1}{4\pi\epsilon} \sum_{i=1}^N \frac{q_i}{|\mathbf{R} - \mathbf{R}_i|} \quad (\text{V}). \quad (4.47)$$

4-5.3 Electric Potential due to Continuous Distributions

For a continuous charge distribution specified over a given volume ν' , across a surface S' , or along a line l' , we (1) replace q_i in Eq. (4.47) with, respectively, $\rho_v d\nu'$, $\rho_s ds'$, and $\rho_l dl'$; (2) convert the summation into an integration; and (3) define $R' = |\mathbf{R} - \mathbf{R}_i|$ as the distance between the integration point and the observation point. These steps lead to the following expressions:

$$V(\mathbf{R}) = \frac{1}{4\pi\epsilon} \int_{\nu'} \frac{\rho_v}{R'} d\nu' \quad (\text{volume distribution}), \quad (4.48a)$$

$$V(\mathbf{R}) = \frac{1}{4\pi\epsilon} \int_{S'} \frac{\rho_s}{R'} ds' \quad (\text{surface distribution}), \quad (4.48b)$$

$$V(\mathbf{R}) = \frac{1}{4\pi\epsilon} \int_{l'} \frac{\rho_l}{R'} dl' \quad (\text{line distribution}). \quad (4.48c)$$

4-5.4 Electric Field as a Function of Electric Potential

In Section 4-5.1, we expressed V in terms of a line integral over $\mathbf{E}$. Now we explore the inverse relationship by examining the differential form of V given by Eq. (4.37):

$$dV = -\mathbf{E} \cdot d\mathbf{l}. \quad (4.49)$$

For a scalar function V , Eq. (3.73) gives

$$dV = \nabla V \cdot d\mathbf{l}, \quad (4.50)$$

where ∇V is the gradient of V . Comparison of Eq. (4.49) with Eq. (4.50) leads to

$$\mathbf{E} = -\nabla V. \quad (4.51)$$

This relationship between V and $\mathbf{E}$ in differential form allows us to determine $\mathbf{E}$ for any charge distribution by first calculating V using the expressions given in Sections 4-5.1 to 4-5.3 and then taking the negative gradient of V .

to find $\mathbf{E}$. The expressions for V , given by Eqs. (4.47) to (4.48c), involve scalar sums and scalar integrals, and as such they are usually much easier to calculate than the vector sums and integrals in the expressions for $\mathbf{E}$ derived in Section 4-3 on the basis of Coulomb's law. Thus, even though the electric potential approach for finding $\mathbf{E}$ is a two-step process, it is computationally simpler to apply than the direct method based on Coulomb's law.

Example 4-7 Electric Field of an Electric Dipole

An **electric dipole** consists of two point charges of equal magnitude and opposite polarity, separated by a small distance, as shown in Fig. 4-13(a). Determine V and $\mathbf{E}$ at any point P in free space, given that P is at a distance $R \gg d$, where d is the spacing between the two charges.

Solution: The electric potential due to a single point charge is given by Eq. (4.45). For the two charges shown in Fig. 4-13(a), application of Eq. (4.47) gives

$$V = \frac{1}{4\pi\epsilon_0} \left(\frac{q}{R_1} + \frac{-q}{R_2} \right) = \frac{q}{4\pi\epsilon_0} \left(\frac{R_2 - R_1}{R_1 R_2} \right).$$

Since $d \ll R$, the lines labeled R_1 and R_2 in Fig. 4-13(a) are approximately parallel to each other, in which case the following approximations apply:

$$R_2 - R_1 \simeq d \cos \theta, \quad R_1 R_2 \simeq R^2.$$

Hence,

$$V = \frac{qd \cos \theta}{4\pi\epsilon_0 R^2}. \quad (4.52)$$

The numerator of Eq. (4.52) can be written as the dot product of $q\mathbf{d}$, where $\mathbf{d}$ is the distance vector from charge $-q$ to charge $+q$, and the unit vector $\hat{\mathbf{R}}$ pointing from the center of the dipole toward the observation point P ,

$$qd \cos \theta = q\mathbf{d} \cdot \hat{\mathbf{R}} = \mathbf{p} \cdot \hat{\mathbf{R}}, \quad (4.53)$$

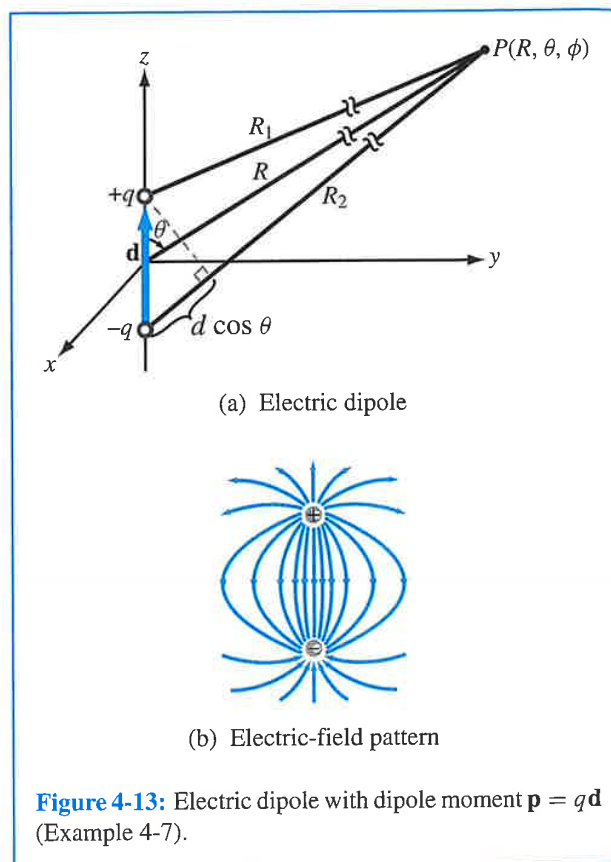


Figure 4-13: Electric dipole with dipole moment $\mathbf{p} = q\mathbf{d}$ (Example 4-7).

where $\mathbf{p} = q\mathbf{d}$ is called the **dipole moment** of the electric dipole. Using Eq. (4.53) in Eq. (4.52) then gives

$$V = \frac{\mathbf{p} \cdot \hat{\mathbf{R}}}{4\pi\epsilon_0 R^2} \quad (\text{electric dipole}). \quad (4.54)$$

In spherical coordinates, Eq. (4.51) is given by

$$\begin{aligned} \mathbf{E} &= -\nabla V \\ &= -\left(\hat{\mathbf{R}} \frac{\partial V}{\partial R} + \hat{\theta} \frac{1}{R} \frac{\partial V}{\partial \theta} + \hat{\phi} \frac{1}{R \sin \theta} \frac{\partial V}{\partial \phi} \right), \end{aligned} \quad (4.55)$$

where we have used the expression for ∇V given on the inside back cover of the book. Upon taking the derivatives of the expression for V given by Eq. (4.52) with respect to R and θ and then substituting the results in Eq. (4.55), we have

$$\mathbf{E} = \frac{qd}{4\pi\epsilon_0 R^3} (\hat{\mathbf{R}} 2 \cos \theta + \hat{\boldsymbol{\theta}} \sin \theta) \quad (\text{V/m}). \quad (4.56)$$

We should note that the expressions for V and $\mathbf{E}$, given by Eqs. (4.54) and (4.56), are applicable only when $R \gg d$. To compute V and $\mathbf{E}$ at points in the vicinity of the two charges making up the dipole, it is necessary to perform the calculations without resorting to the far-distance approximations that led to Eq. (4.52). Such an exact calculation for $\mathbf{E}$ leads to the pattern shown in Fig. 4-13(b).

4-5.5 Poisson's Equation

With $\mathbf{D} = \epsilon \mathbf{E}$, the differential form of Gauss's law given by Eq. (4.26) may be written as

$$\nabla \cdot \mathbf{E} = \frac{\rho_v}{\epsilon}. \quad (4.57)$$

Inserting Eq. (4.51) in Eq. (4.57) gives

$$\nabla \cdot (\nabla V) = -\frac{\rho_v}{\epsilon}. \quad (4.58)$$

In view of the definition for the Laplacian of a scalar function V given by Eq. (3.110) as

$$\nabla^2 V = \nabla \cdot (\nabla V) = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}, \quad (4.59)$$

Eq. (4.58) can be cast in the abbreviated form

$$\nabla^2 V = -\frac{\rho_v}{\epsilon} \quad (\text{Poisson's equation}). \quad (4.60)$$

This is known as **Poisson's equation**. For a volume v' containing a volume charge density distribution ρ_v , the

solution for V derived previously and expressed by Eq. (4.48a) as

$$V = \frac{1}{4\pi\epsilon} \int_{v'} \frac{\rho_v}{R'} dv' \quad (4.61)$$

satisfies Eq. (4.60). If the medium under consideration contains no free charges, Eq. (4.60) reduces to

$$\nabla^2 V = 0 \quad (\text{Laplace's equation}),$$

and it is then referred to as **Laplace's equation**. Poisson's and Laplace's equations are useful for determining the electrostatic potential V in regions at whose boundaries V is known, such as the region between the plates of a capacitor with a specified voltage difference across it.

REVIEW QUESTIONS

Q4.10 What is a conservative field?

Q4.11 Why is the electric potential at a point in space always defined relative to the potential at some reference point?

Q4.12 Explain why Eq. (4.40) is a mathematical statement of Kirchhoff's voltage law.

Q4.13 Why is it usually easier to compute V for a given charge distribution and then find $\mathbf{E}$ from $\mathbf{E} = -\nabla V$ than to compute $\mathbf{E}$ directly by applying Coulomb's law?

Q4.14 What is an electric dipole?

EXERCISE 4.10 Determine the electric potential at the origin in free space due to four charges of $20 \mu\text{C}$ each located at the corners of a square in the x - y plane and whose center is at the origin. The square has sides of 2 m each.

Ans. $V = \sqrt{2} \times 10^{-5} / (\pi\epsilon_0) \quad (\text{V}).$

EXERCISE 4.11 A spherical shell of radius R has a uniform surface charge density ρ_s . Determine the electric potential at the center of the shell.

Ans. $V = \rho_s R / \epsilon$ (V).

4-6 Electrical Properties of Materials

The electromagnetic *constitutive parameters* of a material medium are its electrical permittivity ϵ , magnetic permeability μ , and conductivity σ . A material is said to be *homogeneous* if its constitutive parameters do not vary from point to point, and it is *isotropic* if its constitutive parameters are independent of direction. Most materials exhibit isotropic properties, but some crystals do not. Throughout this book, all materials are assumed to be homogeneous and isotropic. In this chapter, we are concerned with only ϵ and σ . Discussion of μ is deferred to Chapter 5.

The conductivity of a material is a measure of how easily electrons can travel through the material under the influence of an external electric field. Materials are classified as *conductors* (metals) or *dielectrics* (insulators) according to the magnitudes of their conductivities. A conductor has a large number of loosely attached electrons in the outermost shells of the atoms. In the absence of an external electric field, these free electrons move in random directions and with varying speeds. Their random motion produces zero average current through the conductor. Upon applying an external electric field, however, the electrons migrate from one atom to the next along a direction opposite that of the external field. Their movement, which is characterized by an average velocity called the *electron drift velocity* $\mathbf{u}_e$, gives rise to a *conduction current*.

In a dielectric, the electrons are tightly held to the atoms, so much so that it is very difficult to detach them even under the influence of an electric field. Consequently, no current flows through the material. A *perfect*

Table 4-1: Conductivity of some common materials at 20° C.

Material	Conductivity, σ (S/m)
<i>Conductors</i>	
Silver	6.2×10^7
Copper	5.8×10^7
Gold	4.1×10^7
Aluminum	3.5×10^7
Iron	10^7
Mercury	10^6
Carbon	3×10^4
<i>Semiconductors</i>	
Pure germanium	2.2
Pure silicon	4.4×10^{-4}
<i>Insulators</i>	
Glass	10^{-12}
Paraffin	10^{-15}
Mica	10^{-15}
Fused quartz	10^{-17}

dielectric is a material with $\sigma = 0$ and, in contrast, a *perfect conductor* is a material with $\sigma = \infty$.

The conductivity σ of most metals is in the range from 10^6 to 10^7 S/m, compared with 10^{-10} to 10^{-17} S/m for good insulators (Table 4-1). Materials whose conductivities fall between those of conductors and insulators are called *semiconductors*. The conductivity of pure germanium, for example, is 2.2 S/m. Tabulated values of σ are given in Appendix B for some common materials at room temperature (20° C), and a subset is given in Table 4-1.

The conductivity of a material depends on several factors, including temperature and the presence of impurities. In general, σ of metals increases with decreasing temperature, and at very low temperatures in the neighborhood of absolute zero, some conductors become *superconductors* because their conductivities become practically infinite.

REVIEW QUESTIONS

Q4.15 What are the electromagnetic constitutive parameters of a material medium?

Q4.16 What classifies a material as a conductor, semiconductor, or dielectric? What is a superconductor?

Q4.17 What is the conductivity of a perfect dielectric?

4-7 Conductors

The drift velocity $\mathbf{u}_e$ of electrons in a conducting material is related to the externally applied electric field $\mathbf{E}$ through

$$\mathbf{u}_e = -\mu_e \mathbf{E} \quad (\text{m/s}), \quad (4.62a)$$

where μ_e is a material property call the *electron mobility* with units of $(\text{m}^2/\text{V}\cdot\text{s})$. In a semiconductor, current flow is due to the movement of both electrons and holes, and since holes are positive-charge carriers, the *hole drift velocity* $\mathbf{u}_h$ is in the same direction as $\mathbf{E}$,

$$\mathbf{u}_h = \mu_h \mathbf{E} \quad (\text{m/s}), \quad (4.62b)$$

where μ_h is the *hole mobility*. The mobility accounts for the effective mass of a charged particle and the average distance over which the applied electric field can accelerate it before it is stopped by colliding with an atom and must start accelerating over again. From Eq. (4.11), the current density in a medium containing a volume density ρ_v of charges moving with a velocity $\mathbf{u}$ is $\mathbf{J} = \rho_v \mathbf{u}$. In the present case, the current density consists of a component $\mathbf{J}_e$ due to the electrons and a component $\mathbf{J}_h$ due to the holes. Thus, the total *conduction current density* is

$$\mathbf{J} = \mathbf{J}_e + \mathbf{J}_h = \rho_{ve} \mathbf{u}_e + \rho_{vh} \mathbf{u}_h \quad (\text{A/m}^2). \quad (4.63)$$

Use of Eqs. (4.62a) and (4.62b) gives

$$\mathbf{J} = (-\rho_{ve}\mu_e + \rho_{vh}\mu_h)\mathbf{E}, \quad (4.64)$$

where $\rho_{ve} = -N_e e$ and $\rho_{vh} = N_h e$, with N_e and N_h being the number of free electrons and the number of free holes per unit volume, and $e = 1.6 \times 10^{-19}$ C is the absolute charge of a single hole or electron. The quantity inside the parentheses in Eq. (4.64) is defined as the *conductivity* of the material, σ . Thus,

$$\begin{aligned} \sigma &= -\rho_{ve}\mu_e + \rho_{vh}\mu_h \\ &= (N_e\mu_e + N_h\mu_h)e \quad (\text{S/m}) \quad (\text{semiconductor}), \end{aligned} \quad (4.65)$$

and its unit is siemens per meter (S/m). For a good conductor, $N_h\mu_h \ll N_e\mu_e$, and Eq. (4.65) reduces to

$$\sigma = -\rho_{ve}\mu_e = N_e\mu_e e \quad (\text{S/m}) \quad (\text{conductor}). \quad (4.66)$$

In either case, Eq. (4.64) becomes

$$\mathbf{J} = \sigma \mathbf{E} \quad (\text{A/m}^2) \quad (\text{Ohm's law}), \quad (4.67)$$

and it is called the *point form of Ohm's law*. Note that, in a perfect dielectric with $\sigma = 0$, $\mathbf{J} = 0$ regardless of $\mathbf{E}$, and in a perfect conductor with $\sigma = \infty$, $\mathbf{E} = \mathbf{J}/\sigma = 0$ regardless of $\mathbf{J}$. That is,

$$\begin{aligned} \text{Perfect dielectric:} \quad & \mathbf{J} = 0 \\ \text{Perfect conductor:} \quad & \mathbf{E} = 0 \end{aligned}$$

Because σ is on the order of 10^6 S/m for most metals, such as silver, copper, gold, and aluminum (Table 4-1), it is common practice to set $\mathbf{E} = 0$ in metal conductors.

A perfect conductor is an *equipotential* medium, meaning that the electric potential is the same at every point in the conductor. This property follows from the fact that V_{21} , the voltage difference between two points in the conductor, is by definition equal to the line integral of $\mathbf{E}$ between the two points, as indicated by

Eq. (4.39), and since $\mathbf{E} = 0$ everywhere in the perfect conductor, the voltage difference $V_{21} = 0$. The fact that the conductor is an equipotential medium, however, does not necessarily imply that the potential difference between the conductor and some other conductor is zero. Each conductor is an equipotential medium, but the presence of different distributions of charges on their surfaces can generate a potential difference between them.

Example 4-8 Conduction Current in a Copper Wire

A 2-mm-diameter copper wire with conductivity of 5.8×10^7 S/m and electron mobility of 0.0032 ($\text{m}^2/\text{V} \cdot \text{s}$) is subjected to an electric field of 20 (mV/m). Find (a) the volume charge density of free electrons, (b) the current density, (c) the current flowing in the wire, (d) the electron drift velocity, and (e) the volume density of free electrons.

Solution

(a)

$$\rho_{ve} = -\frac{\sigma}{\mu_e} = -\frac{5.8 \times 10^7}{0.0032} = -1.81 \times 10^{10} \text{ (C/m}^3\text{)}.$$

(b)

$$J = \sigma E = 5.8 \times 10^7 \times 20 \times 10^{-3} = 1.16 \times 10^6 \text{ (A/m}^2\text{)}.$$

(c)

$$\begin{aligned} I &= JA \\ &= J \left(\frac{\pi d^2}{4} \right) = 1.16 \times 10^6 \left(\frac{\pi \times 4 \times 10^{-6}}{4} \right) = 3.64 \text{ A.} \end{aligned}$$

(d)

$$u_e = -\mu_e E = -0.0032 \times 20 \times 10^{-3} = -6.4 \times 10^{-5} \text{ m/s.}$$

The minus sign indicates that $\mathbf{u}_e$ is in the opposite direction of $\mathbf{E}$.

(e)

$$N_e = -\frac{\rho_{ve}}{e} = \frac{1.81 \times 10^{10}}{1.6 \times 10^{-19}} = 1.13 \times 10^{29} \text{ electrons/m}^3. \quad \blacksquare$$

EXERCISE 4.12 Determine the density of free electrons in aluminum, given that its conductivity is 3.5×10^7 (S/m) and its electron mobility is 0.0015 ($\text{m}^2/\text{V} \cdot \text{s}$).

Ans. $N_e = 1.46 \times 10^{29}$ electrons/ m^3 .

EXERCISE 4.13 The current flowing through a 100-m-long conducting wire of uniform cross section has a density of 3×10^5 (A/m^2). Find the voltage drop across the length of the wire if the wire material has a conductivity of 2×10^7 (S/m).

Ans. $V = 1.5$ (V).

4-7.1 Resistance

By way of demonstrating the use of the point form of Ohm's law, we shall use it to derive an expression for the resistance R of a conductor of length l and uniform cross section A , as shown in Fig. 4-14. The conductor axis is oriented along the x -axis and extends between points x_1 and x_2 , with $l = x_2 - x_1$. A voltage V applied across the conductor terminals establishes an electric field $\mathbf{E} = \hat{\mathbf{x}}E_x$; the direction of $\mathbf{E}$ is from the point of higher potential (point 1 in Fig. 4-14) to the point of lower potential (point 2). The relation between V and E_x is obtained by applying Eq. (4.39). Thus,

$$\begin{aligned} V &= V_1 - V_2 \\ &= -\int_{x_2}^{x_1} \mathbf{E} \cdot d\mathbf{l} \\ &= -\int_{x_2}^{x_1} \hat{\mathbf{x}}E_x \cdot \hat{\mathbf{x}} dl = E_x l \quad (\text{V}). \end{aligned} \quad (4.68)$$

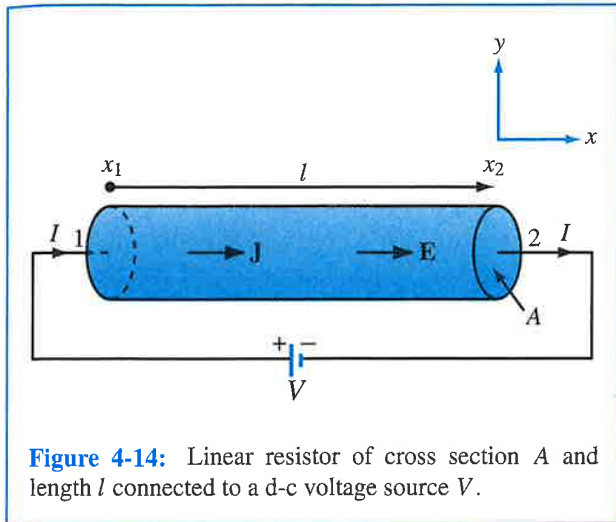


Figure 4-14: Linear resistor of cross section A and length l connected to a d-c voltage source V .

Using Eq. (4.67), the current flowing through the cross section A at x_2 is

$$I = \int_A \mathbf{J} \cdot d\mathbf{s} = \int_A \sigma \mathbf{E} \cdot d\mathbf{s} = \sigma E_x A \quad (\text{A}). \quad (4.69)$$

From $R = V/I$, the ratio of Eq. (4.68) to Eq. (4.69) gives

$$R = \frac{l}{\sigma A} \quad (\Omega). \quad (4.70)$$

We now generalize our result for R to any resistor of arbitrary shape by noting that the voltage V across the resistor is equal to the line integral of $\mathbf{E}$ over a path l between two specified points and the current I is equal to the flux of $\mathbf{J}$ through the surface S of the resistor. Thus,

$$R = \frac{V}{I} = \frac{-\int_l \mathbf{E} \cdot d\mathbf{l}}{\int_S \mathbf{J} \cdot d\mathbf{s}} = \frac{-\int_l \mathbf{E} \cdot d\mathbf{l}}{\int_S \sigma \mathbf{E} \cdot d\mathbf{s}}. \quad (4.71)$$

The reciprocal of R is called the **conductance** G , and the unit of G is (Ω^{-1}) , or siemens (S). For the linear resistor,

$$G = \frac{1}{R} = \frac{\sigma A}{l} \quad (\text{S}). \quad (4.72)$$

Example 4-9 Conductance of Coaxial Cable

The radii of the inner and outer conductors of a coaxial cable of length l are a and b , respectively (Fig. 4-15). The insulation material has conductivity σ . Obtain an expression for G' , the conductance per unit length of the insulation layer.

Solution: Let I be the total current flowing from the inner conductor to the outer conductor through the insulation material. At any radial distance r from the axis of the center conductor, the area through which the current flows is $A = 2\pi r l$. Hence,

$$\mathbf{J} = \hat{\mathbf{r}} \frac{I}{A} = \hat{\mathbf{r}} \frac{I}{2\pi r l}, \quad (4.73)$$

and from $\mathbf{J} = \sigma \mathbf{E}$,

$$\mathbf{E} = \hat{\mathbf{r}} \frac{I}{2\pi \sigma r l}. \quad (4.74)$$

In a resistor, the current flows from higher electric potential to lower potential. Hence, if $\mathbf{J}$ is in the $\hat{\mathbf{r}}$ -direction, the inner conductor must be at a higher potential than the outer conductor. Accordingly, the voltage difference between the conductors is

$$\begin{aligned} V_{ab} &= - \int_b^a \mathbf{E} \cdot d\mathbf{l} = - \int_b^a \frac{I}{2\pi \sigma l} \frac{\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}}{r} dr \\ &= \frac{I}{2\pi \sigma l} \ln \left(\frac{b}{a} \right). \end{aligned} \quad (4.75)$$

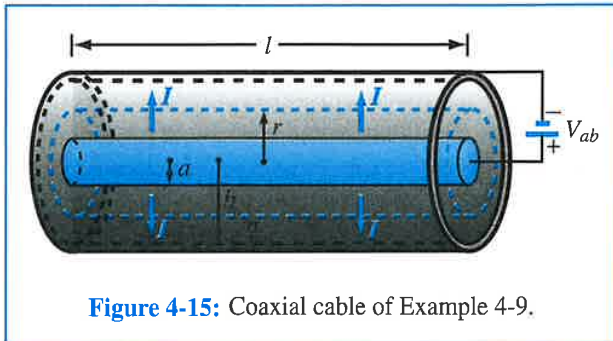


Figure 4-15: Coaxial cable of Example 4-9.

The conductance per unit length is then

$$G' = \frac{G}{l} = \frac{1}{Rl} = \frac{I}{V_{abl}} = \frac{2\pi\sigma}{\ln(b/a)} \quad (\text{S/m}). \quad (4.76)$$

4-7.2 Joule's Law

We now consider the power dissipated in a conducting medium in the presence of an electrostatic field $\mathbf{E}$. The medium contains free electrons and holes with volume charge densities ρ_{ve} and ρ_{vh} , respectively. The electron and hole charge contained in an elemental volume Δv is $q_e = \rho_{ve} \Delta v$ and $q_h = \rho_{vh} \Delta v$, respectively. The electric forces acting on q_e and q_h are $\mathbf{F}_e = q_e \mathbf{E} = \rho_{ve} \mathbf{E} \Delta v$ and $\mathbf{F}_h = q_h \mathbf{E} = \rho_{vh} \mathbf{E} \Delta v$. The work (energy) expended by the electric field in moving q_e a differential distance Δl_e and moving q_h a distance Δl_h is

$$\Delta W = \mathbf{F}_e \cdot \Delta \mathbf{l}_e + \mathbf{F}_h \cdot \Delta \mathbf{l}_h. \quad (4.77)$$

Power P , measured in watts (W), is defined as the time rate of change of energy. The change in power corresponding to ΔW is then

$$\begin{aligned} \Delta P &= \frac{\Delta W}{\Delta t} = \mathbf{F}_e \cdot \frac{\Delta \mathbf{l}_e}{\Delta t} + \mathbf{F}_h \cdot \frac{\Delta \mathbf{l}_h}{\Delta t} \\ &= \mathbf{F}_e \cdot \mathbf{u}_e + \mathbf{F}_h \cdot \mathbf{u}_h \\ &= (\rho_{ve} \mathbf{E} \cdot \mathbf{u}_e + \rho_{vh} \mathbf{E} \cdot \mathbf{u}_h) \Delta v \\ &= \mathbf{E} \cdot \mathbf{J} \Delta v, \end{aligned} \quad (4.78)$$

where $\mathbf{u}_e = \Delta \mathbf{l}_e / \Delta t$ is the electron drift velocity and $\mathbf{u}_h = \Delta \mathbf{l}_h / \Delta t$ is the hole drift velocity. Equation (4.63) was used in the last step of the derivation leading to Eq. (4.78). For a volume v , the total dissipated power is

$$P = \int_v \mathbf{E} \cdot \mathbf{J} d v \quad (\text{W}) \quad (\text{Joule's law}), \quad (4.79)$$

and in view of the relation given by Eq. (4.67),

$$P = \int_v \sigma |\mathbf{E}|^2 d v \quad (\text{W}). \quad (4.80)$$

Equation (4.79) is a mathematical statement of **Joule's law**. For the resistor example considered earlier, $|\mathbf{E}| = E_x$ and its volume is $v = lA$. Separating the volume integral in Eq. (4.80) into a product of a surface integral over A and a line integral over l , we have

$$\begin{aligned} P &= \int_v \sigma |\mathbf{E}|^2 d v \\ &= \int_A \sigma E_x ds \int_l E_x dl \\ &= (\sigma E_x A)(E_x l) = IV \quad (\text{W}), \end{aligned} \quad (4.81)$$

where use was made of Eq. (4.68) for the voltage V and Eq. (4.69) for the current I . With $V = IR$, we obtain the familiar expression

$$P = I^2 R \quad (\text{W}). \quad (4.82)$$

REVIEW QUESTIONS

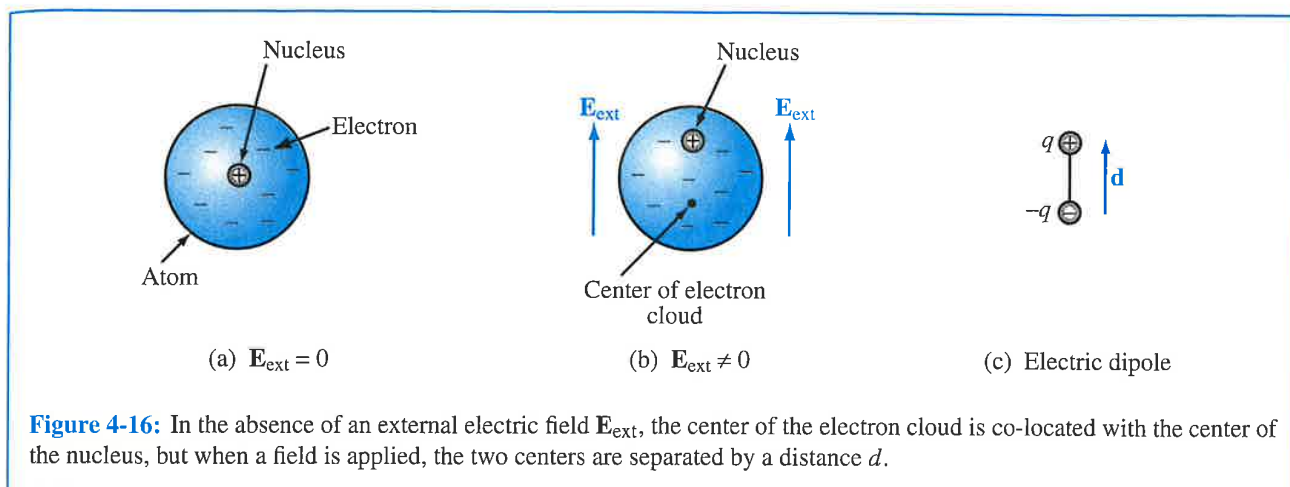
Q4.18 What is the fundamental difference between an insulator, a semiconductor, and a conductor?

Q4.19 Show that the power dissipated in the coaxial cable of Fig. 4-15 is $P = I^2 \ln(b/a)/(2\pi\sigma l)$.

EXERCISE 4.14 A 50-m-long copper wire has a circular cross section with radius $r = 2$ cm. Given that the conductivity of copper is 5.8×10^7 (S/m), determine (a) the resistance R of the wire and (b) the power dissipated in the wire if the voltage across its length is 1.5 (mV).

Ans. (a) $R = 6.9 \times 10^{-4} \Omega$, (b) $P = 3.3$ (mW).

EXERCISE 4.15 Repeat part (b) of Exercise 4.14 by applying Eq. (4.80).



4-8 Dielectrics

As we discussed previously, the fundamental difference between a conductor and a dielectric is that a conductor has loosely held (free) electrons that can migrate through the crystalline structure of the material, whereas the electrons in the outermost shells of a dielectric are strongly bound to the atom. In the absence of an electric field, the electrons in any material form a symmetrical cloud around the nucleus, with the center of the cloud being at the same location as the center of the nucleus, as shown in Fig. 4-16(a). The electric field generated by the positively charged nucleus attracts and holds the electron cloud around it, and the mutual repulsion of the electron clouds of adjacent atoms gives matter its form. When a conductor is subjected to an externally applied electric field, the most loosely bound electrons in each atom can easily jump from one atom to the next, thereby setting up an electric current. In a dielectric, however, an externally applied electric field E_{ext} cannot effect mass migration of charges since none are able to move freely, but it can **polarize** the atoms or molecules in the material by distorting the center of the cloud and the location of the nucleus. The polarization process is illustrated in Fig. 4-16(b). The polarized atom or molecule may be represented by

an electric dipole consisting of charge $+q$ at the center of the nucleus and charge $-q$ at the center of the electron cloud [Fig. 4-16(c)]. Each such dipole sets up a small electric field, pointing from the positively charged nucleus to the center of the equally but negatively charged electron cloud. This **induced** electric field, called a **polarization** field, is weaker than and opposite in direction to E_{ext} . Consequently, the net electric field present in the dielectric material is smaller than E_{ext} . At the microscopic level, each dipole exhibits a dipole moment similar to that described in Example 4-7. Within the dielectric material, the dipoles align themselves in a linear arrangement, as shown in Fig. 4-17. Along the upper and lower edges of the material, the dipole arrangement exhibits a positive surface charge density on the upper surface and a negative density on the lower surface.

The relatively simple picture described in Figs. 4-16 and 4-17 pertains to **nonpolar materials** in which the molecules do not have permanent dipole moments. Nonpolar molecules become polarized only when an external electric field is applied, and when the field is terminated, the molecules return to their original unpolarized state. In some materials, such as water, the molecular structure is such that the molecules possess built-in per-

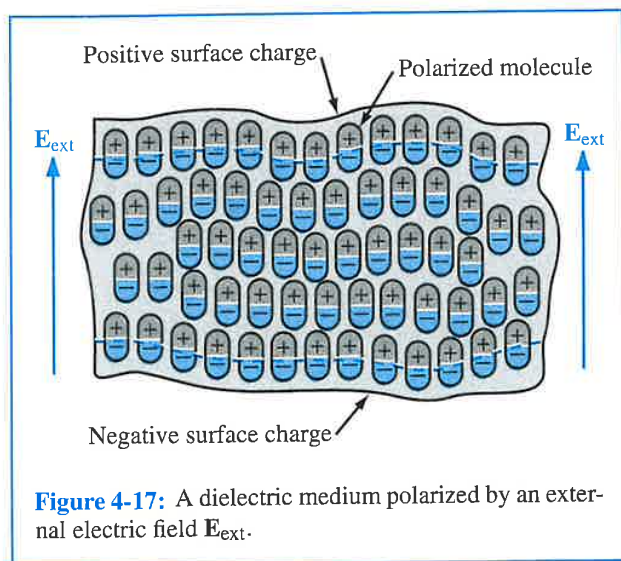


Figure 4-17: A dielectric medium polarized by an external electric field $\mathbf{E}_{\text{ext}}$.

manent dipole moments that are randomly oriented in the absence of an applied electric field. Materials composed of permanent dipoles are called **polar materials**. Owing to their random orientations, the dipoles of polar materials produce no net dipole moment macroscopically (at the macroscopic scale, each point in the material represents a small volume containing thousands of molecules). Under the influence of an applied field, the permanent dipoles tend to align themselves to some extent along the direction of the electric field, in an arrangement somewhat similar to that shown in Fig. 4-17 for nonpolar materials.

Whereas $\mathbf{D}$ and $\mathbf{E}$ are related by ϵ_0 in free space, the presence of these microscopic dipoles in a dielectric material alters that relationship in that material to

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}, \quad (4.83)$$

where $\mathbf{P}$, called the **electric polarization field**, accounts for the polarization properties of the material. The polarization field is produced by the electric field $\mathbf{E}$ and depends on the material properties.

A dielectric medium is said to be **linear** if the magnitude of the induced polarization field is directly proportional to the magnitude of $\mathbf{E}$, and it is said to be **isotropic** if the polarization field and $\mathbf{E}$ are in the same direction. In some crystals, the periodic structure of the material allows more polarization to take place along certain directions, such as the crystal axes, than along others. In such **anisotropic** dielectrics, $\mathbf{E}$ and $\mathbf{D}$ may have different directions. A medium is said to be **homogeneous** if its constitutive parameters (ϵ , μ , and σ) are constant throughout the medium. Our present treatment will be limited to media that are linear, isotropic, and homogeneous. For such media the polarization field is directly proportional to $\mathbf{E}$ and is expressed by the relationship

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E}, \quad (4.84)$$

where χ_e is called the **electric susceptibility** of the material. Inserting Eq. (4.84) into Eq. (4.83), we have

$$\begin{aligned} \mathbf{D} &= \epsilon_0 \mathbf{E} + \epsilon_0 \chi_e \mathbf{E} \\ &= \epsilon_0 (1 + \chi_e) \mathbf{E} = \epsilon \mathbf{E}, \end{aligned} \quad (4.85)$$

which defines the permittivity ϵ of the material as

$$\epsilon = \epsilon_0 (1 + \chi_e). \quad (4.86)$$

As was mentioned earlier, it is often convenient to characterize the permittivity of a material relative to that of free space, ϵ_0 ; this is accommodated by the relative permittivity $\epsilon_r = \epsilon/\epsilon_0$. Values of ϵ_r are listed in Table 4-2 for a few common materials, and a longer list is given in Appendix B. In free space $\epsilon_r = 1$, and **for most conductors** $\epsilon_r \simeq 1$. The dielectric constant of air is approximately 1.0006 at sea level, and it decreases toward unity with increasing altitude. Except in some special circumstances, such as when calculating electromagnetic wave refraction (bending) through the atmosphere over long distances, **air is treated the same as free space**.

The dielectric polarization model presented thus far has placed no restriction on the upper end of the strength of the applied electric field $\mathbf{E}$. In reality, if $\mathbf{E}$ exceeds a

Table 4-2: Relative permittivity (dielectric constant) and dielectric strength of common materials.

Material	Relative Permittivity, ϵ_r	Dielectric Strength, E_{ds} (MV/m)
Air (at sea level)	1.0006	3
Petroleum oil	2.1	12
Polystyrene	2.6	20
Glass	4.5–10	25–40
Quartz	3.8–5	30
Bakelite	5	20
Mica	5.4–6	200

$$\epsilon = \epsilon_r \epsilon_0 \text{ and } \epsilon_0 = 8.854 \times 10^{-12} \text{ F/m.}$$

certain critical value, known as the *dielectric strength* of the material, it will free the electrons completely from the molecules and cause them to accelerate through the material in the form of a conduction current. When this happens, sparking can occur, and the dielectric material can sustain permanent damage due to electron collision with the molecular structure. This abrupt change in behavior is called a *dielectric breakdown*. The dielectric strength E_{ds} is the highest magnitude of $\mathbf{E}$ that the material can sustain without breakdown. Dielectric breakdown can occur in gas, liquid, and solid dielectrics. The associated field strength depends on the material composition, as well as other factors such as temperature and humidity. The dielectric strength for air is 3 (MV/m); for glass, it is 25 to 40 (MV/m); and for mica, it is 200 (MV/m) [see Table 4-2].

A charged thundercloud with an electric potential V , relative to the ground, induces an electric field $E = V/d$ in the air medium between the ground and the cloud, where d is the height of the cloud base above the ground's surface. If V is sufficiently large so that E exceeds the dielectric strength of air, ionization occurs and discharge (lightning) follows. The *breakdown voltage* V_{br} of a parallel-plate capacitor is discussed in Example 4-11.

REVIEW QUESTIONS

Q4.20 What is a polar material? A nonpolar material?

Q4.21 Do $\mathbf{D}$ and $\mathbf{E}$ always point in the same direction? If not, when do they not?

Q4.22 What happens when dielectric breakdown occurs?

4-9 Electric Boundary Conditions

An electric field is said to be spatially continuous if it does not exhibit abrupt changes in either its magnitude or direction as a function of spatial position. Even though the electric field may be continuous in each of two dissimilar media, it may be discontinuous at the boundary between them if surface charge exists along that boundary. Boundary conditions specify how the tangential and normal components of the field in one medium are related to the components of the field across the boundary in another medium. We will derive a general set of boundary conditions, applicable at the interface between *any* two dissimilar media, be they two different dielectrics or a conductor and a dielectric. Also, any of the dielectrics

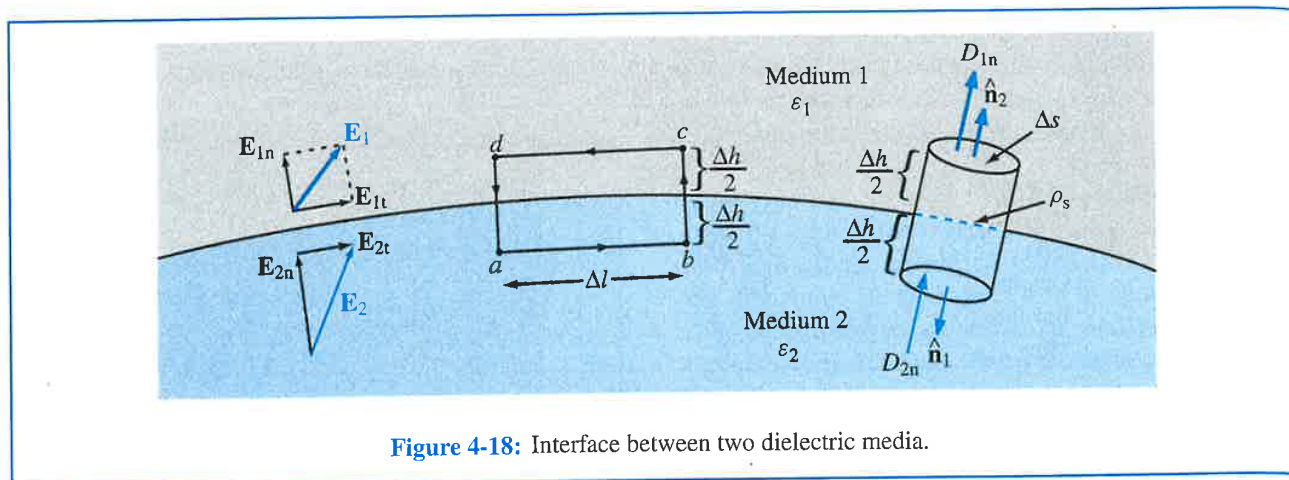


Figure 4-18: Interface between two dielectric media.

may be free space. Even though these boundary conditions will be derived for electrostatic conditions, they will be equally valid for time-varying electric fields. Figure 4-18 shows an interface between medium 1 with permittivity ϵ_1 and medium 2 with permittivity ϵ_2 . In the general case, the boundary may have a surface charge density ρ_s .

To derive the boundary conditions for the tangential components of $\mathbf{E}$ and $\mathbf{D}$, we begin by constructing the closed rectangular loop $abcd$ shown in Fig. 4-18, and then we apply the conservative property of the electric field given by Eq. (4.40), which states that the line integral of the electrostatic field around a closed path is always equal to zero. By letting $\Delta h \rightarrow 0$, the contributions to the line integral by segments bc and da go to zero. Hence,

$$\oint_C \mathbf{E} \cdot d\mathbf{l} = \int_a^b \mathbf{E}_2 \cdot d\mathbf{l} + \int_c^d \mathbf{E}_1 \cdot d\mathbf{l} = 0, \quad (4.87)$$

where $\mathbf{E}_1$ and $\mathbf{E}_2$ are the electric fields in media 1 and 2, respectively. In terms of the tangential and normal directions shown in Fig. 4-18,

$$\mathbf{E}_1 = \mathbf{E}_{1t} + \mathbf{E}_{1n}, \quad (4.88a)$$

$$\mathbf{E}_2 = \mathbf{E}_{2t} + \mathbf{E}_{2n}. \quad (4.88b)$$

Over segment ab , $\mathbf{E}_{2t}$ and $d\mathbf{l}$ have the same direction, but over segment cd , $\mathbf{E}_{1t}$ and $d\mathbf{l}$ are in opposite directions. Consequently, Eq. (4.87) gives

$$E_{2t} \Delta l - E_{1t} \Delta l = 0, \quad (4.89)$$

or

$$E_{1t} = E_{2t} \quad (\text{V/m}). \quad (4.90)$$

Accordingly, *the tangential component of the electric field is continuous across the boundary between any two media*. Since $D_{1t} = \epsilon_1 E_{1t}$ and $D_{2t} = \epsilon_2 E_{2t}$, the boundary condition on the tangential component of the electric flux density is

$$\frac{D_{1t}}{\epsilon_1} = \frac{D_{2t}}{\epsilon_2}. \quad (4.91)$$

Next we apply Gauss's law, as expressed by Eq. (4.29), to determine the boundary conditions on the normal components of $\mathbf{E}$ and $\mathbf{D}$. According to Gauss's law, the total outward flux of $\mathbf{D}$ through the three surfaces of the small cylinder shown in Fig. 4-18 must equal the total charge enclosed in the cylinder. By letting the cylinder's height

$\Delta h \rightarrow 0$, the contribution to the total flux by the side surface goes to zero. Also, even if each of the two media happens to have free or bound volume charge densities, the only charge remaining in the collapsed cylinder is that distributed on the boundary. Thus, $Q = \rho_s \Delta s$, and

$$\oint_S \mathbf{D} \cdot d\mathbf{s} = \int_{\text{top}} \mathbf{D}_1 \cdot \hat{\mathbf{n}}_2 ds + \int_{\text{bottom}} \mathbf{D}_2 \cdot \hat{\mathbf{n}}_1 ds = \rho_s \Delta s, \quad (4.92)$$

where $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$ are the outward normal unit vectors of the bottom and top surfaces, respectively. It is important to remember that *the normal unit vector at the surface of any medium is always defined to be in the outward direction away from that medium*. Since $\hat{\mathbf{n}}_1 = -\hat{\mathbf{n}}_2$, Eq. (4.92) simplifies to

$$\hat{\mathbf{n}}_2 \cdot (\mathbf{D}_1 - \mathbf{D}_2) = \rho_s \quad (\text{C/m}^2). \quad (4.93)$$

With D_{1n} and D_{2n} defined as the normal components of $\mathbf{D}_1$ and $\mathbf{D}_2$ along $\hat{\mathbf{n}}_2$, we have

$$D_{1n} - D_{2n} = \rho_s \quad (\text{C/m}^2). \quad (4.94)$$

Thus, *the normal component of $\mathbf{D}$ changes abruptly at a charged boundary between two different media, and the amount of change is equal to the surface charge density*. The corresponding boundary condition for $\mathbf{E}$ is

$$\varepsilon_1 E_{1n} - \varepsilon_2 E_{2n} = \rho_s. \quad (4.95)$$

In summary, (1) the conservative property of $\mathbf{E}$,

$$\nabla \times \mathbf{E} = 0 \iff \oint_C \mathbf{E} \cdot d\mathbf{l} = 0, \quad (4.96)$$

led to the result that $\mathbf{E}$ has a continuous tangential component across a boundary, and (2) the divergence property of $\mathbf{D}$,

$$\nabla \cdot \mathbf{D} = \rho_v \iff \oint_S \mathbf{D} \cdot d\mathbf{s} = Q, \quad (4.97)$$

led to the result that the normal component of $\mathbf{D}$ changes by ρ_s across the boundary. A summary of the conditions at the boundary between different types of media is given in Table 4-3.

Example 4-10 Application of Boundary Conditions

The x - y plane is a charge-free boundary separating two dielectric media with permittivities ε_1 and ε_2 , as shown in Fig. 4-19. If the electric field in medium 1 is $\mathbf{E}_1 = \hat{\mathbf{x}}E_{1x} + \hat{\mathbf{y}}E_{1y} + \hat{\mathbf{z}}E_{1z}$, find (a) the electric field $\mathbf{E}_2$ in medium 2 and (b) the angles θ_1 and θ_2 .

Solution: (a) Let $\mathbf{E}_2 = \hat{\mathbf{x}}E_{2x} + \hat{\mathbf{y}}E_{2y} + \hat{\mathbf{z}}E_{2z}$. Our task is to find the components of $\mathbf{E}_2$ in terms of the given components of $\mathbf{E}_1$. The normal to the boundary is $\hat{\mathbf{z}}$. Hence, the x and y components of the fields are tangential to the boundary and the z components are normal to the boundary. At a charge-free interface, the tangential components of $\mathbf{E}$ and the normal components of $\mathbf{D}$ are continuous. Consequently,

$$E_{2x} = E_{1x}, \quad E_{2y} = E_{1y},$$

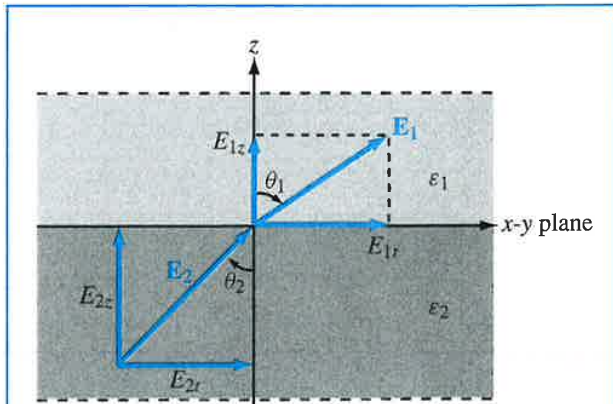


Figure 4-19: Application of boundary conditions at the interface between two dielectric media (Example 4-10).

Table 4-3: Boundary conditions for the electric fields.

Field Component	Any Two Media	Medium 1 Dielectric ϵ_1	Medium 2 Dielectric ϵ_2	Medium 1 Dielectric ϵ_1	Medium 2 Conductor
Tangential $\mathbf{E}$	$E_{1t} = E_{2t}$	$E_{1t} = E_{2t}$	$E_{1t} = E_{2t}$	$E_{1t} = E_{2t} = 0$	
Tangential $\mathbf{D}$	$D_{1t}/\epsilon_1 = D_{2t}/\epsilon_2$	$D_{1t}/\epsilon_1 = D_{2t}/\epsilon_2$	$D_{1t}/\epsilon_1 = D_{2t}/\epsilon_2$	$D_{1t} = D_{2t} = 0$	
Normal $\mathbf{E}$	$\hat{\mathbf{n}} \cdot (\epsilon_1 \mathbf{E}_1 - \epsilon_2 \mathbf{E}_2) = \rho_s$	$\epsilon_1 E_{1n} - \epsilon_2 E_{2n} = \rho_s$	$\epsilon_1 E_{1n} - \epsilon_2 E_{2n} = \rho_s$	$E_{1n} = \rho_s/\epsilon_1$	$E_{2n} = 0$
Normal $\mathbf{D}$	$\hat{\mathbf{n}} \cdot (\mathbf{D}_1 - \mathbf{D}_2) = \rho_s$	$D_{1n} - D_{2n} = \rho_s$	$D_{1n} - D_{2n} = \rho_s$	$D_{1n} = \rho_s$	$D_{2n} = 0$
Notes: (1) ρ_s is the surface charge density at the boundary; (2) normal components of $\mathbf{E}_1$, $\mathbf{D}_1$, $\mathbf{E}_2$, and $\mathbf{D}_2$ are along $\hat{\mathbf{n}}_2$, the outward normal unit vector of medium 2.					

and

$$D_{2z} = D_{1z} \quad \text{or} \quad \epsilon_2 E_{2z} = \epsilon_1 E_{1z}.$$

Hence,

$$\mathbf{E}_2 = \hat{\mathbf{x}}E_{1x} + \hat{\mathbf{y}}E_{1y} + \hat{\mathbf{z}}\frac{\epsilon_1}{\epsilon_2}E_{1z}. \quad (4.98)$$

(b) The tangential components of $\mathbf{E}_1$ and $\mathbf{E}_2$ are $E_{1t} = \sqrt{E_{1x}^2 + E_{1y}^2}$ and $E_{2t} = \sqrt{E_{2x}^2 + E_{2y}^2}$. The angles θ_1 and θ_2 are then given by

$$\tan \theta_1 = \frac{E_{1t}}{E_{1z}} = \frac{\sqrt{E_{1x}^2 + E_{1y}^2}}{E_{1z}},$$

$$\tan \theta_2 = \frac{E_{2t}}{E_{2z}} = \frac{\sqrt{E_{2x}^2 + E_{2y}^2}}{E_{2z}} = \frac{\sqrt{E_{1x}^2 + E_{1y}^2}}{(\epsilon_1/\epsilon_2)E_{1z}},$$

and the two angles are related by

$$\frac{\tan \theta_2}{\tan \theta_1} = \frac{\epsilon_2}{\epsilon_1}. \quad (4.99)$$

EXERCISE 4.16 With reference to Fig. 4-19, find $\mathbf{E}_1$ if $\mathbf{E}_2 = \hat{\mathbf{x}}2 - \hat{\mathbf{y}}3 + \hat{\mathbf{z}}3$ (V/m), $\epsilon_1 = 2\epsilon_0$, and $\epsilon_2 = 8\epsilon_0$. Assume the boundary to be charge free.

Ans. $\mathbf{E}_1 = \hat{\mathbf{x}}2 - \hat{\mathbf{y}}3 + \hat{\mathbf{z}}12$ (V/m).

EXERCISE 4.17 Repeat Exercise 4.16 for a boundary with surface charge density $\rho_s = 3.54 \times 10^{-11}$ (C/m²).

Ans. $\mathbf{E}_1 = \hat{\mathbf{x}}2 - \hat{\mathbf{y}}3 + \hat{\mathbf{z}}14$ (V/m).

4-9.1 Dielectric-Conductor Boundary

Consider the case when medium 1 in Fig. 4-18 is a dielectric and medium 2 is a perfect conductor. In a perfect conductor, $\mathbf{E} = \mathbf{D} = 0$ everywhere in the conductor. Hence, $\mathbf{E}_2 = \mathbf{D}_2 = 0$, which requires the tangential and normal components of $\mathbf{E}_2$ and $\mathbf{D}_2$ to be zero. Consequently, from Eq. (4.90) and Eq. (4.94), the fields in the dielectric medium, at the boundary with the conductor, are given by

$$E_{1t} = D_{1t} = 0, \quad (4.100a)$$

$$D_{1n} = \epsilon_1 E_{1n} = \rho_s. \quad (4.100b)$$

These two boundary conditions can be combined into

$$\mathbf{D}_1 = \epsilon_1 \mathbf{E}_1 = \hat{\mathbf{n}}\rho_s \quad (\text{at conductor surface}), \quad (4.101)$$

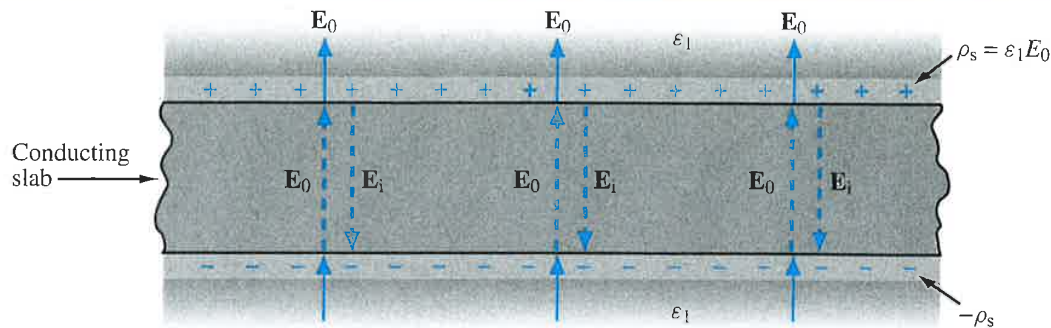


Figure 4-20: When a conducting slab is placed in an external electric field $\mathbf{E}_0$, charges that accumulate on the conductor surfaces induce an internal electric field $\mathbf{E}_i = -\mathbf{E}_0$. Consequently, the total field inside the conductor is zero.

where $\hat{\mathbf{n}}$ is a unit vector directed normally outward from the conducting surface. This means that *the electric field lines point directly away from the conductor surface when ρ_s is positive and directly toward the conductor surface when ρ_s is negative.*

Figure 4-20 shows an infinitely long conducting slab placed in a uniform electric field $\mathbf{E}_0$. The medium above and below the slab has a permittivity ϵ_1 . Because $\mathbf{E}_0$ points away from the upper surface, it induces a positive charge density $\rho_s = \epsilon_1 |\mathbf{E}_0|$ on the upper surface of the slab. On the bottom surface, $\mathbf{E}_0$ points toward the surface, and therefore the induced charge density is $-\rho_s$. The presence of these surface charges induces an electric field $\mathbf{E}_i$ in the conductor, resulting in a total field $\mathbf{E} = \mathbf{E}_0 + \mathbf{E}_i$. To satisfy the condition that $\mathbf{E}$ must be everywhere zero in the conductor, $\mathbf{E}_i$ must equal $-\mathbf{E}_0$.

If we place a metallic sphere in an electrostatic field, as shown in Fig. 4-21, negative charges will accumulate on the lower hemisphere and positive charges will accumulate on the upper hemisphere. The presence of the sphere causes the field lines to bend to satisfy the condition given

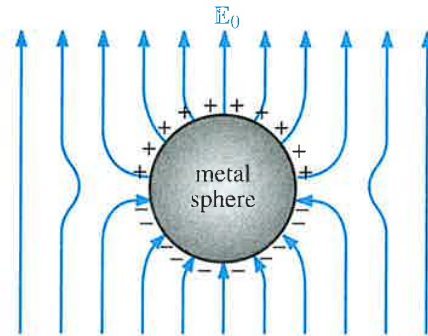


Figure 4-21: Metal sphere placed in an external electric field $\mathbf{E}_0$.

by Eq. (4.101); that is, *$\mathbf{E}$ is always normal to the surface at the conductor boundary.*

4-9.2 Conductor–Conductor Boundary

We now examine the general case of the boundary between two media neither of which are perfect dielectrics

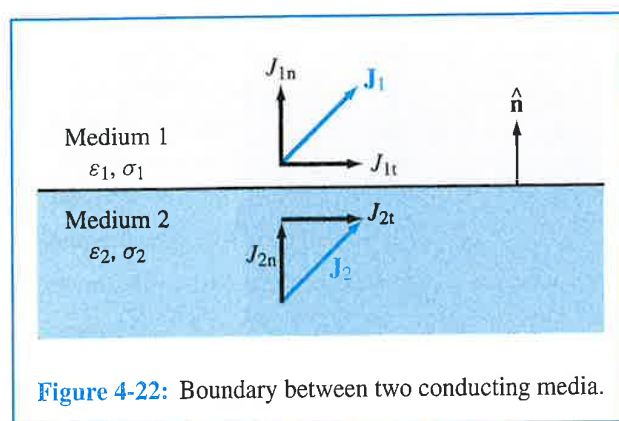


Figure 4-22: Boundary between two conducting media.

or perfect conductors. Depicted in Fig. 4-22, medium 1 has permittivity ϵ_1 and conductivity σ_1 , medium 2 has ϵ_2 and σ_2 , and the interface between them has a surface charge density ρ_s . For the electric fields, Eqs. (4.90) and (4.95) give

$$E_{1t} = E_{2t}, \quad \epsilon_1 E_{1n} - \epsilon_2 E_{2n} = \rho_s. \quad (4.102)$$

Since we are dealing with conducting media, the electric fields give rise to current densities $\mathbf{J}_1$ and $\mathbf{J}_2$, with $\mathbf{J}_1$ being proportional to $\mathbf{E}_1$ and $\mathbf{J}_2$ being proportional to $\mathbf{E}_2$. From $\mathbf{J} = \sigma \mathbf{E}$, we have

$$\frac{J_{1t}}{\sigma_1} = \frac{J_{2t}}{\sigma_2}, \quad \epsilon_1 \frac{J_{1n}}{\sigma_1} - \epsilon_2 \frac{J_{2n}}{\sigma_2} = \rho_s. \quad (4.103)$$

The tangential components J_{1t} and J_{2t} represent currents flowing in the two media in a direction parallel to the boundary, and hence no transfer of charge is involved between them. This is not the case for the normal components. If $J_{1n} \neq J_{2n}$, then a different amount of charge arrives at the boundary than leaves it. Hence, ρ_s cannot remain constant with time, which violates the condition of electrostatics requiring all fields and charges to remain constant. Consequently, *the normal component of $\mathbf{J}$ has to be continuous across the boundary between two differ-*

ent media under electrostatic conditions. Upon setting $J_{1n} = J_{2n}$ in Eq. (4.103), we have

$$J_{1n} \left(\frac{\epsilon_1}{\sigma_1} - \frac{\epsilon_2}{\sigma_2} \right) = \rho_s \quad (\text{electrostatics}). \quad (4.104)$$

REVIEW QUESTIONS

Q4.23 What are the boundary conditions for the electric field at a conductor–dielectric boundary?

Q4.24 Under electrostatic conditions, we require $J_{1n} = J_{2n}$ at the boundary between two conductors. Why?

4-10 Capacitance

When separated by an insulating (dielectric) medium, any two conducting bodies, regardless of their shapes and sizes, form a *capacitor*. If a d-c voltage source is connected to the conductors, as shown in Fig. 4-23 for two arbitrary conductors, charge of equal and opposite polarity is transferred to the conductors' surfaces. The surface of the conductor connected to the positive side of the source will accumulate charge $+Q$, and charge $-Q$ will accumulate on the surface of the other conductor. From our discussion in Section 4-7, *when a conductor has excess charge, it distributes the charge on its surface in such a manner as to maintain a zero electric field everywhere within the conductor*. This ensures that a conductor is an equipotential body, meaning that the electric potential is the same at every point in the conductor. *Capacitance* of a two-conductor capacitor is defined as

$$C = \frac{Q}{V} \quad (\text{C/V or F}), \quad (4.105)$$

where V is the potential (voltage) difference between the conductor with charge $+Q$ and the conductor with charge $-Q$. Capacitance is measured in farads (F), which is equivalent to coulombs per volt (C/V).

The presence of free charges on the conductors' surfaces gives rise to an electric field $\mathbf{E}$, as shown in Fig. 4-23; the field lines originate on the positive charges and terminate on the negative charges, and since the tangential component of $\mathbf{E}$ is always equal to zero at a conductor's surface, $\mathbf{E}$ is always perpendicular to the conducting surfaces. The normal component of $\mathbf{E}$ at any point on the surface of either conductor is given by

$$E_n = \hat{\mathbf{n}} \cdot \mathbf{E} = \frac{\rho_s}{\epsilon} \quad (\text{at conductor surface}), \quad (4.106)$$

where ρ_s is the surface charge density at that point, $\hat{\mathbf{n}}$ is the outward normal unit vector at the same location, and ϵ is the permittivity of the dielectric medium separating the conductors. The charge Q is equal to the integral of ρ_s over surface S [Fig. 4-23]:

$$Q = \int_S \rho_s ds = \int_S \epsilon \hat{\mathbf{n}} \cdot \mathbf{E} ds = \int_S \epsilon \mathbf{E} \cdot d\mathbf{s}, \quad (4.107)$$

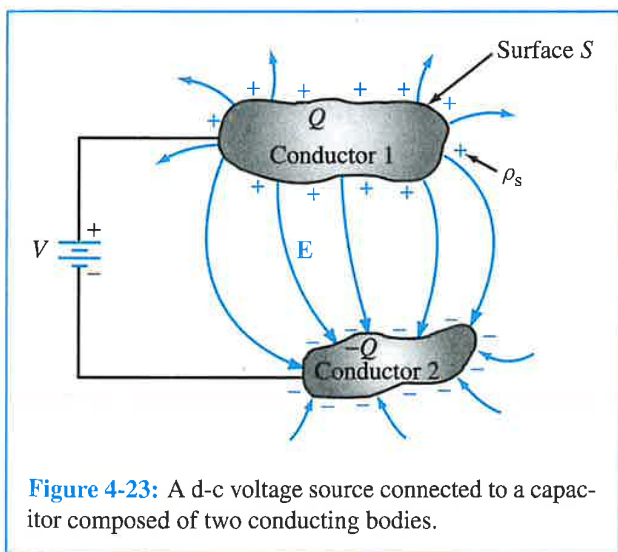


Figure 4-23: A d-c voltage source connected to a capacitor composed of two conducting bodies.

where use was made of Eq. (4.106). The voltage V is related to $\mathbf{E}$ by Eq. (4.39):

$$V = V_{12} = - \int_{P_2}^{P_1} \mathbf{E} \cdot d\mathbf{l}, \quad (4.108)$$

where points P_1 and P_2 are any two points on conductors 1 and 2, respectively. Substituting Eqs. (4.107) and (4.108) into Eq. (4.105) gives

$$C = \frac{\int_S \epsilon \mathbf{E} \cdot d\mathbf{s}}{- \int_l \mathbf{E} \cdot d\mathbf{l}} \quad (\text{F}), \quad (4.109)$$

where l is the integration path from conductor 2 to conductor 1. To avoid making sign errors when applying Eq. (4.109), it is important to remember that surface S is the $+Q$ surface and P_1 is on S . Because $\mathbf{E}$ appears in both the numerator and denominator of Eq. (4.109), *the value of C obtained for any specific capacitor configuration is always independent of $\mathbf{E}$* . In fact, C depends only on the capacitor geometry (sizes, shapes and relative positions of the two conductors) and the permittivity of the insulating material.

If the material between the conductors is not a perfect dielectric (i.e., if it has a small conductivity σ), then current can flow through the material between the conductors, and the material will exhibit a resistance R . The general expression for R for a resistor of arbitrary shape is given by Eq. (4.71):

$$R = \frac{- \int_l \mathbf{E} \cdot d\mathbf{l}}{\int_S \sigma \mathbf{E} \cdot d\mathbf{s}} \quad (\Omega). \quad (4.110)$$

For a medium with uniform σ and ϵ , the product of Eqs. (4.109) and (4.110) gives

$$RC = \frac{\epsilon}{\sigma}. \quad (4.111)$$

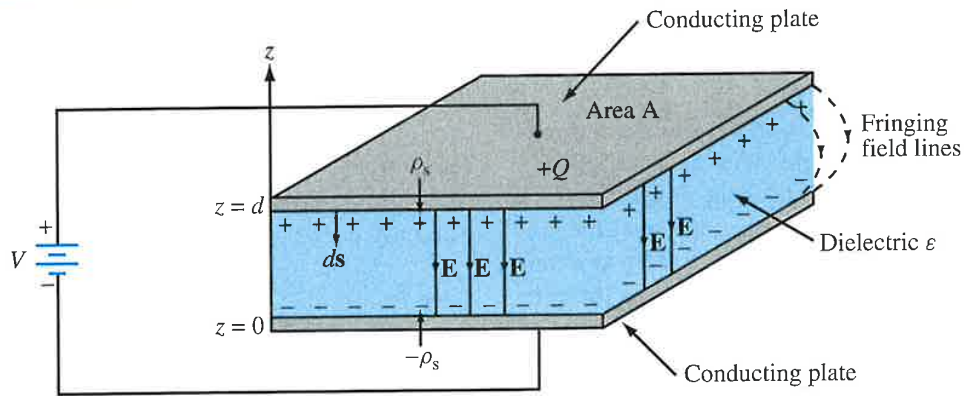


Figure 4-24: A d-c voltage source connected to a parallel-plate capacitor (Example 4-11).

This simple relation allows us to find R if C is known, or vice versa.

Example 4-11 Capacitance and Breakdown Voltage of Parallel-Plate Capacitor

Obtain an expression for the capacitance C of a parallel-plate capacitor comprised of two parallel plates each of surface area A and separated by a distance d . The capacitor is filled with a dielectric material with permittivity ϵ . Also, determine the breakdown voltage if $d = 1$ cm and the dielectric material is quartz.

Solution: In Fig. 4-24, we place the lower plate of the capacitor in the x - y plane and the upper plate in the plane $z = d$. Because of the applied voltage difference V , charge $+Q$ accumulates uniformly on the top plate and $-Q$ accumulates uniformly on the lower plate. In the dielectric medium between the plates, the charges induce a uniform electric field in the $-\hat{z}$ -direction (from positive to negative charges). In addition, some *fringing field lines* will exist near the edges, but their effects may be ignored if the dimensions of the plates are much larger than the separation d between them, because in that case the bulk of the electric field lines will exist in the medium

between the plates. The charge density on the upper plate is $\rho_s = Q/A$. Hence,

$$\mathbf{E} = -\hat{z}E,$$

and from Eq. (4.106), the magnitude of $\mathbf{E}$ at the conductor-dielectric boundary is $E = \rho_s/\epsilon = Q/\epsilon A$. From Eq. (4.108), the voltage difference is

$$V = -\int_0^d \mathbf{E} \cdot d\mathbf{l} = -\int_0^d (-\hat{z}E) \cdot \hat{z}dz = Ed, \quad (4.112)$$

and the capacitance is

$$C = \frac{Q}{V} = \frac{Q}{Ed} = \frac{\epsilon A}{d}, \quad (4.113)$$

where use was made of the relation $E = Q/\epsilon A$.

From $V = Ed$, as given by Eq. (4.112), $V = V_{br}$ when $E = E_{ds}$, the dielectric strength of the material. According to Table 4-2, $E_{ds} = 30$ (MV/m) for quartz. Hence, the breakdown voltage is

$$V_{br} = E_{ds}d = 30 \times 10^6 \times 10^{-2} = 3 \times 10^5 \text{ V.} \quad \square$$

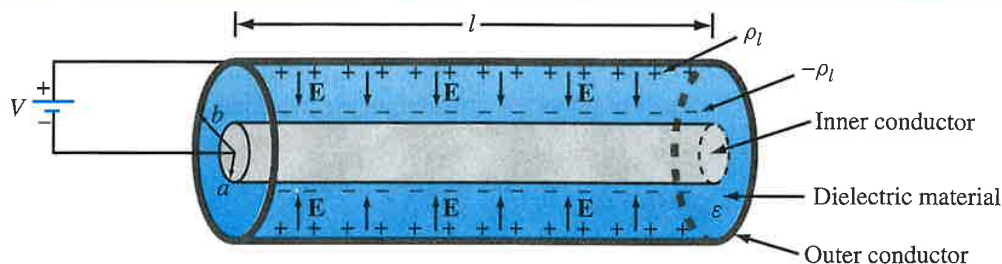


Figure 4-25: Coaxial capacitor filled with insulating material of permittivity ϵ (Example 4-12).

Example 4-12 Capacitance of Coaxial Line

Obtain an expression for the capacitance of the coaxial line shown in Fig. 4-25.

Solution: For a given voltage V across the capacitor (Fig. 4-25), charge $+Q$ and $-Q$ will accumulate on the surfaces of the outer and inner conductors, respectively. We assume that these charges are uniformly distributed along the length of the conductors with line charge density $\rho_l = Q/l$ on the outer conductor and $-\rho_l$ on the inner conductor. Ignoring fringing fields near the ends of the coaxial line, we can construct a cylindrical Gaussian surface in the dielectric, around the inner conductor, with radius r such that $a < r < b$. The inner conductor is a line charge similar to that of Example 4-6, except that the line charge of the inner conductor is negative. With a minus sign added to the expression for $\mathbf{E}$ given by Eq. (4.33), we have

$$\mathbf{E} = -\hat{\mathbf{r}} \frac{\rho_l}{2\pi\epsilon r} = -\hat{\mathbf{r}} \frac{Q}{2\pi\epsilon r l} \quad (4.114)$$

The potential difference V between the outer and inner conductors is

$$\begin{aligned} V &= -\int_a^b \mathbf{E} \cdot d\mathbf{l} = -\int_a^b \left(-\hat{\mathbf{r}} \frac{Q}{2\pi\epsilon r l} \right) \cdot (\hat{\mathbf{r}} dr) \\ &= \frac{Q}{2\pi\epsilon l} \ln\left(\frac{b}{a}\right). \end{aligned} \quad (4.115)$$

The capacitance C is then given by

$$C = \frac{Q}{V} = \frac{2\pi\epsilon l}{\ln(b/a)}. \quad (4.116)$$

The capacitance per unit length of the coaxial line is

$$C' = \frac{C}{l} = \frac{2\pi\epsilon}{\ln(b/a)} \quad (\text{F/m}). \quad (4.117)$$

REVIEW QUESTIONS

Q4.25 How is the capacitance of a two-conductor structure related to the resistance of the insulating material between the conductors?

Q4.26 What are fringing fields and when may they be ignored?

4-11 Electrostatic Potential Energy

When a source is connected to a capacitor, it expends energy in charging up the capacitor. If the capacitor plates are made of a good conductor with effectively zero resistance and if the dielectric separating the two conductors has negligible conductivity, then no real current can flow through the dielectric, and no ohmic losses occur anywhere in the capacitor. Where then does the charging-up energy go? The energy ends up getting stored in the dielectric medium in the form of **electrostatic potential energy**. The amount of stored energy W_e is related to Q , C , and V .

Under the influence of the electric field in the dielectric medium between the two conductors, charge q accumulates on one of the conductors, and an equal and opposite charge accumulates on the other conductor. In effect, charge q has been transferred from one of the conductors to the other. The voltage v across the capacitor is related to q by

$$v = \frac{q}{C}. \quad (4.118)$$

From the basic definition of the electric potential V , the amount of work dW_e required to transfer an additional incremental amount of charge dq is

$$dW_e = v dq = \frac{q}{C} dq. \quad (4.119)$$

If we start with an uncharged capacitor and charge it up from zero charge until a final charge Q has been reached, then the total amount of work performed is

$$W_e = \int_0^Q \frac{q}{C} dq = \frac{1}{2} \frac{Q^2}{C} \quad (\text{J}). \quad (4.120)$$

Using $C = Q/V$, where V is the final voltage, W_e can also be written as

$$W_e = \frac{1}{2} C V^2 \quad (\text{J}). \quad (4.121)$$

For the parallel-plate capacitor discussed in Example 4-11, its capacitance is given by Eq. (4.113) as

$C = \epsilon A/d$, where A is the surface area of each of its plates and d is the separation between them. Also, the voltage V across the capacitor is related to the magnitude of the electric field, E , in the dielectric by $V = Ed$. Using these two expressions in Eq. (4.121) gives

$$W_e = \frac{1}{2} \frac{\epsilon A}{d} (Ed)^2 = \frac{1}{2} \epsilon E^2 (Ad) = \frac{1}{2} \epsilon E^2 \nu, \quad (4.122)$$

where $\nu = Ad$ is the volume of the capacitor.

The **electrostatic energy density** w_e is defined as the electrostatic potential energy W_e per unit volume:

$$w_e = \frac{W_e}{\nu} = \frac{1}{2} \epsilon E^2 \quad (\text{J/m}^3). \quad (4.123)$$

Even though this expression was derived for a parallel-plate capacitor, it is equally valid for any dielectric medium in an electric field E . Furthermore, for any volume ν containing a dielectric ϵ , the total electrostatic potential energy stored in ν is

$$W_e = \frac{1}{2} \int_{\nu} \epsilon E^2 d\nu \quad (\text{J}). \quad (4.124)$$

Returning to the parallel-plate capacitor, the oppositely charged plates are attracted to each other by an electrical force $\mathbf{F}$. The force acting on any system of charges may be obtained from energy considerations. In the material that follows, we will show how $\mathbf{F}$ can be determined from W_e , the electrostatic energy stored in the system by virtue of the presence of the electric charges. The treatment will be general in nature, although we will occasionally use the capacitor example by way of illustration.

If the two plates of the parallel-plate capacitor are allowed to move closer to each other, under the influence of the electric force $\mathbf{F}$, by a differential distance $d\mathbf{l}$, while maintaining the charges on the plates constant, then the mechanical work done by the system (charged capacitor) is

$$dW = \mathbf{F} \cdot d\mathbf{l}. \quad (4.125)$$

For an isolated system, this mechanical work is provided by *expending* electrostatic energy. Hence, dW is equal to the *loss* of energy stored in the dielectric insulating material of the capacitor, or

$$dW = -dW_e. \quad (4.126)$$

From Eq. (3.73), dW_e may be written in terms of the gradient of W_e as

$$dW_e = \nabla W_e \cdot d\mathbf{l}. \quad (4.127)$$

In view of Eq. (4.126), comparison of Eqs. (4.125) and (4.127) leads to

$$\mathbf{F} = -\nabla W_e \quad (\text{N}). \quad (4.128)$$

It is important to note that Eq. (4.128) is predicated on the assumption that the charges in the system are constant.

To apply Eq. (4.128) to the parallel-plate capacitor, we rewrite Eq. (4.120) in the following form:

$$W_e = \frac{1}{2} \frac{Q^2}{C} = \frac{Q^2 z}{2\epsilon A}, \quad (4.129)$$

where we have replaced d with the variable z , representing the vertical spacing between the conducting plates. Use of Eq. (4.129) in Eq. (4.128) gives

$$\mathbf{F} = -\nabla W_e = -\hat{\mathbf{z}} \frac{\partial}{\partial z} \left(\frac{Q^2 z}{2\epsilon A} \right) = -\hat{\mathbf{z}} \left(\frac{Q^2}{2\epsilon A} \right), \quad (4.130)$$

and since $Q = \epsilon A E$, $\mathbf{F}$ can also be written as

$$\mathbf{F} = -\hat{\mathbf{z}} \frac{\epsilon A E^2}{2} \quad (\text{parallel-plate capacitor}). \quad (4.131)$$

REVIEW QUESTIONS

Q4.27 To bring a charge q from infinity to a given point in space, a certain amount of work W is expended. Where does the energy corresponding to W go?

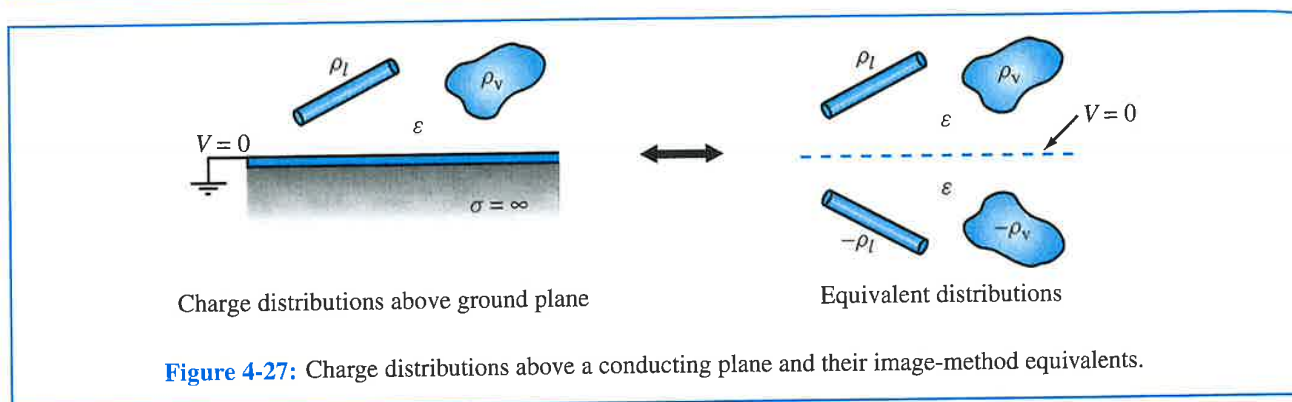
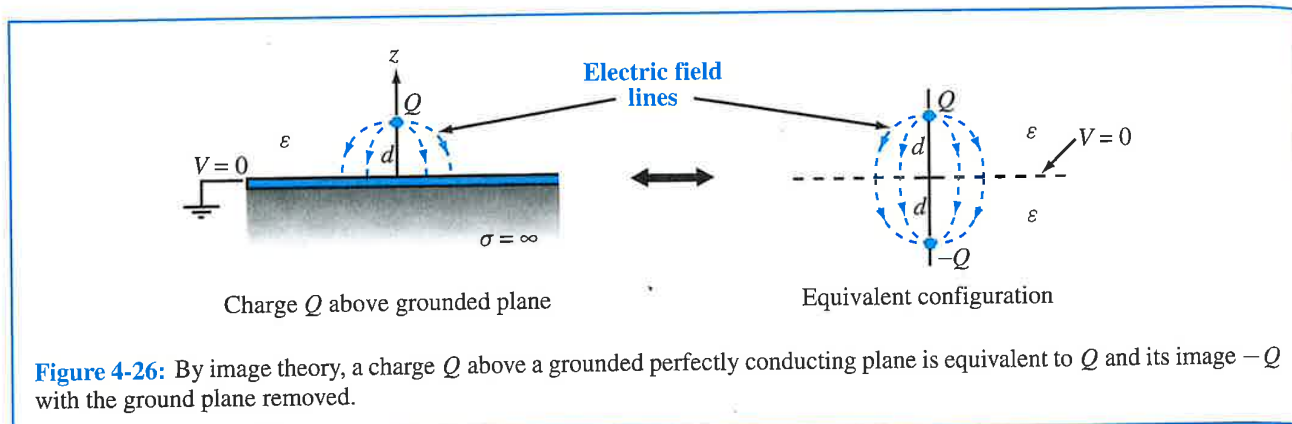
Q4.28 When a voltage source is connected across a capacitor, what is the direction of the electrical force acting on its two conducting surfaces?

EXERCISE 4.18 The radii of the inner and outer conductors of a coaxial cable are 2 cm and 5 cm, respectively, and the insulating material between them has a relative permittivity of 4. The charge density on the outer conductor is $\rho_l = 10^{-4}$ (C/m). Use the expression for $\mathbf{E}$ derived in Example 4-12 to calculate the total energy stored in a 20-cm length of the cable.

Ans. $W_e = 4.1$ J.

4-12 Image Method

Consider a point charge Q at a distance d above a perfectly conducting plane, as shown in the left-hand section of Fig. 4-26. We want to determine V , $\mathbf{E}$, and $\mathbf{D}$ at any point in the space above the grounded conductor, as well as the distribution of surface charge on the conducting plate. Three different methods have been introduced in this chapter for finding $\mathbf{E}$. The first, based on Coulomb's law, requires knowledge of the magnitudes and locations of all the charges contributing to $\mathbf{E}$ at a given point in space. In the present case, the charge Q will induce an unknown and nonuniform distribution of charge on the surface of the conductor. Hence, we cannot utilize Coulomb's method. The second method is based on the application of Gauss's law, and it is equally difficult to use because it is not clear how one would construct a Gaussian surface such that $\mathbf{E}$ is always totally tangential or totally normal at every point on that surface. In the third method, the electric field is found from $\mathbf{E} = -\nabla V$ after solving Poisson's or Laplace's equation for V , subject to the available boundary conditions; that is, $V = 0$ at any point on the grounded conducting surface and at infinity. Although such an approach is feasible in principle, the solution is quite complicated mathematically.



Alternatively, the problem at hand can be solved with great ease using **image theory**, which states that *any given charge configuration above an infinite, perfectly conducting plane is electrically equivalent to the combination of the given charge configuration and its image configuration, with the conducting plane removed*. The image-method equivalent of the charge Q above a conducting plane is shown in the right-hand section of Fig. 4-26. It consists of the charge Q itself and an image charge $-Q$ at a distance $2d$ from Q , with nothing else between them. The electric field due to the two isolated charges can now be easily found at any point (x, y, z) by applying Coulomb's method, as demonstrated by Exam-

ple 4-13. The combination of the two charges will always produce a potential $V = 0$ at every point in the plane where the conducting surface had been. If the charge is in the presence of more than one grounded plane, it is necessary to establish images of the charge relative to each of the planes and then to establish images of each of those images against the remaining planes. The process is continued until the condition $V = 0$ is satisfied at all points on all the grounded planes. The image method applies not only to point charges, but also to any distributions of charge, such as the line and volume distributions depicted in Fig. 4-27.

Example 4-13 Image Method for Charge above Conducting Plane

Use image theory to determine V and $\mathbf{E}$ at an arbitrary point $P(x, y, z)$ in the region $z > 0$ due to a charge Q in free space at a distance d above a grounded conducting plane.

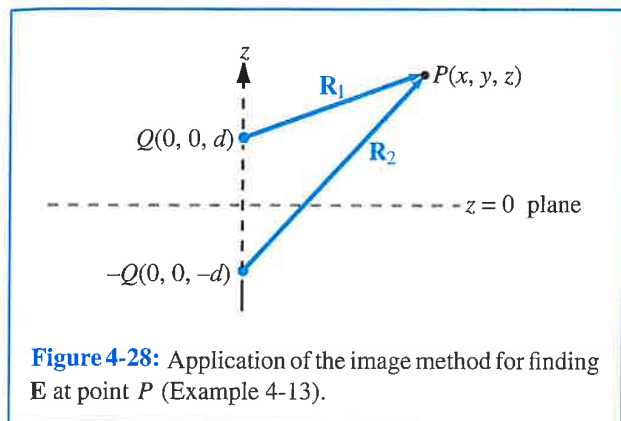


Figure 4-28: Application of the image method for finding $\mathbf{E}$ at point P (Example 4-13).

Solution: In Fig. 4-28, charge Q is at $(0, 0, d)$ and its image $-Q$ is at $(0, 0, -d)$ in Cartesian coordinates. From Eq. (4.19), the electric field at point $P(x, y, z)$ due to the two charges is given by

$$\begin{aligned}\mathbf{E} &= \frac{1}{4\pi\epsilon_0} \left(\frac{Q\mathbf{R}_1}{R_1^3} + \frac{-Q\mathbf{R}_2}{R_2^3} \right) \\ &= \frac{Q}{4\pi\epsilon_0} \left[\frac{\hat{\mathbf{x}}x + \hat{\mathbf{y}}y + \hat{\mathbf{z}}(z-d)}{[x^2 + y^2 + (z-d)^2]^{3/2}} \right. \\ &\quad \left. - \frac{\hat{\mathbf{x}}x + \hat{\mathbf{y}}y + \hat{\mathbf{z}}(z+d)}{[x^2 + y^2 + (z+d)^2]^{3/2}} \right]\end{aligned}$$

for $z \geq 0$. ■

EXERCISE 4.19 Use the result of Example 4-13 to find the surface charge density ρ_s on the surface of the conducting plane.

Ans. $\rho_s = -Qd/[2\pi(x^2 + y^2 + d^2)^{3/2}]$.

REVIEW QUESTIONS

Q4.29 What is the fundamental premise of the image method?

Q4.30 Given a certain charge distribution, what are the various approaches described in this chapter for computing the electric field $\mathbf{E}$ at a given point in space?

CHAPTER HIGHLIGHTS

- Maxwell's equations are the fundamental tenets of electromagnetic theory.
- Under static conditions, Maxwell's equations separate into two uncoupled pairs, with one pair pertaining to electrostatics and the other to magnetostatics.
- Coulomb's law provides an explicit expression for the electric field due to a specified charge distribution.
- Gauss's law states that the total electric field flux through a closed surface is equal to the net charge enclosed by the surface.
- The electrostatic field $\mathbf{E}$ at a point is related to the electric potential V at that point by $\mathbf{E} = -\nabla V$, with V being referenced to zero potential at infinity.
- Because most metals have conductivities on the order of 10^6 (S/m), they are treated in practice as perfect conductors. By the same token, insulators with conductivities smaller than 10^{-10} (S/m) are treated as perfect dielectrics.
- Boundary conditions at the interface between two materials specify the relations between the normal and tangential components of $\mathbf{D}$, $\mathbf{E}$, and $\mathbf{J}$ in one of the materials to the corresponding components in the other.

- The capacitance of a two-conductor body and resistance of the medium between them can be computed from knowledge of the electric field in that medium.
- The electrostatic energy density stored in a dielectric medium is $w_e = \frac{1}{2}\epsilon E^2$ (J/m³).
- When a charge configuration exists above an infinite, perfectly conducting plane, the induced field **E** is the same as that due to the configuration itself and its image with the conducting plane removed.

GLOSSARY OF IMPORTANT TERMS

Provide definitions or explain the meaning of the following terms:

static condition
 electrostatics
 volume, surface, and line charge densities
 current density **J**
 conduction current
 convection current
 electric field intensity **E**
 electric flux density **D**
 Coulomb's law
 Gauss's law
 Gaussian surface
 conservative field
 electric potential
 electric dipole
 dipole moment **p**
 constitutive parameters
 conductivity σ
 conductor
 dielectric material
 semiconductor
 superconductor
 electron drift velocity **u_e**
 hole drift velocity **u_h**
 perfect conductor
 perfect dielectric

electron mobility μ_e
 hole mobility μ_h
 Ohm's law
 conductance G
 Joule's law
 boundary conditions
 polarization vector **P**
 linear material
 isotropic material
 homogeneous material
 electric susceptibility χ_e
 permittivity ϵ
 relative permittivity ϵ_r
 dielectric strength
 dielectric breakdown
 capacitance C
 electrostatic potential energy W_e
 electrostatic energy density w_e
 image method

PROBLEMS

Sections 4-2: Charge and Current Distributions

4.1* A cube 2 m on a side is located in the first octant in a Cartesian coordinate system, with one of its corners at the origin. Find the total charge contained in the cube if the charge density is given by $\rho_v = xy^2e^{-2z}$ (mC/m³).

4.2 Find the total charge contained in a cylindrical volume defined by $r \leq 2$ m and $0 \leq z \leq 3$ m if $\rho_v = 10rz$ (mC/m³).

4.3* Find the total charge contained in a cone defined by $R \leq 2$ m and $0 \leq \theta \leq \pi/4$, given that $\rho_v = 20R^2 \cos^2 \theta$ (mC/m³).

4.4 If the line charge density is given by $\rho_l = 12y^2$ (mC/m), find the total charge distributed on the y -axis from $y = -5$ to $y = 5$.

* Answer(s) available in Appendix D.

 Solution available in CD-ROM.

4.5* Find the total charge on a circular disk defined by $r \leq a$ and $z = 0$ if:

(a) $\rho_s = \rho_{s0} \sin \phi$ (C/m²)

(b) $\rho_s = \rho_{s0} \sin^2 \phi$ (C/m²)

(c) $\rho_s = \rho_{s0} e^{-r}$ (C/m²)

(d) $\rho_s = \rho_{s0} e^{-r} \sin^2 \phi$ (C/m²)

where ρ_{s0} is a constant.

4.6 If $\mathbf{J} = \hat{\mathbf{y}}2xz$ (A/m²), find the current I flowing through a square with corners at (0, 0, 0), (2, 0, 0), (2, 0, 2), and (0, 0, 2).

4.7* If $\mathbf{J} = \hat{\mathbf{R}}25/R$ (A/m²), find I through the surface $R = 5$ m.

4.8 An electron beam shaped like a circular cylinder of radius r_0 carries a charge density given by

$$\rho_v = \left(\frac{-\rho_0}{1+r^2} \right) \quad (\text{C/m}^3)$$

where ρ_0 is a positive constant and the beam's axis is coincident with the z -axis.

(a) Determine the total charge contained in length L of the beam.

(b) If the electrons are moving in the $+z$ -direction with uniform speed u , determine the magnitude and direction of the current crossing the z -plane.

Section 4-3: Coulomb's Law

4.9* A square with sides of 2 m has a charge of $20 \mu\text{C}$ at each of its four corners. Determine the electric field at a point 5 m above the center of the square.

4.10 Three point charges, each with $q = 3 \text{ nC}$, are located at the corners of a triangle in the x - y plane, with one corner at the origin, another at (2 cm, 0, 0), and the third at (0, 2 cm, 0). Find the force acting on the charge located at the origin.

4.11* Charge $q_1 = 4 \mu\text{C}$ is located at (1 cm, 1 cm, 0) and charge q_2 is located at (0, 0, 4 cm). What should q_2 be so that $\mathbf{E}$ at (0, 2 cm, 0) has no y -component?

4.12 A line of charge with uniform density $\rho_l = 4 (\mu\text{C/m})$ exists in air along the z -axis between $z = 0$ and $z = 5$ cm. Find $\mathbf{E}$ at (0, 10 cm, 0).

4.13* Electric charge is distributed along an arc located in the x - y plane and defined by $r = 2$ cm and $0 \leq \phi \leq \pi/4$. If $\rho_l = 5 (\mu\text{C/m})$, find $\mathbf{E}$ at (0, 0, z) and then evaluate it at:

(a) The origin.

(b) $z = 5$ cm

(c) $z = -5$ cm

4.14 A line of charge with uniform density ρ_l extends between $z = -L/2$ and $z = L/2$ along the z -axis. Apply Coulomb's law to obtain an expression for the electric field at any point $P(r, \phi, 0)$ on the x - y plane. Show that your result reduces to the expression given by (4.33) as the length L is extended to infinity.

4.15* Repeat Example 4-5 for the circular disk of charge of radius a , but in the present case, assume the surface charge density to vary with r as

$$\rho_s = \rho_{s0} r^2 \quad (\text{C/m}^2)$$

where ρ_{s0} is a constant.

4.16 Multiple charges at different locations are said to be in equilibrium if the force acting on any one of them is identical in magnitude and direction to the force acting on any of the others. Suppose we have two negative charges, one located at the origin and carrying charge $-9e$, and the other located on the positive x -axis at a distance d from the first one and carrying charge $-36e$. Determine the location, polarity and magnitude of a third charge whose placement would bring the entire system into equilibrium.

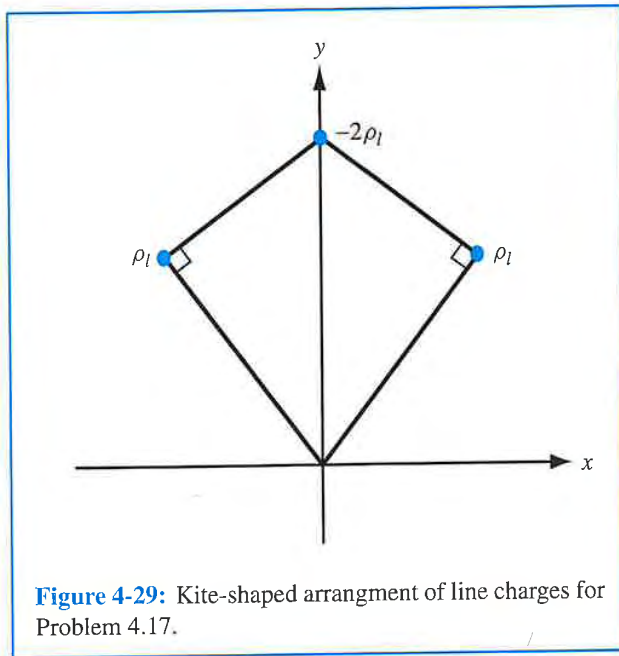


Figure 4-29: Kite-shaped arrangement of line charges for Problem 4.17.

4.17* Three infinite lines of charge, all parallel to the z -axis, are located at the three corners of the kite-shaped arrangement shown in Fig. 4-29. If the two right triangles are symmetrical and of equal corresponding sides, show that the electric field is zero at the origin.

4.18 Three infinite lines of charge, $\rho_{l1} = 5$ (nC/m), $\rho_{l2} = -5$ (nC/m), and $\rho_{l3} = 5$ (nC/m), are all parallel to the z -axis. If they pass through the respective points $(0, -b)$, $(0, 0)$, and $(0, b)$ in the x - y plane, find the electric field at $(a, 0, 0)$. Evaluate your result for $a = 2$ cm and $b = 1$ cm.

4.19 A horizontal strip lying in the x - y plane is of width d in the y -direction and infinitely long in the x -direction. If the strip is in air and has a uniform charge distribution ρ_s , use Coulomb's law to obtain an explicit expression for the electric field at a point P located at a distance h above the centerline of the strip. Extend your result to the special case where d is infinite and compare it with Eq. (4.25).

Section 4-4: Gauss's Law

4.20 Given the electric flux density

$$\mathbf{D} = \hat{x}2(x + y) + \hat{y}(3x - 2y) \quad (\text{C/m}^2)$$

determine

- ρ_v by applying Eq. (4.26).
- The total charge Q enclosed in a cube 2 m on a side, located in the first octant with three of its sides coincident with the x -, y -, and z -axes and one of its corners at the origin.
- The total charge Q in the cube, obtained by applying Eq. (4.29).

4.21* Repeat Problem 4.20 for $\mathbf{D} = \hat{x}xy^2z^3$ (C/m²).

4.22 Charge Q_1 is uniformly distributed over a thin spherical shell of radius a , and charge Q_2 is uniformly distributed over a second spherical shell of radius b , with $b > a$. Apply Gauss's law to find $\mathbf{E}$ in the regions $R < a$, $a < R < b$, and $R > b$.

4.23^{ss} The electric flux density inside a dielectric sphere of radius a centered at the origin is given by

$$\mathbf{D} = \hat{\mathbf{R}}\rho_0 R \quad (\text{C/m}^2)$$

where ρ_0 is a constant. Find the total charge inside the sphere.

4.24 In a certain region of space, the charge density is given in cylindrical coordinates by the function:

$$\rho_v = 20re^{-r} \quad (\text{C/m}^3)$$

Apply Gauss's law to find $\mathbf{D}$.

4.25* An infinitely long cylindrical shell extending between $r = 1$ m and $r = 3$ m contains a uniform charge density ρ_{v0} . Apply Gauss's law to find $\mathbf{D}$ in all regions.

4.26 If the charge density increases linearly with distance from the origin such that $\rho_v = 0$ at the origin and $\rho_v = 10$ C/m³ at $R = 2$ m, find the corresponding variation of $\mathbf{D}$.

Section 4-5: Electric Potential

4.27 A square in the x - y plane in free space has a point charge of $+Q$ at corner $(a/2, a/2)$, the same at corner $(a/2, -a/2)$, and a point charge of $-Q$ at each of the other two corners.

- Find the electric potential at any point P along the x -axis.
- Evaluate V at $x = a/2$.

4.28 The circular disk of radius a shown in Fig. 4-7 has uniform charge density ρ_s across its surface.

- Obtain an expression for the electric potential V at a point $P(0, 0, z)$ on the z -axis.
- Use your result to find $\mathbf{E}$ and then evaluate it for $z = h$. Compare your final expression with (4.24), which was obtained on the basis of Coulomb's law.

4.29* A circular ring of charge of radius a lies in the x - y plane and is centered at the origin. Assume also that the ring is in air and carries a uniform density ρ_l .

- Show that the electrical potential at $(0, 0, z)$ is given by $V = \rho_l a / [2\epsilon_0(a^2 + z^2)^{1/2}]$.
- Find the corresponding electric field $\mathbf{E}$.

4.30 Show that the electric potential difference V_{12} between two points in air at radial distances r_1 and r_2 from an infinite line of charge with density ρ_l along the z -axis is $V_{12} = (\rho_l / 2\pi\epsilon_0) \ln(r_2/r_1)$.

 **4.31*** Find the electric potential V at a location a distance b from the origin in the x - y plane due to a line charge with charge density ρ_l and of length l . The line charge is coincident with the z -axis and extends from $z = -l/2$ to $z = l/2$.

4.32 For the electric dipole shown in Fig. 4-13, $d = 1$ cm and $|\mathbf{E}| = 2$ (mV/m) at $R = 1$ m and $\theta = 0^\circ$. Find $\mathbf{E}$ at $R = 2$ m and $\theta = 90^\circ$.

4.33 For each of the distributions of the electric potential V shown in Fig. 4-30, sketch the corresponding distribution of $\mathbf{E}$ (in all cases, the vertical axis is in volts and the horizontal axis is in meters).

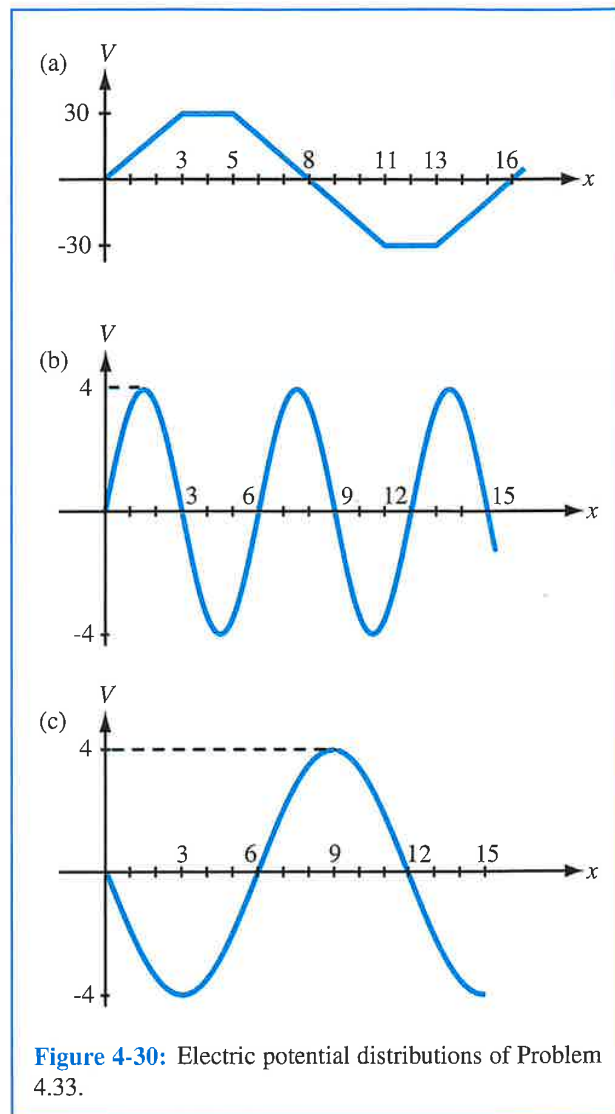


Figure 4-30: Electric potential distributions of Problem 4.33.

4.34 Given the electric field

$$\mathbf{E} = \hat{\mathbf{R}} \frac{12}{R^2} \quad (\text{V/m})$$

find the electric potential of point A with respect to point B where A is at $+2$ m and B at -4 m, both on the z -axis.

4.35* An infinitely long line of charge with uniform density $\rho_l = 6$ (nC/m) lies in the x - y plane parallel to the y -axis at $x = 2$ m. Find the potential V_{AB} at point $A(3 \text{ m}, 0, 4 \text{ m})$ in Cartesian coordinates with respect to point $B(0, 0, 0)$ by applying the result of Problem 4.30.

4.36 The x - y plane contains a uniform sheet of charge with $\rho_{s1} = 0.2$ (nC/m²). A second sheet with $\rho_{s2} = -0.2$ (nC/m²) occupies the plane $z = 6$ m. Find V_{AB} , V_{BC} , and V_{AC} for $A(0, 0, 6 \text{ m})$, $B(0, 0, 0)$, and $C(0, -2 \text{ m}, 2 \text{ m})$.

Section 4-7: Conductors

4.37* A cylindrical bar of silicon has a radius of 2 mm and a length of 5 cm. If a voltage of 5 V is applied between the ends of the bar and $\mu_e = 0.13$ (m²/V·s), $\mu_h = 0.05$ (m²/V·s), $N_e = 1.5 \times 10^{16}$ electrons/m³, and $N_h = N_e$, find the following:

- The conductivity of silicon.
- The current I flowing in the bar.
- The drift velocities $\mathbf{u}_e$ and $\mathbf{u}_h$.
- The resistance of the bar.
- The power dissipated in the bar.

4.38 Repeat Problem 4.37 for a bar of germanium with $\mu_e = 0.4$ (m²/V·s), $\mu_h = 0.2$ (m²/V·s), and $N_e = N_h = 2.4 \times 10^{19}$ electrons or holes/m³.

4.39 A 100-m-long conductor of uniform cross-section has a voltage drop of 2 V between its ends. If the density of the current flowing through it is 7×10^5 (A/m²), identify the material of the conductor.

4.40 A coaxial resistor of length l consists of two concentric cylinders. The inner cylinder has radius a and is made of a material with conductivity σ_1 , and the outer cylinder, extending between $r = a$ and $r = b$, is made of a material with conductivity σ_2 . If the two ends of the resistor are capped with conducting plates, show that the resistance between the two ends is $R = l/[\pi(\sigma_1 a^2 + \sigma_2(b^2 - a^2))]$.

4.41* Apply the result of Problem 4.40 to find the resistance of a 10-cm-long hollow cylinder (Fig. 4-31) made of carbon with $\sigma = 3 \times 10^4$ (S/m).

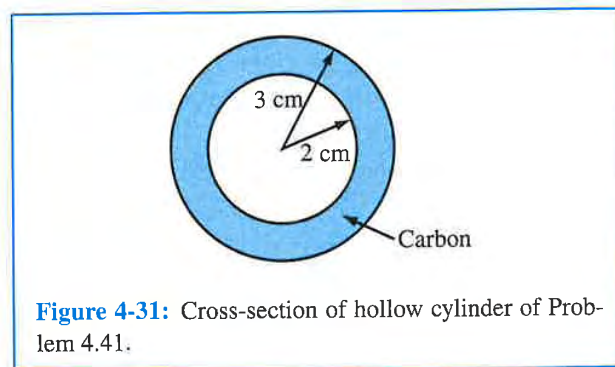


Figure 4-31: Cross-section of hollow cylinder of Problem 4.41.

4.42 A 2×10^{-3} -mm-thick square sheet of aluminum has 10 cm $\times$ 10 cm faces. Find the following:

- The resistance between opposite edges on a square face.
- The resistance between the two square faces. (See Appendix B for the electrical constants of materials.)

Section 4-9: Boundary Conditions

4.43* With reference to Fig. 4-19, find $\mathbf{E}_1$ if $\mathbf{E}_2 = \hat{x}3 - \hat{y}2 + \hat{z}4$ (V/m), $\epsilon_1 = 2\epsilon_0$, $\epsilon_2 = 18\epsilon_0$, and the boundary has a surface charge density $\rho_s = 7.08 \times 10^{-11}$ (C/m²). What angle does $\mathbf{E}_2$ make with the z -axis?

4.44 An infinitely long dielectric cylinder with $\epsilon_{1r} = 4$ and described by $r \leq 10$ cm is surrounded by a material with $\epsilon_{2r} = 8$. If $\mathbf{E}_1 = \hat{r}r^2 - \hat{\phi}r \cos \phi + \hat{z}3$ (V/m) in the cylinder region, find $\mathbf{E}_2$ and $\mathbf{D}_2$ in the surrounding region. Assume that no free charges exist along the cylinder's boundary.

4.45* A 2-cm dielectric sphere with $\epsilon_{1r} = 3$ is embedded in a medium with $\epsilon_{2r} = 9$. If $\mathbf{E}_2 = \hat{R}2 \cos \theta - \hat{\theta}3 \sin \theta$ (V/m) in the surrounding region, find $\mathbf{E}_1$ and $\mathbf{D}_1$ in the sphere.

4.46 If $\mathbf{E} = \hat{R}50$ (V/m) at the surface of a 5-cm conducting sphere centered at the origin, what is the total charge Q on the sphere's surface?

4.47* Figure 4-32 shows three planar dielectric slabs of equal thickness but with different dielectric constants. If $\mathbf{E}_0$ in air makes an angle of 45° with respect to the z -axis, find the angle of $\mathbf{E}$ in each of the other layers.

Sections 4-10 and 4-11: Capacitance and Electrical Energy

4.48 Determine the force of attraction in a parallel-plate capacitor with $A = 10$ cm², $d = 1$ cm, and $\epsilon_r = 4$ if the voltage across it is 50 V.

4.49* Dielectric breakdown occurs in a material whenever the magnitude of the field $\mathbf{E}$ exceeds the dielectric strength anywhere in that material. In the coaxial capacitor of Example 4-12,

- At what value of r is $|\mathbf{E}|$ maximum?
- What is the breakdown voltage if $a = 1$ cm, $b = 2$ cm, and the dielectric material is mica with $\epsilon_r = 6$?

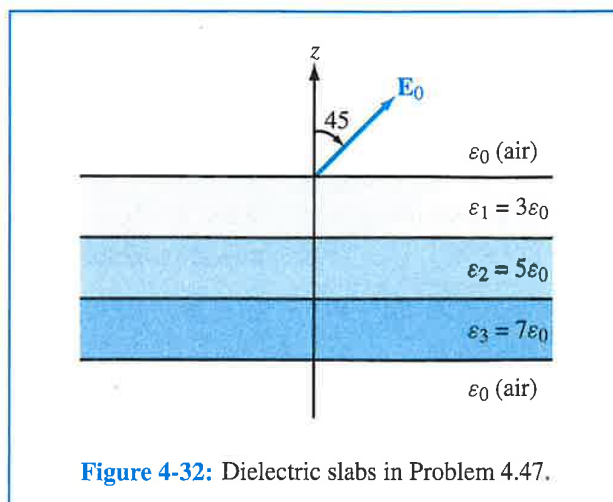


Figure 4-32: Dielectric slabs in Problem 4.47.

4.50 An electron with charge $Q_e = -1.6 \times 10^{-19}$ C and mass $m_e = 9.1 \times 10^{-31}$ kg is injected at a point adjacent to the negatively charged plate in the region between the plates of an air-filled parallel-plate capacitor with separation of 1 cm and rectangular plates each 10 cm² in area (Fig. 4-33). If the voltage across the capacitor is 10 V, find the following:

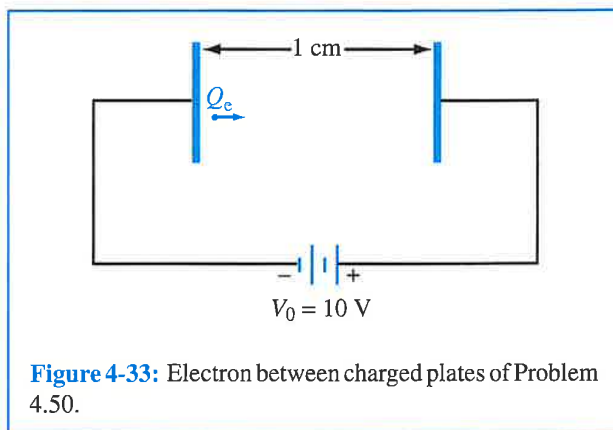


Figure 4-33: Electron between charged plates of Problem 4.50.

- (a) The force acting on the electron.
- (b) The acceleration of the electron.
- (c) The time it takes the electron to reach the positively charged plate, assuming that it starts from rest.

4.51* In a dielectric medium with $\epsilon_r = 4$, the electric field is given by

$$\mathbf{E} = \hat{\mathbf{x}}(x^2 + 2z) + \hat{\mathbf{y}}x^2 - \hat{\mathbf{z}}(y + z) \quad (\text{V/m})$$

Calculate the electrostatic energy stored in the region $-1 \text{ m} \leq x \leq 1 \text{ m}$, $0 \leq y \leq 2 \text{ m}$, and $0 \leq z \leq 3 \text{ m}$.

4.52 Figure 4-34(a) depicts a capacitor consisting of two parallel, conducting plates separated by a distance d .

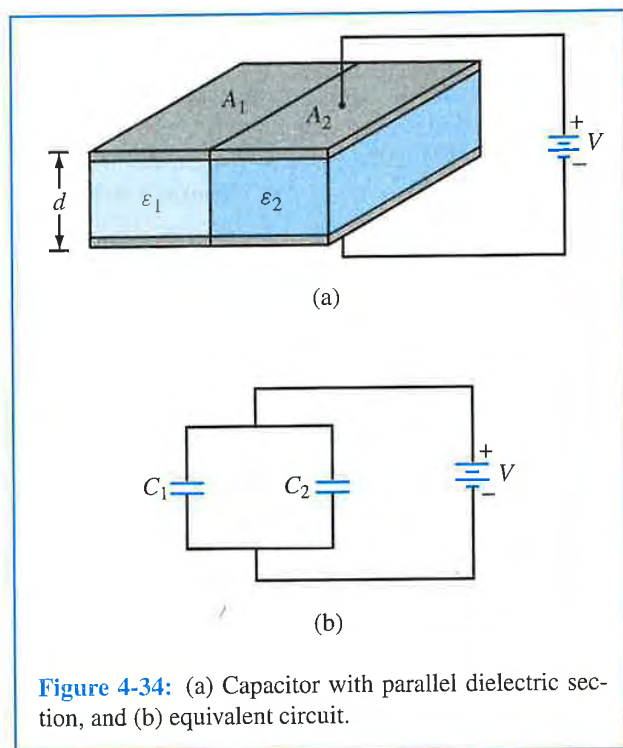


Figure 4-34: (a) Capacitor with parallel dielectric section, and (b) equivalent circuit.

The space between the plates contains two adjacent dielectrics, one with permittivity ϵ_1 and surface area A_1 and another with ϵ_2 and A_2 . The objective of this problem is to show that the capacitance C of the configuration shown in Fig. 4-34(a) is equivalent to two capacitances in parallel, as illustrated in Fig. 4-34(b), with

$$C = C_1 + C_2 \quad (4.132)$$

where

$$C_1 = \frac{\epsilon_1 A_1}{d} \quad (4.133)$$

$$C_2 = \frac{\epsilon_2 A_2}{d} \quad (4.134)$$

To this end, proceed as follows:

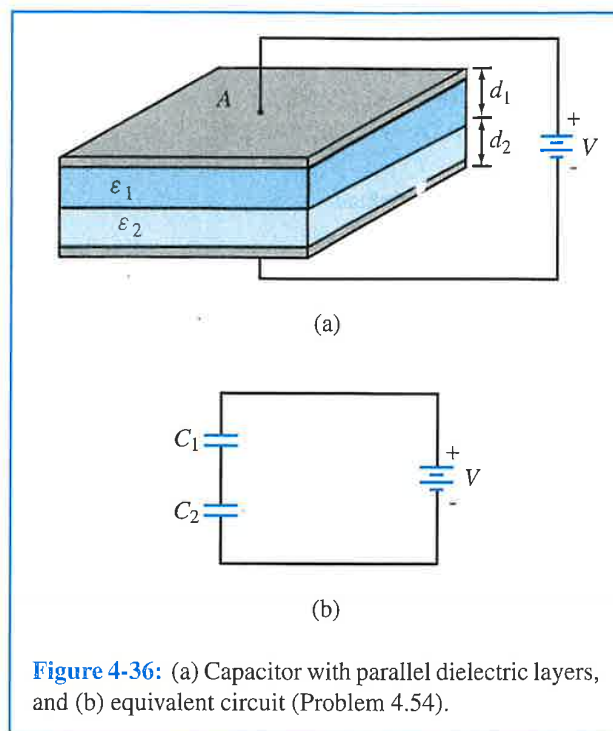
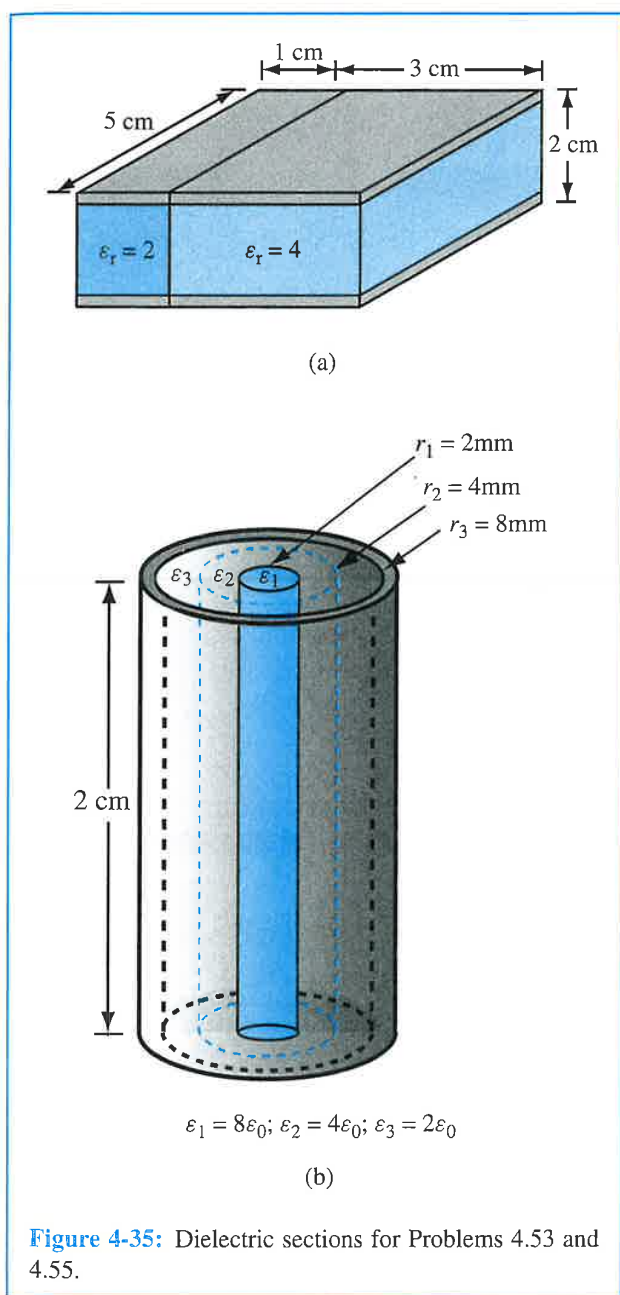
- (a) Find the electric fields $\mathbf{E}_1$ and $\mathbf{E}_2$ in the two dielectric layers.
- (b) Calculate the energy stored in each section and use the result to calculate C_1 and C_2 .
- (c) Use the total energy stored in the capacitor to obtain an expression for C . Show that (4.132) is indeed a valid result.

4.53* Use the result of Problem 4.52 to determine the capacitance for each of the following configurations:

- (a) Conducting plates are on top and bottom faces of the rectangular structure in Fig. 4-35(a).
- (b) Conducting plates are on front and back faces of the structure in Fig. 4-35(a).
- (c) Conducting plates are on top and bottom faces of the cylindrical structure in Fig. 4-35(b).

4.54 The capacitor shown in Fig. 4-36 consists of two parallel dielectric layers. Use energy considerations to show that the equivalent capacitance of the overall capacitor, C , is equal to the series combination of the capacitances of the individual layers, C_1 and C_2 , namely

$$C = \frac{C_1 C_2}{C_1 + C_2} \quad (4.135)$$



where

$$C_1 = \epsilon_1 \frac{A}{d_1}, \quad C_2 = \epsilon_2 \frac{A}{d_2}$$

- (a) Let V_1 and V_2 be the electric potentials across the upper and lower dielectrics, respectively. What are the corresponding electric fields E_1 and E_2 ? By applying the appropriate boundary condition at the interface between the two dielectrics, obtain explicit expressions for E_1 and E_2 in terms of ϵ_1 , ϵ_2 , V , and the indicated dimensions of the capacitor.
- (b) Calculate the energy stored in each of the dielectric layers and then use the sum to obtain an expression for C .
- (c) Show that C is given by Eq. (4.135).