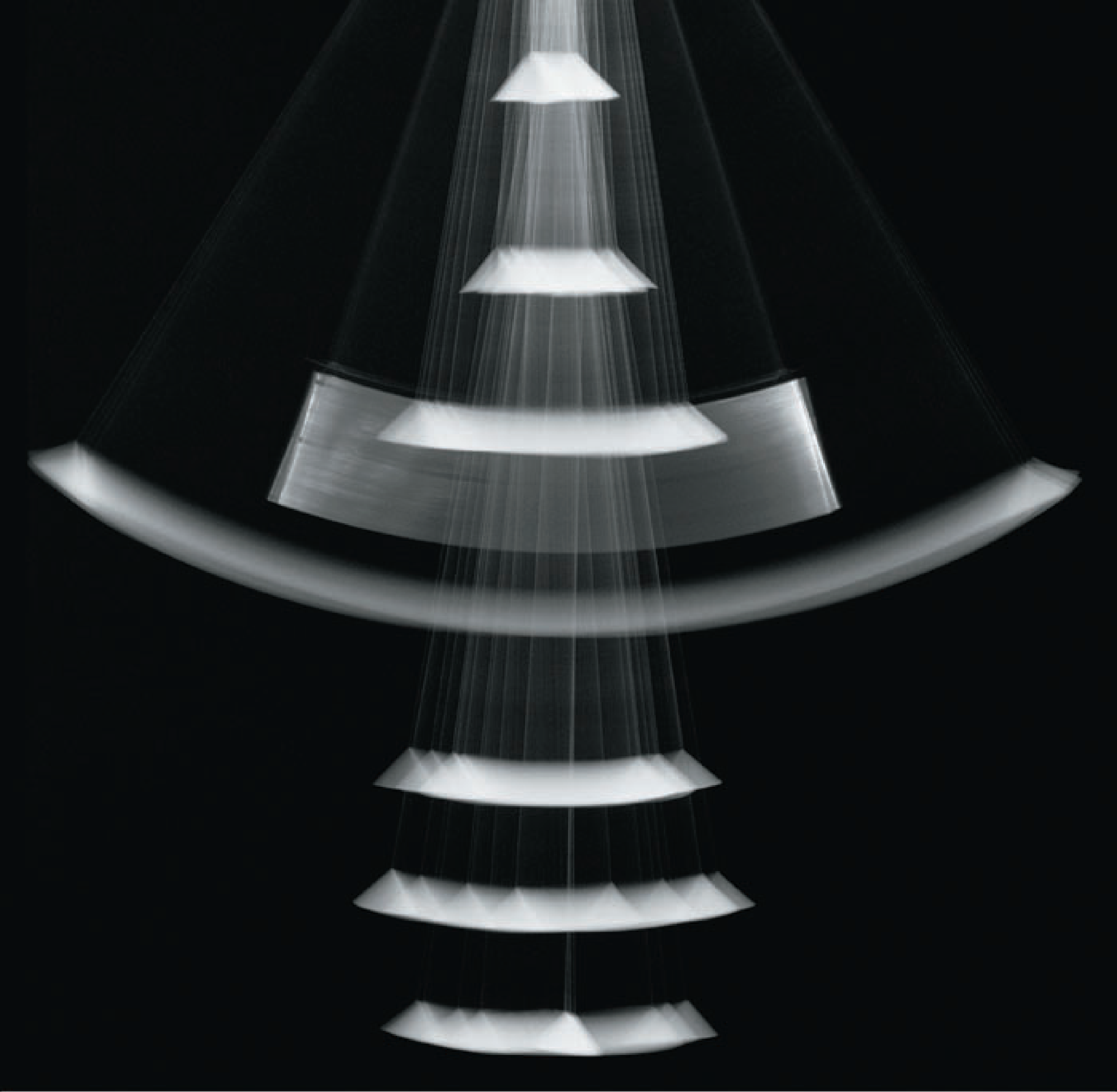


Essential

UNIVERSITY PHYSICS

Richard Wolfson

Second Edition

PHYSICAL CONSTANTS

CONSTANT	SYMBOL	THREE-FIGURE VALUE	BEST KNOWN VALUE*
Speed of light	c	3.00×10^{8} m/s	299 792 458 m/s (exact)
Elementary charge	e	1.60×10^{-19} C	$1.602\ 176\ 487(40)\times10^{-19}$ C
Electron mass	m_e	9.11×10^{-31} kg	$9.109\ 382\ 15(45)\times10^{-31}$ kg
Proton mass	m_p	1.67×10^{-27} kg	$1.672\ 621\ 637(83)\times10^{-27}$ kg
Gravitational constant	G	6.67×10^{-11} N·m^2/kg^2	$6.674\ 28(67)\times10^{-11}$ N·m^2/kg^2
Permeability constant	μ_0	1.26×10^{-6} N/A^2 (H/m)	$4\pi\times10^{-7}$ (exact)
Permittivity constant	ϵ_0	8.85×10^{-12} C^2/N·m^2 (F/m)	$1/\mu_0 c^2$ (exact)
Boltzmann's constant	k	1.38×10^{-23} J/K	$1.380\ 6504(24)\times10^{-23}$ J/K
Universal gas constant	R	8.31 J/K·mol	8.314 472(15) J/K·mol
Stefan–Boltzmann constant	σ	5.67×10^{-8} W/m^2·K^4	$5.670\ 400(40)\times10^{-8}$ W/m^2·K^4
Planck's constant	$h\,(=2\pi\hbar)$	6.63×10^{-34} J·s	$6.626\ 068\ 96(33)\times10^{-34}$ J·s
Avogadro's number	N_{A}	6.02×10^{23} mol^{-1}	$6.022\ 141\ 79(30)\times10^{23}$ mol^{-1}
Bohr radius	a_0	5.29×10^{-11} m	$5.291\ 772\ 0859(36)\times10^{-11}$ m

*Parentheses indicate uncertainties in last decimal places. *Source*: U.S. National Institute of Standards and Technology, 2007 values

SI PREFIXES

POWER	PREFIX	SYMBOL
10^{24}	yotta	Y
10^{21}	zetta	Z
10^{18}	exa	E
10^{15}	peta	P
10^{12}	tera	T
10^{9}	giga	G
10^{6}	mega	M
10^{3}	kilo	k
10^{2}	hecto	h
10^{1}	deca	da
10^{0}	—	—
10^{-1}	deci	d
10^{-2}	centi	c
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f
10^{-18}	atto	a
10^{-21}	zepto	z
10^{-24}	yocto	y

THE GREEK ALPHABET

	UPPERCASE	LOWERCASE
Alpha	A	α
Beta	B	β
Gamma	Γ	γ
Delta	Δ	δ
Epsilon	E	ϵ
Zeta	Z	ζ
Eta	H	η
Theta	Θ	θ
Iota	I	ι
Kappa	K	κ
Lambda	Λ	λ
Mu	M	μ
Nu	N	ν
Xi	Ξ	ξ
Omicron	O	o
Pi	Π	π
Rho	P	ρ
Sigma	Σ	σ
Tau	T	τ
Upsilon	Υ	υ
Phi	Φ	ϕ
Chi	X	χ
Psi	Ψ	ψ
Omega	Ω	ω

Conversion Factors (more conversion factors in Appendix C)

Length

1 in = 2.54 cm
1 mi = 1.609 km
1 ft = 0.3048 m
1 light year = 9.46×10^{15} m

Velocity

1 mi/h = 0.447 m/s
1 m/s = 2.24 mi/h = 3.28 ft/s

Mass, energy, force

1 u = 1.661×10^{-27} kg
1 cal = 4.184 J
1 Btu = 1.054 kJ
1 kWh = 3.6 MJ
1 eV = 1.602×10^{-19} J
1 pound (lb) = 4.448 N
= weight of 0.454 kg

Time

1 day = 86,400 s
1 year = 3.16×10^{7} s

Pressure

1 atm = 101.3 kPa = 760 mm Hg
1 atm = 14.7 lb/in^2

Rotation and angle

1 rad = $180°/\pi$ = 57.3°
1 rev = 360° = 2π rad
1 rev/s = 60 rpm

Magnetic field

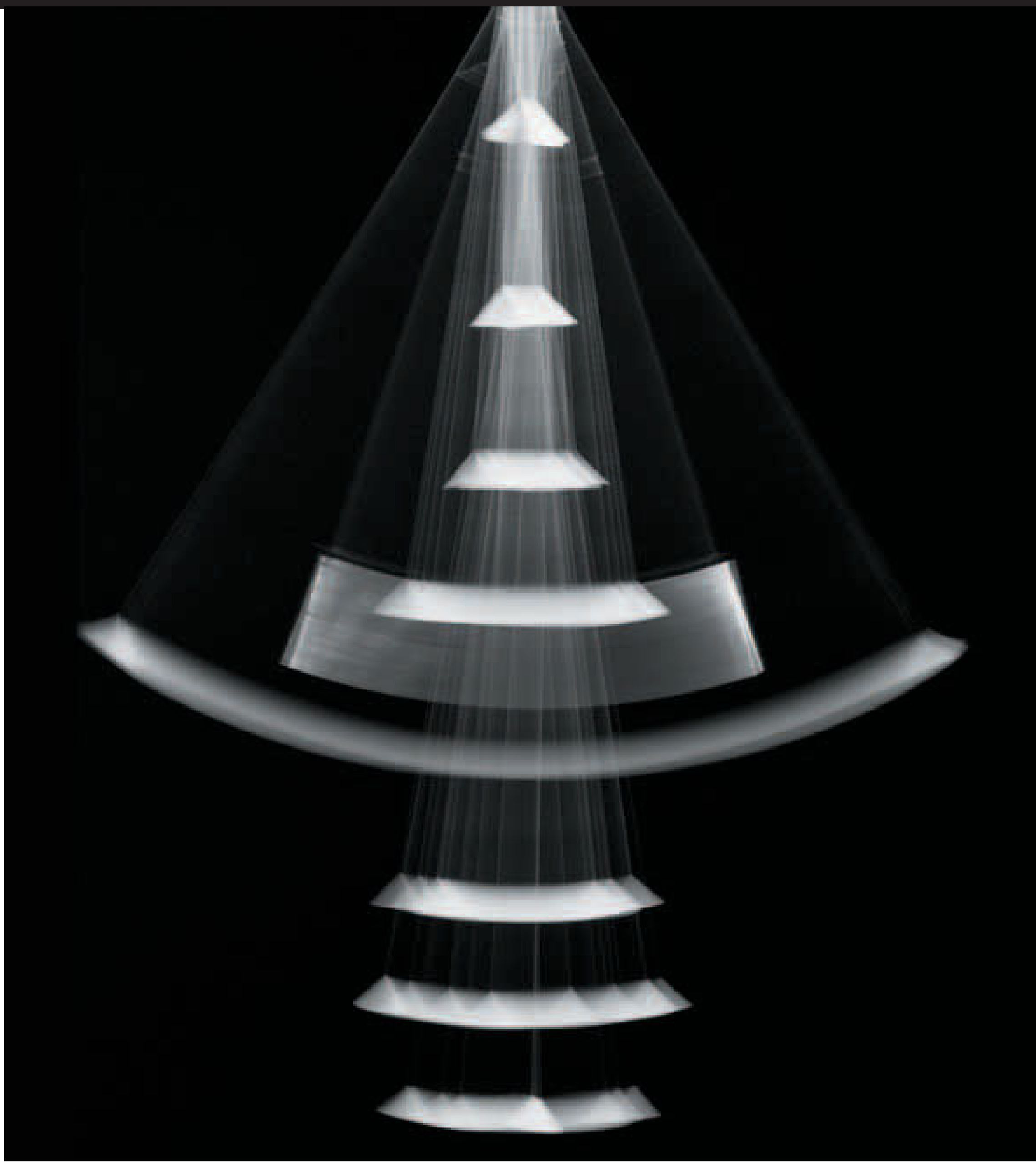
1 gauss = 10^{-4} T

Essential UNIVERSITY PHYSICS

SECOND EDITION

Richard Wolfson

Middlebury College



Addison-Wesley

Boston Columbus Indianapolis New York San Francisco Upper Saddle River
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montréal Toronto
Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Publisher: James Smith
Executive Editor: Nancy Whilton
Project Editor: Martha Steele
Director of Development: Michael Gillespie
Editorial Manager: Laura Kenney
Development Editor: Ashley Eklund
Editorial Assistant: Peter Alston
Managing Editor: Corinne Benson
Production Project Manager: Beth Collins
Production Management and Composition: PreMediaGlobal
Copyeditor: Joanna Dinsmore
Interior and Cover Designer: Derek Bacchus
Illustrators: Rolin Graphics
Photo Researcher: Eric Schrader
Senior Photo Editor: Donna Kalal
Manufacturing Buyer: Jeff Sargent
Senior Marketing Manager: Kerry Chapman
Director of Marketing, Science: Christy Lawrence
Cover Photo Credit: Andrew Lambert Photography / Science Photo Library;
Andrew Lambert Photography / Science Photo Library
Printer: Edwards Brothers

Credits and acknowledgments borrowed from other sources and reproduced, with permission, in this textbook appear on the appropriate page within the text and/or on pages C-1and C-2.

Library of Congress Cataloging-in-Publication Data

Wolfson, Richard.
Essential university physics / Richard Wolfson.—2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-321-70669-0 (v. 1 : alk. paper)
ISBN-10: 0-321-70669-2 (alk. paper)
ISBN-13: 978-0-321-71204-2 (v. 2 : alk. paper)
ISBN-10: 0-321-71204-8 (alk. paper)
1. Physics—Textbooks. I. Title.
QC21. 3.W65 2012
530—dc22

2010039884

1 2 3 4 5 6 7 8 9 10—EB—14 13 12 11 10

ISBN 10: 0-321-70669-2; ISBN 13: 978-0-321-70669-0
(Volume 1 without Mastering Physics)
ISBN 10: 0-321-71204-8; ISBN 13: 978-0-321-71204-2
(Volume 1 with Mastering Physics)
ISBN 10: 0-321-71206-4; ISBN 13: 978-0-321-71206-6
(Volume 1 Exam copy)

ISBN 10: 0-321-70127-5; ISBN 13: 978-0-321-70127-5
(Volume 2 without Mastering Physics)
ISBN 10: 0-321-71174-2; ISBN 13: 978-0-321-71174-8
(Volume 2 with Mastering Physics)
ISBN 10: 0-321-71198-X; ISBN 13: 978-0-321-71198-4
(Volume 2 Exam copy)

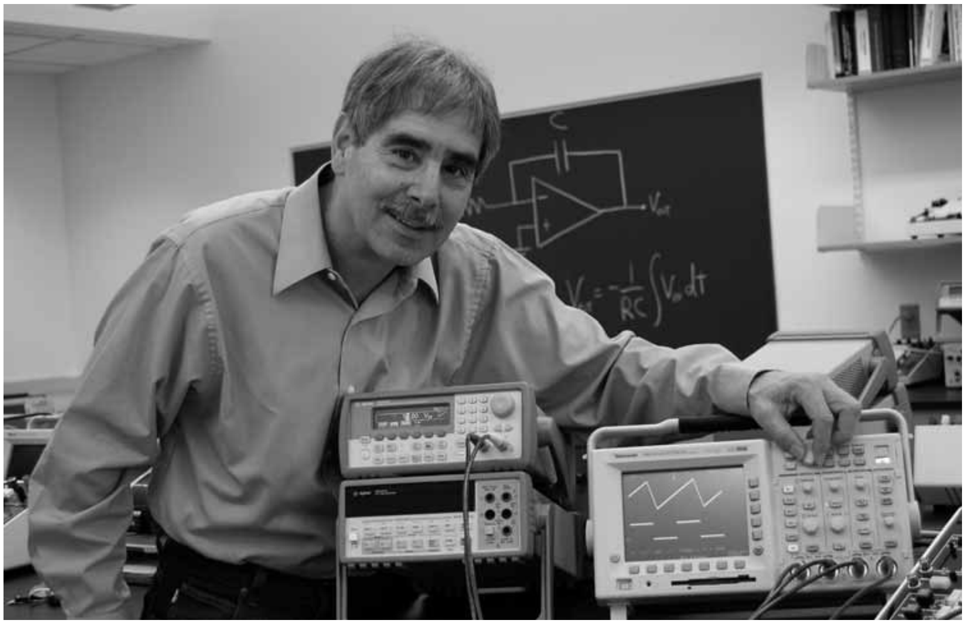
Addison-Wesley
is an imprint of

PEARSON

www.pearsonhighered.com

Brief Contents

About the Author



Richard Wolfson

Richard Wolfson is the Benjamin F. Wissler Professor of Physics at Middlebury College, where he has taught since 1976. He did undergraduate work at MIT and Swarthmore College, and he holds an M.S. degree from the University of Michigan and Ph.D. from Dartmouth. His ongoing research on the Sun's corona and climate change has taken him to sabbaticals at the National Center for Atmospheric Research in Boulder, Colorado; St. Andrews University in Scotland; and Stanford University.

Rich is a committed and passionate teacher. This is reflected in his many publications for students and the general public, including the video series *Einstein's Relativity and the Quantum Revolution: Modern Physics for Nonscientists* (The Teaching Company, 1999), *Physics in Your Life* (The Teaching Company, 2004), and *How the Universe Works: Understanding Physics, from Quark to Galaxy* (The Teaching Company, 2011); books *Nuclear Choices: A Citizen's Guide to Nuclear Technology* (MIT Press, 1993), *Simply Einstein: Relativity Demystified* (W. W. Norton, 2003), and *Energy, Environment, and Climate* (W. W. Norton, 2007); and articles for *Scientific American* and the *World Book Encyclopedia.*

Outside of his research and teaching, Rich enjoys hiking, canoeing, gardening, cooking, and watercolor painting.

Preface to the Instructor

Introductory physics texts have grown ever larger, more massive, more encyclopedic, more colorful, and more expensive. *Essential University Physics* bucks that trend—without compromising coverage, pedagogy, or quality. The text benefits from the author's three decades of teaching introductory physics, seeing firsthand the difficulties and misconceptions that students face as well as the "Got It!" moments when big ideas become clear. It also builds on the author's honing multiple editions of a previous calculus-based textbook and on feedback from hundreds of instructors and students.

Goals of This Book

Physics is the fundamental science, at once fascinating, challenging, and subtle—and yet simple in a way that reflects the few basic principles that govern the physical universe. My goal is to bring this sense of physics alive for students in a range of academic disciplines who need a solid calculus-based physics course—whether they're engineers, physics majors, premeds, biologists, chemists, geologists, mathematicians, computer scientists, or other majors. My own courses are populated by just such a variety of students, and among my greatest joys as a teacher is having students who took a course only because it was required say afterwards that they really enjoyed their exposure to the ideas of physics. More specifically, my goals include:

- Helping students build the analytical and quantitative skills and confidence needed to apply physics in problem solving for science and engineering.
- Adressing key misconceptions and helping students build a stronger conceptual understanding.
- Helping students see the relevance and excitement of the physics they're studying with contemporary applications in science, technology, and everyday life.
- Helping students develop an appreciation of the physical universe at its most fundamental level.
- Engaging students with an informal, conversational writing style that balances precision with approachability.

New to This Edition

We've updated this second edition based on user feedback to expand coverage of a few key topics and abbreviate coverage of some that are less widely taught. We've also added new and revised features to improve conceptual understanding and spotlight relevancy, and have completely overhauled end-of-chapter problem sets. Specific changes include:

- Expanded coverage of impulse (Ch. 9), the addition of the wave equation (Ch.14), and a new statistical treatment of entropy (Ch. 19).
- Reduced coverage of toroids (Ch. 26) and phasors (Ch. 28).
- **New Conceptual Examples** that contain a **Making the Connection** follow-up question applying the concept from the example to a quantitative problem based on a real-world situation.
- **New Applications** to emphasize how physics concepts apply to real-world situations, including biomedical and engineering fields.
- **To Learn/To Know** chapter previews are now called **New Concepts, New Skills/Connecting Your Knowledge**, and have been revised to better emphasize the key concepts in each chapter and to show how they connect with concepts in previous chapters.

- Complete edition Volumes 1–2 (shrinkwrapped) with Mastering-Physics® (ISBN 0-321-714-385): Chapters 1–39
- Volume 1 with Mastering-Physics® (ISBN 0-321-71204-8): Chapters 1–19
- Volume 2 with Mastering-Physics® (ISBN 0-321-71174-2): Chapters 20–39
- Complete edition Volumes 1–2 (shrinkwrapped) (ISBN 0-321-69692-1): Chapters 1–39
- Volume 1 (ISBN 0-321-70669-2): Chapters 1–19
- Volume 2 (ISBN 0-321-70127-5): Chapters 20–39

- End-of-chapter (EOC) problem sets now include **new biomedical-related problems** that investigate physics concepts in biology, medicine, and biomedical technologies, and **new computational problems** that require a computer or graphing calculator.
- Every end-of-chapter problem set now includes **new Passage Problems** similar to the types of questions asked on the MCAT Exam. Each set includes 4 multiple-choice questions based on a passage of text, often with an accompanying figure or graph, and involves biomedical and other real-world scenarios.
- EOC problem sets have been thoroughly revised to include new calculus-based problems, new context-rich problems, and more medium-difficulty problems.
- All quantitative results—both in-text examples and EOC problems—have been checked independently for accuracy by two physicists.
- Finally, we've incorporated new research results and new applications of physics principles wherever they're relevant.

Pedagogical Innovations

This book is *concise*, but it's also *progressive* in its embrace of proven techniques from physics education research and *strategic* in its approach to learning physics. Chapter 1 introduces the IDEA framework for problem solving, and every one of the book's subsequent **worked examples** employs this framework. IDEA—an acronym for Identify, Develop, Evaluate, Assess—is not a "cookbook" method for students to apply mindlessly, but rather a tool for organizing students' thinking and discouraging equation hunting. It begins with an interpretation of the problem and an identification of the key physics concepts involved; develops a plan for reaching the solution; carries out the mathematical evaluation; and assesses the solution to see that it makes sense, to compare the example with others, and to mine additional insights into physics. In nearly all of the text's worked examples, the Develop phase includes making a drawing, and most of these use a hand-drawn style to encourage students to make their own drawings—a step that research suggests they often skip. IDEA provides a common approach to all physics problem solving, an approach that emphasizes the conceptual unity of physics and helps break the typical student view of physics as a hodgepodge of equations and unrelated ideas. In addition to IDEA-based worked examples, other pedagogical features include:

- **Problem-Solving Strategy boxes** that follow the IDEA framework to provide detailed guidance for specific classes of physics problems, such as Newton's second law, conservation of energy, thermal-energy balance, Gauss's law, or multiloop circuits.
- **Tactics boxes** that reinforce specific essential skills such as differentiation, setting up integrals, vector products, drawing free-body diagrams, simplifying series and parallel circuits, or ray tracing.
- **Got It? boxes** that provide quick checks for students to test their conceptual understanding. Many of these use a multiple-choice or quantitative ranking format to probe student misconceptions and facilitate their use with classroom-response systems.
- **Tips** that provide helpful problem-solving hints or warn against common pitfalls and misconceptions.
- **Chapter openers** that include a forward-looking **New Concepts, New Skills** list for the chapter ahead, and a backward-looking **Connecting Your Knowledge** list of important ideas on which the chapter builds. Both lists reference specific chapter sections by number.
- **Applications**, self-contained presentations typically shorter than half a page, provide interesting and contemporary instances of physics in the real world, such as bicycle stability; flywheel energy storage; laser vision correction; ultracapacitors; wind energy; magnetic resonance imaging; global climate change; combined-cycle power generation; circuit models of the cell membrane; CD, DVD, and Blu-ray technologies; and radiocarbon dating.

- **For Thought and Discussion** questions at the end of each chapter designed for peer learning or for self-study to enhance students' conceptual understanding of physics.
- **Annotated figures** that adopt the research-based approach of including simple "instructor's voice" commentary to help students read and interpret pictorial and graphical information.
- **End-of-chapter** problems that begin with simpler exercises keyed to individual chapter sections and ramp up to more challenging and often multistep problems that synthesize chapter material. Context-rich problems focusing on real-world situations are interspersed throughout each problem set.
- **Chapter summaries** that combine text, art, and equations to provide a synthesized overview of each chapter. Each summary is hierarchical, beginning with the chapter's "big picture" ideas, then focusing on key concepts and equations, and ending with a list of "applications"—specific instances or applications of the physics presented in the chapter.

Organization

This contemporary book is *concise*, *strategic*, and *progressive*, but it's *traditional* in its organization. Following the introductory Chapter 1, the book is divided into six parts. Part One (Chapters 2–12) develops the basic concepts of mechanics, including Newton's laws and conservation principles as applied to single particles and multiparticle systems. Part Two (Chapters 13–15) extends mechanics to oscillations, waves, and fluids. Part Three (Chapters 16–19) covers thermodynamics. Part Four (Chapters 20–29) deals with electricity and magnetism. Part Five (Chapters 30–32) treats optics, first in the geometrical optics approximation and then including wave phenomena. Part Six (Chapters 33–39) introduces relativity and quantum physics. Each part begins with a brief description of its coverage, and ends with a conceptual summary and a challenge problem that synthesizes ideas from several chapters.

Essential University Physics is available in two paperback volumes, so students can purchase only what they need—making the low-cost aspect of this text even more attractive. Volume 1 includes Parts One, Two, and Three, mechanics through thermodynamics. Volume 2 contains Parts Four, and Five, and Six, electricity and magnetism along with optics and modern physics.

Instructor Supplements

NOTE: For convenience, all of the following instructor supplements (except the Instructor Resource DVD) can be downloaded from the "Instructor Area," accessed via the left-hand navigation bar of MasteringPhysics® (www.masteringphysics.com).

- The **Instructor Solutions Manual** (ISBN 0-321-69723-5) contains solutions to all end-of-chapter exercises and problems, written in the Interpret/Develop/Evaluate/ Assess (IDEA) problem-solving framework. The solutions are provided in PDF and editable Microsoft® Word formats for Mac and PC, with equations in MathType, and can also be downloaded from the Instructor Resource Center (www.pearsonhighered.com/irc).
- The **Instructor Resource DVD** (ISBN 0-321-71171-8) provides all the figures, photos, and tables from the text in JPEG format. All the problem-solving strategies, Tactics Boxes, key equations, and chapter summaries are provided in PDF and editable Microsoft® Word formats with equations in MathType. Each chapter also has a set of PowerPoint® lecture outlines and "clicker" questions. A comprehensive library of more than 220 applets from **ActivPhysics OnLine™** and a suite of over 70 PhET simulations are also included.
- Mastering PHYSICS™ **MasteringPhysics®** (www.masteringphysics.com) is the most advanced physics homework and tutorial system available. This online homework and tutoring system guides students through the toughest topics in physics with self-paced tutorials that provide individualized coaching. These assignable, in-depth tutorials are designed to coach students with hints and feedback specific to their individual errors. Instructors can also assign end-of-chapter problems from every chapter, including multiple-choice questions, section-specific exercises, and general problems. Quantitative problems can be assigned with numerical answers and randomized values (with sig fig feedback) or solutions.
- **Pearson eText** is available either automatically when MasteringPhysics® is packaged with new books or as a

purchased upgrade online. Users can search for words or phrases, create notes, highlight text, bookmark sections, click on definitions to key terms, and launch ActivPhysics applets and PhET simulations as they read. Professors also have the ability to annotate the text for their course and hide chapters not covered in their syllabi.

- The **Test Bank** (ISBN 0-321-71172-6) contains more than 2000 multiple-choice, true-false, and conceptual questions in TestGen® and Microsoft Word® formats for Mac and PC users. More than half of the questions can be assigned with randomized numerical values. The Test Bank can also be downloaded from www. pearsonhighered.com/irc.

Student Supplements

- The **Student Solutions Manuals, Volume 1 (Chapters 1–19)** (ISBN 0-321-71203-X) and **Volume 2 (Chapters 20–39)** (ISBN 0-321-71205-6) contain detailed solutions to all of the odd-numbered end-of-chapter problems from the textbook. All solutions are written in the Interpret/Develop/Evaluate/Assess (IDEA) problem-solving framework.
- **MasteringPhysics**® (www.masteringphysics.com) is the most advanced physics homework and tutorial system available. This online homework and tutoring system guides students through the most important topics in physics with self-paced tutorials that provide individualized coaching. These assignable, in-depth tutorials are designed to coach students with hints and feedback specific to their individual errors. Instructors can also assign end-of-chapter problems from every chapter including multiple-choice questions, section-specific exercises, and general problems. Quantitative problems can be assigned with numerical answers and randomized values (with sig fig feedback) or solutions.
- **Pearson eText** is available through MasteringPhysics®, either automatically when MasteringPhysics® is packaged with new books or as a purchased upgrade online. Allowing students access to the text wherever they have access to the Internet, Pearson eText comprises the full text with additional interactive features. Users can search for words or phrases, create notes, highlight text, bookmark sections, click on definitions to key terms, and launch ActivPhysics applets and PhET simulations as they read.

Acknowledgments

A project of this magnitude isn't the work of its author alone. First and foremost among those I thank for their contributions are the now several thousand students I've taught in calculus-based introductory physics courses at Middlebury College. Over the years your questions have taught me how to convey physics ideas in many different ways appropriate to your diverse learning styles. You've helped identify the "sticking points" that challenge introductory physics students, and you've showed me ways to help you avoid and "unlearn" the misconceptions that many students bring to introductory physics.

Thanks also go to my Middlebury faculty colleagues and to numerous instructors and students from around the world who have contributed valuable suggestions that were incorporated in the revisions of my earlier introductory physics text, *Physics for Scientists and Engineers* (Wolfson and Pasachoff, third edition: Addison-Wesley, 1999). I've heard you, and you'll find still more of your suggestions implemented in *Essential University Physics*.

Experienced physics instructors thoroughly reviewed every chapter of this book, and reviewers' comments resulted in substantive changes—and sometimes in major rewrites—to the first drafts of the manuscript. We list all these reviewers below. But first, special thanks are due to six individuals who made exceptional contributions to the quality and in some cases the very existence of this book. First is Professor Jay Pasachoff of Williams College, whose willingness more than two decades ago to take a chance on an inexperienced coauthor has made writing introductory physics a large part of my professional career. Dr. Adam Black, physics editor and Ph.D. physicist at Addison-Wesley, had the vision to see promise in a new introductory text that would respond to the rising chorus of complaints about massive, encyclopedic, and expensive physics texts. Brad Patterson, developmental editor for the first edition, brought his graduate-level knowledge of physics to a role that made him a real collaborator and the closest this book has to a coauthor. Brad is responsible for many of the book's innovative features, and it was a pleasure to work with him. We've gone to great lengths to make this book as error-free as possible, and much of the credit for that happy situation goes to Charles Hibbard and Peter W. Murphy. Not only did they check the numbers for every worked example, but they also read the entire book in page proof with a professionally critical eye, and they're responsible for many improvements to both text and art made even at that late stage.

I also wish to thank Martha Steele, Ashley Eklund, Nancy Whilton, and Beth Collins at Addison-Wesley, and Jared Sterzer at PreMediaGlobal, for their highly professional efforts in shepherding this book through its vigorous production schedule. Martha, especially, has been with this project since its first edition, and her cheerful and meticulous attention to detail has made production of this book a smooth and pleasant process. Finally, as always, I thank my family, my colleagues, and my students for the patience they showed during the intensive process of writing and revising this book.

Reviewers

John R. Albright, *Purdue University–Calumet*
Rama Bansil, *Boston University*
Richard Barber, *Santa Clara University*
Linda S. Barton, *Rochester Institute of Technology*
Rasheed Bashirov, *Albertson College of Idaho*
Chris Berven, *University of Idaho*
David Bixler, *Angelo State University*
Ben Bromley, *University of Utah*
Charles Burkhardt, *St. Louis Community College*
Susan Cable, *Central Florida Community College*
George T. Carlson, Jr., *West Virginia Institute of Technology–West Virginia University*
Catherine Check, *Rock Valley College*
Norbert Chencinski, *College of Staten Island*
Carl Covatto, *Arizona State University*
David Donnelly, *Texas State University–San Marcos*
David G. Ellis, *University of Toledo*
Tim Farris, *Volunteer State Community College*
Paula Fekete, *Hunter College of The City University of New York*
Idan Ginsburg, *Harvard University*
James Goff, *Pima Community College*
Austin Hedeman, *University of California–Berkeley*
Mark Hollabaugh, *Normandale Community College*
Rex W. Joyner, *Indiana Institute of Technology*
Nikos Kalogeropoulos, *Borough of Manhattan Community College–The City University of New York*
Kevin T. Kilty, *Laramie County Community College*
Duane Larson, *Bevill State Community College*
Kenneth W. McLaughlin, *Loras College*
Tom Marvin, *Southern Oregon University*
Perry S. Mason, *Lubbock Christian University*
Mark Masters, *Indiana University–Purdue University Fort Wayne*
Jonathan Mitschele, *Saint Joseph's College*
Gregor Novak, *United States Air Force Academy*
Richard Olenick, *University of Dallas*
Robert Philbin, *Trinidad State Junior College*
Russell Poch, *Howard Community College*
Steven Pollock, *Colorado University–Boulder*
James Rabchuk, *Western Illinois University*
George Schmiedeshoff, *Occidental College*
Natalia Semushkina, *Shippensburg University of Pennsylvania*
Anwar Shiekh, *Dine College*
David Slimmer, *Lander University*
Chris Sorensen, *Kansas State University*
Ronald G. Tabak, *Youngstown State University*
Gajendra Tulsian, *Daytona Beach Community College*
Henry Weigel, *Arapahoe Community College*
Arthur W. Wiggins, *Oakland Community College*
Fredy Zypman, *Yeshiva University*

Preface to the Student

Welcome to physics! Maybe you're taking introductory physics because you're majoring in a field of science or engineering that requires a semester or two of physics. Maybe you're premed, and you know that medical schools are increasingly interested in seeing calculus-based physics on your transcript. Perhaps you're really gung-ho and plan to major in physics. Or maybe you want to study physics further as a minor associated with related fields like math or chemistry or to complement a discipline like economics, environmental studies, or even music. Perhaps you had a great high-school physics course, and you're eager to continue. Maybe high-school physics was an academic disaster for you, and you're approaching this course with trepidation. Or perhaps this is your first experience with physics. Whatever your reason for taking introductory physics, welcome!

And whatever your reason, my goals for you are similar: I'd like to help you develop an understanding and appreciation of the physical universe at a deep and fundamental level; I'd like you to become aware of the broad range of natural and technological phenomena that physics can explain; and I'd like to help you strengthen your analytic and quantitative problem-solving skills. Even if you're studying physics only because it's a requirement, I want to help you engage the subject and come away with an appreciation for this fundamental science and its wide applicability. One of my greatest joys as a physics teacher is having students tell me after the course that they had taken it only because it was required, but found they really enjoyed their exposure to the ideas of physics.

Physics is fundamental. To understand physics is to understand how the world works, both in everyday life and on scales of time and space so small and so large as to defy intuition. For that reason I hope you'll find physics fascinating. But you'll also find it challenging. Learning physics will challenge you with the need for precise thinking and language; with subtle interpretations of even commonplace phenomena; and with the need for skillful application of mathematics. But there's also a simplicity to physics, a simplicity that results because there are in physics only a very few really basic principles to learn. Those succinct principles encompass a universe of natural phenomena and technological applications.

I've been teaching introductory physics for decades, and this book distills everything my students have taught me about the many different ways to approach physics; about the subtle misconceptions students often bring to physics; about the ideas and types of problems that present the greatest challenges; and about ways to make physics engaging, exciting, and relevant to your life and interests.

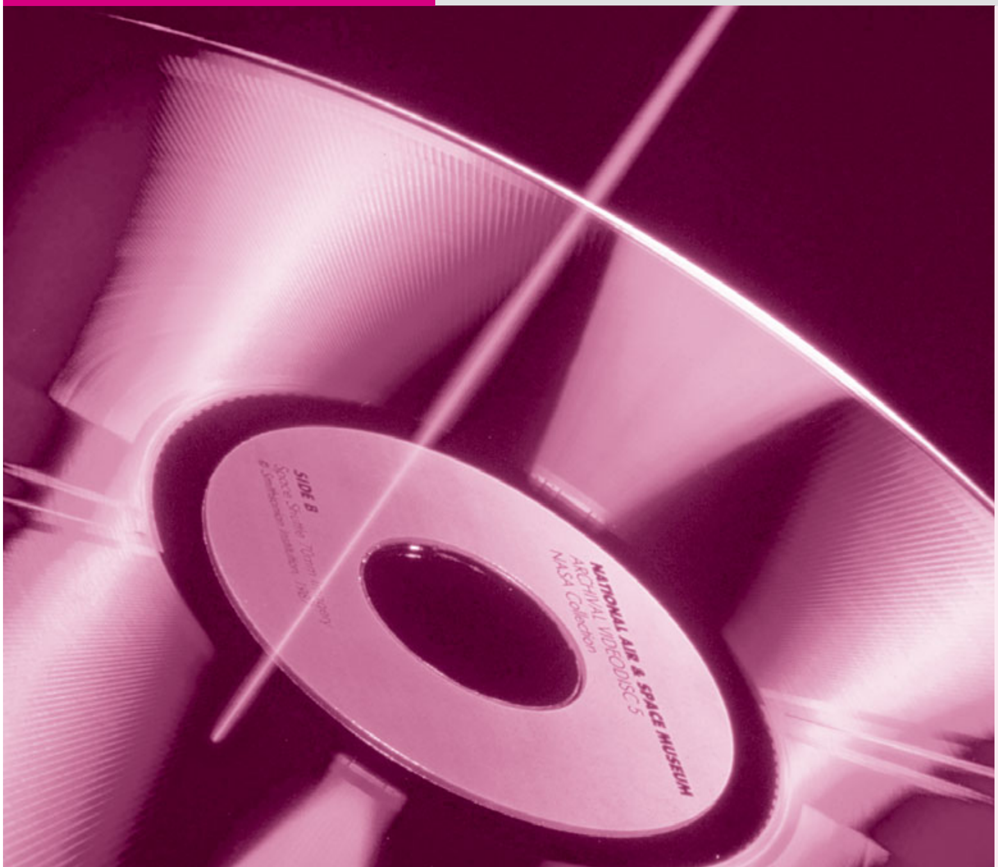
I have some specific advice for you that grows out of my long experience teaching introductory physics. Keeping this advice in mind will make physics easier (but not necessarily easy!), more interesting, and, I hope, more fun:

- *Read* each chapter thoroughly and carefully before you attempt to work any problem assignments. I've written this text with an informal, conversational style to make it engaging. It's not a reference work to be left alone until you need some specific piece of information; rather, it's an unfolding "story" of physics—its big ideas and their applications in quantitative problem solving. You may think physics is hard because it's mathematical, but in my long experience I've found that failure to *read* thoroughly is the biggest single reason for difficulties in introductory physics.
- *Look for the big ideas*. Physics isn't a hodgepodge of different phenomena, laws, and equations to memorize. Rather, it's a few big ideas from which flow myriad applications, examples, and special cases. In particular, don't think of physics as a jumble of equations that you choose among when solving a problem. Rather, identify those few big ideas and the equations that represent them, and try to see how seemingly distinct examples and special cases relate to the big ideas.
- *When working problems*, *re-read* the appropriate sections of the text, paying particular attention to the worked examples. Follow the IDEA strategy described in Chapter 1 and used in every subsequent worked example. Don't skimp on the final Assess step. Always ask: Does this answer make sense? How can I understand my answer in relation to the big principles of physics? How was this problem like others I've worked, or like examples in the text?
- *Don't confuse physics with math*. Mathematics is a tool, not an end in itself. Equations in physics aren't abstract math, but statements about the physical world. Be sure you understand each equation for what it says about physics, not just as an equality between mathematical terms.
- *Work with others*. Getting together informally in a room with a blackboard is a great way to explore physics, to clarify your ideas and help others clarify theirs, and to learn from your peers. I urge you to discuss physics problems together with your classmates, to contemplate together the "For Thought and Discussion" questions at the end of each chapter, and to engage one another in lively dialog as you grow your understanding of physics, the fundamental science.

Detailed Contents

Volume 1 contains Chapters 1–19
Volume 2 contains Chapters 20–39

1 Doing Physics



Which realms of physics are involved in the workings of your DVD player?

New Concepts, New Skills

By the end of this chapter you should be able to

- Describe the different realms of physics and their applications in both natural and technological systems (1.1).
- Explain the SI unit system and convert units (1.2).
- Express and manipulate numbers using scientific notation (1.3).
- Explain the importance of significant figures and handle them in calculations (1.3).
- Make quick order-of-magnitude estimates (1.3).
- Describe the strategic steps used in solving physics problems (1.4).

Connecting Your Knowledge

- Physics is a quantitative science, and we'll begin right off using algebra. Soon we'll add trigonometry and later calculus (which you might be studying now, concurrently with physics). However, you don't need to have taken physics previously to get a full understanding from this book.

You slip a DVD into your player and settle in to watch a movie. The DVD spins, and a precisely focused laser beam "reads" its content. Electronic circuitry processes the information, sending it to your video display and to loudspeakers that turn electrical signals into sound waves. Every step of the way, principles of physics govern the delivery of the movie from DVD to you.

1.1 Realms of Physics

That DVD player is a metaphor for all of **physics**—the science that describes the fundamental workings of physical reality. Physics explains natural phenomena ranging from the behavior of atoms and molecules to thunderstorms and rainbows and on to the evolution of stars, galaxies, and the universe itself. Technological applications of physics are the basis for everything from microelectronics to medical imaging to cars, airplanes, and space flight.

At its most fundamental, physics provides a nearly unified description of all physical phenomena. However, it's convenient to divide physics into distinct realms

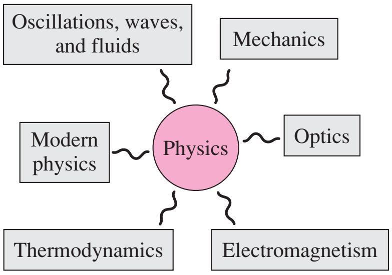


FIGURE 1.1 Realms of physics.

(Fig. 1.1). Your DVD player encompasses essentially all those realms. **Mechanics**, the branch of physics that deals with motion, describes the spinning disc. Mechanics also explains the motion of a car, the orbits of the planets, and the stability of a skyscraper. Part 1 of this book deals with the basic ideas of mechanics.

Those sound waves coming from your loudspeakers represent **wave motion**. Other examples include the ocean waves that pound Earth's coastlines, the wave of standing spectators that sweeps through a football stadium, and the undulations of Earth's crust that spread the energy of an earthquake. Part 2 of this book covers wave motion and other phenomena involving the motion of fluids like air and water.

When you burn your own DVD, the high temperature produced by an intensely focused laser beam alters the material properties of a writable DVD, thus storing video or computer information. That's an example of **thermodynamics**—the study of heat and its effects on matter. Thermodynamics also describes the delicate balance of energy-transfer processes that keeps our planet at a habitable temperature and puts serious constraints on our ability to meet the burgeoning energy demands of modern society. Part 3 comprises four chapters on thermodynamics.

An electric motor spins your DVD, converting electrical energy to the energy of motion. Electric motors are ubiquitous in modern society, running everything from subway trains to washing machines to your computer's hard drive. Conversely, electric generators convert the energy of motion to electricity, providing virtually all of our electrical energy. Motors and generators are two applications of **electromagnetism** in modern technology. Others include computers, audiovisual electronics, microwave ovens, digital watches, and even the humble lightbulb; without these electromagnetic technologies our lives would be very different. Equally electromagnetic are all the wireless technologies that enable modern communications, from satellite TV to cell phones to wireless computer networks, mice, and keyboards. And even light itself is an electromagnetic phenomenon. Part 4 presents the principles of electromagnetism and their many applications.

The precise focusing of laser light in your DVD player allows hours of video to fit on a small plastic disc. The details and limitations of that focusing are governed by the principles of **optics**, the study of light and its behavior. Applications of optics range from simple magnifiers to contact lenses to sophisticated instruments such as microscopes, telescopes, and spectrometers. Optical fibers carry your e-mail, web pages, and music downloads over the global Internet. Natural optical systems include your eye and the raindrops that deflect sunlight to form rainbows. Part 5 of the book explores optical principles and their applications.

That laser light in your DVD player is an example of an electromagnetic wave, but an atomic-level look at the light's interaction with matter reveals particle-like "bundles" of electromagnetic energy. This is the realm of **quantum physics**, which deals with the often counterintuitive behavior of matter and energy at the atomic level. Quantum phenomena also explain how that DVD laser works and, more profoundly, the structure of atoms and the periodic arrangement of the elements that is the basis of all chemistry. Quantum physics is one of the two great developments of **modern physics**. The other is Einstein's **theory of relativity**. Relativity and quantum physics arose during the 20th century, and together they've radically altered our commonsense notions of time, space, and causality. Part 6 of the book surveys the ideas of modern physics.

CONCEPTUAL EXAMPLE 1.1 Car Physics

Name systems in your car that exemplify the different realms of physics.

EVALUATE *Mechanics* is easy; the car is fundamentally a mechanical system whose purpose is motion. Details include starting, stopping, cornering, as well as a host of other motions within mechanical subsystems. Your car's springs and shock absorbers constitute an *oscillatory* system engineered to give a comfortable ride. The car's engine is a prime example of a *thermodynamic* system, converting the energy of burning gasoline into the car's motion. *Electromagnetic* systems range from the starter motor and spark plugs to sophisticated electronic devices that monitor and optimize engine performance. *Optical* principles govern rear- and side-view mirrors and headlights. Increasingly, optical fibers transmit information to critical safety systems. *Modern physics* is less obvious in your car, but ultimately, everything from the chemical reactions of burning gasoline to the atomic-scale operation of automotive electronics is governed by its principles.

1.2 Measurements and Units

"A long way" means different things to a sedentary person, a marathon runner, a pilot, and an astronaut. We need to quantify our measurements. Science uses the **metric system**, with fundamental quantities length, mass, and time measured in meters, kilograms, and seconds, respectively. The modern version of the metric system is **SI**, for Système International d'Unités (International System of Units), which incorporates scientifically precise definitions of the fundamental quantities.

Length

The **meter** was first defined as one ten-millionth of the distance from the equator to the north pole. In 1889 a standard meter was fabricated to replace the Earth-based unit, and in 1960 that gave way to a standard based on the wavelength of light. Such an **operational definition**, a measurement standard based on a laboratory procedure, has the advantage that scientists anywhere can reproduce the standard meter. By the 1970s, the speed of light had become one of the most precisely determined quantities. As a result, in 1983 the meter was given a new operational definition:

> The meter is the length of the path traveled by light in vacuum during a time interval of 1/299,792,458 of a second.

This definition of the meter also means that the speed of light is now a defined quantity; its value is exactly 299,792,458 m/s.

Time

The **second** used to be defined by Earth's rotation, but that's not constant, so it was redefined as a specific fraction of the year 1900. In 1967 the second was given an operational definition involving atomic vibrations:

> The second is the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom.

The device that implements this definition—which will seem a lot less obscure once you've studied some atomic physics—is called an atomic clock.

Mass

Today's mass standard is the least satisfactory. Unlike the operational definitions of length and time, based on procedures that can be repeated anywhere, the unit of mass is defined by a particular object—the international prototype **kilogram** kept at the International Bureau of Weights and Measures at Sèvres, France.

The prototype kilogram is made of a special platinum-iridium alloy that is very hard and not subject to corrosion. Nevertheless, it could change, and in any event comparison with such a standard is less convenient than an operational definition that can be checked in a laboratory. So scientists are developing techniques based on counting the atoms in a given volume, to scale up from the mass of a single atom to a new definition of the kilogram.

Other SI Units

SI includes seven independent base units: In addition to the three we've just defined, there are the ampere (A) for electric current, the kelvin (K) for temperature, the mole (mol) for the amount of a substance, and the candela (cd) for luminosity. Two supplementary units are used to measure angle: the radian (rad) for ordinary angles (Fig. 1.2) and the steradian (sr) for solid angles. Units for all other quantities are derived from these base units.

APPLICATION **Units Matter: A Bad Day on Mars**

In September 1999 the Mars Climate Orbiter was destroyed when the spacecraft passed through Mars's atmosphere and experienced stresses and heating it was not designed to tolerate. Why did this $125-million craft enter the Martian atmosphere when it was supposed to remain in the vacuum of space? NASA identified the "root cause" as a failure to convert the English units one team used to specify rocket thrust to the SI units another team expected. Units matter!

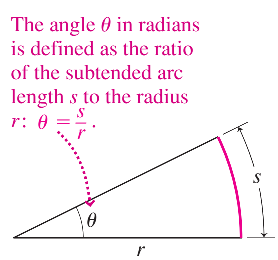


FIGURE 1.2 The radian is the SI unit of angle.

Table 1.1 SI Prefixes

Prefix	Symbol	Power
yotta	Y	10^{24}
zetta	Z	10^{21}
exa	E	10^{18}
peta	P	10^{15}
tera	T	10^{12}
giga	G	10^{9}
mega	M	10^{6}
kilo	k	10^{3}
hecto	h	10^{2}
deca	da	10^{1}
—	—	10^{0}
deci	d	10^{-1}
centi	c	10^{-2}
milli	m	10^{-3}
micro	μ	10^{-6}
nano	n	10^{-9}
pico	p	10^{-12}
femto	f	10^{-15}
atto	a	10^{-18}
zepto	z	10^{-21}
yocto	y	10^{-24}

You could specify the length of a bacterium (e.g., 0.00001 m) or the distance to the next city (e.g., 58,000 m) in meters, but the results are unwieldy—too small in the first case and too large in the latter. So we use prefixes to indicate multiples of the SI base units. For example, the prefix k (for "kilo") means 1000; 1 km is 1000 m, and the distance to the next city is 58 km. Similarly, the prefix μ (the lowercase Greek "mu") means "micro," or 10^{-6}. So our bacterium is 10 μm long. The SI prefixes are listed in Table 1.1, which is repeated inside the front cover. We'll use the prefixes routinely in examples and problems.

When two units are used together, a hyphen appears between them—for example, newton-meter. Each unit has a symbol, such as m for meter or N for newton (the SI unit of force). Symbols are ordinarily lowercase, but those named after people are uppercase. Thus "newton" is written with a small "n" but its symbol is a capital N. The exception is the unit of volume, the liter; since the lowercase "l" is easily confused with the number one, the symbol for liter is a capital L. When two units are multiplied, their symbols are separated by a centered dot: N·m for newton-meter. Division of units is expressed by using the slash (/) or writing with the denominator unit raised to the -1 power. Thus the SI unit of speed is the meter per second, written m/s or $\mathrm{m{\cdot}s^{-1}}$.

Other Unit Systems

The inches, feet, yards, miles, and pounds of the so-called English system still dominate measurement in the United States. Other non-SI units such as the hour are often mixed with English or SI units, as with speed limits in miles per hour or kilometers per hour. In some areas of physics there are good reasons for using non-SI units. We'll discuss these as the need arises and will occasionally use non-SI units in examples and problems. We'll also often find it convenient to use degrees rather than radians for angles. The vast majority of examples and problems in this book, however, use strictly SI units.

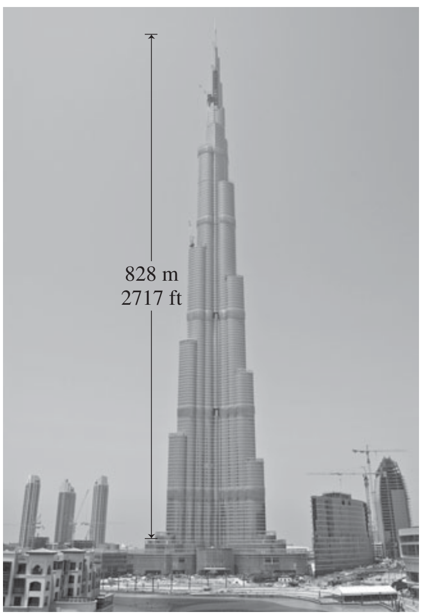


FIGURE 1.3 Dubai's Burj Khalifa is the world's tallest structure.

Changing Units

Sometimes we need to change from one unit system to another—for example, from English to SI. Appendix C contains tables for converting among unit systems; you should familiarize yourself with this and the other appendices and refer to them often.

For example, Appendix C shows that 1 ft = 0.3048 m. Since 1 ft and 0.3048 m represent the same physical distance, multiplying any distance by their ratio will change the units but not the actual physical distance. Thus the height of Dubai's Burj Khalifa (Fig. 1.3)—the world's tallest structure—is 2717 ft or

$$(2717\ \mathrm{ft})\left(\frac{0.3048\ \mathrm{m}}{1\ \mathrm{ft}}\right) = 828.1\ \mathrm{m}$$

Often you'll need to change several units in the same expression. Keeping track of the units through a chain of multiplications helps prevent you from carelessly inverting any of the conversion factors. A numerical answer cannot be correct unless it has the right units!

EXAMPLE 1.1 Changing Units: Speed Limits

Express a 65 mi/h speed limit in meters per second.

EVALUATE In Appendix C, we find that 1 mi = 1609 m, so we can multiply miles by the ratio 1609 m/mi to get meters. Similarly, we use the conversion factor 3600 s/h to convert hours to seconds. So we have

$$65\ \mathrm{mi/h} = \left(\frac{65\ \cancel{\mathrm{mi}}}{\cancel{\mathrm{h}}}\right)\left(\frac{1609\ \mathrm{m}}{\cancel{\mathrm{mi}}}\right)\left(\frac{1\ \cancel{\mathrm{h}}}{3600\ \mathrm{s}}\right) = 29\ \mathrm{m/s}$$

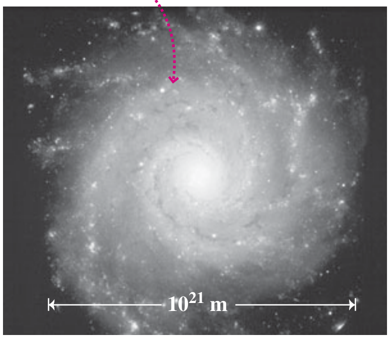
■

1.3 Working with Numbers

Scientific Notation

The range of measured quantities in the universe is enormous; lengths alone go from about 1/1,000,000,000,000,000 m for the radius of a proton to 1,000,000,000,000,000,000,000 m for the size of a galaxy; our telescopes see 100,000 times farther still. Therefore, we frequently express numbers in **scientific notation**, where a reasonable-size number is multiplied by a power of 10. For example, 4185 is 4.185×10^3 and 0.00012 is 1.2×10^{-4}. Table 1.2 suggests the vast range of measurements for the fundamental quantities of length, time, and mass. Take a minute (about 10^2 heartbeats, or 3×10^{-8} of a typical human lifespan) to peruse this table along with Fig. 1.4.

This galaxy is 10^{21} m across and has a mass of ~ 10^{42} kg.



Your movie is stored on a DVD in "pits" only 4×10^{-7} m in size.

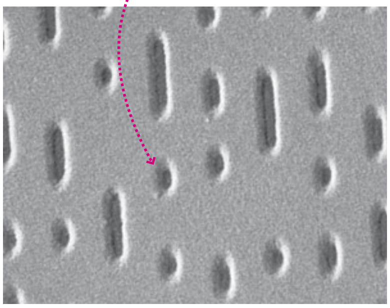


FIGURE 1.4 Large and small.

Table 1.2 Distances, Times, and Masses (rounded to one significant figure)

Radius of observable universe	1×10^{26} m
Earth's radius	6×10^{6} m
Tallest mountain	9×10^{3} m
Height of person	2 m
Diameter of red blood cell	1×10^{-5} m
Size of proton	1×10^{-15} m
Age of universe	4×10^{17} s
Earth's orbital period (1 year)	3×10^{7} s
Human heartbeat	1 s
Wave period, microwave oven	5×10^{-10} s
Time for light to cross a proton	3×10^{-24} s
Mass of Milky Way galaxy	1×10^{42} kg
Mass of mountain	1×10^{18} kg
Mass of human	70 kg
Mass of red blood cell	1×10^{-13} kg
Mass of uranium atom	4×10^{-25} kg
Mass of electron	1×10^{-30} kg

Scientific calculators handle numbers in scientific notation. But straightforward rules allow you to manipulate scientific notation if you don't have such a calculator handy.

TACTICS 1.1 Using Scientific Notation

Addition/Subtraction

To add (or subtract) numbers in scientific notation, first give them the same exponent and then add (or subtract):

$$3.75\times10^6 + 5.2\times10^5 = 3.75\times10^6 + 0.52\times10^6 = 4.27\times10^6$$

Multiplication/Division

To multiply (or divide) numbers in scientific notation, multiply (or divide) the digits and add (or subtract) the exponents:

$$(3.0\times10^8\ \text{m/s})(2.1\times10^{-10}\ \text{s}) = (3.0)(2.1)\times10^{8+(-10)}\ \text{m} = 6.3\times10^{-2}\ \text{m}$$

Powers/Roots

To raise numbers in scientific notation to any power, raise the digits to the given power and multiply the exponent by the power:

$$\sqrt{(3.61\times10^4)^3} = \sqrt{3.61^3\times10^{(4)(3)}} = (47.04\times10^{12})^{1/2}$$
$$= \sqrt{47.04}\times10^{(12)(1/2)} = 6.86\times10^6$$

EXAMPLE 1.2 Scientific Notation: Tsunami Warnings

Earthquake-generated tsunamis are so devastating because the entire ocean, from surface to bottom, participates in the wave motion. The speed of such waves is given by $v = \sqrt{gh}$, where $g = 9.8\ \text{m/s}^2$ is the gravitational acceleration, and h is the depth in meters. Determine a tsunami's speed in 3.0-km deep water.

EVALUATE That 3.0-km depth is 3.0×10^3 m, so we have

$$\begin{aligned} v = \sqrt{gh} &= \left[(9.8\ \text{m/s}^2)(3.0\times10^3\ \text{m})\right]^{1/2} = (29.4\times10^3\ \text{m}^2/\text{s}^2)^{1/2} \\ &= (2.94\times10^4\ \text{m}^2/\text{s}^2)^{1/2} = \sqrt{2.94}\times10^2\ \text{m/s} = 1.7\times10^2\ \text{m/s} \end{aligned}$$

where we wrote $29.4\times10^3\ \text{m}^2/\text{s}^2$ as $2.94\times10^4\ \text{m}^2/\text{s}^2$ in the second line in order to calculate the square root more easily. Converting the speed to km/h gives

$$\begin{aligned} 1.7\times10^2\ \text{m/s} &= \left(\frac{1.7\times10^2\ \cancel{\text{m}}}{\cancel{\text{s}}}\right)\left(\frac{1\ \text{km}}{1.0\times10^3\ \cancel{\text{m}}}\right)\left(\frac{3.6\times10^3\ \cancel{\text{s}}}{\text{h}}\right) \\ &= 6.1\times10^2\ \text{km/h} \end{aligned}$$

This speed—about 600 km/h—shows why even distant coastlines have little time to prepare for the arrival of a tsunami. ■

Significant Figures

How precise is that 1.7×10^2 m/s we calculated in Example 1.2? The two **significant figures** in this number imply that the value is closer to 1.7 than to 1.6 or 1.8. The fewer significant figures, the less precisely we can claim to know a given quantity.

In Example 1.2 we were, in fact, given two significant figures for both quantities. The mere act of calculating can't add precision, so we rounded our answer to two significant figures as well. Calculators and computers often give numbers with many figures, but most of those are usually meaningless.

What's Earth's circumference? It's $2\pi R_E$, of course. And π is approximately 3.14159. . . . But if you only know Earth's radius as 6.37×10^6 m, knowing π to more significant figures doesn't mean you can claim to know the circumference any more precisely. This example suggests a rule for handling calculations involving numbers with different precisions:

> In multiplication and division, the answer should have the same number of significant figures as the least precise of the quantities entering the calculation.

You're engineering an access ramp to a bridge whose main span is 1.248 km long. The ramp will be 65.4 m long. What will be the overall length? A simple calculation gives 1.248 km + 0.0654 km = 1.3134 km. How should you round this? You know the bridge length to ±0.001 km, so an addition this small is significant. Thus your answer should have three digits to the right of the decimal point, giving 1.313 km. Thus:

> In addition and subtraction, the answer should have the same number of digits to the right of the decimal point as the term in the sum or difference that has the smallest number of digits to the right of the decimal point.

In subtraction, this rule can quickly lead to loss of precision, as Example 1.3 illustrates.

EXAMPLE 1.3 Significant Figures: Nuclear Fuel

A uranium fuel rod is 3.241 m long before it's inserted in a nuclear reactor. After insertion, heat from the nuclear reaction has increased its length to 3.249 m. What's the increase in its length?

EVALUATE Subtraction gives 3.249 m − 3.241 m = 0.008 m or 8 mm. Should this be 8 mm or 8.000 mm? Just 8 mm. Subtraction affected only the last digit of the four-significant-figure lengths, leaving only one significant figure in the answer. ■

✓TIP Intermediate Results

Although it's important that your final answer reflect the precision of the numbers that went into it, any intermediate results should have at least one extra significant figure. Otherwise, rounding of intermediate results could alter your answer.

GOT IT? 1.1 Rank the numbers according to (a) their size and (b) the number of significant figures. Some may be of equal rank. 0.0008, 3.14×10^7, 2.998×10^{-9}, 55×10^6, 0.041×10^9

Estimation

Some problems in physics and engineering call for precise numerical answers. We need to know exactly how long to fire a rocket to put a space probe on course toward a distant planet, or exactly what size to cut the tiny quartz crystal whose vibrations set the pulse of a digital watch. But for many other purposes, we need only a rough idea of the size of a physical effect. And rough estimates help check whether the results of more difficult calculations make sense.

EXAMPLE 1.4 Estimation: Counting Brain Cells

Estimate the mass of your brain and the number of cells it contains.

EVALUATE My head is about 6 in. or 15 cm wide, but there's a lot of skull bone in there, so maybe my brain is about 10 cm or 0.1 m across. I don't know its exact shape, but for estimating, I'll take it to be a cube. Then its volume is $(10\,\text{cm})^3 = 1000\,\text{cm}^3$, or $10^{-3}\,\text{m}^3$. I'm mostly water, and water's density is 1 gram per cubic centimeter $(1\,\text{g/cm}^3)$, so my 1000-cm^3 brain has a mass of about 1 kg.

How big is a brain cell? I don't know, but Table 1.2 lists the diameter of a red blood cell as about 10^{-5} m. If brain cells are roughly the same size, then each cell has a volume of approximately $(10^{-5}\,\text{m})^3 = 10^{-15}\,\text{m}^3$. Then the number of cells in my 10^{-3}-m^3 brain is roughly

$$N = \frac{10^{-3}\,\text{m}^3/\text{brain}}{10^{-15}\,\text{m}^3/\text{cell}} = 10^{12}\,\text{cells/brain}$$

Crude though they are, these estimates aren't bad. The average adult brain's mass is about 1.3 kg, and it contains at least 10^{11} cells. ■

1.4 Strategies for Learning Physics

You can learn *about* physics, and you can learn to *do* physics. This book is for science and engineering students, so it emphasizes both. Learning about physics will help you appreciate the role of this fundamental science in explaining both natural and technological phenomena. Learning to do physics will make you adept at solving quantitative problems—finding answers to questions about how the natural world works and about how we forge the technologies at the heart of modern society.

Physics: Challenge and Simplicity

Physics problems can be challenging, calling for clever insight and mathematical agility. That challenge is what gives physics a reputation as a difficult subject. But underlying all of physics is only a handful of basic principles. Because physics is so fundamental, it's also inherently simple. There are only a few basic ideas to learn; if you really understand those, you can apply them in a wide variety of situations. These ideas and their applications are all connected, and we'll emphasize those connections and the underlying simplicity of physics by reminding you how the many examples, applications, and problems are manifestations of the same few basic principles. If you approach physics as a hodgepodge of unrelated laws and equations, you'll miss the point and make things difficult. But if you look for the basic principles, for connections among seemingly unrelated phenomena and problems, then you'll discover the underlying simplicity that reflects the scope and power of physics—the fundamental science.

Problem Solving: The IDEA Strategy

Solving a quantitative physics problem always starts with basic principles or concepts and ends with a precise answer expressed as either a numerical quantity or an algebraic expression. Whatever the principle, whatever the realm of physics, and whatever the specific situation, the path from principle to answer follows four simple steps—steps that make up a comprehensive strategy for approaching all problems in physics. Their acronym, IDEA, will help you remember these steps, and they'll be reinforced as we apply them over and over again in worked examples throughout the book. We'll generally write all four steps separately, although the examples in this chapter cut right to the EVALUATE phase. And in some chapters we'll introduce versions of this strategy tailored to specific material.

The IDEA strategy isn't a "cookbook" formula for working physics problems. Rather, it's a tool for organizing your thoughts, clarifying your conceptual understanding, developing and executing plans for solving problems, and assessing your answers. Here's the big IDEA:

PROBLEM-SOLVING STRATEGY 1.1 Physics Problems

INTERPRET The first step is to *interpret* the problem to be sure you know what it's asking. Then *identify* the applicable concepts and principles—Newton's laws of motion, conservation of energy, the first law of thermodynamics, Gauss's law, and so forth. Also *identify* the players in the situation—the object whose motion you're asked to describe, the forces acting, the thermodynamic system you're to analyze, the charges that produce an electric field, the components in an electric circuit, the light rays that will help you locate an image, and so on.

DEVELOP The second step is to *develop* a plan for solving the problem. It's always helpful and often essential to *draw* a diagram showing the situation. Your drawing should indicate objects, forces, and other physical entities. Labeling masses, positions, forces, velocities, heat flows, electric or magnetic fields, and other quantities will be a big help. Next, *determine* the relevant mathematical formulas—namely, those that contain the quantities you're given in the problem as well as the unknown(s) you're solving for. Don't just grab equations—rather, think about how each reflects the underlying concepts and principles that you've identified as applying to this problem. The plan you develop might include calculating intermediate quantities, finding values in a table or in one of this text's several appendices, or even solving a preliminary problem whose answer you need in order to get your final result.

EVALUATE Physics problems have numerical or symbolic answers, and you need to *evaluate* your answer. In this step you *execute* your plan, going in sequence through the steps you've outlined. Here's where your math skills come in. Use algebra, trig, or calculus, as needed, to solve your equations. It's a good idea to keep all numerical quantities, whether known or not, in symbolic form as you work through the solution of your problem. At the end you can plug in numbers and work the arithmetic to *evaluate* the numerical answer, if the problem calls for one.

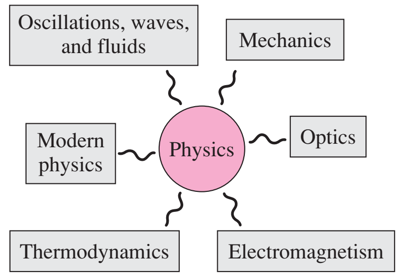
ASSESS Don't be satisfied with your answer until you *assess* whether it makes sense! Are the units correct? Do the numbers sound reasonable? Does the algebraic form of your answer work in obvious special cases, like perhaps "turning off" gravity or making an object's mass zero or infinite? Checking special cases not only helps you decide whether your answer makes sense but also can give you insights into the underlying physics. In worked examples, we'll often use this step to enhance your knowledge of physics by relating the example to other applications of physics.

Don't memorize the IDEA problem-solving strategy. Instead, grow to understand it as you see it applied in examples and as you apply it yourself in working end-of-chapter problems. This book has a number of additional features and supplements, discussed in the Preface, to help you develop your problem-solving skills.

CHAPTER 1 SUMMARY

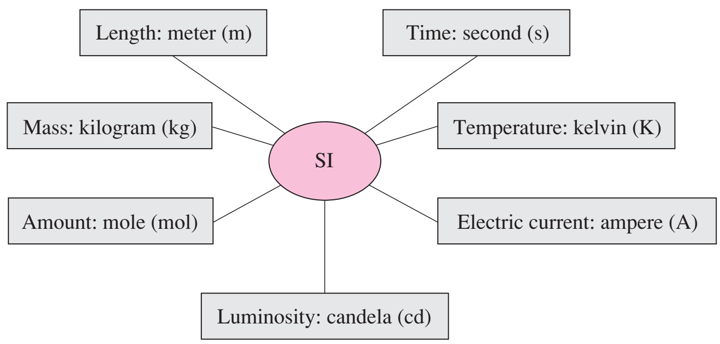
Big Picture

Physics is the fundamental science. It's convenient to consider several realms of physics, which together describe all that's known about physical reality:



Key Concepts and Equations

Numbers describing physical quantities must have units. The SI unit system comprises seven fundamental units:



In addition, physics uses geometric measures of angle.

Numbers are often written with prefixes or in scientific notation to express powers of 10. Precision is shown by the number of significant figures:

$$\text{Earth's radius} = \underbrace{6.37}_{\text{Three significant figures}} \times 10^{\overbrace{6}^{\text{Power of 10}}}\ \text{m} = 6.37\ \underbrace{\text{M}}_{\text{SI prefix for “}\times 10^{6}\text{”}}\text{m}$$

Applications

The IDEA strategy for solving physics problems follows four steps: Interpret, Develop, Evaluate, and Assess.

For Thought and Discussion

1. Explain why measurement standards based on laboratory procedures are preferable to those based on specific objects such as the international prototype kilogram.
2. Which measurement standards are now defined operationally? Which aren't?
3. When a computer that carries seven significant figures adds 1.000000 and 2.5310215, what's its answer? Why?
4. Why doesn't Earth's rotation provide a suitable time standard?
5. To raise a power of 10 to another power, you multiply the exponent by the power. Explain why this works.
6. What facts might a scientist use in estimating Earth's age?
7. How would you determine the length of a curved line?
8. Write $1/x$ as x to some power.
9. Emissions of carbon dioxide from fossil-fuel combustion are often expressed in gigatonnes per year, where 1 tonne $= 1000$ kg. But sometimes CO_2 emissions are given in petagrams per year. How are the two units related?

Exercises and Problems

Exercises

Section 1.2 Measurements and Units

10. The power output of a typical large power plant is 1000 megawatts (MW). Express this result in (a) W, (b) kW, and (c) GW.
11. The diameter of a hydrogen atom is about 0.1 nm, and the diameter of a proton is about 1 fm. How many times bigger than a proton is a hydrogen atom?
12. Use the definition of the meter to determine how far light travels in 1 ns.
13. In nanoseconds, how long is the period of the cesium-133 radiation used to define the second?
14. Lake Baikal in Siberia holds the world's largest quantity of fresh water, about 14 Eg. How many kilograms is that?
15. A hydrogen atom is about 0.1 nm in diameter. How many hydrogen atoms lined up side by side would make a line 1 cm long?
16. How long a piece of wire would you need to form a circular arc subtending an angle of 1.4 rad, if the radius of the arc is 8.1 cm?
17. Making a turn, a jetliner flies 2.1 km on a circular path of radius 3.4 km. Through what angle does it turn?
18. A car is moving at 35.0 mi/h. Express its speed in (a) m/s and (b) ft/s.
19. You have postage for a 1-oz letter but only a metric scale. What's the maximum mass your letter can have, in grams?
20. A year is very nearly $\pi \times 10^7$ s. By what percentage is this figure in error?
21. How many cubic centimeters are in a cubic meter?
22. Since the start of the industrial era, humankind has emitted about half an exagram of carbon to the atmosphere. What's that in tonnes $(\mathrm{t};\ 1\ \mathrm{t} = 1000\ \mathrm{kg})$?
23. A gallon of paint covers 350 ft^2. What's its coverage in m^2/L?
24. Highways in Canada have speed limits of 100 km/h. How does this compare with the 65 mi/h speed limit common in the United States?
25. One m/s is how many km/h?
26. A 3.0-lb box of grass seed will seed 2100 ft^2 of lawn. Express this coverage in m^2/kg.
27. A radian is how many degrees?

Section 1.3 Working with Numbers

28. Add 3.63105 m and 2.13103 km.
29. Divide 4.23103 m/s by 0.57 ms, and express your answer in m/s^2.
30. Add 5.131022 cm and 6.83103 mm, and multiply the result by 1.83104 N (N is the SI unit of force).
31. Find the cube root of 6.4×10^{19} without a calculator.
32. Add 1.46 m and 2.3 cm.
33. You're asked to specify the length of an updated aircraft model for a sales brochure. The original plane was 41 m long; the new model has a 3.6-cm-long radio antenna added to its nose. What length do you put in the brochure?
34. Repeat the preceding exercise, this time using 41.05 m as the airplane's original length.

Problems

35. To see why it's important to carry more digits in intermediate calculations, determine $(\sqrt{3})^3$ to three significant figures in two ways: (a) Find $\sqrt{3}$ and round to three significant figures, then cube and again round; and (b) find $\sqrt{3}$ to four significant figures, then cube and round to three significant figures.
36. You've been hired as an environmental watchdog for a big-city newspaper. You're asked to estimate the number of trees that go into one day's printing, given that half the newsprint comes from recycling, the rest from new wood pulp. What do you report?
37. The average dairy cow produces about 10^4 kg of milk per year. Estimate the number of dairy cows needed to keep the United States supplied with milk.
38. How many Earths would fit inside the Sun?
39. The average American uses electrical energy at the rate of about 1.5 kilowatts (kW). Solar energy reaches Earth's surface at an average rate of about 300 watts on every square meter. What fraction of the United States' land area would have to be covered with 20% efficient solar cells to provide all of our electrical energy?
40. You're writing a biography of the famous physicist Enrico Fermi, who was fond of estimation problems. Here's one problem Fermi posed: What's the number of piano tuners in Chicago? Give your estimate, and explain to your readers how you got it.
41. (a) Estimate the volume of water going over Niagara Falls each second. (b) The falls provides the outlet for Lake Erie; if the falls were shut off, estimate how long it would take Lake Erie to rise 1 m.
42. Estimate the number of air molecules in your dorm room.
43. A human hair is about 100 μm across. Estimate the number of hairs in a typical braid.
44. You're working in the fraud protection division of a credit-card company, and you're asked to estimate the chances that a 16-digit number chosen at random will be a valid credit-card number. What do you answer?
45. Bubble gum's density is about 1 g/cm^3. You blow an 8-g wad of gum into a bubble 10 cm in diameter. What's the bubble's thickness? (*Hint*: Think about spreading the bubble into a flat sheet. The surface area of a sphere is $4\pi r^2$.)
46. The Moon barely covers the Sun during a solar eclipse. Given that Moon and Sun are, respectively, 4×10^5 km and 1.5×10^8 km from Earth, determine how much bigger the Sun's diameter is

than the Moon's. If the Moon's radius is 1800 km, how big is the Sun?

47. The semiconductor chip at the heart of a personal computer is a square 4 mm on a side and contains 10^9 electronic components. (a) What's the size of each component, assuming they're square? (b) If a calculation requires that electrical impulses traverse 10^4 components on the chip, each a million times, how many such calculations can the computer perform each second? (*Hint*: The maximum speed of an electrical impulse is 3×10^8 m/s, close to the speed of light.)
48. Estimate the number of (a) atoms and (b) cells in your body.
49. When we write the number 3.6 as typical of a number with two significant figures, we're saying that the actual value is closer to 3.6 than to 3.5 or 3.7; that is, the actual value lies between 3.55 and 3.65. Show that the percent uncertainty implied by such two-significant-figure precision varies with the value of the number, being the lowest for numbers beginning with 9 and the highest for numbers beginning with 1. In particular, what is the percent uncertainty implied by the numbers (a) 1.1, (b) 5.0, and (c) 9.9?
50. Continental drift occurs at about the rate your fingernails grow. Estimate the age of the Atlantic Ocean, given that the eastern and western hemispheres have been drifting apart.
51. You're driving into Canada and trying to decide whether to fill your gas tank before or after crossing the border. Gas in the United States costs \$2.97/gallon, in Canada it's 94¢/L, and the Canadian dollar is worth 87¢ in U.S. currency. Where should you fill up?
52. In the 1908 London Olympics, the intended 26-mile marathon was extended 385 yards to put the end in front of the royal reviewing stand. This distance subsequently became standard. What's the marathon distance in kilometers, to the nearest meter?
53. Express the following with appropriate units and significant figures: (a) 1.0 m plus 1 mm, (b) 1.0 m times 1 mm, (c) 1.0 m minus 999 mm, (d) 1.0 m divided by 999 mm.
54. You're shopping for a new computer, and a salesperson claims the microprocessor chip in the model you're looking at contains 10 billion electronic components. The chip measures 5 mm on a side and uses 32-nm technology, meaning each component is 32 nm across. Is the salesperson right?
55. Café Milagro sells coffee online. A half-kilogram bag of coffee costs \$8.95, excluding shipping. If you order six bags, the shipping costs \$6.90. What's the cost per bag when you include shipping?
56. The world consumes energy at the rate of about 450 EJ per year, where the joule (J) is the SI energy unit. Convert this figure to watts (W), where 1 W = 1 J/s, and then estimate the average per capita energy consumption rate in watts.

Passage Problems

The human body contains about 10^{14} cells, and the diameter of a typical cell is about 10 μm. Like all ordinary matter, cells are made of atoms; a typical atomic diameter is 0.1 nm.

57. How does the number of atoms in a cell compare with the number of cells in the body?
    a. greater b. smaller c. about the same
58. The volume of a cell is about
    a. 10^{-10} m^3. b. 10^{-15} m^3. c. 10^{-20} m^3. d. 10^{-30} m^3.
59. The mass of a cell is about
    a. 10^{-10} kg. b. 10^{-12} kg. c. 10^{-14} kg. d. 10^{-16} kg.
60. The number of atoms in the body is closest to
    a. 10^{14}. b. 10^{20}. c. 10^{30}. d. 10^{40}.

Answers to Chapter Questions

Answer to Chapter Opening Question

All of them!

Answers to GOT IT? Question

1.1. (a) 2.998×10^{-9}, 0.0008, 3.14×10^7, 0.041×10^9, 55×10^6
(b) 0.0008, 0.041×10^9 and 55×10^6 (with two significant figures each), 3.14×10^7, 2.998×10^{-9}

PART ONE OVERVIEW

Mechanics

Hikers check their position using signals from GPS satellites.

A wilderness hiker uses the Global Positioning System to follow her chosen route. A farmer plows a field with centimeter-scale precision, guided by GPS and saving precious fuel as a result. One scientist uses GPS to track endangered elephants, another to study the accelerated flow of glaciers as Earth's climate warms. Our deep understanding of motion is what lets us use a constellation of satellites, 20,000 km up and moving faster than 10,000 km/h, to find positions on Earth so precisely.

Motion occurs at all scales, from the intricate dance of molecules at the heart of life's cellular mechanics, to the everyday motion of cars, baseballs, and our own bodies, to the trajectories of GPS and TV satellites and of spacecraft exploring the distant planets, to the stately motions of the celestial bodies themselves and the overall expansion of the universe. The study of motion is called **mechanics**. The 11 chapters of Part 1 introduce the physics of motion, first for individual bodies and then for complicated systems whose constituent parts move relative to one another.

We explore motion here from the viewpoint of Newtonian mechanics, which applies accurately in all cases except the subatomic realm and when relative speeds approach that of light. The Newtonian mechanics of Part 1 provides the groundwork for much of the material in subsequent parts, until, in the book's final chapters, we extend mechanics into the subatomic and high-speed realms.

2 Motion in a Straight Line

The server tosses the tennis ball straight up and hits it on its way down. Right at its peak height, the ball has zero velocity, but what's its acceleration?

New Concepts, New Skills

By the end of this chapter you should be able to

- Explain the meanings of and relationships among displacement, velocity, and acceleration in one-dimensional motion (2.1–2.3).
- Calculate average and instantaneous velocities and accelerations (2.1–2.3).
- Solve problems involving motion under constant acceleration, including the acceleration of gravity at Earth's surface (2.4, 2.5).

Connecting Your Knowledge

- From Chapter 1, you should be comfortable working with SI units, and you should be able to convert from other unit systems (1.2).
- You should be able to express quantitative answers with the correct number of significant figures (1.3).
- You should be ready to apply the four problem-solving steps of the IDEA strategy (1.4).

Electrons swarming around atomic nuclei, cars speeding along a highway, blood coursing through your veins, galaxies rushing apart in the expanding universe—all these are examples of matter in motion. The study of motion without regard to its cause is called **kinematics** (from the Greek "kinema," or motion, as in motion pictures). This chapter deals with the simplest case: a single object moving in a straight line. Later, we generalize to motion in more dimensions and with more complicated objects. But the basic concepts and mathematical techniques we develop here continue to apply.

2.1 Average Motion

You drive 15 minutes to a pizza place 10 km away, grab your pizza, and return home in another 15 minutes. You've traveled a total distance of 20 km, and the trip took half an hour, so your **average speed**—distance divided by time—was 40 kilometers per hour. To describe your motion more precisely, we introduce the quantity x that gives your position

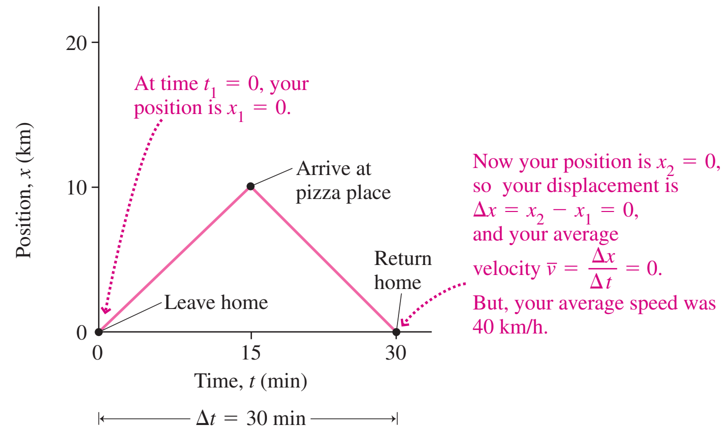


FIGURE 2.1 Position versus time for the pizza trip.

at any time t. We then define **displacement**, Δx, as the net change in position: $\Delta x = x_2 - x_1$, where x_1 and x_2 are your starting and ending positions, respectively. Your **average velocity**, $\bar{v}$, is displacement divided by the time interval:

$$\bar{v} = \frac{\Delta x}{\Delta t} \quad \text{(average velocity)} \tag{2.1}$$

where $\Delta t = t_2 - t_1$ is the interval between your ending and starting times. The bar in $\bar{v}$ indicates an average quantity (and is read "v bar"). The symbol Δ (capital Greek delta) stands for "the change in." For the round trip to the pizza place, your overall displacement was zero and therefore your average velocity was also zero—even though your average *speed* was not (Fig. 2.1).

Directions and Coordinate Systems

It matters whether you go north or south, east or west. Displacement therefore includes not only *how far* but also *in what direction*. For motion in a straight line, we can describe both properties by taking position coordinates x to be positive going in one direction from some origin, and negative in the other. This gives us a one-dimensional **coordinate system**. The choice of coordinate system—both of origin and of which direction is positive—is entirely up to you. The coordinate system isn't physically real; it's just a convenience we create to help in the mathematical description of motion.

Figure 2.2 shows some midwestern cities that lie on a north-south line. We've established a coordinate system with northward direction positive and origin at Kansas City. Arrows show displacements from Houston to Des Moines and from International Falls to Des Moines; the former is approximately $+1300$ km, and the latter is approximately -750 km, with the minus sign indicating a southward direction. Suppose the Houston-to-Des Moines trip takes 2.6 hours by plane; then the average velocity is $(1300\text{ km})/(2.6\text{ h}) = 500$ km/h. If the International Falls-to-Des Moines trip takes 10 h by car, then the average velocity is $(-750\text{ km})/(10\text{ h}) = -75$ km/h; again, the minus sign indicates southward.

In calculating average velocity, all that matters is the overall displacement. Maybe that trip from Houston to Des Moines was a nonstop flight going 500 km/h. Or maybe it involved a faster plane that stopped for half an hour in Kansas City. Maybe the plane even went first to Minneapolis, then backtracked to Des Moines. No matter: The displacement remains 1300 km and, as long as the total time is 2.6 h, the average velocity remains 500 km/h.

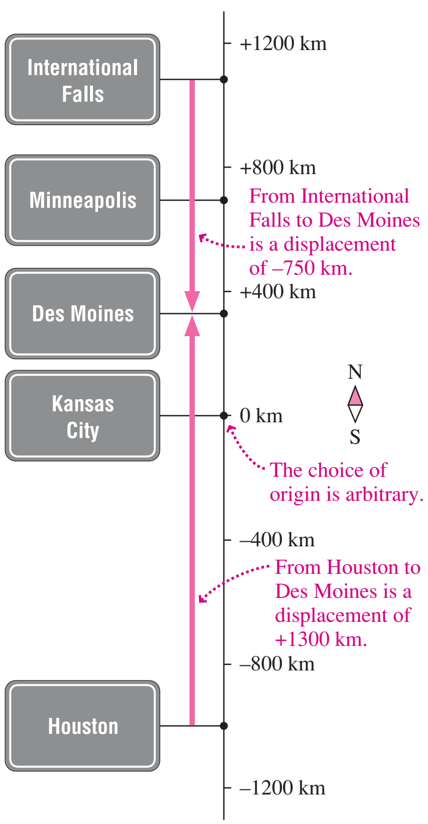


FIGURE 2.2 Describing motion in the central United States.

GOT IT? 2.1 We just described three trips from Houston to Des Moines: (a) direct; (b) with a stop in Kansas City, and (c) via Minneapolis. For which of these trips is the average speed the same as the average velocity? Where the two differ, which is greater?

EXAMPLE 2.1 Speed and Velocity: Flying with a Connection

To get a cheap flight from Houston to Kansas City—a distance of 1000 km—you have to connect in Minneapolis, 700 km north of Kansas City. The flight to Minneapolis takes 2.2 h, then you have a 30-min layover, and then a 1.3-h flight to Kansas City. What are your average velocity and your average speed on this trip?

INTERPRET We interpret this as a one-dimensional kinematics problem involving the distinction between velocity and speed, and we identify three distinct travel segments: the two flights and the layover. We identify the key concepts as speed and velocity; their distinction is clear from our pizza example.

DEVELOP Figure 2.2 is our drawing. We determine that Equation 2.1, $\bar{v} = \Delta x/\Delta t$, will give the average velocity, and that the average speed is the total distance divided by the total time. We develop our plan: Find the displacement and the total time, and use those values to get the average velocity; then find the total distance traveled and use that along with the total time to get the average speed.

EVALUATE You start in Houston and end up in Kansas City, for a displacement of 1000 km—regardless of how far you actually traveled. The total time for the three segments is $\Delta t = 2.2\text{ h} + 0.50\text{ h} + 1.3\text{ h} = 4.0\text{ h}$. Then the average velocity, from Equation 2.1, is

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{1000\text{ km}}{4.0\text{ h}} = 250\text{ km/h}$$

However, that Minneapolis connection means you've gone an extra 2×700 km, for a total distance of 2400 km in 4 hours. Thus your average speed is $(2400\text{ km})/(4.0\text{ h}) = 600$ km/h, more than twice your average velocity.

ASSESS Make sense? Average velocity depends only on the net displacement between the starting and ending points. Average speed takes into account the actual distance you travel—which can be a lot longer on a circuitous trip like this one. So it's entirely reasonable that the average speed should be greater. ■

2.2 Instantaneous Velocity

Geologists determine the velocity of a lava flow by dropping a stick into the lava and timing how long it takes the stick to go a known distance (Fig. 2.3*a*). Dividing the distance by the time then gives the average velocity. But did the lava flow faster at the beginning of the interval? Or did it speed up and slow down again? Clearly, velocity can change over time. To understand motion in all its detail, we need to know the velocity at each instant.

Geologists could explore that detail with a series of observations taken over smaller intervals of time and distance (Fig. 2.3*b*). As the size of the intervals shrinks, a more detailed picture of the motion emerges. In the limit of very small intervals, we're measuring the velocity at a single instant. This is the **instantaneous velocity**, or simply the **velocity.** The magnitude of the instantaneous velocity is the **instantaneous speed**.

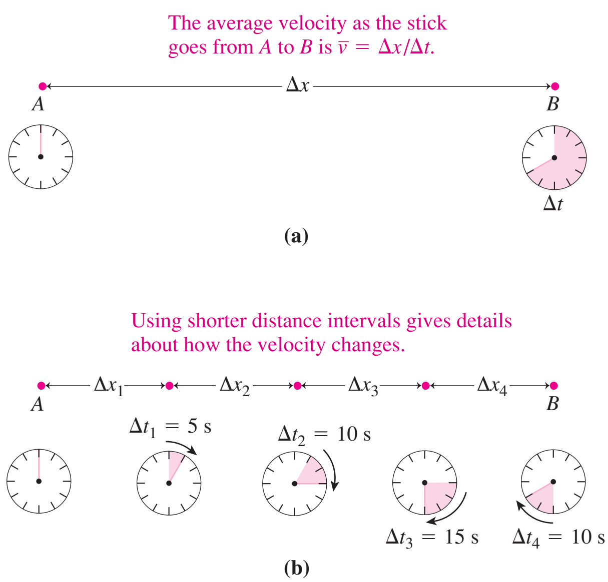


FIGURE 2.3 Determining the velocity of a lava flow.

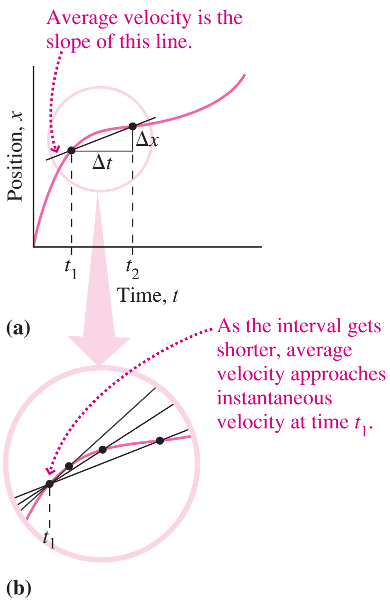


FIGURE 2.4 Position-versus-time graph for the motion in Fig. 2.3.

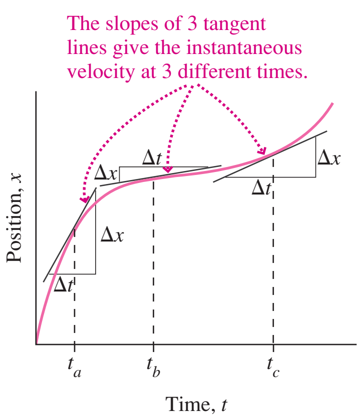


FIGURE 2.5 The instantaneous velocity is the slope of the tangent line.

You might object that it's impossible to achieve that limit of an arbitrarily small time interval. With observational measurements that's true, but calculus lets us go there. Figure 2.4*a* is a plot of position versus time for the stick in the lava flow shown in Fig. 2.3. Where the curve is steep, the position changes rapidly with time—so the velocity is greater. Where the curve is flatter, the velocity is lower. Study the clocks in Fig. 2.3*b* and you'll see that the stick starts out moving rapidly, then slows, and then speeds up a bit at the end. The curve in Fig. 2.4*a* reflects this behavior.

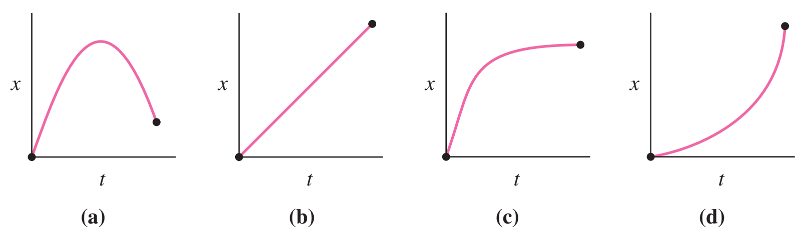
Suppose we want the instantaneous velocity at the time marked t_1 in Fig. 2.4*a*. We can approximate this quantity by measuring the displacement Δx over the interval Δt between t_1 and some later time t_2: the ratio $\Delta x/\Delta t$ is then the average velocity over this interval. Note that this ratio is the slope of a line drawn through points on the curve that mark the ends of the interval.

Figure 2.4*b* shows what happens as we make the time interval Δt arbitrarily small: Eventually, the line between the two points becomes indistinguishable from the tangent line to the curve. That tangent line has the same slope as the curve right at the point we're interested in, and therefore it defines the instantaneous velocity at that point. We write this mathematically by saying that the instantaneous velocity is the limit, as the time interval Δt becomes arbitrarily close to zero, of the ratio of displacement Δx to Δt:

$$v = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} \tag{2.2a}$$

You can imagine making the interval Δt as close to zero as you like, getting ever better approximations to the instantaneous velocity. Given a graph of position versus time, an easy approach is to "eyeball" the tangent line to the graph at a point you're interested in; its slope is the instantaneous velocity (Fig. 2.5).

GOT IT? 2.2 The figures show position-versus-time graphs for four objects. Which object is moving with constant speed? Which reverses direction? Which starts slowly and then speeds up?



Given position as a mathematical function of time, calculus provides a quick way to find instantaneous velocity. In calculus, the result of the limiting process described in Equation 2.2a is called the **derivative** of x with respect to t and is given the symbol dx/dt:

$$\frac{dx}{dt} = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t}$$

The quantities dx and dt are called **infinitesimals**; they represent vanishingly small quantities that result from the limiting process. We can then write Equation 2.2a as

$$v = \frac{dx}{dt} \qquad \text{(instantaneous velocity)} \tag{2.2b}$$

Given position x as a function of time t, calculus shows how to find the velocity $v = dx/dt$. Consult Tactics 2.1 if you haven't yet seen derivatives in your calculus class, or if you need a refresher.

TACTICS 2.1 **Taking Derivatives**

You don't have to go through an elaborate limiting process every time you want to find an instantaneous velocity. That's because calculus provides formulas for the derivatives of common functions. For example, any function of the form $x = bt^n$, where b and n are constants, has the derivative

$$\frac{dx}{dt} = nbt^{n-1} \tag{2.3}$$

Appendix A lists derivatives of other common functions.

EXAMPLE 2.2 **Instantaneous Velocity: A Rocket Ascends**

The altitude of a rocket in the first half-minute of its ascent is given by $x = bt^2$, where the constant b is $2.90\ \text{m/s}^2$. Find a general expression for the rocket's velocity as a function of time and from it the instantaneous velocity at $t = 20\ \text{s}$. Also find an expression for the average velocity, and compare your two velocity expressions.

INTERPRET We interpret this as a problem involving the comparison of two distinct but related concepts: instantaneous velocity and average velocity. We identify the rocket as the object whose velocities we're interested in.

DEVELOP Equation 2.2b, $v = dx/dt$, gives the instantaneous velocity and Equation 2.1, $\bar{v} = \Delta x/\Delta t$, gives the average velocity. Our plan is to use Equation 2.3, $dx/dt = nbt^{n-1}$, to evaluate the derivative that gives the instantaneous velocity. Then we can use Equation 2.1 for the average velocity, but first we'll need to determine the displacement from the equation we're given for the rocket's position.

EVALUATE Applying Equation 2.2b with position given by $x = bt^2$ and using Equation 2.3 to evaluate the derivative, we have

$$v = \frac{dx}{dt} = \frac{d(bt^2)}{dt} = 2bt$$

for the instantaneous velocity. Evaluating at $t = 20\ \text{s}$ with $b = 2.90\ \text{m/s}^2$ gives $v = 116\ \text{m/s}$. For the average velocity we need the total displacement at 20 s. Since $x = bt^2$, Equation 2.1 gives

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{bt^2}{t} = bt$$

where we've used $x = bt^2$ for Δx and t for Δt because both position and time are taken to be zero at liftoff. Comparison with our earlier result shows that the average velocity from liftoff to any particular time is exactly half the instantaneous velocity at that time.

ASSESS Make sense? Yes. The rocket's speed is always increasing, so its velocity at the end of any time interval is greater than the average velocity over that interval. The fact that the average velocity is exactly half the instantaneous velocity results from the quadratic (t^2) dependence of position on time.

✓TIP Language

Language often holds clues to the meaning of physical concepts. In this example we speak of the *instantaneous* velocity *at* a particular time. That wording should remind you of the limiting process that focuses on a single instant. In contrast, we speak of the *average* velocity *over* a time interval, since averaging explicitly involves a range of times.

■

2.3 Acceleration

When velocity changes, as in Example 2.2, an object is said to undergo **acceleration**. Quantitatively, we define acceleration as the rate of change of velocity, just as we defined velocity as the rate of change of position. The **average acceleration** over a time interval Δt is

$$\bar{a} = \frac{\Delta v}{\Delta t} \quad \text{(average acceleration)} \tag{2.4}$$

where Δv is the change in velocity and the bar on $\bar{a}$ indicates that this is an average value. Just as we defined instantaneous velocity through a limiting procedure, we define **instantaneous acceleration** as

$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt} \quad \text{(instantaneous acceleration)} \tag{2.5}$$

As we did with velocity, we also use the term *acceleration* alone to mean instantaneous acceleration.

In one-dimensional motion, acceleration is either in the direction of the velocity or opposite it. In the former case the accelerating object speeds up, whereas in the latter it slows (Fig. 2.6).

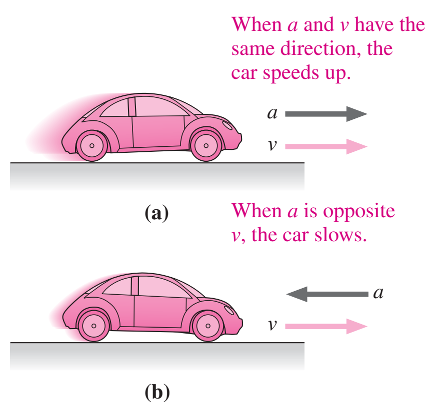


FIGURE 2.6 Acceleration and velocity.

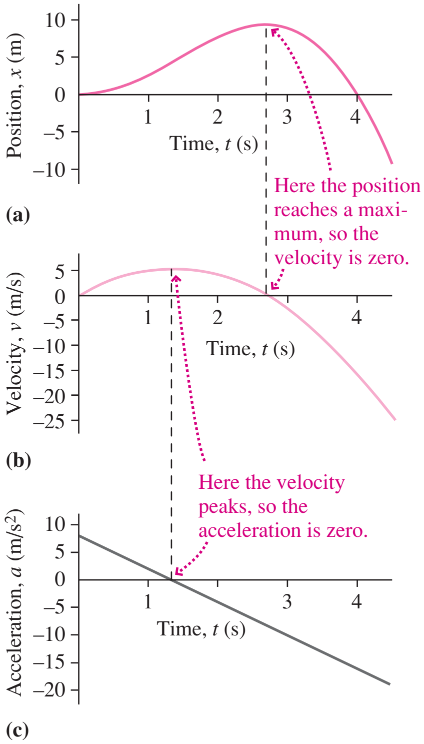


FIGURE 2.7 (a) Position, (b) velocity, and (c) acceleration versus time.

Although slowing is sometimes called *deceleration*, it's simpler to use *acceleration* to describe the time rate of change of velocity no matter what's happening. With two-dimensional motion, we'll find much richer relationships between the directions of velocity and acceleration.

Since acceleration is the rate of change of velocity, its units are (distance per time) per time, or distance/time^2. In SI, that's m/s^2. Sometimes acceleration is given in mixed units; for example, a car going from 0 to 60 mi/h in 10 s has an average acceleration of 6 mi/h/s.

Position, Velocity, and Acceleration

Figure 2.7 shows graphs of position, velocity, and acceleration for an object undergoing one-dimensional motion. In Fig. 2.7*a* the rise and fall of the position-versus-time curve show that the object first moves away from the origin, reverses, then reaches the origin again at $t = 4$ s. It then continues moving into the region $x < 0$. Velocity, shown in Fig. 2.7*b*, is the slope of the position-versus-time curve in Fig. 2.7*a*. Note that the velocity is great where the curve in Fig. 2.7*a* is steep—that is, where position is changing most rapidly. At the peak of the position curve, the object is momentarily at rest as it reverses, so there the position curve is flat and the velocity is zero. After the object reverses, at about 2.7 s, it's heading in the negative x-direction and so its velocity is negative.

Just as velocity is the slope of the position-versus-time curve, acceleration is the slope of the velocity-versus-time curve. Initially that slope is positive—velocity is increasing—but

CONCEPTUAL EXAMPLE 2.1 Acceleration Without Velocity?

Can an object be accelerating even though it's not moving?

EVALUATE Figure 2.7 shows that velocity is the *slope* of the position curve—and the slope depends on how the position is *changing*, not on its actual value. Similarly, acceleration depends only on the rate of *change* of velocity, not on velocity itself. So there's no intrinsic reason why there can't be acceleration at an instant when velocity is zero.

ASSESS Figure 2.8, which shows a ball thrown straight up, is a case in point. Right at the peak of its flight, the ball's velocity is instantaneously zero. But just before the peak it's moving upward, and just after it's moving downward. No matter how small a time interval you consider, the velocity is always changing. Therefore, the ball is accelerating, even right at the instant its velocity is zero.

MAKING THE CONNECTION Just 0.010 s before it peaks, the ball in Fig. 2.8 is moving upward at 0.098 m/s; 0.010 s after it peaks, it's moving downward with the same speed. What's its average acceleration over this 0.02-s interval?

EVALUATE Equation 2.4 gives the average acceleration: $\bar{a} = \Delta v/\Delta t = (-0.098\ \text{m/s} - 0.098\ \text{m/s})/(0.020\ \text{s}) = -9.8\ \text{m/s}^2$. Here we've implicitly chosen a coordinate system with a positive upward direction, so both the initial velocity and the acceleration are negative. The time interval is so small that our result must be close to the instantaneous acceleration right at the peak—when the velocity is zero. You might recognize 9.8 m/s^2 as the acceleration due to Earth's gravity.

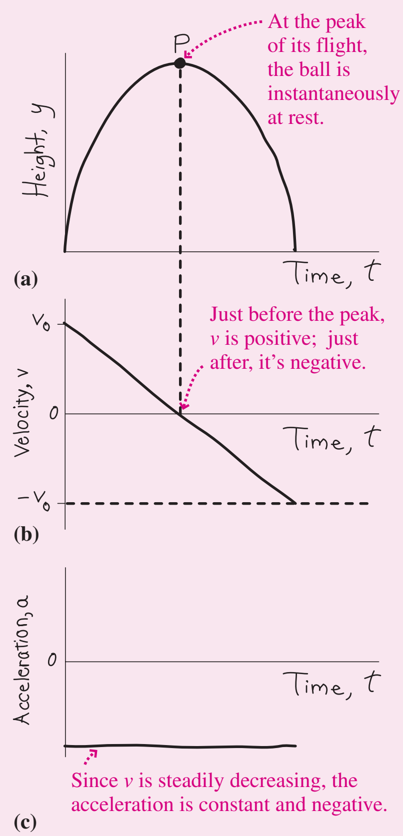


FIGURE 2.8 Our sketch for Conceptual Example 2.1.

eventually it peaks at the point of maximum velocity and zero acceleration and then it decreases. That velocity decrease corresponds to a negative acceleration, as shown clearly in the region of Fig. 2.7*c* beyond about 1.3 s.

Acceleration is the rate of change of velocity, and velocity is the rate of change of position. That makes acceleration the rate of change of the rate of change of position. Mathematically, acceleration is the **second derivative** of position with respect to time. Symbolically, we write the second derivative as d^2x/dt^2; this is just a symbol and doesn't mean that anything is actually squared. Then the relationship among acceleration, velocity, and position can be written

$$a = \frac{dv}{dt} = \frac{d}{dt}\left(\frac{dx}{dt}\right) = \frac{d^2x}{dt^2} \tag{2.6}$$

2.4 Constant Acceleration

The description of motion has an especially simple form when acceleration is constant. Suppose an object starts at time $t = 0$ with some initial velocity v_0 and constant acceleration a. Later, at some time t, it has velocity v. Because the acceleration doesn't change, its average and instantaneous values are identical, so we can write

$$a = \bar{a} = \frac{\Delta v}{\Delta t} = \frac{v - v_0}{t - 0}$$

or, rearranging,

$$v = v_0 + at \qquad \text{(for constant acceleration only)} \tag{2.7}$$

This equation says that the velocity changes from its initial value by an amount that is the product of acceleration and time.

✓TIP Know Your Limits

Many equations we develop are special cases of more general laws, and they're limited to special circumstances. Equation 2.7 is a case in point: It applies *only when acceleration is constant*.

Having determined velocity as a function of time, we now consider position. With constant acceleration, velocity increases steadily—and thus the average velocity over an interval is the average of the velocities at the beginning and the end of that interval. So we can write

$$\bar{v} = \tfrac{1}{2}(v_0 + v) \tag{2.8}$$

for the average velocity over the interval from $t = 0$ to some later time when the velocity is v. We can also write the average velocity as the change in position divided by the time interval. Suppose that at time 0 our object was at position x_0. Then its average velocity over a time interval from 0 to time t is

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{x - x_0}{t - 0}$$

where x is the object's position at time t. Equating this expression for $\bar{v}$ with the expression in Equation 2.8 gives

$$x = x_0 + \bar{v}t = x_0 + \tfrac{1}{2}(v_0 + v)t \tag{2.9}$$

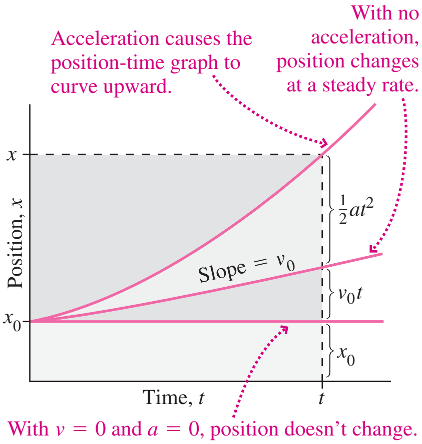


FIGURE 2.9 Meaning of the terms in Equation 2.10.

Table 2.1 Equations of Motion for Constant Acceleration

Equation	**Contains**	**Number**
$v = v_0 + at$	v, a, t; no x	2.7
$x = x_0 + \frac{1}{2}(v_0 + v)t$	x, v, t; no a	2.9
$x = x_0 + v_0t + \frac{1}{2}at^2$	x, a, t; no v	2.10
$v^2 = v_0^2 + 2a(x - x_0)$	x, v, a; no t	2.11

But we already found the instantaneous velocity v that appears in this expression; it's given by Equation 2.7. Substituting and simplifying then give the position as a function of time:

$$x = x_0 + v_0t + \tfrac{1}{2}at^2 \qquad \text{(for constant acceleration only)} \tag{2.10}$$

Does Equation 2.10 make sense? With no acceleration $(a = 0)$, position would increase linearly with time, at a rate given by the initial velocity v_0. With constant acceleration, the additional term $\frac{1}{2}at^2$ describes the effect of the ever-changing velocity; time is squared because the longer the object travels, the faster it moves, so the more distance it covers in a given time. Figure 2.9 shows the meaning of the terms in Equation 2.10.

How much runway do I need to land a jetliner, given touchdown speed and a constant acceleration? A question like this involves position, velocity, and acceleration without explicit mention of time. So we solve Equation 2.7 for time, $t = (v - v_0)/a$, and substitute this expression for t in Equation 2.9 to write

$$x - x_0 = \frac{1}{2}\frac{(v_0 + v)(v - v_0)}{a}$$

or, since $(a + b)(a - b) = a^2 - b^2$,

$$v^2 = v_0^2 + 2a(x - x_0) \tag{2.11}$$

Equations 2.7, 2.9, 2.10, and 2.11 link all possible combinations of position, velocity, and acceleration for motion with constant acceleration. We summarize them in Table 2.1, and remind you that they apply *only* in the case of constant acceleration.

Using the Equations of Motion

The equations in Table 2.1 fully describe motion under constant acceleration. Don't regard them as separate laws, but recognize them as complementary descriptions of a single underlying phenomenon—one-dimensional motion with *constant acceleration*. Having several equations provides convenient starting points for approaching problems. Don't memorize these equations, but grow familiar with them as you work problems. We now offer a strategy for solving problems about one-dimensional motion with *constant acceleration* using these equations.

PROBLEM-SOLVING STRATEGY 2.1 Motion with Constant Acceleration

INTERPRET Interpret the problem to be sure it asks about motion with *constant acceleration*. Next, identify the object(s) whose motion you're interested in.

DEVELOP Draw a diagram with appropriate labels, and choose a coordinate system. For instance, sketch the initial and final physical situations, or draw a position-versus-time graph. Then determine which equations of motion from Table 2.1 contain the quantities you're given and will be easiest to solve for the unknown(s).

EVALUATE Solve the equations in symbolic form and then evaluate numerical quantities.

ASSESS Does your answer make sense? Are the units correct? Do the numbers sound reasonable? What happens in special cases—for example, when a distance, velocity, acceleration, or time becomes very large or very small?

The next two examples are typical of problems involving constant acceleration. Example 2.3 is a straightforward application of the equations we've just derived to a single object. Example 2.4 involves two objects, in which case we need to write equations describing the motions of both objects.

EXAMPLE 2.3 Motion with Constant Acceleration: Landing a Jetliner

A jetliner touches down at 270 km/h. The plane then decelerates (i.e., undergoes acceleration directed opposite its velocity) at 4.5 m/s^2. What's the minimum runway length on which this aircraft can land?

INTERPRET We interpret this as a problem involving one-dimensional motion with constant acceleration and identify the airplane as the object of interest.

DEVELOP We determine that Equation 2.11, $v^2 = v_0^2 + 2a(x - x_0)$, relates distance, velocity, and acceleration, so our plan is to solve that equation for the minimum runway length. We want the airplane to come to a stop, so the final velocity v is 0, and v_0 is the initial touchdown velocity. If x_0 is the touchdown point, then the quantity $x - x_0$ is the distance we're interested in; we'll call this Δx.

EVALUATE Setting $v = 0$ and solving Equation 2.11 then give

$$\Delta x = \frac{-v_0^2}{2a} = \frac{-[(270\ \text{km/h})(1000\ \text{m/km})(1/3600\ \text{h/s})]^2}{(2)(-4.5\ \text{m/s}^2)} = 625\ \text{m}$$

Note that we used a negative value for the acceleration because the plane's acceleration is directed opposite its velocity—which we chose as the positive x-direction. We also converted the speed to m/s for compatibility with the SI units given for acceleration.

ASSESS Make sense? That 625 m is just over one-third of a mile, which seems a bit short. However, this is an absolute minimum with no margin of safety. For full-size jetliners, the standard for minimum landing runway length is about 5000 feet or 1.5 km.

✓TIP Be Careful with Mixed Units

Frequently problems are stated in units other than SI. Although it's possible to work consistently in other units, when in doubt, convert to SI. In this problem, the acceleration is originally in SI units but the velocity isn't—a sure indication of the need for conversion.

■

EXAMPLE 2.4 Motion with Two Objects: Speed Trap!

A speeding motorist zooms through a 50 km/h zone at 75 km/h (that's 21 m/s) without noticing a stationary police car. The police officer immediately heads after the speeder, accelerating at 2.5 m/s^2. When the officer catches up to the speeder, how far down the road are they, and how fast is the police car going?

INTERPRET We interpret this as *two* problems involving one-dimensional motion with constant acceleration. We identify the objects in question as the speeding car and the police car. Their motions are related because we're interested in the point where the two coincide.

DEVELOP It's helpful to draw a sketch showing qualitatively the position-versus-time graphs for the two cars. Since the speeding car moves with constant speed, its graph is a straight line. The police car is accelerating from rest, so its graph starts flat and gets increasingly steeper. Our sketch in Fig. 2.10 shows clearly the point we're interested in, when the two cars coincide for the second time. Equation 2.10, $x = x_0 + v_0t + \frac{1}{2}at^2$, gives position versus time with constant acceleration. Our plan is (1) to write versions of this equation specialized to each car, (2) to equate the resulting position expressions to find the time when the cars coincide, and (3) to find the corresponding position and the police car's velocity. For the latter we'll use Equation 2.7, $v = v_0 + at$.

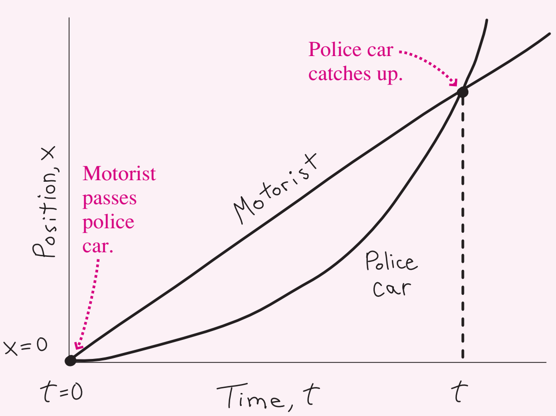


FIGURE 2.10 Our sketch of position versus time for the cars in Example 2.4.

EVALUATE Let's take the origin to be the point where the speeder passes the police car and $t = 0$ to be the corresponding time, as marked in Fig. 2.10. Then $x_0 = 0$ in Equation 2.10 for both cars, while the speeder has no acceleration and the police car has no initial velocity. Thus our two versions of Equation 2.10 are

$$x_s = v_{s0}t \ \ (\text{speeder}) \quad \text{and} \quad x_p = \tfrac{1}{2}a_pt^2 \ (\text{police car})$$

Equating x_s and x_p tells when the speeder and the police car are at the same place, so we write $v_{s0}t = \frac{1}{2}a_pt^2$. This equation is satisfied when $t = 0$ or $t = 2v_{s0}/a_p$. Why two answers? We asked for *any* times when the two cars are in the same place. That includes the initial encounter at $t = 0$ as well as the later time $t = 2v_{s0}/a_p$ when the police car catches the speeder; both points are shown on our sketch. *Where* does this occur? We can evaluate using $t = 2v_{s0}/a_p$ in the speeder's equation:

$$x_s = v_{s0}t = v_{s0}\frac{2v_{s0}}{a_p} = \frac{2v_{s0}^2}{a_p} = \frac{(2)(21\ \text{m/s})^2}{2.5\ \text{m/s}^2} = 350\ \text{m}$$

Equation 2.7 then gives the police car's speed at this time:

$$v_p = a_pt = a_p\frac{2v_{s0}}{a_p} = 2v_{s0} = 150\ \text{km/h}$$

ASSESS Make sense? As Fig. 2.10 shows, the police car starts from rest and undergoes constant acceleration, so it has to be going faster at the point where the two cars meet. In fact, it's going twice as fast—again, as in Example 2.2, that's because the police car's position depends quadratically on time. That quadratic dependence also tells us that the police car's position-versus-time graph in Fig. 2.10 is a parabola. ■

GOT IT? 2.3 The police car in Example 2.4 starts with zero velocity and is going at twice the car's velocity when it catches up to the car. So at some intermediate instant it must be going at the same velocity as the car. Is that instant (a) halfway between the times when the two cars coincide, (b) closer to the time when the speeder passes the stationary police car, or (c) closer to the time when the police car catches the speeder?

FIGURE 2.11 Strobe photo of a falling ball. Successive images are farther apart, showing that the ball is accelerating.

2.5 The Acceleration of Gravity

Drop an object, and it falls at an increasing rate, accelerating because of gravity (Fig. 2.11). The acceleration is constant for objects falling near Earth's surface, and furthermore it has the same value for all objects. This value, the **acceleration of gravity**, is designated g and is approximately 9.8 m/s^2 near Earth's surface.

The acceleration of gravity applies strictly only in **free fall**—motion under the influence of gravity alone. Air resistance, in particular, may dramatically alter the motion, giving the false impression that gravity acts differently on lighter and heavier objects. As early as the year 1600, Galileo is reputed to have shown that all objects have the same acceleration by dropping objects off the Leaning Tower of Pisa. Astronauts have verified that a feather and a hammer fall with the same acceleration on the airless Moon—although that acceleration is less than on Earth.

Although g is approximately constant near Earth's surface, it varies slightly with latitude and even local geology. The variation with altitude becomes substantial over distances of tens to hundreds of kilometers. But nearer Earth's surface it's a good approximation to take g as strictly constant. Then an object in free fall undergoes constant acceleration, and the equations of Table 2.1 apply. In working gravitational problems, we usually replace x with y to designate the vertical direction. If we make the arbitrary but common choice that the upward direction is positive, then acceleration a becomes $-g$ because the acceleration is downward.

EXAMPLE 2.5 Constant Acceleration Due to Gravity: Cliff Diving

A diver drops from a 10-m-high cliff. At what speed does he enter the water, and how long is he in the air?

INTERPRET This is a case of constant acceleration due to gravity, and the diver is the object of interest. The diver drops a known distance starting from rest, and we want to know the speed and time when he hits the water.

DEVELOP Figure 2.12 is a sketch showing what the diver's position versus time should look like. We've incorporated what we know: the initial position 10 m above the water, the start from rest, and the downward acceleration that results in a parabolic position-versus-time curve. Given the dive height, Equation 2.11, $v^2 = v_0^2 + 2a(x - x_0)$, determines the speed v. Since the diver starts from rest, $v_0 = 0$ and the equation becomes $v^2 = -2g(y - y_0)$. So our plan is first to solve for the speed at the water; then use Equation 2.7, $v = v_0 + at$, to get the time.

EVALUATE Our sketch shows that we've chosen $y = 0$ at the water, so $y_0 = 10$ m and Equation 2.11 gives

$$|v| = \sqrt{-2g(y - y_0)} = \sqrt{(-2)(9.8\text{ m/s}^2)(0\text{ m} - 10\text{ m})}$$
$$= 14\text{ m/s}$$

This is the magnitude of the velocity, hence the absolute value sign; the actual value is $v = -14$ m/s, with the minus sign indicating downward motion. Knowing the initial and final velocities, we use Equation 2.7 to find how long the dive takes. Solving that equation for t gives

$$t = \frac{v_0 - v}{g} = \frac{0\text{ m/s} - (-14\text{ m/s})}{9.8\text{ m/s}^2} = 1.4\text{ s}$$

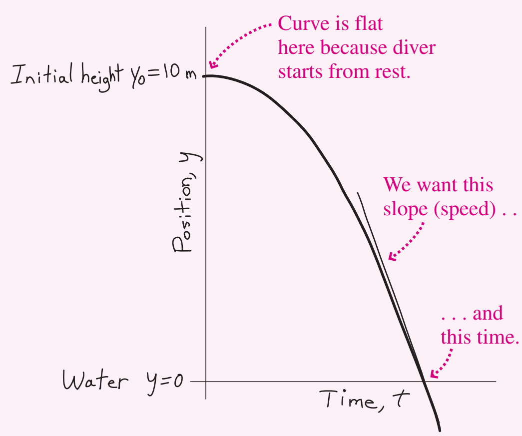


FIGURE 2.12 Our sketch for Example 2.5.

Note the careful attention to signs here; we wrote v with its negative sign and used $a = -g$ in Equation 2.7 because we defined downward to be the negative direction in our coordinate system.

ASSESS Make sense? Our expression for v gives a higher speed with a greater acceleration or a greater distance $y - y_0$—both as expected. Our approach here isn't the only one possible; we could also have found the time by solving Equation 2.10 and then evaluating the speed using Equation 2.7. ■

In Example 2.5 the diver was moving downward, and the downward gravitational acceleration steadily increased his speed. But, as Conceptual Example 2.1 suggested, the acceleration of gravity is downward regardless of an object's motion. Throw a ball straight up, and it's accelerating *downward* even while moving *upward*. Since velocity and acceleration are in opposite directions, the ball slows until it reaches its peak, then pauses instantaneously, and then gains speed as it falls. All the while its acceleration is 9.8 m/s^2 downward.

EXAMPLE 2.6 Constant Acceleration Due to Gravity: Tossing a Ball

You toss a ball straight up at 7.3 m/s; it leaves your hand at 1.5 m above the floor. Find when it hits the floor, the maximum height it reaches, and its speed when it passes your hand on the way down.

INTERPRET We have constant acceleration due to gravity, and here the object of interest is the ball. We want to find time, height, and speed.

DEVELOP The ball starts by going up, eventually comes to a stop, and then heads downward. Figure 2.13 is a sketch of the height versus time that we expect, showing what we know and the three quantities we're after. Equation 2.10, $x = x_0 + v_0 t + \frac{1}{2}at^2$, determines position (here, height y) as a function of time, so our plan is to use that equation to find the time the ball hits the floor. Then we can use Equation 2.11, $v^2 = v_0^2 + 2a(x - x_0)$, to find the height at which $v = 0$—that is, the peak height. Finally, Equation 2.11 will also give us the speed at any height, letting us answer the question about the speed when the ball passes the height of 1.5 m on its way down.

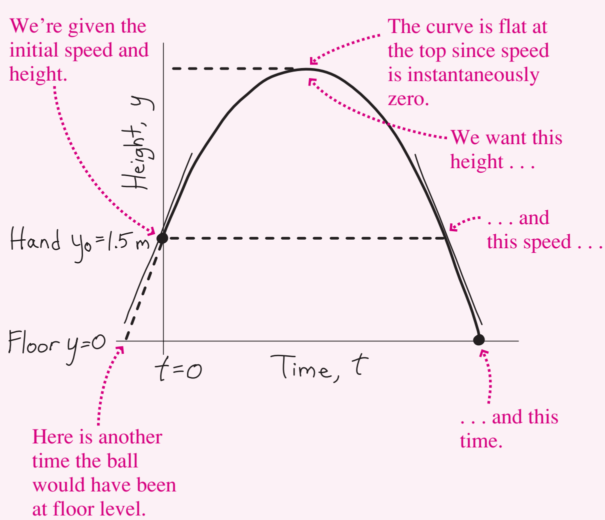


FIGURE 2.13 Our sketch for Example 2.6.

EVALUATE Our sketch shows that we've taken $y = 0$ at the floor, so when the ball is at the floor, Equation 2.10 becomes $0 = y_0 + v_0 t - \frac{1}{2}gt^2$, which we can solve for t using the quadratic formula (Appendix A; $t = \left(v_0 \pm \sqrt{v_0^2 + 2y_0 g}\right)/g$). Here v_0 is the initial velocity, 7.3 m/s; it's positive because the motion is initially upward. The initial position is the hand height so $y_0 = 1.5$ m, and g of course is 9.8 m/s^2 (we accounted for the downward acceleration by putting $a = -g$ in Equation 2.10). Putting in these numbers gives $t = 1.7$ s or -0.18 s; the answer we want is 1.7 s. At the peak of its flight, the ball's velocity is instantaneously zero because it's moving neither up nor down. So we set $v^2 = 0$ in Equation 2.11 to get $0 = v_0^2 - 2g(y - y_0)$. Solving for y then gives the peak height:

$$y = y_0 + \frac{v_0^2}{2g} = 1.5\text{ m} + \frac{(7.3\text{ m/s})^2}{(2)(9.8\text{ m/s}^2)} = 4.2\text{ m}$$

To find the speed when the ball reaches 1.5 m on the way down, we set $y = y_0$ in Equation 2.11. The result is $v^2 = v_0^2$, so $v = \pm v_0$ or ± 7.3 m/s. Once again, there are two answers. The equation has given us *all* the velocities the ball has at 1.5 m—including the initial upward velocity and the later downward velocity. We've shown here that an upward-thrown object returns to its initial height with the same speed it had initially.

ASSESS Make sense? With no air resistance to sap the ball of its energy, it seems reasonable that the ball comes back down with the same speed—a fact we'll explore further when we introduce energy conservation in Chapter 7. But why are there two answers for time and velocity? Equation 2.10 doesn't "know" about your hand or the floor; it "assumes" the ball has always been undergoing downward acceleration g. We asked of Equation 2.10 when the ball would be at $y = 0$. The second answer, 1.7 s, was the one we wanted. But if the ball had always been in free fall, it would also have been on the floor 0.18 s earlier, heading upward. That's the meaning of the other answer, -0.18 s, as we've indicated on our sketch. Similarly, Equation 2.11 gave us all the velocities the ball had at a height of 1.5 m, including both the initial upward velocity and the later downward velocity. ■

✓TIP Multiple Answers

Frequently the mathematics of a problem gives more than one answer. Think about what each answer means before discarding it! Sometimes an answer isn't consistent with the physical assumptions of the problem, but other times all answers are meaningful even if they aren't all what you're looking for.

GOT IT? 2.4 Standing on a roof, you simultaneously throw one ball straight up and drop another from rest. Which hits the ground first? Which hits the ground moving faster?

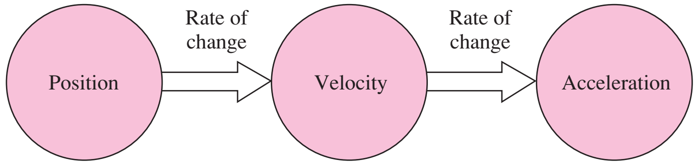
APPLICATION Keeping Time

The NIST-F1 atomic clock, shown here with its developers, sets the U.S. standard of time. The clock gets its remarkable accuracy by monitoring a supercold clump of freely falling cesium atoms for what is, in this context, a long time period of about 1 second. The atom clump is put in free fall by a more sophisticated version of the ball toss in Example 2.6. In the NIST-F1 clock, laser beams gently "toss" the ball of atoms upward with a speed that gives it an up-and-down travel time of about 1 second (see Problem 64). For this reason NIST-F1 is called an **atomic fountain clock**. In the photo you can see the clock's towerlike structure that accommodates this atomic fountain.

CHAPTER 2 SUMMARY

Big Picture

The big ideas here are those of **kinematics**—the study of motion without regard to its cause. **Position, velocity,** and **acceleration** are the quantities that characterize motion:



Key Concepts and Equations

Average velocity and acceleration involve changes in position and velocity, respectively, occurring over a time interval Δt:

$$\bar{v} = \frac{\Delta x}{\Delta t}$$

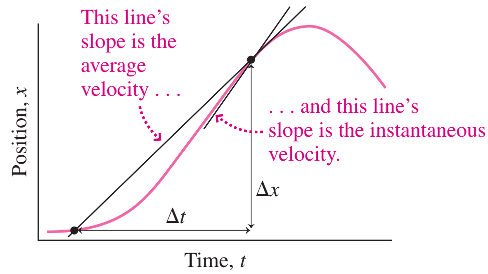
$$\bar{a} = \frac{\Delta v}{\Delta t}$$

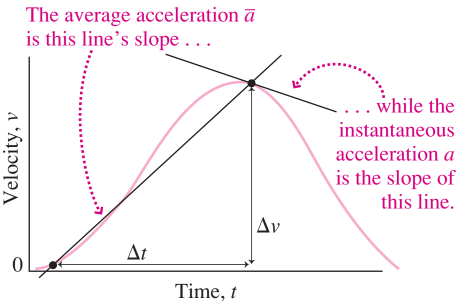
Here Δx is the **displacement**, or change in position, and Δv is the change in velocity.

Instantaneous values are the limits of infinitesimally small time intervals and are given by calculus as the time derivatives of position and velocity:

$$v = \frac{dx}{dt}$$

$$a = \frac{dv}{dt}$$





Applications

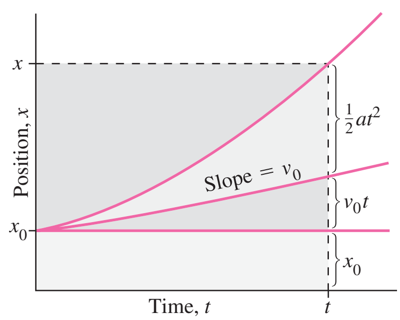
Constant acceleration is a special case that yields simple equations describing one-dimensional motion:

$$v = v_0 + at$$

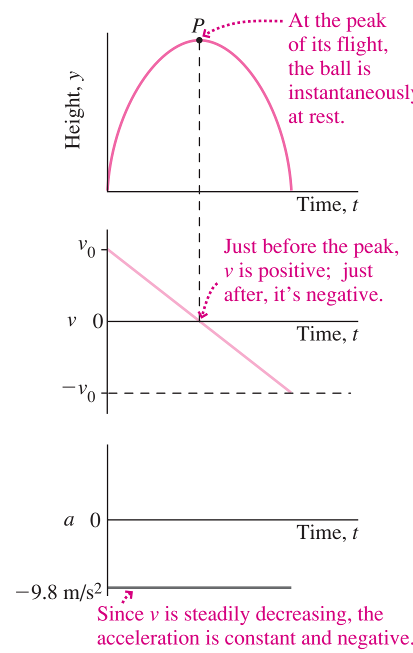
$$x = x_0 + v_0 t + \tfrac{1}{2}at^2$$

$$v^2 = v_0^2 + 2a(x - x_0)$$

These equations apply only in the case of constant acceleration.



An important example is the acceleration of gravity, essentially constant near Earth's surface, with magnitude approximately $9.8\ \mathrm{m/s^2}$.



For Thought and Discussion

1. Under what conditions are average and instantaneous velocity equal?
2. Does a speedometer measure speed or velocity?
3. You check your odometer at the beginning of a day's driving and again at the end. Under what conditions would the difference between the two readings represent your displacement?
4. Consider two possible definitions of average speed: (a) the average of the values of the instantaneous speed over a time interval and (b) the magnitude of the average velocity. Are these definitions equivalent? Give two examples to demonstrate your conclusion.
5. Is it possible to be at position $x = 0$ and still be moving?
6. Is it possible to have zero velocity and still be accelerating?
7. If you know the initial velocity v_0 and the initial and final heights y_0 and y, you can use Equation 2.10 to solve for the time t when the object will be at height y. But the equation is quadratic in t, so you'll get two answers. Physically, why is this?
8. Starting from rest, an object undergoes acceleration given by $a = bt$, where t is time and b is a constant. Can you use bt for a in Equation 2.10 to predict the object's position as a function of time? Why or why not?
9. In which of the velocity-versus-time graphs shown in Fig. 2.14 would the average velocity over the interval shown equal the average of the velocities at the ends of the interval?

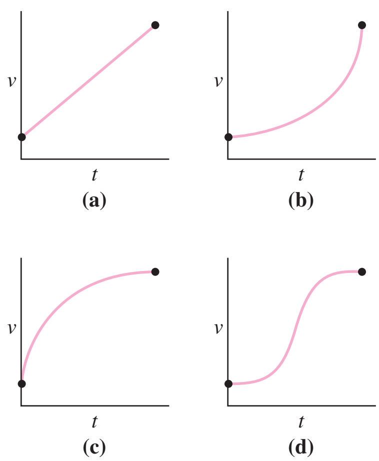


FIGURE 2.14 For Thought and Discussion 9

10. If you travel in a straight line at 50 km/h for 1 h and at 100 km/h for another hour, is your average velocity 75 km/h? If not, is it more or less?
11. If you travel in a straight line at 50 km/h for 50 km and then at 100 km/h for another 50 km, is your average velocity 75 km/h? If not, is it more or less?

Exercises and Problems

Exercises

Section 2.1 Average Motion

12. In 2009, Usain Bolt of Jamaica set a world record in the 100-m dash with a time of 9.58 s. What was his average speed?
13. The standard 26-mile, 385-yard marathon dates to 1908, when the Olympic marathon started at Windsor Castle and finished before the Royal Box at London's Olympic Stadium. Today's top marathoners achieve times around 2 hours, 4 minutes for the standard marathon. (a) What's the average speed of a marathon run in this time? (b) Marathons before 1908 were typically about 25 miles. How much longer does the race last today as a result of the extra mile and 385 yards, assuming it's run at the average speed?
14. Starting from home, you bicycle 24 km north in 2.5 h and then turn around and pedal straight home in 1.5 h. What are your (a) displacement at the end of the first 2.5 h, (b) average velocity over the first 2.5 h, (c) average velocity for the homeward leg of the trip, (d) displacement for the entire trip, and (e) average velocity for the entire trip?
15. The Voyager 1 spacecraft is expected to continue broadcasting data until at least 2020, when it will be some 14 billion miles from Earth. How long will it take Voyager's radio signals, traveling at the speed of light, to reach Earth from this distance?
16. In 2008, Australian Emma Snowsill set an unofficial record in the women's Olympic triathlon, completing the 1.5-km swim, 40-km bicycle ride, and 10-km run in 1 h, 58 min, 27.66 s. What was her average speed?
17. Taking Earth's orbit to be a circle of radius 1.5×10^8 km, determine Earth's orbital speed in (a) meters per second and (b) miles per second.
18. What's the conversion factor from meters per second to miles per hour?

Section 2.2 Instantaneous Velocity

19. On a single graph, plot distance versus time for the first two trips from Houston to Des Moines described on page 14. For each trip, identify graphically the average velocity and, for each segment of the trip, the instantaneous velocity.
20. For the motion plotted in Fig. 2.15, estimate (a) the greatest velocity in the positive x-direction, (b) the greatest velocity in the negative x-direction, (c) any times when the object is instantaneously at rest, and (d) the average velocity over the interval shown.

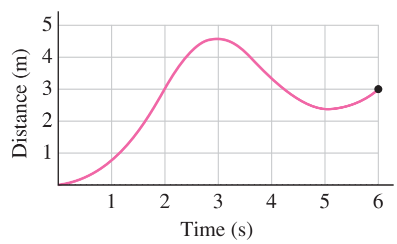


FIGURE 2.15 Exercise 20

21. A model rocket is launched straight upward. Its altitude y as a function of time is given by $y = bt - ct^2$, where $b = 82$ m/s, $c = 4.9$ m/s^2, t is the time in seconds, and y is in meters. (a) Use differentiation to find a general expression for the rocket's velocity as a function of time. (b) When is the velocity zero?

Section 2.3 Acceleration

22. A giant eruption on the Sun propels solar material from rest to 450 km/s over a period of 1 h. Find the average acceleration.
23. Starting from rest, a subway train first accelerates to 25 m/s, then brakes. Forty-eight seconds after starting, it's moving at 17 m/s. What's its average acceleration in this 48-s interval?
24. A space shuttle's main engines cut off 8.5 min after launch, at which time its speed is 7.6 km/s. What's the shuttle's average acceleration during this interval?

25. An egg drops from a second-story window, taking 1.12 s to fall and reaching 11.0 m/s just before hitting the ground. On contact, the egg stops completely in 0.131 s. Calculate the average magnitudes of its acceleration while falling and while stopping.
26. An airplane's takeoff speed is 320 km/h. If its average acceleration is 2.9 m/s^2, how much time is it accelerating down the runway before it lifts off?
27. ThrustSSC, the world's first supersonic car, accelerates from rest to 1000 km/h in 16 s. What's its acceleration?

Section 2.4 Constant Acceleration

28. You're driving at 70 km/h when you apply constant acceleration to pass another car. Six seconds later, you're doing 80 km/h. How far did you go in this time?
29. Differentiate both sides of Equation 2.10, and show that you get Equation 2.7.
30. An X-ray tube gives electrons constant acceleration over a distance of 15 cm. If their final speed is 1.2×10^7 m/s, what are (a) the electrons' acceleration and (b) the time they spend accelerating?
31. A rocket rises with constant acceleration to an altitude of 85 km, at which point its speed is 2.8 km/s. (a) What's its acceleration? (b) How long does the ascent take?
32. Starting from rest, a car accelerates at a constant rate, reaching 88 km/h in 12 s. Find (a) its acceleration and (b) how far it goes in this time.
33. A car moving initially at 50 mi/h begins slowing at a constant rate 100 ft short of a stoplight. If the car comes to a full stop just at the light, what is the magnitude of its acceleration?
34. BIO In a medical X-ray tube, electrons are accelerated to a velocity of 10^8 m/s and then slammed into a tungsten target. As they stop, the electrons' rapid acceleration produces X rays. If the time for an electron to stop is on the order of 10^{-9} s, approximately how far does it move while stopping?
35. The Barringer meteor crater in Arizona is 180 m deep and 1.2 km in diameter. Fragments of the meteor lie just below the bottom of the crater. If these fragments negatively accelerated at a constant rate of 4×10^5 m/s^2 as they plowed through Earth, what was the meteor's speed at impact?
36. You're driving at speed v_0 when you spot a stationary moose on the road, a distance d ahead. Find an expression for the magnitude of the acceleration you need if you're to stop before hitting the moose.

Section 2.5 The Acceleration of Gravity

37. You drop a rock into a deep well and 4.4 s later hear a splash. How far down is the water? Neglect the travel time of sound.
38. Your friend is sitting 6.5 m above you on a tree branch. How fast should you throw an apple so it just reaches her?
39. A model rocket leaves the ground, heading straight up at 49 m/s. (a) What's its maximum altitude? Find its speed and altitude at (b) 1 s, (c) 4 s, and (d) 7 s.
40. A foul ball leaves the bat going straight up at 23 m/s. (a) How high does it rise? (b) How long is it in the air? Neglect the distance between bat and ground.
41. A Frisbee is lodged in a tree 6.5 m above the ground. A rock thrown from below must be going at least 3 m/s to dislodge the Frisbee. How fast must such a rock be thrown upward if it leaves the thrower's hand 1.3 m above the ground?
42. Space pirates kidnap an earthling and hold him on one of the solar system's planets. With nothing else to do, the prisoner amuses himself by dropping his watch from eye level (170 cm) to the floor. He observes that the watch takes 0.95 s to fall. On what planet is he being held? (*Hint*: Consult Appendix E.)

Problems

43. You allow 40 min to drive 25 mi to the airport, but you're caught in heavy traffic and average only 20 mi/h for the first 15 min. What must your average speed be on the rest of the trip if you're to make your flight?
44. A base runner can get from first to second base in 3.4 s. If he leaves first as the pitcher throws a 90 mi/h fastball the 61-ft distance to the catcher, and if the catcher takes 0.45 s to catch and rethrow the ball, how fast does the catcher have to throw the ball to second base to make an out? Home plate to second base is the diagonal of a square 90 ft on a side.
45. You drive 4600 km from coast to coast of the United States at 65 mi/h (105 km/h), stopping an average of 30 min for rest after every 2 h of driving. (a) What's your average velocity for the entire trip? (b) How long does the trip take?
46. You can run 9.0 m/s, 20% faster than your brother. How much head start should you give him in order to have a tie race over 100 m?
47. A jetliner leaves San Francisco for New York, 4600 km away. With a strong tailwind, its speed is 1100 km/h. At the same time, a second jet leaves New York for San Francisco. Flying into the wind, it makes only 700 km/h. When and where do the two planes pass?
48. An object's position is given by $x = bt + ct^3$, where $b = 1.50$ m/s, $c = 0.640$ m/s^3, and t is time in seconds. To study the limiting process leading to the instantaneous velocity, calculate the object's average velocity over time intervals from (a) 1.00 s to 3.00 s, (b) 1.50 s to 2.50 s, and (c) 1.95 s to 2.05 s. (d) Find the instantaneous velocity as a function of time by differentiating, and compare its value at 2 s with your average velocities.
49. An object's position as a function of time t is given by $x = bt^4$, with b a constant. Find an expression for the instantaneous velocity, and show that the average velocity over the interval from $t = 0$ to any time t is one-fourth of the instantaneous velocity at t.
50. In a drag race, the position of a car as a function of time is given by $x = bt^2$, with $b = 2.000$ m/s^2. In an attempt to determine the car's velocity midway down a 400-m track, two observers stand at the 180-m and 220-m marks and note when the car passes. (a) What value do the two observers compute for the car's velocity over this 40-m stretch? Give your answer to four significant figures. (b) By what percentage does this observed value differ from the instantaneous value at $x = 200$ m?
51. An object's position is given by $x = bt^3$, with x in meters, t in seconds, and $b = 1.5$ m/s^3. Determine (a) the instantaneous velocity and (b) the instantaneous acceleration at the end of 2.5 s. Find (c) the average velocity and (d) the average acceleration during the first 2.5 s.
52. Squaring Equation 2.7 gives an expression for v^2. Equation 2.11 also gives an expression for v^2. Equate the two expressions, and show that the resulting equation reduces to Equation 2.10.
53. On packed snow, computerized antilock brakes can reduce a car's stopping distance by 55%. By what percentage is the stopping time reduced?
54. A particle leaves its initial position x_0 at time $t = 0$, moving in the positive x-direction with speed v_0 but undergoing acceleration of magnitude a in the negative x-direction. Find expressions for (a) the time when it returns to x_0 and (b) its speed when it passes that point.
55. A hockey puck moving at 32 m/s slams through a wall of snow 35 cm thick. It emerges moving at 18 m/s. Assuming constant acceleration, find (a) the time the puck spends in the snow and (b) the thickness of a snow wall that would stop the puck entirely.

56. Amtrak's 20th-Century Limited is en route from Chicago to New York at 110 km/h when the engineer spots a cow on the track. The train brakes to a halt in 1.2 min, stopping just in front of the cow. (a) What is the magnitude of the train's acceleration? (b) What's the direction of the acceleration? (c) How far was the train from the cow when the engineer applied the brakes?
57. A jetliner touches down at 220 km/h and comes to a halt 29 s later. What's the shortest runway on which this aircraft can land?
58. A motorist suddenly notices a stalled car and slams on the brakes, negatively accelerating at $6.3\ \text{m/s}^2$. Unfortunately, this isn't enough, and a collision ensues. From the damage sustained, police estimate that the car was going 18 km/h at the time of the collision. They also measure skid marks 34 m long. (a) How fast was the motorist going when the brakes were first applied? (b) How much time elapsed from the initial braking to the collision?
59. A racing car undergoing constant acceleration covers 140 m in 3.6 s. (a) If it's moving at 53 m/s at the end of this interval, what was its speed at the beginning of the interval? (b) How far did it travel from rest to the end of the 140-m distance?
60. The maximum braking acceleration of a car on a dry road is about $8\ \text{m/s}^2$. If two cars move head-on toward each other at 88 km/h (55 mi/h), and their drivers brake when they're 85 m apart, will they collide? If so, at what relative speed? If not, how far apart will they be when they stop? Plot distance versus time for both cars on a single graph.
61. After 35 min of running, at the 9-km point in a 10-km race, you find yourself 100 m behind the leader and moving at the same speed. What should your acceleration be if you're to catch up by the finish line? Assume that the leader maintains constant speed.
62. You're speeding at 85 km/h when you notice that you're only 10 m behind the car in front of you, which is moving at the legal speed limit of 60 km/h. You slam on your brakes, and your car negatively accelerates at $4.2\ \text{m/s}^2$. Assuming the other car continues at constant speed, will you collide? If so, at what relative speed? If not, what will be the distance between the cars at their closest approach?
63. Airbags cushioned the Mars rover Spirit's landing, and the rover bounced some 15 m vertically after its first impact. Assuming no loss of speed at contact with the Martian surface, what was Spirit's impact speed?
64. Calculate the speed with which cesium atoms must be "tossed" in the NIST-F1 atomic clock so that their up-and-down travel time is 1.0 s. (See the Application on page 24.)
65. A falling object travels one-fourth of its total distance in the last second of its fall. From what height was it dropped?
66. You're on a NASA team engineering a probe to land on Jupiter's moon Io, and your job is to specify the impact speed the probe can tolerate without damage. Rockets will bring the probe to a halt 100 m above the surface, after which it will fall freely. What speed do you specify? (Consult Appendix E.)
67. You're atop a building of height h, and a friend is poised to drop a ball from a window at $h/2$. Find an expression for the speed at which you should simultaneously throw a ball downward, so the two hit the ground at the same time.
68. A castle's defenders throw rocks down on their attackers from a 15-m-high wall, with initial speed 10 m/s. How much faster are the rocks moving when they hit the ground than if they were simply dropped?
69. Two divers jump from a 3.00-m platform. One jumps upward at 1.80 m/s, and the second steps off the platform as the first passes it on the way down. (a) What are their speeds as they hit the water? (b) Which hits the water first and by how much?
70. A balloon is rising at 10 m/s when its passenger throws a ball straight up at 12 m/s relative to the balloon. How much later does the passenger catch the ball?
71. Landing on the Moon, a spacecraft fires its rockets and comes to a complete stop just 12 m above the lunar surface. It then drops freely to the surface. How long does it take to fall, and what's its impact speed? (*Hint*: Consult Appendix E.)
72. You're at mission control for a rocket launch, deciding whether to let the launch proceed. A band of clouds 5.3 km thick extends upward from 1.9 km altitude. The rocket will accelerate at $4.6\ \text{m/s}^2$, and it isn't allowed to be out of sight for more than 30 s. Should you allow the launch?
73. You're an investigator for the National Transportation Safety Board, examining a subway accident in which a train going at 80 km/h collided with a slower train traveling in the same direction at 25 km/h. Your job is to determine the relative speed of the collision, to help establish new crash standards. The faster train's "black box" shows that it began negatively accelerating at $2.1\ \text{m/s}^2$ when it was 50 m from the slower train, while the slower train continued at constant speed. What do you report?
74. You toss a book into your dorm room, just clearing a windowsill 4.2 m above the ground. (a) If the book leaves your hand 1.5 m above the ground, how fast must it be going to clear the sill? (b) How long after it leaves your hand will it hit the floor, 0.87 m below the windowsill?
75. Consider an object traversing a distance L, part of the way at speed v_1 and the rest of the way at speed v_2. Find expressions for the average speeds when the object moves at each of the two speeds (a) for half the total *time* and (b) for half the *distance*.
76. A particle's position as a function of time is given by $x = x_0 \sin \omega t$, where x_0 and ω are constants. (a) Find expressions for the velocity and acceleration. (b) What are the maximum values of velocity and acceleration? (*Hint*: Consult the table of derivatives in Appendix A.)
77. Ice skaters, ballet dancers, and basketball players executing vertical leaps often give the illusion of "hanging" almost motionless near the top of the leap. To see why this is, consider a leap to maximum height h. Of the total time spent in the air, what fraction is spent in the upper half (i.e., at $y > \frac{1}{2}h$)?
78. You're staring idly out your dorm window when you see a water balloon fall past. If the balloon takes 0.22 s to cross the 1.3-m-high window, from what height above the window was it dropped?
79. A police radar's effective range is 1.0 km, and your radar detector's range is 1.9 km. You're going 110 km/h in a 70 km/h zone when the radar detector beeps. At what rate must you negatively accelerate to avoid a speeding ticket?
80. An object starts moving in a straight line from position x_0, at time $t = 0$, with velocity v_0. Its acceleration is given by $a = a_0 + bt$, where a_0 and b are constants. Find expressions for (a) the instantaneous velocity and (b) the position, as functions of time.
81. You're a consultant on a movie set, and the producer wants a car to drop so that it crosses the camera's field of view in time Δt. The field of view has height h. Derive an expression for the height above the top of the field of view from which the car should be released.
82. (a) For the ball in Example 2.6, find its velocity just before it hits the floor. (b) Suppose you had tossed a second ball straight down at 7.3 m/s (from the same place 1.5 m above the floor). What would its velocity be just before it hits the floor? (c) When would the second ball hit the floor? (Interpret any multiple answers.)

83. Your roommate is an aspiring novelist and asks your opinion on a matter of physics. The novel's central character is kept awake at night by a leaky faucet. The sink is 19.6 cm below the faucet. At the instant one drop leaves the faucet, another strikes the sink below and two more are in between on the way down. How many drops per second are keeping the protagonist awake?
84. You and your roommate plot to drop water balloons on students entering your dorm. Your window is 20 m above the sidewalk. You plan to place an X on the sidewalk to mark the spot a student must be when you drop the balloon. You note that most students approach the dorm at about 2 m/s. How far from the impact point do you place the X?
85. Derive Equation 2.10 by integrating Equation 2.7 over time. You'll have to interpret the constant of integration.

Passage Problems

A wildlife biologist is studying the hunting patterns of tigers. She anesthetizes a tiger and attaches a GPS collar to track its movements. The collar transmits data on the tiger's position and velocity. Figure 2.16 shows the tiger's velocity as a function of time as it moves on a one-dimensional path.

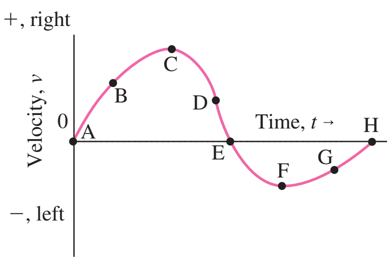


FIGURE 2.16 The tiger's velocity (Passage Problems 86–90)

86. At which marked point(s) is the tiger not moving?
    a. E only
    b. A, E, and H
    c. C and F
    d. none of the points (it's always moving)
87. At which marked point(s) is the tiger not accelerating?
    a. E only
    b. A, E, and H
    c. C and F
    d. all of the points (it's never accelerating)
88. At which point does the tiger have the greatest speed?
    a. B
    b. C
    c. D
    d. F
89. At which point does the tiger's acceleration have the greatest magnitude?
    a. B
    b. C
    c. D
    d. F
90. At which point is the tiger farthest from its starting position at $t = 0$?
    a. C
    b. E
    c. F
    d. H

Answers to Chapter Questions

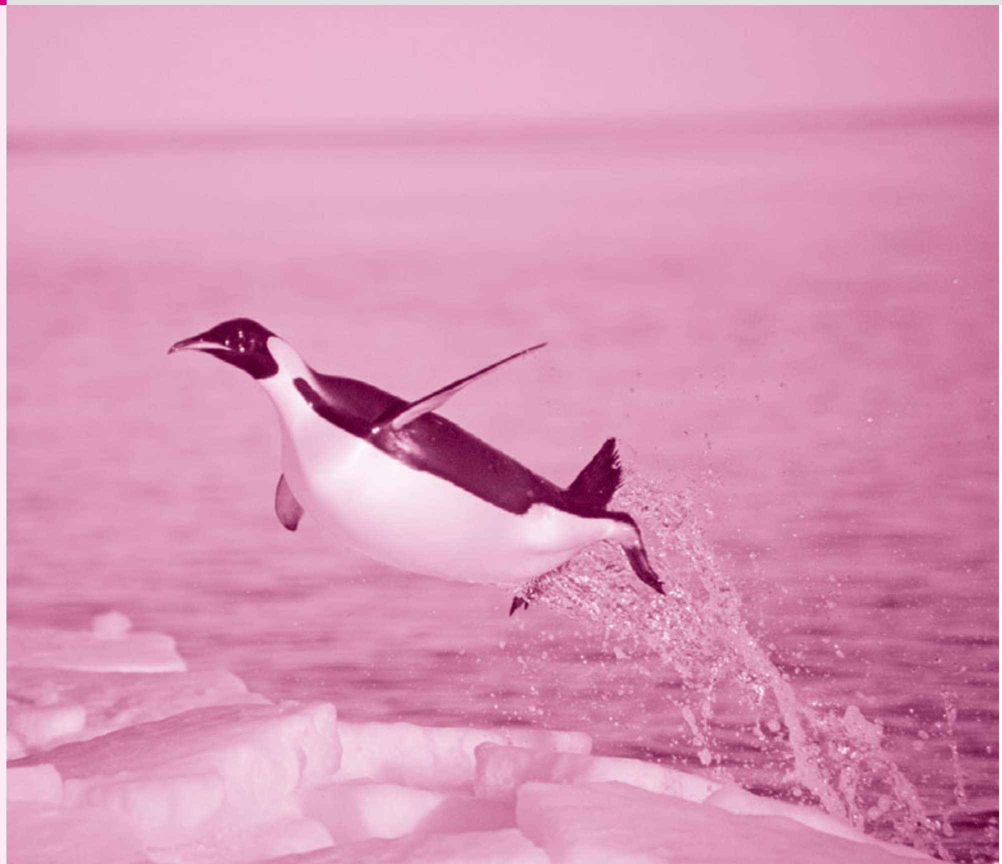
Answer to Chapter Opening Question

Although the ball's velocity is zero at the top of its motion, its acceleration is $-9.8\ \text{m/s}^2$, as it is throughout the toss.

Answers to GOT IT? Questions

2.1. (a) and (b); average speed is greater for (c).
2.2. (b) moves with constant speed; (a) reverses; (d) speeds up.
2.3. (a) halfway between the times. Because its acceleration is constant, the police car's speed increases by equal amounts in equal times. So it gets from 0 to half its final velocity—which is twice the car's velocity—in half the total time.
2.4. The dropped ball hits first; the thrown ball hits moving faster.

3 Motion in Two and Three Dimensions



At what angle should this penguin leave the water to maximize the range of its jump?

New Concepts, New Skills

By the end of this chapter you should be able to

- Use vectors to describe position, velocity, and acceleration in three dimensions (3.1, 3.2).
- Add and subtract vectors, and multiply them by scalars (3.1).
- Explain how the effects of acceleration depend on the direction of acceleration relative to velocity (3.2).
- Transform velocities to different reference frames (3.3).
- Solve quantitative problems involving motion in two dimensions with constant acceleration, including projectile motion with the acceleration of gravity (3.4, 3.5).
- Explain why circular motion necessarily entails acceleration, and solve quantitative problems involving uniform circular motion (3.6).

Connecting Your Knowledge

- You should understand the concepts of position, velocity, and acceleration in one dimension (2.1–2.3).
- You should know how to solve problems in one-dimensional motion with constant acceleration (2.4).
- You should be familiar with the acceleration of gravity near Earth's surface, and be able to apply it to one-dimensional motion under the influence of gravity (2.5).

What's the speed of an orbiting satellite? How should I leap to win the long-jump competition? How do I engineer a curve in the road for safe driving? These and many other questions involve motion in more than one dimension. In this chapter we extend the ideas of one-dimensional motion to these more complex—and more interesting—situations.

3.1 Vectors

We've seen that quantities describing motion have direction as well as magnitude. In Chapter 2, a simple plus or minus sign took care of direction. But now, in two or three dimensions, we need a way to account for all possible directions. We do this with mathematical quantities called **vectors**, which express both magnitude and direction. Vectors stand in contrast to **scalars**, which are quantities that have no direction.

Position and Displacement

The simplest vector quantity is position. Given an origin, we can characterize any position in space by drawing an arrow from the origin to that position. That arrow is a pictorial representation of a **position vector**, which we call $\vec{r}$. The arrow over the r indicates that this is a vector quantity, and it's crucial to include the arrow whenever you're dealing with vectors. Figure 3.1 shows a position vector in a two-dimensional coordinate system; this vector describes a point a distance of 2 m from the origin, in a direction 30° from the horizontal axis.

Suppose you walk from the origin straight to the point described by the vector $\vec{r}_1$ in Fig. 3.1, and then you turn right and walk another 1 m. Figure 3.2 shows how you can tell where you end up. Draw a second vector whose length represents 1 m and that points to the right; we'll call this vector $\Delta\vec{r}$ because it's a **displacement vector**, representing a *change* in position. Put the tail of $\Delta\vec{r}$ at the head of the vector $\vec{r}_1$; then the head of $\Delta\vec{r}$ shows your ending position. The result is the same as if you had walked straight from the origin to this position. So the new position is described by a third vector $\vec{r}_2$, as indicated in Fig. 3.2. What we've just described is **vector addition**. To add two vectors, put the second vector's tail at the head of the first; the sum is then the vector that extends from the tail of the first vector to the head of the second, as does $\vec{r}_2$ in Fig. 3.2.

A vector has both magnitude and direction—but because that's all the information it contains, it doesn't matter where it starts. So you're free to move a vector around to form vector sums. Figure 3.3 shows some examples of vector addition and also shows that vector addition obeys simple rules you know for regular arithmetic.

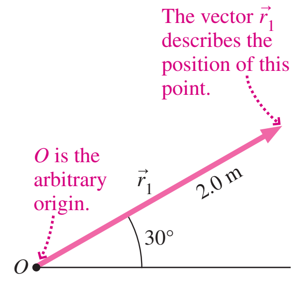


FIGURE 3.1 A position vector $\vec{r}_1$.

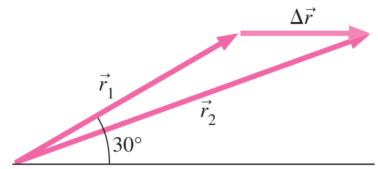


FIGURE 3.2 Vectors $\vec{r}_1$ and $\Delta\vec{r}$ sum to $\vec{r}_2$.

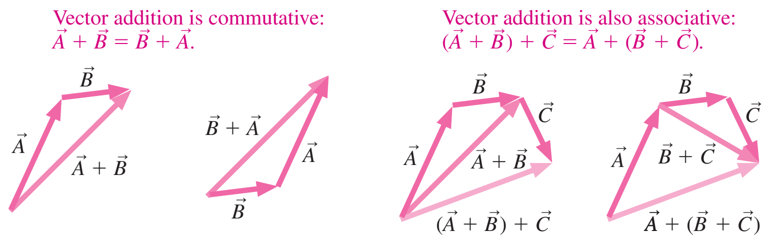


FIGURE 3.3 Vector addition is commutative and associative.

Multiplication

You and I jog in the same direction, but you go twice as far. Your displacement vector, $\vec{B}$, is twice as long as my displacement vector, $\vec{A}$; mathematically, $\vec{B} = 2\vec{A}$. That's what it means to multiply a vector by a scalar; simply rescale the magnitude of the vector by that scalar. If the scalar is negative, then the vector direction reverses—and that provides a way to subtract vectors. In Fig. 3.2, for example, you can see that $\vec{r}_1 = \vec{r}_2 + (-1)\Delta\vec{r}$, or simply $\vec{r}_1 = \vec{r}_2 - \Delta\vec{r}$. Later, we'll see ways to multiply two vectors, but for now the only multiplication we consider is a vector multiplied by a scalar.

Vector Components

You can always add vectors graphically, as shown in Fig. 3.2, or you can use geometric relationships like the laws of sines and cosines to accomplish the same thing algebraically. In both these approaches, you specify a vector by giving its magnitude and direction. But often it's more convenient instead to describe vectors using their **components** in a given coordinate system.

A **coordinate system** is a framework for describing positions in space. It's a mathematical construct, and you're free to choose whatever coordinate system you want. You've already seen **Cartesian** or **rectangular coordinate systems**, in which a pair of numbers (x, y) represents each point in a plane. You could also think of each point as representing the head of a position vector, in which case the numbers x and y are the vector components. The components tell how much of the vector is in the x-direction and how much is in the y-direction. Not all vectors represent actual positions in space; for example, there are velocity, acceleration, and force vectors. The lengths of these vectors represent the magnitudes of the corresponding physical quantities. For an arbitrary vector quantity $\vec{A}$, we designate the components A_x and A_y (Fig. 3.4). Note that the components themselves aren't vectors but scalars.

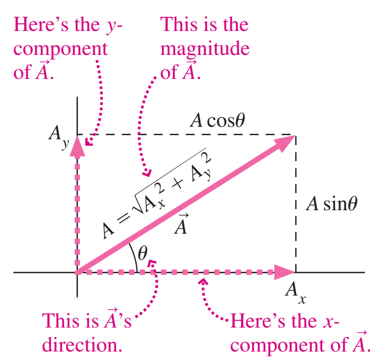


FIGURE 3.4 Magnitude/direction and component representations of vector $\vec{A}$.

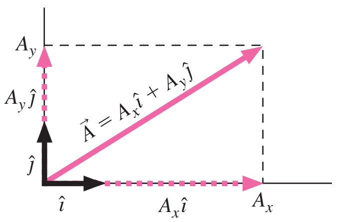
In two dimensions it takes two quantities to specify a vector—either its magnitude and direction or its components. They're related by the Pythagorean theorem and the definitions of the trig functions, as shown in Fig. 3.4:

$$A = \sqrt{A_x^2 + A_y^2} \quad \text{and} \quad \tan\theta = \frac{A_y}{A_x} \qquad (\text{vector magnitude and direction}) \tag{3.1}$$

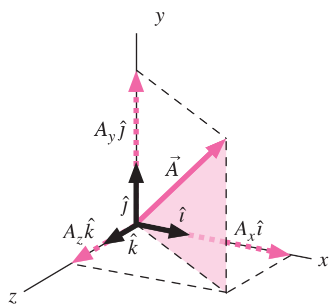
Without the arrow above it, a vector's symbol stands for the vector's magnitude. Going the other way, we have

$$A_x = A\cos\theta \quad \text{and} \quad A_y = A\sin\theta \qquad (\text{vector components}) \tag{3.2}$$

If a vector $\vec{A}$ has zero magnitude, we write $\vec{A} = \vec{0}$, where the vector arrow on the zero indicates that both components must be zero.



(a)



(b)

FIGURE 3.5 Vectors in (a) a plane and (b) space, expressed using unit vectors.

Unit Vectors

It's cumbersome to say "a vector of magnitude 2 m at 30° to the x-axis" or, equivalently, "a vector whose x- and y-components are 1.73 m and 1.0 m, respectively." We can express this more succinctly using the **unit vectors** $\hat{\imath}$ and $\hat{\jmath}$ (read as "i hat"). These unit vectors have magnitude 1, no units, and point along the x- and y-axes, respectively. In three dimensions we add a third unit vector, $\hat{k}$, along the z-axis. Any vector in the x-direction can be written as some number—perhaps with units, such as meters or meters per second—times the unit vector $\hat{\imath}$, and analogously in the y-direction using $\hat{\jmath}$. That means any vector in a plane can be written as a sum involving the two unit vectors: $\vec{A} = A_x\hat{\imath} + A_y\hat{\jmath}$ (Fig. 3.5*a*). Similarly, any vector in space can be written with the three unit vectors (Fig. 3.5*b*).

The unit vectors convey only direction; the numbers that multiply them give size and units. Together they provide compact representations of vectors, including units. The displacement vector $\vec{r}_1$ in Fig. 3.1, for example, is $\vec{r}_1 = 1.7\hat{\imath} + 1.0\hat{\jmath}$ m.

EXAMPLE 3.1 Unit Vectors: Taking a Drive

You drive to a city 160 km from home, going 35° north of east. Express your new position in unit vector notation, using an east-west/north-south coordinate system.

INTERPRET We interpret this as a problem about writing a vector in unit vector notation, given its magnitude and direction.

DEVELOP Unit vector notation multiplies a vector's x- and y-components by the unit vectors $\hat{\imath}$ and $\hat{\jmath}$ and sums the results, so we draw a sketch showing those components (Fig. 3.6). Our plan is to solve for the two components, multiply by the unit vectors, and then add. Equations 3.2 determine the components.

EVALUATE We have $x = r\cos\theta = (160\text{ km})(\cos 35°) = 131$ km and $y = r\sin\theta = (160\text{ km})(\sin 35°) = 92$ km. Then the position of the city is

$$\vec{r} = 131\hat{\imath} + 92\hat{\jmath}\ \text{km}$$

ASSESS Make sense? Figure 3.6 suggests that the x-component should be longer than the y component, as our answer indicates. Our sketch shows the component values and the final answer. Note that we treat $131\hat{\imath} + 92\hat{\jmath}$ as a single vector quantity, labeling it at the end with the appropriate unit, km. ■

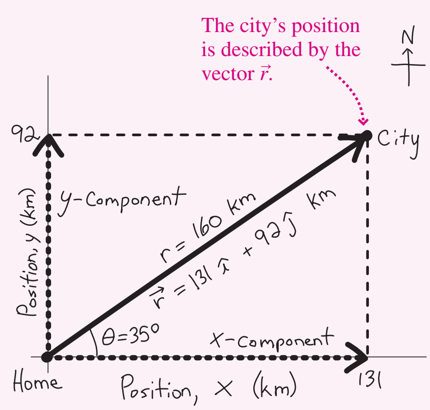


FIGURE 3.6 Our sketch for Example 3.1.

Vector Arithmetic with Unit Vectors

Vector addition is simple with unit vectors: Just add the corresponding components. If $\vec{A} = A_x\hat{\imath} + A_y\hat{\jmath}$ and $\vec{B} = B_x\hat{\imath} + B_y\hat{\jmath}$, for example, then their sum is

$$\vec{A} + \vec{B} = (A_x\hat{\imath} + A_y\hat{\jmath}) + (B_x\hat{\imath} + B_y\hat{\jmath}) = (A_x + B_x)\hat{\imath} + (A_y + B_y)\hat{\jmath}$$

Subtraction and multiplication by a scalar are similarly straightforward.

GOT IT? 3.1 Which vector describes a displacement of 10 units in a direction 30° below the positive x-axis? (a) $10\hat{\imath} - 10\hat{\jmath}$, (b) $5.0\hat{\imath} - 8.6\hat{\jmath}$, (c) $8.6\hat{\imath} - 5.0\hat{\jmath}$, (d) $10(\hat{\imath} + \hat{\jmath})$

3.2 Velocity and Acceleration Vectors

We defined velocity in one dimension as the rate of change of position. In two or three dimensions it's the same thing, except now the change in position—displacement—is a vector. So we write

$$\overline{\vec{v}} = \frac{\Delta\vec{r}}{\Delta t} \quad \text{(average velocity vector)} \tag{3.3}$$

for the average velocity, in analogy with Equation 2.1. Here division by Δt simply means multiplying by $1/\Delta t$. As before, instantaneous velocity is given by a limiting process:

$$\vec{v} = \lim_{\Delta t \to 0}\frac{\Delta\vec{r}}{\Delta t} = \frac{d\vec{r}}{dt} \quad \text{(instantaneous velocity vector)} \tag{3.4}$$

Again, that derivative $d\vec{r}/dt$ is shorthand for the result of the limiting process, taking ever smaller time intervals Δt and the corresponding displacements $\Delta\vec{r}$. Another way to look at Equation 3.4 is in terms of components. If $\vec{r} = x\hat{\imath} + y\hat{\jmath}$, then we can write

$$\vec{v} = \frac{d\vec{r}}{dt} = \frac{dx}{dt}\hat{\imath} + \frac{dy}{dt}\hat{\jmath} = v_x\hat{\imath} + v_y\hat{\jmath}$$

where the velocity components v_x and v_y are the derivatives of the position components.

Acceleration is the rate of change of velocity, so we write

$$\overline{\vec{a}} = \frac{\Delta\vec{v}}{\Delta t} \quad \text{(average acceleration vector)} \tag{3.5}$$

for the average acceleration and

$$\vec{a} = \lim_{\Delta t \to 0}\frac{\Delta\vec{v}}{\Delta t} = \frac{d\vec{v}}{dt} \quad \text{(instantaneous acceleration vector)} \tag{3.6}$$

for the instantaneous acceleration. We can also express instantaneous acceleration in components, as we did for velocity:

$$\vec{a} = \frac{d\vec{v}}{dt} = \frac{dv_x}{dt}\hat{\imath} + \frac{dv_y}{dt}\hat{\jmath} = a_x\hat{\imath} + a_y\hat{\jmath}$$

Velocity and Acceleration in Two Dimensions

Motion in a straight line may or may not involve acceleration, but motion on curved paths in two or three dimensions is *always* accelerated motion. Why? Because moving in multiple dimensions means *changing direction*—and *any* change in velocity, including direction,

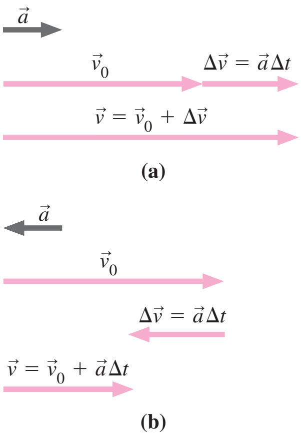


FIGURE 3.7 When $\vec{v}$ and $\vec{a}$ are colinear, only the speed changes.

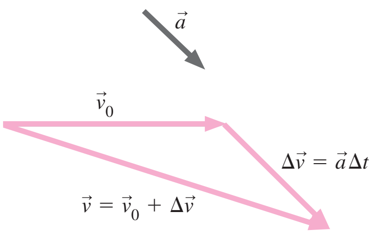


FIGURE 3.8 In general, acceleration changes both the magnitude and the direction of velocity.

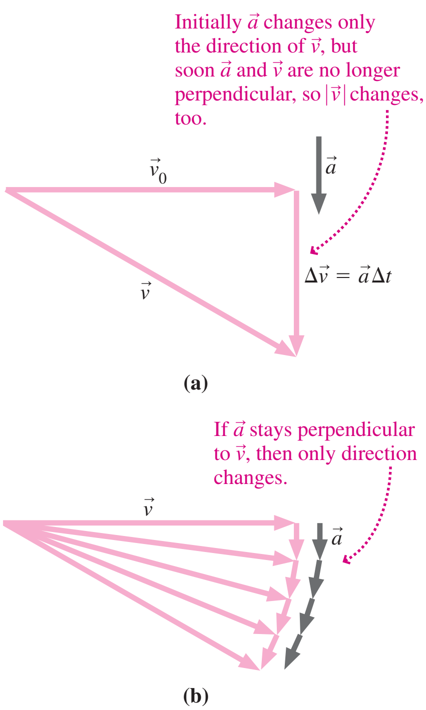


FIGURE 3.9 Acceleration that is always perpendicular to velocity changes only the direction.

involves acceleration. Get used to thinking of acceleration as meaning more than "speeding up" or "slowing down." It can equally well mean "changing direction," whether or not speed is also changing. Whether acceleration results in a speed change, a direction change, or both depends on the relative orientation of the velocity and acceleration vectors.

Suppose you're driving down a straight road at speed v_0 when you step on the gas to give a constant acceleration $\vec{a}$ for a time Δt. Equation 3.5 shows that the change in your velocity is $\Delta\vec{v} = \vec{a}\Delta t$. In this case the acceleration is in the same direction as your velocity and, as Fig. 3.7*a* shows, the result is an increase in the magnitude of your velocity; that is, you speed up. Step on the brake, and your acceleration is opposite your velocity, and you slow down (Fig. 3.7*b*).

✓TIP Vectors Tell It All

Are you thinking there should be a minus sign in Fig. 3.7*b* because the speed is decreasing? Nope: Vectors have both magnitude and direction, and the vector addition $\vec{v} = \vec{v}_0 + \vec{a}\Delta t$ tells it all. In Fig. 3.7*b*, $\Delta\vec{v}$ points to the left, and that takes care of the "subtraction."

In two dimensions acceleration and velocity can be at any angle. In general, acceleration then changes both the magnitude and the direction of the velocity (Fig. 3.8). Particularly interesting is the case when $\vec{a}$ is perpendicular to $\vec{v}$; then only the direction of motion changes. If acceleration is constant—in both magnitude and direction—then the two vectors won't stay perpendicular once the direction of $\vec{v}$ starts to change, and the magnitude will change, too. But in the special case where acceleration changes direction so it's always perpendicular to velocity, then it's strictly true that only the direction of motion changes. Figure 3.9 illustrates this point, which we'll soon explore quantitatively.

GOT IT? 3.2 An object is accelerating downward. Which, if any, of the following must be true? (a) The object cannot be moving upward. (b) The object cannot be moving in a straight line. (c) The object is moving directly downward. (d) If the object's motion is instantaneously horizontal, it can't continue to be so.

3.3 Relative Motion

You stroll down the aisle of a plane, walking toward the front at a leisurely 4 km/h. Meanwhile the plane is moving relative to the ground at 1000 km/h. Obviously, then, you're moving at 1004 km/h relative to the ground. As this example suggests, velocity is meaningful only when we know the answer to the question, Velocity relative to what? That "what" is called a **frame of reference**. Often we know an object's velocity relative to one frame of reference—for example, your velocity relative to the plane—and we want to know its velocity relative to some other reference frame—in this case the ground. In this one-dimensional case, we can simply add the two velocities. If you had been walking toward the back of the plane, then the two velocities would have opposite signs and you would be going at 996 km/h relative to the ground.

The same idea works in two dimensions, but here we need to recognize that velocity is a vector. Suppose that airplane is flying with velocity $\vec{v}\,'$ relative to the air. If a wind is blowing, then the air is moving with some velocity $\vec{V}$ relative to the ground. The plane's velocity $\vec{v}$ relative to the ground is the vector sum of its velocity relative to the air and the air's velocity relative to the ground:

$$\vec{v} = \vec{v}\,' + \vec{V} \qquad \text{(relative velocity)} \tag{3.7}$$

Here we use lowercase letters for the velocities of an object relative to two different reference frames; we distinguish the two with the prime on one of the velocities. The capital $\vec{V}$ is the relative velocity between the two frames. In general, Equation 3.7 lets us use the velocity of an object in one reference frame to find its velocity relative to another frame—provided we know that relative velocity $\vec{V}$. Example 3.2 illustrates the application of this idea to aircraft navigation.

EXAMPLE 3.2 Relative Velocity: Navigating a Jetliner

A jetliner flies at 960 km/h relative to the air. It's going from Houston to Omaha, 1290 km northward. At cruising altitude a wind is blowing eastward at 190 km/h. In what direction should the plane fly? How long will the trip take?

INTERPRET This is a problem involving relative velocities. We identify the given information: the plane's speed, but not its direction, in the reference frame of the air; the plane's direction, but not its speed, in the reference frame of the ground; and the wind velocity, both speed and direction.

DEVELOP Equation 3.7, $\vec{v} = \vec{v}' + \vec{V}$, applies, and we identify $\vec{v}$ as the plane's velocity relative to the ground, $\vec{v}'$ as its velocity relative to the air, and $\vec{V}$ as the wind velocity. Equation 3.7 shows that $\vec{v}'$ and $\vec{V}$ add vectorially to give $\vec{v}$; that, with the given information, helps us draw the situation (Fig. 3.10). Measuring the angle of $\vec{v}'$ and the length of $\vec{v}$ in the diagram would then give the answers. However, we'll work the problem algebraically using vector components. Since the plane is flying northward and the wind is blowing eastward, a suitable coordinate system has x-axis eastward and y-axis northward. Our plan is to work out the vector components in these coordinates and then apply Equation 3.7.

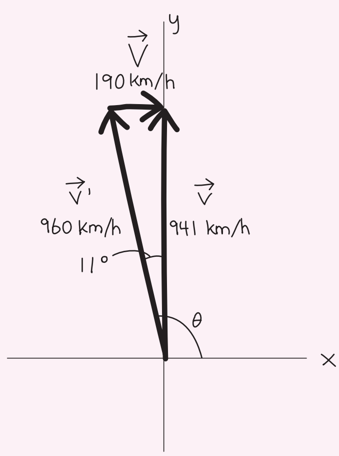


FIGURE 3.10 Our vector diagram for Example 3.2.

EVALUATE Using Equations 3.2 for the vector components, we can express the three vectors as

$$\vec{v}' = v'\cos\theta\,\hat{\imath} + v'\sin\theta\,\hat{\jmath}, \quad \vec{V} = V\hat{\imath}, \quad \text{and} \quad \vec{v} = v\hat{\jmath}$$

Here we know the magnitude v' of the velocity $\vec{v}'$, but we don't know the angle θ. We know the magnitude V of the wind velocity $\vec{V}$, and we also know its direction—toward the east. So $\vec{V}$ has only an x-component. Meanwhile we want the velocity $\vec{v}$ relative to the ground to be purely northward, so it has only a y-component—although we don't know its magnitude v. We're now ready to put the three velocities into Equation 3.7. Since two vectors are equal only if all their components are equal, we can express the vector Equation 3.7 as two separate scalar equations for the x- and y-components:

$$x\text{-component:} \quad v'\cos\theta + V = 0$$

$$y\text{-component:} \quad v'\sin\theta + 0 = v$$

The rest is math, evaluating the unknowns θ and v. Solving the x equation gives

$$\theta = \cos^{-1}\left(-\frac{V}{v'}\right) = \cos^{-1}\left(-\frac{190\text{ km/h}}{960\text{ km/h}}\right) = 101.4°$$

This angle is measured from the x-axis (eastward; see Fig. 3.10), so it amounts to a flight path 11° west of north. We can then evaluate v from the y equation:

$$v = v'\sin\theta = (960\text{ km/h})(\sin 101.4°) = 941\text{ km/h}$$

That's the plane's speed relative to the ground. Going 1290 km will then take $(1290\text{ km})/(941\text{ km/h}) = 1.4$ h.

ASSESS Make sense? The plane's heading of 11° west of north seems reasonable compensation for an eastward wind blowing at 190 km/h, given the plane's airspeed of 960 km/h. If there were no wind, the trip would take 1 h, 20 min (1290 km divided by 960 km/h), so our time of 1 h, 24 min with the wind makes sense. ■

3.4 Constant Acceleration

When acceleration is constant, the individual components of the acceleration vector are themselves constant. Furthermore, the component of acceleration in one direction has no effect on the motion in a perpendicular direction (Fig. 3.11, next page). Then with constant acceleration, the separate components of the motion must obey the constant-acceleration formulas we developed in Chapter 2 for one-dimensional motion. Using vector notation, we can then generalize Equations 2.7 and 2.10 to read

$$\vec{v} = \vec{v}_0 + \vec{a}t \qquad \text{(for constant acceleration only)} \tag{3.8}$$

$$\vec{r} = \vec{r}_0 + \vec{v}_0 t + \tfrac{1}{2}\vec{a}t^2 \qquad \text{(for constant acceleration only)} \tag{3.9}$$

where $\vec{r}$ is the position vector. In two dimensions, each of these vector equations represents a pair of scalar equations describing constant acceleration in two mutually perpendicular directions. Equation 3.9, for example, contains the pair $x = x_0 + v_{x0}t + \frac{1}{2}a_x t^2$ and

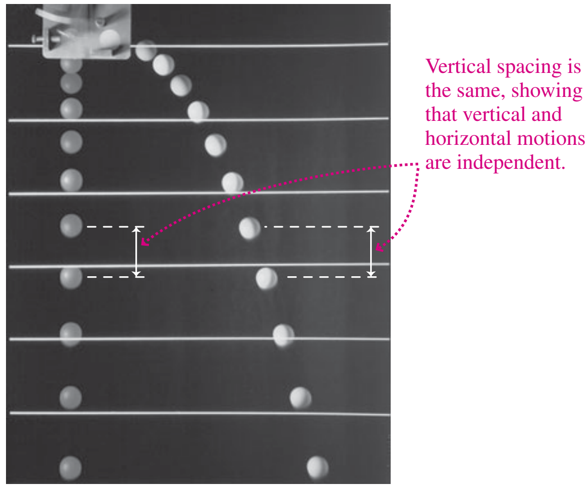


FIGURE 3.11 Two marbles, one dropped and the other projected horizontally.

$y = y_0 + v_{y0}t + \frac{1}{2}a_y t^2$. (Remember that the components of the displacement vector $\vec{r}$ are just the coordinates x and y.) In three dimensions there would be a third equation for the z-component. Starting with these vector forms of the equations of motion, you can apply Problem-Solving Strategy 2.1 to problems in two or three dimensions.

EXAMPLE 3.3 Acceleration in Two Dimensions: Windsurfing

You're windsurfing at 7.3 m/s when a gust hits, accelerating your sailboard at 0.82 m/s^2 at 60° to your original direction. If the gust lasts 8.7 s, what's the board's displacement during this time?

INTERPRET This is a problem involving constant acceleration in two dimensions. The key concept is that motion in perpendicular directions is independent, so we can treat the problem as involving two separate one-dimensional motions.

DEVELOP Equation 3.9, $\vec{r} = \vec{r}_0 + \vec{v}_0 t + \frac{1}{2}\vec{a}t^2$, will give the board's displacement. We need a coordinate system, so we take the x-axis along the board's initial motion, with the origin at the point where the gust first hits. Our plan is to find the components of the acceleration vector and then apply the two components of Equation 3.9 to get the components of the displacement. In Fig. 3.12 we draw the acceleration vector to determine its components.

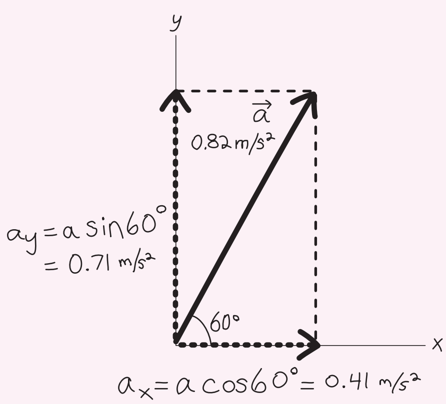


FIGURE 3.12 Our sketch of the sailboard's acceleration components.

EVALUATE With the x-direction along the initial velocity, $\vec{v}_0 = 7.3\hat{\imath}$ m/s. As Fig. 3.12 shows, the acceleration is $\vec{a} = 0.41\hat{\imath} + 0.71\hat{\jmath}$ m/s^2. Our choice of origin gives $x_0 = y_0 = 0$, so the two components of Equation 3.9 are

$$x = v_{x0}t + \tfrac{1}{2}a_x t^2 = 79.0 \text{ m}$$
$$y = \tfrac{1}{2}a_y t^2 = 26.9 \text{ m}$$

where we used the appropriate components of $\vec{a}$ and where $t = 8.7$ s. The new position vector is then $\vec{r} = x\hat{\imath} + y\hat{\jmath} = 79.0\hat{\imath} + 26.9\hat{\jmath}$ m, giving a net displacement of $r = \sqrt{x^2 + y^2} = 83$ m.

ASSESS Make sense? Figure 3.13 shows how the acceleration deflects the sailboard from its original path and also increases its speed somewhat. Since the acceleration makes a fairly large angle with the initial velocity, the change in direction is the greater effect. ■

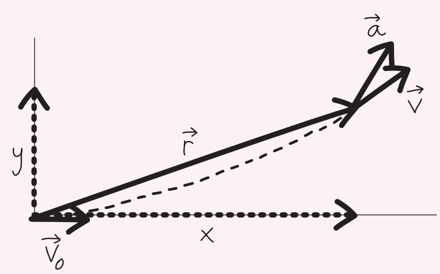


FIGURE 3.13 Our sketch of the displacement $\vec{r}$ and curved path (dashed) for the sailboard.

GOT IT? 3.3 An object is moving initially in the $+x$-direction. Which of the following accelerations, all acting for the same time interval, will cause the greatest change in its speed? In its direction? (a) $10\hat{\imath}\ \mathrm{m/s^2}$, (b) $10\hat{\jmath}\ \mathrm{m/s^2}$, (c) $10\hat{\imath} + 5\hat{\jmath}\ \mathrm{m/s^2}$, (d) $2\hat{\imath} - 8\hat{\jmath}\ \mathrm{m/s^2}$

3.5 Projectile Motion

A **projectile** is an object that's launched into the air and then moves predominantly under the influence of gravity. Examples are numerous; baseballs, jets of water, fireworks, missiles, ejecta from volcanoes, drops of ink in an ink-jet printer, and leaping dolphins are all projectiles.

To treat projectile motion, we make two simplifying assumptions: (1) We neglect any variation in the direction or magnitude of the gravitational acceleration, and (2) we neglect air resistance. The first assumption is equivalent to neglecting Earth's curvature, and is valid for projectiles whose displacements are small compared with Earth's radius. Air resistance has a more variable effect; for dense, compact objects it's often negligible, but for objects whose ratio of surface area to mass is large—like ping-pong balls and parachutes—air resistance dramatically alters the motion.

To describe projectile motion, it's convenient to choose a coordinate system with the y-axis vertically upward and the x-axis horizontal. With the only acceleration provided by gravity, $a_x = 0$ and $a_y = -g$, so the components of Equations 3.8 and 3.9 become

$$\left.\begin{aligned} v_x &= v_{x0} && (3.10)\\ v_y &= v_{y0} - gt && (3.11)\\ x &= x_0 + v_{x0}t && (3.12)\\ y &= y_0 + v_{y0}t - \tfrac{1}{2}gt^2 && (3.13)\end{aligned}\right\} \text{(for constant gravitational acceleration)}$$

We take g to be positive, and account for the downward direction using minus signs. Equations 3.10–3.13 tell us mathematically what Fig. 3.14 tells us physically: Projectile motion comprises two perpendicular and independent components—horizontal motion with constant velocity and vertical motion with constant acceleration.

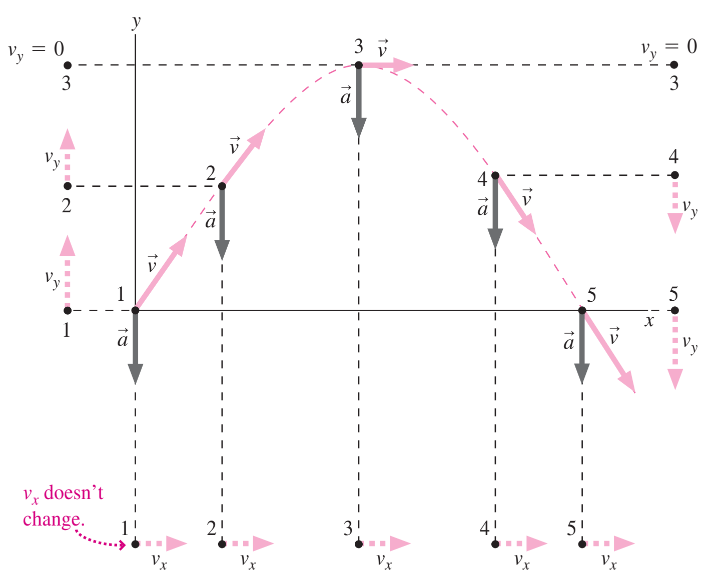


FIGURE 3.14 Velocity and acceleration at five points on a projectile's path. Also shown are horizontal and vertical components.

PROBLEM-SOLVING STRATEGY 3.1 Projectile Motion

INTERPRET Make sure that you have a problem involving the constant acceleration of gravity near Earth's surface, and that the motion involves both horizontal and vertical components. Identify the object or objects in question and whatever initial or final positions and velocities are given. Know what quantities you're being asked to find.

DEVELOP Establish a horizontal/vertical coordinate system, and write the separate components of the equations of motion (Equations 3.10–3.13). The equations for different components will be linked by a common variable—namely, time. Draw a sketch showing the initial motion and a rough trajectory.

EVALUATE Solve your individual equations simultaneously for the unknowns of the problem.

ASSESS Check that your answer makes sense. Consider special cases, like purely vertical or horizontal initial velocities. Because the equations of motion are quadratic in time, you may have two answers. One answer may be the one you want, but you gain more insight into physics if you consider the meaning of the second answer, too.

EXAMPLE 3.4 Finding the Horizontal Distance: Washout!

A raging flood has washed away a section of highway, creating a gash 1.7 m deep. A car moving at 31 m/s goes straight over the edge. How far from the edge of the washout does it land?

INTERPRET This is a problem involving projectile motion, and it asks for the horizontal distance the car moves after it leaves the road. We're given the car's initial speed and direction (horizontal) and the distance it falls.

DEVELOP Figure 3.15*a* shows the situation, and we've sketched the essentials in Fig. 3.15*b*. Since there's no horizontal acceleration, Equation 3.12, $x = x_0 + v_{x0}t$, would determine the unknown distance if we knew the time. But horizontal and vertical motions are independent, so we can find the time until the car hits the ground from the vertical motion alone, as determined by Equation 3.13, $y = y_0 + v_{y0}t - \frac{1}{2}gt^2$. So our plan is to get the time from Equation 3.13 and then use that time in Equation 3.12 to get the horizontal distance. If we choose the origin as the bottom of the washout, then $y_0 = 1.7$ m. Then we want the time when $y = 0$.

EVALUATE With $v_{y0} = 0$, we solve Equation 3.13 for t:

$$t = \sqrt{\frac{2y_0}{g}} = \sqrt{\frac{(2)(1.7\text{ m})}{(9.8\text{ m/s}^2)}} = 0.589\text{ s}$$

During this time the car continues to move horizontally at $v_{x0} = 31$ m/s, so Equation 3.12 gives $x = v_{x0}t = (31\text{ m/s})(0.589\text{ s}) = 18$ m.

ASSESS Make sense? About half a second to drop 1.7 m or about 6 ft seems reasonable, and at 31 m/s an object will go somewhat farther than 15 m in this time.

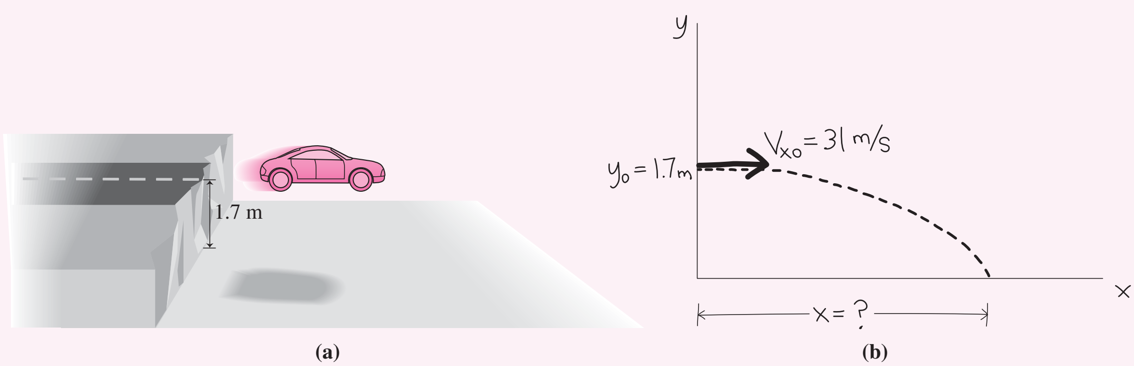


FIGURE 3.15 (a) The highway and car, and (b) our sketch.

✓TIP Multistep Problems

Example 3.4 asked for the horizontal distance the car traveled. For that we needed the time—which we weren't given. This is a common situation in all but the simplest physics problems. You need to work through several steps to get the answer—in this case solving first for the unknown time and then for the distance. In essence, we solved two problems in Example 3.4: the first involving vertical motion and the second horizontal motion.

Instead of calculating t numerically in Example 3.4, we could have kept $t = \sqrt{2y_0/g}$ in symbolic form until the end. That would avoid roundoff error and having to keep track of numerical digits and units. And you can often gain more physical insight from an answer that's expressed symbolically before you put in the numbers.

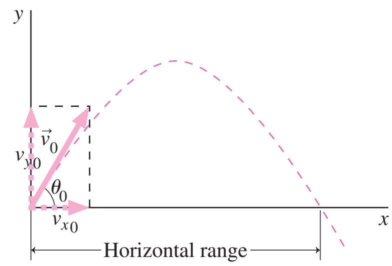


FIGURE 3.16 Parabolic trajectory of a projectile.

Projectile Trajectories

We're often interested in the path, or **trajectory**, of a projectile without the details of where it is at each instant of time. We can specify the trajectory by giving the height y as a function of the horizontal position x. Consider a projectile launched from the origin at some angle θ_0 to the horizontal, with initial speed v_0. As Fig. 3.16 suggests, the components of the initial velocity are $v_{x0} = v_0 \cos\theta_0$ and $v_{y0} = v_0 \sin\theta_0$. Then Equations 3.12 and 3.13 become

$$x = v_0 \cos\theta_0 t \quad \text{and} \quad y = v_0 \sin\theta_0 t - \tfrac{1}{2}gt^2$$

Solving the x equation for the time t gives

$$t = \frac{x}{v_0 \cos\theta_0}$$

Using this result in the y equation, we have

$$y = v_0 \sin\theta_0\left(\frac{x}{v_0 \cos\theta_0}\right) - \tfrac{1}{2}g\left(\frac{x}{v_0 \cos\theta_0}\right)^2$$

or

$$y = x \tan\theta_0 - \frac{g}{2v_0^2 \cos^2\theta_0}x^2 \qquad \text{(projectile trajectory)} \qquad (3.14)$$

Equation 3.14 gives a mathematical description of the projectile's trajectory. Since y is a quadratic function of x, the trajectory is a parabola.

APPLICATION **Pop Flies, Line Drives, and Hang Times**

Although air resistance significantly influences baseball trajectories, to a first approximation baseballs behave like projectiles. For a given speed off the bat, this means a pop fly's "hang time" is much greater than that of a nearly horizontal line drive, and that makes the fly ball much easier to catch (see photo).

EXAMPLE 3.5 Finding the Trajectory: Out of the Hole

A construction worker stands in a 2.6-m-deep hole, 3.1 m from the edge of the hole. He tosses a hammer to a companion outside the hole. If the hammer leaves his hand 1.0 m above the bottom of the hole at an angle of 35°, what's the minimum speed it needs to clear the edge of the hole? How far from the edge of the hole does it land?

INTERPRET We're concerned about *where* an object is but not *when*, so we interpret this as a problem about the trajectory—specifically, the minimum-speed trajectory that just grazes the edge of the hole.

DEVELOP We draw the situation in Fig. 3.17. Equation 3.14 determines the trajectory, so our plan is to find the speed that makes the trajectory pass just over the edge of the hole at $x = 3.1$ m, $y = 1.6$ m, where Fig. 3.17 shows that we've chosen a coordinate system with its origin at the worker's hand.

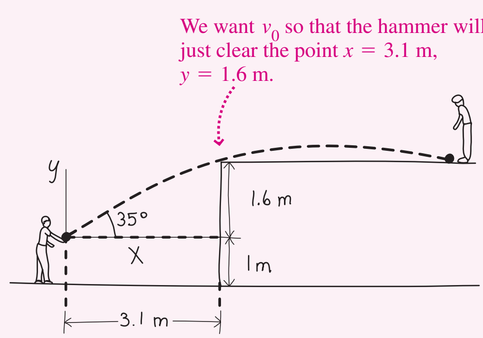


FIGURE 3.17 Our sketch for Example 3.5.

(continued)

EVALUATE To find the minimum speed we solve Equation 3.14 for v_0, using the coordinates of the hole's edge for x and y:

$$v_0 = \sqrt{\frac{gx^2}{2\cos^2\theta_0\,(x\tan\theta_0 - y)}} = 11\ \text{m/s}$$

To find where the hammer lands, we need to know the horizontal position when $y = 1.6$ m. Rearranging Equation 3.14 into the standard form for a quadratic equation gives $(g/2v_0^2\cos^2\theta_0)x^2 - (\tan\theta_0)x + y = 0$. Applying the quadratic formula (Appendix A) gives $x = 3.1$ m and $x = 8.7$ m; the second value is the one we want. That 8.7 m is the distance from our origin at the worker's hand, and amounts to 8.7 m − 3.1 m = 5.6 m from the hole's edge.

ASSESS Make sense? The other answer to the quadratic, $x = 3.1$ m, is a clue that we did the problem correctly. That 3.1 m is the distance to the edge of the hole. The fact that we get this position when we ask for a vertical height of 1.6 m confirms that the trajectory does indeed just clear the edge of the hole.

The Range of a Projectile

How far will a soccer ball go if I kick it at 12 m/s at 50° to the horizontal? If I can throw a rock at 15 m/s, can I get it across a 30-m-wide pond? How far off vertical can a rocket's trajectory be and still land within 50 km of its launch point? As in these examples, we're frequently interested in the **horizontal range** of a projectile—that is, how far it moves horizontally over level ground.

For a projectile launched on level ground, we can determine when the projectile will return to the ground by setting $y = 0$ in Equation 3.14:

$$0 = x\tan\theta_0 - \frac{g}{2v_0^2\cos^2\theta_0}x^2 = x\left(\tan\theta_0 - \frac{gx}{2v_0^2\cos^2\theta_0}\right)$$

There are two solutions: $x = 0$, corresponding to the launch point, and

$$x = \frac{2v_0^2}{g}\cos^2\theta_0\tan\theta_0 = \frac{2v_0^2}{g}\sin\theta_0\cos\theta_0$$

But $\sin 2\theta_0 = 2\sin\theta_0\cos\theta_0$, so this becomes

$$x = \frac{v_0^2}{g}\sin 2\theta_0 \qquad \text{(horizontal range)} \tag{3.15}$$

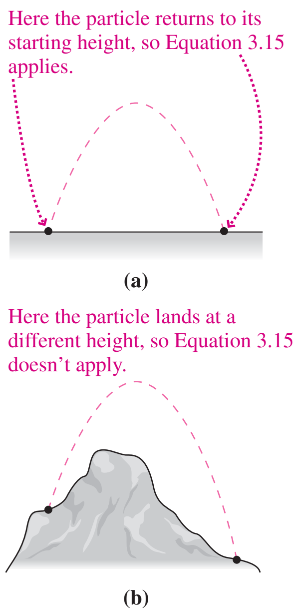


FIGURE 3.18 Equation 3.15 applies in (a) but not in (b).

✓TIP Know Your Limits

We emphasize that Equation 3.15 gives the *horizontal* range—the distance a projectile travels horizontally before returning *to its starting height*. From the way it was derived—setting $y = 0$—you can see that it does *not* give the horizontal distance when the projectile returns to a different height (Fig. 3.18).

The maximum range occurs when $\sin 2\theta = 1$ in Equation 3.15, which occurs when $\theta = 45°$. As Fig. 3.19 suggests, the range for a given launch speed v_0 is equal for angles equally spaced on either side of 45°—as you can prove in Problem 66.

CONCEPTUAL EXAMPLE 3.1 Projectile Flight Times

The ranges in Fig. 3.19 are equal for angles on either side of 45°. How do the flight times compare?

EVALUATE We're being asked about the times projectiles spend on the trajectories shown. Since horizontal and vertical motions are independent, flight time depends on how high the projectile goes. So we can argue from the vertical motions that the trajectory with the higher launch angle takes longer. We can also argue from horizontal motions: Horizontal distances of the paired trajectories are the same, but the lower trajectory has a greater horizontal velocity component, so again the lower trajectory takes less time.

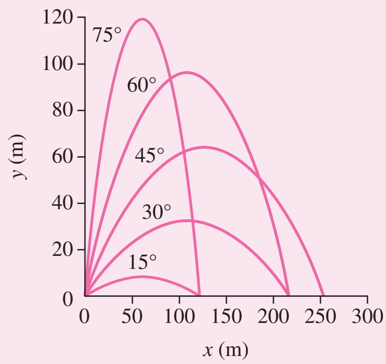


FIGURE 3.19 Trajectories for a projectile launched at 50 m/s.

ASSESS Consider the extreme cases of near-vertical and near-horizontal trajectories. The former goes nearly straight up and down, taking a relatively long time but returning essentially to its starting point. The latter hardly gets anywhere because it immediately hits the ground right at its starting point, so it takes just about no time!

MAKING THE CONNECTION Find the flight times for the 30° and 60° trajectories in Fig. 3.19.

EVALUATE The range of Equation 3.15 is also equal to the horizontal velocity v_x multiplied by the time: $v_x t = v_0^2 \sin 2\theta_0/g$. Using $v_{x0} = v_0 \cos\theta_0$ and solving for t gives $t = 2v_0 \sin\theta_0/g$. Using Fig. 3.19's $v_0 = 50$ m/s yields $t_{30} = 5.1$ s and $t_{60} = 8.8$ s. You can explore this time difference more generally in Problem 61.

✓TIP Know the Fundamentals

Equations 3.14 and 3.15 for a projectile's trajectory and range are useful, but they're not fundamental equations of physics. Both follow directly from the equations for constant acceleration. If you think that specialized results like Equations 3.14 and 3.15 are on an equal footing with more fundamental equations and principles, then you're seeing physics as a hodgepodge of equations and missing the big picture of a science with a few underlying principles from which all else follows.

EXAMPLE 3.6 Projectile Range: Probing the Atmosphere

After a short engine firing, an atmosphere-probing rocket reaches 4.6 km/s. If the rocket must land within 50 km of its launch site, what's the maximum allowable deviation from a vertical trajectory?

INTERPRET Although we're asked about the launch angle, the 50-km criterion is a clue that we can interpret this as a problem about the horizontal range. That "short engine firing" means we can neglect the distance over which the rocket fires and consider it a projectile that leaves the ground at $v_0 = 4.6$ km/s.

DEVELOP Equation 3.15, $x = (v_0^2/g)\sin 2\theta_0$, determines the horizontal range, so our plan is to solve that equation for θ_0 with range $x = 50$ km.

EVALUATE We have $\sin 2\theta_0 = gx/v_0^2 = 0.0232$. There are two solutions, corresponding to $2\theta_0 = 1.33°$ and $2\theta_0 = 180° - 1.33°$. The second is the one we want, giving a launch angle $\theta_0 = 90° - 0.67°$. Therefore the launch angle must be within 0.67° of vertical.

ASSESS Make sense? At 4.6 km/s, this rocket goes quite high, so with even a small deviation from vertical it will land far from its launch point. Again we've got two solutions. The one we rejected is like the low trajectories of Fig. 3.19; although it gives a 50-km range, it isn't going to get our rocket high into the atmosphere. ■

3.6 Uniform Circular Motion

An important case of accelerated motion in two dimensions is **uniform circular motion**—that of an object describing a circular path at constant speed. Although the speed is constant, the motion is accelerated because the *direction* of the velocity is changing.

Uniform circular motion is common. Many spacecraft are in circular orbits, and the orbits of the planets are approximately circular. Earth's daily rotation carries you around in uniform circular motion. Pieces of rotating machinery describe uniform circular motion, and you're temporarily in circular motion as you drive around a curve. Electrons undergo circular motion in magnetic fields.

Here we derive an important relationship among the acceleration, speed, and radius of uniform circular motion. Figure 3.20 shows several velocity vectors for an object moving with speed v around a circle of radius r. Velocity vectors are tangent to the circle, indicating

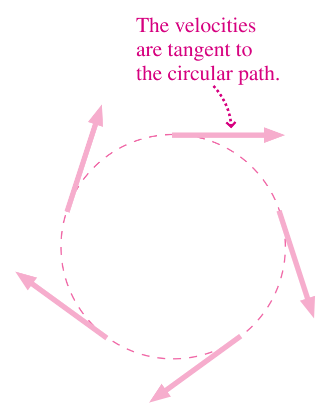


FIGURE 3.20 Velocity vectors in circular motion are tangent to the circular path.

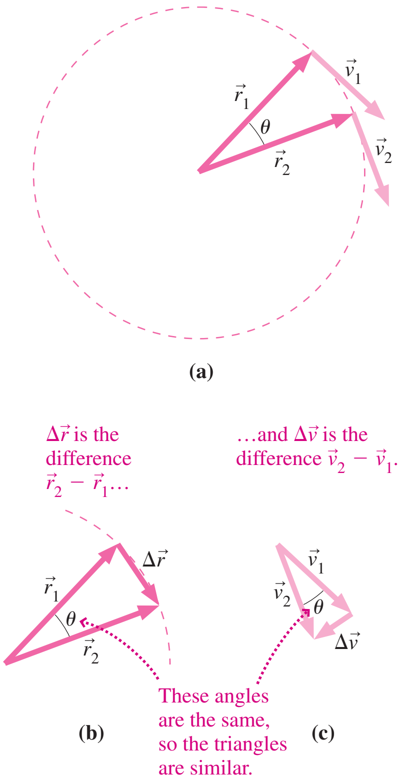


FIGURE 3.21 Position and velocity vectors for two nearby points on the circular path.

the instantaneous direction of motion. In Fig. 3.21*a* we focus on two nearby points described by position vectors $\vec{r}_1$ and $\vec{r}_2$, where the velocities are $\vec{v}_1$ and $\vec{v}_2$. Figures 3.21*b* and *c* show the corresponding displacement $\Delta\vec{r} = \vec{r}_2 - \vec{r}_1$ and velocity difference $\Delta\vec{v} = \vec{v}_2 - \vec{v}_1$.

Because $\vec{v}_1$ is perpendicular to $\vec{r}_1$, and $\vec{v}_2$ is perpendicular to $\vec{r}_2$, the angles θ shown in all three parts of Fig. 3.21 are the same. Therefore, the triangles in Fig. 3.21*b* and *c* are similar, and we can write

$$\frac{\Delta v}{v} = \frac{\Delta r}{r}$$

Now suppose the angle θ is small, corresponding to a short time interval Δt for motion from position $\vec{r}_1$ to $\vec{r}_2$. Then the length of the vector $\Delta\vec{r}$ is approximately the length of the circular arc joining the endpoints of the position vectors, as suggested in Fig. 3.21*b*. The length of this arc is the distance the object travels in the time Δt, or $v\Delta t$, so $\Delta r \simeq v\Delta t$. Then the relation between similar triangles becomes

$$\frac{\Delta v}{v} \simeq \frac{v\,\Delta t}{r}$$

Rearranging this equation gives an approximate expression for the magnitude of the average acceleration:

$$\bar{a} = \frac{\Delta v}{\Delta t} \simeq \frac{v^2}{r}$$

Taking the limit $\Delta t \to 0$ gives the instantaneous acceleration; in this limit the angle θ approaches 0, the circular arc and $\Delta\vec{r}$ become indistinguishable, and the relation $\Delta r \simeq v\,\Delta t$ becomes exact. So we have

$$a = \frac{v^2}{r} \qquad \text{(uniform circular motion)} \qquad (3.16)$$

for the magnitude of the instantaneous acceleration of an object moving in a circle of radius r at constant speed v. What about its direction? As Fig. 3.21*c* suggests, $\Delta\vec{v}$ is very nearly perpendicular to both velocity vectors; in the limit $\Delta t \to 0$, $\Delta\vec{v}$ and the acceleration $\Delta\vec{v}/\Delta t$ become exactly perpendicular to the velocity. The direction of the acceleration vector is therefore toward the center of the circle.

Clearly, our geometric argument would work for any point on the circle, so we conclude that the acceleration has constant magnitude v^2/r and always points toward the center of the circle. Isaac Newton coined the term *centripetal* to describe this center-pointing acceleration. However, we'll use that term sparingly because we want to emphasize that centripetal acceleration is fundamentally no different from any other acceleration: It's simply a vector describing the rate of change of velocity.

Does Equation 3.16 make sense? Yes. An increase in speed v means the time Δt for a given change in direction of the velocity becomes shorter. Not only that, but the associated change $\Delta\vec{v}$ in velocity is larger. These two effects combine to give an acceleration that depends on the *square* of the speed. On the other hand, an increase in the radius with a fixed speed increases the time Δt associated with a given change in velocity, so the acceleration is inversely proportional to the radius.

✓TIP Circular Motion and Constant Acceleration

The direction toward the center changes as an object moves around a circular path, so the acceleration vector is *not constant*, even though its magnitude is. Uniform circular motion is *not* motion with constant acceleration, and our constant-acceleration equations *do not apply*. In fact, we know that constant acceleration in two dimensions implies a parabolic trajectory, not a circle.

EXAMPLE 3.7 Uniform Circular Motion: The International Space Station

Find the orbital period (the time to complete one orbit) of the International Space Station in its circular orbit at altitude 400 km, where the acceleration of gravity is 89% of its surface value.

INTERPRET This is a problem about uniform circular motion.

DEVELOP Given the radius and acceleration, we could use Equation 3.16, $a = v^2/r$, to determine the orbital speed. But we're given the altitude, not the orbital radius, and we want the period, not the speed. So our plan is to write the speed in terms of the period and use the result in Equation 3.16. The orbital altitude is the distance from Earth's surface, so we'll need to add Earth's radius to get the orbital radius r.

EVALUATE The speed v is the orbital circumference, $2\pi r$, divided by the period T. Using this in Equation 3.16 gives

$$a = \frac{v^2}{r} = \frac{(2\pi r/T)^2}{r} = \frac{4\pi^2 r}{T^2}$$

Appendix E lists Earth's radius as $R_E = 6.37$ Mm, giving an orbital radius $r = R_E + 400$ km $= 6.77$ Mm. Solving our acceleration expression for the period then gives $T = \sqrt{4\pi^2 r/a} = 5536\text{ s} = 92\text{ min}$, where we used $a = 0.89g$.

ASSESS Make sense? You've probably heard that astronauts orbit Earth in about an hour and a half, experiencing multiple sunrises and sunsets in a 24-hour day. Our answer of 92 min is certainly consistent with that. There's no choice here; for a given orbital radius, Earth's size and mass determine the period. Because astronauts' orbits are limited to a few hundred kilometers, a distance small compared with R_E, variations in g and T are minimal. Any such "low Earth orbit" has a period of approximately 90 min. At higher altitudes, gravity diminishes significantly and periods lengthen; the Moon, for example, orbits in 27 days. We'll discuss orbits more in Chapter 8. ■

EXAMPLE 3.8 Uniform Circular Motion: Engineering a Road

An engineer is designing a flat, horizontal road for an 80 km/h speed limit (that's 22.2 m/s). If the maximum acceleration of a vehicle on this road is 1.5 m/s^2, what's the minimum safe radius for curves in the road?

INTERPRET Even though a curve is only a portion of a circle, we can still interpret this problem as involving uniform circular motion.

DEVELOP Equation 3.16, $a = v^2/r$, determines the acceleration given the speed and radius. Here we have the acceleration and speed, so our plan is to solve for the radius.

EVALUATE Using the given numbers, we have $r = v^2/a = (22.2\text{ m/s})^2/1.5\text{ m/s}^2 = 329$ m.

ASSESS Make sense? A speed of 80 km/h is pretty fast, so we need a wide curve to keep the required acceleration below its design value. If the curve is sharper, vehicles may slide off the road. We'll see more clearly in subsequent chapters how vehicles manage to negotiate high-speed curves. ■

Nonuniform Circular Motion

What if an object moves in a circular path but its speed changes? Then it has components of acceleration both perpendicular and parallel to its velocity. The former, the **radial acceleration** a_r, is what changes the direction to keep the object in circular motion. Its magnitude is still v^2/r, with v now the instantaneous speed. The parallel component of acceleration, also called **tangential acceleration** a_t because it's tangent to the circle, changes the speed but not the direction. Its magnitude is therefore the rate of change of speed, or dv/dt. Figure 3.22 shows these two acceleration components for a car rounding a curve.

Finally, what if the radius of a curved path changes? At any point on a curve we can define a **radius of curvature**. Then the radial acceleration is still v^2/r, and it can vary if either v or r changes along the curve. The tangential acceleration is still tangent to the curve, and it still describes the rate of change of speed. So it's straightforward to generalize the ideas of uniform circular motion to cases where the motion is nonuniform either because the speed changes, or because the radius changes, or both.

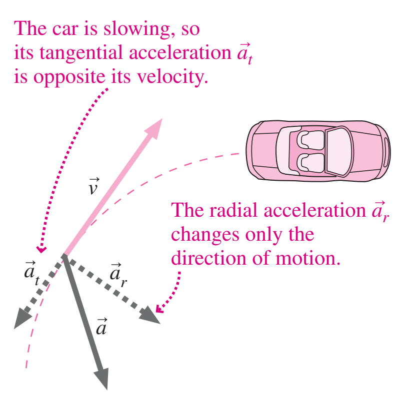
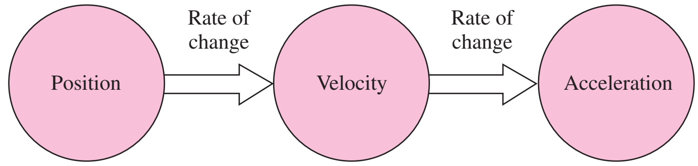


FIGURE 3.22 Acceleration of a car that slows as it rounds a curve.

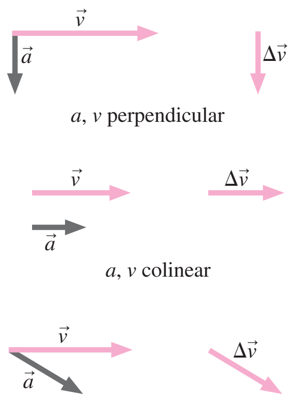
CHAPTER 3 SUMMARY

Big Picture

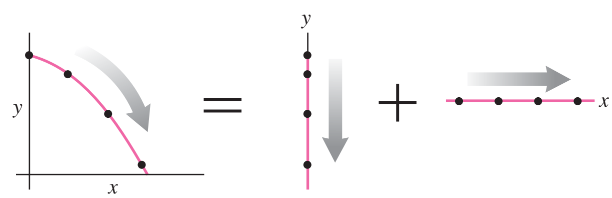
Quantities characterizing motion in two and three dimensions have both **magnitude** and **direction** and are described by **vectors**. Position, velocity, and acceleration are all vector quantities, related as they are in one dimension:



These vector quantities need not have the same direction. In particular, acceleration that's perpendicular to velocity changes the direction but not the magnitude of the velocity. Acceleration that's colinear changes only the magnitude of the velocity. In general, both change.



Components of motion in two perpendicular directions are independent. This reduces problems in two and three dimensions to sets of one-dimensional problems that can be solved with the methods of Chapter 2.



Key Concepts and Equations

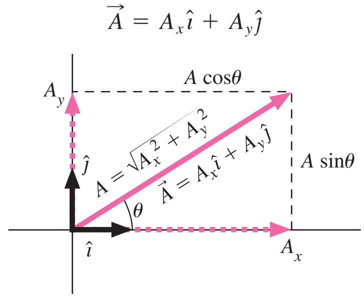
Vectors can be described by magnitude and direction or by components. In two dimensions these representations are related by

$$A = \sqrt{A_x^2 + A_y^2} \quad \text{and} \quad \theta = \tan^{-1}\frac{A_y}{A_x}$$

$$A_x = A\cos\theta \quad \text{and} \quad A_y = A\sin\theta$$

A compact way to express vectors involves unit vectors that have magnitude 1, have no units, and point along the coordinate axes:

$$\vec{A} = A_x\hat{\imath} + A_y\hat{\jmath}$$



Velocity is the rate of change of the position vector $\vec{r}$:

$$\vec{v} = \frac{d\vec{r}}{dt}$$

Acceleration is the rate of change of velocity:

$$\vec{a} = \frac{d\vec{v}}{dt}$$

Applications

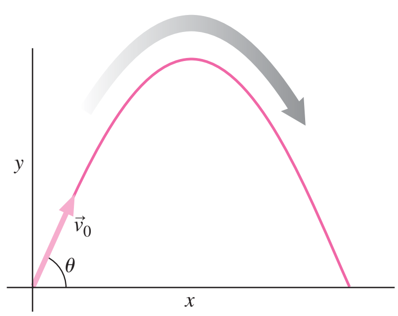
When acceleration is constant, motion is described by vector equations that generalize the one-dimensional equations of Chapter 2:

$$\vec{v} = \vec{v}_0 + \vec{a}t \qquad \vec{r} = \vec{r}_0 + \vec{v}_0 t + \tfrac{1}{2}\vec{a}t^2$$

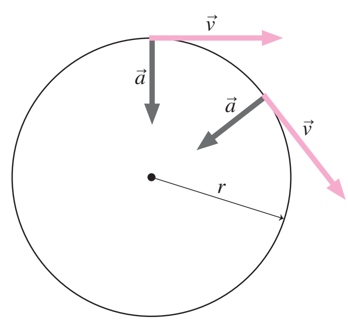
An important application of constant-acceleration motion in two dimensions is **projectile motion** under the influence of gravity.

Projectile trajectory:

$$y = x\tan\theta_0 - \frac{g}{2v_0^2\cos^2\theta_0}x^2$$



In **uniform circular motion** the magnitudes of velocity and acceleration remain constant, but their directions continually change. For an object moving in a circular path of radius r, the magnitudes of $\vec{a}$ and $\vec{v}$ are related by $a = v^2/r$.



For homework assigned on MasteringPhysics, go to www.masteringphysics.com | BIO Biology and/or medical-related problems | C Computer problems

For Thought and Discussion

1. Under what conditions is the magnitude of the vector sum $\vec{A} + \vec{B}$ equal to the sum of the magnitudes of the two vectors?
2. Can two vectors of equal magnitude sum to zero? How about two vectors of unequal magnitude?
3. Repeat Question 2 for three vectors.
4. Can an object have a southward acceleration while moving northward? A westward acceleration while moving northward?
5. You're a passenger in a car rounding a curve. The driver claims the car isn't accelerating because the speedometer reading is unchanging. Explain why the driver is wrong.
6. In what sense is Equation 3.8 really two (or three) equations?
7. Is a projectile's speed constant throughout its parabolic trajectory?
8. Is there any point on a projectile's trajectory where velocity and acceleration are perpendicular?
9. How is it possible for an object to be moving in one direction but accelerating in another?
10. You're in a bus moving with constant velocity on a level road when you throw a ball straight up. When the ball returns, does it land ahead of you, behind you, or back at your hand? Explain.

Exercises and Problems

Exercises

Section 3.1 Vectors

11. You walk west 220 m, then north 150 m. What are the magnitude and direction of your displacement vector?
12. An ion in a mass spectrometer follows a semicircular path of radius 15.2 cm. What are (a) the distance it travels and (b) the magnitude of its displacement?
13. A migrating whale follows the west coast of Mexico and North America toward its summer home in Alaska. It first travels 360 km northwest to just off the coast of northern California, and then turns due north and travels 400 km toward its destination. Determine graphically the magnitude and direction of its displacement.
14. Vector $\vec{A}$ has magnitude 3.0 m and points to the right; vector $\vec{B}$ has magnitude 4.0 m and points vertically upward. Find the magnitude and direction of vector $\vec{C}$ such that $\vec{A} + \vec{B} + \vec{C} = \vec{0}$.
15. Use unit vectors to express a displacement of 120 km at 29° counterclockwise from the x-axis.
16. Find the magnitude of the vector $34\hat{\imath} + 13\hat{\jmath}$ m and determine its angle to the x-axis.
17. (a) What's the magnitude of $\hat{\imath} + \hat{\jmath}$? (b) What angle does it make with the x-axis?

Section 3.2 Velocity and Acceleration Vectors

18. You're heading an international effort to save Earth from an asteroid heading toward us at 15 km/s. Your team mounts a rocket on the asteroid and fires it for 10 min, after which the asteroid is moving at 19 km/s at 28° to its original path. In a news conference, what do you report for the acceleration imparted to the asteroid?
19. An object is moving at 18 m/s at 220° counterclockwise from the x-axis. Find the x- and y-components of its velocity.
20. A car drives north at 40 mi/h for 10 min, then turns east and goes 5.0 mi at 60 mi/h. Finally, it goes southwest at 30 mi/h for 6.0 min. Determine the car's (a) displacement and (b) average velocity for this trip.
21. An object's velocity is $\vec{v} = ct^3\hat{\imath} + d\hat{\jmath}$, where t is time and c and d are positive constants with appropriate units. What's the direction of the object's acceleration?
22. A car, initially going eastward, rounds a 90° curve and ends up heading southward. If the speedometer reading remains constant, what's the direction of the car's average acceleration vector?
23. What are (a) the average velocity and (b) the average acceleration of the tip of the 2.4-cm-long hour hand of a clock in the interval from noon to 6 PM? Use unit vector notation, with the x-axis pointing toward 3 and the y-axis toward noon.
24. An ice skater is gliding along at 2.4 m/s, when she undergoes an acceleration of magnitude 1.1 m/s^2 for 3.0 s. After that she's moving at 5.7 m/s. Find the angle between her acceleration vector and her initial velocity.
25. An object is moving in the x-direction at 1.3 m/s when it undergoes an acceleration $\vec{a} = 0.52\hat{\jmath}$ m/s^2. Find its velocity vector after 4.4 s.

Section 3.3 Relative Motion

26. You're a pilot beginning a 1500-km flight. Your plane's speed is 1000 km/h, and air traffic control says you'll have to head 15° west of south to maintain a southward course. If the flight takes 100 min, what's the wind velocity?
27. You wish to row straight across a 63-m-wide river. You can row at a steady 1.3 m/s relative to the water, and the river flows at 0.57 m/s. (a) What direction should you head? (b) How long will it take you to cross the river?
28. A plane with airspeed 370 km/h flies perpendicularly across the jet stream, its nose pointed into the jet stream at 32° from the perpendicular direction of its flight. Find the speed of the jet stream.
29. A flock of geese is attempting to migrate due south, but the wind is blowing from the west at 5.1 m/s. If the birds can fly at 7.5 m/s relative to the air, what direction should they head?

Section 3.4 Constant Acceleration

30. The position of an object as a function of time is $\vec{r} = (3.2t + 1.8t^2)\hat{\imath} + (1.7t - 2.4t^2)\hat{\jmath}$ m, with t in seconds. Find the object's acceleration vector.
31. You're sailboarding at 6.5 m/s when a wind gust hits, lasting 6.3 s accelerating your board at 0.48 m/s^2 at 35° to your original direction. Find the magnitude and direction of your displacement during the gust.

Section 3.5 Projectile Motion

32. You toss an apple horizontally at 8.7 m/s from a height of 2.6 m. Simultaneously, you drop a peach from the same height. How long does each take to reach the ground?
33. A carpenter tosses a shingle horizontally off an 8.8-m-high roof at 11 m/s. (a) How long does it take the shingle to reach the ground? (b) How far does it move horizontally?
34. An arrow fired horizontally at 41 m/s travels 23 m horizontally. From what height was it fired?
35. Droplets in an ink-jet printer are ejected horizontally at 12 m/s and travel a horizontal distance of 1.0 mm to the paper. How far do they fall in this interval?
36. Protons drop 1.2 μm over the 1.7-km length of a particle accelerator. What's their approximate average speed?
37. If you can hit a golf ball 180 m on Earth, how far can you hit it on the Moon? (Your answer will be an underestimate because it neglects air resistance on Earth.)

Section 3.6 Uniform Circular Motion

38. How fast would a car have to round a 75-m-radius turn for its acceleration to be numerically equal to that of gravity?
39. Estimate the acceleration of the Moon, which completes a nearly circular orbit of 385,000 km radius in 27 days.

40. Global Positioning System (GPS) satellites circle Earth at altitudes of approximately 20,000 km, where the gravitational acceleration has 5.8% of its surface value. To the nearest hour, what's the orbital period of the GPS satellites?

Problems

41. Two vectors $\vec{A}$ and $\vec{B}$ have the same magnitude A and are at right angles. Find the magnitudes of (a) $\vec{A} + 2\vec{B}$ and (b) $3\vec{A} - \vec{B}$.
42. Vector $\vec{A}$ has magnitude 1.0 m and points 35° clockwise from the x-axis. Vector $\vec{B}$ has magnitude 1.8 m. Find the direction of $\vec{B}$ such that $\vec{A} + \vec{B}$ is in the y-direction.
43. Let $\vec{A} = 15\hat{\imath} - 40\hat{\jmath}$ and $\vec{B} = 31\hat{\jmath} + 18\hat{k}$. Find $\vec{C}$ such that $\vec{A} + \vec{B} + \vec{C} = \vec{0}$.
44. **BIO** A biologist looking through a microscope sees a bacterium at $\vec{r}_1 = 2.2\hat{\imath} + 3.7\hat{\jmath} - 1.2\hat{k}\ \mu\text{m}$. After 6.2 s, it's at $\vec{r}_2 = 4.6\hat{\imath} + 1.9\hat{k}\ \mu\text{m}$. Find (a) its average velocity, expressed in unit vectors, and (b) its average speed.
45. A particle's position is $\vec{r} = (ct^2 - 2dt^3)\hat{\imath} + (2ct^2 - dt^3)\hat{\jmath}$, where c and d are positive constants. Find expressions for times $t > 0$ when the particle is moving in (a) the x-direction and (b) the y-direction.
46. For the particle in Problem 45, is there any time $t > 0$ when the particle is (a) at rest and (b) accelerating in the x-direction? If either answer is "yes," find the time(s).
47. Attempting to stop on a slippery road, a car moving at 80 km/h skids at 30° to its initial motion, stopping in 3.9 s. Determine the average acceleration in m/s^2, in coordinates with the x-axis in the direction of the original motion and the y-axis toward the side to which the car skids.
48. An object undergoes acceleration $2.3\hat{\imath} + 3.6\hat{\jmath}$ m/s^2 for 10 s. At the end of this time, its velocity is $33\hat{\imath} + 15\hat{\jmath}$ m/s. (a) What was its velocity at the beginning of the 10-s interval? (b) By how much did its speed change? (c) By how much did its direction change? (d) Show that the speed change is not given by the magnitude of the acceleration multiplied by the time. Why not?
49. The Singapore Flyer is the world's largest Ferris wheel. Its diameter is 150 m and it rotates once every 30 min. Find the magnitudes of (a) the average velocity and (b) the average acceleration at the wheel's rim, over a 5.0-min interval.
50. A ferryboat sails between towns directly opposite each other on a river, moving at speed v' relative to the water. (a) Find an expression for the angle it should head at if the river flows at speed V. (b) What's the significance of your answer if $V > v'$?
51. The sum of two vectors, $\vec{A} + \vec{B}$, is perpendicular to their difference, $\vec{A} - \vec{B}$. How do the vectors' magnitudes compare?
52. Write an expression for a unit vector at 45° clockwise from the x-axis.
53. An object is initially moving in the x-direction at 4.5 m/s, when it undergoes an acceleration in the y-direction for a period of 18 s. If the object moves equal distances in the x- and y-directions during this time, what's the magnitude of its acceleration?
54. A particle leaves the origin with initial velocity $\vec{v}_0 = 11\hat{\imath} + 14\hat{\jmath}$ m/s, undergoing constant acceleration $\vec{a} = -1.2\hat{\imath} + 0.26\hat{\jmath}$ m/s^2. (a) When does the particle cross the y-axis? (b) What's its y-coordinate at the time? (c) How fast is it moving, and in what direction?
55. A kid fires a squirt gun horizontally from 1.6 m above the ground. It hits another kid 2.1 m away square in the back, 0.93 m above the ground. What was the water's initial speed?
56. A projectile has horizontal range R on level ground and reaches maximum height h. Find an expression for its initial speed.
57. You throw a baseball at a 45° angle to the horizontal, aiming at a friend who's sitting in a tree a distance h above level ground. At the instant you throw your ball, your friend drops another ball. (a) Show that the two balls will collide, no matter what your ball's initial speed, provided it's greater than some minimum value. (b) Find an expression for that minimum speed.
58. In a chase scene, a movie stuntman runs horizontally off the flat roof of one building and lands on another roof 1.9 m lower. If the gap between the buildings is 4.5 m wide, how fast must he run to cross the gap?
59. Standing on the ground 3.0 m from a building, you want to throw a package from your 1.5-m shoulder level to someone in a window 4.2 m above the ground. At what speed and angle should you throw the package so it just barely clears the windowsill?
60. Derive a general formula for the horizontal distance covered by a projectile launched horizontally at speed v_0 from height h.
61. Consider two projectiles launched on level ground with the same speed, at angles $45° \pm \alpha$. Show that the ratio of their flight times is $\tan(\alpha + 45°)$.
62. You toss a protein bar to your hiking companion located 8.6 m up a 39° slope, as shown in Fig. 3.23. Determine the initial velocity vector so the bar reaches your friend moving horizontally.

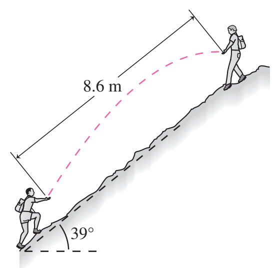


FIGURE 3.23 Problem 62

63. Prove that a projectile launched on level ground reaches maximum height midway along its trajectory.
64. A projectile launched at angle θ to the horizontal reaches maximum height h. Show that its horizontal range is $4h/\tan\theta$.
65. As an expert witness, you're testifying in a case involving a motorcycle accident. A motorcyclist driving in a 60-km/h zone hit a stopped car on a level road. The motorcyclist was thrown from his bike and landed 39 m down the road. You're asked whether he was speeding. What's your answer?
66. Show that, for a given initial speed, the horizontal range of a projectile is the same for launch angles $45° + \alpha$ and $45° - \alpha$.
67. A basketball player is 15 ft horizontally from the center of the basket, which is 10 ft off the ground. At what angle should the player aim the ball from a height of 8.2 ft with a speed of 26 ft/s?
68. Two projectiles are launched simultaneously from the same point, with different launch speeds and angles. Show that no combination of speeds and angles will permit them to land simultaneously and at the same point.
69. A jet is diving vertically downward at 1200 km/h. If the pilot can withstand a maximum acceleration of $5g$ (i.e., 5 times Earth's gravitational acceleration) before losing consciousness, at what height must the plane start a quarter turn to pull out of the dive? Assume the speed remains constant.
70. Your alpine rescue team is using a slingshot to send an emergency medical packet to climbers stranded on a ledge, as shown in Fig. 3.24; your job is to calculate the launch speed. What do you report?

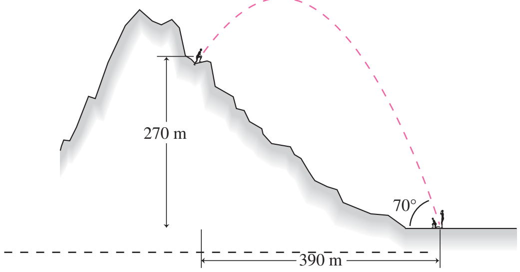


FIGURE 3.24 Problem 70

71. If you can throw a stone straight up to height h, what's the maximum horizontal distance you could throw it over level ground?
72. In a conversion from military to peacetime use, a missile with maximum horizontal range 180 km is being adapted for studying Earth's upper atmosphere. What is the maximum altitude it can achieve if launched vertically?
73. A soccer player can kick the ball 28 m on level ground, with its initial velocity at 40° to the horizontal. At the same initial speed and angle to the horizontal, what horizontal distance can the player kick the ball on a 15° upward slope?
74. A diver leaves a 3-m board on a trajectory that takes her 2.5 m above the board and then into the water 2.8 m horizontally from the end of the board. At what speed and angle did she leave the board?
75. Using calculus, you can find a function's maximum or minimum by differentiating and setting the result to zero. Do this for Equation 3.15, differentiating with respect to θ, and thus verify that the maximum range occurs for $\theta = 45°$.
76. You're a consulting engineer specializing in athletic facilities, and you've been asked to help design the Olympic ski jump pictured in Fig. 3.25. Skiers will leave the jump at 28 m/s and 9.5° below the horizontal, and land 55 m horizontally from the end of the jump. Your job is to specify the slope of the ground so skiers' trajectories make an angle of only 3.0° with the ground on landing, ensuring their safety. What slope do you specify?

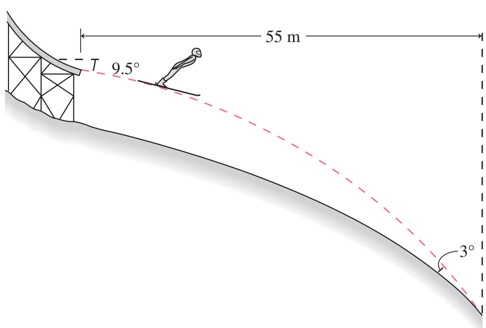


FIGURE 3.25 Problem 76

77. Differentiate the trajectory Equation 3.14 to find its slope, $\tan\theta = dy/dx$, and show that the slope is in the direction of the projectile's velocity, as given by Equations 3.10 and 3.11.
78. Your medieval history class is constructing a trebuchet, a catapult-like weapon for hurling stones at enemy castles. The plan is to launch stones off a 75-m-high cliff, with initial speed 36 m/s. Some members of the class think a 45° launch angle will give the maximum range, but others claim the cliff height makes a difference. What do you give for the angle that will maximize the range?
79. Generalize Problem 78 to find an expression for the angle that will maximize the range of a projectile launched with speed v_0 from height h above level ground.
80. (a) Show that the position of a particle on a circle of radius R with its center at the origin is $\vec{r} = R(\cos\theta\hat{\imath} + \sin\theta\hat{\jmath})$, where θ is the angle the position vector makes with the x-axis. (b) If the particle moves with constant speed v starting on the x-axis at $t = 0$, find an expression for θ in terms of time t and the period T to complete a full circle. (c) Differentiate the position vector twice with respect to time to find the acceleration, and show that its magnitude is given by Equation 3.16 and its direction is toward the center of the circle.

Passage Problems

Alice (A), Bob (B), and Carrie (C) all start from their dorm and head for the library for an evening study session. Alice takes a straight path, while the paths Bob and Carrie follow are portions of circular arcs, as shown in Fig. 3.26. Each student walks at a constant speed. All three leave the dorm at the same time, and they arrive simultaneously at the library.

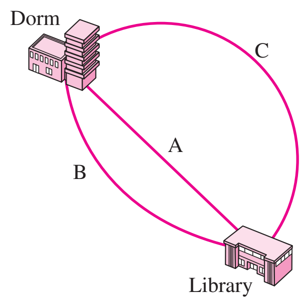


FIGURE 3.26 Passage Problems 81–84

81. Which statement characterizes the distances the students travel?
   a. They're equal.
   b. C > A > B
   c. C > B > A
   d. B > C > A
82. Which statement characterizes the students' displacements?
   a. They're equal.
   b. C > A > B
   c. C > B > A
   d. B > C > A
83. Which statement characterizes their average speeds?
   a. They're equal.
   b. C > A > B
   c. C > B > A
   d. B > C > A
84. Which statement characterizes their accelerations while walking (not starting and stopping)?
   a. They're equal.
   b. None accelerates.
   c. A > B > C
   d. C > B > A
   e. B > C > A
   f. There's not enough information to decide.

Answers to Chapter Questions

Answer to Chapter Opening Question

Assuming negligible air resistance, the penguin should leave the water at a 45° angle.

Answers to GOT IT? Questions

3.1. (c).
3.2. (d) only.
3.3. (c) gives the greatest change in speed; (b) gives the greatest change in direction.

4 Force and Motion

What forces govern the motion of the sailboard?

New Concepts, New Skills

By the end of this chapter you should be able to

- Explain the concept of force and its role in causing *change* in motion (4.1).
- Describe the fundamental forces of physics (4.3).
- State Newton's three laws of motion (4.2, 4.6).
- Describe the force of gravity and the distinction between mass and weight (4.4).
- Apply Newton's laws to one-dimensional motion (4.5, 4.6).

Connecting Your Knowledge

- Newton's laws relate force and acceleration. Therefore you need a solid understanding of acceleration, here based on the one-dimensional analysis of Chapter 2 (2.3).

An interplanetary spacecraft moves effortlessly, yet its engines shut down years ago. Why does it keep moving? A baseball heads toward the batter. The batter swings, and suddenly the ball is heading toward left field. Why did its motion change?

Questions about the "why" of motion are the subject of **dynamics**. Here we develop the basic laws that answer those questions. Isaac Newton first stated these laws more than 300 years ago, yet they remain a vital part of physics and engineering today, helping us guide spacecraft to distant planets, develop better cars, and manipulate the components of individual cells.

4.1 The Wrong Question

We began this chapter with two questions: one about why a spacecraft *moved* and the other about why a baseball's motion *changed*. For nearly 2000 years following the work of Aristotle (384–322 BCE), the first question—Why do things move?—was the crucial one. And the answer seemed obvious: It took a force—a push or a pull—to keep something moving. This idea makes sense: Stop exerting yourself when jogging, and you stop moving; take your foot off the gas pedal, and your car soon stops. Everyday experience seems to suggest that Aristotle was right, and most of us carry in our heads the Aristotelian idea that motion requires a cause—something that pushes or pulls on a moving object to keep it going.

Actually, "What keeps things moving?" is the wrong question. In the early 1600s, Galileo Galilei did experiments that convinced him that a moving object has an intrinsic "quantity of motion" and needs no push to keep it moving (Fig. 4.1). Instead of answering "What keeps things moving?," Galileo declared that the question needs no answer. In so doing, he set the stage for centuries of progress in physics, beginning with the achievements of Issac Newton and culminating in the work of Albert Einstein.

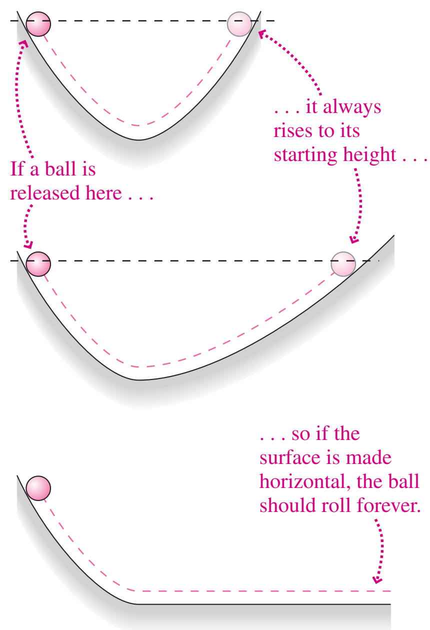


FIGURE 4.1 Galileo considered balls rolling on inclines and concluded that a ball on a horizontal surface should roll forever.

The Right Question

Our first question—about why the spacecraft keeps moving—is the wrong question. So what's the right question? It's the second one, about why the baseball's motion *changed*. Dynamics isn't about what causes motion itself; it's about what causes *changes* in motion. Changes include starting and stopping, speeding up and slowing down, and changing direction. Any *change* in motion begs an explanation, but motion itself does not. Get used to this important idea and you'll have a much easier time with physics. But if you remain a "closet Aristotelian," secretly looking for causes of motion, you'll find it difficult to understand and apply the simple laws that actually govern motion.

Galileo identified the right question about motion. But it was Isaac Newton who formulated the quantitative laws describing how motion changes. We use those laws today for everything from designing antilock braking systems, to building skyscrapers, to guiding spacecraft.

4.2 Newton's First and Second Laws

What caused the baseball's motion to change? Obviously, it was the bat's push. The term **force** descibes a push or a pull. And the essence of dynamics is simply this:

> Force causes change in motion.

We'll soon quantify this idea, writing equations and solving numerical problems. But the essential point is in the simple sentence above. If you want to change an object's motion, you need to apply a force. If you see an object's motion change, you know there's a force acting. Contrary to Aristotle, and probably to your own intuitive sense, it does *not* take a force to keep something in unchanging motion; force is needed *only* to *change* an object's motion.

The Net Force

You can push a ball left or right, up or down. Your car's tires can push the car forward or backward, or make it round a curve. Force has direction and is a vector quantity. Furthermore, more than one force can act on an object. We call the individual forces on an object **interaction forces** because they always involve other objects interacting with the object in question. In Fig. 4.2*a*, for example, the interaction forces are exerted by the people pushing the car. In Fig. 4.2*b* the interaction forces include the force of air on the plane, the engine force from the hot exhaust gases, and Earth's gravitational force.

We now explore in more detail the relation between force and change in motion. Experiment shows that what matters is the **net force**, meaning the vector sum of all individual interaction forces acting on an object. If the net force on an object isn't zero, then the object's motion must be changing—in direction or speed or both (Fig. 4.2*a*). If the net force on an object is zero—no matter what individual interaction forces contribute to the net force—then the object's motion is unchanging (Fig. 4.2*b*).

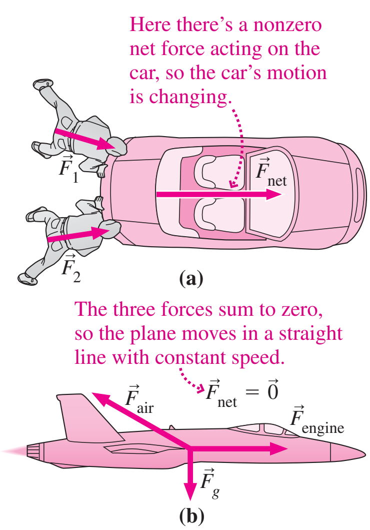


FIGURE 4.2 The net force determines the change in an object's motion.

Newton's First Law

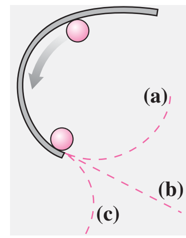
The basic idea that force causes change in motion is the essence of **Newton's first law**:

> **Newton's first law of motion:** A body in uniform motion remains in uniform motion, and a body at rest remains at rest, unless acted on by a nonzero net force.

The word "uniform" here is essential; **uniform motion** means unchanging motion—that is, motion in a straight line at constant speed. The phrase "a body at rest" isn't really necessary because rest is just the special case of uniform motion with zero speed, but we include it for consistency with Newton's original statement.

The first law says that uniform motion is a perfectly natural state, requiring no explanation. Again, the word "uniform" is crucial. The first law does *not* say that an object moving in a circle will continue to do so without a nonzero net force; in fact, it says that an object moving in a circle—or in any other curved path—*must* be subject to a nonzero net force because its motion is changing.

GOT IT? 4.1 On a horizontal tabletop is a curved barrier that exerts a force on a ball, guiding its motion in a circular path as shown. After the ball leaves the barrier, which of the dashed paths shown does it follow?



Newton's first law is simplicity itself, but it's counter to our Aristotelian preconceptions; after all, your car soon stops when you take your foot off the gas. But because the motion changes, that just means—as the first law says—that there must be a nonzero net force acting. That force is often a "hidden" one, like friction, that isn't as obvious as the push or pull of muscle. Go to an ice show or hockey game, where frictional forces are minimal, and the first law becomes a lot clearer.

Newton's Second Law

Newton's second law quantifies the relation between force and change in motion. Newton reasoned that the product of mass and velocity was the best measure of an object's "quantity of motion." The modern term is **momentum**, and we write

$$\vec{p} = m\vec{v} \qquad \text{(momentum)} \tag{4.1}$$

for the momentum of an object with mass m and velocity $\vec{v}$. As the product of a scalar (mass) and a vector (velocity), momentum is itself a vector quantity. Newton's second law relates the rate of change of an object's momentum to the net force acting on that object:

Newton's second law of motion: The rate at which a body's momentum changes is equal to the net force acting on the body:

$$\vec{F}_{\text{net}} = \frac{d\vec{p}}{dt} \qquad \text{(Newton's 2}^{\text{nd}}\text{ law)} \tag{4.2}$$

When a body's mass remains constant, we can use the definition of momentum, $\vec{p} = m\vec{v}$, to write

$$\vec{F}_{\text{net}} = \frac{d\vec{p}}{dt} = \frac{d(m\vec{v})}{dt} = m\frac{d\vec{v}}{dt}$$

But $d\vec{v}/dt$ is the acceleration $\vec{a}$, so

$$\vec{F}_{\text{net}} = m\vec{a} \qquad \text{(Newton's 2}^{\text{nd}}\text{ law, constant mass)} \tag{4.3}$$

Although Newton originally wrote his second law in the form 4.2, which remains the most general form, the form 4.3 is more widely recognized.

Newton's second law includes the first law as the special case $\vec{F}_{net} = \vec{0}$. In this case Equation 4.3 gives $\vec{a} = \vec{0}$, so an object's velocity doesn't change.

✓**TIP** Understanding Newton

To apply Newton's law successfully, you have to understand the terms in Equation 4.3. On the left is the net force $\vec{F}_{net}$—the vector sum of all real, physical interaction forces acting on an object. On the right is $m\vec{a}$—not a force but the product of the object's mass and acceleration. The equal sign says that they have the same value, not that they're the same thing. So don't go adding an extra force $m\vec{a}$ when you're applying Newton's second law.

Mass, Inertia, and Force

Because it takes force to change an object's motion, the first law says that objects naturally resist changes in motion. The term **inertia** describes this resistance, and for that reason the first law is also called the **law of inertia**. Just as we describe a sluggish person as having a lot of inertia, so an object that is hard to start moving—or hard to stop once started—has a lot of inertia. If we solve the second law for the acceleration $\vec{a}$, we find that $\vec{a} = \vec{F}/m$—showing that a given force is *less* effective in changing the motion of a *more massive* object (Fig. 4.3). The mass m that appears in Newton's laws is thus a measure of an object's inertia and determines the object's response to a given force.

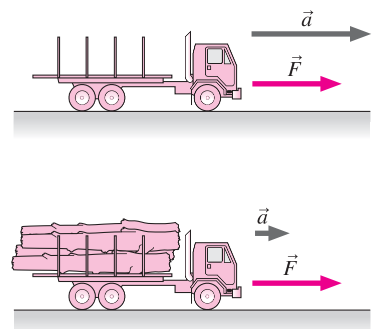


FIGURE 4.3 The loaded truck has greater mass—more inertia—so its acceleration is smaller when the same force is applied.

By comparing the acceleration of a known and an unknown mass, we can determine the unknown mass. From Newton's second law for a force of magnitude F,

$$F = m_{known}a_{known} \quad \text{and} \quad F = m_{unknown}a_{unknown}$$

where we're interested only in magnitudes so we don't use vectors. Equating these two expressions for the same force, we get

$$\frac{m_{unknown}}{m_{known}} = \frac{a_{known}}{a_{unknown}} \tag{4.4}$$

Equation 4.4 is an operational definition of mass; it shows how, given a known mass and a fixed force, we can determine other masses.

The force required to accelerate a 1-kg mass at the rate of 1 m/s^2 is defined to be 1 **newton** (N). Equation 4.3 shows that 1 N is equivalent to 1 kg · m/s^2. Other common force units are the English pound (lb, equal to 4.448 N) and the dyne, a metric unit equal to 10^{-5} N. A 1-N force is rather small; you can readily exert forces measuring hundreds of newtons with your own body.

EXAMPLE 4.1 Force from Newton: A Car Accelerates

A 1200-kg car accelerates from rest to 20 m/s in 7.8 s, moving in a straight line with constant acceleration. (a) Find the net force acting on the car. (b) If the car then rounds a bend 85 m in radius at a steady 20 m/s, what net force acts on it?

INTERPRET In this problem we're asked to evaluate the net force on a car (a) when it undergoes constant acceleration and (b) when it rounds a turn. In both cases the net force is entirely horizontal, so we need to consider only the horizontal component of Newton's law.

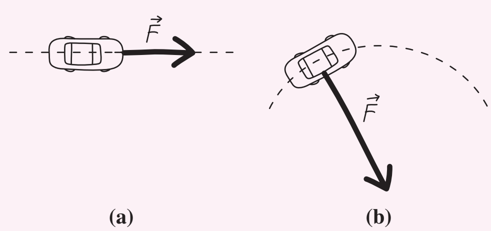


FIGURE 4.4 Our sketch of the net force on the car of Example 4.1.

DEVELOP Figure 4.4 shows the horizontal force acting on the car in each case; since this is the net force, it's equal to the car's mass multiplied by its acceleration. We aren't actually given the acceleration in this problem, but for (a) we know the change in speed and the time involved, so we can write $a = \Delta v/\Delta t$. For (b) we're given the speed and the radius of the turn; since the car is in uniform circular motion, Equation 3.16 applies, and we have $a = v^2/r$.

(continued)

EVALUATE We solve for the unknown acceleration and evaluate the numerical answers for both cases:

$$\text{(a)} \qquad F_{\text{net}} = ma = m\frac{\Delta v}{\Delta t} = (1200\text{ kg})\left(\frac{20\text{ m/s}}{7.8\text{ s}}\right) = 3.1\text{ kN}$$

$$\text{(b)} \qquad F_{\text{net}} = ma = m\frac{v^2}{r} = (1200\text{ kg})\frac{(20\text{ m/s})^2}{85\text{ m}} = 5.6\text{ kN}$$

ASSESS First, the units worked out; they were actually $\text{kg} \cdot \text{m/s}^2$, but that defines the newton. The answers came out in thousands of N, but we moved the decimal point three places and changed to kN for convenience. And the numbers seem to make sense; we mentioned that 1 newton is a rather small force, so it's not surprising to find forces on cars measured in kilonewtons.

Note that Newton's law doesn't distinguish between forces that change an object's speed, as in (a), and forces that change its direction, as in (b). Newton's law relates force, mass, and acceleration in *all* cases.

GOT IT? 4.2 A nonzero net force acts on an object. Does that mean the object necessarily moves in the same direction as the net force?

Inertial Reference Frames

Why don't flight attendants serve beverages when an airplane is accelerating down the runway? For one thing, their beverage cart wouldn't stay put, but would accelerate toward the back of the plane even in the absence of a net force. So is Newton's first law wrong? No, but Newton's laws don't apply in an accelerating airplane. With respect to the ground, in fact, the beverage cart is doing just what Newton says it should: It remains in its original state of motion, while all around it plane and passengers accelerate toward takeoff.

In Section 3.3 we defined a reference frame as a system against which we measure velocities; more generally, a reference frame is the "background" in which we study physical reality. Our airplane example shows that Newton's laws don't work in all reference frames; in particular, they're not valid in accelerating frames. Where they are valid is in reference frames undergoing uniform motion—called **inertial reference frames** because only in these frames does the law of inertia hold. In a noninertial frame like an accelerating airplane, a car rounding a curve, or a whirling merry-go-round, an object at rest doesn't remain at rest, even when no force is acting. A good test for an inertial frame is to check whether Newton's first law is obeyed—that is, whether an object at rest remains at rest, and an object in uniform motion remains in uniform motion, when no force is acting on it.

Strictly speaking, our rotating Earth is not an inertial frame, and therefore Newton's laws aren't exactly valid on Earth. But Earth's rotation has an insignificant effect on most motions of interest, so we can usually treat Earth as an inertial reference frame. An important exception is the motion of oceans and atmosphere; here, scientists must take Earth's rotation into account.

If Earth isn't an inertial frame, what is? That's a surprisingly subtle question, and it pointed Einstein toward his general theory of relativity. The law of inertia is intimately related to questions of space, time, and gravity—questions whose answers lie in Einstein's theory. We'll look briefly at that theory in Chapter 33.

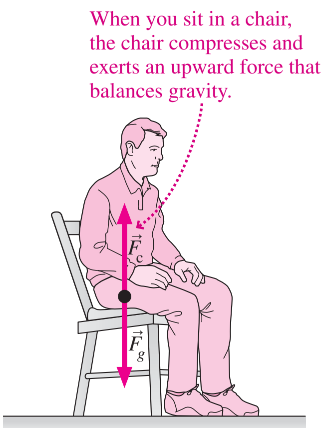


FIGURE 4.5 A compression force.

4.3 Forces

The most familiar forces are pushes and pulls you apply yourself, but passive objects can apply forces, too. A car collides with a parked truck and comes to a stop. Why? Because the truck exerts a force on it. The Moon circles Earth rather than moving in a straight line. Why? Because Earth exerts a gravitational force on it. You sit in a chair and don't fall to the floor. Why not? Because the chair exerts an upward force on you, countering gravity.

Some forces, like those you apply with your muscles, can have values that you choose. Other forces take on values determined by the situation. When you sit in the chair shown in Fig. 4.5, the downward force of gravity on you causes the chair to compress slightly. The chair acts like a spring and exerts an upward force. When the chair compresses enough that the upward force is equal in magnitude to the downward force of gravity, there's no net force and you sit without accelerating. The same thing happens with **tension forces** when objects are suspended from ropes or cables—the ropes stretch until the force they exert balances the force of gravity (Fig. 4.6).

Forces like the pull you exert on your rolling luggage, the force of a chair on your body, and the force a baseball exerts on a bat are **contact forces** because the force is exerted through direct contact. Other forces, like gravity and electric and magnetic forces, are **action-at-a-distance forces** because they seemingly act between distant objects, like Earth and the Moon. Actually, the distinction isn't clear-cut; at the microscopic level, contact forces involve action-at-a-distance electric forces between molecules. And the action-at-a-distance concept itself is troubling. How can Earth "reach out" across empty space and pull on the Moon? Later we'll look at an approach to forces that avoids this quandary.

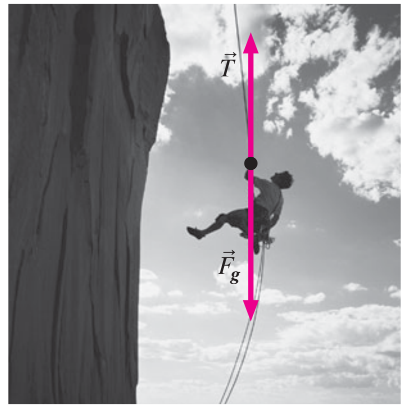


FIGURE 4.6 The climbing rope exerts an upward tension force $\vec{T}$ that balances the force of gravity.

The Fundamental Forces

Gravity, tension forces, compression forces, contact forces, electric forces, friction forces—how many kinds of forces are there? At present, physicists identify three basic forces: the gravitational force, the electroweak force, and the strong force.

Gravity is the weakest of the fundamental forces, but because it acts attractively between all matter, gravity's effect is cumulative. That makes gravity the dominant force in the large-scale universe, determining the structure of planets, stars, galaxies, and the universe itself.

The **electroweak force** subsumes **electromagnetism** and the **weak nuclear force**. Virtually all the nongravitational forces we encounter in everyday life are electromagnetic, including contact forces, friction, tension and compression forces, and the forces that bind atoms into chemical compounds. The weak nuclear force is less obvious, but it's crucial in the Sun's energy production—providing the energy that powers life on Earth.

The **strong force** describes how particles called **quarks** bind together to form protons, neutrons, and a host of less-familiar particles. The force that joins protons and neutrons to make atomic nuclei is a residue of the strong force between their constituent quarks. Although the strong force isn't obvious in everyday life, it's ultimately responsible for the structure of matter. If its strength were slightly different, atoms more complex than helium couldn't exist, and the universe would be devoid of life!

Unifying the fundamental forces is a major goal of physics. Over the centuries we've come to understand seemingly disparate forces as manifestations of a more fundamental underlying force. Figure 4.7 suggests that the process continues, as physicists attempt first to unify the strong and electroweak forces, and then ultimately to add gravity to give a "Theory of Everything."

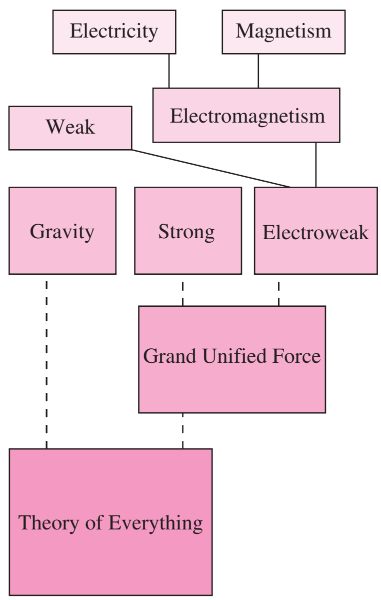


FIGURE 4.7 Unification of forces is a major theme in physics.

4.4 The Force of Gravity

Newton's second law shows that mass is a measure of a body's resistance to changes in motion—its inertia. A body's mass is an intrinsic property; it doesn't depend on location. If my mass is 65 kg, it's 65 kg on Earth, in an orbiting spacecraft, or on the Moon. That means no matter where I am, a force of 65 N gives me an acceleration of 1 m/s^2.

We commonly use the term "weight" to mean the same thing as mass. In physics, though, **weight** is the *force* that gravity exerts on a body. Near Earth's surface, a freely falling body accelerates downward at 9.8 m/s^2; we designate this acceleration vector by $\vec{g}$. Newton's second law, $\vec{F} = m\vec{a}$, then says that the force of gravity on a body of mass m is $m\vec{g}$; this force is the body's weight:

$$\vec{w} = m\vec{g} \qquad \text{(weight)} \tag{4.5}$$

With my 65-kg mass, my weight near Earth's surface is then $(65\ \text{kg})(9.8\ \text{m/s}^2)$ or 640 N. On the Moon, where the acceleration of gravity is only 1.6 m/s^2, I would weigh only 100 N. And in the remote reaches of intergalactic space, far from any gravitating object, my weight would be essentially zero.

EXAMPLE 4.2 Mass and Weight: Exploring Mars

The Phoenix spacecraft that landed on Mars in 2008 weighed 3.43 kN on Earth. What were its mass and weight on Mars?

INTERPRET Here we're asked about the relation between mass and weight, and the object we're interested in is the Phoenix spacecraft.

DEVELOP Equation 4.5 describes the relation between mass and weight. Writing this equation in scalar form because we're interested only in magnitudes, we have $w = mg$.

EVALUATE First we want to find mass from weight, so we solve for m using the Earth weight and Earth's gravity:

$$m = \frac{w}{g} = \frac{3.43 \text{ kN}}{9.8 \text{ m/s}^2} = 350 \text{ kg}$$

This mass is the same everywhere, so the weight on Mars is given by $w = mg_{\text{Mars}} = (350 \text{ kg})(3.74 \text{ m/s}^2) = 1.31 \text{ kN}$. Here we found the acceleration of gravity on Mars in Appendix E.

ASSESS Make sense? Sure, Mars's gravitational acceleration is lower than Earth's, and so is the spacecraft's weight on Mars. ■

One reason we confuse mass and weight is the common use of the SI unit kilogram to describe "weight." At the doctor's office you may be told that you "weigh" 55 kg. You don't; you have a mass of 55 kg, so your weight is $(55 \text{ kg})(9.8 \text{ m/s}^2)$ or 540 N. The unit of force in the English system is the pound, so giving your weight in pounds is correct.

That we confuse mass and weight at all results from the remarkable fact that the gravitational acceleration of all objects is the same. This makes a body's *weight*, a gravitational property, proportional to its *mass*, a measure of its inertia in terms that have nothing to do with gravity. First inferred by Galileo from his experiments with falling bodies, this relation between gravitation and inertia seemed a coincidence until the early 20th century. Finally Albert Einstein showed how that simple relation reflects the underlying geometry of space and time in a way that intimately links gravitation and acceleration.

Weightlessness

Aren't astronauts "weightless"? Not according to our definition. At the altitude of the International Space Station, the acceleration of gravity has about 89% of its value at Earth's surface, so the gravitational forces $m\vec{g}$ on the station and its occupants are almost as large as on Earth. But the astronauts *seem* weightless, and indeed they *feel* weightless (Fig. 4.8). What's going on?

Imagine yourself in an elevator whose cable has broken and is dropping freely downward with the gravitational acceleration g. In other words, the elevator and its occupant are in **free fall**, with only the force of gravity acting. If you let go of a book, it too falls freely with acceleration g. But so does everything else around it—and therefore the book

FIGURE 4.8 These astronauts only *seem* weightless.

stays put relative to you (Fig. 4.9*a*). To you, the book seems "weightless," since it doesn't seem to fall when you let go of it. And you're "weightless" too; if you jump off the elevator's floor, you float to the ceiling rather than falling back. Of course you, the book, and the elevator are *all* falling, but because all have the same acceleration that isn't obvious to you. The gravitational force is still acting; it's making you fall. So you really do have weight, and your condition is best termed **apparent weightlessness**.

A falling elevator is a dangerous place; your state of apparent weightlessness would end with a deadly smash caused by nongravitational contact forces when you hit the ground. But apparent weightlessness occurs permanently in a state of free fall that doesn't intersect Earth—as in an orbiting spacecraft (Fig. 4.9*b*). It's not being in outer space that makes astronauts seem weightless; it's that they, like our hapless elevator occupant, are in free fall—moving under the influence of the gravitational force alone. The condition of apparent weightlessness in orbiting spacecraft is sometimes called "microgravity."

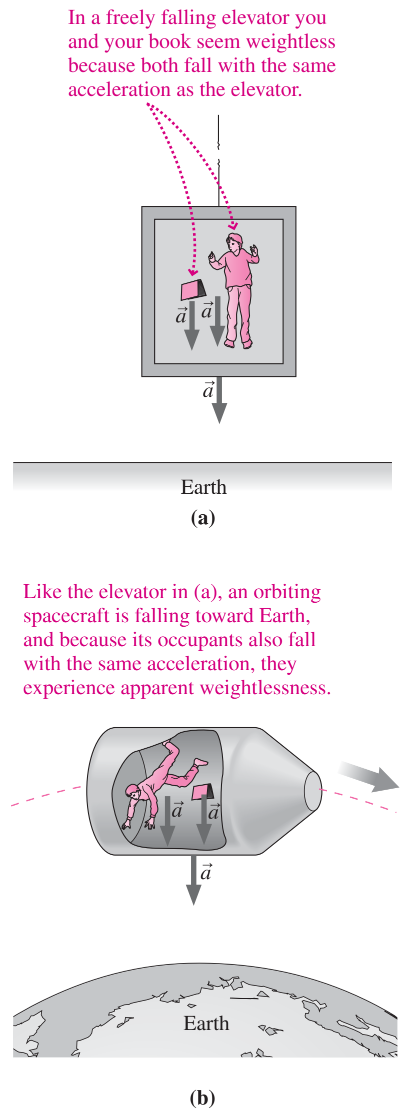


FIGURE 4.9 Objects in free fall appear weightless because they all experience the same acceleration.

GOT IT? 4.3 A popular children's book explains the weightlessness astronauts experience by saying there's no gravity in space. If there were no gravity in space, what would be the motion of a space shuttle, a satellite, or, for that matter, the Moon?

4.5 Using Newton's Second Law

The interesting problems involving Newton's second law are those where more than one force acts on an object. To apply the second law, we then need the net force. For an object of constant mass, the second law relates the net force and the acceleration:

$$\vec{F}_{\text{net}} = m\vec{a}$$

Using Newton's second law with multiple forces is easier if we draw a **free-body diagram**, a simple diagram that shows only the object of interest and the forces acting on it.

TACTICS 4.1 Drawing a Free-Body Diagram

Drawing a free-body diagram, which shows the forces acting on an object, is the key to solving problems with Newton's laws. To make a free-body diagram:

1. Identify the object of interest and all the forces acting on it.
2. Represent the object as a dot.
3. Draw the vectors for *only* those forces acting *on* the object, with their tails all starting on the dot.

Figure 4.10 shows two examples where we reduce physical scenarios to free-body diagrams. We often add a coordinate system to the free-body diagram so that we can express force vectors in components.

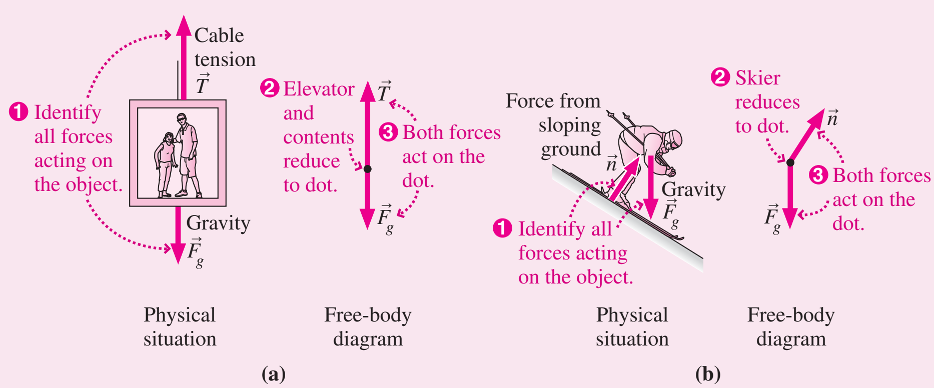


FIGURE 4.10 Free-body diagrams. (a) A one-dimensional situation like those we discuss in this chapter. (b) A two-dimensional situation. We'll deal with such cases in Chapter 5.

Our IDEA strategy applies to Newton's laws as it does to other physics problems. For the second law, we can elaborate on the four IDEA steps:

PROBLEM-SOLVING STRATEGY 4.1 Newton's Second Law

INTERPRET Interpret the problem to be sure that you know what it's asking and that Newton's second law is the relevant concept. Identify the object of interest and all the individual interaction forces acting on it.

DEVELOP Draw a free-body diagram as described in Tactics 4.1. Develop your solution plan by writing Newton's second law, $\vec{F}_{\text{net}} = m\vec{a}$, with $\vec{F}_{\text{net}}$ expressed as the sum of the forces you've identified. Then choose a coordinate system so you can express Newton's law in components.

EVALUATE At this point the physics is done, and you're ready to execute your plan by solving Newton's second law and evaluating the numerical answer(s), if called for. Even in the one-dimensional problems of this chapter, remember that Newton's law is a vector equation; that will help you get the signs right. You need to write the components of Newton's law in the coordinate system you chose, and then solve the resulting equation(s) for the quantity(ies) of interest.

ASSESS Assess your solution to see that it makes sense. Are the numbers reasonable? Do the units work out correctly? What happens in special cases—for example, when a mass, force, or acceleration becomes very small or very large, or an angle becomes 0° or 90°?

EXAMPLE 4.3 Newton's Second Law: In the Elevator

A 740-kg elevator accelerates upward at 1.1 m/s^2, pulled by a cable of negligible mass. Find the tension force in the cable.

INTERPRET In this problem we're asked to evaluate one of the forces on an object. First we identify the object of interest. Although the problem asks about the cable tension, it's the elevator on which that tension acts, so the elevator is the object of interest. Next, we identify the forces acting on the elevator. There are two: the downward force of gravity $\vec{F}_g$ and the upward cable tension $\vec{T}$.

DEVELOP Figure 4.11*a* shows the elevator accelerating upward; Fig. 4.11*b* is a free-body diagram representing the elevator as a dot with the two force vectors acting on it. The applicable equation is Newton's second law, $\vec{F}_{\text{net}} = m\vec{a}$, with $\vec{F}_{\text{net}}$ given by the sum of the forces we've identified:

$$\vec{F}_{\text{net}} = \vec{T} + \vec{F}_g = m\vec{a} \tag{4.6}$$

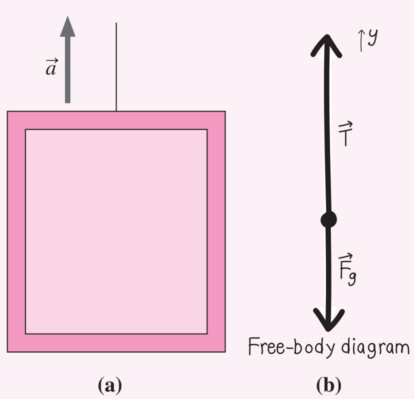


FIGURE 4.11 The forces on the elevator are the cable tension $\vec{T}$ and gravity $\vec{F}_g$.

✓TIP Vectors Tell It All

Don't be tempted to put a minus sign in this equation because one force is downward. You don't have to worry about signs until you write the components of a vector equation in the coordinate system you chose.

Now we need to choose a coordinate system. Here all the forces are vertical, so we choose our *y*-axis pointing upward.

EVALUATE Now we're ready to rewrite Newton's second law—Equation 4.6 in this case—in our coordinate system. Formally, we remove the vector signs and add coordinate subscripts—just *y* in this case:

$$T_y + F_{gy} = ma_y \tag{4.7}$$

Still no need to worry about signs. Now, what is T_y? Since the tension points upward and we've chosen that to be the positive direction, the component of tension in the *y*-direction is its magnitude T. What about F_{gy}? Gravity points downward, so this component is negative. Furthermore, we know that the magnitude of the gravitational force is mg. So $F_{gy} = -mg$. Then our Newton's law equation becomes

$$T - mg = ma_y$$

so

$$T = ma_y + mg = m(a_y + g) \tag{4.8}$$

For the numbers given, this equation yields

$$T = m(a_y + g) = (740\text{ kg})(1.1\text{ m/s}^2 + 9.8\text{ m/s}^2) = 8.1\text{ kN}$$

ASSESS We can see that this answer makes sense—and learn a lot more about physics—from the algebraic form of the answer in Equation 4.8. Consider some special cases: If the acceleration a_y were zero, then the net force on the elevator would have to be zero. In that case Equation 4.8 gives $T = mg$. Of course: The cable is then supporting the elevator's weight mg but not exerting any additional force to accelerate it.

On the other hand, if the elevator is accelerating upward, then the cable has to provide an extra force in addition to the weight; that's why the tension becomes $ma_y + mg$. Numerically, our answer of 8.1 kN is *greater* than the elevator's weight—and the cable had better be strong enough to handle the extra force.

Finally, if the elevator is accelerating downward, then a_y is negative, and the cable tension is *less* than the weight. In free fall, $a_y = -g$, and the cable tension would be zero.

You might have reasoned out this problem in your head. But we did it very thoroughly because the strategy we followed will let you solve all problems involving Newton's second law, even if they're much more complicated. If you always follow this strategy and don't try to find shortcuts, you'll become confident in using Newton's second law. ■

GOT IT? 4.4 For each of the following situations, would the cable tension in Example 4.3 be greater than, less than, or equal to the elevator's weight? (a) elevator starts moving upward, accelerating from rest; (b) elevator decelerates to a stop while moving upward; (c) elevator starts moving downward, accelerating from rest; (d) elevator slows to a stop while moving downward; (e) elevator is moving upward with constant speed

CONCEPTUAL EXAMPLE 4.1 At the Equator

When you stand on a scale, the scale pushes up to support you, and the scale reading shows the force with which it's pushing. If you stand on a scale at Earth's equator, is the reading greater or less than your weight?

EVALUATE The question asks about the force the scale exerts on you, in comparison to your weight (the gravitational force on you). Figure 4.12 is our sketch, showing the scale force upward and the gravitational force downward, toward Earth's center. You're in circular motion about Earth's center, so the direction of your acceleration is toward the center (downward). According to Newton's second law, the net force and acceleration are in the same direction. The only two forces acting on you are the downward force of gravity and the upward force of the scale. For them to sum to a net force that's downward, the force of gravity—your weight—must be larger. Therefore, the scale reading must be less than your weight.

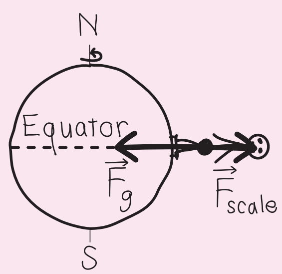


FIGURE 4.12 Our sketch for Conceptual Example 4.1.

ASSESS Make sense? Yes: If the two forces had equal magnitudes, the net force would be zero—inconsistent with the fact that you're accelerating. And if the scale force were greater, you'd be accelerating in the wrong direction! The same effect occurs everywhere except at the poles, but its analysis is more complicated because the acceleration is toward Earth's axis, not the center.

MAKING THE CONNECTION By what percentage is your apparent weight (the scale reading) at the equator less than your actual weight?

EVALUATE Using Earth's radius R_E from Appendix E, and its 24-hour rotation period, you can find your acceleration: From Equation 3.16, it's v^2/R_E. Following Problem-Solving Strategy 4.1 and working in a coordinate system with the vertical direction upward, you'll find that Newton's second law becomes $F_{scale} - mg = -m\dfrac{v^2}{R_E}$, or $F_{scale} = mg - mv^2/R_E$. So the scale reading differs from your weight mg by mv^2/R_E. Working the numbers shows that's a difference of only 0.34%. Note that this result doesn't depend on your mass m.

4.6 Newton's Third Law

Push your book across your desk, and you feel the book push back (Fig. 4.13*a*). Kick a ball with bare feet, and your toes hurt. Why? You exert a force on the ball, and the ball exerts a force back on you. A rocket engine exerts forces that expel hot gases out of its nozzle—and the hot gases exert a force on the rocket, accelerating it forward (Fig. 4.13*b*).

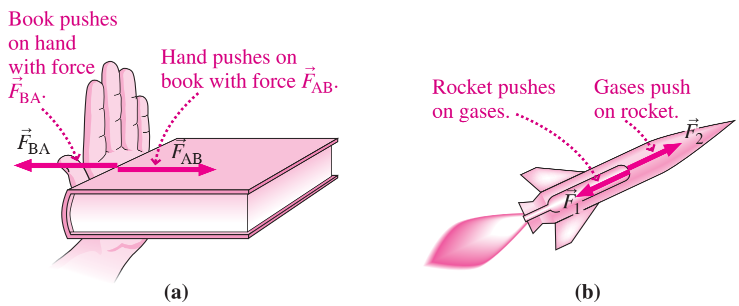


FIGURE 4.13 Newton's third law says that forces always come in pairs. With objects in contact, both forces act at the contact point. To emphasize that the two forces act on *different* objects, we draw them slightly displaced.

Whenever one object exerts a force on a second object, the second object also exerts a force on the first. The two forces are in opposite directions, but they have equal magnitudes. This fact constitutes **Newton's third law** of motion. The familiar expression "for every action there is an equal and opposite reaction" is Newton's 17th-century language. But there's really no distinction between "action" and "reaction"; both are always present. In modern language, the third law states:

Newton's third law of motion If object A exerts a force on object B, then object B exerts an oppositely directed force of equal magnitude on A.

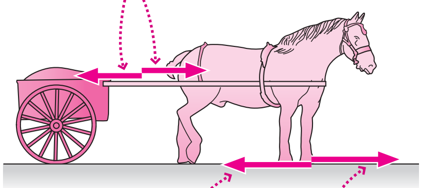


FIGURE 4.14 The horse-and-cart dilemma: The horse pulls on the cart, and the cart pulls back on the horse with a force of equal magnitude. So how can the pair ever get moving? No problem: The *net* force on the horse involves forces from *different* third-law pairs. Their magnitudes aren't equal and the horse experiences a net force in the forward direction.

Newton's third law is about forces between objects. It says that such forces always occur in pairs—that it's not possible for object A to exert a force on object B without B exerting a force back on A. You can now see why we coined the term "interaction forces"—when there's force between two objects, it's always a true *inter*action, with both objects exerting forces and both experiencing forces. We'll use the terms **interaction force pair** and **third-law pair** for the two forces described by Newton's third law.

It's crucial to recognize that the forces of a third-law pair act on *different* objects; the force $\vec{F}_{AB}$ of object A acts on object B, and the force $\vec{F}_{BA}$ of B acts on A. The forces have equal magnitudes and opposite directions, but they don't cancel to give zero net force *because they don't act on the same object*. In Fig. 4.13*a*, for example, $\vec{F}_{AB}$ is the force the hand exerts on the book. There's no other horizontal force acting on the book, so the net force on the book is nonzero and the book accelerates. Failure to recognize that the two forces of a third-law pair act on different objects leads to a contradiction, embodied in the famous horse-and-cart dilemma illustrated in Fig. 4.14.

APPLICATION Hollywood Goes Weightless

The film *Apollo 13* shows Tom Hanks and his fellow actors floating weightlessly around the cabin of their movie-set spacecraft. What special effects did Hollywood use here? None. The actors' apparent weightlessness was the real thing. But even Hollywood's budget wasn't enough to buy a space-shuttle flight. So the producers rented NASA's weightlessness training aircraft, aptly dubbed the "vomit comet." This airplane executes parabolic trajectories that mimic the free-fall motion of a projectile, so its occupants experience apparent weightlessness.

Movie critics marveled at how *Apollo 13* "simulated the weightlessness of outer space." Nonsense! The actors were in free fall just like the real astronauts on board the real Apollo 13, and they experienced exactly the same physical phenomenon—apparent weightlessness when moving under the influence of gravity alone.

EXAMPLE 4.4 Newton's Third Law: Pushing Books

On a frictionless horizontal surface, you push with force $\vec{F}$ on a book of mass m_1 that in turn pushes on a book of mass m_2 (Fig. 4.15*a*). What force does the second book exert on the first?

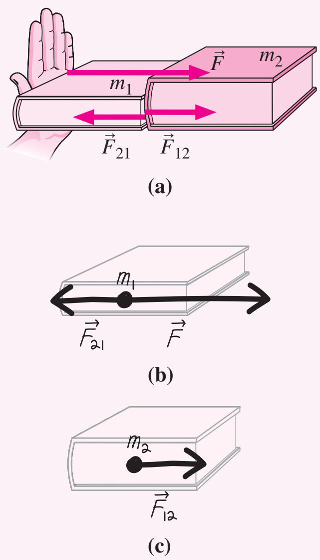


FIGURE 4.15 Horizontal forces on the books of Example 4.4. Not shown are the vertical forces of gravity and the normal force from the surface supporting the books.

INTERPRET This problem is about the interaction between two objects, so we identify both books as objects of interest.

DEVELOP In a problem with multiple objects, it's a good idea to draw a separate free-body diagram for each object. We've done that in Fig. 4.15*b* and *c*, keeping very light images of the books themselves. Now, we're asked about the force the second book exerts on the first. Newton's third law would give us that force if we knew the force the first book exerts on the second. Since that's the only horizontal force acting on book 2, we could get it from Newton's *second* law if we knew the acceleration of book 2. So here's our plan: (1) Find the acceleration of book 2; (2) use Newton's second law to find the net force on book 2, which in this case is the single force $\vec{F}_{12}$; and (3) apply Newton's third law to get $\vec{F}_{21}$, which is what we're looking for.

EVALUATE (1) The total mass of the two books is $m_1 + m_2$, and the net force applied to the combination is $\vec{F}$. Newton's second law, $\vec{F} = m\vec{a}$, gives

$$\vec{a} = \frac{\vec{F}}{m} = \frac{\vec{F}}{m_1 + m_2}$$

for the acceleration of both books, including book 2. (2) Now that we know book 2's acceleration, we use Newton's second law to find $\vec{F}_{12}$, which we recognize as the net force on book 2:

$$\vec{F}_{12} = m_2\vec{a} = m_2\frac{\vec{F}}{m_1 + m_2} = \frac{m_2}{m_1 + m_2}\vec{F}$$

(3) Finally, the forces the books exert on each other constitute a third-law pair, so we have

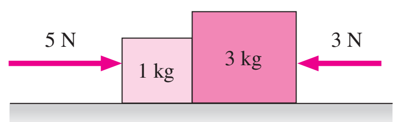
$$\vec{F}_{21} = -\vec{F}_{12} = -\frac{m_2}{m_1 + m_2}\vec{F}$$

ASSESS You can see that this result makes sense by considering the first book. It too undergoes acceleration $\vec{a} = \vec{F}/(m_1 + m_2)$, but there are *two* forces acting on it: the applied force $\vec{F}$ and the force $\vec{F}_{21}$ from the second book. So the net force on the first book is

$$\vec{F} + \vec{F}_{21} = \vec{F} - \frac{m_2}{m_1 + m_2}\vec{F} = \frac{m_1}{m_1 + m_2}\vec{F} = m_1\vec{a}$$

consistent with Newton's second law. Our result shows that Newton's second and third laws are both necessary for a fully consistent description of the motion.

GOT IT? 4.5 The figure shows two blocks with two forces acting on the pair. Is the net force on the larger block (a) greater than 2 N, (b) equal to 2 N, or (c) less than 2 N?



A contact force such as the force between the books in Example 4.4 is called a **normal force** (symbol $\vec{n}$) because it acts at right angles to the surfaces in contact. Other examples of normal forces are the upward force that a table or bridge exerts on objects it supports, and the force perpendicular to a sloping surface supporting an object (Fig. 4.16).

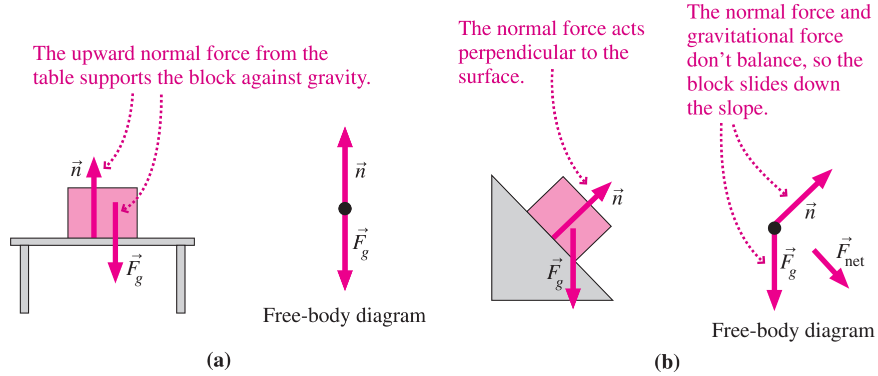


FIGURE 4.16 Normal forces. Also shown in each case is the gravitational force.

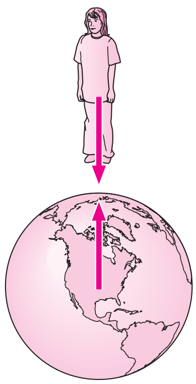


FIGURE 4.17 Gravitational forces on you and on Earth form a third-law pair. Figure is obviously not to scale!

Newton's third law also applies to forces like gravity that don't involve direct contact. Since Earth exerts a downward force on you, the third law says that you exert an equal upward force on Earth (Fig. 4.17). If you're in free fall, then Earth's gravity causes you to accelerate toward Earth. Earth, too, accelerates toward you—but it's so massive that this acceleration is negligible.

Measuring Force

Newton's third law provides a convenient way to measure forces using the tension or compression force in a spring. A spring stretches or compresses in proportion to the force exerted on it. By Newton's third law, the force *on* the spring is equal and opposite to the force the spring exerts on whatever is stretching or compressing it (Fig. 4.18). The spring's stretch or compression thus provides a measure of the force on whatever object is attached to the spring.

In an **ideal spring**, the stretch or compression is directly proportional to the force exerted by the spring. **Hooke's law** expresses this proportionality mathematically:

$$F_{sp} = -kx \qquad \text{(Hooke's law, ideal spring)} \tag{4.9}$$

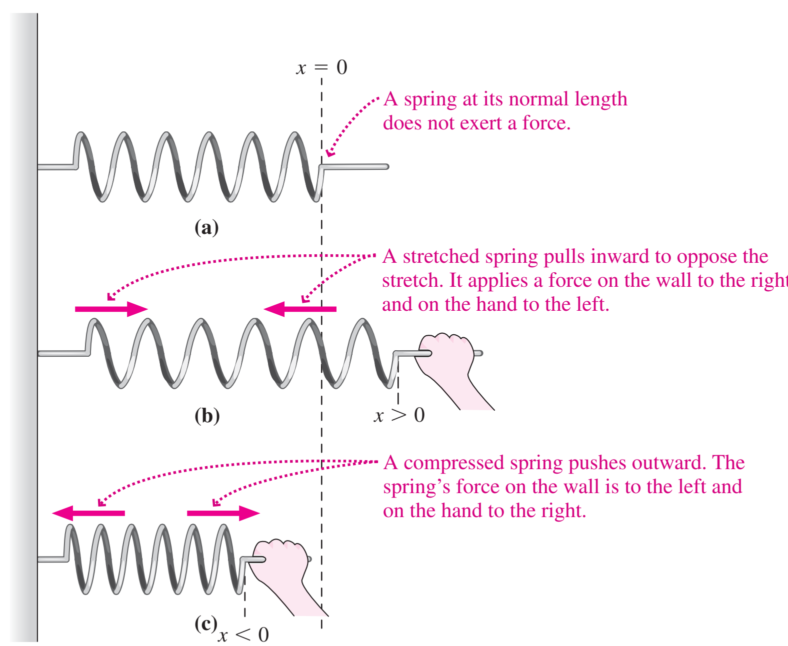


FIGURE 4.18 A spring responds to stretching or compression with an oppositely directed force.

Here F_{sp} is the spring force, x is the distance the spring has been stretched or compressed from its normal length, and k is the **spring constant**, which measures the "stiffness" of the spring. Its units are N/m. The minus sign shows that the spring force is *opposite* the distortion of the spring: Stretch it, and the spring responds with a force *opposite* the stretching force; compress it, and the spring pushes back against the compressing force. Real springs obey Hooke's law only up to a point; stretch it too much, and a spring will deform and eventually break.

A **spring scale** is a spring with an indicator and a scale calibrated in force units (Fig. 4.19). Common examples include many bathroom scales, hanging scales in supermarkets, and laboratory spring scales. Even electronic scales are spring scales, with their "springs" materials that produce electrical signals when deformed by an applied force.

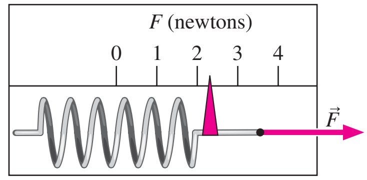


FIGURE 4.19 A spring scale.

Hang an object on a spring scale, and the spring stretches until its force counters the gravitational force on the object. Or, with a stand-on scale, the spring compresses until it supports you against gravity. Either way, the spring force is equal in magnitude to the weight mg, and thus the spring indicator provides a measure of weight. Given g, this procedure also provides the object's mass.

Be careful, though: A spring scale provides the true weight only if the scale isn't accelerating; otherwise, the scale reading is only an **apparent weight**. Weigh yourself in an accelerating elevator and you may be horrified or delighted, depending on the direction of the acceleration. Conceptual Example 4.1 made this point qualitatively, and Example 4.5 does so quantitatively.

EXAMPLE 4.5 True and Apparent Weight: A Helicopter Ride

A helicopter rises vertically, carrying concrete for a ski-lift foundation. A 35-kg bag of concrete sits in the helicopter on a spring scale whose spring constant is 3.4 kN/m. By how much does the spring compress (a) when the helicopter is at rest and (b) when it's accelerating upward at 1.9 m/s^2?

INTERPRET This problem is about concrete, a spring scale, and a helicopter. Ultimately, that means it's about mass, force, and acceleration—the content of Newton's laws. We're clearly interested in the spring and the concrete mass resting on it, which share the motion of the helicopter. We identify two forces acting on the concrete: gravity and the spring force $\vec{F}_{sp}$.

DEVELOP As with any Newton's law problem, we start with a free-body diagram (Fig. 4.20). We then write Newton's second law in its vector form

$$\vec{F}_{net} = \vec{F}_{sp} + \vec{F}_g = m\vec{a}$$

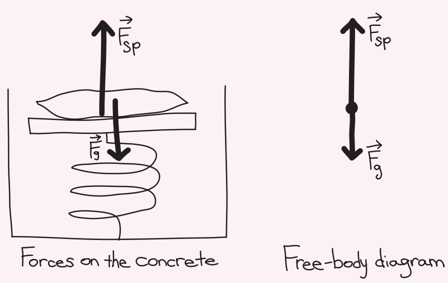


FIGURE 4.20 Our drawings for Example 4.5.

Vectors tell it all; don't worry about signs at this point. Our equation expresses all the physics of the situation, but before we can move on to the solution, we need to choose a coordinate system. Here it's convenient to take the y-axis vertically upward.

EVALUATE The forces are in the vertical direction, so we're concerned with only the y-component of Newton's law: $F_{spy} + F_{gy} = ma_y$. The spring force is upward and, from Hooke's law, it has magnitude kx, so $F_{spy} = kx$. Gravity is downward with magnitude mg, so $F_{gy} = -mg$. The y-component of Newton's law then becomes $kx - mg = ma_y$, which we solve to get

$$x = \frac{m(a_y + g)}{k}$$

Putting in the numbers (a) with the helicopter at rest $(a_y = 0)$ and (b) with $a_y = 1.9\ \text{m/s}^2$ gives

(a) $$x = \frac{m(a_y + g)}{k} = \frac{(35\ \text{kg})(0 + 9.8\ \text{m/s}^2)}{3400\ \text{N/m}} = 10\ \text{cm}$$

(b) $$x = \frac{(35\ \text{kg})(1.9\ \text{m/s}^2 + 9.8\ \text{m/s}^2)}{3400\ \text{N/m}} = 12\ \text{cm}$$

ASSESS Why is the answer to (b) larger? Because, just as with the cable in Example 4.3, the spring needs to provide an additional force to accelerate the concrete upward. ■

GOT IT? 4.6 (a) Would the answer to (a) in Example 4.5 change if the helicopter were not at rest but moving upward at constant speed? (b) Would the answer to (b) change if the helicopter were moving *downward* but still accelerating *upward*?

CHAPTER 4 SUMMARY

Big Picture

The big idea of this chapter—and of all Newtonian mechanics—is that **force** causes *change* in motion, not motion itself. Uniform motion—straight line, constant speed—needs no cause or explanation. Any deviation, in speed or direction, requires a **net force**. This idea is the essence of Newton's first and second laws. Combined with Newton's third law, these laws provide a consistent description of motion.

Newton's First Law

A body in uniform motion remains in uniform motion, and a body at rest remains at rest, unless acted on by a nonzero net force.

This law is implicit in Newton's second law.

Newton's Second Law

The rate at which a body's momentum changes is equal to the net force acting on the body.

Here **momentum** is the "quantity of motion," the product of mass and velocity.

Newton's Third Law

If object A exerts a force on object B, then object B exerts an oppositely directed force of equal magnitude on A.

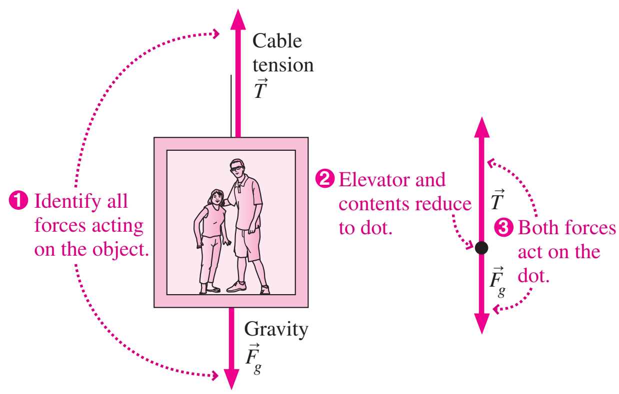
Newton's third law says that forces come in pairs.

Solving Problems with Newton's Laws

INTERPRET Interpret the problem to be sure that you know what it's asking and that Newton's second law is the relevant concept. Identify the object of interest and all the individual **interaction forces** acting on it.

DEVELOP Draw a **free-body diagram** as described in Tactics 4.1. Develop your solution plan by writing Newton's second law, $\vec{F}_{\text{net}} = m\vec{a}$, with $\vec{F}_{\text{net}}$ expressed as the sum of the forces you've identified. Then choose a coordinate system so you can express Newton's law in components.

EVALUATE At this point the physics is done, and you're ready to execute your plan by solving Newton's second law and evaluating the numerical answer(s), if called for. Remember that even in the one-dimensional problems of this chapter, Newton's law is a vector equation; that will help you get the signs right. You need to write the components of Newton's law in the coordinate system you chose, and then solve the resulting equation(s) for the quantity(ies) of interest.



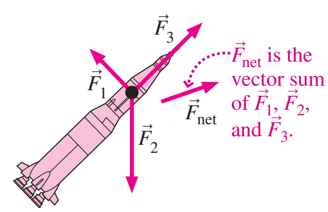
ASSESS Assess your solution to see that it makes sense. Are the numbers reasonable? Do the units work out correctly? What happens in special cases—for example, when a mass, a force, an acceleration, or an angle gets very small or very large?

Key Concepts and Equations

Mathematically, Newton's second law is $\vec{F}_{\text{net}} = d\vec{p}/dt$, where $\vec{p} = m\vec{v}$ is an object's momentum, and $\vec{F}_{\text{net}}$ is the sum of all the individual forces acting on the object. When an object has constant mass, the second law takes the familiar form

$$\vec{F}_{\text{net}} = m\vec{a} \qquad \text{(Newton's second law)}$$

Newton's second law is a *vector* equation. To use it correctly, you must write the components of the equation in a chosen coordinate system. In one-dimensional problems the result is a single equation.



Applications

The force of gravity on an object is its **weight**. Since all objects at a given location experience the same gravitational acceleration, weight is proportional to mass:

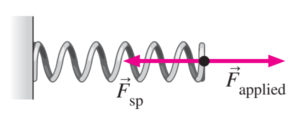
$$\vec{w} = m\vec{g} \qquad \text{(weight on Earth)}$$

In an accelerated reference frame, an object's **apparent weight** differs from its actual weight; in particular, an object in free fall experiences **apparent weightlessness**.

Springs are convenient force-measuring devices, stretching or compressing in response to the applied force. For an ideal spring, the stretch or compression is directly proportional to the force:

$$F_{\text{sp}} = -kx \qquad \text{(Hooke's law)}$$

where k is the **spring constant,** with units of N/m.



For Thought and Discussion

1. Distinguish the Aristotelian and Galilean/Newtonian views of the natural state of motion.
2. A ball bounces off a wall with the same speed it had before it hit the wall. Has its momentum changed? Has a force acted on the ball? Has a force acted on the wall? Relate your answers to Newton's laws of motion.
3. We often use the term "inertia" to describe human sluggishness. How is this usage related to the meaning of "inertia" in physics?
4. Does a body necessarily move in the direction of the net force acting on it?
5. A truck crashes into a stalled car. A student trying to explain the physics of this event claims that no forces are involved; the car was just "in the way" so it got hit. Comment.
6. A barefoot astronaut kicks a ball, hard, across a space station. Does the ball's apparent weightlessness mean the astronaut's toes don't hurt? Explain.
7. The surface gravity on Jupiter's moon Io is one-fifth that on Earth. What would happen to your weight and to your mass if you were on Io?
8. In paddling a canoe, you push water backward with your paddle. What force actually propels the canoe forward?
9. Is it possible for a nonzero net force to act on an object without the object's speed changing? Explain.
10. As your plane accelerates down the runway, you take your keys from your pocket and suspend them by a thread. Do they hang vertically? Explain.
11. A driver tells passengers to buckle their seatbelts, invoking the law of inertia. What's that got to do with seatbelts?

Exercises and Problems

Exercises

Section 4.2 Newton's First and Second Laws

12. A subway train's mass is 1.5×10^6 kg. What force is required to accelerate the train at $2.5\ \mathrm{m/s^2}$?
13. A 61-Mg railroad locomotive can exert a 0.12-MN force. At what rate can it accelerate (a) by itself and (b) when pulling a 1.4-Gg train?
14. A small plane accelerates down the runway at $7.2\ \mathrm{m/s^2}$. If its propeller provides an 11-kN force, what's the plane's mass?
15. A car leaves the road traveling at 110 km/h and hits a tree, coming to a stop in 0.14 s. What average force does a seatbelt exert on a 60-kg passenger during this collision?
16. By how much does the force required to stop a car increase if the initial speed is doubled while the stopping distance remains the same?
17. BIO Kinesin is a "motor protein" responsible for moving materials within living cells. If it exerts a 6.0-pN force, what acceleration will it give a molecular complex with mass 3.0×10^{-18} kg?
18. Starting from rest, a 940-kg racing car covers 400 m in 4.95 s. Find the average force on the car.
19. In an egg-dropping contest, a student encases an 85-g egg in a Styrofoam block. If the force on the egg can't exceed 1.5 N, and if the block hits the ground at 1.2 m/s, by how much must the Styrofoam compress on impact?
20. In a front-end collision, a 1300-kg car with shock-absorbing bumpers can withstand a maximum force of 65 kN before damage occurs. If the maximum speed for a nondamaging collision is 10 km/h, by how much must the bumper be able to move relative to the car?

Section 4.4 The Force of Gravity

21. Show that the units of acceleration can be written as N/kg. Why does it make sense to give g as 9.8 N/kg when talking about mass and weight?
22. Your spaceship crashes on one of the Sun's planets. Fortunately, the ship's scales are intact and show that your weight is 532 N. If your mass is 60 kg, where are you? (*Hint:* Consult Appendix E.)
23. Your friend can barely lift a 35-kg concrete block on Earth. How massive a block could she lift on the Moon?
24. A cereal box says "net weight 340 grams." What's the actual weight (a) in SI units and (b) in ounces?
25. You're a safety engineer for a bridge spanning the U.S.-Canadian border. U.S. specifications permit a maximum load of 10 tons. What load limit should you specify on the Canadian side, where "weight" is given in kilograms?
26. The gravitational acceleration at the International Space Station's altitude is about 89% of its surface value. What's the weight of a 68-kg astronaut at this altitude?

Section 4.5 Using Newton's Second Law

27. A 50-kg parachutist descends at a steady 40 km/h. What force does air exert on the parachute?
28. A 930-kg motorboat accelerates away from a dock at $2.3\ \mathrm{m/s^2}$. Its propeller provides a 3.9-kN thrust force. What drag force does the water exert on the boat?
29. An elevator accelerates downward at $2.4\ \mathrm{m/s^2}$. What force does the elevator's floor exert on a 52-kg passenger?
30. At 560 metric tons, the Airbus A-380 is the world's largest airliner. What's the upward force on an A-380 when the plane is (a) flying at constant altitude and (b) accelerating upward at $1.1\ \mathrm{m/s^2}$?
31. You're an engineer working on Ares I, NASA's replacement for the space shuttles. Performance specs call for a first-stage rocket capable of accelerating a total mass of 630 Mg vertically from rest to 7200 km/h in 2.0 min. You're asked to determine the required engine thrust (force) and the force exerted on a 75-kg astronaut during liftoff. What do you report?
32. You step into an elevator, and it accelerates to a downward speed of 9.2 m/s in 2.1 s. How does your apparent weight during this time compare with your actual weight?

Section 4.6 Newton's Third Law

33. What upward gravitational force does a 5600-kg elephant exert on Earth?
34. Your friend's mass is 65 kg. If she jumps off a 120-cm-high table, how far does Earth move toward her as she falls?
35. What force is necessary to stretch a spring 48 cm, if its spring constant is 270 N/m?
36. A 35-N force is applied to a spring with spring constant $k = 220$ N/m. How much does the spring stretch?
37. A spring with spring constant $k = 340$ N/m is used to weigh a 6.7-kg fish. How far does the spring stretch?

Problems

38. A 1.25-kg object is moving in the x-direction at 17.4 m/s. Just 3.41 s later, it's moving at 26.8 m/s at 34.0° to the x-axis. Find the magnitude and direction of the force applied during this time.

39. An airplane encounters sudden turbulence, and you feel momentarily lighter. If your apparent weight seems to be about 70% of your normal weight, what are the magnitude and direction of the plane's acceleration?
40. A 74-kg tree surgeon rides a "cherry picker" lift to reach the upper branches of a tree. What force does the lift exert on the surgeon when it's (a) at rest; (b) moving upward at a steady 2.4 m/s; (c) moving downward at a steady 2.4 m/s; (d) accelerating upward at $1.7\ \text{m/s}^2$; (e) accelerating downward at $1.7\ \text{m/s}^2$?
41. A dancer executes a vertical jump during which the floor pushes up on his feet with a force 50% greater than his weight. What's his upward acceleration?
42. Find expressions for the force needed to bring an object of mass m from rest to speed v (a) in time Δt and (b) over distance Δx.
43. An elevator moves upward at $5.2\ \text{m/s}$. What's its minimum stopping time if the passengers are to remain on the floor?
44. A 2.50-kg object is moving along the x-axis at 1.60 m/s. As it passes the origin, two forces $\vec{F}_1$ and $\vec{F}_2$ are applied, both in the y-direction (plus or minus). The forces are applied for 3.00 s, after which the object is at $x = 4.80$ m, $y = 10.8$ m. If $\vec{F}_1 = 15.0\hat{j}$N, what's $\vec{F}_2$?
45. Blocks of 1.0, 2.0, and 3.0 kg are lined up on a frictionless table, as shown in Fig. 4.21, with a 12-N force applied to the leftmost block. What force does the middle block exert on the rightmost one?

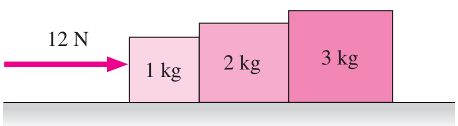


FIGURE 4.21 Problem 45

46. A child pulls an 11-kg wagon with a horizontal handle whose mass is 1.8 kg, accelerating the wagon and handle at $2.3\ \text{m/s}^2$. Find the tension forces at each end of the handle. Why are they different?
47. A 2200-kg airplane pulls two gliders, the first of mass 310 kg and the second of mass 260 kg, down the runway with acceleration $1.9\ \text{m/s}^2$ (Fig. 4.22). Neglecting the mass of the two ropes and any frictional forces, determine (a) the horizontal thrust of the plane's propeller; (b) the tension force in the first rope; (c) the tension force in the second rope; and (d) the net force on the first glider.

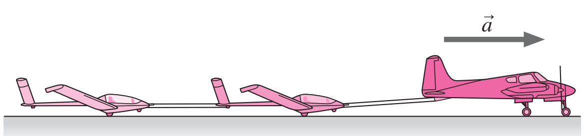


FIGURE 4.22 Problem 47

48. **BIO** A biologist is studying the growth of rats on the Space Station. To determine a rat's mass, she puts it in a 320-g cage, attaches a spring scale, and pulls so that the scale reads 0.46 N. If rat and cage accelerate at $0.40\ \text{m/s}^2$, what's the rat's mass?
49. An elastic towrope has spring constant 1300 N/m. It's connected between a truck and a 1900-kg car. As the truck tows the car, the rope stretches 55 cm. Starting from rest, how far do the truck and the car move in 1 min? Assume the car experiences negligible friction.
50. A 2.0-kg mass and a 3.0-kg mass are on a horizontal frictionless surface, connected by a massless spring with spring constant $k = 140$ N/m. A 15-N force is applied to the larger mass, as shown in Fig. 4.23. How much does the spring stretch from its equilibrium length?

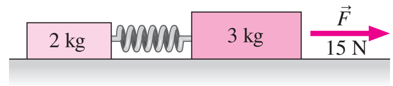


FIGURE 4.23 Problem 50

51. You're an automotive engineer designing the "crumple zone" of a new car—the region that compresses as the car comes to a stop in a head-on collision. If the maximum allowable force on a passenger in a 70-km/h collision is 20 times the passenger's weight, what do you specify for the amount of compression in the crumple zone?
52. **BIO** Frogs' tongues dart out to catch insects, with maximum tongue accelerations of about $250\ \text{m/s}^2$. What force is needed to give a 500-mg tongue such an acceleration?
53. Two large crates, with masses 640 kg and 490 kg, are connected by a stiff, massless spring $(k = 8.1\ \text{kN/m})$ and propelled along an essentially frictionless factory floor by a horizontal force applied to the more massive crate. If the spring compresses 5.1 cm, what's the applied force?
54. What force do the blades of a 4300-kg helicopter exert on the air when the helicopter is (a) hovering at constant altitude; (b) dropping at 21 m/s with speed decreasing at $3.2\ \text{m/s}^2$; (c) rising at 17 m/s with speed increasing at $3.2\ \text{m/s}^2$; (d) rising at a steady 15 m/s; (e) rising at 15 m/s with speed decreasing at $3.2\ \text{m/s}^2$?
55. What engine thrust (force) is needed to accelerate a rocket of mass m (a) downward at $1.40g$ near Earth's surface; (b) upward at $1.40g$ near Earth's surface; (c) at $1.40g$ in interstellar space, far from any star or planet?
56. Your engineering firm is asked to specify the maximum load for the elevators in a new building. Each elevator has mass 490 kg when empty and maximum acceleration $2.24\ \text{m/s}^2$. The elevator cables can withstand a maximum tension of 19.5 kN before breaking. For safety, you need to ensure that the tension never exceeds two-thirds of that value. What do you specify for the maximum load? How many 70-kg people is that?
57. An F-16 jet fighter has mass 12 Mg and engine thrust 132 kN. An Airbus A-380 has mass 560 Mg and total engine thrust 1.5 MN. Could either aircraft climb vertically with no lift from its wings? If so, what vertical acceleration could it achieve?
58. Two springs have the same unstretched length but different spring constants, k_1 and k_2. (a) If they're connected side by side and stretched a distance x, as shown in Fig. 4.24a, show that the force exerted by the combination is $(k_1 + k_2)x$. (b) If they're connected end to end (Fig. 4.24b) and the combination is stretched a distance x, show that they exert a force $k_1k_2x/(k_1 + k_2)$.

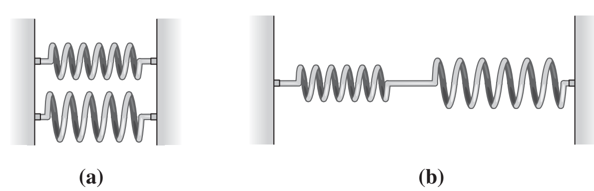


FIGURE 4.24 Problem 58

59. Although we usually write Newton's second law for one-dimensional motion in the form $F = ma$, which holds when mass is constant, a more fundamental version is $F = \dfrac{d(mv)}{dt}$. Consider

an object whose mass is changing, and use the product rule for derivatives to show that Newton's law then takes the form $F = ma + v\dfrac{dm}{dt}$.

60. A railroad car is being pulled beneath a grain elevator that dumps grain at the rate of 450 kg/s. Use the result of Problem 59 to find the force needed to keep the car moving at a constant 2.0 m/s.
61. A block 20% more massive than you hangs from a rope that goes over a frictionless, massless pulley. With what acceleration must you climb the other end of the rope to keep the block from falling?
62. You're asked to calibrate a device used to measure vertical acceleration in helicopters. The device consists of a mass m hanging from a massless spring of constant k. Your job is to express the acceleration as a function of the spring's stretch Δy from its equilibrium length. What's your expression?
63. Your airplane is caught in a brief, violent downdraft. To your amazement, pretzels rise vertically off your seatback tray, and you estimate their upward acceleration relative to the plane at 2 m/s^2. What's the plane's downward acceleration?
64. You're assessing the Engineered Material Arresting System (EMAS) at New York's JFK airport. The system consists of a 132-m-long bed of crushable cement blocks, designed to stop aircraft from sliding off the runway in emergencies. The EMAS can exert a 300-kN force on a 55-Mg jetliner that hits the system at 36 m/s. Can it stop the plane before it plows through all the blocks?
65. Two masses are joined by a massless string. A 30-N force applied vertically to the upper mass gives the system a constant upward acceleration of 3.2 m/s^2. If the string tension is 18 N, what are the two masses?
66. A mass M hangs from a uniform rope of length L and mass m. Find an expression for the rope tension as a function of the distance y measured downward from the top of the rope.
67. "Jerk" is the rate of change of acceleration, and it's what can make you sick on an amusement park ride. In a particular ride, a car and passengers with total mass M are subject to a force given by $F = F_0 \sin \omega t$, where F_0 and ω are constants. Find an expression for the maximum jerk.

Passage Problems

Laptop computers are equipped with accelerometers that sense when the device is dropped and then put the hard drive into a protective mode. Your computer geek friend has written a program that reads the accelerometer and calculates the laptop's apparent weight. You're amusing yourself with this program on a long plane flight. Your laptop weighs just 5 pounds, and for a long time that's what the program reports. But then the "Fasten Seatbelt" light comes on as the plane encounters turbulence. For the next 12 seconds, your laptop reports rapid changes in apparent weight, as shown in Fig. 4.25.

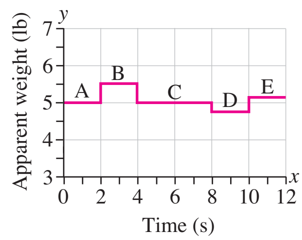


FIGURE 4.25 The laptop's apparent weight (Passage Problems 68–71).

68. At the first sign of turbulence, the plane's acceleration
    a. is upward.
    b. is downward.
    c. is impossible to tell from the graph.
69. The plane's vertical acceleration has its greatest magnitude
    a. during interval B.
    b. during interval C.
    c. during interval D.
70. During interval C, you can conclude for certain that the plane is
    a. at rest.
    b. accelerating upward.
    c. accelerating downward.
    d. moving with constant vertical velocity.
71. The magnitude of the greatest vertical acceleration the plane undergoes during the time shown on the graph is approximately
    a. 0.5 m/s^2.
    b. 1.0 m/s^2.
    c. 5 m/s^2.
    d. 10 m/s^2.

Answers to Chapter Questions

Answer to Chapter Opening Question

The human body exerts a contact force; wind and water are fluids that exert pressure forces; gravity is an action-at-a-distance force between Earth and the sailboard.

Answers to GOT IT? Questions

4.1. (b).
4.2. No. Look at Fig. 4.4*b*.
4.3. All would move in straight lines.
4.4. (a) greater; (b) less; (c) less; (d) greater; (e) equal.
4.5. (c) less than 2 N.
4.6. (a) No, the acceleration is still 0; (b) no, the direction of velocity is irrelevant (this situation would occur if the helicopter were moving downward but slowing).

5 Using Newton's Laws

Why doesn't the roller coaster fall off its loop-the-loop track?

New Concepts, New Skills

By the end of this chapter you should be able to

- Use Newton's second law to solve problems involving the motion of a single object in two dimensions under the influence of multiple forces (5.1).
- Solve Newton's law problems involving multiple objects (5.2).
- Explain that circular motion is just a special case of Newton's second law, and solve circular-motion problems involving multiple forces (5.3).
- Describe the force of friction, both static and kinetic, and solve Newton's law problems in which one of the forces is friction (5.4).
- Explain drag forces qualitatively (5.5).

Connecting Your Knowledge

- You first met Newton's laws in Chapter 4, and you should be familiar with all three of them (4.2, 4.6).
- This chapter builds especially on applications of Newton's second law, now generalizing to the case where the forces acting on an object no longer lie along a line (4.5).

Chapter 4 introduced Newton's three laws of motion and used them in one-dimensional applications. Now we apply Newton's laws in two dimensions. This material is at the heart of Newtonian physics, from textbook problems to systems that guide spacecraft to distant planets. The chapter consists largely of examples, to help you learn to apply Newton's laws and also to appreciate their wide range of applicability. We also introduce frictional forces and elaborate on circular motion. As you study the diverse examples, keep in mind that they all follow from the underlying principles embodied in Newton's laws.

5.1 Using Newton's Second Law

Newton's second law, $\vec{F}_{\text{net}} = m\vec{a}$, is the cornerstone of mechanics. We can use it to develop faster skis, engineer skyscrapers, design safer roads, compute a rocket's thrust, and solve myriad other practical problems.

We'll work Example 5.1 in great detail, applying Problem-Solving Strategy 4.1. Follow this example closely, and try to understand how our strategy is grounded in Newton's basic statement that the net force on an object determines that object's acceleration.

EXAMPLE 5.1 Newton's Law in Two Dimensions: Skiing

A skier of mass $m = 65$ kg glides down a slope at angle $\theta = 32°$, as shown in Fig. 5.1. Find (a) the skier's acceleration and (b) the force the snow exerts on the skier. The snow is so slippery that you can neglect friction.

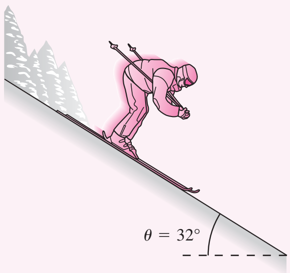


FIGURE 5.1 What's the skier's acceleration?

INTERPRET This problem is about the skier's motion, so we identify the skier as the object of interest. Next, we identify the forces acting on the object. In this case there are just two: the downward force of gravity and the normal force the ground exerts on the skier. As always, the normal force is perpendicular to the surfaces in contact—in this case perpendicular to the slope.

DEVELOP Our strategy for using Newton's second law calls for drawing a free-body diagram that shows only the object and the forces acting on it; that's Fig. 5.2. Determining the relevant equation is straightforward here: It's Newton's second law, $\vec{F}_{\text{net}} = m\vec{a}$. We write Newton's law explicitly for the forces we've identified:

$$\vec{F}_{\text{net}} = \vec{n} + \vec{F}_g = m\vec{a}$$

To apply Newton's law in two dimensions, we need to choose a coordinate system so that we can write this vector equation in components. Since the coordinate system is just a mathematical construct, you're free to choose any coordinate system you like—but a smart choice can make the problem a lot easier. In this example, the normal force is perpendicular to the slope and the skier's acceleration is along the slope. That means two of the three vectors in Newton's law will have only a single nonzero component if we choose a coordinate system with axes parallel and perpendicular to the slope. In the standard horizontal/vertical system, gravity would have a single component but we would have to break both the acceleration and the normal force into two components each (see Problem 32). The choice doesn't affect physical reality and you could work the problem either way, but the "tilted" coordinate system makes the math easier. We sketched this coordinate system on the free-body diagram in Fig. 5.2.

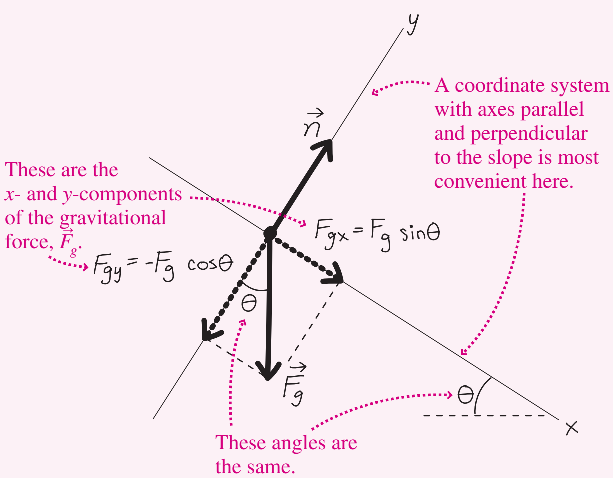


FIGURE 5.2 Our free-body diagram for the skier.

EVALUATE The rest is math. First, we write the components of Newton's law in our coordinate system. That means writing a version of the equation for each coordinate direction by removing the arrows indicating vector quantities and adding subscripts for the coordinate directions:

$$x\text{-component:} \quad n_x + F_{gx} = ma_x$$

$$y\text{-component:} \quad n_y + F_{gy} = ma_y$$

Don't worry about signs until the next step, when we actually evaluate the individual terms in these equations. Let's begin with the x equation. With the x-axis parallel and the y-axis perpendicular to the slope, the normal force has only a y-component, so $n_x = 0$. Meanwhile, the acceleration points downslope—that's the positive x-direction—so $a_x = a$, the magnitude of the acceleration. Only gravity has two nonzero components and, as Fig. 5.2 shows, trigonometry gives $F_{gx} = F_g \sin\theta$. But F_g, the magnitude of the gravitational force, is just mg, so $F_{gx} = mg \sin\theta$. This component has a positive sign because our x-axis slopes downward. Then, with $n_x = 0$, the x equation becomes

$$x\text{-component:} \quad mg \sin\theta = ma$$

On to the y equation. The normal force points in the positive y-direction, so $n_y = n$, the magnitude of the normal force. The acceleration has no component perpendicular to the slope, so $a_y = 0$. Figure 5.2 shows that $F_{gy} = -F_g \cos\theta = -mg \cos\theta$, so the y equation is

$$y\text{-component:} \quad n - mg \cos\theta = 0$$

Now we can evaluate to get the answers. The x equation solves directly to give

$$a = g \sin\theta = (9.8 \text{ m/s}^2)(\sin 32°) = 5.2 \text{ m/s}^2$$

which is the acceleration we were asked to find in (a). Next, we solve the y equation to get $n = mg \cos\theta$. Putting in the numbers gives $n = 540$ N. This is the answer to (b), the force the snow exerts on the skier.

ASSESS A look at two special cases shows that these results make sense. First, suppose $\theta = 0°$, so the surface is horizontal. Then the x equation gives $a = 0$, as expected. The y equation gives $n = mg$, showing that a horizontal surface exerts a force that just balances the skier's weight. At the other extreme, consider $\theta = 90°$, so the slope is a vertical cliff. Then the skier falls freely with acceleration g, as expected. In this case $n = 0$ because there's no contact between skier and slope. At intermediate angles, the slope's normal force lessens the effect of gravity, resulting in a lower acceleration. As the x equation shows, that acceleration is independent of the skier's mass—just as in the case of a vertical fall. The force exerted by the snow—here $mg \cos\theta$, or 540 N—is less than the skier's weight mg because the slope has to balance only the perpendicular component of the gravitational force.

If you understand this example, you should be able to apply Newton's second law confidently in other problems involving motion with forces in two dimensions.

Sometimes we're interested in finding the conditions under which an object won't accelerate. Examples are engineering problems, such as ensuring that bridges and buildings don't fall down, and physiology problems involving muscles and bones. Next we give a wilder example.

EXAMPLE 5.2 Objects at Rest: Bear Precautions

To protect her 17-kg pack from bears, a camper hangs it from ropes between two trees (Fig. 5.3). What's the tension in each rope?

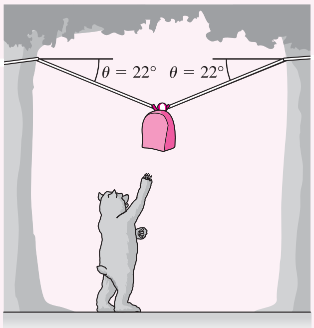


FIGURE 5.3 Bear precautions.

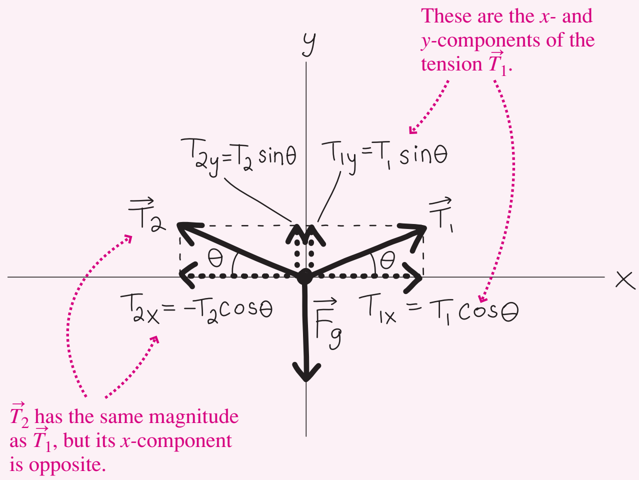


FIGURE 5.4 Our free-body diagram for the pack.

INTERPRET Here the pack is the object of interest. The only forces acting on it are gravity and tension forces in the two halves of the rope. To keep the pack from accelerating, they must sum to zero net force.

DEVELOP Figure 5.4 is our free-body diagram for the pack. The relevant equation is again Newton's second law, $\vec{F}_{\text{net}} = m\vec{a}$—this time with $\vec{a} = \vec{0}$. For the three forces acting on the pack, Newton's law is then $\vec{T}_1 + \vec{T}_2 + \vec{F}_g = \vec{0}$. Next, we need a coordinate system. The two rope tensions point in different directions that aren't perpendicular, so it doesn't make sense to align a coordinate axis with either of them. Instead, a horizontal/vertical system is simplest.

EVALUATE First we need to write Newton's law in components. Formally, we have $T_{1x} + T_{2x} + F_{gx} = 0$ and $T_{1y} + T_{2y} + F_{gy} = 0$ for the component equations. Figure 5.4 shows the components of the tension forces, and we see that $F_{gx} = 0$ and $F_{gy} = -F_g = -mg$. So our component equations become

$$x\text{-component:} \quad T_1 \cos\theta - T_2 \cos\theta = 0$$

$$y\text{-component:} \quad T_1 \sin\theta + T_2 \sin\theta - mg = 0$$

The x equation tells us something that's apparent from the symmetry of the situation: Since the angle θ is the same for both halves of the rope, the magnitudes T_1 and T_2 of the tension forces are the same. Let's just call the magnitude T: $T_1 = T_2 = T$. Then the first two terms in the y equation are equal, and the equation becomes $2T\sin\theta - mg = 0$, which gives

$$T = \frac{mg}{2\sin\theta} = \frac{(17\text{ kg})(9.8\text{ m/s}^2)}{2\sin 22^\circ} = 220\text{ N}$$

ASSESS Make sense? Let's look at some special cases. With $\theta = 90^\circ$, the rope hangs vertically, $\sin\theta = 1$, and the tension in each half of the rope is $\frac{1}{2}mg$. Of course: Each piece of the rope supports half the pack's weight. But as θ gets smaller, the ropes become more horizontal and the tension increases. That's because the vertical tension components together still have to support the pack's weight—but now there's a horizontal component as well, increasing the overall tension. Ropes break if the tension becomes too great, and in this example that means the rope's so-called breaking tension must be considerably greater than the pack's weight. If $\theta = 0$, in fact, the tension would become infinite—demonstrating that it's impossible to support a weight with a purely horizontal rope. ■

EXAMPLE 5.3 Objects at Rest: Restraining a Ski Racer

A starting gate acts horizontally to restrain a 60-kg ski racer on a frictionless 30° slope (Fig. 5.5). What horizontal force does the starting gate apply to the skier?

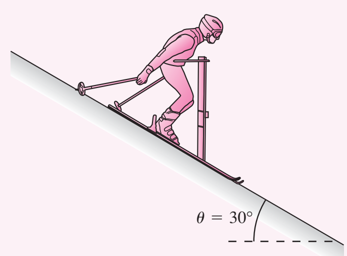


FIGURE 5.5 Restraining a skier.

INTERPRET Again, we want the skier to remain unaccelerated. The skier is the object of interest, and we identify three forces acting: gravity, the normal force from the slope, and a horizontal restraining force $\vec{F}_h$ that we're asked to find.

DEVELOP Figure 5.6 is our free-body diagram. The applicable equation is Newton's second law. Again, we want $\vec{a} = \vec{0}$, so with the forces we identified, $\vec{F}_{net} = m\vec{a}$ becomes $\vec{F}_h + \vec{n} + \vec{F}_g = \vec{0}$. Developing our solution strategy, we choose a coordinate system. With two forces now either horizontal or vertical, a horizontal/vertical system makes the most sense; we've shown this on Fig. 5.6.

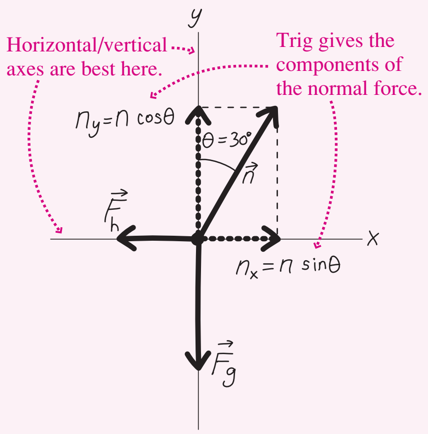


FIGURE 5.6 Our free-body diagram for the restrained skier.

EVALUATE As usual, the component equations follow directly from the vector form of Newton's law: $F_{hx} + n_x + F_{gx} = 0$ and $F_{hy} + n_y + F_{gy} = 0$. Figure 5.6 gives the components of the normal force and shows that $F_{hx} = -F_h$, $F_{gy} = -F_g = -mg$, and $F_{gx} = F_{hy} = 0$. Then the component equations become

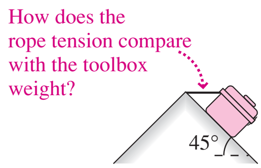
$$x:\quad -F_h + n\sin\theta = 0 \qquad y:\quad n\cos\theta - mg = 0$$

There are two unknowns here—namely, the horizontal force we're looking for and the normal force n. We can solve the y equation to get $n = mg/\cos\theta$. Using this expression in the x equation and solving for F_h then give the answer:

$$F_h = \frac{mg}{\cos\theta}\sin\theta = mg\tan\theta = (60\text{ kg})(9.8\text{ m/s}^2)(\tan 30°) = 340\text{ N}$$

ASSESS Again, let's look at the extreme cases. With $\theta = 0$, we have $F_h = 0$. Of course! It doesn't take any force to restrain a skier on flat ground. But as the slope becomes more vertical, $\tan\theta \rightarrow \infty$, and in the vertical limit, it becomes impossible to restrain the skier with a purely horizontal force. ■

GOT IT? 5.1 A roofer's toolbox rests on an essentially frictionless metal roof with a 45° slope, secured by a horizontal rope as shown. Is the rope tension (a) greater than, (b) less than, or (c) equal to the box's weight?



5.2 Multiple Objects

In the preceding examples there was a single object of interest. But often we have several objects whose motion is linked. Our Newton's law strategy still applies, with extensions to handle multiple objects.

PROBLEM-SOLVING STRATEGY 5.1 Newton's Second Law and Multiple Objects

INTERPRET Interpret the problem to be sure that you know what it's asking and that Newton's second law is the relevant concept. Identify the *multiple* objects of interest and all the individual interaction forces acting on *each* object. Finally, identify *connections* between the objects and the resulting *constraints* on their motions.

DEVELOP Draw a *separate* free-body diagram showing all the forces acting on *each* object. Develop your solution plan by writing Newton's law, $\vec{F}_{net} = m\vec{a}$, separately for each object, with $\vec{F}_{net}$ expressed as the sum of the forces acting on that object. Then choose a coordinate system appropriate to each object, so you can express each Newton's law equation in components. The coordinate systems for different objects don't need to have the same orientation.

EVALUATE At this point the physics is done, and you're ready to execute your plan by solving the equations and evaluating the numerical answer(s), if called for. Write the components of Newton's law for each object in the coordinate system you chose for each. You can then solve the resulting equations for the quantity(ies) you're interested in, using the connections you identified to relate the quantities that appear in the equations for the different objects.

ASSESS Assess your solution to see whether it makes sense. Are the numbers reasonable? Do the units work out correctly? What happens in special cases—for example, when a mass, a force, an acceleration, or an angle gets very small or very large?

EXAMPLE 5.4 Multiple Objects: Rescuing a Climber

A 70-kg climber finds himself dangling over the edge of an ice cliff, as shown in Fig. 5.7. Fortunately, he's roped to a 940-kg rock located 51 m from the edge of the cliff. Unfortunately, the ice is frictionless, and the climber accelerates downward. What's his acceleration, and how much time does he have before the rock goes over the edge? Neglect the rope's mass.

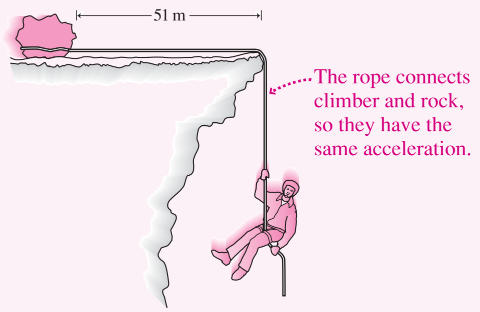
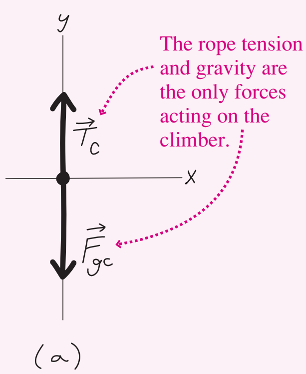


FIGURE 5.7 A climber in trouble.

INTERPRET We need to find the climber's acceleration, and from that we can get the time before the rock goes over the edge. We identify two objects of interest: the climber and the rock, and we note that the rope connects them. There are two forces on the climber: gravity and the upward rope tension. There are three forces on the rock: gravity, the normal force from the surface, and the rightward-pointing rope tension.

DEVELOP Figure 5.8 shows a free-body diagram for each object. Newton's law applies to each, so we write two vector equations:

$$\text{climber:} \quad \vec{T}_c + \vec{F}_{gc} = m_c\vec{a}_c$$
$$\text{rock:} \quad \vec{T}_r + \vec{F}_{gr} + \vec{n} = m_r\vec{a}_r$$



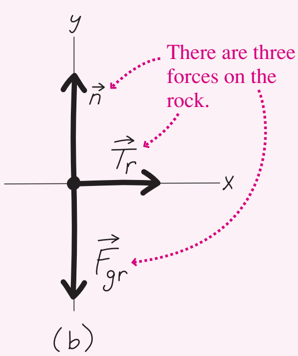


FIGURE 5.8 Our free-body diagrams for (a) the climber and (b) the rock.

where the subscripts c and r stand for climber and rock, respectively. All forces are either horizontal or vertical, so we can use the same horizontal/vertical coordinate system for both objects, as shown in Fig. 5.8.

EVALUATE Again, the component equations follow directly from the vector forms. There are no horizontal forces on the climber, so only the y equation is significant. We're skilled enough now to skip the intermediate step of writing the components without their actual expressions, and we see from Fig. 5.8a that the y-component of Newton's law for the climber becomes $T_c - m_c g = m_c a_c$. For the rock, the only horizontal force is the tension, pointing to the right or positive x-direction, so the rock's x equation is $T_r = m_r a_r$. Since it's on a horizontal surface, the rock has no vertical acceleration, so its y equation is $n - m_r g = 0$. In writing these equations, we haven't added the subscripts x and y because each vector has only a single nonzero component. Now we need to consider the connection between rock and climber. That's the rope, and its presence means that the magnitude of both accelerations is the same. Calling that magnitude a, we can see from Fig. 5.8 that $a_r = a$ and $a_c = -a$. The value for the rock is positive because $\vec{T}_r$ points to the right, which we defined as the positive x-direction; the value for the climber is negative because he's accelerating downward, which we defined as the negative y-direction. The rope, furthermore, has negligible mass, so the tension throughout it must be the same (more on this point just after the example). Therefore, the tension forces on rock and climber have equal magnitude T, so $T_c = T_r = T$. Putting this all together gives us three equations:

$$\text{climber, } y\text{:} \quad T - m_c g = -m_c a$$
$$\text{rock, } x\text{:} \quad T = m_r a$$
$$\text{rock, } y\text{:} \quad n - m_r g = 0$$

The rock's x equation gives the tension, which we can substitute into the climber's equation to get $m_r a - m_c g = -m_c a$. Solving for a then gives the answer:

$$a = \frac{m_c g}{m_c + m_r} = \frac{(70\text{ kg})(9.8\text{ m/s}^2)}{(70\text{ kg} + 940\text{ kg})} = 0.679\text{ m/s}^2$$

We didn't need the rock's y equation, which just says that the normal force supports the rock's weight.

ASSESS Again, let's look at special cases. Suppose the rock's mass is zero; then our expression gives $a = g$. In this case there's no rope tension and the climber plummets in free fall. Also, acceleration decreases as the rock's mass increases, so with an infinitely massive rock, the climber would dangle without accelerating. You can see physically why our expression for acceleration makes sense. The gravitational force $m_c g$ acting on the climber has to accelerate both rock and climber—whose combined mass is $m_c + m_r$. The result is an acceleration of $m_c g/(m_c + m_r)$.

We're not quite done because we were also asked for the time until the rock goes over the cliff, putting the climber in real trouble. We interpret this as a problem in one-dimensional motion from Chapter 2, and we determine that Equation 2.10, $x = x_0 + v_0 t + \frac{1}{2}at^2$, applies. With $x_0 = 0$ and $v_0 = 0$, we have $x = \frac{1}{2}at^2$. We evaluate by solving for t and using the acceleration we found along with $x = 51$ m for the distance from the rock to the cliff edge:

$$t = \sqrt{\frac{2x}{a}} = \sqrt{\frac{(2)(51\text{ m})}{0.679\text{ m/s}^2}} = 12\text{ s}$$ ■

✓TIP Ropes and Tension Forces

Tension forces can be confusing. In Example 5.4, the rock pulls on one end of the rope and the climber pulls on the other. So why isn't the rope tension the sum of these forces? And why is it important to neglect the rope's mass? The answers lie in the meaning of tension.

Figure 5.9 shows a situation similar to Example 5.4, with two people pulling on opposite ends of a rope with forces of 1 N each. You might think the rope tension is then 2 N, but it's not. To see why, consider the part of the rope that's highlighted in Fig. 5.9*b*. To the left is the hand pulling leftward with 1 N. The rope isn't accelerating, so there must be a 1-N force pulling to the right on the highlighted piece. The remainder of the rope provides that force. We could have divided the rope anywhere, so we conclude that every part of the rope exerts a 1-N force on the adjacent rope. That 1-N force is what we mean by the rope tension.

As long as the rope isn't accelerating, the net force on it must be zero, so the forces at the two ends have the same magnitude. That conclusion would hold even if the rope were accelerating—provided it had negligible mass. That's often a good approximation in situations involving tension forces. But if a rope, cable, or chain has significant mass and is accelerating, then the tension force differs at the two ends. That difference, according to Newton's second law, is what accelerates the rope.

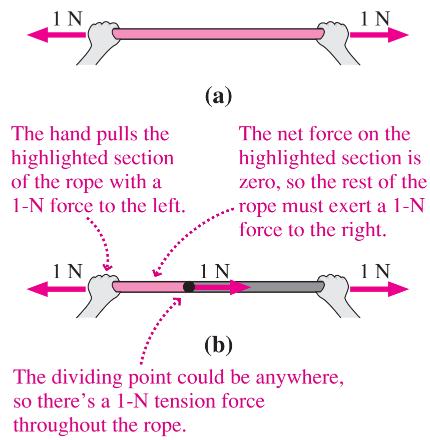
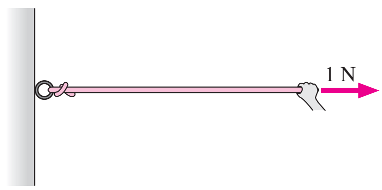


FIGURE 5.9 Understanding tension forces.

GOT IT? 5.2 In the figure we've replaced one hand with a hook attaching the rope to a wall. On the right, the hand still pulls with a 1-N force. Now what are (a) the rope tension and (b) the force exerted by the hook on the rope?



5.3 Circular Motion

A car rounds a curve. A satellite circles Earth. A proton whirls around a giant particle accelerator. Since they're not going in straight lines, Newton tells us that a force acts on each (Fig. 5.10). We know from Section 3.6 that the acceleration of an object moving with constant speed v in a circular path of radius r has magnitude v^2/r and points toward the center of the circle. Newton's second law then tells us that the magnitude of the net force on an object of mass m in circular motion is

$$F_{\text{net}} = ma = \frac{mv^2}{r} \quad \text{(uniform circular motion)} \tag{5.1}$$

The force is in the same direction as the acceleration—toward the center of the circular path. For that reason it's sometimes called the **centripetal force**, meaning center-seeking (from the Latin *centrum*, "center," and *petere*, "to seek").

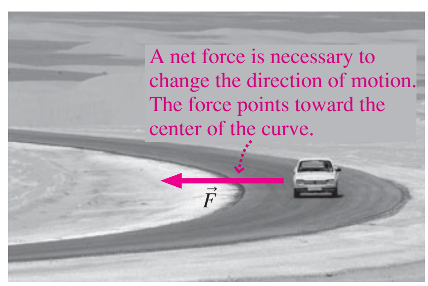


FIGURE 5.10 A car rounds a turn on the Trans-Sahara highway.

✓TIP Look for Real Forces

Centripetal force is *not* some new kind of force. It's just the name for *any* forces that keep an object in circular motion—which are always real, physical forces. Common examples of forces involved in circular motion include the gravitational force on a satellite, friction between tires and road, magnetic forces, tension forces, normal forces, and combinations of these and other forces.

Newton's second law describes circular motion exactly as it does any other motion: by relating net force, mass, and acceleration. Therefore, we can analyze circular motion with the same strategy we've used in other Newton's law problems.

EXAMPLE 5.5 Circular Motion: Whirling a Ball on a String

A ball of mass m whirls around in a horizontal circle at the end of a massless string of length L (Fig. 5.11). The string makes an angle θ with the horizontal. Find the ball's speed and the string tension.

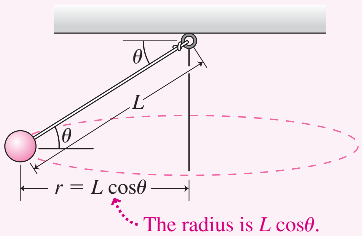


FIGURE 5.11 A ball whirling on a string.

INTERPRET This problem is similar to the Newton's law problems we worked involving force and acceleration. The object of interest is the ball, and only two forces are acting on it: gravity and the string tension.

DEVELOP Figure 5.12 is our free-body diagram showing the two forces we've identified. The relevant equation is Newton's second law, which becomes

$$\vec{T} + \vec{F}_g = m\vec{a}$$

The ball's path is in a horizontal plane, so its acceleration is horizontal. Then two of the three vectors in our problem—$\vec{F}_g$ and $\vec{a}$—are horizontal or vertical, so in developing our strategy, we choose a horizontal/vertical coordinate system.

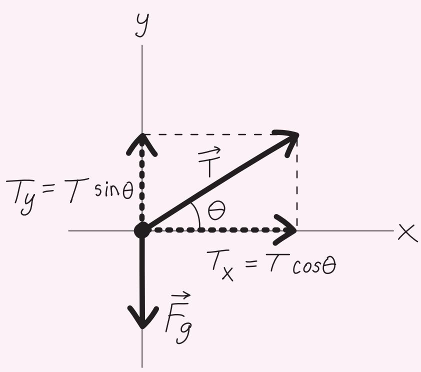


FIGURE 5.12 Our free-body diagram for the whirling ball.

✓TIP Real Forces Only!

Were you tempted to draw a third force in Fig. 5.12, perhaps pointing outward to balance the other two? *Don't!* Because the ball is accelerating, the net force is nonzero and the individual forces *do not balance*. Or maybe you were tempted to draw an inward-pointing force, mv^2/r. *Don't!* The quantity mv^2/r is *not* another force; it's just the product of mass and acceleration that appears in Newton's law. Students often complicate problems by introducing forces that aren't there. That makes physics seem harder than it is!

EVALUATE We now need the x- and y-components of Newton's law. Figure 5.12 shows that $F_{gy} = -F_g = -mg$ and gives tension components in terms of trig functions. The acceleration is purely horizontal, so $a_y = 0$, and since the ball is in circular motion, $a_x = v^2/r$. But what's r? Not the string length L, but the radius of the ball's circular path. Figure 5.11 shows that the radius is $L\cos\theta$. With all these expressions, the components of Newton's law become

$$x: \quad T\cos\theta = \frac{mv^2}{L\cos\theta} \qquad\qquad y: \quad T\sin\theta - mg = 0$$

We can get the tension directly from the y equation: $T = mg/\sin\theta$. Using this result in the x equation lets us solve for the speed v:

$$v = \sqrt{\frac{TL\cos^2\theta}{m}} = \sqrt{\frac{(mg/\sin\theta)L\cos^2\theta}{m}} = \sqrt{\frac{gL\cos^2\theta}{\sin\theta}}$$

ASSESS In the special case $\theta = 90°$, the string hangs vertically; here $\cos\theta = 0$, so $v = 0$. There's no motion, and the string tension equals the ball's weight. But as the string becomes increasingly horizontal, both speed and tension increase. And, just as in Example 5.2, the tension becomes very great as the string approaches horizontal. Here the string tension has two jobs to do: Its vertical component supports the ball against gravity, while its horizontal component keeps the ball in its circular path. The vertical component is always equal to mg, but as the string approaches horizontal, that becomes an insignificant part of the overall tension—and thus the tension and speed grow very large. ■

EXAMPLE 5.6 Circular Motion: Engineering a Road

Roads designed for high-speed travel have banked curves to give the normal force a component toward the center of the curve. That lets cars turn without relying on friction between tires and road. At what angle should a road with 200-m curvature radius be banked for travel at 90 km/h (25 m/s)?

INTERPRET This is another example involving circular motion and Newton's second law. Although we're asked about the road, a car on the road is the object we're interested in, and we need to design the road so the car can round the curve without needing a frictional force. That means the only forces on the car are gravity and the normal force.

DEVELOP Figure 5.13 shows the physical situation, and Fig. 5.14 is our free-body diagram for the car. Newton's second law is the applicable equation, and here it becomes $\vec{n} + \vec{F}_g = m\vec{a}$. Unlike the skier of Example 5.1, the car isn't accelerating down the slope, so a horizontal/vertical coordinate system makes the most sense.

EVALUATE First we write Newton's law in components. Gravity has only a vertical component, $F_{gy} = -mg$ in our coordinate system, and Fig. 5.14 shows the two components of the normal force. The acceleration is purely horizontal and points toward the center of the curve; in our coordinate system that's the positive x-direction. Since the car is in circular motion, the magnitude of the acceleration is v^2/r. So the components of Newton's law become

$$x:\quad n\sin\theta = \frac{mv^2}{r} \qquad y:\quad n\cos\theta - mg = 0$$

where the 0 on the right-hand side of the y equation reflects the fact that we don't want the car to accelerate in the vertical direction. Solving the y equation gives $n = mg/\cos\theta$. Then using this result in the x equation gives $mg\sin\theta/\cos\theta = mv^2/r$, or $g\tan\theta = v^2/r$. The mass canceled, which is good news because it means our banked road will work for a vehicle of any mass. Now we can solve for the banking angle:

$$\theta = \tan^{-1}\left(\frac{v^2}{gr}\right) = \tan^{-1}\left(\frac{(25\ \text{m/s})^2}{(9.8\ \text{m/s}^2)(200\ \text{m})}\right) = 18°$$

ASSESS Make sense? At low speed v or large radius r, the car's motion changes gently and it doesn't take a large force to keep it on its circular path. But as v increases or r decreases, the required force increases and so does the banking angle. That's because the horizontal component of the normal force is what keeps the car in circular motion, and the steeper the angle, the greater that component. ■

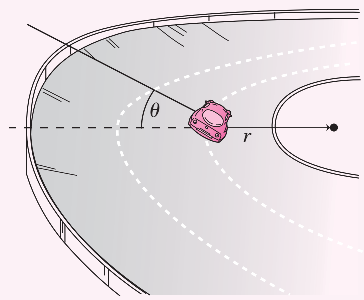


FIGURE 5.13 Car on a banked curve.

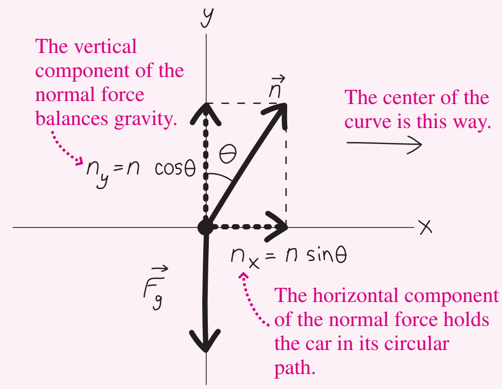


FIGURE 5.14 Our free-body diagram for the car on a banked curve.

EXAMPLE 5.7 Circular Motion: Looping the Loop

The "Great American Revolution" roller coaster at Valencia, California, includes a loop-the-loop section whose radius is 6.3 m at the top (see the chapter opening photo). What's the minimum speed for a roller-coaster car at the top of the loop if it's to stay on the track?

INTERPRET Again, we have circular motion described by Newton's second law. We're asked about the minimum speed for the car to stay on the track. What does it mean to stay on the track? It means there must be a normal force between car and track; otherwise, the two aren't in contact. So we can identify two forces acting on the car: gravity and the normal force from the track.

DEVELOP Figure 5.15 shows the physical situation. The situation is especially simple at the top of the track, where both forces point in the same direction. We show this in our free-body diagram, Fig. 5.16 (next page). Since that common direction is downward, it makes sense to choose a coordinate system with the y-axis *downward*. The applicable equation is Newton's second law, and with the two forces we've identified, that becomes $\vec{n} + \vec{F}_g = m\vec{a}$.

EVALUATE With both forces in the same direction, we need only the y-component of Newton's law. With the downward direction positive, $n_y = n$ and $F_{gy} = mg$. At the top of the loop, the car is in circular motion, so its acceleration is toward the center—downward—and has magnitude v^2/r. So $a_y = v^2/r$, and the y-component of Newton's law becomes

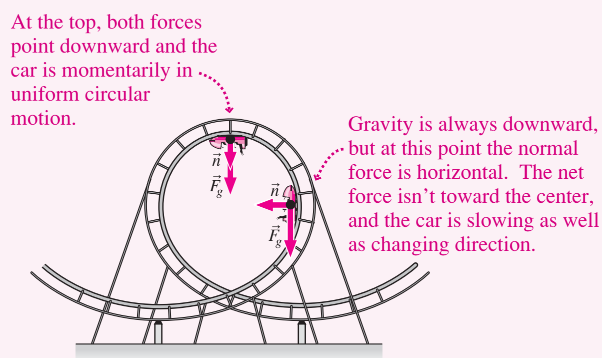


FIGURE 5.15 Forces on the roller-coaster car.

$$n + mg = \frac{mv^2}{r}$$

Solving for the speed gives $v = \sqrt{(nr/m) + gr}$. Now, the minimum possible speed for contact with the track occurs when n gets arbitrarily

(continued)

small right at the top of the track, so we find this minimum limit by setting $n = 0$. Then the answer is

$$v_{\min} = \sqrt{gr} = \sqrt{(9.8\ \text{m/s}^2)(6.3\ \text{m})} = 7.9\ \text{m/s}$$

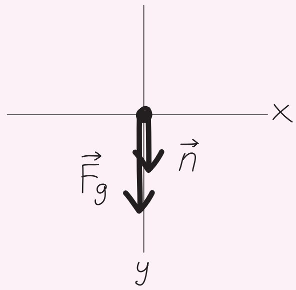


FIGURE 5.16 Our free-body diagram at the top of the loop.

ASSESS Do you see what's happening here? With the minimum speed, the normal force vanishes at the top of the loop, and gravity alone provides the force that keeps the object in its circular path. Since the motion is circular, that force must have magnitude mv^2/r. But the force of gravity alone is mg, and $v_{\min} = \sqrt{gr}$ follows directly from equating those two quantities. A car moving any slower than $v_{\min}$ would lose contact with the track and go into the parabolic trajectory of a projectile. For a car moving faster, there would be a nonzero normal force contributing to the downward acceleration at the top of the loop. In the "Great American Revolution," the actual speed at the loop's top is 9.7 m/s to provide a margin of safety. As with many problems involving gravity, the mass cancels. That's a good thing because it means the safe speed doesn't depend on the number or mass of the riders. ■

✓TIP Force and Motion

We've said this before, but it's worth noting again: Force doesn't cause motion but rather *change* in motion. The direction of an object's motion need not be the direction of the force on the object. That's true in Example 5.7, where the car is moving horizontally at the top of the loop while subject to a downward force. What *is* in the same direction as the force is the *change* in motion, here embodied in the center-directed acceleration of circular motion.

CONCEPTUAL EXAMPLE 5.1 Bad Hair Day

What's wrong with this cartoon showing riders on a loop-the-loop roller coaster (Fig. 5.17)?

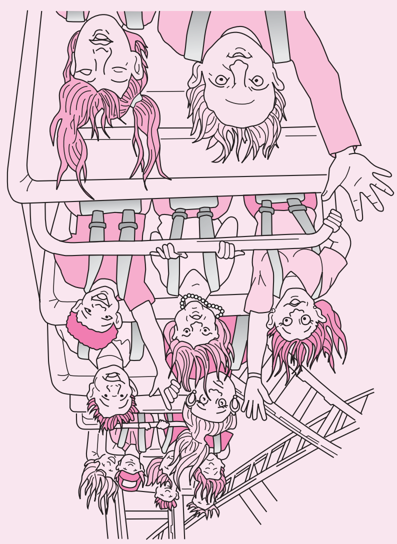


FIGURE 5.17 Conceptual Example 5.1.

EVALUATE Our objects of interest are the riders near the top of the roller coaster. We need to know the forces on them; one is obviously gravity. If the roller coaster is moving faster than Example 5.7's minimum speed—and it better be, for safety—then there are also normal forces from the seats as well as internal forces acting to accelerate parts of the riders' bodies.

Newton's law relates net force and acceleration: $\vec{F} = m\vec{a}$. This equation implies that the net force and acceleration must be in the same direction. At the top of the loop that direction is downward. Every part of the riders' bodies must therefore experience a net downward force. Again, Example 5.7 shows that the minimum force is that of gravity alone; for safety, there must be additional downward forces.

Now focus on the riders' hair, shown hanging downward. Forces on an individual hair are gravity and tension, and our safety argument shows that they should both point in the same direction—namely, downward—to provide a downward force stronger than gravity alone. How, then, can the riders' hair hang downward? That implies an *upward* tension force, inconsistent with our argument. The artist should have drawn the hair "hanging" upward.

ASSESS Make sense? Yes: To the riders, it feels like up is down! They feel the normal force of the seat pushing down, and their hairs experience a downward-pointing tension force. Even though the riders wear seatbelts, they don't need them: If the speed exceeds Example 5.7's minimum, then they feel tightly bound to their seats. Is there some mysterious new force that pushes them against their seats and that pulls their hair up? No! Newton's second law says the net force on the riders is in the direction of their acceleration—namely, downward. And for safety, that net force must be greater than gravity. It's those additional downward forces—the normal force from the seat and the tension force in the hair—that make up feel like down.

MAKING THE CONNECTION Suppose the riders feel like they weigh 50% of what they weigh at rest on the ground. How does the roller coaster's speed compare with Example 5.7's minimum?

EVALUATE In Example 5.7, we found the speed in terms of the normal force n and other quantities: $v = \sqrt{(nr/m) + gr}$. An apparent weight 50% of normal implies that $n = mg/2$. Then $v = \sqrt{(gr/2) + gr} = \sqrt{3/2}\sqrt{gr}$. Example 5.7 shows that the minimum speed is $\sqrt{gr}$, so our result is $\sqrt{3/2} \simeq 1.22$ times the minimum speed. Of course, that 50% apparent weight the riders feel is upward!

5.4 Friction

Your everyday experience of motion seems inconsistent with Newton's first law. Slide a book across the table, and it stops. Take your foot off the gas, and your car coasts to a stop. But Newton's law is right, so these examples show that some force must be acting. That force is **friction**, a force that opposes the relative motion of two surfaces in contact.

On Earth, we can rarely ignore friction. Some 20% of the gasoline burned in your car is used to overcome friction inside the engine. Friction causes wear and tear on machinery and clothing. But friction is also useful; without it, you couldn't drive or walk.

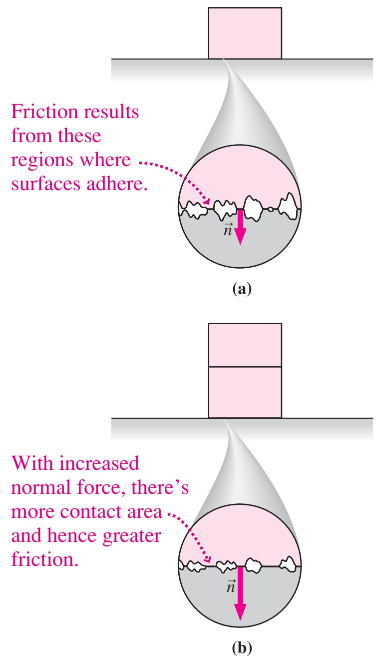


FIGURE 5.18 Friction originates in the contact between two surfaces.

The Nature of Friction

Friction is ultimately an electrical force between molecules in different surfaces. When two surfaces are in contact, microscopic irregularities adhere, as shown in Fig. 5.18*a*. At the macroscopic level, the result is a force that opposes any relative movement of the surfaces.

Experiments show that the magnitude of the frictional force depends on the normal force between surfaces in contact. Figure 5.18*b* shows why this makes sense: As the normal forces push the surfaces together, the actual contact area increases. There's more adherence, and this increases the frictional force.

At the microscopic level, friction is complicated. The simple equations we'll develop here provide approximate descriptions of frictional forces. Friction is important in everyday life, but it's not one of the fundamental physical interactions.

Frictional Forces

Try pushing a heavy trunk across the floor. At first nothing happens. Push harder; still nothing. Finally, as you push even harder, the trunk starts to slide—and you may notice that once it gets going, you don't have to push quite so hard. Why is that?

With the trunk at rest, microscopic contacts between trunk and floor solidify into relatively strong bonds. As you start pushing, you distort those bonds without breaking them; they respond with a force that opposes your applied force. This is the force of **static friction**, $\vec{f}_s$. As you increase the applied force, static friction increases equally, as shown in Fig. 5.19, and the trunk remains at rest. Experimentally, we find that the maximum static-friction force is proportional to the normal force between surfaces, and we write

$$f_s \le \mu_s n \qquad \text{(static friction)} \tag{5.2}$$

Here the proportionality constant μ_s (lowercase Greek mu, with the subscript s for "static") is the **coefficient of static friction**, a quantity that depends on the two surfaces. The $\le$ sign indicates that the force of static friction ranges from zero up to the maximum value on the right-hand side.

Eventually you push hard enough to break the bonds between trunk and floor, and the trunk begins to move; this is the point in Fig. 5.19 where the force suddenly drops. Now the microscopic bonds don't have time to strengthen, so the force needed to overcome them isn't so great. This weaker force between surfaces in relative motion is the force of **kinetic friction**, $\vec{f}_k$. Again, it's proportional to the normal force between the surfaces:

$$f_k = \mu_k n \qquad \text{(kinetic friction)} \tag{5.3}$$

where now the proportionality constant is μ_k, the **coefficient of kinetic friction**. Because kinetic friction is weaker, the coefficient of kinetic friction for a given pair of surfaces is less than the coefficient of static friction. Cross-country skiers exploit that fact by using waxes that provide a high coefficient of static friction for pushing against the snow and for climbing hills, while the lower kinetic friction permits effortless gliding.

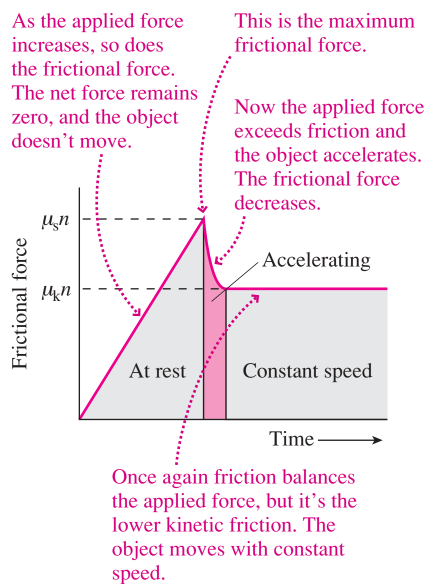


FIGURE 5.19 Behavior of frictional forces.

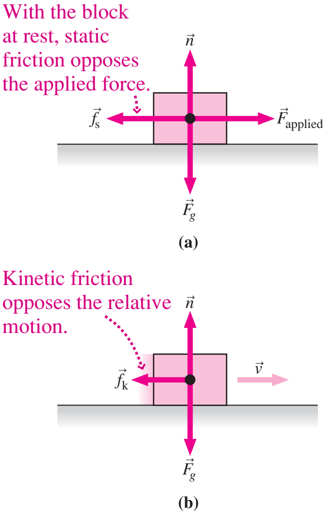


FIGURE 5.20 Direction of frictional forces.

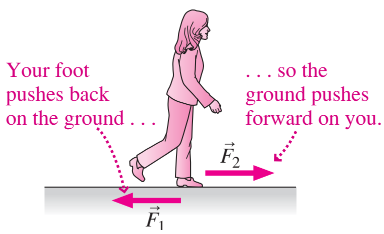


FIGURE 5.21 Walking.

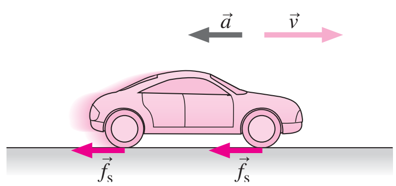


FIGURE 5.22 Friction stops the car.

Equations 5.2 and 5.3 give only the magnitudes of the frictional forces. The direction of the frictional force is parallel to the two surfaces, in the direction that opposes any applied force (Fig. 5.20*a*) or the surfaces' relative motion (Fig. 5.20*b*).

Since they describe proportionality between the magnitudes of two forces, the coefficients of friction are dimensionless. Typical values of μ_k range from less than 0.01 for smooth or lubricated surfaces to about 1.5 for very rough ones. Rubber on dry concrete—vital in driving an automobile—has μ_k about 0.8 and μ_s as high as 1. A waxed ski on dry snow has $\mu_k \approx 0.04$, while the synovial fluid that lubricates your body's joints reduces μ_k to a low 0.003.

If you push a moving object with a force equal to the opposing force of kinetic friction, then the net force is zero and, according to Newton, the object moves at constant speed. Since friction is nearly always present, but not as obvious as the push of a hand or the pull of a rope, you can see why it's so easy to believe that force is needed to make things move—rather than, as Newton recognized, to make them accelerate.

We emphasize that the equations describing friction are empirical expressions that approximate the effects of complicated but more basic interactions at the microscopic level. Our friction equations have neither the precision nor the fundamental character of Newton's laws.

Applications of Friction

Static friction plays a vital role in everyday activities such as walking and driving. As you walk, your foot contacting the ground is momentarily at rest, pushing back against the ground. By Newton's third law, the ground pushes forward, accelerating you forward (Fig. 5.21). Both forces of the third-law pair arise from static friction between foot and ground. On a frictionless surface, walking is impossible.

Similarly, the tires of an accelerating car push back on the road. If they aren't slipping, the bottom of each tire is momentarily at rest (more on this in Chapter 10). Therefore the force is static friction. The third law then requires a frictional force of the road pushing forward on the tires; that's what accelerates the car. Braking is the opposite: The tires push forward, and the road pushes back to decelerate the car (Fig. 5.22). The brakes affect only the wheels; it's friction between tires and road that stops the car. You know this if you've applied your brakes on an icy road!

EXAMPLE 5.8 Frictional Forces: Stopping a Car

The kinetic- and static-friction coefficients between a car's tires and a dry road are 0.61 and 0.89, respectively. If the car is traveling at 90 km/h (25 m/s) on a level road, determine the minimum stopping distance and the stopping distance with the wheels fully locked and the car skidding.

INTERPRET Since we're asked about the stopping distance, this is ultimately a question about accelerated motion in one dimension—the subject of Chapter 2. But here friction causes that acceleration, so we have a Newton's law problem. The car is the object of interest, and we identify three forces: gravity, the normal force, and friction.

DEVELOP Figure 5.23 is our free-body diagram. We have a two-part problem here: First, we need to use Newton's second law to find the acceleration, and then we can use Equation 2.11, $v^2 = v_0^2 + 2a\Delta x$, to relate distance and acceleration. With the three forces acting on the car, Newton's law becomes $\vec{F}_g + \vec{n} + \vec{f}_f = m\vec{a}$. A horizontal/vertical coordinate system is most appropriate for the components of Newton's law.

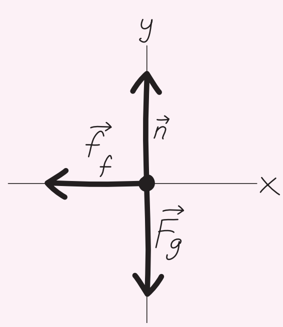


FIGURE 5.23 Our free-body diagram for the braking car.

EVALUATE The only horizontal force is friction, which points in the $-x$-direction and has magnitude μn, where μ can be either the kinetic- or the static-friction coefficient. The normal force and gravity act in the vertical direction, so the component equations are

$$x: \quad -\mu n = ma_x \qquad y: \quad -mg + n = 0$$

Solving the y equation for n and substituting in the x equation give the acceleration: $a_x = -\mu g$. We then use this result in Equation 2.11 and solve for the stopping distance Δx. With final speed $v = 0$, this gives

$$\Delta x = \frac{v_0^2}{-2a_x} = \frac{v_0^2}{2\mu g}$$

Using the numbers given, we get $\Delta x = 36$ m for the minimum stopping distance (no skid; static friction) and 52 m for the car skidding with its wheels locked (kinetic friction). The difference could well be enough to prevent an accident.

ASSESS Our result $a_x = -\mu g$ shows that a higher friction coefficient leads to a larger acceleration; this makes sense because friction is what causes the acceleration. What happened to the car's mass? A more massive car requires a larger frictional stopping force for the same acceleration—but friction depends on the normal force, and the latter is greater in proportion to the car's mass. Thus the stopping distance doesn't depend on mass.

This example shows that stopping distance increases as the *square* of the speed. That's one reason high speeds are dangerous: Doubling your speed quadruples your stopping distance! ■

EXAMPLE 5.9 Frictional Forces: Steering

A level road makes a 90° turn with radius 73 m. What's the maximum speed for a car to negotiate this turn when the road is dry $(\mu_s = 0.88)$ and when the road is snow covered $(\mu_s = 0.21)$?

INTERPRET This example is similar to Example 5.8, but now the frictional force acts perpendicular to the car's motion, keeping it in a circular path. Because the car isn't moving in the direction of the force, we're dealing with *static* friction. The car is the object of interest, and again the forces are gravity, the normal force, and friction.

DEVELOP Figure 5.24 is our free-body diagram. Newton's law is the applicable equation, and we're dealing with the acceleration v^2/r that occurs in circular motion. With the three forces acting on the car, Newton's law is $\vec{F}_g + \vec{n} + \vec{f}_s = m\vec{a}$. A horizontal/vertical coordinate system is most appropriate, and now it's most convenient to take the x-axis in the direction of the acceleration—namely, toward the center of the curve.

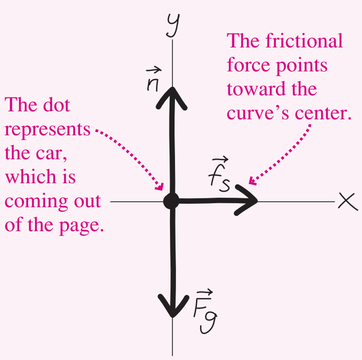


FIGURE 5.24 Our free-body diagram for the cornering car.

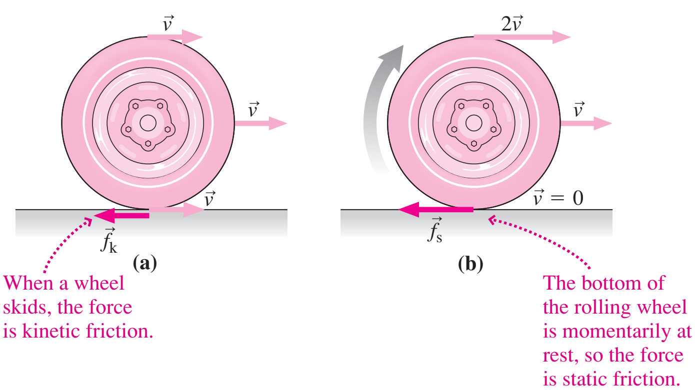
EVALUATE Again, the only horizontal force is friction, with magnitude $\mu_s n$. Here it points in the positive x-direction, as does the acceleration of magnitude v^2/r. So the x-component of Newton's law is $\mu_s n = mv^2/r$. There's no vertical acceleration, so the y-component is $-mg + n = 0$. Solving for n and using the result in the x equation give $\mu_s mg = mv^2/r$. Again the mass cancels, and we solve for v to get

$$v = \sqrt{\mu_s g r}$$

Putting in the numbers, we get $v = 25$ m/s (90 km/h) for the dry road and 12 m/s (44 km/h) for the snowy road. Exceed these speeds, and your car inevitably moves in a path with a larger radius—and that means going off the road!

ASSESS Once again, it makes sense that the car's mass doesn't matter. A more massive car needs a larger frictional force, and it gets what it needs because its larger mass results in a larger normal force. The safe speed increases with the curve radius r, and that, too, makes sense: A larger radius means a gentler turn, with less acceleration at a given speed. So less frictional force is needed. ■

APPLICATION Antilock Brakes



Today's cars have computer-controlled antilock braking systems (ABS). These systems exploit the fact that static friction is greater than kinetic friction. Slam on the brakes of a non-ABS car and the wheels lock and skid without turning. The force between tires and road is then *kinetic* friction (part a in the figure). But if you pump the brakes to keep the wheels from skidding, then it's the greater force of *static* friction (part b).

ABS improves on this brake-pumping strategy with a computer that independently controls the brakes at each wheel, keeping each just on the verge of slipping. Drivers of ABS cars should slam the brakes hard in an emergency; the ensuing clatter indicates the ABS system is working.

Although ABS can reduce the stopping distance, its real significance is in preventing vehicles from skidding out of control as can happen when you apply the brakes with some wheels on ice and others on pavement. Increasingly, today's cars incorporate their computer-controlled brakes into sophisticated systems that enhance stability during emergency maneuvers.

EXAMPLE 5.10 Friction on a Slope: Avalanche!

A storm dumps new snow on a ski slope. The coefficient of static friction between the new snow and the older snow underneath is 0.46. What's the maximum slope angle to which the new snow can adhere?

INTERPRET The problem asks about an angle, but it's friction that holds the new snow to the old, so this is really a problem about the maximum possible static friction. We aren't given an object, but we can model the new snow as a slab of mass m resting on a slope of unknown angle θ. The forces on the slab are gravity, the normal force, and static friction $\vec{f}_s$.

DEVELOP Figure 5.25 shows the model, and Fig. 5.26 is our free-body diagram. Newton's second law is the applicable equation, here with $\vec{a} = \vec{0}$, giving $\vec{F}_g + \vec{n} + \vec{f}_s = \vec{0}$. We also need the maximum static-friction force, given in Equation 5.2, $f_{s\,max} = \mu_s n$. As in Example 5.1, a tilted coordinate system is simplest and is shown in Fig. 5.26.

EVALUATE With the positive x-direction downslope, Fig. 5.26 shows that the x-component of gravity is $F_g \sin\theta = mg \sin\theta$, while the frictional force acts upslope ($-x$-direction) and has maximum magnitude $\mu_s n$; therefore, $f_{sx} = -\mu_s n$. So the x-component of Newton's law is $mg \sin\theta - \mu_s n = 0$. We can read the y-component from Fig. 5.26:

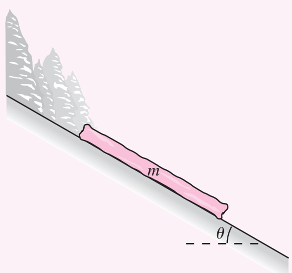


FIGURE 5.25 A layer of snow, modeled as a slab on a sloping surface.

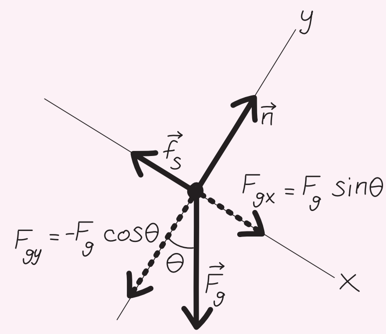


FIGURE 5.26 Our free-body diagram for the snow slab.

$-mg \cos\theta + n = 0$. Solving the y equation gives $n = mg \cos\theta$. Using this result in the x equation then yields $mg \sin\theta - \mu_s mg \cos\theta = 0$. Both m and g cancel, and we have $\sin\theta = \mu_s \cos\theta$ or, since $\tan\theta = \sin\theta/\cos\theta$,

$$\tan\theta = \mu_s$$

For the numbers in this example, we get $\theta = \tan^{-1}\mu_s = \tan^{-1}(0.46) = 25°$.

ASSESS Make sense? Sure: The steeper the slope, the greater the friction needed to keep the snow from sliding. Two effects are at work here: First, as the slope steepens, so does the component of gravity along the slope. Second, as the slope steepens, the normal force gets smaller, and that reduces the frictional force for a given friction coefficient. Note here that the normal force is not simply the weight mg of the snow; again, that's because of the sloping surface.

The real avalanche danger comes at angles slightly smaller than our answer $\tan\theta = \mu_s$, where a thick snowpack can build up. Changes in the snow's composition with temperature may decrease the friction coefficient and unleash an avalanche. ■

EXAMPLE 5.11 Friction: Dragging a Trunk

You drag a trunk of mass m across a level floor using a massless rope that makes an angle θ with the horizontal (Fig. 5.27). Given a kinetic-friction coefficient μ_k, what rope tension is required to move the trunk at constant speed?

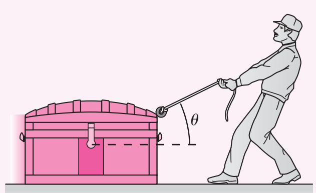


FIGURE 5.27 Dragging a trunk.

INTERPRET Even though the trunk is moving, it isn't accelerating, so here's another problem involving Newton's law with zero acceleration. The object is the trunk, and now four forces act: gravity, the normal force, friction, and the rope tension.

DEVELOP Figure 5.28 is our free-body diagram showing all four forces acting on the trunk. The relevant equation is Newton's law.

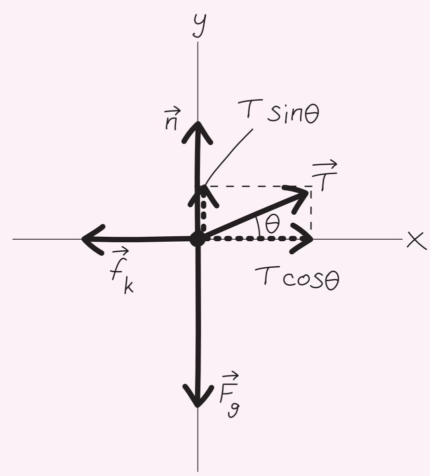


FIGURE 5.28 Our free-body diagram for the trunk.

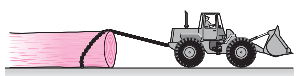
With no acceleration, it is $\vec{F}_g + \vec{n} + \vec{f}_k + \vec{T} = \vec{0}$, with the magnitude of kinetic friction given by $f_k = \mu_k n$. All vectors except the tension force are horizontal or vertical, so we choose a horizontal/vertical coordinate system.

EVALUATE From Fig. 5.28, we can write the components of Newton's law: $T\cos\theta - \mu_k n = 0$ in the x-direction and $T\sin\theta - mg + n = 0$ in the y-direction. This time the unknown T appears in both equations. Solving the y equation for n gives $n = mg - T\sin\theta$. Putting this n in the x equation then yields $T\cos\theta - \mu_k(mg - T\sin\theta) = 0$. Factoring terms involving T and solving, we arrive at the answer:

$$T = \frac{\mu_k mg}{\cos\theta + \mu_k \sin\theta}$$

ASSESS Make sense? Without friction, we wouldn't need any force to move the trunk at constant speed, and indeed our expression gives $T = 0$ in this case. If $\theta = 0$, then $\sin\theta = 0$ and we get $T = \mu_k mg$. Of course: In this case the normal force equals the weight, so the frictional force is $\mu_k mg$. Since the frictional force is horizontal and with $\theta = 0$ we're pulling horizontally, this is also the magnitude of the tension force. At intermediate angles, two effects come into play: First, the upward component of tension helps support the trunk's weight, and that means less normal force is needed. With less normal force, there's less friction—making the trunk easier to pull. But as the angle increases, less of the tension is horizontal and that means a larger tension force is needed to overcome friction. In combination, these two effects mean there's an optimum angle at which the rope tension is a minimum. Problem 66 explores this point further. ■

GOT IT? 5.3 The figure shows a logging vehicle pulling a redwood log. Is the frictional force in this case (a) less than, (b) equal to, or (c) greater than the weight multiplied by the coefficient of friction?



5.5 Drag Forces

Friction isn't the only "hidden" force that robs objects of their motion and obscures Newton's first law. Objects moving through fluids like water or air experience **drag forces** that oppose the relative motion of object and fluid. Ultimately, drag results from collisions between fluid molecules and the object. The drag force depends on several factors, including fluid density and the object's cross-sectional area and speed.

Terminal Speed

When an object falls from rest, its speed is initially low and so is the velocity-dependent drag force. It therefore accelerates downward with nearly the gravitational acceleration g. But as the object gains speed, the drag force increases—until eventually the drag force and gravity have equal magnitudes. At that point the net force on the object is zero, and it falls with constant speed, called its **terminal speed**.

Because the drag force depends on an object's area and the gravitational force depends on its mass, the terminal speed is lower for lighter objects with large areas. A parachute, for example, is designed specifically to have a large surface area that results, typically, in a terminal speed around 5 m/s. A ping-pong ball and a golf ball have about the same size and therefore the same area, but the ping-pong ball's much lower mass leads to a terminal speed of about 10 m/s compared with the golf ball's 50 m/s. For an irregularly shaped object, the drag and thus the terminal speed depend on how large a surface area the object presents to the air. Skydivers exploit this effect to vary their rates of fall.

Drag and Projectile Motion

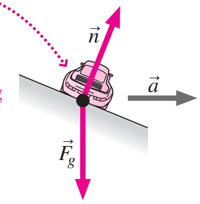
In Chapter 3, we consistently neglected air resistance—the drag force of air—in projectile motion. Calculating drag effects on projectiles is not trivial and usually requires computer calculations. The net effect, though, is obvious: Air resistance decreases the range of a projectile. Despite the physicist's need for computer calculations, others—especially athletes—have a feel for drag forces that lets them play their sports by judging correctly the trajectory of a projectile under the influence of drag forces.

CHAPTER 5 SUMMARY

Big Picture

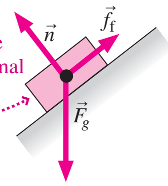
The big idea here is the same as in Chapter 4—namely, that Newton's laws are a universal description of motion, in which force causes not motion itself but change in motion. Here we focus on Newton's second law, extended to the richer and more complex examples of motion in two dimensions. To use Newton's law, we now sum forces that may point in different directions, but the result is the same: The net force determines an object's acceleration.

Here's a car on a banked turn. The forces on it don't sum to zero because the car is accelerating toward the center of the turn.



Common forces include gravity, the normal force from surfaces, tension forces, and a force introduced here: friction. Important examples are those where an object is accelerating, including in circular motion, and those where there's no acceleration and therefore the net force is zero.

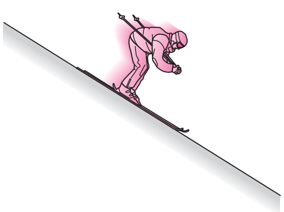
A block sits at rest on a slope. The three forces—gravity, normal force, and friction—sum to zero.



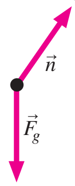
Solving Problems with Newton's Laws

The problem-solving strategy in this chapter is exactly the same as in Chapter 4, except that in two dimensions the choice of coordinate system and the division of forces into components become crucial steps. You usually need both component equations to solve a problem.

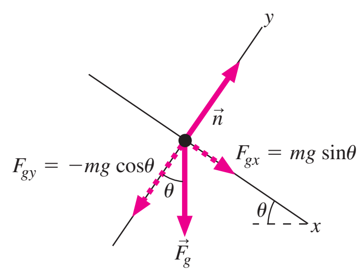
A skier on a frictionless slope



Free-body diagram showing the two forces acting



Coordinate system and vector components



$$\vec{F} = m\vec{a} \rightarrow \vec{n} + \vec{F}_g = m\vec{a} \rightarrow \begin{cases} n_x + F_{gx} = ma_x \\ n_y + F_{gy} = ma_y \end{cases} \rightarrow \begin{cases} mg\sin\theta = ma_x \\ n - mg\cos\theta = 0 \end{cases}$$

Key Concepts and Equations

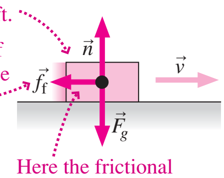
Newton's second law, $\vec{F}_{net} = m\vec{a}$, is the key equation in this chapter. It's crucial to remember that it's a *vector* equation, representing a pair of scalar equations for its two components in two dimensions.

Applications

Friction acts between surfaces to oppose their relative motion, and its strength depends on the normal force $\vec{n}$ acting perpendicular to them. When surfaces aren't actually in relative motion, the force is **static friction**, whose value ranges from zero to a maximum value $\mu_s n$ as needed to oppose any applied force: $f_s \le \mu_s n$. Here μ_s is the **coefficient of static friction**, which depends on the nature of the two surfaces. For surfaces in relative motion, the force is **kinetic friction**, given by $f_k = \mu_k n$, where the **coefficient of kinetic friction** is less than the coefficient of static friction.

A block moving to the right experiences a frictional force to the left.

The magnitude of the frictional force depends on the normal force: $f_f = \mu n$.



Here the frictional force is a little less than the normal force, so μ is a little less than 1.

For Thought and Discussion

1. Compare the net force on a heavy trunk when it's (a) at rest on the floor; (b) being slid across the floor at constant speed; (c) being pulled upward in an elevator whose cable tension equals the combined weight of the elevator and trunk; and (d) sliding down a frictionless ramp.
2. The force of static friction acts only between surfaces at rest. Yet that force is essential in walking and in accelerating or braking a car. Explain.
3. A jet plane flies at constant speed in a vertical circular loop. At what point in the loop does the seat exert the greatest force on the pilot? The least force?
4. In cross-country skiing, skis should easily glide forward but should remain at rest when the skier pushes back against the snow. What frictional properties should the ski wax have to achieve this goal?
5. Why do airplanes bank when turning?
6. Why is it easier for a child to stand nearer the inside of a rotating merry-go-round?
7. Gravity pulls a satellite toward Earth's center. So why doesn't the satellite actually fall to Earth?
8. Explain why a car with ABS brakes can have a shorter stopping distance.
9. A fishing line has a 20-lb breaking strength. Is it possible to break the line while reeling in a 15-lb fish? Explain.
10. Two blocks rest on slopes of unequal angles, connected by a rope passing over a pulley (Fig. 5.29). If the blocks have equal masses, will they remain at rest? Why? Neglect friction.

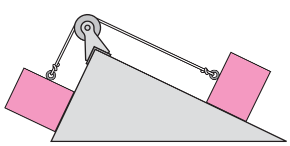


FIGURE 5.29 For Thought and Discussion 10; Exercises 19 and 20

11. You're on a plane undergoing a banked turn, so steep that out the window you see the ground below. Yet your pretzels stay put on the seatback tray, rather than sliding downward. Why?

Exercises and Problems

Exercises

Section 5.1 Using Newton's Second Law

12. Two forces, both in the x-y plane, act on a 1.5-kg mass that accelerates at 7.3 m/s^2 in a direction 30° counterclockwise from the x-axis. One force has magnitude 6.8 N and points in the $+x$-direction. Find the other force.
13. Two forces act on a 3.1-kg mass that undergoes acceleration $\vec{a} = 0.91\hat{\imath} - 0.27\hat{\jmath}$ m/s^2. If one force is $-1.2\hat{\imath} - 2.5\hat{\jmath}$ N, what's the other?
14. At what angle should you tilt an air table to simulate free fall at the Moon's surface, where $g = 1.6$ m/s^2?
15. A skier starts from rest at the top of a 24° slope 1.3 km long. Neglecting friction, how long does it take to reach the bottom?
16. A tow truck is connected to a 1400-kg car by a cable that makes a 25° angle to the horizontal. If the truck accelerates at 0.57 m/s^2, what's the magnitude of the cable tension? Neglect friction and the cable's mass.
17. BIO Studies of gymnasts show that their high rate of injuries to the Achilles tendon is due to tensions in the tendon that typically reach 10 times body weight. That force is provided by a pair of muscles, each exerting a force at 25° to the vertical, with their horizontal components opposite. For a 55-kg gymnast, find the force in each of these muscles.

Section 5.2 Multiple Objects

18. Your 12-kg baby sister pulls on the bottom of the tablecloth with all her weight. On the table, 60 cm from the edge, is a 6.8-kg roast turkey. (a) What's the turkey's acceleration? (b) From the time your sister starts pulling, how long do you have to intervene before the turkey goes over the edge? Neglect friction.
19. If the left-hand slope in Fig. 5.29 makes a 60° angle with the horizontal, and the right-hand slope makes a 20° angle, how should the masses compare if the objects are not to slide along the frictionless slopes?
20. Suppose the angles shown in Fig. 5.29 are 60° and 20°. If the left-hand mass is 2.1 kg, what should the right-hand mass be so that it accelerates (a) downslope at 0.64 m/s^2 and (b) upslope at 0.76 m/s^2?
21. Two unfortunate climbers, roped together, are sliding freely down an icy mountainside. The upper climber (mass 75 kg) is on a slope at 12° to the horizontal, but the lower climber (mass 63 kg) has gone over the edge to a steeper slope at 38°. (a) Assuming frictionless ice and a massless rope, what's the acceleration of the pair? (b) The upper climber manages to stop the slide with an ice ax. After the climbers have come to a complete stop, what force must the ax exert against the ice?

Section 5.3 Circular Motion

22. Suppose the Moon were held in its orbit not by gravity but by tension in a massless cable. Estimate the magnitude of the cable tension. (*Hint:* See Appendix E.)
23. Show that the force needed to keep a mass m in a circular path of radius r with period T is $4\pi^2 mr/T^2$.
24. A 940-g rock is whirled in a horizontal circle at the end of a 1.30-m-long string. (a) If the breaking strength of the string is 120 N, what's the minimum angle the string can make with the horizontal? (b) At this minimum angle, what's the rock's speed?
25. You're investigating a subway accident in which a train derailed while rounding an unbanked curve of radius 132 m, and you're asked to estimate whether the train exceeded the 45-km/h speed limit for this curve. You interview a passenger who had been standing and holding onto a strap; she noticed that an unused strap was hanging at about a 15° angle to the vertical just before the accident. What do you conclude?
26. A tetherball on a 1.7-m rope is struck so that it goes into circular motion in a horizontal plane, with the rope making a 15° angle to the horizontal. What's the ball's speed?
27. An airplane goes into a turn 3.6 km in radius. If the banking angle required is 28° from the horizontal, what's the plane's speed?

Section 5.4 Friction

28. Movers slide a 73-kg file cabinet along a floor where the coefficient of kinetic friction is 0.81. What's the frictional force on the cabinet?

29. A hockey puck is given an initial speed of 14 m/s. If it comes to rest in 56 m, what's the coefficient of kinetic friction?
30. Starting from rest, a skier slides 100 m down a 28° slope. How much longer does the run take if the coefficient of kinetic friction is 0.17 instead of 0?
31. What frictional coefficient is needed to keep a car moving at 90 km/h on a 120-m-radius unbanked turn?

Problems

32. Repeat Example 5.1, this time using a horizontal/vertical coordinate system.
33. A block is launched with initial speed 2.2 m/s up a 35° frictionless ramp. How far up the ramp does it slide?
34. **BIO** In the process of mitosis (cell division), two motor proteins pull on a spindle pole, each with a 7.3-pN force. The two force vectors make a 65° angle. What's the magnitude of the force the two motor proteins exert on the spindle pole?
35. A 15-kg monkey hangs from the middle of a massless rope, each half of which makes an 8° angle with the horizontal. What's the rope tension? Compare with the monkey's weight.
36. A camper hangs a 26-kg pack between two trees using separate ropes of different lengths, as shown in Fig. 5.30. Find the tension in each rope.

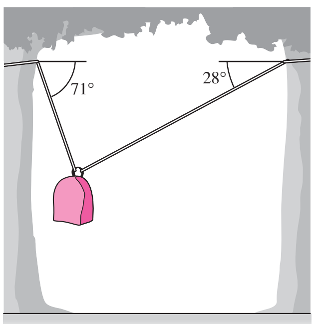


FIGURE 5.30 Problem 36

37. A mass m_1 undergoes circular motion of radius R on a horizontal frictionless table, connected by a massless string through a hole in the table to a second mass m_2 (Fig. 5.31). If m_2 is stationary, find expressions for (a) the string tension and (b) the period of the circular motion.

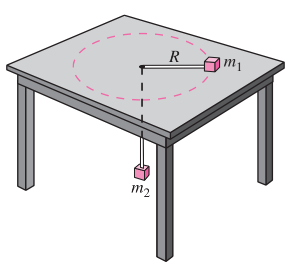


FIGURE 5.31 Problem 37

38. **BIO** Patients with severe leg breaks are often placed in *traction*, with an external force countering muscles that would pull too hard on the broken bones. In the arrangement shown in Fig. 5.32, the mass m is 4.8 kg, and the pulleys can be considered massless and frictionless. Find the horizontal traction force applied to the leg.

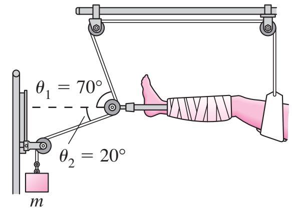


FIGURE 5.32 Problem 38

39. Riders on the "Great American Revolution" loop-the-loop roller coaster of Example 5.7 wear seatbelts as the roller coaster negotiates its 6.3-m-radius loop at 9.7 m/s. At the top of the loop, what are the magnitude and direction of the force exerted on a 60-kg rider (a) by the roller-coaster seat and (b) by the seatbelt? (c) What would happen if the rider unbuckled at this point?
40. A 45-kg skater rounds a 5.0-m-radius turn at 6.3 m/s. (a) What are the horizontal and vertical components of the force the ice exerts on her skate blades? (b) At what angle can she lean without falling over?
41. When a plane turns, it banks as shown in Fig. 5.33 to give the wings' lifting force a horizontal component that turns the plane. If a plane is flying level at 950 km/h and the banking angle is not to exceed 40°, what's the minimum curvature radius for the turn?

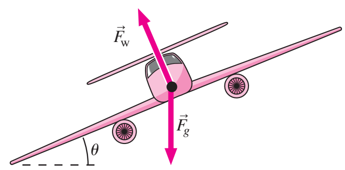


FIGURE 5.33 Problem 41

42. You whirl a bucket of water in a vertical circle of radius 85 cm. What's the minimum speed that will keep the water from falling out?
43. A child sleds down a 12° slope at constant speed. What's the frictional coefficient between slope and sled?
44. The handle of a 22-kg lawnmower makes a 35° angle with the horizontal. If the coefficient of friction between lawnmower and ground is 0.68, what magnitude of force, applied in the direction of the handle, is required to push the mower at constant velocity? Compare with the mower's weight.
45. Repeat Example 5.4, now assuming that the coefficient of kinetic friction between rock and ice is 0.057.
46. A bat crashes into the vertical front of an accelerating subway train. If the frictional coefficient between bat and train is 0.86, what's the minimum acceleration of the train that will allow the bat to remain in place?
47. The coefficient of static friction between steel train wheels and steel rails is 0.58. The engineer of a train moving at 140 km/h spots a stalled car on the tracks 150 m ahead. If he applies the brakes so the wheels don't slip, will the train stop in time?
48. A bug crawls outward from the center of a CD spinning at 200 revolutions per minute. The coefficient of static friction between the bug's sticky feet and the disc surface is 1.2. How far does the bug get from the center before slipping?
49. A 310-g paperback book rests on a 1.2-kg textbook. A force is applied to the textbook, and the two books accelerate together from rest to 96 cm/s in 0.42 s. The textbook is then brought to a

stop in 0.33 s, during which time the paperback slides off. Within what range does the coefficient of static friction between the two books lie?

50. Children sled down a 41-m-long hill inclined at 25°. At the bottom, the slope levels out. If the coefficient of friction is 0.12, how far do the children slide on the level ground?
51. In a typical front-wheel-drive car, 70% of the car's weight rides on the front wheels. If the coefficient of friction between tires and road is 0.61, what's the car's maximum acceleration?
52. A police officer investigating an accident estimates that a moving car hit a stationary car at 25 km/h. If the moving car left skid marks 47 m long, and if the coefficient of kinetic friction is 0.71, what was the initial speed of the moving car?
53. A slide inclined at 35° takes bathers into a swimming pool. With water sprayed onto the slide to make it essentially frictionless, a bather spends only one-third as much time on the slide as when it's dry. What's the coefficient of friction on the dry slide?
54. You try to move a heavy trunk, pushing down and forward at an angle of 50° below the horizontal. Show that, no matter how hard you push, it's impossible to budge the trunk if the coefficient of static friction exceeds 0.84.
55. A block is shoved up a 22° slope with an initial speed of 1.4 m/s. The coefficient of kinetic friction is 0.70. (a) How far up the slope will the block get? (b) Once stopped, will it slide back down?
56. At the end of a factory production line, boxes start from rest and slide down a 30° ramp 5.4 m long. If the slide can take no more than 3.3 s, what's the maximum allowed frictional coefficient?
57. You're in traffic court, arguing against a speeding citation. You entered a 210-m-radius banked turn designed for 80 km/h, which was also the posted speed limit. The road was icy, yet you stayed in your lane, so you argue that you must have been going at the design speed. But police measurements show there was a frictional coefficient $\mu = 0.15$ between tires and road. Is it possible you were speeding, and if so by how much?
58. A space station is in the shape of a hollow ring, 450 m in diameter (Fig. 5.34). At how many revolutions per minute should it rotate in order to simulate Earth's gravity—that is, so the normal force on an astronaut at the outer edge would equal the astronaut's weight on Earth?

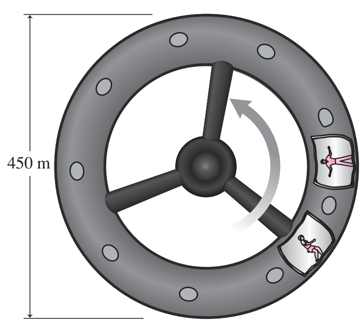


FIGURE 5.34 Problem 58

59. In a loop-the-loop roller coaster, show that a car moving too slowly would leave the track at an angle ϕ given by $\cos\phi = v^2/rg$, where ϕ is the angle made by a vertical line through the center of the circular track and a line from the center to the point where the car leaves the track.
60. Find an expression for the minimum frictional coefficient needed to keep a car with speed v on a banked turn of radius R designed for speed v_0.
61. An astronaut is training in an earthbound centrifuge that consists of a small chamber whirled horizontally at the end of a 5.1-m-long shaft. The astronaut places a notebook on the vertical wall of the chamber and it stays in place. If the coefficient of static friction is 0.62, what's the minimum rate at which the centrifuge must be revolving?
62. You stand on a spring scale at the north pole and again at the equator. Which scale reading will be lower, and by what percentage will it be lower than the higher reading? Assume g has the same value at pole and equator.
63. Driving in thick fog on a horizontal road, you spot a tractor-trailer truck jackknifed across the road. To avert a collision, you could brake to a stop or swerve in a circular arc, as suggested in Fig. 5.35. Which option offers the greater margin of safety? Assume that there is the same coefficient of static friction in both cases, and that you maintain constant speed if you swerve.

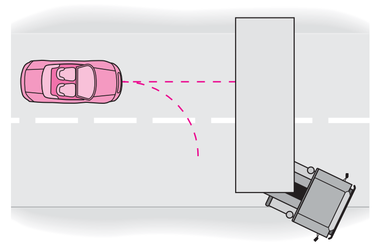


FIGURE 5.35 Problem 63

64. A block is projected up an incline at angle θ. It returns to its initial position with half its initial speed. Show that the coefficient of kinetic friction is $\mu_k = \frac{3}{5}\tan\theta$.
65. A 2.1-kg mass is connected to a spring with spring constant $k = 150$ N/m and unstretched length 18 cm. The two are mounted on a frictionless air table, with the free end of the spring attached to a frictionless pivot. The mass is set into circular motion at 1.4 m/s. Find the radius of its path.
66. Take $\mu_k = 0.75$ in Example 5.11, and plot the tension force in units of the trunk's weight, as a function of the rope angle θ (that is, plot T/mg versus θ). Use your plot to determine (a) the minimum tension necessary to move the trunk and (b) the angle at which this minimum tension should be applied.
67. Repeat the preceding problem for an arbitrary value of μ_k, by using calculus to find the minimum force needed to move the trunk with constant speed.
68. Moving through a liquid, an object of mass m experiences a resistive drag force proportional to its velocity, $F_{drag} = -bv$, where b is a constant. (a) Find an expression for the object's speed as a function of time, when it starts from rest and falls vertically through the liquid. (b) Show that it reaches a terminal velocity mg/b.
69. Suppose the object in Problem 68 had an initial velocity in the horizontal direction equal to the terminal speed, $v_{x0} = mg/b$. Show that the horizontal distance it can go is limited to $x_{max} = mv_{x0}/b$, and find an expression for its trajectory (y as a function of x).
70. A block is launched with speed v_0 up a slope making an angle θ with the horizontal; the coefficient of kinetic friction is μ_k. (a) Find an expression for the distance d the block travels along the slope. (b) Use calculus to determine the angle that minimizes d.
71. A florist asks you to make a window display with two hanging pots as shown in Fig. 5.36. The florist is adamant that the strings

be as invisible as possible, so you decide to use fishing line but want to use the thinnest line you can. Will fishing line that can withstand 100 N of tension work?

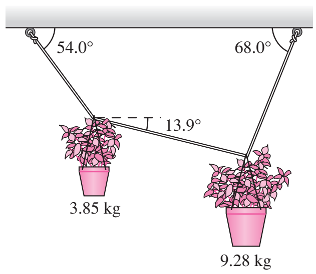


FIGURE 5.36 Problem 71

72. You're at the state fair. A sideshow barker claims that the star of the show can throw a 7.3-kg Olympic-style hammer "faster than a speeding bullet." You recall that bullets travel at several hundreds of meters per second. The burly hammer thrower whirls the hammer in a circle that you estimate to be 2.4 m in diameter. You guess the chain holding the hammer makes an angle of 10° with the horizontal. When the hammer flies off, is it really moving faster than a bullet?
73. One of the limiting factors in high-performance aircraft is the acceleration to which the pilot can be subjected without blacking out; it's measured in "gees," or multiples of the gravitational acceleration. The F-22 Raptor fighter can achieve Mach 1.8 (1.8 times the speed of sound, which is about 340 m/s). Suppose a pilot dives in a circle and pulls up. If the pilot can't exceed 6*g*, what's the tightest circle (smallest radius) in which the plane can turn?

Passage Problems

A *spiral* is an ice-skating position in which the skater glides on one foot with the other foot held above hip level. It's a required element in women's singles figure skating competition and is related to the arabesque performed in ballet. Figure 5.37 shows skater Sarah Hughes executing a spiral during her gold-medal performance at the Winter Olympics in Salt Lake City.

FIGURE 5.37 Passage Problems 74–77

74. From the photo, you can conclude that the skater is
    a. executing a turn to her left.
    b. executing a turn to her right.
    c. moving in a straight line out of the page.
75. The net force on the skater
    a. points to her left.
    b. points to her right.
    c. is zero.
76. If the skater were to execute the same maneuver but at higher speed, the tilt evident in the photo would be
    a. less.
    b. greater.
    c. unchanged.
77. The tilt angle θ that the skater's body makes with the vertical is given approximately by $\theta = \tan^{-1}(0.5)$. From this you can conclude that the skater's centripetal acceleration has approximate magnitude
    a. 0.
    b. 0.5 m/s^2.
    c. 5 m/s^2.
    d. can't be determined without knowing the skater's speed

Answers to Chapter Questions

Answer to Chapter Opening Question

The roller coaster is "falling," in the sense that it's *accelerating* downward, but that doesn't mean it has to be *moving* downward. The acceleration comes in response to the downward force of gravity and the normal force from the track, both of which are needed to keep the roller coaster moving in its circular path. Quantitatively, the net force is equal to the mass multiplied by the acceleration, so Newton's second law of motion is perfectly satisfied.

Answers to GOT IT? Questions

5.1. (c) Equal—but only because of the 45° slope. At larger angles, the tension would be greater than the weight; at smaller angles, less.
5.2. (a) 1 N; (b) 1 N—the left hand in Fig. 5.9 and the hook in this figure play exactly the same role, balancing the 1-N tension force in the rope.
5.3. (c) Greater because the chain is pulling downward, making the normal force greater than the log's weight.

6 Work, Energy, and Power

Climbing a mountain, these cyclists do work against gravity. Does that work depend on the route chosen?

New Concepts, New Skills

By the end of this chapter you should be able to

- Explain the concept of work and evaluate the work done by constant forces (6.1).
- Evaluate the work done by forces that vary with position (6.2).
- Explain the concept of kinetic energy and its relation to work (6.3).
- Describe the relation between energy and power (6.4).

Connecting Your Knowledge

- The idea of work builds on the concept of force. You should have a firm understanding of force from Chapters 4 and 5, and you should be able to distinguish individual forces from the net force on an object (4.1–4.3).
- The relation between work and kinetic energy follows from Newton's second law. Chapters 4 and 5 have given you a solid understanding of this law and its applications (4.2, 4.5, 5.1).

Figure 6.1*a* shows a skier starting from rest at the top of a uniform slope. What's the skier's speed at the bottom? You can solve this problem by applying Newton's second law to find the skier's constant acceleration and then the speed. But what about the skier in Fig. 6.1*b*? Here the slope is continuously changing and so is the acceleration. Constant-acceleration equations don't apply, so solving for the details of the skier's motion is difficult.

There are many cases where motion involves changing forces and accelerations. In this chapter, we introduce the important physical concepts of **work** and **energy**. These powerful concepts enable us to "shortcut" the detailed application of Newton's law to analyze these more complex situations. We begin with the concept of work.

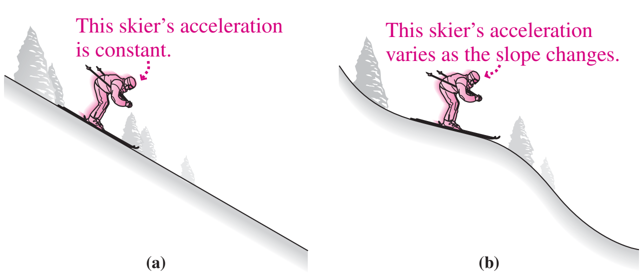


FIGURE 6.1 Two skiers.

6.1 Work

We all have an intuitive sense of the term *work*. Carrying a piece of furniture upstairs involves work. The heavier the furniture or the higher the stairs, the greater the work. Pushing a stalled car involves work. The harder you push or the farther you push, the more work you do. The precise definition of work reflects our intuition:

For an object moving in one dimension, the work W done on the object by constant applied force $\vec{F}$ is

$$W = F_x \Delta x \tag{6.1}$$

where F_x is the component of the force in the direction of the object's motion and Δx is the object's displacement.

The force $\vec{F}$ need not be the net force. If you're interested, for example, in how much work *you* must do to drag a heavy box across the floor, then $\vec{F}$ is the force *you apply* and W is the work *you do*.

Equation 6.1 shows that the SI unit of work is the newton-meter $(\text{N}\cdot\text{m})$. One newton-meter is given the name **joule**, in honor of the 19th-century British physicist and brewer James Joule.

According to Equation 6.1, the person pushing the car in Fig. 6.2*a* does work equal to the force he applies times the distance the car moves. But the person pulling the suitcase in Fig. 6.2*b* does work equal to only the horizontal component of the force she applies times the distance the suitcase moves. Furthermore, by our definition, the waiter of Fig. 6.2*c* does no work on the tray. Why not? Because the force on the tray is vertical while the tray's displacement is horizontal; there's no component of force in the direction of the tray's motion.

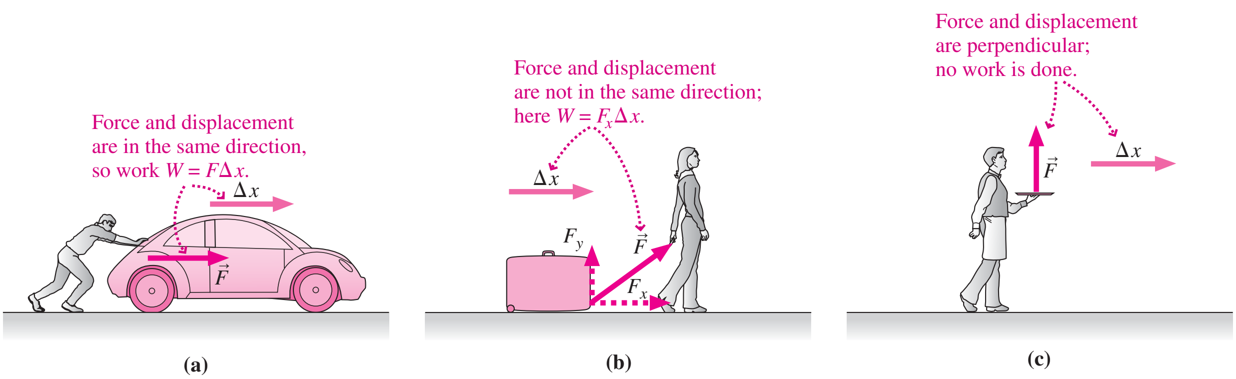


FIGURE 6.2 Work depends on the orientation of force and displacement.

Work can be positive or negative (Fig. 6.3). When a force acts in the same general direction as the motion, it does positive work. A force acting at 90° to the motion does no work. And when a force acts to oppose motion, it does negative work.

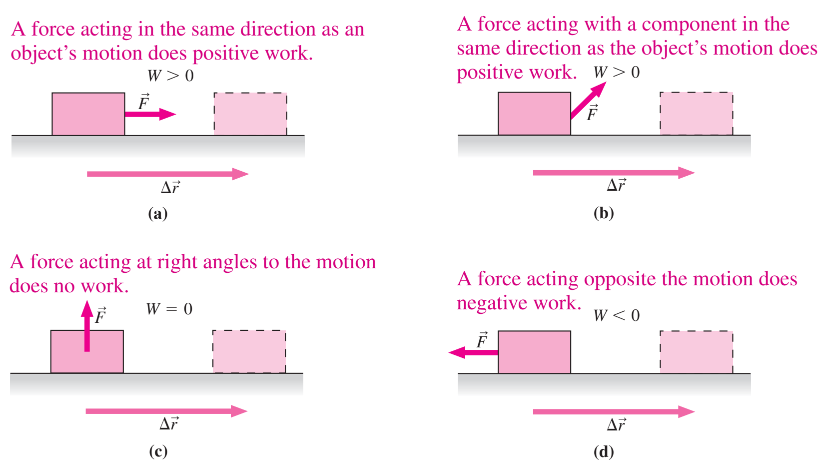


FIGURE 6.3 The sign of the work depends on the relative directions of force and motion. We use $\Delta\vec{r}$ here to indicate that the displacement can be any vector.

EXAMPLE 6.1 Calculating Work: Pushing a Car

The person in Fig. 6.2*a* pushes with a force of 650 N, moving the car a distance of 4.3 m. How much work does he do?

INTERPRET This problem is about work. We identify the car as the object *on* which the work is done and the person as the agent *doing* the work.

DEVELOP Figure 6.2*a* is our drawing. Equation 6.1, $W = F_x \Delta x$, is the relevant equation, so our plan is to apply that equation. The force is in the same direction as the displacement, so 650 N is the component we need.

EVALUATE We apply Equation 6.1 to get

$$W = F_x \Delta x = (650\text{ N})(4.3\text{ m}) = 2.8\text{ kJ}$$

ASSESS Make sense? The units work out, with newtons times meters giving joules—here expressed in kilojoules for convenience. ■

EXAMPLE 6.2 Calculating Work: Pulling a Suitcase

The airline passenger in Fig. 6.2*b* exerts a 60-N force on her suitcase, pulling at 35° to the horizontal. How much work does she do in pulling the suitcase 45 m on a level floor?

INTERPRET Again, this example is about work—here done *by* the passenger *on* the suitcase.

DEVELOP Equation 6.1, $W = F_x \Delta x$, applies here, but because the displacement is horizontal while the force isn't, we need to find the horizontal force component. We've redrawn the force vector in Fig. 6.4 to determine F_x.

EVALUATE Applying Equation 6.1 to the *x*-component from Fig. 6.4, we get

$$W = F_x \Delta x = [(60\text{ N})(\cos 35°)](45\text{ m}) = 2.2\text{ kJ}$$

ASSESS The answer of 2.2 kJ is less than the product of 60 N and 45 m, and that makes sense because only the *x*-component of that 60-N force contributes to the work. ■

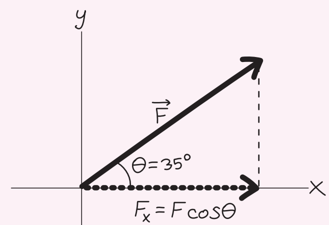


FIGURE 6.4 Our sketch for Example 6.2.

Work and the Scalar Product

Work is a *scalar* quantity; it's specified completely by a single number and has no direction. But Fig. 6.2 shows clearly that work involves a relation between two *vectors*: the force $\vec{F}$ and the displacement, designated more generally by $\Delta\vec{r}$. If θ is the angle between these two vectors, then the component of the force along the direction of motion is $F\cos\theta$, and the work is

$$W = (F\cos\theta)(\Delta r) = F\,\Delta r\cos\theta \quad (6.2)$$

This equation is a generalization of our definition 6.1. If we choose the *x*-axis along $\Delta\vec{r}$, then $\Delta r = \Delta x$ and $F\cos\theta = F_x$, so we recover Equation 6.1.

Equation 6.2 shows that work is the product of the magnitudes of the vectors $\vec{F}$ and $\Delta\vec{r}$ and the cosine of the angle between them. This combination occurs so often that it's given a special name: the **scalar product** of two vectors.

> The scalar product of any two vectors $\vec{A}$ and $\vec{B}$ is defined as
> $$\vec{A}\cdot\vec{B} = AB\cos\theta \quad (6.3)$$
> where A and B are the magnitudes of the vectors and θ is the angle between them.

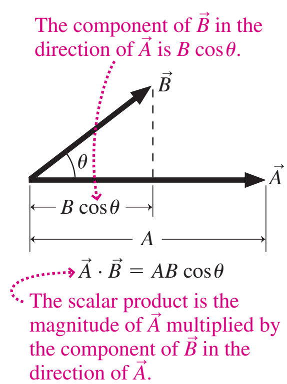


FIGURE 6.5 Geometric interpretation of the scalar product.

The term *scalar product* should remind you that $\vec{A}\cdot\vec{B}$ is itself a *scalar*, even though it's formed from two vectors. A centered dot designates the scalar product; for this reason, it's also called the **dot product**. Figure 6.5 gives a geometric interpretation.

The scalar product is commutative: $\vec{A}\cdot\vec{B} = \vec{B}\cdot\vec{A}$, and it's also distributive: $\vec{A}\cdot(\vec{B}+\vec{C}) = \vec{A}\cdot\vec{B} + \vec{A}\cdot\vec{C}$. With vectors expressed in unit vector notation, Problem 46 shows how the distributive law gives a simple form for the scalar product. If $\vec{A} = A_x\hat{\imath} + A_y\hat{\jmath} + A_z\hat{k}$ and $\vec{B} = B_x\hat{\imath} + B_y\hat{\imath} + B_z\hat{k}$, then

$$\vec{A} \cdot \vec{B} = A_xB_x + A_yB_y + A_zB_z \tag{6.4}$$

Comparing Equation 6.2 with Equation 6.3 shows that the work done by a constant force $\vec{F}$ moving an object through a straight-line displacement $\Delta\vec{r}$ is

$$W = \vec{F} \cdot \Delta\vec{r} \tag{6.5}$$

As the examples below show, either Equation 6.3 or Equation 6.4 can be used in evaluating the dot product in this expression for work.

EXAMPLE 6.3 Work and the Scalar Product: A Tugboat

A tugboat pushes a cruise ship with force $\vec{F} = 1.2\hat{\imath} + 2.3\hat{\jmath}$ MN, moving the ship along a straight path with displacement $\Delta\vec{r} = 380\hat{\imath} + 460\hat{\jmath}$ m. Find (a) the work done by the tugboat and (b) the angle between the force and displacement.

INTERPRET Part (a) is about calculating work given force and displacement in unit vector notation. Part (b) is less obvious, but knowing that work involves the angle between force and displacement provides a clue, suggesting that the answer to (a) may lead us to (b).

DEVELOP Figure 6.6 is a sketch of the two vectors that will serve as a check on our final answer. For (a), we want to use Equation 6.5, $W = \vec{F} \cdot \Delta\vec{r}$, with the scalar product given by Equation 6.4. That will give us the work W. We also have the vectors $\vec{F}$ and $\Delta\vec{r}$, so we can find their magnitudes. That suggests a strategy for (b): Given the work and the vector magnitudes, we can write Equation 6.3 with a single unknown, the angle θ that we're asked to find.

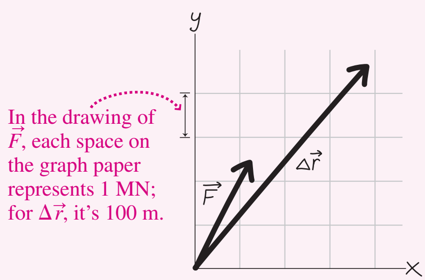


FIGURE 6.6 Our sketch of the vectors in Example 6.3.

EVALUATE For (a), we use Equations 6.5 and 6.4 to write

$$\begin{aligned} W &= \vec{F} \cdot \Delta\vec{r} = F_x\,\Delta x + F_y\,\Delta y \\ &= (1.2\text{ MN})(380\text{ m}) + (2.3\text{ MN})(460\text{ m}) = 1510\text{ MJ} \end{aligned}$$

The first equality is from Equation 6.5; the second gives the scalar product in unit vector form from Equation 6.4. Δx and Δy are the components of the displacement $\Delta\vec{r}$. Now that we have the work, we can get the angle. The magnitude of a vector comes from the Pythagorean theorem, as expressed in Equation 3.1. Then we have $F = \sqrt{F_x^2 + F_y^2} = \sqrt{(1.2\text{ MN})^2 + (2.3\text{ MN})^2} = 2.59$ MN; a similar calculation gives $\Delta r = 597$ m. Now we solve Equation 6.3 for θ:

$$\theta = \cos^{-1}\left(\frac{W}{F\,\Delta r}\right) = \cos^{-1}\left(\frac{1510\text{ MJ}}{(2.59\text{ MN})(597\text{ m})}\right) = 12°$$

ASSESS This small angle is consistent with our sketch in Fig. 6.6. And it makes good physical sense: A tugboat is most efficient when pushing in the direction the ship is supposed to go. Note how the units work out in that last calculation: MJ in the numerator and MN·m in the denominator. But 1 N·m is 1 J, so that's MJ in the denominator, too, giving the dimensionless cosine. ■

GOT IT? 6.1 Two objects are each displaced the same distance, one by a force F pushing in the direction of motion and the other by a force $2F$ pushing at 45° to the direction of motion. Which force does more work?

6.2 Forces That Vary

Often the force applied to an object varies with position. Important examples include electric and gravitational forces, which vary with the distance between interacting objects. The force of a spring that we encountered in Chapter 4 provides another example; as the spring stretches, the force increases.

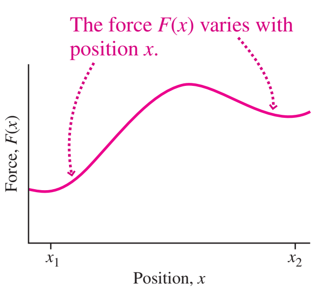


FIGURE 6.7 A varying force.

Figure 6.7 is a plot of a force F that varies with position x. We want to find the work done as an object moves from x_1 to x_2. We can't simply write $F(x_2 - x_1)$; since the force varies, there's no single value for F. What we can do, though, is divide the region into rectangles of width Δx, as shown in Fig. 6.8*a*. If we make Δx small enough, the force will be nearly constant over the width of each rectangle (Fig. 6.8*b*). Then the work ΔW done in moving the width Δx of one such rectangle is approximately $F(x)\,\Delta x$, where $F(x)$ is the force at the midpoint x of that rectangle. We write $F(x)$ to show explicitly that the force is a function of position. Note that the quantity $F(x)\,\Delta x$ is the area of the rectangle expressed in the appropriate units (N·m, or, equivalently, J).

Suppose there are N rectangles. Let x_i be the midpoint of the ith rectangle. Then the total work done in moving from x_1 to x_2 is given approximately by the sum of the individual amounts of work ΔW_i associated with each rectangle, or

$$W \simeq \sum_{i=1}^{N} \Delta W_i = \sum_{i=1}^{N} F(x_i)\,\Delta x \tag{6.6}$$

How good is this approximation? That depends on how small we make the rectangles. Suppose we let them get arbitrarily small. Then the number of rectangles must grow arbitrarily large. In the limit of infinitely many infinitesimally small rectangles, the approximation in Equation 6.6 becomes exact (Fig. 6.8*c*). Then we have

$$W = \lim_{\Delta x \to 0} \sum_i F(x_i)\,\Delta x \tag{6.7}$$

where the sum is over all the infinitesimal rectangles between x_1 and x_2. The quantity on the right-hand side of Equation 6.7 is the **definite integral** of the function $F(x)$ over the interval from x_1 to x_2. We introduce special symbolism for the limiting process of Equation 6.7:

$$W = \int_{x_1}^{x_2} F(x)\,dx \qquad \left(\begin{array}{c}\text{work done by a varying}\\ \text{force in one dimension}\end{array}\right) \tag{6.8}$$

Equation 6.8 means exactly the same thing as Equation 6.7: It tells us to divide the interval from x_1 to x_2 into many small rectangles of width Δx, to multiply the value of the function $F(x)$ at each rectangle by the width Δx, and to sum those products. As we take arbitrarily many arbitrarily small rectangles, the result of this process gives us the value of the definite integral. You can think of the symbol $\int$ in Equation 6.8 as standing for "sum" and the symbol dx as a limiting case of arbitrarily small Δx. The definite integral has a simple geometric interpretation: It's the area under the curve $F(x)$ between the limits x_1 and x_2 (Fig. 6.8*c*).

Computers approximate the infinite sum implied in Equation 6.8 using a large number of very small rectangles. But calculus often provides a better way.

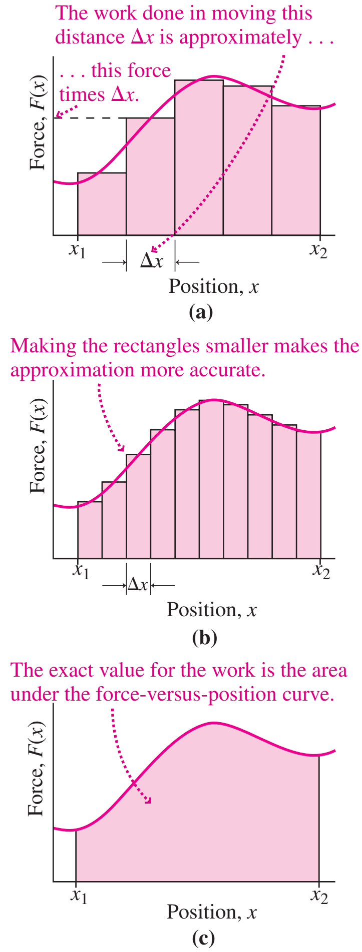


FIGURE 6.8 Work done by a varying force.

TACTICS 6.1 **Integrating**

In your calculus course you've learned, or will soon learn, that integrals and derivatives are inverses. In Section 2.2, you saw that the derivative of x^n is nx^{n-1}; therefore, the integral of x^n is $(x^{n+1})/(n+1)$, as you can verify by differentiating. We determine the value of a definite integral by evaluating this expression at upper and lower limits and subtracting:

$$\int_{x_1}^{x_2} x^n\,dx = \left.\frac{x^{n+1}}{n+1}\right|_{x_1}^{x_2} = \frac{x_2^{n+1}}{n+1} - \frac{x_1^{n+1}}{n+1} \tag{6.9}$$

where the middle term is a shorthand notation for the difference given in the rightmost term. A review of integration and a table of common integrals are given in Appendix A.

Stretching a Spring

A spring provides an important example of a force that varies with position. We've seen that an ideal spring exerts a force proportional to its displacement from equilibrium: $F = -kx$, where k is the spring constant and the minus sign shows that the spring force is opposite the direction of the displacement. It's not just coiled springs that we're interested in here; many physical systems, from molecules to skyscrapers to stars, behave as though they contain springs. The work and energy considerations we develop here apply to those systems as well.

The force exerted *by* a stretched spring is $-kx$, so the force exerted *on* the spring *by* the external stretching force is $+kx$. If we let $x = 0$ be one end of the spring at equilibrium

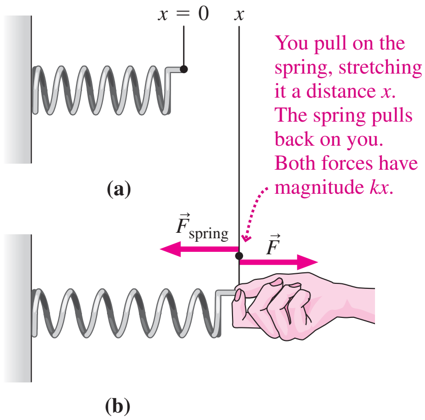


FIGURE 6.9 Stretching a spring.

and if we hold the other end fixed and pull the spring until its free end is at a new position x, as shown in Fig. 6.9, then Equation 6.8 shows that the work done on the spring by the external force is

$$W = \int_0^x F(x)\,dx = \int_0^x kx\,dx = \tfrac{1}{2}kx^2\Big|_0^x = \tfrac{1}{2}kx^2 - \tfrac{1}{2}k(0)^2 = \tfrac{1}{2}kx^2 \qquad (6.10)$$

where we used Equation 6.9 to evaluate the integral. The more we stretch the spring, the greater the force we must apply—and that means we must do more work for a given amount of additional stretch. Figure 6.10 shows graphically why the work depends quadratically on the displacement. Although we used the word *stretch* in developing Equation 6.10, the result applies equally to compressing a spring a distance x from equilibrium.

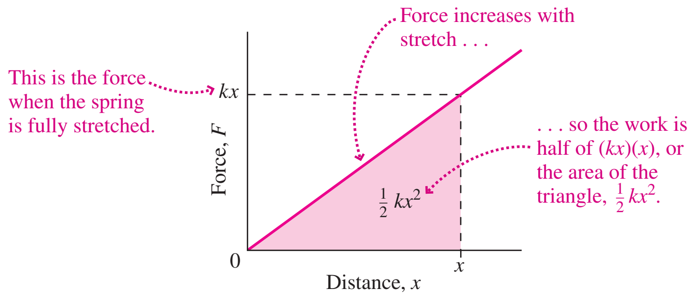


FIGURE 6.10 Work done in stretching a spring.

EXAMPLE 6.4 The Spring Force: Bungee Jumping

An elastic cord used in bungee jumping is normally 11 m long and has spring constant $k = 250$ N/m. At the lowest point in a jump, the cord length has doubled. How much work has been done on the cord?

INTERPRET The bungee cord behaves like a spring—as we can tell because we're given its spring constant. So this example is about the work done in stretching a spring. We're told the 11-m-long cord length doubles in length, so it's stretched another 11 m.

DEVELOP Equation 6.10 gives the work done in stretching the cord a distance x from its unstretched configuration.

EVALUATE Applying Equation 6.10 gives

$$W = \tfrac{1}{2}kx^2 = \left(\tfrac{1}{2}\right)(250\ \text{N/m})(11\ \text{m})^2 = 15\ \text{kN}\cdot\text{m} = 15\ \text{kJ}$$

ASSESS As you'll see shortly, that's just about equal to the work done by gravity on a 70-kg person dropping the 22-m distance from the attachment point of the cord to its full stretched extent. We'll see in the next chapter why this is no coincidence. ■

CONCEPTUAL EXAMPLE 6.1 Bungee Details

In Example 6.4, is the work done as the cord stretches its final meter greater than, less than, or equal to the work done in the first meter of stretch?

EVALUATE We're asked to compare the work done during the beginning and end of the bungee cord's stretch. We know that work is the area under the force-distance curve. We've sketched the force-distance curve in Fig. 6.11, highlighting the first and last meters. Clearly, the area associated with the last meter of stretch is much larger. Therefore, the work is greater.

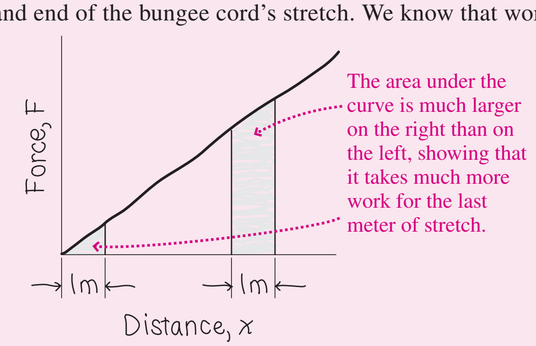


FIGURE 6.11 Conceptual Example 6.1.

ASSESS Makes sense! Once the cord has stretched 10 m, it exerts a large force. That makes it much harder to stretch farther—and thus the final meter requires a lot of work. The first meter takes much less work because at first the cord exerts very little force.

MAKING THE CONNECTION Find the work involved in stretching during the first and last meters, and compare.

EVALUATE We can use Equation 6.10, but instead of the limits 0 and x, we'll use 0 and 1 m for the first meter of stretch, and 10 m and 11 m for the last meter. The results are 125 J and 2.6 kJ. Stretching the final meter takes more than 20 times the work required for the first meter!

EXAMPLE 6.5 A Varying Friction Force: Rough Sliding

Workers pushing a 180-kg trunk across a level floor encounter a 10-m-long region where the floor becomes increasingly rough. The coefficient of kinetic friction here is given by $\mu_k = \mu_0 + ax^2$, where $\mu_0 = 0.17$, $a = 0.0062\ \text{m}^{-2}$, and x is the distance from the beginning of the rough region. How much work does it take to push the trunk across the region?

INTERPRET This example asks for the work needed to push the trunk. To move the trunk at constant speed, the workers must apply a force equal in magnitude to the frictional force. That force varies with position, so we're dealing with a varying force.

DEVELOP Our drawing, the force-position curve in Fig. 6.12, emphasizes that we have a varying force. Therefore, we have to integrate using Equation 6.8, $W = \int_{x_1}^{x_2} F(x)\,dx$. And we need to know the frictional force, which is given by Equation 5.3: $f_k = \mu_k n$. On a level floor, the normal force is equal in magnitude to the weight, mg, so Equation 6.8 becomes $W = \int_{x_1}^{x_2} \mu_k mg\,dx = \int_{x_1}^{x_2} mg(\mu_0 + ax^2)\,dx$.

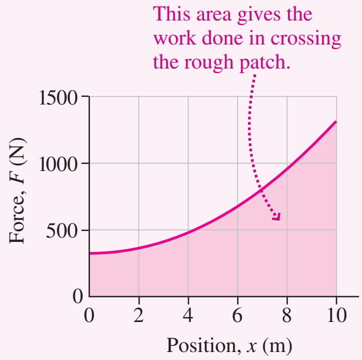


FIGURE 6.12 Force versus position for Example 6.5.

EVALUATE We evaluate the integral using Equation 6.9. Actually, we have two integrals here: one of dx alone and the other of $x^2\,dx$. According to Equation 6.9, the former gives x and the latter $x^3/3$. So the result is

$$W = \int_{x_1}^{x_2} mg(\mu_0 + ax^2)\,dx = mg\left(\mu_0 x + \tfrac{1}{3}ax^3\right)\Big|_{x_1}^{x_2}$$

$$= mg\left[\left(\mu_0 x_2 + \tfrac{1}{3}ax_2^3\right) - \left(\mu_0 x_1 + \tfrac{1}{3}ax_1^3\right)\right]$$

Putting in the values given for μ_0, a, and m, using $g = 9.8\ \text{m/s}^2$, and taking $x_1 = 0$ and $x_2 = 10$ m for the endpoints of the rough interval, we get 6.6 kJ for our answer.

ASSESS Is this answer reasonable? Figure 6.12 shows that the maximum force is approximately 1.3 kN. If this force acted over the entire 10-m interval, the work would be about 13 kJ. But it's approximately half that because the coefficient of kinetic friction and therefore the force start out quite low. You can see that the area under the curve in Fig. 6.12 is about half the area of the full rectangle, so our answer of 6.6 kJ makes sense. ■

✓TIP Don't Just Multiply!

When force depends on position, there's no single value for the force, so you can't just multiply force by distance to get work. You need either to integrate, as in Example 6.5, or to use a result that's been derived by integration, as with the equation $W = \frac{1}{2}kx^2$ used in Example 6.4.

Force and Work in Two and Three Dimensions

Sometimes a force varies in both magnitude and direction or an object moves on a curved path; either way, the angle between force and motion may vary. Then we have to take the scalar products of the force $\vec{F}$ with small displacements $\Delta\vec{r}$, writing $\Delta W = \vec{F}\cdot\Delta\vec{r}$ for the work involved in one such small displacement. Adding them all gives the total work, which in the limit of very small displacements becomes a **line integral**:

$$W = \int_{\vec{r}_1}^{\vec{r}_2} \vec{F}\cdot d\vec{r} \qquad (6.11)$$

where the integral is taken over a specific path between positions $\vec{r}_1$ and $\vec{r}_2$. We won't pursue line integrals further here, but they'll be useful in later chapters.

Work Done Against Gravity

When an object moves upward or downward on an arbitrary path, the angle between its displacement and the gravitational force varies. But here we don't really need the line integral of Equation 6.11 because we can consider any path as consisting of small horizontal and vertical steps (Fig. 6.13). Only the vertical steps contribute to the work, which then becomes simply $W = mgh$, where h is the total height the object rises—a result that's

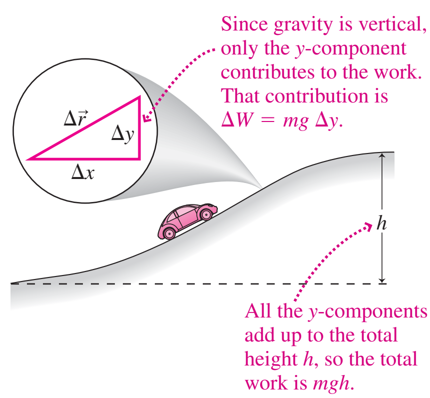


FIGURE 6.13 A car climbs a hill with varying slope.

independent of the particular path taken. (As in our earlier work with gravity, this result holds only near Earth's surface, where we can neglect the variation in gravity with height.)

GOT IT? 6.2 Three forces have magnitudes in newtons that are numerically equal to these quantities: (a) x, (b) x^2, and (c) $\sqrt{x}$, where x is the position in meters. Each force acts on an object as it moves from $x = 0$ to $x = 1$ m. Notice that all three forces have the same values at the two endpoints—namely, 0 N and 1 N. Which force does the most work? Which does the least?

6.3 Kinetic Energy

Closely related to work is **energy**—one of the most important concepts in all of physics. Here we introduce the energy associated with motion, or **kinetic energy**. Our goal is to relate kinetic energy and work. We'll start by evaluating the net work done on an object, and then apply Newton's second law. The net work is done by all the forces acting on an object, so we use the net force in our expression for work. We'll consider one-dimensional motion, with force and displacement along the same line. In that case, Equation 6.8 gives the net work:

$$W_{\text{net}} = \int F_{\text{net}}\,dx$$

But the net force can be written in terms of Newton's second law: $F_{\text{net}} = ma$, or $F_{\text{net}} = m\,dv/dt$, so

$$W_{\text{net}} = \int m\frac{dv}{dt}\,dx$$

The quantities dv, dt, and dx arose as the limits of small numbers Δv, Δt, and Δx. In calculus, you've seen that the limit of a product or quotient is the product or quotient of the individual terms involved. For these reasons, we can rearrange the symbols dv, dt, and dx to rewrite our expression in the form

$$W_{\text{net}} = \int m\,dv\,\frac{dx}{dt}$$

But $dx/dt = v$, so we have

$$W_{\text{net}} = \int mv\,dv$$

The integral here is like $\int x\,dx$, which we evaluate by raising the exponent and dividing by the new exponent. What about the limits? Suppose our object starts at some speed v_1 and ends at v_2. Then we have

$$W_{\text{net}} = \int_{v_1}^{v_2} mv\,dv = \tfrac{1}{2}mv^2\Big|_{v_1}^{v_2} = \tfrac{1}{2}mv_2^2 - \tfrac{1}{2}mv_1^2 \qquad (6.12)$$

Equation 6.12 shows that an object has associated with it a quantity $\frac{1}{2}mv^2$ that changes when, and only when, net work is done on the object. This quantity plays a vital role in physics and is called the object's **kinetic energy**:

The kinetic energy K of an object of mass m moving at speed v is

$$K = \tfrac{1}{2}mv^2 \qquad (6.13)$$

Like velocity, kinetic energy is a relative term; its value depends on the reference frame in which it's measured. But unlike velocity, kinetic energy is a *scalar*. And since it depends on the *square* of the velocity, kinetic energy is never negative. All moving objects possess kinetic energy.

Equation 6.12 equates the change in an object's kinetic energy with the net work done on the object, a result known as the **work-energy theorem**:

Work-energy theorem: The change in an object's kinetic energy is equal to the net work done on the object:

$$\Delta K = W_{\text{net}} \tag{6.14}$$

Equations 6.12 and 6.14 are equivalent statements of the work-energy theorem.

We've seen that work can be positive or negative; so, therefore, can *changes* in kinetic energy. If I stop a moving object, for example, I reduce its kinetic energy from $\frac{1}{2}mv^2$ to zero—a change $\Delta K = -\frac{1}{2}mv^2$. So I do negative work by applying a force directed opposite to the motion. By Newton's third law, the object exerts an equal but oppositely directed force on me, therefore doing positive work $\frac{1}{2}mv^2$ *on me*. So an object of mass m moving at speed v can do work equal to its initial kinetic energy, $\frac{1}{2}mv^2$, if it's brought to rest.

EXAMPLE 6.6 Work and Kinetic Energy: Passing Zone

A 1400-kg car enters a passing zone and accelerates from 70 to 95 km/h. (a) How much work is done on the car? (b) If the car then brakes to a stop, how much work is done on it?

INTERPRET Here we're asked about work, but we aren't given any forces as we were in previous examples. However, we now know the work-energy theorem, which relates work and kinetic energy. Kinetic energy depends on speed, which we're given. So this is a problem involving the work-energy theorem.

DEVELOP The relevant equation is Equation 6.14 or its more explicit form, Equation 6.12. Since we're given speeds, it's easiest to work with Equation 6.12.

EVALUATE For (a), Equation 6.12 gives

$$\begin{aligned} W_{\text{net}} &= \tfrac{1}{2}mv_2^2 - \tfrac{1}{2}mv_1^2 = \tfrac{1}{2}m\left(v_2^2 - v_1^2\right) \\ &= \left(\tfrac{1}{2}\right)(1400\ \text{kg})\left[(26.4\ \text{m/s})^2 - (19.4\ \text{m/s})^2\right] = 220\ \text{kJ} \end{aligned}$$

where we converted the speeds to meters per second before doing the calculation. The work-energy theorem applies equally to the braking car in (b), for which $v_1 = 26.4$ m/s and $v_2 = 0$. Here we have

$$\begin{aligned} W_{\text{net}} &= \tfrac{1}{2}m\left(v_2^2 - v_1^2\right) = \left(\tfrac{1}{2}\right)(1400\ \text{kg})\left[0^2 - (26.4\ \text{m/s})^2\right] \\ &= -490\ \text{kJ} \end{aligned}$$

ASSESS Make sense? Yes. There's a greater change in speed and thus in kinetic energy in the braking case, so the magnitude of the work involved is greater. Our second answer is negative because stopping the car means applying a force that *opposes* its motion—and that means negative work done on the car. ■

GOT IT? 6.3 For each situation, tell whether the net work done on a soccer ball is positive, negative, or zero. Justify your answers using the work-energy theorem. (a) You carry the ball out to the field, walking at constant speed. (b) You kick the stationary ball, starting it flying through the air. (c) The ball rolls along the field, gradually coming to a halt.

Energy Units

Since work is equal to the change in kinetic energy, the units of energy are the same as those of work. In SI, the unit of energy is therefore the joule, equal to 1 newton-meter. In science, engineering, and everyday life, though, you'll encounter other energy units. Scientific units include the **erg**, used in the centimeter-gram-second system of units and equal to 10^{-7} J; the **electron-volt**, used in nuclear, atomic, and molecular physics; and the **calorie**, used in thermodynamics and to describe the energies of chemical reactions. English units include the **foot-pound** and the **British thermal unit** (Btu); the latter is commonly used in engineering of heating and cooling systems. Your electric company charges you for energy use in **kilowatt-hours** $(\text{kW}\cdot\text{h})$; we'll see in the next section how this unit relates to the SI joule. Appendix C contains an extensive table of energy units and conversion factors as well as the energy contents of common fuels.

6.4 Power

Climbing a flight of stairs requires the same amount of work no matter how fast you go. But it's harder to *run* up the stairs than to walk. Harder in what sense? In the sense that you do the same work in a shorter time; the *rate* at which you do the work is greater. We define **power** as the rate of doing work:

If an amount of work ΔW is done in time Δt, then the average power $\overline{P}$ is

$$\overline{P} = \frac{\Delta W}{\Delta t} \qquad \text{(average power)} \qquad (6.15)$$

Often the rate of doing work varies with time. Then we define the **instantaneous power** as the average power taken in the limit of an arbitrarily small time interval Δt:

$$P = \lim_{\Delta t \to 0} \frac{\Delta W}{\Delta t} = \frac{dW}{dt} \qquad \text{(instantaneous power)} \qquad (6.16)$$

Equations 6.15 and 6.16 both show that the units of power are joules/second. One J/s is given the name **watt** (W) in honor of James Watt, a Scottish engineer and inventor who was instrumental in developing the steam engine as a practical power source. Watt himself defined another unit, the horsepower. One horsepower (hp) is about 746 J/s or 746 W.

EXAMPLE 6.7 Power: Climbing Mount Washington

A 55-kg hiker ascends New Hampshire's Mount Washington, making the vertical rise of 1300 m in 2 h. A 1500-kg car drives up the Mount Washington Auto Road, taking half an hour. Neglecting energy lost to friction, what's the average power output for each?

INTERPRET This example is about power, which we identify as the *rate* at which hiker and car expend energy. So we need to know the work done by each and the corresponding time.

DEVELOP Equation 6.15, $\overline{P} = \Delta W/\Delta t$, is relevant, since we want the *average* power. To use this equation we'll need to find the work done in climbing the mountain. As you learned in Section 6.2, work done against gravity is independent of the path taken and is given by mgh, where h is the total height of the climb.

EVALUATE We apply Equation 6.15 in the two cases:

$$\overline{P}_{\text{hiker}} = \frac{\Delta W}{\Delta t} = \frac{(55\text{ kg})(9.8\text{ m/s}^2)(1300\text{ m})}{(2.0\text{ h})(3600\text{ s/h})} = 97\text{ W}$$

$$\overline{P}_{\text{car}} = \frac{\Delta W}{\Delta t} = \frac{(1500\text{ kg})(9.8\text{ m/s}^2)(1300\text{ m})}{(0.50\text{ h})(3600\text{ s/h})} = 11\text{ kW}$$

ASSESS Do these values make sense? A power of 97 W is typical of the sustained long-term output of the human body, as you can confirm by considering a typical daily diet of 2000 "calories" (actually kilocalories; see Exercise 31). The car's output amounts to 14 hp, which you may find low, given that the car's engine is probably rated at several hundred horsepower. But cars are notoriously inefficient machines, with only a small fraction of the rated horsepower available to do useful work. Most of the rest is lost to friction and heating. ■

When power is constant, so the average power and instantaneous power are the same, Equation 6.15 shows that the amount of work W done in time Δt is

$$W = P\,\Delta t \qquad (6.17)$$

When power isn't constant, we can consider small amounts of work ΔW, each taken over so small a time interval Δt that the power is nearly constant. Adding all these amounts of work and taking the limit as Δt becomes arbitrarily small, we have

$$W = \lim_{\Delta t \to 0} \sum P\,\Delta t = \int_{t_1}^{t_2} P\,dt \qquad (6.18)$$

where t_1 and t_2 are the beginning and end of the time interval over which we calculate the work.

EXAMPLE 6.8 Energy and Power: Yankee Stadium

Each of the 884 floodlights at Yankee Stadium uses electrical energy at the rate of 1650 W. How much does it cost to run these lights during a 5-h night game, if electricity costs 21¢/kW·h?

INTERPRET We're given a single floodlight's power consumption and the cost of electricity per kilowatt-hour, a unit of energy. So this problem is about calculating energy given power and time, with a little economics thrown in.

DEVELOP Since the power is constant, we can calculate the energy used over time with Equation 6.17, $W = P\,\Delta t$.

EVALUATE At 1.65 kW each, all 884 floodlights use energy at the rate $(884)(1.65\text{ kW}) = 1459\text{ kW}$. Then the total for a 5-h game is

$$W = P\,\Delta t = (1459\text{ kW})(5\text{ h}) = 7295\text{ kW}\cdot\text{h}$$

The cost is then $(7295\text{ kW}\cdot\text{h})(\$0.21/\text{kW}\cdot\text{h}) = \$1532$.

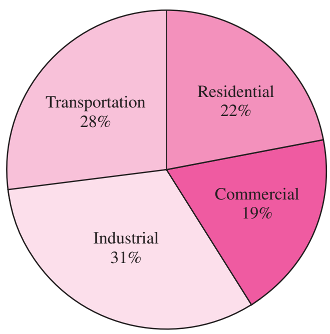
ASSESS Do we have the right units here? Yes. With power in kilowatts and time in hours, the energy comes out immediately in kilowatt-hours. ■

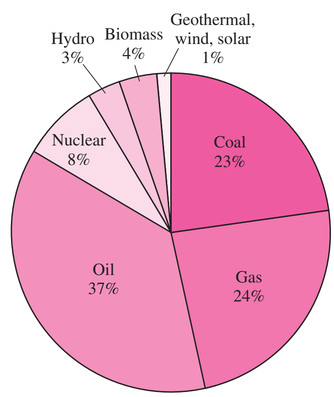
APPLICATION Energy and Society

Humankind's rate of energy consumption is a matter of concern, especially given our dependence on fossil fuels whose carbon dioxide emissions threaten global climate change. Just how rapidly are we using energy?

Example 6.7 suggests that the average power output of the human body is approximately 100 W. Before our species harnessed fire and domesticated animals, that was all the power available to each of us. But in today's high-energy societies, we use energy at a much greater rate. For the average citizen of the United States in the early 21st century, for example, the rate of energy consumption is about 11 kW—the equivalent of more than a hundred human bodies. The rate is lower in most other industrialized countries, but it still amounts to many tens of human bodies' worth.

What do we do with all that energy? And where does it come from? The first pie chart shows that most goes for industry and transportation, with lesser amounts used in the residential and commercial sectors. The second chart is a stark reminder that our energy supply is neither diversified nor renewable, with some 84% coming from the fossil fuels coal, oil, and natural gas. That's going to have to change in the coming decades, as a result of both limited fossil-fuel resources and the environmental consequences of fossil-fuel combustion. Much of what you learn in an introductory physics course has direct relevance to the energy challenges we face today.





✓TIP Don't Confuse Energy and Power

Is that 11-kW per-capita energy consumption per year, per day, or what? That question reflects a common confusion of energy and power. Power is the *rate* of energy use; it doesn't need any "per time" attached to it. It's just 11 kW, period.

Power and Velocity

We can derive an expression relating power, applied force, and velocity by noting that the work dW done by a force $\vec{F}$ acting on an object that undergoes an infinitesimal displacement $d\vec{r}$ follows from Equation 6.5:

$$dW = \vec{F} \cdot d\vec{r}$$

Dividing both sides by the associated time interval dt gives the power:

$$P = \frac{dW}{dt} = \vec{F} \cdot \frac{d\vec{r}}{dt}$$

But $d\vec{r}/dt$ is the velocity $\vec{v}$, so

$$P = \vec{F} \cdot \vec{v} \tag{6.19}$$

EXAMPLE 6.9 Power and Velocity: Bicycling

Riding your 14-kg bicycle at a steady 18 km/h (5.0 m/s), you experience a 30-N force from air resistance. If your mass is 68 kg, what power must you supply on level ground and going up a 5° incline?

INTERPRET This example asks about power in two different situations: one with air resistance alone and the other when climbing. We identify the forces involved as air resistance and gravity. You need to exert forces of equal magnitude to overcome them.

DEVELOP Given that we have force and velocity, Equation 6.19, $P = \vec{F} \cdot \vec{v}$, applies. The force you apply to propel the bicycle is in the same direction as its motion, so $\vec{F} \cdot \vec{v}$ in that equation becomes just Fv.

EVALUATE On level ground, we have $P = Fv = (30\ \text{N})(5.0\ \text{m/s}) = 150\ \text{W}$. Climbing the hill, you have to exert an additional force to overcome the downslope component of gravity, which in Example 5.1 we found to be $mg \sin\theta$. So here we have

$$\begin{aligned} P &= Fv = (F_{\text{air}} + mg \sin\theta)v \\ &= [30\ \text{N} + (82\ \text{kg})(9.8\ \text{m/s}^2)(\sin 5°)](5.0\ \text{m/s}) = 500\ \text{W} \end{aligned}$$

where we used your combined mass, body plus bicycle.

ASSESS Both numbers make sense. The values go from a little to a lot more than your body's average power output of around 100 W, and as you've surely experienced, even a modest slope takes much more cycling effort than level ground. ■

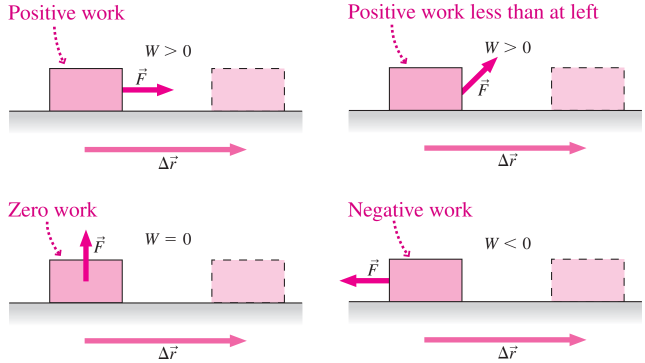
GOT IT? 6.4 A newspaper reports that a new power plant will produce "50 megawatts per hour." What's wrong with this statement?

CHAPTER 6 SUMMARY

Big Picture

Work and **kinetic energy** are the big ideas here. A force acting on an object does work when the object undergoes a displacement and the force has a component in the direction of that displacement. A force at right angles to the displacement does no work, and a force with a component opposite the displacement does negative work.

Kinetic energy is the energy associated with an object's motion. An object's kinetic energy changes only when net work is done on the object.



Key Concepts and Equations

Work is the product of force and displacement, but only the component of force in the direction of displacement counts toward the work.

For a constant force and displacement in the x-direction,

$$W = F_x \Delta x \qquad \text{(constant force only)}$$

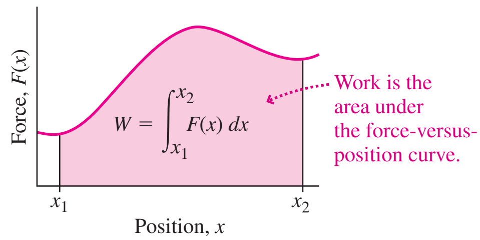
More generally, for a constant force $\vec{F}$ and arbitrary displacement $\Delta\vec{r}$, the work is

$$W = \vec{F} \cdot \Delta\vec{r} = F\,\Delta r \cos\theta \qquad \text{(constant force only)}$$

Here F and Δr are the magnitudes of the force and displacement vectors, and θ is the angle between them. We've written work here using the shorthand notation of the scalar product, defined for any two vectors $\vec{A}$ and $\vec{B}$ as the product of their magnitudes and the cosine of the angle between them:

$$\vec{A} \cdot \vec{B} = AB\cos\theta \qquad \text{(scalar product)}$$

When force varies with position, calculating the work involves integrating. In one dimension:



Most generally, work is the **line integral** of a varying force over an arbitrary path: $W = \int \vec{F} \cdot d\vec{r}$

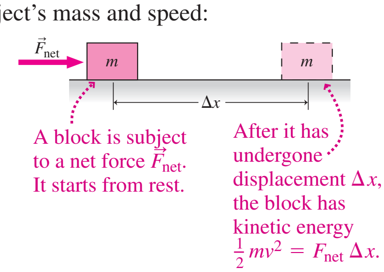
Kinetic energy is a scalar quantity that depends on an object's mass and speed:

$$K = \tfrac{1}{2}mv^2$$

The **work-energy theorem** states that the change in an object's kinetic energy is equal to the net work done on it:

$$\Delta K = W_{\text{net}} \qquad \text{(work-energy theorem)}$$

The unit of energy and work is the joule (J), equal to 1 newton-meter.

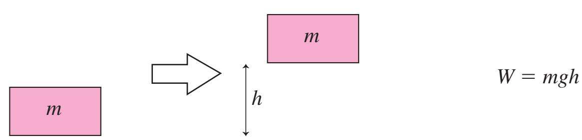


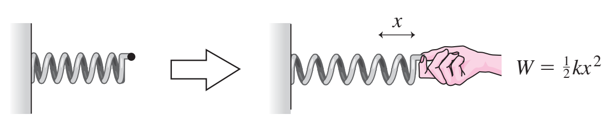
Power is the rate at which work is done or energy is used. The unit of power is the **watt** (W), equal to 1 joule/second.

$$P = \frac{dW}{dt} = \vec{F} \cdot \vec{v}$$

Applications

Common applications of work done against everyday forces are the work mgh needed to raise an object of mass m a distance h against gravity, and the work $\frac{1}{2}kx^2$ needed to stretch or compress a spring of spring constant k a distance x from its equilibrium length.





For Thought and Discussion

1. Give two examples of situations in which you might think you're doing work but in which, in the technical sense, you do no work.
2. If the scalar product of two nonzero vectors is zero, what can you conclude about their relative directions?
3. Must you do work to whirl a ball around on the end of a string? Explain.
4. If you pick up a suitcase and put it down, how much total work have you done on the suitcase? Does your answer change if you pick up the suitcase and drop it?
5. You want to raise a piano a given height using a ramp. With a fixed, nonzero coefficient of friction, will you have to do more work if the ramp is steeper or more gradual? Explain.
6. Does the gravitational force of the Sun do work on a planet in a circular orbit? On a comet in an elliptical orbit? Explain.
7. A pendulum bob swings back and forth on the end of a string, describing a circular arc. Does the tension force in the string do any work?
8. Does your car's kinetic energy change if you drive at constant speed for 1 hour?
9. A watt-second is a unit of what quantity? Relate it to a more standard SI unit.
10. A truck is moving northward at 55 mi/h. Later, it's moving eastward at the same speed. Has its kinetic energy changed? Has work been done on the truck? Has a force acted on the truck? Explain.

Exercises and Problems

Exercises

Section 6.1 Work

11. How much work do you do as you exert a 75-N force to push a shopping cart through a 12-m-long supermarket aisle?
12. If the coefficient of kinetic friction is 0.21, how much work do you do when you slide a 50-kg box at constant speed across a 4.8-m-wide room?
13. A crane lifts a 650-kg beam vertically upward 23 m and then swings it eastward 18 m. How much work does the crane do? Neglect friction, and assume the beam moves with constant speed.
14. The world's highest waterfall, the Cherun-Meru in Venezuela, has a total drop of 980 m. How much work does gravity do on a cubic meter of water dropping down the Cherun-Meru?
15. A meteorite plunges to Earth, embedding itself 75 cm in the ground. If it does 140 MJ of work in the process, what average force does the meteorite exert on the ground?
16. An elevator of mass m rises a vertical distance h with upward acceleration equal to one-tenth g. Find an expression for the work the elevator cable does on the elevator.
17. Show that the scalar product obeys the distributive law: $\vec{A}\cdot(\vec{B}+\vec{C}) = \vec{A}\cdot\vec{B} + \vec{A}\cdot\vec{C}$.
18. Find the work done by a force $\vec{F} = 1.8\hat{\imath} + 2.2\hat{\jmath}$ N as it acts on an object moving from the origin to the point $56\hat{\imath} + 31\hat{\jmath}$ m.
19. To push a stalled car, you apply a 470-N force at 17° to the car's motion, doing 860 J of work in the process. How far do you push the car?

Section 6.2 Forces That Vary

20. Find the total work done by the force shown in Fig. 6.14 as the object on which it acts moves (a) from $x = 0$ to $x = 3$ km and (b) from $x = 3$ km to $x = 4$ km.

FIGURE 6.14 Exercise 20

21. How much work does it take to stretch a spring with $k = 200$ N/m (a) 10 cm from equilibrium and (b) from 10 cm to 20 cm from equilibrium?
22. Uncompressed, the spring for an automobile suspension is 45 cm long. It needs to be fitted into a space 32 cm long. If the spring constant is 3.8 kN/m, how much work does a mechanic have to do to fit the spring?
23. You do 8.5 J of work to stretch a spring with $k = 190$ N/m, starting with the spring unstretched. How far does the spring stretch?
24. BIO Spider silk is a remarkable elastic material. A particular strand has spring constant 70 mN/m, and it stretches 9.6 cm when a fly hits it. How much work did the fly's impact do on the silk strand?

Section 6.3 Kinetic Energy

25. What's the kinetic energy of a 2.4×10^5-kg airplane cruising at 900 km/h?
26. A cyclotron accelerates protons from rest to 21 Mm/s. How much work does it do on each proton?
27. At what speed must a 950-kg subcompact car be moving to have the same kinetic energy as a 3.2×10^4-kg truck going 20 km/h?
28. A 60-kg skateboarder comes over the top of a hill at 5.0 m/s and reaches 10 m/s at the bottom. Find the total work done on the skateboarder between the top and bottom of the hill.
29. After a tornado, a 0.50-g drinking straw was found embedded 4.5 cm in a tree. Subsequent measurements showed that the tree exerted a stopping force of 70 N on the straw. What was the straw's speed?
30. From what height would you have to drop a car for its impact to be equivalent to a 20-mi/h collision?

Section 6.4 Power

31. A typical human diet is "2000 calories" per day, where the "calorie" describing food energy is actually 1 kilocalorie. Express 2000 kcal/day in watts.
32. A horse plows a 200-m-long furrow in 5.0 min, exerting a 750-N force. Find its power output, measured in watts and in horsepower.
33. A typical car battery stores about 1 kW·h of energy. What's its power output if it drains completely in (a) 1 minute, (b) 1 hour, and (c) 1 day?
34. A sprinter completes a 100-m dash in 10.6 s, doing 22.4 kJ of work. What's her average power output?
35. How much work can a 3.5-hp lawnmower engine do in 1 h?
36. A 75-kg long-jumper takes 3.1 s to reach a prejump speed of 10 m/s. What's his power output?
37. Estimate your power output as you do deep knee bends at the rate of one per second.
38. In midday sunshine, solar energy strikes Earth at the rate of about 1 kW/m^2. How long would it take a perfectly efficient solar collector of 15-m^2 area to collect 40 kW·h of energy? (*Note:* This is roughly the energy content of a gallon of gasoline.)

39. It takes about 20 kJ to melt an ice cube. A typical microwave oven produces 900 W of microwave power. How long will it take a typical microwave to melt the ice cube?
40. Which consumes more energy, a 1.2-kW hair dryer used for 10 min or a 7-W night-light left on for 24 h?

Problems

41. You slide a box of books at constant speed up a 30° ramp, applying a force of 200 N directed up the slope. The coefficient of sliding friction is 0.18. (a) How much work have you done when the box has risen 1 m vertically? (b) What's the mass of the box?
42. Two people push a stalled car at its front doors, each applying a 280-N force at 25° to the forward direction, as shown in Fig. 6.15. How much work does each person do in pushing the car 5.6 m?

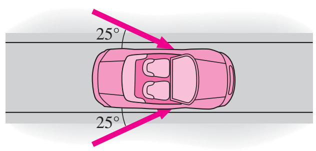


FIGURE 6.15 Problem 42

43. BIO You're at the gym, doing arm raises. With each rep, you lift a 20-N weight 55 cm. (a) How many raises must you do before you've expended 200 kcal of work (see Problem 31)? (b) If your workout takes 1.0 min, what's your average power output?
44. A locomotive does 7.9×10^{11} J of work in pulling a 3.4×10^6-kg train 180 km. Find the average force in the coupling between the locomotive and the rest of the train.
45. You pull a box 23 m horizontally, using the rope shown in Fig. 6.16. If the rope tension is 120 N, and if the rope does 2500 J of work on the box, what angle θ does the rope make with the horizontal?

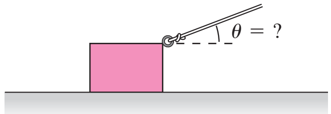


FIGURE 6.16 Problem 45

46. (a) Find the scalar products $\hat{\imath}\cdot\hat{\imath}$, $\hat{\jmath}\cdot\hat{\jmath}$, and $\hat{k}\cdot\hat{k}$. (b) Find $\hat{\imath}\cdot\hat{\jmath}$, $\hat{\jmath}\cdot\hat{k}$, and $\hat{k}\cdot\hat{\imath}$. (c) Use the distributive law to multiply out the scalar product of two arbitrary vectors $\vec{A} = A_x\hat{\imath} + A_y\hat{\jmath} + A_z\hat{k}$ and $\vec{B} = B_x\hat{\imath} + B_y\hat{\jmath} + B_z\hat{k}$, and use the results of (a) and (b) to verify Equation 6.4.
47. (a) Find the scalar product of the vectors $a\hat{\imath} + b\hat{\jmath}$ and $b\hat{\imath} - a\hat{\jmath}$, where a and b are arbitrary constants. (b) What's the angle between the two vectors?
48. Looking to cut costs, the airline you work for asks you to investigate the efficiency of the tractors that push aircraft away from the gates. One model is supposed to do no more than 10 MJ of work in pushing a 747 aircraft 25 m. If the tractor exerts a 0.42-MN force, does it meet its specifications?
49. How much work does a force $\vec{F} = 67\hat{\imath} + 23\hat{\jmath} + 55\hat{k}$ N do as it acts on a body moving in a straight line from $\vec{r}_1 = 16\hat{\imath} + 31\hat{\jmath}$ m to $\vec{r}_2 = 21\hat{\imath} + 10\hat{\jmath} + 14\hat{k}$ m?
50. A force $\vec{F}$ acts in the x-direction, its magnitude given by $F = ax^2$, where x is in meters and $a = 5.0\ \mathrm{N/m^2}$. Find the work done by this force as it acts on a particle moving from $x = 0$ to $x = 6.0$ m.
51. A certain amount of work is required to stretch spring A a certain distance. Twice as much work is required to stretch spring B half that distance. Compare the spring constants of the two.
52. A force with magnitude $F = a\sqrt{x}$ acts in the x-direction, where $a = 9.5\ \mathrm{N/m^{1/2}}$. Calculate the work this force does as it acts on an object moving from (a) $x = 0$ to $x = 3.0$ m; (b) 3.0 m to 6.0 m; and (c) 6.0 m to 9.0 m.
53. The force exerted by a rubber band is given approximately by
$$F = F_0\left[\frac{L_0 - x}{L_0} - \frac{L_0^2}{(L_0 + x)^2}\right]$$
where L_0 is the unstretched length, x is the stretch, and F_0 is a constant. Find the work needed to stretch the rubber band a distance x.
54. You put your little sister (mass m) on a swing whose chains have length L and pull slowly back until the swing makes an angle ϕ with the vertical. Show that the work you do is $mgL(1 - \cos\phi)$.
55. Two unknown elementary particles pass through a detection chamber. If they have the same kinetic energy and their mass ratio is 4 : 1, what's the ratio of their speeds?
56. A tractor tows a plane from its airport gate, doing 8.7 MJ of work. The link from the plane to the tractor makes a 22° angle with the plane's motion, and the tension in the link is 0.41 MN. How far does the tractor move the plane?
57. A force pointing in the x-direction is given by $F = F_0(x/x_0)$, where F_0 and x_0 are constants and x is position. Find an expression for the work done by this force as it acts on an object moving from $x = 0$ to $x = x_0$.
58. A force pointing in the x-direction is given by $F = ax^{3/2}$, where $a = 0.75\ \mathrm{N/m^{3/2}}$. Find the work done by this force as it acts on an object moving from $x = 0$ to $x = 14$ m.
59. Two vectors have equal magnitude, and their scalar product is one-third the square of their magnitude. Find the angle between them.
60. At what rate can a half-horsepower well pump deliver water to a tank 60 m above the water level in the well? Give your answer in kg/s and gal/min.
61. The rate at which the United States imports oil, expressed in terms of the energy content of the imported oil, is about 800 GW. Using the "Energy Content of Fuels" table in Appendix C, convert this figure to gallons per day.
62. BIO By measuring oxygen uptake, sports physiologists have found that long-distance runners' power output is given approximately by $P = m(bv - c)$, where m and v are the runner's mass and speed, and b and c are constants given by $b = 4.27$ J/kg·m and $c = 1.83$ W/kg. Determine the work done by a 54-kg runner who runs a 10-km race at 5.2 m/s.
63. You're writing performance specifications for a new car model. The 1750-kg car delivers energy to its drive wheels at the rate of 35 kW. Neglecting air resistance, what do you list for the greatest speed at which it can climb a 4.5° slope?
64. A 1400-kg car ascends a mountain road at a steady 60 km/h, against a 450-N force of air resistance. If the engine supplies energy to the drive wheels at the rate of 38 kW, what's the slope angle of the road?
65. You do 2.2 kJ of work pushing a trunk at constant speed 3.1 m along a ramp inclined upward at 22°. What's the frictional coefficient between trunk and ramp?
66. BIO (a) Find the work done in lifting 1 L of blood (mass 1 kg) from the foot to the head of a 1.7-m-tall person. (b) If blood circulates through the body at the rate of 5.0 L/min, estimate the heart's power output. (Your answer underestimates the power by a factor of about 5 because it neglects fluid friction and other factors.)
67. (a) What power is needed to push a 95-kg crate at 0.62 m/s along a horizontal floor where the coefficient of friction is 0.78? (b) How much work is done in pushing the crate 11 m?
68. You mix flour into bread dough, exerting a 45-N force on the spoon, which you move at 0.29 m/s. (a) What power do you supply? (b) How much work do you do if you stir for 1.0 min?
69. A machine does work at a rate given by $P = ct^2$, where $c = 18\ \mathrm{W/s^2}$ and t is time. Find the work done between $t = 10$ s and $t = 20$ s.

70. **BIO** A typical bumblebee has mass 0.25 mg. It beats its wings 100 times per second, and the wings undergo an average displacement of about 1.5 mm. When the bee is hovering over a flower, the average force between wings and air must support the bee's weight. Estimate the average power the bee expends in hovering.
71. You're trying to decide whether to buy an energy-efficient 225-W refrigerator for \$1150 or a standard 425-W model for \$850. The standard model will run 20% of the time, but better insulation means the energy-efficient model will run 11% of the time. If electricity costs 9.5¢/kW·h, how long would you have to own the energy-efficient model to make up the difference in cost? Neglect interest you might earn on your money.
72. Your friend does five reps with a barbell, on each rep lifting 45 kg 0.50 m. She claims the work done is enough to "burn off" a chocolate bar with energy content 230 kcal (see Problem 31). Is that true? If not, how many lifts would it take?
73. A machine delivers power at a decreasing rate $P = P_0 t_0^2/(t + t_0)^2$, where P_0 and t_0 are constants. The machine starts at $t = 0$ and runs forever. Show that it nevertheless does only a finite amount of work, equal to $P_0 t_0$.
74. A locomotive accelerates a freight train of total mass M from rest, applying constant power P. Determine the speed and position of the train as functions of time, assuming all the power goes to increasing the train's kinetic energy.
75. A force given by $F = b/\sqrt{x}$ acts in the x-direction, where b is a constant with the units N·m$^{1/2}$. Show that even though the force becomes arbitrarily large as x approaches zero, the work done in moving from x_1 to x_2 remains finite even as x_1 approaches zero. Find an expression for that work in the limit $x_1 \to 0$.
76. **BIO** You're assisting a cardiologist in planning a stress test for a 75-kg patient. The test involves rapid walking on an inclined treadmill, and the patient is to reach a peak power output of 350 W. If the patient's maximum walking speed is 8.0 km/h, what should be the treadmill's inclination angle?
77. You're an engineer for a company that makes bungee-jump cords, and you're asked to develop a formula for the work involved in stretching cords to double their length. Your cords have force-distance relations described by $F = -(kx + bx^2 + cx^3 + dx^4)$, where k, b, c, and d are constants. (a) Given a cord with unstretched length L_0, what's your formula? (b) Evaluate the work done in doubling the stretch of a 10-m cord with $k = 420$ N/m, $b = -86$ N/m^2, $c = 12$ N/m^3, and $d = -0.50$ N/m^4.
78. You push an object of mass m slowly, partway up a loop-the-loop track of radius R, starting from the bottom, and ending at a height $h < R$ above the bottom. The coefficient of friction between the object and the track is a constant μ. Show that the work you do against friction is $\mu mg\sqrt{2hR - h^2}$.
79. A particle moves from the origin to the point $x = 3$ m, $y = 6$ m along the curve $y = ax^2 - bx$, where $a = 2$ m^{-1} and $b = 4$. It's subject to a force $cxy\hat{\imath} + d\hat{\jmath}$, where $c = 10$ N/m^2 and $d = 15$ N. Calculate the work done by the force.
80. Repeat Problem 79 for the following cases: (a) the particle moves first along the x-axis from the origin to the point (3 m, 0) and then parallel to the y-axis until it reaches (3 m, 6 m); (b) it moves first along the y-axis from the origin to the point (0, 6 m) and then parallel to the x-axis until it reaches (3 m, 6 m).
81. You're an expert witness in a medical malpractice lawsuit. A hospital patient's leg slipped off a stretcher and his heel hit the floor. The defense attorney for the hospital claims the leg, with mass 8 kg, hit the floor with a force equal to the weight of the leg—about 80 N—and any damage was due to a prior injury. You argue that the leg and heel dropped freely for 0.7 m, then hit the floor and stopped in 2 cm. What do you tell the jury about the force on the heel?

Passage Problems

The energy in a batted baseball comes from the power delivered while the bat is in contact with the ball. The most powerful hitters can supply some 10 horsepower during the brief contact time, propelling the ball to over 100 miles per hour. Figure 6.17 shows data taken from a particular hit, giving the power the bat delivers to the ball as a function of time.

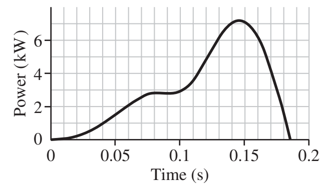


FIGURE 6.17 Passage Problems 82–85

82. Which of the following is greatest at the peak of the curve?
    a. the ball's kinetic energy
    b. the ball's speed
    c. the rate at which the bat supplies energy to the ball
    d. the total work the bat has done on the ball
83. The ball has its maximum speed at about
    a. 85 ms.
    b. 145 ms.
    c. 185 ms.
    d. whenever the force is greatest.
84. As a result of being hit, the ball's kinetic energy increases by about
    a. 550 J.
    b. 1.3 kJ.
    c. 7.0 kJ.
    d. You can't tell because you don't know its speed coming from the pitcher.
85. The force on the ball is greatest approximately
    a. at 185 ms.
    b. at the peak in Fig. 6.17.
    c. before the peak in Fig. 6.17.
    d. after the peak in Fig. 6.17 but before 185 ms.

Answers to Chapter Questions

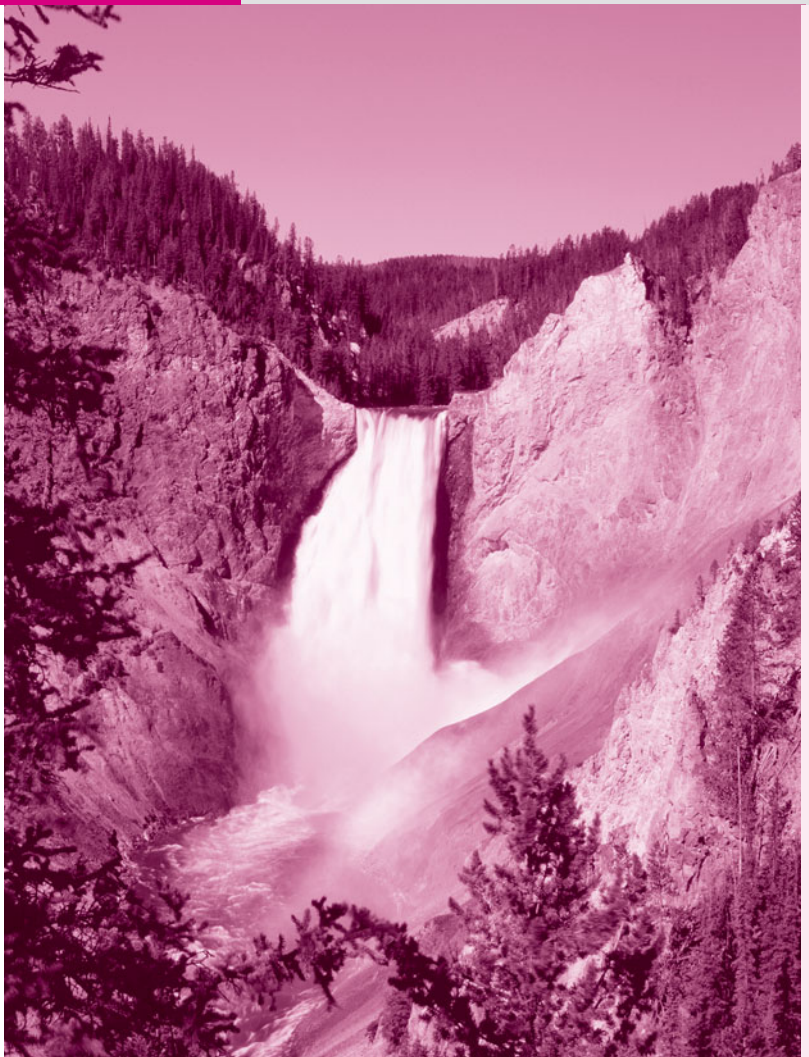
Answer to Chapter Opening Question

No. The work done against gravity in climbing a particular height is independent of the path. A rider on a bicycle with a combined mass of 80 kg does roughly 400 kilojoules or 100 kilocalories of work against gravity regardless of the path up a 500-m mountain. To climb such a mountain in 20 minutes, the rider's power output must exceed 300 watts.

Answers to GOT IT? Questions

6.1. The force $2F$ does $\sqrt{2}$ more work than F does. That's because $2F$'s component along the direction of motion is $2F \cos 45°$, or $2F\sqrt{2}/2 = F\sqrt{2}$.
6.2. (c) $\sqrt{x}$ does the most work. (b) x^2 does the least. You can see this by plotting these two functions from $x = 0$ to $x = 1$ and comparing the areas under each. The case of x is intermediate.
6.3. (a) Kinetic energy doesn't change, so the net work done on the ball is zero. (b) Kinetic energy increases, so the net work is positive. (c) Kinetic energy decreases, so the net work is negative.
6.4. The megawatt is a unit of power; the "per time" is already built in. A correct statement would be that the power plant will produce "energy at the rate of 50 megawatts."

7 Conservation of Energy



How many different energy conversions take place as the Yellowstone River plunges over Yellowstone Falls?

New Concepts, New Skills

By the end of this chapter you should be able to

- Explain the difference between conservative and nonconservative forces (7.1).
- Describe the concept of potential energy, and calculate potential energy given a conservative force as a function of position (7.2).
- Articulate and apply the principle of conservation of mechanical energy (7.3).
- Show the relation between force and energy using potential-energy curves (7.4).

The rock climber of Fig. 7.1*a* (next page) does work as she ascends the vertical cliff. So does the mover of Fig. 7.1*b* (next page), as he pushes a heavy chest across the floor. But there's a difference. If the rock climber lets go, down she goes; the work she put into the climb comes back as the kinetic energy of her fall. If the mover lets go of the chest, though, he and the chest stay right where they are.

This contrast highlights a distinction between two types of forces, called *conservative* and *nonconservative*. From that distinction we'll develop one of the most important principles in all of physics: **conservation of energy**. In this chapter, we consider conservation of **mechanical energy**, which includes kinetic energy, introduced in Chapter 6, and **potential energy**, which we'll introduce here. In later chapters, we'll expand the conservation-of-energy principle to include other forms of energy.

Connecting Your Knowledge

- The material of this chapter builds on the idea of work done against conservative forces. You should be familiar with the concept of work as introduced in Chapter 6 (6.1, 6.2).
- You should be able to calculate work for both constant and position-varying forces (6.1, 6.2).
- Kinetic energy is crucial to an understanding of the all-important conservation-of-energy principle (6.3).

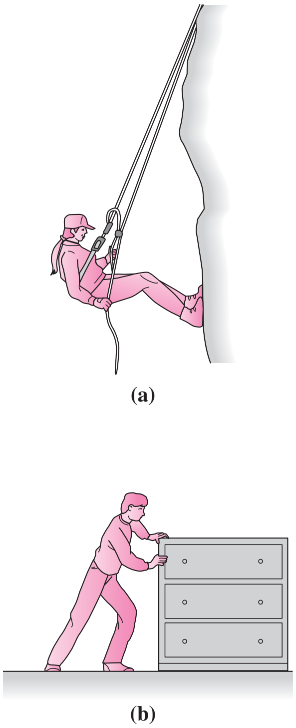


FIGURE 7.1 Both the rock climber and the mover do work, but only the climber can recover that work as kinetic energy.

7.1 Conservative and Nonconservative Forces

Both the climber and the mover in Fig. 7.1 are working against external forces—gravity for the climber and friction for the mover. The difference is this: If the climber lets go, the gravitational force "gives back" the work she did; that work manifests itself as kinetic energy. But the frictional force doesn't "give back" the mover's work; that work can't be recovered as kinetic energy.

A **conservative force** is a force like gravity or a spring that "gives back" work done against it. A force like friction is **nonconservative**. A more precise mathematical sense of this distinction comes from considering the work involved in moving an object over a closed path—one that ends where it started. Suppose our rock climber ascends a cliff of height h and then descends to her starting point. How much work has the gravitational force done on her? That force has magnitude mg and points downward. As she climbs up, the force is directed opposite to her motion, so gravity does work $-mgh$. As she descends, the force is in the same direction as her motion, so the gravitational work is $+mgh$. The total work that gravity does on the climber as she traverses the closed path up and down the cliff is therefore zero.

Now suppose the mover in Fig. 7.1*b* pushes the chest a distance L across a room, discovers it's the wrong room, and pushes the chest back to the door. Like the climber, the chest describes a closed path. How much work is done by the frictional force acting over this path? That force has magnitude μn, where the normal force in this case of horizontal motion is mg. But the force of kinetic friction *always* acts to oppose the motion, so it *always* does negative work. With frictional force μmg opposing the motion, the work done as the chest crosses the width L of the room is $-\mu mgL$. As the chest comes back, the work is also $-\mu mgL$, so the total work done by the frictional force is $-2\mu mgL$.

The difference between our answers for the work done by the gravitational and frictional forces acting over closed paths provides one precise definition of the distinction between conservative and nonconservative forces:

> When the total work done by a force F acting as an object moves over any closed path is zero, the force is conservative. Mathematically,
>
> $$\oint \vec{F} \cdot d\vec{r} = 0 \qquad \text{(conservative force)} \qquad (7.1)$$

This expression comes from the most general formula for work, Equation 6.11, $W = \int_{\vec{r}_1}^{\vec{r}_2} \vec{F} \cdot d\vec{r}$, which we introduced in Chapter 6. The circle on the integral sign in Equation 7.1 indicates that the integral is to be taken over a *closed* path. A force for which the integral is nonzero, like friction, is nonconservative.

Equation 7.1 suggests a related property of conservative forces. Suppose we move an object along the straight path between points A and B shown in Fig. 7.2, along which a conservative force acts; let the work done by the conservative force be W_{AB}. Since the work done over any closed path is zero, the work W_{BA} done in moving back from B to A must be $-W_{AB}$, whether we return along the straight path or the curved path or any other path. So, going from A to B involves work W_{AB}, regardless of the path taken. In other words:

> The work done by a conservative force in moving between two points is independent of the path taken; mathematically, $\int_A^B \vec{F} \cdot d\vec{r}$ depends only on the endpoints A and B, not on the path between them.

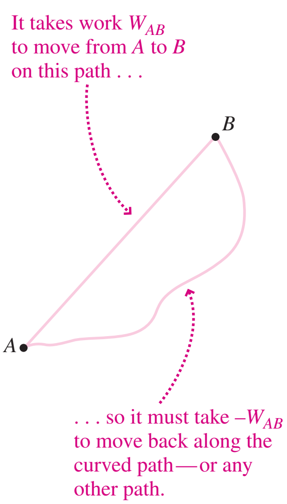


FIGURE 7.2 The work done by a conservative force is independent of path.

In contrast, the work done by a nonconservative force does depend on the path. On a frictional surface, for example, the least work is done over a straight-line path; any other path involves more work.

Important examples of conservative forces include gravity and the static electric force. The force of an ideal spring—fundamentally an electric force—is also conservative.

Nonconservative forces include friction and the electric force in the presence of time-varying magnetic effects, which we'll encounter in Chapter 27.

GOT IT? 7.1 Suppose it takes the same amount of work to push a trunk across a rough floor as it does to lift a weight the same distance straight upward. How do the amounts of work compare if the trunk and weight are moved instead on curved paths between the same starting and ending points?

7.2 Potential Energy

Work done against a conservative force is somehow "stored," in the sense that we can get it back again in the form of kinetic energy. The climber in Fig. 7.1*a* is acutely aware of that "stored work"; it gives her the potential for a dangerous fall. *Potential* is an appropriate word here: We can consider the "stored work" as **potential energy U**, in the sense that it can be released as kinetic energy.

We define potential energy formally in terms of the work done by a conservative force. Specifically:

The change ΔU_{AB} in potential energy associated with a conservative force is the negative of the work done by that force as it acts over any path from point A to point B:

$$\Delta U_{AB} = -\int_A^B \vec{F} \cdot d\vec{r} \qquad \text{(potential energy)} \qquad (7.2)$$

Why the minus sign? Because potential energy represents stored work. If a conservative force does *positive* work (as does gravity on a falling object), then potential energy must decrease—and that means ΔU must be *negative*. Conversely, if a conservative force does *negative* work (as does gravity on a weight being lifted), then energy is stored, and ΔU must be *positive*. The minus sign in Equation 7.2 handles both these cases. We'll often drop the subscript AB and write simply ΔU for potential-energy change. Keeping the subscript is important, though, when we need to be clear about whether we're going from A to B or from B to A.

Changes in potential energy are all that ever matter physically; the actual value of potential energy is meaningless. Often, though, it's convenient to establish a reference point at which the potential energy is defined to be zero. When we say "the potential energy U," we really mean the potential-energy difference ΔU between that reference point and whatever other point we're considering. Our rock climber, for example, might find it convenient to take the zero of potential energy at the base of the cliff. But the choice is purely for convenience; only potential-energy *differences* really matter.

Equation 7.2 is a completely general definition of potential energy, applicable in all circumstances. Often, though, we can consider a path where force and displacement are parallel (or antiparallel). Then Equation 7.2 simplifies to

$$\Delta U = -\int_{x_1}^{x_2} F(x)\, dx \qquad (7.2a)$$

where x_1 and x_2 are the starting and ending points on the x-axis, taken to coincide with the path. When the force is constant, this equation simplifies further to

$$\Delta U = -F(x_2 - x_1) \qquad (7.2b)$$

✓TIP Understand Your Equations

Equation 7.2b provides a very simple expression for potential-energy changes, but it applies *only* when the force is constant. Equation 7.2b is a special case of Equation 7.2a that follows because a constant force can be taken outside the integral.

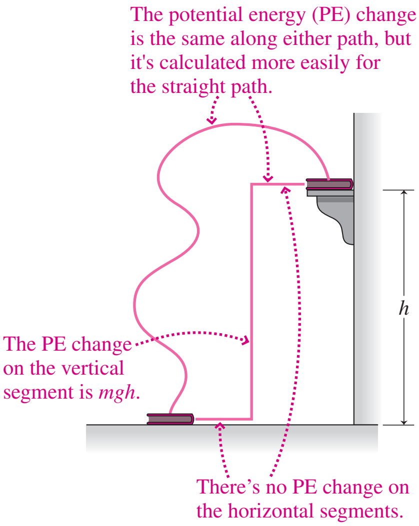


FIGURE 7.3 A good choice of path makes it easier to calculate the potential-energy change.

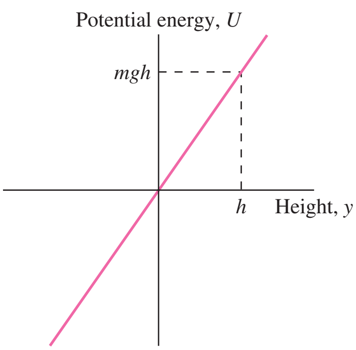


FIGURE 7.4 Gravitational force is constant, so potential energy increases linearly with height.

Gravitational Potential Energy

We're frequently moving things up and down, causing changes in potential energy. Figure 7.3 shows two possible paths for a book that's lifted from the floor to a shelf of height h. Since the gravitational force is conservative, we can use either path to calculate the potential-energy change. It's easiest to use the path consisting of straight segments. No work or potential-energy change is associated with the horizontal motion, since the gravitational force is perpendicular to the motion. For the vertical lift, the force of gravity is constant and Equation 7.2b gives immediately $\Delta U = mgh$, where the minus sign in Equation 7.2b cancels with the minus sign associated with the *downward* direction of gravity. This result is quite general: When a mass m undergoes a vertical displacement Δy near Earth's surface, its gravitational potential energy changes by

$$\Delta U = mg\,\Delta y \qquad (\text{gravitational potential energy}) \tag{7.3}$$

The quantity Δy can be positive or negative, depending on whether the object moves up or down; correspondingly, the potential energy can either increase or decrease. We emphasize that Equation 7.3 applies *near Earth's surface*—that is, for distances small compared with Earth's radius. That assumption allows us to treat the gravitational force as constant over the path. We'll explore the more general case in the next chapter.

We've found the *change* in the book's potential energy, but what about the potential energy itself? That depends on where we define the zero of potential energy. If we choose $U = 0$ at the floor, then $U = mgh$ on the shelf. But we could just as well take $U = 0$ at the shelf; then the book's potential energy on the floor would be $-mgh$. Negative potential energies arise frequently, and that's OK because only *differences* in potential energy really matter. Figure 7.4 shows a plot of potential energy versus height with $U = 0$ taken at the floor. The *linear* increase in potential energy with height reflects the *constant* gravitational force.

EXAMPLE 7.1 Gravitational Potential Energy: Riding the Elevator

A 55-kg engineer leaves her office on the 33rd floor of a skyscraper and takes an elevator up to the 59th floor. Later she descends to street level. If the engineer chooses the zero of potential energy at her office and if the distance from one floor to the next is 3.5 m, what's the engineer's potential energy (a) in her office, (b) on the 59th floor, and (c) at street level?

INTERPRET This is a problem about gravitational potential energy relative to a specified point of zero energy—namely, the engineer's office.

DEVELOP Equation 7.3, $\Delta U = mg\,\Delta y$, gives the change in gravitational energy associated with a change Δy in vertical position. We're given positions in floors, not meters, so we need to convert using the given factor 3.5 m per floor.

EVALUATE (a) In her office, the engineer's potential energy is zero, since she defined it that way. (b) The 59th floor is $59 - 33 = 26$ floors higher, so the potential energy there is

$$U_{59} = mg\,\Delta y = (55\text{ kg})(9.8\text{ m/s}^2)(26\text{ floors})(3.5\text{ m/floor}) = 49\text{ kJ}$$

Note that we can write U rather than ΔU because we're calculating the potential-energy change from the place where $U = 0$. (c) The street level is 32 floors *below* the engineer's office, so

$$U_{\text{street}} = mg\,\Delta y = (55\text{ kg})(9.8\text{ m/s}^2)(-32\text{ floors})(3.5\text{ m/floor}) = -60\text{ kJ}$$

ASSESS Makes sense: When the engineer goes *up*, the potential energy relative to her office is positive; when she goes *down*, it's negative. And the distance down is a bit farther, so the magnitude of the change is greater going down. ■

APPLICATION **Pumped Storage**

Electricity is a wonderfully versatile form of energy, but it's not easy to store. Large electric power plants are most efficient when operated continuously, yet the demand for power fluctuates. Renewable energy sources like wind and solar vary, not necessarily with demand. Energy storage can help in both cases. Today, the only practical way to store large amounts of excess electrical energy is to convert it to gravitational potential energy. In so-called pumped-storage facilities, surplus electric power pumps water from a lower reservoir to a higher one, thereby increasing the water's gravitational potential energy. When power demand is high, water runs back down, turning the pump motors into generators that produce electricity. The photograph here shows the Northfield Mountain Pumped Storage Project in Massachusetts. You can explore this facility quantitatively in Problem 28.

Elastic Potential Energy

When you stretch or compress a spring or other elastic object, you do work against the spring force, and that work ends up stored as **elastic potential energy**. For an ideal spring, the force is $F = -kx$, where x is the distance the spring is stretched from equilibrium, and the minus sign shows that the force opposes the stretching or compression. Since the force varies with position, we use Equation 7.2a to evaluate the potential energy:

$$\Delta U = -\int_{x_1}^{x_2} F(x)\,dx = -\int_{x_1}^{x_2} (-kx)\,dx = \tfrac{1}{2}kx_2^2 - \tfrac{1}{2}kx_1^2$$

where x_1 and x_2 are the initial and final values of the stretch. If we take $U = 0$ when $x = 0$—that is, when the spring is neither stretched nor compressed—then we can use this result to write the potential energy at an arbitrary stretch (or compression) x as

$$U = \tfrac{1}{2}kx^2 \qquad \text{(elastic potential energy)} \qquad (7.4)$$

Comparison with Equation 6.10, $W = \frac{1}{2}kx^2$, shows that this is equal to the work done in stretching the spring. Of course: That work gets stored as potential energy. Figure 7.5 shows potential energy as a function of the stretch or compression of a spring. The *parabolic* shape of the potential-energy curve reflects the *linear* change of the spring force with stretch or compression.

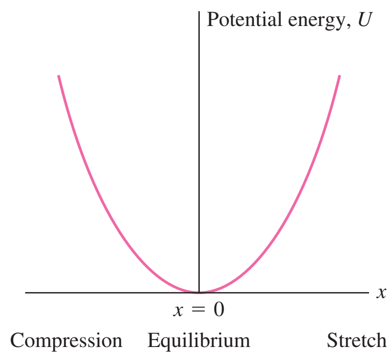


FIGURE 7.5 The potential-energy curve for a spring is a parabola.

EXAMPLE 7.2 Energy Storage: Springs Versus Gasoline

A car's suspension consists of springs with an overall effective spring constant of 120 kN/m. How much would you have to compress the springs to store the same amount of energy as in 1 gram of gasoline?

INTERPRET This problem is about the energy stored in a spring, as compared with the chemical energy of gasoline.

DEVELOP Equation 7.4, $U = \frac{1}{2}kx^2$, gives a spring's stored energy when it's been compressed a distance x. Here we want that energy to equal the energy in 1 gram of gasoline. We can get that value from the "Energy Content of Fuels" table in Appendix C, which lists 44 MJ/kg for gasoline.

EVALUATE At 44 MJ/kg, the energy in 1 g of gasoline is 44 kJ. Setting this equal to the spring energy $\frac{1}{2}kx^2$ and solving for x, we get

$$x = \sqrt{\frac{2U}{k}} = \sqrt{\frac{(2)(44\ \text{kJ})}{120\ \text{kN/m}}} = 86\ \text{cm}$$

ASSESS This answer is absurd. A car's springs couldn't compress anywhere near that far before the underside of the car hit the ground. And 1 g isn't much gasoline. This example shows that springs, though useful energy-storage devices, can't possibly compete with chemical fuels.

EXAMPLE 7.3 Elastic Potential Energy: A Climbing Rope

Ropes used in rock climbing are "springy" so that they cushion a fall. A particular rope exerts a force $F = -kx + bx^2$, where $k = 223$ N/m, $b = 4.10$ N/m^2, and x is the stretch. Find the potential energy stored in this rope when it's been stretched 2.62 m, taking $U = 0$ at $x = 0$.

INTERPRET Like Example 7.2, this one is about elastic potential energy. But this one isn't so easy because the rope isn't a simple $F = -kx$ spring for which we already have a potential-energy formula.

DEVELOP Because the rope force varies with stretch, we'll have to integrate. Since force and displacement are in the same direction, we can use Equation 7.2a, $\Delta U = -\int_{x_1}^{x_2} F(x)\,dx$. But that's not so much a formula as a strategy for deriving one.

EVALUATE Applying Equation 7.2 to this particular rope, we have

$$U = -\int_{x_1}^{x_2} F(x)\,dx = -\int_0^x (-kx + bx^2)\,dx = \tfrac{1}{2}kx^2 - \tfrac{1}{3}bx^3\Big|_0^x$$
$$= \tfrac{1}{2}kx^2 - \tfrac{1}{3}bx^3$$
$$= \left(\tfrac{1}{2}\right)(223\text{ N/m})(2.62\text{ m})^2 - \left(\tfrac{1}{3}\right)(4.1\text{ N/m}^2)(2.62\text{ m})^3$$
$$= 741\text{ J}$$

ASSESS This result is about 3% less than the potential energy $U = \frac{1}{2}kx^2$ of an ideal spring with the same spring constant. This shows the effect of the extra term $+bx^2$, whose positive sign reduces the restoring force and thus the work needed to stretch the spring. ■

GOT IT? 7.2 Gravitational force actually decreases with height, but that decrease is negligible near Earth's surface. To account for the decrease, would the exact value for the potential-energy change associated with a height change h be (a) greater than, (b) less than, or (c) equal to mgh, where g is the gravitational acceleration at Earth's surface?

7.3 Conservation of Mechanical Energy

The work-energy theorem, introduced in Chapter 6, shows that the change ΔK in a body's kinetic energy is equal to the net work done on it:

$$\Delta K = W_{\text{net}}$$

Consider separately the work W_c done by conservative forces and the work W_{nc} done by nonconservative forces. Then

$$\Delta K = W_c + W_{nc}$$

We've defined the change in potential energy ΔU as the negative of the work done by conservative forces. So we can write

$$\Delta K = -\Delta U + W_{nc}$$

or

$$\Delta K + \Delta U = W_{nc} \qquad (7.5)$$

We define the sum of the kinetic and potential energy as the **mechanical energy**. Then Equation 7.5 shows that the change in mechanical energy is equal to the work done by nonconservative forces.

In the absence of nonconservative forces, the mechanical energy is unchanged:

$$\Delta K + \Delta U = 0 \qquad (7.6)$$

and, equivalently, (conservation of mechanical energy)

$$K + U = \text{constant} = K_0 + U_0 \qquad (7.7)$$

where K_0 and U_0 are the kinetic and potential energy of a body at some point, and K and U are their values when the body is at any other point. Equations 7.6 and 7.7 express the **law of conservation of mechanical energy**. They show that, in the absence of nonconservative forces, the mechanical energy $K + U$ remains always the same. The kinetic energy K may change, but that change is always compensated by an equal but opposite change in potential energy.

Conservation of mechanical energy is a powerful principle. Throughout physics, from the subatomic realm through practical problems in engineering and on to astrophysics, the principle of energy conservation is widely used in solving problems that would be intractable without it.

PROBLEM-SOLVING STRATEGY 7.1 Conservation of Energy

Using energy conservation to solve problems is straightforward: Equation 7.7 basically tells it all. Our IDEA problem-solving strategy adapts well to such problems.

INTERPRET First, interpret the problem to be sure that conservation of mechanical energy applies. Are all the forces conservative? If so, mechanical energy is conserved. Next, identify a point at which you know both the kinetic and the potential energy; then you know the total mechanical energy, which is what's conserved. If the problem doesn't do so and it's not implicit in the equations you use, you may need to identify the zero of potential energy—although that's your own arbitrary choice. You also need to identify the quantity the problem is asking for, and the situation in which it has the value you're after. The quantity may be the energy itself or a related quantity like height, speed, or spring compression.

DEVELOP Draw your object first in the situation where you know its energies and then in the situation that contains the unknown. It's helpful to draw simple bar charts suggesting the relative sizes of the potential- and kinetic-energy terms; we'll show you how in several examples. Then you're ready to set up the quantitative statement of mechanical energy conservation, Equation 7.7: $K + U = K_0 + U_0$. Consider which of the four terms you know or can calculate from the given information. You'll probably need secondary equations like the expressions for kinetic energy and for various forms of potential energy. Consider how the quantity you're trying to find is related to an energy.

EVALUATE Write Equation 7.7 for your specific problem, including expressions for kinetic or potential energy that contain the quantity you're after. Solving is then a matter of algebra.

ASSESS As usual, ask whether your answer makes physical sense. Does it have the right units? Are the numbers reasonable? Do the signs make sense? Is your answer consistent with the bar charts in your drawing?

EXAMPLE 7.4 Energy Conservation: Tranquilizing an Elephant

A biologist uses a spring-loaded gun to shoot tranquilizer darts into an elephant. The gun's spring has $k = 940$ N/m and is compressed a distance $x_0 = 25$ cm before firing a 38-g dart. Assuming the gun is pointed horizontally, at what speed does the dart leave the gun?

INTERPRET We're dealing with a spring, assumed ideal, so conservation of mechanical energy applies. We identify the initial state—dart at rest, spring fully compressed—as the point where we know both kinetic and potential energy. The state we're then interested in is when the dart just leaves the gun, when its potential energy has been converted to kinetic energy and before gravity has changed its vertical position.

DEVELOP In Fig. 7.6 we've sketched the two states, giving the potential and kinetic energy for each. We've also sketched bar graphs showing the relative sizes of the energies. To use the statement of energy conservation, Equation 7.7, we also need expressions for the kinetic energy $\left(\frac{1}{2}mv^2\right)$ and the spring potential energy ($\frac{1}{2}kx^2$; Equation 7.4). Incidentally, using Equation 7.4 implicitly sets the zero of elastic potential energy when the spring is in its equilibrium position. We might

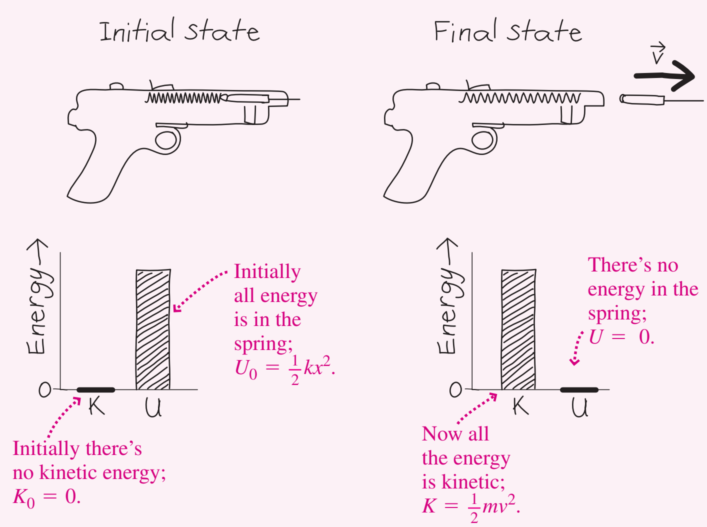


FIGURE 7.6 Our sketches for Example 7.4, showing bar charts for the initial and final states.

(continued)

as well set the zero of gravitational energy at the height of the gun, since there's no change in the dart's vertical position between our initial and final states.

EVALUATE We're now ready to write Equation 7.7, $K + U = K_0 + U_0$. We know three of the terms in this equation: The initial kinetic energy K_0 is 0, since the dart is initially at rest. The initial potential energy is that of the compressed spring, $U_0 = \frac{1}{2}kx_0^2$. The final potential energy is $U = 0$ because the spring is now in its equilibrium position and we've taken the gravitational potential energy to be zero. What we don't know is the final kinetic energy, but we do know that it's given by $K = \frac{1}{2}mv^2$. So Equation 7.7 becomes $\frac{1}{2}mv^2 + 0 = 0 + \frac{1}{2}kx^2$, which solves to give

$$v = \sqrt{\frac{k}{m}}x_0 = \left(\sqrt{\frac{940\text{ N/m}}{0.038\text{ kg}}}\right)(0.25\text{ m}) = 39\text{ m/s}$$

ASSESS Take a look at the answer in algebraic form; it says that a stiffer spring or a greater compression will give a higher dart speed. Increasing the dart mass, on the other hand, will decrease the speed. All this makes good physical sense. And the outcome shows quantitatively what our bar charts suggest—that the dart's energy starts out all potential and ends up all kinetic. ■

Example 7.4 shows the power of the conservation-of-energy principle. If you had tried to find the answer using Newton's law, you would have been stymied by the fact that the spring force and thus the acceleration of the dart vary continuously. But you don't need to worry about those details; all you want is the final speed, and energy conservation gets you there, shortcutting the detailed application of $\vec{F} = m\vec{a}$.

EXAMPLE 7.5 Conservation of Energy: A Spring and Gravity

The spring in Fig. 7.7 has $k = 140$ N/m. A 50-g block is placed against the spring, which is compressed 11 cm. When the block is released, how high up the slope does it rise? Neglect friction.

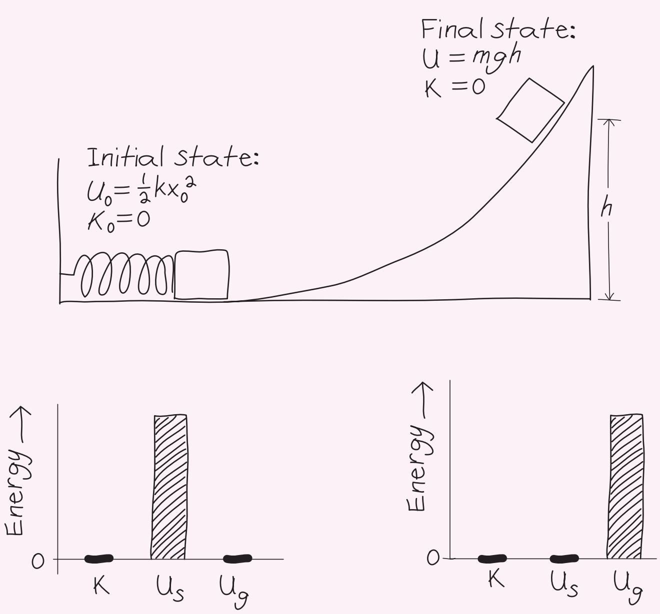


FIGURE 7.7 Our sketches for Example 7.5.

INTERPRET This example is similar to Example 7.4, but now we have changes in both elastic and gravitational potential energy. Since friction is negligible, we can apply conservation of energy. We identify the initial state as the block at rest against the compressed spring; the final state is the block momentarily at rest at its topmost point on the slope. We'll take the zero of gravitational potential energy at the bottom.

DEVELOP Figure 7.7 shows the initial and final states, along with bar charts for each. We've drawn separate bars for the spring and gravitational potential energies, U_s and U_g. Now apply Equation 7.7, $K + U = K_0 + U_0$.

EVALUATE In both states the block is at rest, so kinetic energy is zero. In the initial state we know the potential energy U_0: It's the spring energy $\frac{1}{2}kx^2$. We don't know the final-state potential energy, but we do know that it's gravitational energy—and with the zero of potential energy at the bottom, it's $U = mgh$. With $K = K_0 = 0$, $U_0 = \frac{1}{2}kx^2$, and $U = mgh$, Equation 7.7 reads $0 + mgh = 0 + \frac{1}{2}kx^2$. We then solve for the unknown h to get

$$h = \frac{kx^2}{2mg} = \frac{(140\text{ N/m})(0.11\text{ m})^2}{(2)(0.050\text{ kg})(9.8\text{ m/s}^2)} = 1.7\text{ m}$$

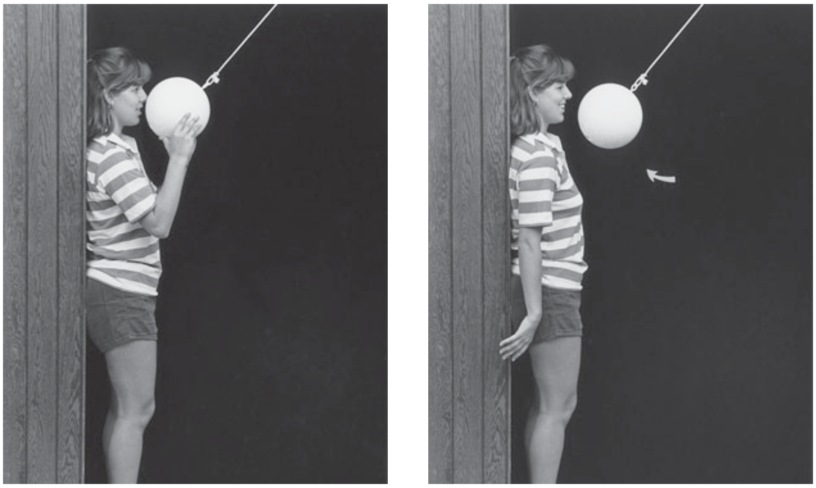
ASSESS Again, the answer in algebraic form makes sense; the stiffer the spring or the more it's compressed, the higher the block will go. But if the block is more massive or gravity is stronger, then the block won't get as far.

✓TIP Save Steps

You might be tempted to solve first for the block's speed when it leaves the spring and then equate $\frac{1}{2}mv^2$ to mgh to find the height. You could—but conservation of energy shortcuts all the details, getting you right from the initial to the final state. As long as energy is conserved, you don't need to worry about what happens in between.

■

GOT IT? 7.3 A bowling ball is tied to the end of a long rope and suspended from the ceiling. A student stands at one side of the room and holds the ball to her nose, then releases it from rest. Should she duck as it swings back?



Nonconservative Forces

In these examples we've assumed that energy is strictly conserved. In the everyday world of friction and other nonconservative forces, conservation of energy is sometimes a good approximation and sometimes not. When it's not, we can still apply our strategy, but now Equation 7.5 shows that we need to subtract any energy lost to nonconservative forces.

EXAMPLE 7.6 Nonconservative Forces: A Sliding Block

A block of mass m is launched from a spring of constant k that's initially compressed a distance x_0. After leaving the spring, the block slides on a horizontal surface with frictional coefficient μ. Find an expression for the distance the block slides before coming to rest.

INTERPRET The presence of friction means that mechanical energy isn't conserved. But we can still identify the kinetic and potential energy in the initial state: The kinetic energy is zero and the potential energy is that of the spring. In the final state, there's no mechanical energy at all. The nonconservative frictional force does negative work on the block, reducing its total energy. The block comes to rest when all its initial energy is gone.

DEVELOP Figure 7.8 shows the situation. With $K_0 = 0$, we determine the total initial energy from Equation 7.4, $U_0 = \frac{1}{2}kx_0^2$. The work W_f done by friction follows from Equation 6.1, $W = F_x\,\Delta x$. Here the frictional force has magnitude $f_f = \mu n = \mu mg$ and so, with its direction opposite the displacement, the frictional work is negative: $W_f = -\mu mg\,\Delta x$. This is the work W_{nc} in Equation 7.5, $\Delta K + \Delta U = W_{nc}$. The initial and final states here have no kinetic energy, so $\Delta K = 0$. Then the block will have lost all its initial energy when $\Delta U = -U_0$. Therefore, Equation 7.5 becomes $-\frac{1}{2}kx_0^2 = -\mu mg\,\Delta x$.

EVALUATE We solve this equation for the unknown distance Δx to get $\Delta x = kx_0^2/2\mu mg$. Since we weren't given numbers, there's nothing further to evaluate.

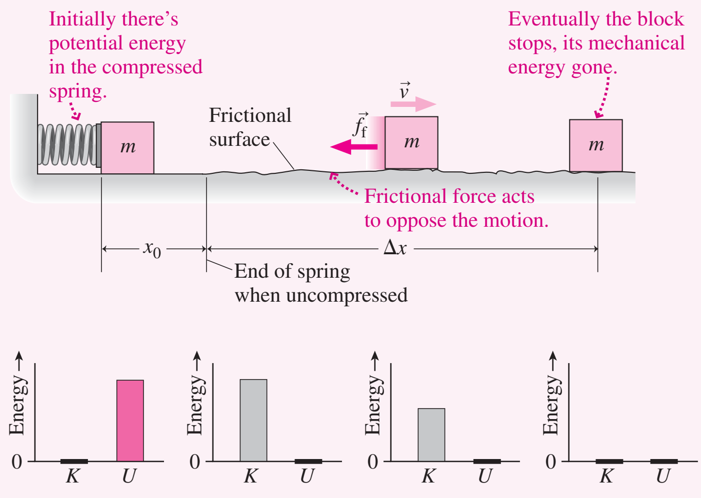


FIGURE 7.8 Intermediate bar charts show gradual loss of mechanical energy.

ASSESS Make sense? The stiffer the spring or the more it's compressed, the farther the block goes. The greater the friction or the normal force mg, the sooner the block stops. If $\mu = 0$, mechanical energy is once again conserved; then our result shows that the block would slide forever. ■

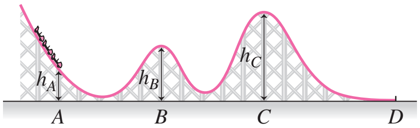


FIGURE 7.9 A roller-coaster track.

7.4 Potential-Energy Curves

Figure 7.9 shows a frictionless roller-coaster track. How fast must a car be coasting at point *A* if it's to reach point *D*? Conservation of energy provides the answer. To get to *D*, the car must clear peak *C*. Clearing *C* requires that the car's total energy exceed its potential energy at *C*; that is, $\frac{1}{2}mv_A^2 + mgh_A > mgh_C$, where we've taken the zero of potential energy at the bottom of the track. Solving for v_A gives $v_A > \sqrt{2g(h_C - h_A)}$. If v_A satisfies this inequality, the car will reach *C* with some kinetic energy remaining and will coast over the peak.

Figure 7.9 is a drawing of the actual roller-coaster track. But because gravitational potential energy is directly proportional to height, it's also a plot of potential energy versus position: a **potential-energy curve.** Conceptual Example 7.1 shows how we can study the car's motion by plotting its total energy on the same graph as the potential-energy curve.

CONCEPTUAL EXAMPLE 7.1 Potential-Energy Curves

Figure 7.10 plots the potential energy of our roller coaster, along with three possible values for its total mechanical energy. Since mechanical energy is conserved in the absence of nonconservative forces, the total-energy curve is a horizontal line. Use these graphs to describe the motion of a roller-coaster car, initially at point *A* and moving to the right.

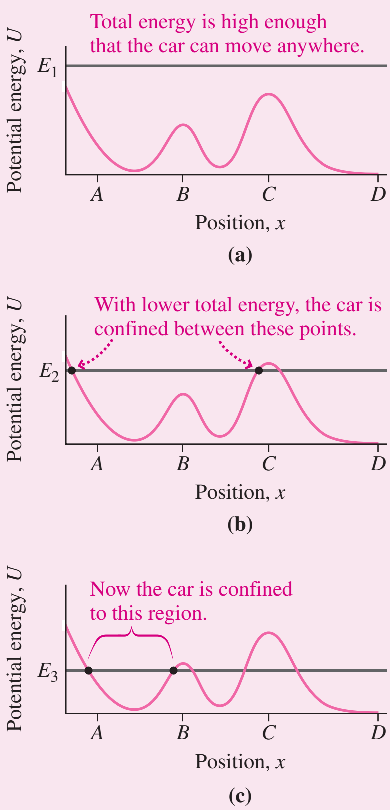


FIGURE 7.10 Potential and total energy for a roller-coaster car.

EVALUATE We're assuming there are no nonconservative forces (an approximation for a real roller coaster), so mechanical energy is conserved. In each figure, the sum of kinetic and potential energy therefore remains equal to the value set by the line indicating the total energy. When the roller-coaster car rises, its potential energy increases and its kinetic energy consequently decreases. But as long as its potential energy remains below its total energy, it still has kinetic energy and is still moving. Anywhere its potential energy equals its total energy, it has no kinetic energy and is momentarily at rest.

In Fig. 7.10*a* the car's total energy exceeds the maximum potential energy. Therefore it can move anywhere from its initial position at *A*. Since it's initially moving to the right, it will clear peaks *B* and *C* and will end up at *D* still moving to the right—and, since *D* is lower than *A*, it will be moving faster than it was at *A*.

In Fig. 7.10*b* the highest peak in the potential-energy curve exceeds the total energy; so does the very leftmost portion of the curve. Therefore, the car will move rightward from *A*, clearing peak *B*, but will come to a stop just before peak *C*, at the point where its potential energy equals its total energy—its right-hand turning point. Then it will roll back down to the left, again clearing peak *B* and climbing to another turning point where the potential-energy curve and total-energy line again intersect. Absent friction, it will run back and forth between the two turning points.

In Fig. 7.10*c* the total energy is lower, and the car can't clear peak *B*. So now it will run back and forth between the two turning points we've marked.

ASSESS Make sense? Yes: The higher the total energy, the larger the extent of the car's allowed motion. That's because, for a given potential energy, it has more energy available as kinetic energy.

MAKING THE CONNECTION Find a condition on the speed at *A* that will allow the car to move beyond peak *B*.

EVALUATE With a total energy equal to U_B, the car could just barely clear peak *B*. Its initial energy is $\frac{1}{2}mv_A^2 + mgh_A$, where v_A and h_A are the speed and height at *A*, and where we've taken the zero of potential energy at the bottom of the curve. Requiring that this quantity exceed $U_B = mgh_B$ then gives $v_A > \sqrt{2g(h_B - h_A)}$.

Even though the car in Figs. 7.10*b* and *c* can't get to *D*, its total energy still exceeds the potential energy at *D*. But it's blocked from reaching *D* by the **potential barrier** of peak *C*. We say that it's **trapped** in a **potential well** between its turning points.

Potential-energy curves are useful even with nongravitational forces where there's no direct correspondence with hills and valleys. The terminology used here—potential barriers, wells, and trapping—remains appropriate in such cases and indeed is widely used throughout physics.

Figure 7.11 shows the potential energy of a pair of hydrogen atoms as a function of their separation. This energy is associated with attractive and repulsive electrical forces involving the electrons and the nuclei of the two atoms. The potential-energy curve exhibits a potential well, showing that the atoms can form a **bound system** in which they're unable to separate fully. That bound system is a hydrogen molecule (H_2). The minimum energy, -7.6×10^{-19} J, corresponds to the molecule's equilibrium separation of 0.074 nm. It's convenient to define the zero of potential energy when the atoms are infinitely far apart; Fig. 7.11 then shows that any total energy less than zero results in a bound system. But if the total energy is greater than zero, the atoms are free to move arbitrarily far apart, so they don't form a molecule.

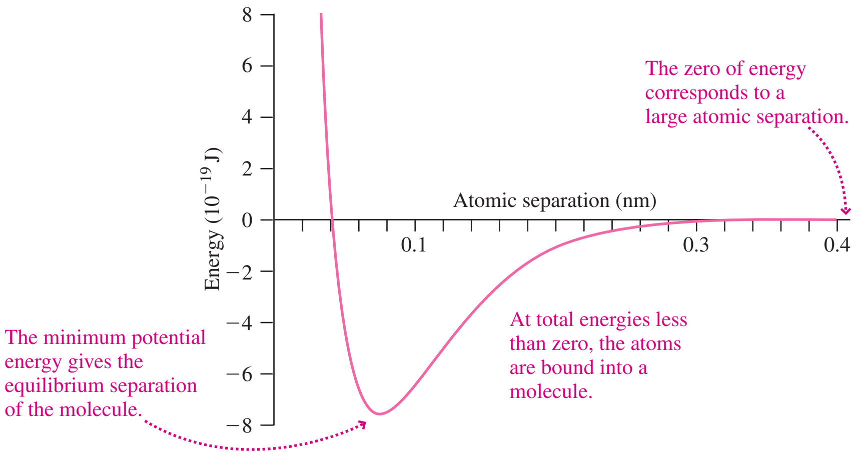


FIGURE 7.11 Potential-energy curve for two hydrogen atoms.

EXAMPLE 7.7 Molecular Energy: Finding Atomic Separation

Very near the bottom of the potential well in Fig. 7.11, the potential energy of the two hydrogen atoms is given approximately by $U = U_0 + a(x - x_0)^2$, where $U_0 = -0.760$ aJ, $a = 286$ aJ/nm^2, and $x_0 = 0.0741$ nm is the equilibrium separation. What range of atomic separations is allowed if the total energy is -0.717 aJ?

INTERPRET This problem sounds complicated, with strange units and talk of molecular energies. But it's about just what's shown in Figs. 7.10 and 7.11. Specifically, we're given the total energy and asked to find the turning points—the points where the line representing total energy intersects the potential-energy curve. If the units look strange, remember the SI prefixes (there's a table inside the front cover), which we use to avoid writing large powers of 10. Here 1 aJ $= 10^{-18}$ J and 1 nm $= 10^{-9}$ m.

DEVELOP Figure 7.12 is a plot of the potential-energy curve from the function we've been given. The straight line represents the total energy E. The turning points are the values of atomic separation where the two curves intersect. We could read them off the graph, or we can solve algebraically by setting the total energy equal to the potential energy.

EVALUATE With the potential energy given by $U = U_0 + a(x - x_0)^2$ and the total energy E, we have turning points when $E = U_0 + a(x - x_0)^2$. We could solve directly for x, but then we'd have to use the quadratic formula. Solving for $x - x_0$ is easier:

$$x - x_0 = \pm\sqrt{\frac{E - U_0}{a}} = \pm\sqrt{\frac{-0.717\text{ aJ} - (-0.760\text{ aJ})}{286\text{ aJ/nm}^2}}$$
$$= \pm 0.0123\text{ nm}$$

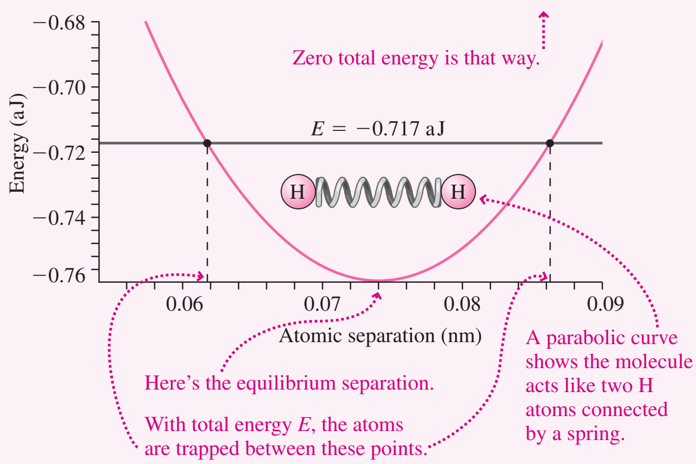


FIGURE 7.12 Analyzing the hydrogen molecule.

Then the turning points are at $x_0 \pm 0.0123$ nm—namely, 0.0864 nm and 0.0618 nm.

ASSESS Make sense? A look at Fig. 7.12 shows that we've correctly located the turning points. The fact that its potential-energy curve is parabolic (like a spring's $U = \frac{1}{2}kx^2$) shows that the molecule can be modeled approximately as two atoms joined by a spring. Chemists frequently use such models and even talk of the "spring constant" of the bond joining atoms into a molecule. ■

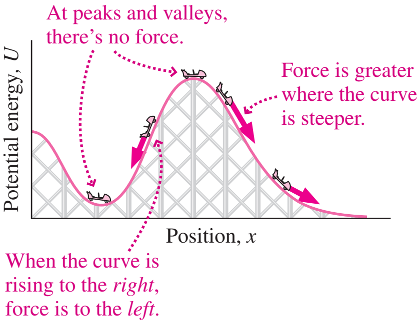


FIGURE 7.13 Force depends on the *slope* of the potential-energy curve.

Force and Potential Energy

The roller-coaster track in Fig. 7.9 traces the potential-energy curve for a car on the track. But it also shows the force acting to accelerate the car: Where the graph is steep—that is, where the potential energy is changing rapidly—the force is greatest. At the peaks and valleys, the force is zero. So it's the *slope* of the potential-energy curve that tells us about the force (Fig. 7.13).

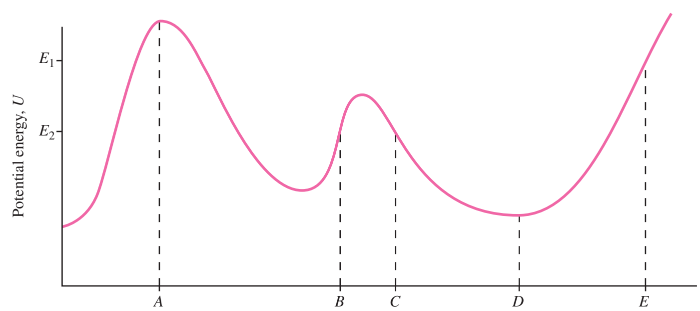
Just how strong is this force? Consider a small change Δx, so small that the force is essentially constant over this distance. Then we can use Equation 7.2b to write $\Delta U = -F_x \Delta x$, or $F_x = -\Delta U/\Delta x$. In the limit $\Delta x \to 0$, $\Delta U/\Delta x$ becomes the derivative, and we have

$$F_x = -\frac{dU}{dx} \tag{7.8}$$

This equation makes mathematical as well as physical sense. We've already written potential energy as the *integral* of force over distance, so it's no surprise that force is the *derivative* of potential energy. Equation 7.8 gives the force component in the x-direction only. In a three-dimensional situation, we'd have to take derivatives of potential energy with respect to y and z to find the full force vector.

Why the minus sign in Equation 7.8? You can see the answer in the molecular energy curve of Fig. 7.11, where pushing the atoms too close together—moving to the *left* of equilibrium—results in a repulsive force to the *right*, and pulling them apart—moving to the *right*—gives an attractive force to the *left*. You can see the same thing for the roller coaster in Fig. 7.13. In both cases the forces tend to drive the system back toward a minimum-energy state. We'll explore such minimum-energy equilibrium states further in Chapter 12.

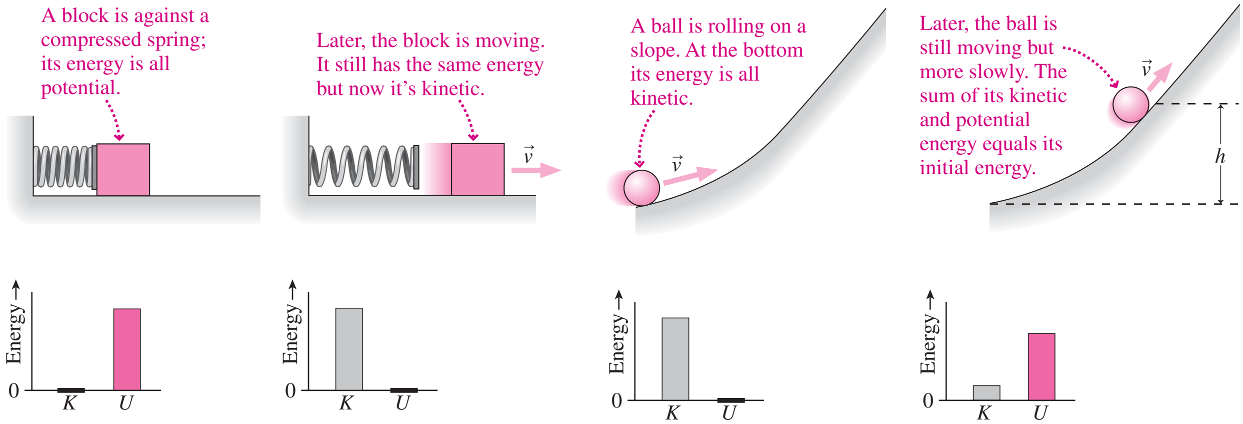
GOT IT? 7.4 The figure shows the potential energy for an electron in a microelectronic device. From among the labeled points, find (a) the point where the force on the electron is greatest; (b) the rightmost position possible if the electron has total energy E_1; (c) the leftmost position possible if the electron has total energy E_2 and starts out to the right of D; (d) a point where the force on the electron is zero; and (e) a point where the force on the electron points to the left. In some cases there may be more than one answer.



CHAPTER 7 SUMMARY

Big Picture

The big idea here is conservation of mechanical energy—the principle that the total energy of a system subject to only conservative forces cannot change. Energy may change from kinetic to potential, and vice versa, but the total remains constant. Applying conservation of energy means understanding the concept of potential energy as stored energy that results when work is done against a conservative force.



Key Concepts and Equations

The important new concept here is potential energy, defined as the negative of the work done by a conservative force. Only the change ΔU has physical significance. Expressions for potential energy include:

$$\Delta U_{AB} = -\int_A^B \vec{F} \cdot d\vec{r}$$

This one is the most general, but it's mathematically involved. The force can vary over an arbitrary path between points A and B.

$$\Delta U = -\int_{x_1}^{x_2} F(x)dx$$

This is a special case, when force and displacement are in the same direction and force may vary with position.

$$\Delta U = -F\,(x_2 - x_1)$$

This is the most specialized case, where the force is constant.

Given the concept of potential energy, the principle of conservation of mechanical energy follows from the work-energy theorem of Chapter 6. Here's the mathematical statement of energy conservation:

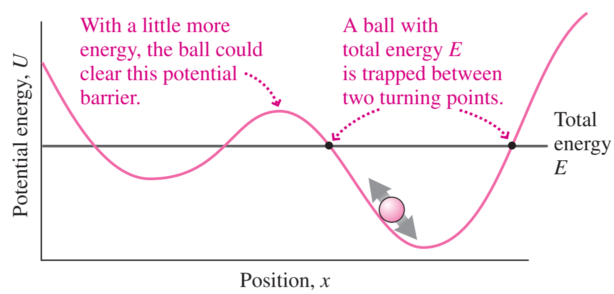
$$K + U = K_0 + U_0$$

K and U are the kinetic and potential energy at some point where we don't know one of these quantities.

The total energy is conserved, as indicated by the equal sign.

K_0 and U_0 are the kinetic and potential energy at some point where both are known. $K_0 + U_0$ is the *total energy*.

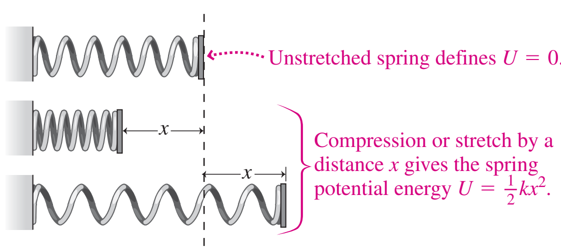
We can describe a wide range of systems—from molecules to roller coasters to planets—in terms of **potential-energy curves**. Knowing the total energy then lets us find **turning points** that determine the range of motion available to the system.

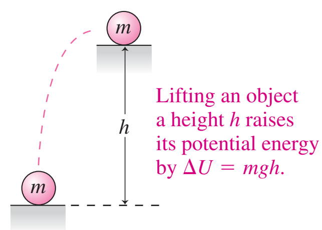


Applications

Two important cases of potential energy are the elastic potential energy of a spring, $U = \frac{1}{2}kx^2$, and the gravitational potential energy change, $\Delta U = mgh$, associated with lifting an object of mass m through a height h.

The former is limited to ideal springs for which $F = -kx$, the latter to the proximity of Earth's surface, where the variation of gravity with height is negligible.





For Thought and Discussion

1. Figure 7.14 shows force vectors at different points in space for two forces. Which is conservative and which nonconservative? Explain.

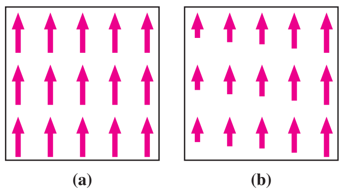


FIGURE 7.14 For Thought and Discussion 1; Problem 29

2. Is the conservation-of-energy principle related to Newton's laws, or is it an entirely separate physical principle? Discuss.
3. Why can't we define a potential energy associated with friction?
4. Can potential energy be negative? Can kinetic energy? Can total mechanical energy? Explain.
5. If the potential energy is zero at a given point, must the force also be zero at that point? Give an example.
6. If the force is zero at a given point, must the potential energy also be zero at that point? Give an example.
7. If the difference in potential energy between two points is zero, does that necessarily mean that an object moving between those points experiences no force?
8. A tightrope walker follows an essentially horizontal rope between two mountain peaks of equal altitude. A climber descends from one peak and climbs the other. Compare the work done by the gravitational force on the tightrope walker and the climber.
9. If conservation of energy is a law of nature, why do we have programs—like mileage requirements for cars or insulation standards for buildings—designed to encourage energy conservation?

Exercises and Problems

Exercises

Section 7.1 Conservative and Nonconservative Forces

10. Determine the work done by the frictional force in moving a block of mass m from point 1 to point 2 over the two paths shown in Fig. 7.15. The coefficient of friction has the constant value over the surface. *Note:* The diagram lies in a horizontal plane.

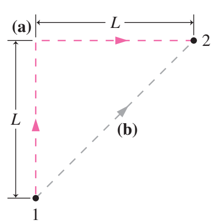


FIGURE 7.15 Exercises 10 and 11

11. Now take Fig. 7.15 to lie in a vertical plane, and find the work done by the gravitational force as an object moves from point 1 to point 2 over each of the paths shown.

Section 7.2 Potential Energy

12. Rework Example 7.1, now taking the zero of potential energy at street level.
13. Find the potential energy of a 70-kg hiker (a) atop New Hampshire's Mount Washington, 1900 m above sea level, and (b) in Death Valley, California, 86 m below sea level. Take the zero of potential energy at sea level.
14. You fly from Boston's Logan Airport, at sea level, to Denver, altitude 1.6 km. Taking your mass as 65 kg and the zero of potential energy at Boston, what's your gravitational potential energy (a) at the plane's 11-km cruising altitude and (b) in Denver?
15. A 60-kg hiker ascending 1250-m-high Camel's Hump mountain in Vermont has potential energy -240 kJ; the zero of potential energy is taken at the mountaintop. What's her altitude?
16. How much energy can be stored in a spring with $k = 320$ N/m if the maximum allowed stretch is 18 cm?
17. How far would you have to stretch a spring with $k = 1.4$ kN/m for it to store 210 J of energy?
18. BIO A biophysicist grabs the ends of a DNA strand with optical tweezers and stretches it 26 μm. How much energy is stored in the stretched molecule if its spring constant is 0.046 pN/μm?

Section 7.3 Conservation of Mechanical Energy

19. A skier starts down a frictionless 32° slope. After a vertical drop of 25 m, the slope temporarily levels out and then slopes down at 20°, dropping an additional 38 m vertically before leveling out again. Find the skier's speed on the two level stretches.
20. A 10,000-kg Navy jet lands on an aircraft carrier and snags a cable to slow it down. The cable is attached to a spring with $k = 40$ kN/m. If the spring stretches 25 m to stop the plane, what was its landing speed?
21. A 120-g arrow is shot vertically from a bow whose effective spring constant is 430 N/m. If the bow is drawn 71 cm before shooting, to what height does the arrow rise?
22. In a railroad yard, a 35,000-kg boxcar moving at 7.5 m/s is stopped by a spring-loaded bumper mounted at the end of the level track. If $k = 2.8$ MN/m, how far does the spring compress in stopping the boxcar?
23. You work for a toy company, and you're designing a spring-launched model rocket. The launching apparatus has room for a spring that can be compressed 14 cm, and the rocket's mass is 65 g. If the rocket is to reach an altitude of 35 m, what should you specify for the spring constant?

Section 7.4 Potential-Energy Curves

24. A particle slides along the frictionless track shown in Fig. 7.16, starting at rest from point A. Find (a) its speed at B, (b) its speed at C, and (c) the approximate location of its right-hand turning point.

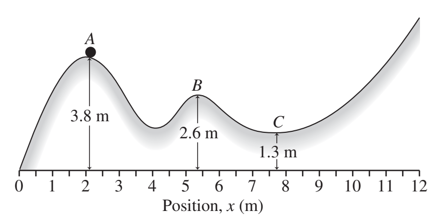


FIGURE 7.16 Exercise 24

25. A particle slides back and forth on a frictionless track whose height as a function of horizontal position x is $y = ax^2$, where $a = 0.92\ \text{m}^{-1}$. If the particle's maximum speed is 8.5 m/s, find its turning points.
26. Figure 7.17 shows a particle's potential-energy curve. Find the force on the particle at each of the labeled curve segments.

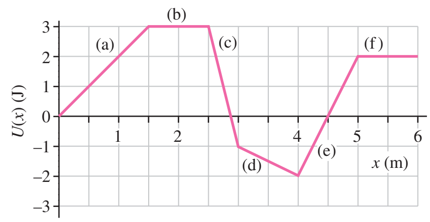


FIGURE 7.17 Exercise 26

27. A particle is trapped in a potential well described by $U(x) = 16x^2 - b$, with U in joules, x in meters, and $b = 4.0$ J. Find the force on the particle when it's at (a) $x = 2.1$ m, (b) $x = 0$, and (c) $x = -1.4$ m.

Problems

28. The reservoir at Northfield Mountain Pumped Storage Project is 214 m above the pump/generators and holds 2.1×10^{10} kg of water (see Application on p. 105). The generators can produce electrical energy at the rate of 1.08 GW. Find (a) the total gravitational energy stored in the reservoir, taking zero potential energy at the generators, and (b) the length of time the station can generate power before the reservoir is drained.
29. The force in Fig. 7.14*a* is given by $\vec{F}_\text{a} = F_0\hat{\jmath}$, where F_0 is a constant. The force in Fig. 7.14*b* is given by $\vec{F}_\text{b} = F_0(x/a)\hat{\jmath}$, where the origin is at the lower left corner of the box, a is the width of the square box, and x increases horizontally to the right. Determine the work done by forces $\vec{F}_\text{a}$ and $\vec{F}_\text{b}$ on an object moved counterclockwise around each box, starting at the lower left corner.
30. An incline makes an angle θ with the horizontal. Find the gravitational potential energy associated with a mass m located a distance x measured along the incline. Take the zero of potential energy at the bottom of the incline.
31. A 1.50-kg brick measures 20.0 cm × 8.00 cm × 5.50 cm. Taking the zero of potential energy when the brick lies on its broadest face, what's its potential energy (a) when the brick is standing on end and (b) when it's balanced on its 8-cm edge? (*Note:* You can treat the brick as though all its mass is concentrated at its center.)
32. A carbon monoxide molecule can be modeled as a carbon atom and an oxygen atom connected by a spring. If a displacement of the carbon by 1.6×10^{-12} m from its equilibrium position relative to the oxygen increases the molecule's potential energy by 0.015 eV, what's the spring constant?
33. A more accurate expression for the force law of the rope in Example 7.3 is $F = -kx + bx^2 - cx^3$, where k and b have the values given in Example 7.3 and $c = 3.1\ \text{N/m}^3$. Find the energy stored in stretching the rope 2.62 m. By what percentage does your result differ from that of Example 7.3?
34. **BIO** For small stretches, the Achilles tendon can be modeled as an ideal spring. Experiments using a particular tendon showed that it stretched 2.66 mm when a 125-kg mass was hung from it. (a) Find the spring constant of this tendon. (b) How much would it have to stretch to store 50.0 J of energy?
35. The force exerted by an unusual spring when it's compressed a distance x from equilibrium is $F = -kx - cx^3$, where $k = 220$ N/m and $c = 3.1\ \text{N/m}^3$. Find the stored energy when it's been compressed 15 cm.
36. The force on a particle is given by $\vec{F} = A\hat{\imath}/x^2$, where A is a positive constant. (a) Find the potential-energy difference between two points x_1 and x_2, where $x_1 > x_2$. (b) Show that the potential-energy difference remains finite even when $x_1 \to \infty$.
37. A particle moves along the x-axis under the influence of a force $F = ax^2 + b$, where a and b are constants. Find its potential energy as a function of position, taking $U = 0$ at $x = 0$.
38. As a highway engineer, you're asked to design a runaway truck lane on a mountain road. The lane will head uphill at 30° and should be able to accommodate a 16,000-kg truck with failed brakes entering the lane at 110 km/h. How long should you make the lane? Neglect friction.
39. A spring of constant k, compressed a distance x, is used to launch a mass m up a frictionless slope at angle θ. Find an expression for the maximum distance along the slope that the mass moves after leaving the spring.
40. A child is on a swing whose 3.2-m-long chains make a maximum angle of 50° with the vertical. What's the child's maximum speed?
41. With $x - x_0 = h$ and $a = g$, Equation 2.11 gives the speed of an object thrown downward with initial speed v_0 after it's dropped a distance h: $v = \sqrt{v_0^2 + 2gh}$. Use conservation of energy to derive the same result.
42. **BIO** The *nuchal ligament* is a cord-like structure that runs along the back of the neck and supports much of the head's weight in animals like horses and cows. The ligament is extremely stiff for small stretches, but loosens as it stretches further, thus functioning as a biological shock absorber. Figure 7.18 shows the force-distance curve for a particular nuchal ligament; the curve can be modeled approximately by the expression $F(x) = 0.43x - 0.033x^2 + 0.00086x^3$, with F in kN and x in cm. Find the energy stored in the ligament when it's been stretched (a) 7.5 cm and (b) 15 cm.

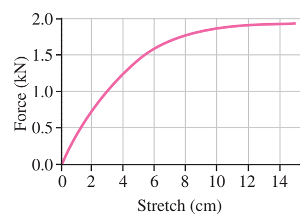


FIGURE 7.18 Problem 42

43. A 200-g block slides back and forth on a frictionless surface between two springs, as shown in Fig. 7.19. The left-hand spring has $k = 130$ N/m and its maximum compression is 16 cm. The right-hand spring has $k = 280$ N/m. Find (a) the maximum compression of the right-hand spring and (b) the speed of the block as it moves between the springs.

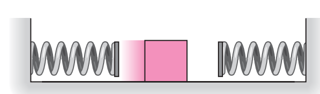


FIGURE 7.19 Problem 43

44. Current automotive standards call for bumpers that sustain essentially no damage in a 4-km/h collision with a stationary object. As an automotive engineer, you'd like to improve on that. You've developed a spring-mounted bumper with effective

spring constant 1.3 MN/m. The springs can compress up to 5.0 cm before damage occurs. For a 1400-kg car, what do you claim as the maximum collision speed?

45. A block slides on the frictionless loop-the-loop track shown in Fig. 7.20. Find the minimum height h at which it can start from rest and still make it around the loop.

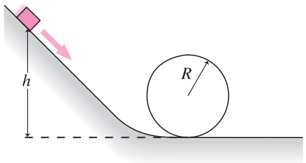


FIGURE 7.20 Problem 45

46. The maximum speed of the pendulum bob in a grandfather clock is 0.55 m/s. If the pendulum makes a maximum angle of 8.0° with the vertical, what's the pendulum's length?
47. A mass m is dropped from height h above the top of a spring of constant k mounted vertically on the floor. Show that the spring's maximum compression is given by $(mg/k)(1 + \sqrt{1 + 2kh/mg})$.
48. A particle with total energy 3.5 J is trapped in a potential well described by $U = 7.0 - 8.0x + 1.7x^2$, where U is in joules and x in meters. Find its turning points.
49. (a) Derive an expression for the potential energy of an object subject to a force $F_x = ax - bx^3$, where $a = 5$ N/m and $b = 2$ N/m^3, taking $U = 0$ at $x = 0$. (b) Graph the potential-energy curve for $x > 0$ and use it to find the turning points for an object whose total energy is -1 J.
50. In ionic solids such as NaCl (salt), the potential energy of a pair of ions takes the form $U = b/r^n - a/r$, where r is the separation of the ions. For NaCl, a and b have the SI values 4.04×10^{-28} and 5.52×10^{-98}, respectively, and $n = 8.22$. Find the equilibrium separation in NaCl.
51. Repeat Exercise 19 for the case when the coefficient of kinetic friction on both slopes is 0.11, while the level stretches remain frictionless.
52. As an energy-efficiency consultant, you're asked to assess a pumped-storage facility. Its reservoir sits 140 m above its generating station and holds 8.5×10^9 kg of water. The power plant generates 330 MW of electric power while draining the reservoir over an 8.0-h period. Its efficiency is the percentage of the stored potential energy that gets converted to electricity. What efficiency do you report?
53. A spring of constant $k = 340$ N/m is used to launch a 1.5-kg block along a horizontal surface whose coefficient of sliding friction is 0.27. If the spring is compressed 18 cm, how far does the block slide?
54. A bug slides back and forth in a bowl 11 cm deep, starting from rest at the top, as shown in Fig. 7.21. The bowl is frictionless except for a 1.5-cm-wide sticky patch on its flat bottom, where the coefficient of friction is 0.61. How many times does the bug cross the sticky region?

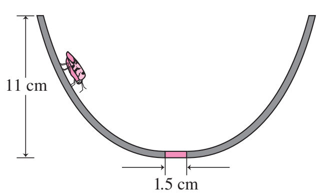


FIGURE 7.21 Problem 54

55. A 190-g block is launched by compressing a spring of constant $k = 200$ N/m by 15 cm. The spring is mounted horizontally, and the surface directly under it is frictionless. But beyond the equilibrium position of the spring end, the surface has frictional coefficient $\mu = 0.27$. This frictional surface extends 85 cm, followed by a frictionless curved rise, as shown in Fig. 7.22. After it's launched, where does the block finally come to rest? Measure from the left end of the frictional zone.

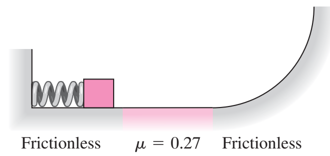


FIGURE 7.22 Problem 55

56. A block slides down a frictionless incline that terminates in a 45° ramp, as shown in Fig. 7.23. Find an expression for the horizontal range x shown in the figure as a function of the heights h_1 and h_2.

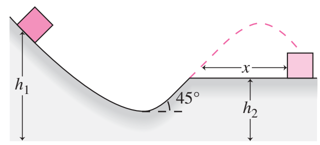


FIGURE 7.23 Problem 56

57. An 840-kg roller-coaster car is launched from a giant spring with $k = 31$ kN/m into a frictionless loop-the-loop track of radius 6.2 m, as shown in Fig. 7.24. What's the minimum spring compression that will ensure the car stays on the track?

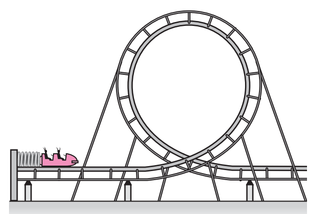


FIGURE 7.24 Problem 57

58. A particle slides back and forth in a frictionless bowl whose height is given by $h = 0.18x^2$, with x and h in meters. Find the x coordinates of its turning points if the particle's maximum speed is 47 cm/s.
59. A child sleds down a frictionless hill whose vertical drop is 7.2 m. At the bottom is a level but rough stretch where the coefficient of kinetic friction is 0.51. How far does she slide across the level stretch?
60. A bug lands on top of the frictionless, spherical head of a bald man. It begins to slide down his head (Fig. 7.25). Show that the

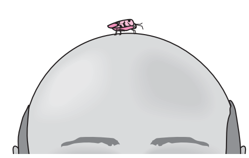


FIGURE 7.25 Problem 60

bug leaves the head when it has dropped a vertical distance one-third of the head's radius.

61. A particle of mass m is subject to a force $\vec{F} = (a\sqrt{x})\hat{\imath}$, where a is a constant. The particle is initially at rest at the origin and is given a slight nudge in the positive x-direction. Find an expression for its speed as a function of position x.
62. A 17-m-long vine hangs vertically from a tree on one side of a 10-m-wide gorge, as shown in Fig. 7.26. Tarzan runs up, hoping to grab the vine, swing over the gorge, and drop vertically off the vine to land on the other side. How fast must he run?

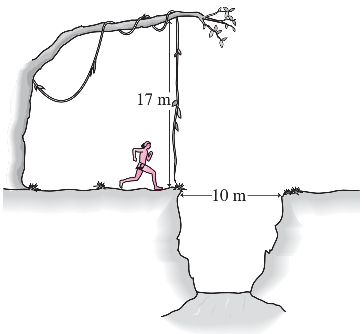


FIGURE 7.26 Problem 62

63. A block of weight 4.5 N is launched up a 30° inclined plane 2.0 m long by a spring with $k = 2.0$ kN/m and maximum compression 10 cm. The coefficient of kinetic friction is 0.50. Does the block reach the top of the incline? If so, how much kinetic energy does it have there? If not, how close to the top, along the incline, does it get?
64. Your engineering department is asked to evaluate the performance of a new 250-hp sports car. You know that 30% of the engine's power can be converted to mechanical energy of the 1500-kg car, and that the power delivered is independent of the car's velocity. What do you report for the time it will take to accelerate from rest to 60 mi/h on a level road?
65. Your roommate is writing a science fiction novel and asks your advice about a plot point. Her characters are mining ore on the Moon and launching it toward Earth. Bins with 1000 kg of ore will be launched by a large spring, to be compressed 15 m. It takes a speed of 2.4 km/s to escape the Moon's gravity. What do you tell her is an appropriate spring constant?
66. You have a summer job at your university's zoology department, where you'll be working with an animal behavior expert. She's assigned you to study videos of different animals leaping into the air. Your task is to compare their power outputs as they jump. You'll have the mass m of each animal from data collected in the field. From the videos, you'll be able to measure both the vertical distance d over which the animal accelerates when it pushes off the ground and the maximum height h it reaches. Your task is to find an algebraic expression for power in terms of these parameters.
67. Biomechanical engineers developing artificial limbs for prosthetic and robotic applications have developed a two-spring design for their replacement Achilles tendon. The first spring has constant k and the second ak, where $a > 1$. When the artificial tendon is stretched from $x = 0$ to $x = x_1$, only the first spring is engaged. For $x > x_1$, a mechanism engages the second spring, giving a configuration like that described in part (a) of Chapter 4's Problem 58. Find an expression for the energy stored in the artificial tendon when it's stretched a distance $2x_1$.

Passage Problems

Nuclear fusion is the process that powers the Sun. Fusion occurs when two low-mass atomic nuclei fuse together to make a larger nucleus, in the process releasing substantial energy. This is hard to achieve because atomic nuclei carry positive electric charge, and their electrical repulsion makes it difficult to get them close enough for the short-range nuclear force to bind them into a single nucleus. Figure 7.27 shows the potential-energy curve for fusion of two deuterons (heavy hydrogen nuclei). The energy is measured in million electron volts (MeV), a unit commonly used in nuclear physics, and the separation is in femtometers $(1 \text{ fm} = 10^{-15} \text{ m})$.

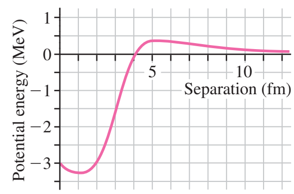


FIGURE 7.27 Potential energy for two deuterons (Passage Problems 68–71)

68. The force between the deuterons is zero at approximately
    a. 3 fm.
    b. 4 fm.
    c. 5 fm.
    d. the force is never zero.
69. In order for initially two widely separated deuterons to get close enough to fuse, their kinetic energy must be about
    a. 0.1 MeV.
    b. 3 MeV.
    c. −3 MeV.
    d. 0.3 MeV.
70. The energy available in fusion is the energy difference between that of widely separated deuterons and the bound deutrons after they've "fallen" into the deep potential well shown in the figure. That energy is about
    a. 0.3 MeV.
    b. 1 MeV.
    c. 3.3 MeV.
    d. 3.6 MeV.
71. When two deuterons are 4 fm apart, the force acting on them
    a. is repulsive.
    b. is attractive.
    c. is zero.
    d. can't be determined from the graph.

Answers to Chapter Questions

Answer to Chapter Opening Question

Potential energy turns into kinetic energy, sound, and heat.

Answers to GOT IT? Questions

7.1. On the curved paths, the work is greater for the trunk. The gravitational force is conservative, so the work is independent of path. The frictional force is nonconservative, however.

7.2. (b) The potential-energy change will be slightly less because at greater heights, the gravitational force is lower and so, therefore, is the work done in traversing a given distance.

7.3. No. Mechanical energy is conserved, so if the ball is released from rest, it cannot climb higher than its initial height.

7.4. (a) B; (b) E; (c) C; (d) A or D; (e) B or E.

8 Gravity

This TV dish points at a satellite in a fixed position in the sky. How does the satellite manage to stay at that position?

New Concepts, New Skills

By the end of this chapter you should be able to

- Calculate the gravitational force between two objects (8.2).
- Describe orbital motion and circular orbits quantitatively (8.3).
- Compute changes in gravitational potential energy over large distances (8.4).
- Relate kinetic and potential energy in the presence of gravity, and explain the concept of escape speed (8.4).
- Describe qualitatively the concept of the gravitational field (8.5).

Connecting Your Knowledge

- Much of the material in this chapter is an application of the concepts of potential and kinetic energy, in particular the principle of conservation of mechanical energy (7.2, 7.3).
- You should understand that calculating potential energy for position-varying forces requires integration, as described in Equations 7.2a and 7.2b and Example 7.3 (7.2).

Gravity is the most obvious of nature's fundamental forces. Theories of gravity have brought us new understandings of the nature and evolution of the universe. We've used our knowledge of gravity to explore the solar system and to engineer a host of space-based technologies. In nearly all applications we still use the theory of gravity that Isaac Newton developed in the 1600s. Only in the most extreme astrophysical situations or where—as with Global Positioning System satellites—we need exquisite precision do we use the successor to Newtonian gravitation, namely, Einstein's general theory of relativity.

8.1 Toward a Law of Gravity

Newton's theory of gravity was the culmination of two centuries of scientific revolution that began in 1543 with Polish astronomer Nicolaus Copernicus's radical suggestion that the planets orbit not Earth but the Sun. Fifty years after Copernicus's work was published, the Danish noble Tycho Brahe began a program of accurate planetary observations. After Tycho's death in 1601, his assistant Johannes Kepler worked to make sense of the observations. Success came when Kepler took a radical step: He gave up the long-standing idea that the planets moved in perfect circles. Kepler summarized his new insights in three laws, described in Fig. 8.1. Kepler based his laws solely on observation and gave no theoretical explanation. So Kepler knew *how* the planets moved, but not *why*.

Shortly after Kepler published his first two laws, Galileo trained his first telescopes on the heavens. Among his discoveries were four moons orbiting Jupiter, sunspots that blemished the supposedly perfect sphere of the Sun, and the phases of Venus (Fig. 8.2). His observations called into question the notion that all celestial objects were perfect and also lent credence to the Copernican view of the Sun as the center of planetary motion.

By Newton's time the intellectual climate was ripe for the culmination of the revolution that had begun with Copernicus. Legend has it that Newton was sitting under an apple tree when an apple struck him on the head, causing him to discover gravity. That story is probably a myth, but if it were true the other half would be that Newton was staring at the Moon when the apple struck. Newton's genius was to recognize that *the motion of the apple and the motion of the Moon were the same, that both were "falling" toward Earth under the influence of the same force*. Newton called this force **gravity**, from the Latin *gravitas*, "heaviness." In one of the most sweeping syntheses in human thought, Newton inferred that everything in the universe, on Earth and in the celestial realm, obeys the same physical laws.

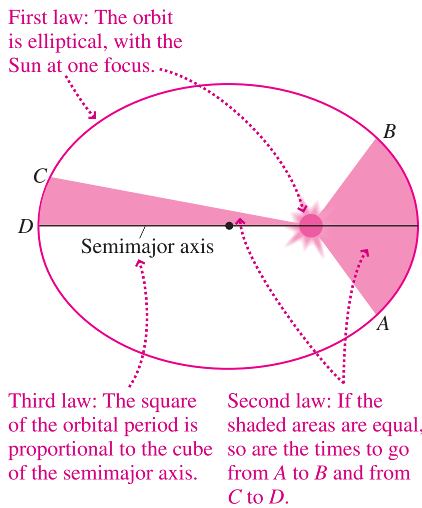


FIGURE 8.1 Kepler's laws.

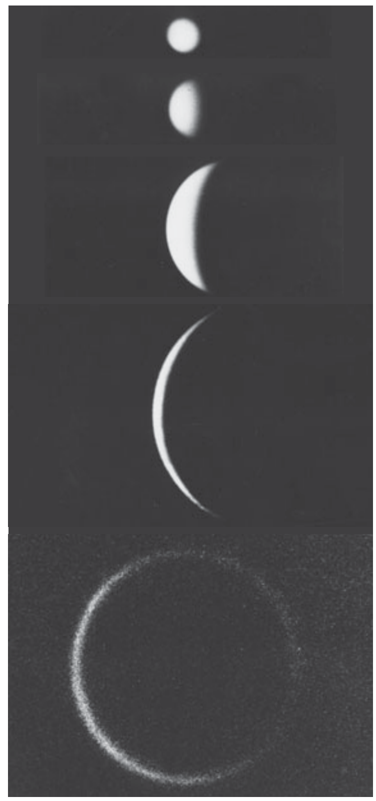


FIGURE 8.2 Phases of Venus. In an Earth-centered system, Venus would always appear the same size because of its constant distance from Earth.

8.2 Universal Gravitation

Newton generalized his new understanding of gravity to suggest that any two particles in the universe exert attractive forces on each other, with magnitude given by

$$F = \frac{Gm_1m_2}{r^2} \qquad \text{(universal gravitation)} \tag{8.1}$$

Here m_1 and m_2 are the particle masses, r the distance between them, and G the **constant of universal gravitation**, whose value—determined after Newton's time—is $6.67 \times 10^{-11}\ \mathrm{N \cdot m^2/kg^2}$. The constant G is truly universal; observation and theory suggest that it has the same value throughout the universe.

The force of gravity acts *between* two particles; that is, m_1 exerts an attractive force on m_2, and m_2 exerts an equal but oppositely directed force on m_1. The two forces therefore obey Newton's third law.

Newton's law of universal gravitation applies strictly only to point particles that have no extent. But, as Newton showed using his newly developed calculus, it also holds for spherically symmetric objects of any size if the distance r is measured from their centers. It also applies approximately to arbitrarily shaped objects provided the distance between them is large compared with their sizes. For example, the gravitational force of Earth on the space station is given accurately by Equation 8.1 because (1) Earth is essentially spherical and (2) the station, though irregular in shape, is vastly smaller than its distance from Earth's center.

EXAMPLE 8.1 The Acceleration of Gravity: On Earth and in Space

Use the law of universal gravitation to find the acceleration of gravity at Earth's surface, at the 380-km altitude of the International Space Station, and on the surface of Mars.

INTERPRET The problem statement tells us this is about universal gravitation, but what's that got to do with the acceleration of gravity? The gravitational force is what causes that acceleration, so we can interpret this problem as being about the force between Earth (or Mars) and some arbitrary mass.

DEVELOP Since the problem involves universal gravitation, Equation 8.1 applies. But we're asked about acceleration, not force. Newton's second law, $F = ma$, relates the two. So our plan is to use Equation 8.1,

(continued)

$F = Gm_1m_2/r^2$, to find the gravitational force on an arbitrary mass and then use Newton's second law to get the acceleration. There's another bit of planning: We need to find the masses of Earth and Mars and their radii. Astrophysical data like these are in Appendix E.

EVALUATE Equation 8.1 gives the force a planet of mass M exerts on an arbitrary mass m a distance r from the planet's center: $F = GMm/r^2$. (Here we set m_1 in Equation 8.1 to the large planetary mass M, and m_2 to the smaller mass m.) But Newton's second law says that this force is equal to the product of mass and acceleration for a body in free fall, so we can write $ma = GMm/r^2$. The mass m cancels, and we're left with the acceleration:

$$a = \frac{GM}{r^2} \tag{8.2}$$

The distance r is measured from the *center* of the object providing the gravitational force, so to find the acceleration at Earth's surface we use R_E, the radius of the Earth, for r. Taking R_E and M_E from Appendix E, we have

$$a = \frac{GM_E}{R_E{}^2} = \frac{(6.67 \times 10^{-11}\ \mathrm{N{\cdot}m^2/kg^2})(5.97 \times 10^{24}\ \mathrm{kg})}{(6.37 \times 10^6\ \mathrm{m})^2} = 9.81\ \mathrm{m/s^2}$$

This, of course, is the value of g—the acceleration due to gravity at Earth's surface.

At the space station's altitude, we have $r = R_E + 380$ km, so

$$a = \frac{GM_E}{r^2} = \frac{(6.67 \times 10^{-11}\ \mathrm{N{\cdot}m^2/kg^2})(5.97 \times 10^{24}\ \mathrm{kg})}{(6.37 \times 10^6\ \mathrm{m} + 380 \times 10^3\ \mathrm{m})^2} = 8.74\ \mathrm{m/s^2}$$

A similar calculation using Appendix E data yields 3.75 m/s^2 for the acceleration of gravity at the surface of Mars.

ASSESS As we've seen, our result for Earth is just what we expect. The acceleration at the space station is lower but still about 90% of the surface value. This confirms Chapter 4's point that weightlessness doesn't mean the absence of gravity. Rather, as Equation 8.2 shows, an object's gravitational acceleration is independent of its mass—so all objects "fall" together. Finally, our answer for Mars is lower than for Earth, as befits its smaller mass—although not as much lower as mass alone would imply. That's because Mars is also smaller, so r in the denominator of Equation 8.2 is lower. ■

✓TIP *G and g*

Don't confuse G and g! Both quantities are associated with gravity, but G is a universal constant, while g describes the gravitational acceleration at a particular place—namely, Earth's surface—and its value depends on Earth's size and mass.

The variation of gravitational acceleration with distance from Earth's center provided Newton with a clue that the gravitational force should vary as the inverse square of the distance. Newton knew the Moon's orbital period and distance from Earth; from these he could calculate its orbital speed and thus its acceleration v^2/r. Newton found—as you can in Exercise 12—that the Moon's acceleration is about 1/3600 the gravitational acceleration g at Earth's surface. The Moon is about 60 times farther from Earth's center than is Earth's surface; since $60^2 = 3600$, the decrease in gravitational acceleration with distance from Earth's center is consistent with a gravitational force that varies as $1/r^2$.

TACTICS 8.1 **Understanding "Inverse Square"**

Newton's universal gravitation is the first of several inverse-square force laws you'll encounter, and it's important to understand what this term means. In Equation 8.1 the distance r between the two masses is *squared*, and it occurs in the *denominator*; hence the force depends on the *inverse square* of the distance. Double the distance and the force drops to $1/2^2$, or 1/4 of its original value. Triple the distance and the force drops to $1/3^2$, or 1/9. Although you can always grind through the arithmetic of Equation 8.1, you should use these simple ratio calculations whenever possible. The same considerations apply to gravitational acceleration, since it's proportional to force (Fig. 8.3).

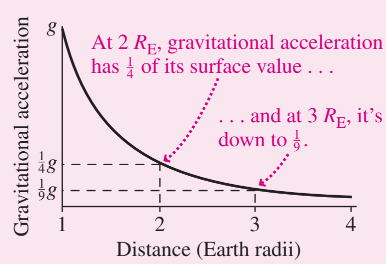


FIGURE 8.3 Meaning of the inverse-square law.

GOT IT? 8.1 Suppose the distance between two objects is cut in half. Is the gravitational force between them (a) quartered, (b) halved, (c) doubled, or (d) quadrupled?

The Cavendish Experiment: Weighing the Earth

Given the mass and radius of the Earth and the measured value of g, we could use Equation 8.1 to determine the universal constant G. Unfortunately, the only way to determine Earth's mass accurately is to measure its gravitational effect and then use Equation 8.1. But that requires knowing G.

To determine G, we need to measure the gravitational force of a *known* mass. Given the weak gravitational force of normal-size objects, this is a challenging task. It was accomplished in 1798 through an ingenious experiment by the British physicist Henry Cavendish. Cavendish mounted two 5-cm-diameter lead spheres on the ends of a rod suspended from a thin fiber. He then brought two 30-cm lead spheres nearby (Fig. 8.4). Their gravitational attraction caused a slight movement of the small spheres, twisting the fiber. Knowing the properties of the fiber, Cavendish could determine the force. With the known masses and their separation, he then used Equation 8.1 to calculate G. His result determined the mass of the Earth; indeed, his published paper was entitled "On Weighing the Earth."

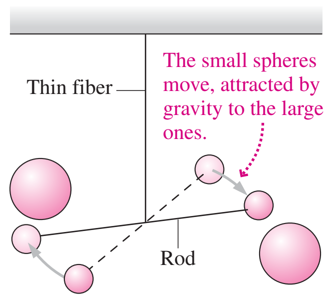


FIGURE 8.4 The Cavendish experiment to determine G.

Gravity is the weakest of the fundamental forces, and, as the Cavendish experiment suggests, the gravitational force between everyday objects is negligible. Yet gravity shapes the large-scale structure of matter and indeed the entire universe. Why, if it's so weak? The answer is that gravity, unlike the stronger electric force, is always attractive; there's no such thing as "negative mass." So large concentrations of matter produce substantial gravitational effects. Electric charge, in contrast, can be positive or negative, and electric effects in normal-sized objects tend to cancel. We'll explore this distinction further in Chapter 20.

8.3 Orbital Motion

Orbital motion occurs when gravity is the dominant force acting on a body. It's not just planets and spacecraft that are in orbit. An individual astronaut floating outside the space station is orbiting Earth. The Sun itself orbits the center of the galaxy, taking about 200 million years to complete one revolution. If we neglect air resistance, even a baseball is temporarily in orbit. Here we discuss quantitatively the special case of circular orbits; then we describe qualitatively the general case.

Newton's genius was to recognize that the Moon is held in its circular orbit by the same force that pulls an apple to the ground. From there, it was a short step for Newton to realize that human-made objects could be put into orbit. Nearly 300 years before the first artificial satellites, he imagined a projectile launched horizontally from a high mountain (Fig. 8.5). The projectile moves in a curve, as gravity pulls it from the straight-line path it would follow if no force were acting. As its initial speed is increased, the projectile travels farther before striking Earth. Finally, there comes a speed for which the projectile's path bends in a way that exactly follows Earth's curvature. It's then in **circular orbit**, continuing forever unless a nongravitational force acts.

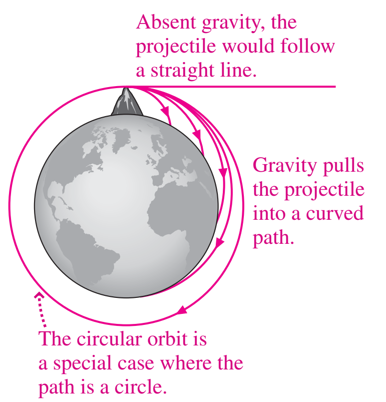


FIGURE 8.5 Newton's "thought experiment" showing that projectile and orbital motions are essentially the same.

Why doesn't an orbiting object fall toward Earth? It does! Under the influence of gravity, it gets ever closer to Earth than it would be on a straight-line path. It's behaving exactly as Newton's second law requires of an object under the influence of a force—by accelerating. For a *circular* orbit, that acceleration amounts to a change in the direction, but not the magnitude, of the orbiting object's velocity.

Remember that Newton's laws aren't so much about *motion* as they are about *changes* in motion. To ask why a satellite doesn't fall to Earth is to adopt the archaic Aristotelian view. The Newtonian question is this: Why doesn't the satellite move in a straight line? And the answer is simple: because a force is acting. That force—gravity—is exactly analogous to the tension force that keeps a ball on a string whirling in its circular path.

We can analyze circular orbits quantitatively because we know that a force of magnitude mv^2/r is required to keep an object of mass m and speed v in a circular path of radius r. In the case of an orbit, that force is gravity, so we have

$$\frac{GMm}{r^2} = \frac{mv^2}{r}$$

where m is the mass of the orbiting object and M the mass of the object about which it's orbiting. We assume here that $M \gg m$, so the gravitating object can be considered essentially at rest—a reasonable approximation with Earth satellites or planets orbiting the much more massive Sun. Solving for the orbital speed gives

$$v = \sqrt{\frac{GM}{r}} \qquad \text{(speed, circular orbit)} \qquad (8.3)$$

Often we're interested in the **orbital period**, or the time to complete one orbit. In one period T, the orbiting object moves the orbital circumference $2\pi r$, so its speed is $v = 2\pi r/T$. Squaring Equation 8.3 then gives

$$\left(\frac{2\pi r}{T}\right)^2 = \frac{GM}{r}$$

or

$$T^2 = \frac{4\pi^2 r^3}{GM} \qquad \text{(orbital period, circular orbit)} \qquad (8.4)$$

In deriving Equation 8.4, we've proved Kepler's third law—that the square of the orbital period is proportional to the cube of the semimajor axis—for the special case of a circular orbit whose semimajor axis is its radius.

Note that orbital speed and period are independent of the orbiting object's mass m—another indication that all objects experience the same gravitational acceleration. Astronauts, for example, have the same orbital parameters as the space station. That's why astronauts seem weightless inside the station and why they don't float away if they step outside.

EXAMPLE 8.2 Orbital Speed and Period: The Space Station

The International Space Station is in a circular orbit at altitude 380 km. What are its orbital speed and period?

INTERPRET This problem involves the speed and period of a circular orbit about Earth.

DEVELOP We can compute the orbit's radius and then use Equation 8.3, $v = \sqrt{GM/r}$, to find the speed and Equation 8.4, $T^2 = 4\pi^2 r^3/GM$, to find the period because the orbit is circular.

EVALUATE As always, the distance is measured from the center of the gravitating body, so r in these equations is Earth's 6.37-Mm radius plus the station's 380-km altitude. So we have

$$v = \sqrt{\frac{GM_E}{r}} = \sqrt{\frac{(6.67 \times 10^{-11}\ \text{N}\cdot\text{m}^2/\text{kg}^2)(5.97 \times 10^{24}\ \text{kg})}{6.37 \times 10^6\ \text{m} + 380 \times 10^3\ \text{m}}}$$
$$= 7.7\ \text{km/s}$$

or about 17,000 mi/h. We can get the orbital period from the speed and radius, or directly from Equation 8.4, $T = \sqrt{4\pi^2 r^3/GM_E}$. Using the numbers in the calculation for v gives $T = 5.5 \times 10^3$ s, or about 90 min.

ASSESS Make sense? Both answers have the correct units, and 90 min seems reasonable for the period of an orbit at a small fraction of the Moon's distance from Earth. Astronauts who want a circular orbit 380 km up have no choice but this speed and period. In fact, for any "near-Earth" orbit, with altitude much less than Earth's radius, the orbital period is about 90 min. If there were no air resistance and if you could throw a baseball fast enough, it too would go into orbit, skimming Earth's surface with a roughly 90-min period. ■

Example 8.2 shows that the near-Earth orbital period is about 90 min. The Moon, on the other hand, takes 27 days to complete its nearly circular orbit. So there must be a distance where the orbital period is 24 h—the same as Earth's rotation. A satellite at this distance will remain fixed with respect to Earth's surface provided its orbit is parallel to the equator. TV, weather, and communication satellites are often placed in this **geosynchronous orbit**.

EXAMPLE 8.3 Geosynchronous Orbit: Finding the Altitude

What altitude is required for geosynchronous orbit?

INTERPRET Here we're given an orbital period—24 h or 86,400 s—and asked to find the corresponding altitude for a circular orbit.

DEVELOP Equation 8.4, $T^2 = 4\pi^2 r^3/GM$, relates the period T and distance r from Earth's center. Our plan is to solve for r and then subtract Earth's radius to find the altitude (distance from the surface).

EVALUATE Solving for r, we get

$$r = \left(\frac{GM_E T^2}{4\pi^2}\right)^{1/3}$$
$$= \left[\frac{(6.67 \times 10^{-11}\ \text{N}\cdot\text{m}^2/\text{kg}^2)(5.97 \times 10^{24}\ \text{kg})(8.64 \times 10^4\ \text{s})^2}{4\pi^2}\right]^{1/3}$$
$$= 4.22 \times 10^7\ \text{m}$$

or 42,200 km from Earth's center. Subtracting Earth's radius then gives an altitude of about 36,000 km, or 22,000 miles.

ASSESS Make sense? This is a lot higher than the 90-min low-Earth orbit, but a lot lower than the Moon's 385,000 km distance. Our answer defines one of the most valuable pieces of "real estate" in space—a place where satellites appear suspended over a fixed spot on Earth. The dish antenna in this chapter's opening photo points to such a satellite, positioned 22,000 mi over the equator. A more careful calculation would use Earth's so-called sidereal rotation period, measured with respect to the distant stars rather than the Sun. Because Earth isn't a perfect sphere, geosynchronous satellites drift slightly and therefore fire small rockets every few weeks to stay in position. ■

Elliptical Orbits

Using his laws of motion and gravity, Newton was able to prove Kepler's assertion that the planets move in elliptical paths with the Sun at one focus. Circular orbits represent the special case where the two foci of the ellipse coincide, so the distance from the gravitating center remains constant. Most planetary orbits are nearly, but not quite, circular; Earth's distance from the Sun, for example, varies by about 3% throughout the year. But the orbits of comets and other smaller bodies are often highly elliptical (Fig. 8.6). Their orbital speeds vary, as they gain speed "falling" toward the Sun, whip quickly around the Sun at the point of closest approach (**perihelion**), and then "climb" ever more slowly to their most distant point (**aphelion**) before returning.

In Chapter 3, we showed that the trajectory of a projectile is a parabola. But our derivation neglected Earth's curvature and the associated variation in g with altitude. In fact, a projectile is just like any orbiting body. If we neglect air resistance, it too describes an elliptical orbit with Earth's center at one focus. Only for trajectories small compared with Earth's radius are the true elliptical path and the parabola of Chapter 3 essentially indistinguishable (Fig. 8.7).

Are missiles and baseballs really in orbit? Yes. But their orbits happen to intersect the Earth. At that point, nongravitational forces put an end to orbital motion. If Earth suddenly shrank to the size of a grapefruit (but kept the same mass), a baseball would continue happily in orbit, as Fig. 8.7 suggests. Newton's ingenious intuition was correct: Barring air resistance, there's truly no difference between the motion of everyday objects near Earth and the motion of celestial objects.

FIGURE 8.6 Orbits of most known comets, like the one shown here, are highly elliptical.

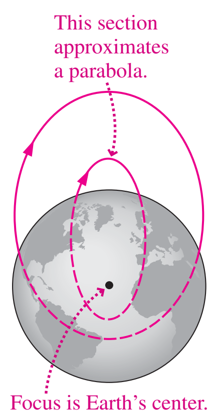


FIGURE 8.7 Projectile trajectories are actually elliptical.

Open Orbits

With elliptical and circular orbits, the motion repeats indefinitely because the orbit is a closed path. But closed orbits aren't the only possibility. Imagine again Newton's thought experiment—only now fire the projectile faster than necessary for a circular orbit (Fig. 8.8). The projectile rises higher than before, describing an ellipse with its low point at the launch site. Faster, and the ellipse gets more elongated. But with great enough initial speed, the projectile describes a trajectory that takes it ever farther from Earth. We'll see in the next section how energy determines the type of orbit.

GOT IT? 8.2 Suppose the paths in Fig. 8.8 are the paths of four projectiles. Rank each path (circular, elliptical, parabolic, and hyperbolic) according to the initial speed of the corresponding projectile. Assume all are launched from their common point at the top of the figure.

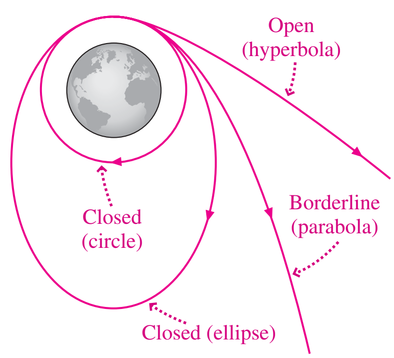


FIGURE 8.8 Closed and open orbits.

8.4 Gravitational Energy

How much energy does it take to boost a satellite to geosynchronous altitude? Our simple answer mgh won't do here, since g varies substantially over the distance involved. So, as we found in Chapter 7, we have to integrate to determine the potential energy.

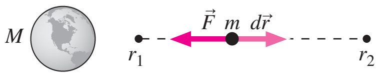


FIGURE 8.9 Finding the potential-energy change requires integration.

Figure 8.9 shows two points at distances r_1 and r_2 from the center of a gravitating mass M, in this case Earth. Equation 7.2 gives the change in potential energy associated with moving a mass m from r_1 to r_2:

$$\Delta U_{12} = -\int_{r_1}^{r_2} \vec{F} \cdot d\vec{r}$$

Here the force points radially inward and has magnitude GMm/r^2, while the path element $d\vec{r}$ points radially outward. Then $\vec{F} \cdot d\vec{r} = -(GMm/r^2)\,dr$, where the minus sign comes from the factor $\cos 180°$ in the dot product of oppositely directed vectors. Then the potential energy difference is

$$\Delta U_{12} = \int_{r_1}^{r_2} \frac{GMm}{r^2}\,dr = GMm \int_{r_1}^{r_2} r^{-2}\,dr = GMm \left.\frac{r^{-1}}{-1}\right|_{r_1}^{r_2} = GMm\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \quad (8.5)$$

Does this make sense? Yes: For $r_1 < r_2$, ΔU_{12} is positive, showing that potential energy increases with height—consistent with our simpler result $\Delta U = mgh$ near Earth's surface. Although we derived Equation 8.5 for two points on a radial line, Fig. 8.10 shows that it holds for any two points at distances r_1 and r_2 from the gravitating center.

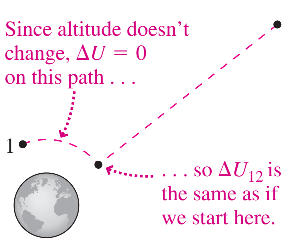


FIGURE 8.10 Gravity is conservative, so we can use any path to evaluate the potential-energy change. Only the radial part of the path contributes to ΔU.

EXAMPLE 8.4 Gravitational Potential Energy: Steps to the Moon

Materials to construct an 11,000-kg lunar observatory are boosted from Earth to geosynchronous orbit. There they are assembled and then launched to the Moon, 385,000 km from Earth. Compare the work that must be done against Earth's gravity on the two legs of the trip.

INTERPRET This problem asks about work done against gravity, a conservative force.

DEVELOP As we saw in Chapter 7, the work done against a conservative force is equal to the change in potential energy; here that change is given by Equation 8.5. For the first leg, we have $r_1 = R_E$ and then, from Example 8.3, $r_2 = 42{,}200$ km.

EVALUATE Since the quantity GM_Em that appears in Equation 8.5 will be used in both steps, we calculate it first: $GM_Em = 4.38 \times 10^{18}\ \text{N}\cdot\text{m}^2$. Then for the first step we have

$$\begin{aligned} W &= \Delta U_{12} = GM_Em\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \\ &= (4.38 \times 10^{18}\ \text{N}\cdot\text{m}^2)\left(\frac{1}{6.37 \times 10^6\ \text{m}} - \frac{1}{4.22 \times 10^7\ \text{m}}\right) \\ &= 5.8 \times 10^{11}\ \text{J} \end{aligned}$$

From geosynchronous orbit to the Moon, a similar calculation gives

$$\begin{aligned} W &= (4.38 \times 10^{18}\ \text{N}\cdot\text{m}^2)\left(\frac{1}{4.22 \times 10^7\ \text{m}} - \frac{1}{3.85 \times 10^8\ \text{m}}\right) \\ &= 9.2 \times 10^{10}\ \text{J} \end{aligned}$$

ASSESS Make sense? Even though the second leg is much longer, the rapid drop-off in the gravitational force means that less work is required than for the shorter boost to geosynchronous altitude. Our calculations here include only the work done against Earth's gravity; additional energy would be required to attain a circular geosynchronous orbit. On the other hand, the Moon's gravitational attraction would lower the required energy somewhat.

The Zero of Potential Energy

Equation 8.5 has an interesting feature: The potential-energy difference remains finite even when the points are infinitely far apart, as you can see by setting either r_1 or r_2 to infinity. Although the gravitational force always acts, it weakens so rapidly that its effect is finite over even infinite distances. This property makes it convenient to set the zero of potential energy at infinity. Setting $r_1 = \infty$ and dropping the subscript on r_2, we then have an expression for the potential energy at an arbitrary distance r from a gravitating center:

$$U(r) = -\frac{GMm}{r} \quad \text{(gravitational potential energy)} \tag{8.6}$$

The potential energy is negative because we chose $U = 0$ at $r = \infty$. Any other point is closer to the gravitating center and therefore has lower potential energy.

Knowing the gravitational potential energy allows us to apply the powerful conservation-of-energy principle. Figure 8.11 shows the potential-energy curve given by Equation 8.6. Superposing three values of total energy E shows that orbits with $E < 0$ have a turning point where they intersect the potential-energy curve, and are therefore closed. Orbits with $E > 0$, in contrast, are open because they never intersect the curve and therefore extend to infinity.

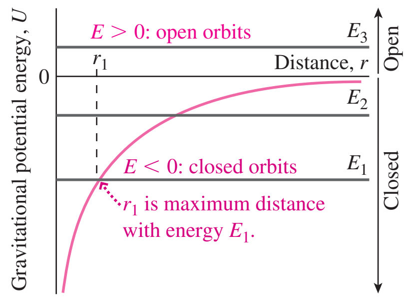


FIGURE 8.11 A gravitational potential-energy curve.

EXAMPLE 8.5 Conservation of Energy: Blast Off!

A rocket is launched vertically upward at 3.1 km/s. How high does it go?

INTERPRET This sounds like a problem from Chapter 2, but here gravity changes, so the acceleration isn't constant. The conservation-of-energy principle lets us cut through those details, so we can apply the methods of Chapter 7. "How high does it go?" in the problem statement means we're dealing with the initial launch state and a final state where the rocket is momentarily at rest at the top of its trajectory.

DEVELOP Equation 7.7 describes conservation of energy: $K + U = K_0 + U_0$. Here we're given speed v at the bottom, so $K_0 = \frac{1}{2}mv^2$. We're going to be using Equation 8.6, $U(r) = -GMm/r$, for potential energy, and that's already established the zero of potential energy at infinity. So U_0 isn't zero but is given by Equation 8.6 with r equal to Earth's radius. Finally, at the top, $K = 0$ and U is also given by Equation 8.6, but now we don't know r. Our plan is to solve for that r and from it get the rocket's altitude. Figure 8.12 shows "before" and "after" diagrams with bar graphs, like those we introduced in Chapter 7.

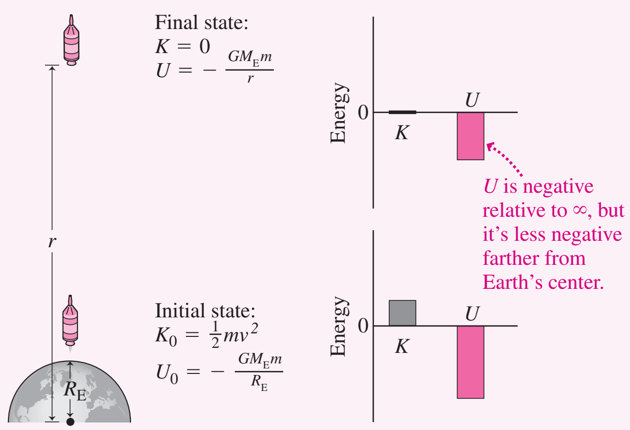


FIGURE 8.12 Diagrams for Example 8.5.

EVALUATE With our values of kinetic and potential energy, $K + U = K_0 + U_0$ becomes

$$-\frac{GM_Em}{r} = \frac{1}{2}mv_0^2 - \frac{GM_Em}{R_E}$$

where m is the rocket's mass, r is the distance from Earth's center at the peak, and Earth's radius R_E is the distance at launch. Solving for r gives

$$\begin{aligned} r &= \left(\frac{1}{R_E} - \frac{v_0^2}{2GM_E}\right)^{-1} \\ &= \left(\frac{1}{6.37\times10^6\ \text{m}} - \frac{(3100\ \text{m/s})^2}{2(6.67\times10^{-11}\ \text{N}\cdot\text{m}^2/\text{kg}^2)(5.97\times10^{24}\ \text{kg})}\right)^{-1} \\ &= 6.90\ \text{Mm} \end{aligned}$$

Again, this is the distance from Earth's center; subtracting Earth's radius then gives a peak altitude of 530 km.

ASSESS Make sense? Yes. Our answer of 530 km is significantly greater than the 490 km you'd get assuming a potential-energy change of $\Delta U = mgh$. That's because the decreasing gravitational force lets the rocket go higher before all its kinetic energy becomes potential.

✓TIP All Conservation-of-Energy Problems Are the Same

This problem is essentially the same as throwing a ball straight up and solving for its maximum height using $U = mgh$ for the potential energy. The only difference is the more complicated potential-energy function $U = -GMm/r$, used here because the variation in gravity is significant over the rocket's trajectory. Recognize what's common to all similar problems, and you'll begin to see how physics really is based on just a few simple principles.

■

Escape Speed

What goes up comes down, right? Not always! Figure 8.11 shows that an object with total energy zero or greater can escape infinitely far from a gravitating body, never to return. Consider an object of mass m at the surface of a gravitating body of mass M and radius r. It has gravitational potential energy given by Equation 8.6, $U = -GMm/r$. Toss it upward with speed v, and it's also got kinetic energy $\frac{1}{2}mv^2$. Its total energy will be zero if

$$0 = K + U = \tfrac{1}{2}mv^2 - \frac{GMm}{r}$$

The speed v here that makes the total energy zero is called the **escape speed** because an object with this speed or greater has enough energy to escape forever from the gravitating body. Solving for v in the preceding equation gives the escape speed:

$$v_{\text{esc}} = \sqrt{\frac{2GM}{r}} \qquad \text{(escape speed)} \tag{8.7}$$

At Earth's surface, $v_{\text{esc}} = 11.2$ km/s. Earth-orbiting spacecraft have lower speeds. Moon-bound astronauts go at just under v_{esc}, so if anything goes wrong (as with Apollo 13), they can return to Earth. Planetary spacecraft have speeds greater than v_{esc}. The Pioneer and Voyager missions to the outer planets gained enough additional energy in their encounters with Jupiter that they now have escape speed relative to the Sun and will coast indefinitely through interstellar space.

Energy in Circular Orbits

In the special case of a circular orbit, kinetic and potential energies are related in a simple way. In Section 8.3, we found that the speed in a circular orbit is given by

$$v^2 = \frac{GM}{r}$$

where r is the distance from a gravitating center of mass M. So the kinetic energy of the orbiting object is

$$K = \tfrac{1}{2}mv^2 = \frac{GMm}{2r}$$

while the potential energy is given by Equation 8.6:

$$U = -\frac{GMm}{r}$$

Comparing these two expressions shows that $U = -2K$ for a circular orbit. The total energy is therefore

$$E = U + K = -2K + K = -K \tag{8.8a}$$

or, equivalently,

$$E = \tfrac{1}{2}U = -\frac{GMm}{2r} \tag{8.8b}$$

The total energy in these equations is negative, showing that circular orbits are—obviously—bound orbits. We stress that these results apply only to *circular* orbits; in elliptical orbits, there's a continuous interchange between kinetic and potential energy as the orbiting object moves relative to the gravitating center.

Equation 8.8a shows that *higher* kinetic energy corresponds to *lower* total energy. This surprising result occurs because *higher* orbital speed corresponds to a *lower* orbit, with lower potential energy.

CONCEPTUAL EXAMPLE 8.1 **Space Maneuvers**

Astronauts heading for the International Space Station find themselves in the right circular orbit, but well behind the station. How should they maneuver to catch up?

EVALUATE To catch up, the astronauts will have to go faster than the space station. That means increasing their kinetic energy—and, as we've just seen, that corresponds to *lowering* their total energy. So they'll need to drop into a lower orbit.

Figure 8.13 shows the catch-up sequence. The astronauts fire their rocket backward, decreasing their energy and dropping briefly into a lower-energy elliptical orbit. They then fire their rocket to circularize the orbit. Now they're in a lower-energy but faster orbit than the space station. When they're correctly positioned, they fire their rocket to boost themselves into a higher-energy elliptical orbit, then fire again to circularize that orbit in the vicinity of the station.

ASSESS Our solution sounds counterintuitive—as if a car, to speed up, had to apply its brakes. But that's what's needed here, thanks to the interplay between kinetic and potential energy in circular orbits.

MAKING THE CONNECTION Suppose the astronauts reach the space station's 380-km altitude, but find themselves one-fourth of an orbit behind the station. If the maneuver described above drops their spacecraft into a 320-km circular orbit, how many orbits must they make before catching up with the station? Neglect the time involved in transferring between circular orbits.

EVALUATE Applying Equation 8.4 gives periods $T_1 = 92.0$ min for the space station and $T_2 = 90.8$ min for the astronauts in their lower orbit. So with each orbit the astronauts gain 1.2 min on the station. They've got to make up one-fourth of an orbit, or 23 min. That will take $(23 \text{ min})/(1.2 \text{ min/orbit}) = 19$ orbits, or just over a day.

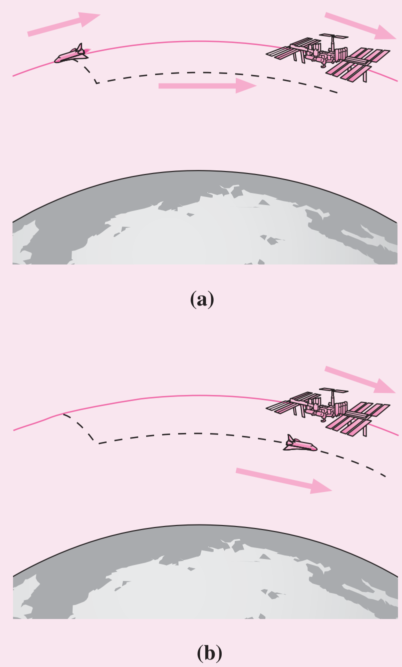


FIGURE 8.13 Playing catch-up with the space station.

GOT IT? 8.3 Two identical spacecraft A and B are in circular orbits about Earth, with B at a higher altitude. Which of the following statements are true? (a) B has greater total energy; (b) B is moving faster; (c) B takes longer to complete its orbit; (d) B has greater potential energy; (e) a larger proportion of B's total energy is potential energy.

8.5 The Gravitational Field

Our description of gravity so far suggests that a massive body like Earth somehow "reaches out" across empty space to pull on objects like falling apples, satellites, or the Moon. This view—called **action-at-a-distance**—has bothered both physicists and philosophers for centuries. How can the Moon, for example, "know" about the presence of the distant Earth?

An alternative view holds that Earth creates a **gravitational field** and that objects respond to the field in their immediate vicinity. The field is described by vectors that give the force per unit mass that would arise at each point if a mass were placed there. Near Earth's surface, for instance, the gravitational field vectors point vertically downward and have magnitude 9.8 N/kg. We can express this field vectorially by writing

$$\vec{g} = -g\hat{j} \qquad \text{(gravitational field near Earth's surface)} \qquad (8.9)$$

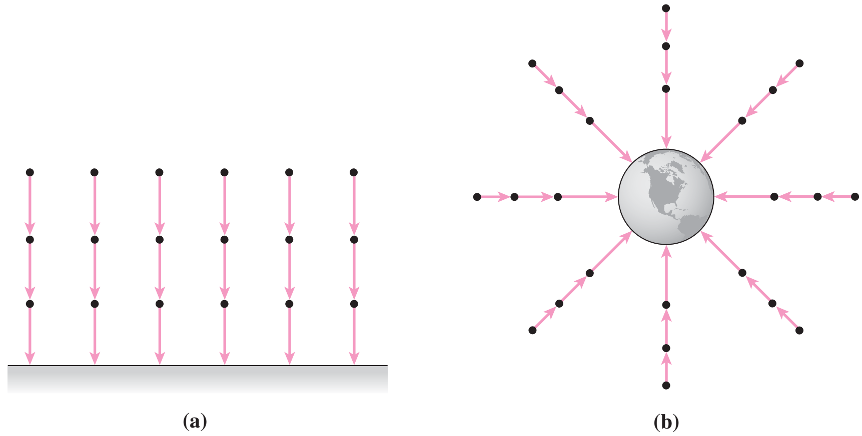


FIGURE 8.14 Gravitational field vectors at points (a) near Earth's surface and (b) on a larger scale.

More generally, the field points toward a spherical gravitating center, and its strength decreases inversely with the square of the distance:

$$\vec{g} = -\frac{GM}{r^2}\hat{r} \qquad (\text{gravitational field of a spherical mass } M) \tag{8.10}$$

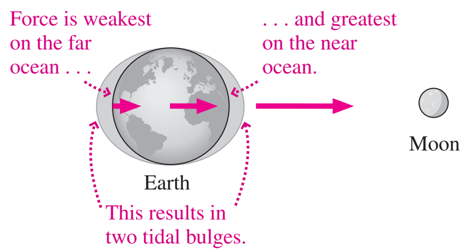
where $\hat{r}$ is a unit vector that points radially outward. Figure 8.14 shows pictorial representations of Equations 8.9 and 8.10. You can show that the units of gravitational field (N/kg) are equivalent to those of acceleration $(\mathrm{m/s^2})$, so the field is really just a vectorial representation of g, the local acceleration of gravity.

What do we gain by this field description? As long as we deal with situations where nothing changes, the action-at-a-distance and field descriptions are equivalent. But what if, for example, Earth suddenly gains mass? How does the Moon know to adjust its orbit? Under the field view, its orbit doesn't change immediately; instead, it takes a small but nonzero time for the information about the more massive Earth to propagate out to the Moon. The Moon always responds to the gravitational field *in its immediate vicinity*, and it takes a short time for the field itself to change. That description is consistent with Einstein's notion that instantaneous transmission of information is impossible; the action-at-a-distance view is not.

More generally, the field view provides a powerful way of describing interactions in physics. We'll see fields again when we study electricity and magnetism, and you'll find that fields aren't just mathematical or philosophical conveniences but are every bit as real as matter itself.

APPLICATION Tides

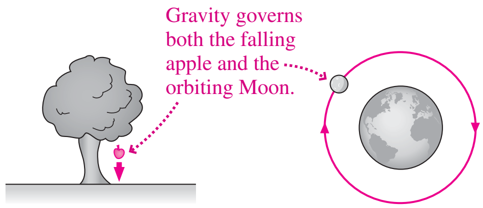
If the gravitational field were uniform, all parts of a freely falling object would experience exactly the same acceleration. But gravity does vary, and the result is a force—not from gravity itself but from *changes* in gravity with position—that tends to stretch or compress an object. Ocean tides result from this **tidal force**, as the nonuniform gravitational forces of Sun and Moon stretch the oceans and create bulges that move across Earth as the planet rotates. The figure shows that the greatest force is on the ocean nearest the Moon, causing one tidal bulge. The solid Earth experiences an intermediate force, pulling it away from the ocean on the far side. The water that's "left behind" forms a second bulge opposite the Moon. The bulges shown are highly exaggerated. Furthermore, shoreline effects and the differing relative positions of the Moon and Sun complicate this simple picture that suggests two equal high tides and two equal low tides a day. Tidal forces also cause internal heating of satellites like Jupiter's moon Io and contribute to the formation of planetary rings.

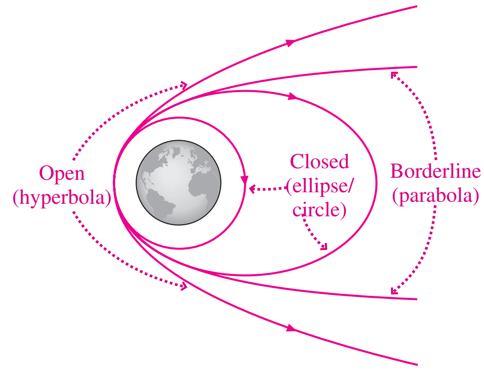


CHAPTER 8 SUMMARY

Big Picture

The big idea here is **universal gravitation**—an attractive force that acts between all matter with a strength that depends directly on the product of two interacting masses and inversely on the square of the distance between them. Gravitation is responsible for the familiar behavior of falling objects and also for the orbits of planets and satellites. Depending on energy, orbits may be closed (elliptical/circular) or open (hyperbolic/parabolic).

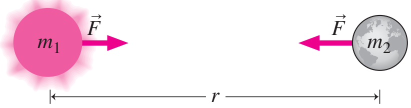




Key Concepts and Equations

Mathematically, Newton's law of universal gravitation describes the attractive force F between two masses m_1 and m_2 located a distance r apart:

$$F = \frac{Gm_1m_2}{r^2} \qquad \text{(universal gravitation)}$$



This equation applies to point masses of negligible size and to spherically symmetric masses of any size. It's an excellent approximation for any objects whose size is much smaller than their separation. In all cases, r is measured from the centers of the gravitating objects.

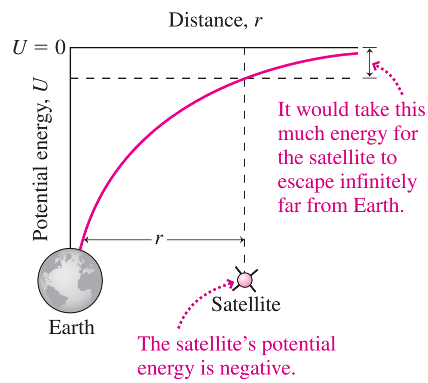
Because the strength of gravity varies with distance, potential-energy changes over large distances aren't just a product of force and distance. Integration shows that the potential energy change ΔU involved in moving a mass m originally a distance r_1 from the center of a mass M to a distance r_2 is

$$\Delta U = GMm\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \qquad \text{(change in potential energy)}$$

With gravity, it's convenient to choose the zero of potential energy at infinity; then

$$U = -\frac{GMm}{r} \qquad \text{(potential energy, } U = 0 \text{ at infinity)}$$

for the potential energy of a mass m located a distance r from the center of a mass M.

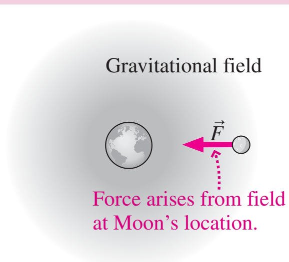


Applications

A total energy—kinetic plus potential—of zero marks the dividing line between closed and open orbits. An object located a distance r from a gravitating mass M must have at least the **escape speed** to achieve an open orbit and escape M's vicinity forever:

$$v_{\text{esc}} = \sqrt{\frac{2GM}{r}}$$

The **gravitational field** concept provides a way to describe gravity that avoids the troublesome action-at-a-distance. A gravitating mass creates a field in the space around it, and a second mass responds to the field in its immediate vicinity.



Circular orbits are readily analyzed using Newton's laws and concepts from circular motion. A circular orbit of radius r about a mass M has a period given by

$$T^2 = \frac{4\pi^2 r^3}{GM}$$

Its kinetic and potential energies are related by $U = -2K$. Total energy is negative, as appropriate for a closed orbit, and the object actually moves faster the lower its total energy.

A special orbit is the **geosynchronous orbit**, parallel to Earth's equator at an altitude of about 36,000 km. Here the orbital period is 24 h, so a satellite in geosynchronous orbit appears from Earth's surface to be fixed in the sky. TV, communications, and weather satellites use geosynchronous orbits.

For Thought and Discussion

1. What do Newton's apple and the Moon have in common?
2. Explain the difference between G and g.
3. When you stand on Earth, the distance between you and Earth is zero. So why isn't the gravitational force infinite?
4. The force of gravity on an object is proportional to the object's mass, yet all objects fall with the same gravitational acceleration. Why?
5. A friend who knows nothing about physics asks what keeps an orbiting satellite from falling to Earth. Give an answer that will satisfy your friend.
6. Could you put a satellite in an orbit that keeps it stationary over the south pole? Explain.
7. Why are satellites generally launched eastward and from low latitudes? (*Hint:* Think about Earth's rotation.)
8. Given Earth's mass, the Moon's distance and orbital period, and the value of G, could you calculate the Moon's mass? If yes, how? If no, why not?
9. How should a satellite be launched so that its orbit takes it over every point on the (rotating) Earth?
10. Does the gravitational force of the Sun do work on a planet in a circular orbit? In an elliptical orbit? Explain.

Exercises and Problems

Exercises

Section 8.2 Universal Gravitation

11. Space explorers land on a planet that has the same mass as Earth, but find they weigh twice as much as they would on Earth. What's the planet's radius?
12. Use data for the Moon's orbit from Appendix E to compute the Moon's acceleration in its circular orbit, and verify that the result is consistent with Newton's law of gravitation.
13. To what fraction of its current radius would Earth have to shrink (with no change in mass) for the gravitational acceleration at its surface to triple?
14. Calculate the gravitational acceleration at the surface of (a) Mercury and (b) Saturn's moon Titan.
15. Two identical lead spheres with their centers 14 cm apart attract each other with a 0.25-μN force. Find their mass.
16. What's the approximate value of the gravitational force between a 67-kg astronaut and a 73,000-kg spacecraft when they're 84 m apart?
17. A sensitive gravimeter is carried to the top of Chicago's Willis (formerly Sears) Tower, where its reading for the acceleration of gravity is 1.36 mm/s^2 lower than at street level. Find the building's height.

Section 8.3 Orbital Motion

18. At what altitude will a satellite complete a circular orbit of Earth in 2.0 h?
19. Find the speed of a satellite in geosynchronous orbit.
20. Mars's orbit has a diameter 1.52 times that of Earth's orbit. How long does it take Mars to orbit the Sun?
21. Calculate the orbital period for Jupiter's moon Io, which orbits 4.22×10^5 km from the planet's center.
22. An astronaut hits a golf ball horizontally from the top of a lunar mountain so fast that it goes into circular orbit. What's its orbital period?
23. The Mars Reconnaissance Orbiter circles the red planet with a 112-min period. What's the spacecraft's altitude?

Section 8.4 Gravitational Energy

24. Earth's distance from the Sun varies from 147 Gm at perihelion to 152 Gm at aphelion because its orbit isn't quite circular. Find the change in potential energy as Earth goes from perihelion to aphelion.
25. So-called suborbital missions take scientific instruments into space for brief periods without the expense of getting into orbit; their trajectories are often simple "up and down" vertical paths. How much energy does it take to launch a 230-kg instrument on a vertical trajectory that peaks at 1800 km altitude?
26. A rocket is launched vertically upward from Earth's surface at 5.1 km/s. What's its maximum altitude?
27. What vertical launch speed is necessary to get a rocket to an altitude of 1100 km?
28. Find the energy necessary to put 1 kg, initially at rest on Earth's surface, into geosynchronous orbit.
29. What's the total mechanical energy associated with Earth's orbital motion?
30. The escape speed from a planet of mass 2.9×10^{24} kg is 7.1 km/s. Find the planet's radius.
31. Determine escape speeds from (a) Jupiter's moon Callisto and (b) a neutron star, with the Sun's mass crammed into a sphere of radius 6.0 km. See Appendix E for relevant data.
32. To what radius would Earth have to shrink, with no change in mass, for escape speed at its surface to be 30 km/s?

Problems

33. The gravitational acceleration at a planet's surface is 22.5 m/s^2. Find the acceleration at an altitude equal to half the planet's radius.
34. BIO One of the longest-standing athletic records is Cuban Javier Sotomayor's 2.45-m high jump. How high could Sotomayor jump on (a) Mars and (b) Earth's Moon?
35. You're the navigator on a spaceship studying an unexplored planet. Your ship has just gone into a circular orbit around the planet, and you determine that the gravitational acceleration at your orbital altitude is half what it would be at the surface. What do you report for your altitude, in terms of the planet's radius?
36. If you're standing on the ground 15 m directly below the center of a spherical water tank containing 4×10^6 kg of water, by what fraction is your weight reduced due to the water's gravitational attraction?
37. Given the Moon's orbital radius of 384,400 km and period of 27.3 days, calculate its acceleration in its circular orbit, and compare with the acceleration of gravity at Earth's surface. Show that the Moon's acceleration is lower by the ratio of the square of Earth's radius to the square of the Moon's orbital radius, thus confirming the inverse-square law for the gravitational force.
38. Equation 7.8 relates force to the derivative of potential energy. Use this fact to differentiate Equation 8.6 for gravitational potential energy, and show that you recover Newton's law of gravitation.

39. During the Apollo Moon landings, one astronaut remained with the command module in lunar orbit, about 130 km above the surface. For half of each orbit, this astronaut was completely cut off from the rest of humanity as the spacecraft rounded the far side of the Moon. How long did this period last?
40. A white dwarf is a collapsed star with roughly the Sun's mass compressed into the size of Earth. What would be (a) the orbital speed and (b) the orbital period for a spaceship in orbit just above the surface of a white dwarf?
41. Given that our Sun orbits the galaxy with a period of 200 My at 2.6×10^{20} m from the galactic center, estimate the galaxy's mass. Assume (incorrectly) that the galaxy is essentially spherical and that most of its mass lies interior to the Sun's orbit.
42. You're preparing an exhibit for the Golf Hall of Fame, and you realize that the longest golf shot in history was Astronaut Alan Shepard's lunar drive. Shepard, swinging single-handed with a golf club attached to a lunar sample scoop, claimed his ball went "miles and miles." The record for a single-handed golf shot on Earth is 257 m. Could Shepard's ball really have gone "miles and miles"? Assume the ball's initial speed is independent of gravitational acceleration.
43. Exact solutions for gravitational problems involving more than two bodies are notoriously difficult. One solvable problem involves a configuration of three equal-mass objects spaced in an equilateral triangle. Forces due to their mutual gravitation cause the configuration to rotate. Suppose three identical stars, each of mass M, form a triangle of side L. Find an expression for the period of their orbital motion.
44. Satellites A and B are in circular orbits, with A twice as far from Earth's center as B. How do their orbital periods compare?
45. The asteroid Pasachoff orbits the Sun with period 1417 days. Find the semimajor axis of its orbit from Kepler's third law. Use Earth's orbital radius and period, respectively, as your units of distance and time.
46. We still don't have a permanent solution for the disposal of radioactive waste. As a nuclear waste specialist with the Department of Energy, you're asked to evaluate a proposal to shoot waste canisters into the Sun. You need to report the speed at which a canister, dropped from rest in the vicinity of Earth's orbit, would hit the Sun. What's your answer?
47. At perihelion in February 1986, Comet Halley was 8.79×10^7 km from the Sun and was moving at 54.6 km/s. What was Halley's speed when it crossed Neptune's orbit in 2006?
48. Neglecting air resistance, to what height would you have to fire a rocket for the constant-acceleration equations of Chapter 2 to give a height in error by 1%? Would those equations overestimate or underestimate the height?
49. Show that an object released from rest very far from Earth reaches Earth's surface at essentially escape speed.
50. By what factor must an object's speed in circular orbit be increased to reach escape speed from its orbital altitude?
51. You're in charge of tracking celestial objects that might pose a danger to Earth. Astronomers have discovered a new comet that's moving at 45 km/s as it crosses Earth's orbit. Determine whether the comet will again return to Earth's vicinity.
52. Two meteoroids are 250,000 km from Earth's center and moving at 2.1 km/s. One is headed straight for Earth, while the other is on a path that will come within 8500 km of Earth's center (Fig. 8.15). Find the speed (a) of the first meteoroid when it strikes Earth and (b) of the second meteoroid at its closest approach. (c) Will the second meteoroid ever return to Earth's vicinity?

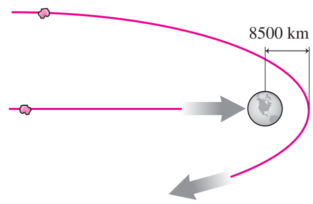


FIGURE 8.15 Problem 52

53. Neglecting Earth's rotation, show that the energy needed to launch a satellite of mass m into circular orbit at altitude h is $\left(\frac{GM_Em}{R_E}\right)\left(\frac{R_E + 2h}{2(R_E + h)}\right)$.
54. A projectile is launched vertically upward from a planet of mass M and radius R; its initial speed is twice the escape speed. Derive an expression for its speed as a function of the distance r from the planet's center.
55. A spacecraft is in circular orbit 5500 km above Earth's surface. How much will its altitude decrease if it moves to a new circular orbit where (a) its orbital speed is 10% higher or (b) its orbital period is 10% shorter?
56. Two meteoroids are 160,000 km from Earth's center and heading straight toward Earth, one at 10 km/s, the other at 20 km/s. At what speeds will they strike Earth?
57. Two rockets are launched from Earth's surface, one at 12 km/s and the other at 18 km/s. How fast is each moving when it crosses the Moon's orbit?
58. A satellite is in an elliptical orbit at altitudes ranging from 230 to 890 km. At its highest point, it's moving at 7.23 km/s. How fast is it moving at its lowest point?
59. A missile's trajectory takes it to a maximum altitude of 1200 km. If its launch speed is 6.1 km/s, how fast is it moving at the peak of its trajectory?
60. A 720-kg spacecraft has total energy −0.53 TJ and is in circular orbit around the Sun. Find (a) its orbital radius, (b) its kinetic energy, and (c) its speed.
61. Mercury's orbital speed varies from 38.8 km/s at aphelion to 59.0 km/s at perihelion. If the planet is 6.99×10^{10} m from the Sun's center at aphelion, how far is it at perihelion?
62. Show that the form $\Delta U = mg\,\Delta r$ follows from Equation 8.5 when $r_1 \simeq r_2$. [*Hint:* Write $r_2 = r_1 + \Delta r$ and apply the binomial approximation (Appendix A).]
63. Two satellites are in geosynchronous orbit but in diametrically opposite positions (Fig. 8.16). In order to catch up with the other, one satellite descends into a lower circular orbit (see

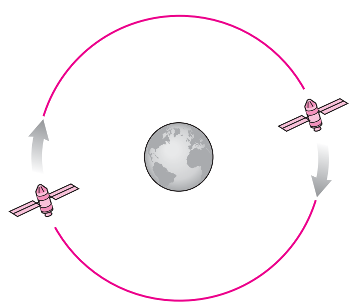


FIGURE 8.16 Problem 63

Conceptual Example 8.1 for a description of this maneuver). How far should it descend if it's to catch up in 10 orbits? Neglect rocket firing times and time spent moving between the two circular orbits.

64. We derived Equation 8.4 on the assumption that the massive gravitating center remains fixed. Now consider two objects with equal mass M orbiting each other, as shown in Fig. 8.17. Show that the orbital period is given by $T^2 = 2\pi^2 d^3/GM$, where d is the distance between the objects.

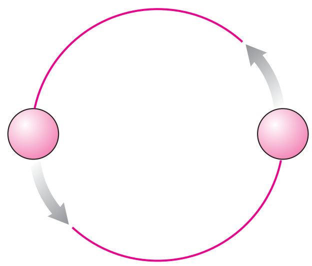


FIGURE 8.17 Problem 64

65. Tidal effects in the Earth-Moon system cause the Moon's orbital period to increase at a current rate of about 35 ms per century. Assuming the Moon's orbit is circular, to what rate of change in the Earth-Moon distance does this correspond? (*Hint:* Differentiate Kepler's third law, Equation 8.4, and consult Appendix E.)
66. As a member of the 2040 Olympic committee, you're considering a new sport: asteroid jumping. On Earth, world-class high jumpers routinely clear 2 m. Your job is to make sure athletes jumping from asteroids will return to the asteroid. Make the simplifying assumption that asteroids are spherical, with average density 2500 kg/m^3. For safety, make sure even a jumper capable of 3 m on Earth will return to the surface. What do you report for the minimum asteroid diameter? **BIO**
67. The Olympic Committee is keeping you busy! You're now asked to consider a proposal for lunar hockey. The record speed for a hockey puck is 168 km/h. Is there any danger that hockey pucks will go into lunar orbit?
68. Tidal forces are proportional to the variation in gravity with position. By differentiating Equation 8.1, estimate the ratio of the tidal forces due to the Sun and the Moon. Compare your answer with the ratio of the gravitational forces that the Sun and Moon exert on Earth. Use data from Appendix E.
69. Spacecraft that study the Sun are often placed at the so-called L1 Lagrange point, located sunward of Earth on the Sun-Earth line. L1 is the point where Earth's and Sun's gravity together produce an orbital period of one year, so that a spacecraft at L1 stays fixed relative to Earth as both planet and spacecraft orbit the Sun. This placement ensures an uninterrupted view of the Sun, without being periodically eclipsed by Earth as would occur in Earth orbit. Find L1's location relative to Earth. (*Hint*: This problem calls for numerical methods or solving a higher-order polynomial equation.) **C**

Passage Problems

The Global Positioning System (GPS) uses a "constellation" of some 30 satellites to provide accurate positioning for any point on Earth (Fig. 8.18). GPS receivers time radio signals traveling at the speed of light from three of the satellites to find the receiver's position. Signals from one or more additional satellites provide corrections, eliminating the need for high-accuracy clocks in individual GPS receivers. GPS satellites are in circular orbits at 20,200 km altitude.

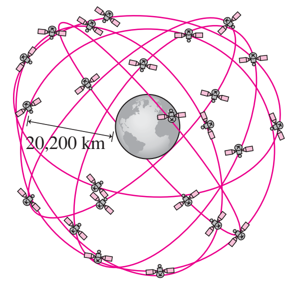


FIGURE 8.18 GPS satellites (Passage Problems 70–73)

70. What's the approximate orbital period of GPS satellites?
    a. 90 min
    b. 8 h
    c. 12 h
    d. 24 h
    e. 1 week
71. What's the approximate speed of GPS satellites?
    a. 9.8 m/s
    b. 500 m/s
    c. 1.7 km/s
    d. 4 km/s
    e. 12 km/s
72. What's the approximate escape speed at GPS orbital distance?
    a. 4 km/s
    b. 5.5 km/s
    c. 6.3 km/s
    d. 9.8 km/s
    e. 11 km/s
73. The current generation of GPS satellites has masses of 844 kg. What's the approximate total energy of such a satellite?
    a. 6 GJ
    b. 3 GJ
    c. -3 GJ
    d. -6 GJ
    e. -8 GJ

Answers to Chapter Questions

Answer to Chapter Opening Question

The satellite orbits Earth in 24 hours, so from Earth's surface it appears at a fixed position in the sky.

Answers to GOT IT? Questions

8.1. (d) Quadrupled. If the original distance were r, the original force would be proportional to $1/r^2$. At half that distance, the force is proportional to $1/(r/2)^2 = 4/r^2$.
8.2. Hyperbolic $>$ parabolic $>$ elliptical $>$ circular.
8.3. (a), (c), and (d). Since B has higher total energy, it must have lower kinetic energy and is therefore moving slower. B is farther from the gravitating body, so its potential energy is higher—still negative, but less so than A's. For circular orbits, the ratio of potential energy to total energy is always the same—namely, $U = 2E$.

9 Systems of Particles

As the skier flies through the air, most parts of his body follow complex trajectories. But one special point follows a parabola. What's that point, and why is it special?

New Concepts, New Skills

By the end of this chapter you should be able to

- Find the center of mass of systems of individual particles and of continuous distributions of matter (9.1).
- Explain the principle of momentum conservation, and apply it to systems of particles (9.2).
- Describe the difference between inelastic and elastic collisions, and apply the appropriate conservation laws to analyze each (9.4–9.6).

Connecting Your Knowledge

- The material in this chapter draws largely on the concept of momentum, introduced in Chapter 4 (4.2).
- Momentum is intimately connected with Newton's second and third laws, also introduced in Chapter 4 (4.2, 4.6).

So far we've treated objects as point particles, ignoring the fact that most objects are composed of smaller parts. Here we deal explicitly with systems of many particles. These include **rigid bodies**—objects such as baseballs, cars, and planets whose constituent particles are stuck together in fixed orientations—as well as systems like human bodies, exploding fireworks, or flowing rivers, whose parts move relative to one another. In subsequent chapters we'll look at specific instances of many-particle systems, including the rotational motion of rigid bodies (Chapter 10) and the behavior of fluids (Chapter 15).

9.1 Center of Mass

The motion of the skier in the photo above is complex, with each part of his body moving on a different path. But the superimposed curve shows one point following the parabola we expect of a projectile (Section 3.5). This point is the **center of mass**, an average position of all the mass making up the skier. Since the net force on the skier as a whole is gravity, the photo suggests that the center of mass obeys Newton's second law, $\vec{F}_{\text{net}} = M\vec{a}_{\text{cm}}$, where M is the skier's total mass and $\vec{a}_{\text{cm}}$ is the acceleration of the center of mass. (We'll use the subscript cm for quantities associated with the center of mass.) To find the center of mass, we therefore need to locate a point whose acceleration obeys $\vec{F}_{\text{net}} = M\vec{a}_{\text{cm}}$, with $\vec{F}_{\text{net}}$ the net force on the entire system.

Consider a system of many particles. To find the center of mass, we want an equation like Newton's second law that involves the total mass of the system and the net force on the entire system. If we apply Newton's second law to the ith particle in the system, we have

$$\vec{F}_i = m_i\vec{a}_i = m_i\frac{d^2\vec{r}_i}{dt^2} = \frac{d^2 m_i\vec{r}_i}{dt^2}$$

where $\vec{F}_i$ is the net force on the particle, m_i is its mass, and we've written the acceleration $\vec{a}_i$ as the second derivative of the position $\vec{r}_i$. The total force on the system is the sum of the forces acting on all N particles. We write this sum compactly using the summation symbol Σ:

$$\vec{F}_{\text{total}} = \sum_{i=1}^{N}\vec{F}_i = \sum_{i=1}^{N}\frac{d^2 m_i\vec{r}_i}{dt^2}$$

where the sum runs over all particles composing the system, from $i = 1$ to N. But the sum of derivatives is the derivative of the sum, so

$$\vec{F}_{\text{total}} = \frac{d^2\left(\sum m_i\vec{r}_i\right)}{dt^2}$$

We can now put this equation in the form of Newton's second law. Multiplying and dividing the right-hand side by the total mass $M = \sum m_i$, and distributing this constant M through the differentiation, we have

$$\vec{F}_{\text{total}} = M\frac{d^2}{dt^2}\left(\frac{\sum m_i\vec{r}_i}{M}\right) \tag{9.1}$$

Equation 9.1 has a form like Newton's law applied to the total mass if we define

$$\vec{r}_{\text{cm}} = \frac{\sum m_i\vec{r}_i}{M} \qquad (\text{center of mass}) \tag{9.2}$$

Then the derivative in Equation 9.1 becomes $d^2\vec{r}_{\text{cm}}/dt^2$, which we recognize as the center-of-mass acceleration, $\vec{a}_{\text{cm}}$. So now Equation 9.1 reads $\vec{F}_{\text{total}} = M\vec{a}_{\text{cm}}$. This is almost Newton's law—but not quite, because the force here is the sum of all the forces acting on all the particles of the system, and we want just the net **external force**—the net force applied from *outside* the system. We can write the force $\vec{F}_{\text{total}}$ as

$$\vec{F}_{\text{total}} = \sum\vec{F}_{\text{ext}} + \sum\vec{F}_{\text{int}}$$

where $\sum\vec{F}_{\text{ext}}$ is the sum of all the external forces and $\sum\vec{F}_{\text{int}}$ the sum of the internal forces. According to Newton's third law, each of the internal forces has an equal but oppositely directed force that itself acts on a particle of the system and is therefore included in the sum $\sum\vec{F}_{\text{int}}$. (Each external force is also part of a third-law pair, but forces paired with the external forces act *outside* the system and therefore aren't included in the sum.) Added vectorially, the internal forces therefore cancel in pairs, so $\sum\vec{F}_{\text{int}} = \vec{0}$, and the force $\vec{F}_{\text{total}}$ in Equation 9.1 is just the net *external* force applied to the system. So the point $\vec{r}_{\text{cm}}$ defined in Equation 9.2 does obey Newton's law, written in the form

$$\vec{F}_{\text{net ext}} = M\vec{a}_{\text{cm}} = M\frac{d^2\vec{r}_{\text{cm}}}{dt^2} \tag{9.3}$$

where $\vec{F}_{\text{net ext}}$ is the net external force applied to the system and M is the total mass.

We've defined the center of mass $\vec{r}_{\text{cm}}$ so we can apply Newton's second law to the entire system rather than to each individual particle. As far as its overall motion is concerned, a complex system acts as though all its mass were concentrated at the center of mass.

Finding the Center of Mass

Equation 9.2 shows that the center-of-mass position is an average of the positions of the individual particles, weighted by their masses. For a one-dimensional system, Equation 9.2 becomes $x_{\text{cm}} = \sum m_i x_i / M$; in two and three dimensions, there are similar equations for the center-of-mass coordinates y_{cm} and z_{cm}. Finding the center of mass (CM) is a matter of establishing a coordinate system and then using the components of Equation 9.2.

EXAMPLE 9.1 CM in One Dimension: Weightlifting

Find the center of mass of a barbell consisting of 50-kg and 80-kg weights at the opposite ends of a 1.5-m-long bar of negligible mass.

INTERPRET This is a problem about center of mass. We identify the system as consisting of two "particles"—namely, the two weights.

DEVELOP Figure 9.1 shows the barbell. Here, with just two particles, we have a one-dimensional situation and Equation 9.2, $\vec{r}_{\text{cm}} = \sum m_i \vec{r}_i / M$, becomes $x_{\text{cm}} = (m_1 x_1 + m_2 x_2)/(m_1 + m_2)$. Before we can apply this equation, however, we need a coordinate system. Of course, any coordinate system will do—but a smart choice makes the math easier. Let's take $x = 0$ at the 50-kg mass, so the term $m_1 x_1$ becomes zero.

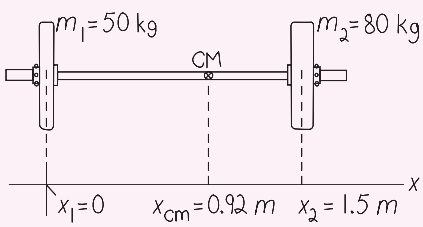


FIGURE 9.1 Our sketch of the barbell.

Our plan is then to find the center-of-mass coordinate x_{cm} using our one-dimensional version of Equation 9.2.

EVALUATE With $x = 0$ at the left end of the barbell, the coordinate of the 80-kg mass is $x_2 = 1.5$ m. So our equation becomes

$$x_{\text{cm}} = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2} = \frac{m_2 x_2}{m_1 + m_2} = \frac{(80 \text{ kg})(1.5 \text{ m})}{(50 \text{ kg} + 80 \text{ kg})} = 0.92 \text{ m}$$

where the equation simplified because of our choice $x_1 = 0$.

ASSESS As Fig. 9.1 shows, this result makes sense: The center of mass is closer to the heavier weight. If the weights had been equal, the center of mass would have been right in the middle.

✓TIP Choosing the Origin

Choosing the origin at one of the masses here conveniently makes one of the terms in the sum $\sum m_i x_i$ zero. But, as always, the choice of origin is purely for convenience and doesn't influence the actual physical location of the center of mass. Exercise 14 demonstrates this point, repeating Example 9.1 with a different origin.

■

EXAMPLE 9.2 CM in Two Dimensions: A Space Station

Figure 9.2 shows a space station consisting of three modules arranged in an equilateral triangle, connected by struts of length L and of negligible mass. Two modules have mass m, the other $2m$. Find the center of mass.

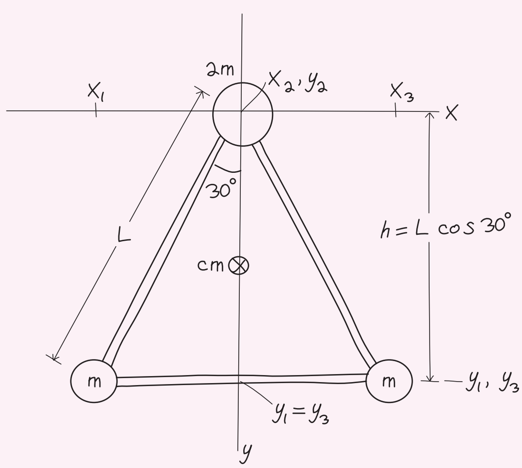


FIGURE 9.2 Our sketch of the space station.

INTERPRET We're after the center of mass of the system consisting of the three modules.

DEVELOP Figure 9.2 is our drawing. We'll use Equation 9.2, $\vec{r}_{\text{cm}} = \sum m_i \vec{r}_i / M$, to find the center-of-mass coordinates x_{cm} and y_{cm}. A sensible coordinate system has the origin at the module with mass $2m$ and the y-axis downward, as shown in Fig. 9.2.

EVALUATE Labeling the modules from left to right, we see that $x_1 = -L \sin 30° = -\frac{1}{2}L$, $y_1 = L \cos 30° = L\sqrt{3}/2$; $x_2 = y_2 = 0$; and $x_3 = -x_1 = \frac{1}{2}L$, $y_3 = y_1 = L\sqrt{3}/2$. Writing explicitly the x- and y-components of Equation 9.2 for this case gives

$$x_{\text{cm}} = \frac{m x_1 + m x_3}{4m} = \frac{m(x_1 - x_1)}{4m} = 0$$

$$y_{\text{cm}} = \frac{m y_1 + m y_3}{4m} = \frac{2m y_1}{4m} = \frac{1}{2} y_1 = \frac{\sqrt{3}}{4} L \simeq 0.43L$$

Although there are three "particles" here, our choice of coordinate system left only two nonzero terms in the numerator, both associated with the same mass m. The more massive module is still in the problem, though; its mass $2m$ contributes to make the total mass M in the denominator equal to $4m$.

(continued)

ASSESS That $x_{\mathrm{cm}} = 0$ is apparent from symmetry (more on this in the following Tip). How about the result for y_{cm}? We have $2m$ at the top of the triangle, and $m + m = 2m$ at the bottom—so shouldn't the center of mass lie midway up the triangle? It does! Expressing the center of mass in terms of the triangle side L obscures this fact. The triangle's height is $h = L\cos 30° = L\sqrt{3}/2$, and our answer for y_{cm} is indeed half this value. We marked the CM on Fig. 9.2.

✓TIP Exploit Symmetries

It's no accident that x_{cm} here lies on the vertical line that bisects the triangle; after all, the triangle is symmetric about that line, so its mass is distributed evenly on either side. Exploit symmetry whenever you can; that can save you a lot of computation throughout physics!

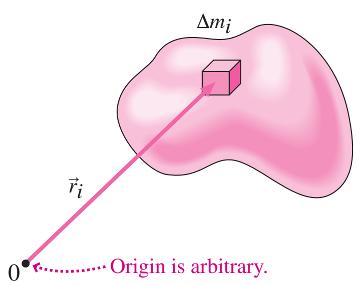


FIGURE 9.3 A chunk of continuous matter, showing one mass element Δm_i and its position vector $\vec{r}_i$.

Continuous Distributions of Matter

We've expressed the center of mass as a sum over individual particles. Ultimately, matter *is* composed of individual particles. But it's often convenient to consider that it's continuously distributed; we don't want to deal with 10^{23} atoms to find the center of mass of a macroscopic object! We can consider continuous matter to be composed of individual pieces of mass Δm_i, with position vectors $\vec{r}_i$; we call these pieces **mass elements** (Fig. 9.3). The center of mass of the entire chunk is then given by Equation 9.2: $\vec{r}_{\mathrm{cm}} = \left(\sum \Delta m_i\, \vec{r}_i\right)/M$, where $M = \sum \Delta m_i$ is the total mass. In the limit as the mass elements become arbitrarily small, this expression becomes an integral:

$$\vec{r}_{\mathrm{cm}} = \lim_{\Delta m_i \to 0} \frac{\sum \Delta m_i \vec{r}_i}{M} = \frac{\int \vec{r}\, dm}{M} \qquad \left(\begin{array}{c}\text{center of mass,}\\ \text{continuous matter}\end{array}\right) \tag{9.4}$$

where the integration is over the entire object. Like the sum in Equation 9.2, the integral of the vector $\vec{r}$ stands for three separate integrals for the components of the center-of-mass position.

EXAMPLE 9.3 Continuous Matter: An Aircraft Wing

A supersonic aircraft wing is an isosceles triangle of length L, width w, and negligible thickness. It has mass M, distributed uniformly over the wing. Where's its center of mass?

INTERPRET Here the matter is distributed continuously, so we need to integrate to find the center of mass. We identify an axis of symmetry through the wing, which we designate the x-axis. By symmetry, the center of mass lies along this x-axis, so $y_{\mathrm{cm}} = 0$ and we'll need to calculate only x_{cm}.

DEVELOP Figure 9.4 shows the wing. Equation 9.4 applies, and we need only the x-component because the y-component is evident from symmetry. The x-component of Equation 9.4 is $x_{\mathrm{cm}} = \left(\int x\, dm\right)/M$. Developing a plan for dealing with an integral like this requires some thought; we'll first do the work and then summarize the general steps involved.

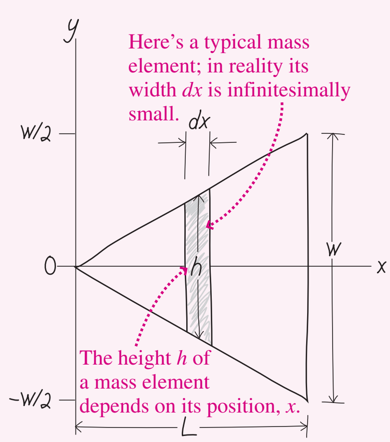


FIGURE 9.4 Our sketch of the supersonic aircraft wing.

Our goal is to find an appropriate mass element dm in terms of the infinitesimal coordinate interval dx. As shown in Fig. 9.4, here it's easiest to use a vertical strip of width dx. Each such strip has a different height h, depending on its position x. If we choose a coordinate system with origin at the wing apex, then, as you can see from the figure, the height grows linearly from 0 at $x = 0$ to w at $x = L$. So $h = (w/L)x$. Now the strip is infinitesimally narrow, so the sloping edges don't matter and its area is that of a very thin rectangle—namely, $h\,dx = (w/L)x\,dx$. The strip's mass dm is then the same fraction of the total wing mass M as its area is of the total wing area $\frac{1}{2}wL$; that is,

$$\frac{dm}{M} = \frac{(w/L)x\,dx}{\frac{1}{2}wL} = \frac{2x\,dx}{L^2}$$

so $dm = 2Mx\,dx/L^2$.

In the integral we weight each mass element dm by its distance x from the origin, and then sum—that is, integrate—over all mass elements. So, from Equation 9.4, we have

$$x_{\mathrm{cm}} = \frac{1}{M}\int x\,dm = \frac{1}{M}\int_0^L x\left(\frac{2Mx}{L^2}dx\right) = \frac{2}{L^2}\int_0^L x^2\,dx$$

As always, constants can come outside the integral. We set the limits 0 and L to cover all the mass elements in the wing. Now we're finally ready to find x_{cm}.

EVALUATE The hard part is done. All that's left is to evaluate the integral:

$$x_{cm} = \frac{2}{L^2}\int_0^L x^2\,dx = \frac{2}{L^2}\frac{x^3}{3}\bigg|_0^L = \frac{2L^3}{3L^2} = \frac{2}{3}L$$

ASSESS Make sense? Yes. Our answer puts the center of mass toward the back of the wing where, because of its increasing width, most of the mass lies. In a complicated calculation like this one, it's reassuring to see that the answer is a quantity with the units of length. ■

TACTICS 9.1 Setting Up an Integral

An integral like $\int x\,dm$ can be confusing because you see both x and dm after the integral sign and they don't seem related. But they are, and here's how to proceed:

1. Find a suitable shape for your mass elements, preferably one that exploits any symmetry in the situation. One dimension of the elements should involve an infinitesimal interval in one of the coordinates x, y, or z. In Example 9.3, the mass elements were strips, symmetric about the wing's centerline and with width dx.
2. Find an expression for the infinitesimal area of your mass elements (in a one-dimensional problem it would be the length; in a three-dimensional problem, the volume). In Example 9.3, the infinitesimal area of each mass element was the strip height h multiplied by the width dx.
3. Form ratios that relate the infinitesimal coordinate interval to the physical quantity in the integral—which in Example 9.3 is the mass element dm. Here we formed the ratio of the area of a mass element to the total area, and equated that to the ratio of dm to the total mass M.
4. Solve your ratio statement for the infinitesimal quantity, in this case dm, that appears in your integral. Then you're ready to evaluate the integral.

Sometimes you'll be given a density—mass per volume, per area, or per length—and then in place of steps 3 and 4 you find dm by multiplying the density by the infinitesimal volume, area, or length you identified in step 2.

Although we described this procedure in the context of Example 9.3, it also applies to other integrals you'll encounter in different areas of physics.

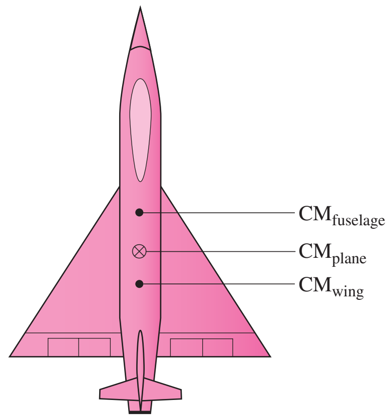


FIGURE 9.5 The center of mass of the airplane is found by treating the wing and fuselage as point particles located at their respective centers of mass.

With more complex objects, it's convenient to find the centers of mass of sub-parts and then treat those as point particles to find the center of mass of the entire object (Fig. 9.5).

The center of mass need not lie within an object, as Fig. 9.6 shows. High jumpers exploit this fact as they straddle the bar with arms and legs dangling on either side (Fig. 9.7). Although the jumper's entire body clears the bar, his center of mass doesn't need to!

GOT IT? 9.1 A thick wire is bent into a semicircle, as shown in Fig. 9.6. Which of the points shown is the center of mass?

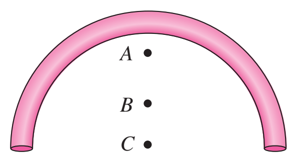


FIGURE 9.6 Got it? The center of mass lies outside the semicircular wire, but which point is it?

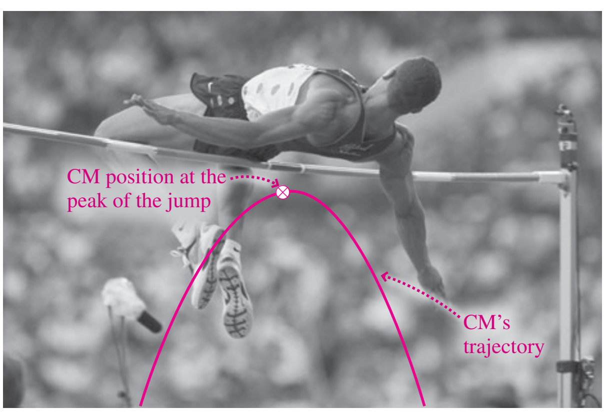


FIGURE 9.7 A high jumper clears the bar, but his center of mass doesn't!

Motion of the Center of Mass

We defined the center of mass so its motion obeys Newton's law $\vec{F}_{\text{net ext}} = M\vec{a}_{\text{cm}}$, with $\vec{F}_{\text{net ext}}$ the net external force on the system and M the total mass. When gravity is the only external force, the center of mass follows the trajectory of a point particle. But if the net external force is zero, then the center-of-mass acceleration $\vec{a}_{\text{cm}}$ is also zero, and the center of mass moves with constant velocity. In the special case of a system at rest, the center of mass remains at rest despite any motions of its internal parts.

EXAMPLE 9.4 CM Motion: Circus Train

Jumbo, a 4.8-t elephant, stands near one end of a 15-t railcar at rest on a frictionless horizontal track. (Here t is for tonne, or metric ton, equal to 1000 kg.) Jumbo walks 19 m toward the other end of the car. How far does the car move?

INTERPRET We're asked about the car's motion, but we can interpret this problem as being fundamentally about the center of mass. We identify the relevant system as comprising Jumbo and the car. Because there's no net external force acting on the system, its center of mass can't move.

DEVELOP Figure 9.8*a* shows the initial situation. The symmetric car has its CM at its center (here we care only about the x-component). Let's take a coordinate system with $x = 0$ at this point—that is, at the *initial* location of the car's center. After the car moves, its center will be somewhere else! Equation 9.2 applies—here in the simpler one-dimensional, two-object form we used in Example 9.1: $x_{\text{cm}} = (m_{\text{J}}x_{\text{J}} + m_{\text{c}}x_{\text{c}})/M$, where we use the subscripts J and c for Jumbo and the car, respectively, and where M is the total mass. We have a before/after situation in which the CM position can't change, so we'll write two versions of this expression, before and after Jumbo's walk. We'll then set them equal to state mathematically that the CM itself doesn't move.

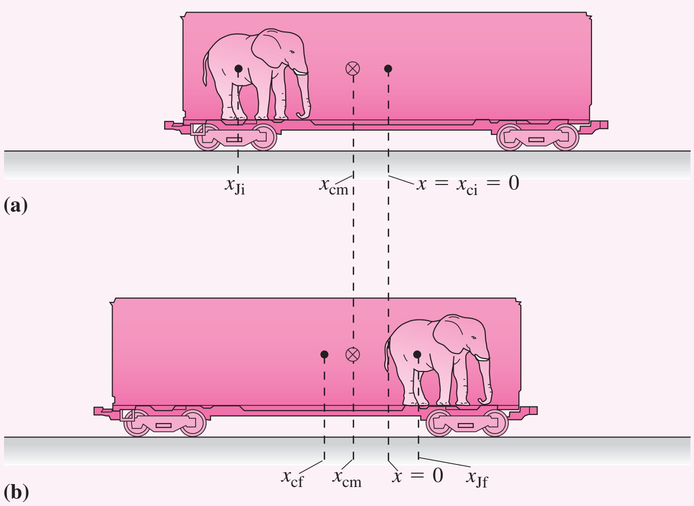


FIGURE 9.8 Jumbo walks, but the center of mass doesn't move.

We chose the coordinates so that $x_{\text{ci}} = 0$, where i designates the initial state, so our initial expression is $x_{\text{cm}} = m_{\text{J}}x_{\text{Ji}}/M$. After Jumbo's walk, our final expression is $x_{\text{cm}} = (m_{\text{J}}x_{\text{Jf}} + m_{\text{c}}x_{\text{cf}})/M$, with f for final. We don't know either coordinate here, but we do know that Jumbo walks 19 m *with respect to the car*. The elephant's final position x_{Jf} is therefore 19 m to the right of x_{Ji}, adjusted by the car's displacement. Therefore Jumbo ends up at $x_{\text{Jf}} = x_{\text{Ji}} + 19\text{ m} + x_{\text{cf}}$. You might think we need a minus sign because the car moves to the left. That's true, but the sign of x_{cf} will take care of that. Trust algebra! So our final expression is

$$x_{\text{cm}} = \frac{m_{\text{J}}x_{\text{Jf}} + m_{\text{c}}x_{\text{cf}}}{M} = \frac{m_{\text{J}}(x_{\text{Ji}} + 19\text{ m} + x_{\text{cf}}) + m_{\text{c}}x_{\text{cf}}}{M}$$

EVALUATE Finally, we execute our plan, equating the two expressions for the unchanging position of the center of mass. The total mass M cancels, and we're left with $m_{\text{J}}x_{\text{Ji}} = m_{\text{J}}(x_{\text{Ji}} + 19\text{ m} + x_{\text{cf}}) + m_{\text{c}}x_{\text{cf}}$. We aren't given x_{Ji}, but the term $m_{\text{J}}x_{\text{Ji}}$ is on both sides of this equation, so it cancels, leaving $0 = m_{\text{J}}(19\text{ m} + x_{\text{cf}}) + m_{\text{c}}x_{\text{cf}}$. We solve for the unknown x_{cf} to get

$$x_{\text{cf}} = -\frac{(19\text{ m})m_{\text{J}}}{(m_{\text{J}} + m_{\text{c}})} = -\frac{(19\text{ m})(4.8\text{ t})}{(4.8\text{ t} + 15\text{ t})} = -4.6\text{ m}$$

The minus sign here indicates a displacement to the left, as we anticipated (Fig. 9.8*b*). Because the masses appear only in ratios, we didn't need to convert to kilograms.

ASSESS The car's 4.6-m displacement is quite a bit less than Jumbo's (which is 19 m − 4.6 m, or 14.4 m relative to the ground). That makes sense because Jumbo is considerably less massive than the car. ■

9.2 Momentum

In Chapter 4 we defined the linear momentum $\vec{p}$ of a particle as $\vec{p} = m\vec{v}$, and we first wrote Newton's law in the form $\vec{F} = d\vec{p}/dt$. We suggested that this form would play an important role in many-particle systems. We're now ready to explore that role.

The momentum of a system of particles is the vector sum of the individual momenta: $\vec{P} = \sum \vec{p}_i = \sum m_i\vec{v}_i$, where m_i and $\vec{v}_i$ are the masses and velocities of the individual particles. But we really don't want to keep track of all the particles in the system. Is there a simpler way to express the total momentum? There is, and it comes from writing the individual velocities as time derivatives of position: $\vec{v} = d\vec{r}/dt$. Then

$$\vec{P} = \sum m_i \frac{d\vec{r}_i}{dt} = \frac{d}{dt}\sum m_i \vec{r}_i$$

where the last step follows because the individual particle masses are constant and because the sum of derivatives is the derivative of the sum. In Section 9.1, we defined the center-of-mass position $\vec{r}_{\rm cm}$ as $\sum m_i\vec{r}_i/M$, where M is the total mass. So the total momentum becomes

$$\vec{P} = \frac{d}{dt}M\vec{r}_{\rm cm}$$

or, assuming the system mass M remains constant,

$$\vec{P} = M\frac{d\vec{r}_{\rm cm}}{dt} = M\vec{v}_{\rm cm} \qquad (9.5)$$

where $\vec{v}_{\rm cm} = d\vec{r}_{\rm cm}/dt$ is the center-of-mass velocity. So a system's momentum is given by an expression similar to that of a single particle; it's the product of the system's mass and its velocity—that is, the velocity of its center of mass. If this seems obvious, watch out! We'll see soon that the same is *not* true for the system's total energy.

If we differentiate Equation 9.5 with respect to time, we have

$$\frac{d\vec{P}}{dt} = M\frac{d\vec{v}_{\rm cm}}{dt} = M\vec{a}_{\rm cm}$$

where $\vec{a}_{\rm cm}$ is the center-of-mass acceleration. But we defined the center of mass so its motion obeyed Newton's second law, $\vec{F} = M\vec{a}_{\rm cm}$, with $\vec{F}$ the net external force on the system. So we can write simply

$$\vec{F}_{\rm net\ ext} = \frac{d\vec{P}}{dt} \qquad (9.6)$$

showing that the momentum of a system of particles changes only if there's a net external force on the system. Remember the hidden role of Newton's third law in all this: Only because forces *internal* to the system cancel in pairs can we ignore them and consider just the external force.

Conservation of Momentum

In the special case when the net external force is zero, Equation 9.6 gives $d\vec{P}/dt = \vec{0}$, so

$$\vec{P} = \text{constant} \qquad \text{(conservation of linear momentum)} \qquad (9.7)$$

Equation 9.7 describes **conservation of linear momentum**, one of the most fundamental laws of physics:

> **Conservation of linear momentum:** When the net external force on a system is zero, the total momentum $\vec{P}$ of the system—the vector sum of the individual momenta $m\vec{v}$ of its constituent particles—remains constant.

Momentum conservation holds no matter how many particles are involved and no matter how they're moving. It applies to systems ranging from atomic nuclei to pool balls, from colliding cars to galaxies. Although we derived Equation 9.7 from Newton's laws, momentum conservation is even more basic, since it applies to subatomic and nuclear systems where the laws and even the language of Newtonian physics are hopelessly inadequate. The following examples show the range and power of momentum conservation.

GOT IT? 9.2 A 500-g fireworks rocket is moving with velocity $\vec{v} = 60\hat{j}$ m/s at the instant it explodes. If you were to add the momentum vectors of all its fragments just after the explosion, what would be the result?

CONCEPTUAL EXAMPLE 9.1 Conservation of Momentum: Kayaking

Jess (mass 53 kg) and Nick (mass 72 kg) sit in a 26-kg kayak at rest on frictionless water. Jess tosses Nick a 17-kg pack, giving it horizontal speed 3.1 m/s relative to the water. What's the kayak's speed after Nick catches the pack? Why can you answer without doing any calculations?

EVALUATE Figure 9.9 shows the kayak before Jess tosses the pack and again after Nick catches it. The water is frictionless, so there's no net external force on the system, which comprises Jess, Nick, the kayak, and the pack. Since there's no net external force, the system's momentum is conserved. Everything is initially at rest, so that momentum is zero. Therefore, it's also zero after Nick catches the pack. At that point Jess, Nick, pack, and kayak are all at rest with respect to each other.

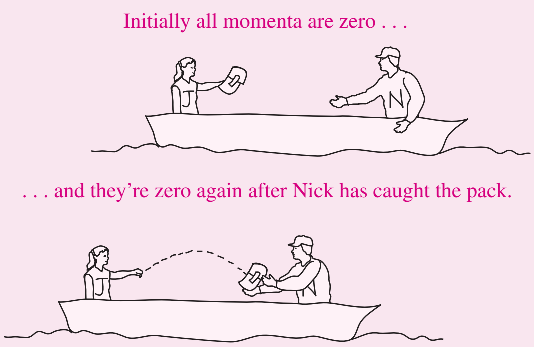


FIGURE 9.9 Our sketch for Conceptual Example 9.1.

ASSESS We didn't need any calculations here because the powerful conservation-of-momentum principle relates the initial and final states, without our having to know what happens in between.

MAKING THE CONNECTION What's the kayak's speed while the pack is in the air?

EVALUATE Momentum conservation still applies, and the system's total momentum is still zero. Now it consists of the pack's momentum $m_p\vec{v}_p$ and the momentum $(m_J + m_N + m_k)\vec{v}_k$ of Jess, Nick, and kayak, with common velocity $\vec{v}_k$ (Fig. 9.10). Sum these momenta, set the sum to zero, and solve, using the given quantities, to get $v_k = -0.35$ m/s. Here we've dropped vector signs; the minus sign then shows that the kayak's velocity is opposite the pack's. Since kayak and passengers are much more massive than the pack, it makes sense that their speed is lower.

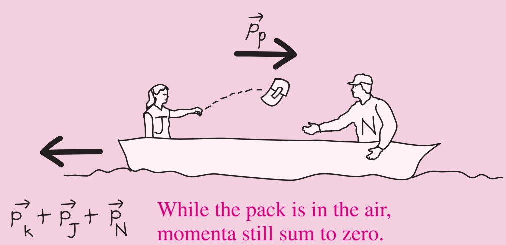


FIGURE 9.10 Our sketch for Making the Connection 9.1.

EXAMPLE 9.5 Conservation of Momentum: Radioactive Decay

A lithium-5 nucleus (^{5}Li) is moving at 1.6 Mm/s when it decays into a proton (^{1}H, or p) and an alpha particle (^{4}He, or α). [Superscripts are the total numbers of nucleons and give the approximate masses in unified atomic mass units (u).] The alpha particle is detected moving at 1.4 Mm/s, at 33° to the original velocity of the ^{5}Li nucleus. What are the magnitude and direction of the proton's velocity?

INTERPRET Although the physical situation here is entirely different from the preceding example, we interpret this one, too, as being about momentum conservation. But there are two differences: First, in this case the total momentum isn't zero, and, second, this situation involves two dimensions. The fundamental principle is the same, however: In the absence of external forces, a system's total momentum can't change. Whether a pack gets tossed or a nucleus decays makes no difference.

DEVELOP Figure 9.11 shows what we know: the velocities for the Li and He nuclei. You can probably guess that the proton must emerge with a downward momentum component, but we'll let the math confirm that. We determine that Equation 9.7, $\vec{P}$ = constant, applies, with the constant equal to the ^{5}Li momentum. After the decay, we have two momenta to account for, so Equation 9.7 becomes

$$m_{Li}\vec{v}_{Li} = m_p\vec{v}_p + m_\alpha\vec{v}_\alpha$$

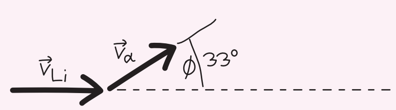


FIGURE 9.11 Our sketch for Example 9.5: what we're given.

Let's choose the x-axis along the direction of $\vec{v}_{\mathrm{Li}}$. Then the two components of the momentum conservation equation become

$$x\text{-component:} \qquad m_{\mathrm{Li}}v_{\mathrm{Li}} = m_p v_{px} + m_\alpha v_{\alpha x}$$

$$y\text{-component:} \qquad 0 = m_p v_{py} + m_\alpha v_{\alpha y}$$

Our plan is to solve these equations for the unknowns v_{px} and v_{py}. From these we can get the magnitude and direction of the proton's velocity.

EVALUATE From Fig. 9.11 it's evident that $v_{\alpha x} = v_\alpha \cos\phi$ and $v_{\alpha y} = v_\alpha \sin\phi$. So we can solve our two equations to get

$$v_{px} = \frac{m_{\mathrm{Li}}v_{\mathrm{Li}} - m_\alpha v_{\alpha x}}{m_p} = \frac{m_{\mathrm{Li}}v_{\mathrm{Li}} - m_\alpha v_\alpha \cos\phi}{m_p}$$

$$= \frac{(5.0\text{ u})(1.6\text{ Mm/s}) - (4.0\text{ u})(1.4\text{ Mm/s})(\cos 33°)}{1.0\text{ u}}$$

$$= 3.30\text{ Mm/s}$$

$$v_{py} = -\frac{m_\alpha v_{\alpha y}}{m_p} = -\frac{m_\alpha v_\alpha \sin\phi}{m_p}$$

$$= \frac{(4.0\text{ u})(1.4\text{ Mm/s})(\sin 33°)}{1.0\text{ u}} = -3.05\text{ Mm/s}$$

Thus the proton's speed $v_p = \sqrt{v_{px}^2 + v_{py}^2} = 4.5$ Mm/s, and its direction is $\theta = \tan^{-1}(v_{py}/v_{px}) = -43°$. Note that here, as in Example 9.4, the masses appear only in ratios so we don't need to change units.

ASSESS Make sense? That negative θ tells us the proton's velocity is downward, as we anticipated. Figure 9.12 makes our result clear. Here we multiplied the velocities by the masses to get momentum vectors. The two momenta after the decay event have equal but opposite vertical components, reflecting that the total momentum of the system never had a vertical component. And the two horizontal components sum to give the initial momentum of the lithium nucleus. Momentum is indeed conserved.

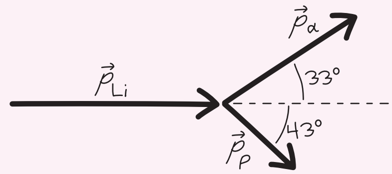


FIGURE 9.12 Our momentum diagram for Example 9.5.

Defining the System: Internal and External Forces

Whether a force is internal or external to a system depends on how we choose to define the system. Sometimes we're interested in the force on a particular object, so it's convenient to define that object as our system. If an external force is acting, then Newton's law says that the system's momentum will change at a rate equal to the net external force. Example 9.6 makes this point.

EXAMPLE 9.6 Changing Momentum: Fighting a Fire

A firefighter directs a stream of water against the window of a burning building, hoping to break the window so water can get to the fire. The hose delivers water at the rate of 45 kg/s, and the water hits the window moving horizontally at 32 m/s. After hitting the window, the water drops vertically. What horizontal force does the water exert on the window?

INTERPRET We're asked about the window, but we're told a lot more about the water. The water stops at the window, so clearly the window exerts a force on the water—and by Newton's third law, that force is equal in magnitude to the force we're after—namely, the force of the water on the window. So we identify the water as our system.

DEVELOP Newton's law in the form $\vec{F} = d\vec{P}/dt$ applies to the water. So our plan is to find the rate at which the water's momentum changes. By Newton's second law, that's equal to the window's force on the water, and by Newton's third law, that's equal to the water's force on the window.

EVALUATE The water strikes the window at 32 m/s, so each kilogram of water loses 32 kg·m/s of momentum. Water strikes the window at the rate of 45 kg/s, so the rate at which it loses momentum to the window is

$$\frac{dP}{dt} = (45\text{ kg/s})(32\text{ m/s}) = 1400\text{ kg}\cdot\text{m/s}^2$$

By Newton's second law, that's equal to the force on the water, and by the third law, that in turn is equal in magnitude to the force on the window. So the window experiences a 1400-N force from the water. Since the window is rigidly attached to the building and the Earth, it doesn't experience significant acceleration—until it breaks and the glass fragments accelerate violently.

ASSESS 1400 N is about twice the weight of a typical person, and a fire hose produces quite a blast of water, so this number seems reasonable. Check the units, too: 1 kg·m/s^2 is equal to 1 N, so our answer does have the units of force.

GOT IT? 9.3 Two skaters toss a basketball back and forth on frictionless ice. Which of the following does not change: (a) the momentum of an individual skater; (b) the momentum of the basketball; (c) the momentum of the system consisting of one skater and the basketball; (d) the momentum of the system consisting of both skaters and the basketball?

APPLICATION **Rockets**

Rockets provide propulsion in the vacuum of space, where there's nothing for a wheel or propeller to push against. If no external forces act, total momentum stays constant. As the rocket's exhaust carries away momentum, the result is an equal but oppositely directed momentum gain for the rocket. The rate of momentum change is the force on the rocket, which engineers call *thrust*. As with the fire hose in Example 9.6, thrust is the product of the exhaust rate dM/dt and exhaust speed v_{ex}: $F = v_{ex}dM/dt$. Because the rocket has to carry the mass it's going to exhaust, the most efficient rockets use high exhaust velocities and therefore need less fuel.

What actually propels the rocket? It's ultimately hot gases inside the rocket engine pushing on the front of the engine chamber. The rocket doesn't "push against" anything outside itself; all the pushing is done *inside* the rocket engine, accelerating the rocket forward. That's why rockets work just fine in the vacuum of space.

9.3 Kinetic Energy of a System

We've seen how the momentum of a many-particle system is determined entirely by the motion of its center of mass; the detailed behavior of the individual particles doesn't matter. For example, a firecracker sliding on ice has the same total momentum before and after it explodes.

The same, however, is *not* true of a system's kinetic energy. Energetically, that firecracker is very different after it explodes; internal potential energy has become kinetic energy of the fragments. Nevertheless, the center-of-mass concept remains useful in categorizing the kinetic energy associated with a system of particles.

The total kinetic energy of a system is the sum of the kinetic energies of the constituent particles: $K = \sum \frac{1}{2}m_i v_i^2$. But the velocity $\vec{v}_i$ of a particle can be written as the vector sum of the center-of-mass velocity $\vec{v}_{cm}$ and a velocity $\vec{v}_{i\,rel}$ of that particle relative to the center of mass: $\vec{v}_i = \vec{v}_{cm} + \vec{v}_{i\,rel}$. Then the total kinetic energy of the system is

$$K = \sum \tfrac{1}{2}m_i(\vec{v}_{cm} + \vec{v}_{i\,rel})\cdot(\vec{v}_{cm} + \vec{v}_{i\,rel}) = \sum \tfrac{1}{2}m_i v_{cm}^2 + \sum m_i\vec{v}_{cm}\cdot\vec{v}_{i\,rel} + \sum \tfrac{1}{2}m_i v_{i\,rel}^2 \quad (9.8)$$

Let's examine the three sums making up the total kinetic energy. Since the center-of-mass speed v_{cm} is common to all particles, it can be factored out of the first sum, so $\sum \frac{1}{2}m_i v_{cm}^2 = \frac{1}{2}v_{cm}^2 \sum m_i = \frac{1}{2}Mv_{cm}^2$, where M is the total mass. This is the kinetic energy of a particle with mass M moving at speed v_{cm}, so we call it K_{cm}, the **kinetic energy of the center of mass**.

The center-of-mass velocity can also be factored out of the second term in Equation 9.8, giving $\sum m_i\vec{v}_{cm}\cdot\vec{v}_{i\,rel} = \vec{v}_{cm}\cdot\sum m_i\vec{v}_{i\,rel}$. Because the $\vec{v}_{i\,rel}$'s are the particle velocities relative to the center of mass, the sum here is the total momentum relative to the center of mass. But that's zero, so the entire second term in Equation 9.8 is zero.

The third term in Equation 9.8, $\sum \frac{1}{2}m_i v_{i\,rel}^2$, is the sum of the individual kinetic energies measured in a frame of reference moving with the center of mass. We call this term K_{int}, the **internal kinetic energy**.

With the middle term gone, Equation 9.8 shows that the kinetic energy of a system breaks into two terms:

$$K = K_{cm} + K_{int} \qquad \text{(kinetic energy of a system)} \qquad (9.9)$$

The first term, the kinetic energy of the center of mass, depends only on the center-of-mass motion. In our firecracker example, K_{cm} doesn't change when the firecracker explodes. The second term, the internal kinetic energy, depends only on the motions of the individual particles relative to the center of mass. The explosion dramatically increases this internal energy.

9.4 Collisions

A **collision** is a brief, intense interaction between objects. Examples abound: automobile collisions; collisions of balls on a pool table; the collision of a tennis ball and racket, baseball and bat, or football and foot; an asteroid colliding with a planet; and collisions of high-energy particles that probe the fundamental structure of matter. Less obvious are collisions among galaxies that last a hundred million years, the interaction of a spacecraft with a planet as the craft gains energy for a voyage to the outer solar system, and the repulsive interaction of two protons that approach but never touch. All these collisions meet two criteria. First, they're brief, lasting but a short time in the overall context of the colliding objects' motions. On a pool table, the collision time is short compared with the time it takes for a ball to roll across the table. An automobile collision lasts a fraction of a second. A baseball spends far more time coming from the pitcher than it does interacting with the bat. And even 10^8 years is short compared with the lifetime of a galaxy. Second, collisions are intense: Forces among the interacting objects are far larger than any external forces that may be acting on the system. External forces are therefore negligible during the collision, so the total momentum of the colliding objects remains unchanged.

Impulse

The forces between colliding objects are *internal* to the system comprising those objects, so they can't alter the total momentum. But they dramatically alter the motions of the colliding objects. How much depends on the magnitude of the force and how long it's applied. If $\overline{\vec{F}}$ is the average force acting on one object during a collision that lasts for time Δt, then Newton's second law reads $\overline{\vec{F}} = \Delta\vec{p}/\Delta t$ or

$$\Delta\vec{p} = \overline{\vec{F}}\,\Delta t \tag{9.10a}$$

The product of average force and time that appears in this equation is called **impulse**. It's given the symbol $\vec{J}$, and its units are newton-seconds. Equation 9.10a shows that a given impulse results in an equal change in momentum.

An impulse $\vec{J}$ produces the same momentum change regardless of whether it involves a larger force exerted over a shorter time or a smaller force exerted over a longer time. The force in a collision usually isn't constant and can fluctuate wildly. In that case, we find the impulse by integrating the force over time, so the momentum change becomes

$$\Delta\vec{p} = \vec{J} = \int \vec{F}(t)\,dt \qquad \text{(impulse)} \tag{9.10b}$$

Although we introduced impulse in the context of collisions, it's useful in other situations involving intense forces applied over short times. For example, small rocket engines are characterized by the impulse they impart.

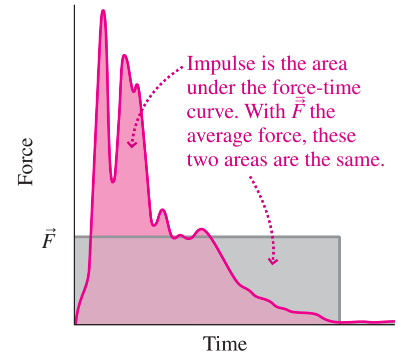
Energy in Collisions

Kinetic energy may or may not be conserved in a collision. If it is, then the collision is **elastic**; if not, it's **inelastic**. An elastic collision requires that the forces between colliding objects be conservative; then kinetic energy is stored briefly as potential energy, and released when the collision is over. Interactions at the atomic and nuclear scales are often truly elastic. In the macroscopic realm, nonconservative forces produce heat or permanently deform the colliding objects, either way robbing the colliding system of kinetic energy. But even many macroscopic collisions are close enough to elastic that we can neglect energy loss during the collision.

GOT IT? 9.4 Which of the following qualifies as a collision? Of the collisions, which are nearly elastic and which inelastic? (a) a basketball rebounds off the backboard; (b) two magnets approach, their north poles facing; they repel and reverse direction without touching; (c) a basketball flies through the air on a parabolic trajectory; (d) a truck strikes a parked car and the two slide off together, crumpled metal hopelessly intertwined; (e) a snowball splats against a tree, leaving a lump of snow adhering to the bark.

APPLICATION **Crash Tests**

Automotive engineers perform crash tests to assess the safety of their vehicles. Sensors measure the rapidly varying forces as the test car collides with a fixed barrier. The graph below is a force-versus-time curve from a typical crash test; impulse is the area under the curve. In addition to force sensors on the vehicle, accelerometers in crash-test dummies determine the maximum accelerations of the heads and other body parts to assess potential injuries.



9.5 Totally Inelastic Collisions

In a **totally inelastic collision**, the colliding objects stick together to form a single object. Even then, kinetic energy is usually not all lost. But a totally inelastic collision entails the maximum energy loss consistent with momentum conservation. The motion after a totally inelastic collision is determined entirely by momentum conservation, and that makes totally inelastic collisions easy to analyze.

Consider masses m_1 and m_2 with initial velocities $\vec{v}_1$ and $\vec{v}_2$ that undergo a totally inelastic collision. After colliding, they stick together to form a single object of mass $m_1 + m_2$ and final velocity $\vec{v}_f$. Conservation of momentum states that the initial and final momenta of this system must be the same:

$$m_1\vec{v}_1 + m_2\vec{v}_2 = (m_1 + m_2)\vec{v}_f \qquad \text{(totally inelastic collision)} \tag{9.11}$$

Given four of the five quantities m_1, $\vec{v}_1$, m_2, $\vec{v}_2$, and $\vec{v}_f$, we can solve for the fifth.

EXAMPLE 9.7 An Inelastic Collision: Hockey

The hockey captain, a physics major, decides to measure the puck's speed. He loads a small Styrofoam chest with sand, giving a total mass of 6.4 kg. He places it at rest on frictionless ice. The 160-g puck strikes the chest and embeds itself in the Styrofoam. The chest moves off at 1.2 m/s. What was the puck's speed?

INTERPRET This is a totally inelastic collision. We identify the system as consisting of puck and chest. Initially, all the system's momentum is in the puck; after the collision, it's in the combination puck + chest. In this case of a single nonzero velocity before collision and a single velocity after, momentum conservation requires that both motions be in the same direction. Therefore, we have a one-dimensional problem.

DEVELOP Figure 9.13 is a sketch of the situation before and after the collision. With a totally inelastic collision, Equation 9.11—the statement of momentum conservation—tells it all. In our one-dimensional situation, this equation becomes $m_p v_p = (m_p + m_c)v_c$, where the subscripts p and c stand for puck and chest, respectively.

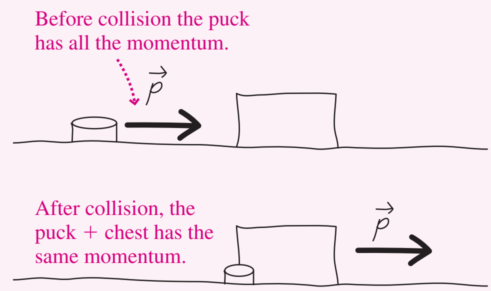


FIGURE 9.13 Our sketch for Example 9.7.

EVALUATE Here we want the initial puck velocity, so we solve for v_p:

$$v_p = \frac{(m_p + m_c)v_c}{m_p} = \frac{(0.16\text{ kg} + 6.4\text{ kg})(1.2\text{ m/s})}{0.16\text{ kg}} = 49\text{ m/s}$$

ASSESS Make sense? Yes: The puck's mass is small, so it needs a much higher speed to carry the same momentum as the much more massive chest. Variations on this technique are often used to determine speeds that would be difficult to measure directly. ■

EXAMPLE 9.8 Conservation of Momentum: Fusion

In a fusion reaction, two deuterium nuclei (^{2}H) join to form helium (^{4}He). Initially, one of the deuterium nuclei is moving at 3.5 Mm/s, the second at 1.8 Mm/s at a 64° angle to the velocity of the first. Find the speed and direction of the helium nucleus.

INTERPRET Although the context is very different, this is another totally inelastic collision. But here both objects are initially moving, and in different directions, so we have a two-dimensional situation. We identify the system as consisting of initially the two deuterium nuclei and finally the single helium nucleus. We're asked for the final velocity of the helium, expressed as magnitude (speed) and direction.

DEVELOP Figure 9.14 shows the situation. Momentum is conserved, so Equation 9.11 applies; solving that equation for $\vec{v}_f$ gives $\vec{v}_f = (m_1\vec{v}_1 + m_2\vec{v}_2)/(m_1 + m_2)$. In two dimensions, this represents two equations for the two components of $\vec{v}_f$. We need a coordinate system, and Fig. 9.14 shows our choice, with the x-axis along the motion of the first deuterium nucleus. We need the components of the initial velocities in order to apply our equation for $\vec{v}_f$.

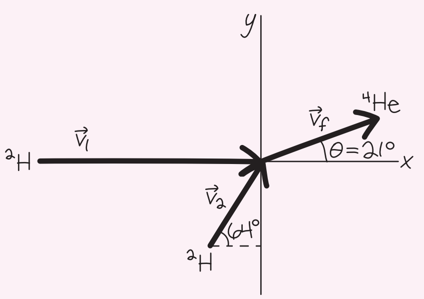


FIGURE 9.14 Our sketch of the velocity vectors for Example 9.8.

EVALUATE With $\vec{v}_1$ in the x-direction, we have $v_{1x} = 3.5$ Mm/s and $v_{1y} = 0$. Figure 9.14 shows that $v_{2x} = (1.8\text{ Mm/s})(\cos 64°) = 0.789$ Mm/s and $v_{2y} = (1.8\text{ Mm/s})(\sin 64°) = 1.62$ Mm/s. So the components of our equation become

$$v_{fx} = \frac{m_1 v_{1x} + m_2 v_{2x}}{m_1 + m_2} = \frac{(2\text{ u})(3.5\text{ Mm/s}) + (2\text{ u})(0.789\text{ Mm/s})}{2\text{ u} + 2\text{ u}} = 2.14\text{ Mm/s}$$

$$v_{fy} = \frac{m_1 v_{1y} + m_2 v_{2y}}{m_1 + m_2} = \frac{0 + (2\text{ u})(1.62\text{ Mm/s})}{2\text{ u} + 2\text{ u}} = 0.809\text{ Mm/s}$$

As in Example 9.5, the superscripts are the nuclear masses in u, and because the mass units cancel, there's no need to convert to kilograms.

From these velocity components we can get the speed and direction: $v_f = \sqrt{v_{fx}^2 + v_{fy}^2} = 2.3$ Mm/s and $\theta = \tan^{-1}(v_{fy}/v_{fx}) = 21°$. We show this final velocity on the diagram in Fig. 9.14.

ASSESS In this example the two incident particles have the same masses, so their velocities are proportional to their momenta. Figure 9.14 shows that the total initial momentum is largely horizontal, with a smaller vertical component, so the 21° angle of the final velocity makes sense. The magnitude of $\vec{v}_f$ also makes sense: Now the total momentum is contained in a single, more massive particle, so we expect a final speed comparable to the initial speeds. ■

EXAMPLE 9.9 The Ballistic Pendulum

The ballistic pendulum measures the speeds of fast-moving objects like bullets. It consists of a wooden block of mass M suspended from vertical strings (Fig. 9.15). A bullet of mass m strikes and embeds itself in the block, and the block swings upward through a vertical distance h. Find an expression for the bullet's speed.

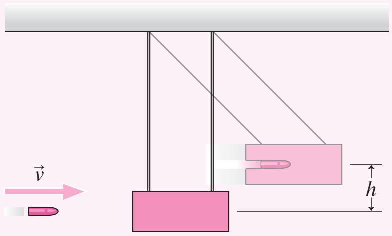


FIGURE 9.15 A ballistic pendulum (Example 9.9).

INTERPRET Interpreting this example is a bit more involved. We actually have two separate events: the bullet striking the block and the subsequent rise of the block. We can interpret the first event as a one-dimensional totally inelastic collision, as in Example 9.7. Momentum is conserved during this event but, because the collision is inelastic, energy is not. Then the block rises, and now a net external force—from string tension and gravity—acts to change the momentum. But gravity is conservative, and the string tension does no work, so now mechanical energy is conserved.

DEVELOP Figure 9.15 is our drawing. Our plan is to separate the two parts of the problem and then to combine the results to get our final answer. First is the inelastic collision; here momentum is conserved, so Equation 9.11 applies. In one dimension, that reads $mv = (m + M)V$, where v is the initial bullet speed and V is the speed of the block with embedded bullet just after the collision. Solving gives $V = mv/(m + M)$. Now the block swings upward. Momentum isn't conserved, but mechanical energy is. Setting the zero of potential energy in the block's initial position, we have $U_0 = 0$ and—using the situation just after the collision as the initial state—$K_0 = \frac{1}{2}(m + M)V^2$. At the peak of its swing the block is momentarily at rest, so $K = 0$. But it's risen a height h, so its potential energy is $U = (m + M)gh$. Conservation of mechanical energy reads $K_0 + U_0 = K + U$—in this case, $\frac{1}{2}(m + M)V^2 = (m + M)gh$.

EVALUATE Now we've got two equations describing the two parts of the problem. Using our expression for V from momentum conservation in the energy-conservation equation, we get

$$\frac{1}{2}\left(\frac{mv}{m + M}\right)^2 = gh$$

Solving for the bullet speed v then gives our answer:

$$v = \left(\frac{m + M}{m}\right)\sqrt{2gh}$$

ASSESS Make sense? Yes: The smaller the bullet mass m, the higher velocity it must have to carry a given momentum; that's reflected by the factor m alone in the denominator. The higher the rise h, obviously, the greater the bullet speed. But the speed scales not as h itself but as $\sqrt{h}$. That's because kinetic energy—which turned into potential energy of the rise—depends on velocity *squared*. ■

9.6 Elastic Collisions

We've seen that momentum is essentially conserved in any collision. In an elastic collision, kinetic energy is conserved as well. In the most general case of a two-body collision, we consider two objects of masses m_1 and m_2, moving initially with velocities $\vec{v}_{1i}$ and $\vec{v}_{2i}$, respectively. Their final velocities after collision are $\vec{v}_{1f}$ and $\vec{v}_{2f}$. Then the conservation statements for momentum and kinetic energy become

$$m_1\vec{v}_{1i} + m_2\vec{v}_{2i} = m_1\vec{v}_{1f} + m_2\vec{v}_{2f} \tag{9.12}$$

and

$$\tfrac{1}{2}m_1v_{1i}^2 + \tfrac{1}{2}m_2v_{2i}^2 = \tfrac{1}{2}m_1v_{1f}^2 + \tfrac{1}{2}m_2v_{2f}^2 \tag{9.13}$$

Given initial velocities, we'd like to predict the outcome of a collision. In the totally inelastic two-dimensional collision, we had enough information to solve the problem. Here, in the two-dimensional elastic case, we have the two components of the momentum conservation equation 9.12 and the single scalar equation for energy conservation 9.13. But we have four unknowns—the magnitudes and directions of both final velocities. With three equations and four unknowns, we don't have enough information to solve the general two-dimensional elastic collision. Later we'll see how other information can help solve such problems. First, though, we look at the special case of a one-dimensional elastic collision.

Elastic Collisions in One Dimension

When two objects collide head-on, the internal forces act along the same line as the incident motion, and the objects' subsequent motion must therefore be along that same line

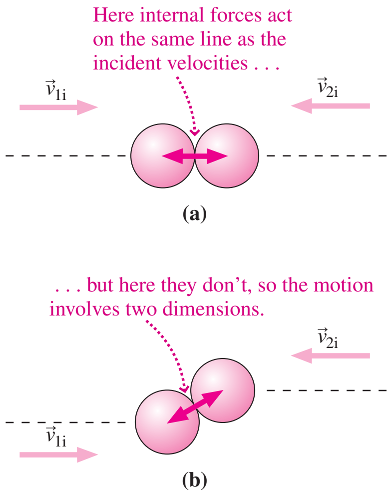


FIGURE 9.16 Only a head-on collision is one-dimensional.

(Fig. 9.16*a*). Although such one-dimensional collisions are a special case, they do occur and they provide much insight into the more general case.

In the one-dimensional case, the momentum conservation equation 9.12 has only one nontrivial component:

$$m_1 v_{1i} + m_2 v_{2i} = m_1 v_{1f} + m_2 v_{2f} \tag{9.12a}$$

where the v's stand for velocity components, rather than magnitudes, and can therefore be positive or negative. If we collect together the terms in Equations 9.12a and 9.13 that are associated with each mass, we have

$$m_1(v_{1i} - v_{1f}) = m_2(v_{2f} - v_{2i}) \tag{9.12b}$$

and

$$m_1(v_{1i}^2 - v_{1f}^2) = m_2(v_{2f}^2 - v_{2i}^2) \tag{9.13a}$$

But $a^2 - b^2 = (a + b)(a - b)$, so Equation 9.13a can be written

$$m_1(v_{1i} - v_{1f})(v_{1i} + v_{1f}) = m_2(v_{2f} - v_{2i})(v_{2f} + v_{2i}) \tag{9.13b}$$

Dividing the left and right sides of Equation 9.13b by the corresponding sides of Equation 9.12b then gives

$$v_{1i} + v_{1f} = v_{2f} + v_{2i}$$

Rearranging shows that

$$v_{1i} - v_{2i} = v_{2f} - v_{1f} \tag{9.14}$$

What does this equation tell us? Both sides describe the relative velocity between the two particles; the equation therefore shows that the relative speed remains unchanged after the collision, although the direction reverses. If the two objects are approaching at a relative speed of 5 m/s, then after collision they'll separate at 5 m/s.

Continuing our search for the final velocities, we solve Equation 9.14 for v_{2f}:

$$v_{2f} = v_{1i} - v_{2i} + v_{1f}$$

and use this result in Equation 9.12a:

$$m_1 v_{1i} + m_2 v_{2i} = m_1 v_{1f} + m_2(v_{1i} - v_{2i} + v_{1f})$$

Solving for v_{1f} then gives

$$v_{1f} = \frac{m_1 - m_2}{m_1 + m_2} v_{1i} + \frac{2m_2}{m_1 + m_2} v_{2i} \tag{9.15a}$$

Problem 69 asks you to show similarly that

$$v_{2f} = \frac{2m_1}{m_1 + m_2} v_{1i} + \frac{m_2 - m_1}{m_1 + m_2} v_{2i} \tag{9.15b}$$

Equations 9.15 are our desired result, expressing the final velocities in terms of the initial velocities alone.

To see that these results make sense, we suppose that $v_{2i} = 0$. (This really isn't a special case, since we can always work in a reference frame with m_2 initially at rest.) We then consider the three special cases of one-dimensional elastic collisions illustrated in Fig. 9.17.

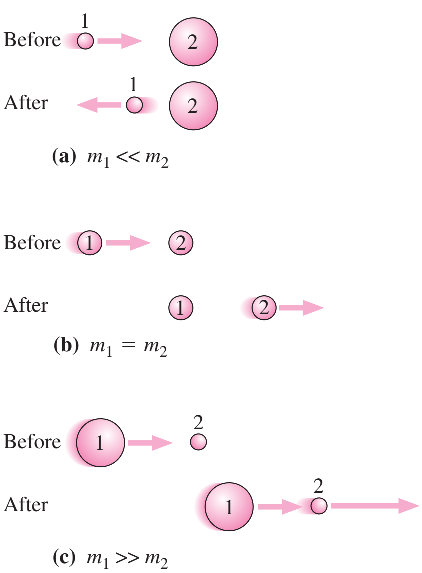


FIGURE 9.17 Special cases of elastic collisions in one dimension.

Case 1: $m_1 \ll m_2$ (Fig. 9.17*a*) Picture a ping-pong ball colliding with a bowling ball, or any object colliding elastically with a perfectly rigid surface. If we set $v_{2i} = 0$ in Equations 9.15, and drop m_1 as being negligible compared with m_2, Equations 9.15 become simply

$$v_{1f} = -v_{1i}$$

and

$$v_{2f} = 0$$

That is, the lighter object rebounds with no change in speed, while the heavier object remains at rest. Does this make sense in light of the conservation laws that Equations 9.15 are supposed to reflect? Clearly energy is conserved: The kinetic energy of m_2 remains zero and the kinetic energy $\frac{1}{2}m_1v_1^2$ is unchanged. But what about momentum? The momentum of the lighter object has changed, from m_1v_{1i} to $-m_1v_{1i}$. But momentum *is* conserved; the momentum given up by the lighter object is absorbed by the heavier object. In the limit of an arbitrarily large m_2, the heavier object can absorb huge amounts of momentum mv without acquiring significant speed. If we "back off" from the extreme case that m_1 can be neglected altogether compared with m_2, we would find that a lighter object striking a heavier one rebounds with reduced speed and that the heavier object begins moving slowly in the opposite direction.

Case 2: $m_1 = m_2$ (Fig. 9.17*b*) Again with $v_{2i} = 0$, Equations 9.15 now give

$$v_{1f} = 0$$

and

$$v_{2f} = v_{1i}$$

So the first object stops abruptly, transferring all its energy and momentum to the second. For purposes of energy transfer, two equal-mass particles are perfectly "matched." We'll encounter analogous instances of energy transfer "matching" when we discuss wave motion and again in connection with electric circuits.

Case 3: $m_1 \gg m_2$ (Fig. 9.17*c*) Now Equations 9.15 give

$$v_{1f} = v_{1i}$$

and

$$v_{2f} = 2v_{1i}$$

where we've neglected m_2 compared with m_1. So here the more massive object barrels right on with no change in motion, while the lighter one heads off with twice the speed of the massive one. This result is entirely consistent with our earlier claim that the relative speed remains unchanged in a one-dimensional elastic collision. How are momentum and energy conserved in this case? In the extreme limit where we neglect the mass m_2, its energy and momentum are negligible. Essentially all the energy and momentum remain with the more massive object, and both these quantities are essentially unchanged in the collision. In the less extreme case where an object of finite mass strikes a less massive object initially at rest, both objects move off in the initial direction of the incident object, with the lighter one moving faster.

EXAMPLE 9.10 Elastic Collisions: Nuclear Engineering

Nuclear power reactors include a substance called a *moderator*, whose job is to slow the neutrons liberated in nuclear fission, making them more likely to induce additional fission and thus sustain a nuclear chain reaction. A Canadian reactor design uses so-called *heavy water* as its moderator. In heavy water, ordinary hydrogen atoms are replaced by deuterium, the rare form of hydrogen whose nucleus consists of a proton and a neutron. The mass of this *deuteron* is thus about 2 u, compared with a neutron's 1 u. Find the fraction of a neutron's kinetic energy that's transferred to an initially stationary deuteron in a head-on elastic collision.

INTERPRET We have a head-on collision, so we're dealing with a one-dimensional situation. The system of interest consists of the neutron and the deuteron. We're not told much else except the masses of the two particles. That should be enough, though, because we're not asked for the final velocities but rather for a ratio of related quantities—namely, kinetic energies.

DEVELOP Since we have a one-dimensional elastic collision, Equations 9.15 apply. We're asked for the fraction of the neutron's kinetic energy that gets transferred to the deuteron, so we need to express the deuteron's final velocity in terms of the neutron's initial velocity. If we take the neutron to be particle 1, then we want Equation 9.15b. With the deuteron initially at rest, $v_{2i} = 0$ and the equation becomes $v_{2f} = 2m_1v_{1i}/(m_1 + m_2)$. Our plan is to use this equation to determine the kinetic-energy ratio.

EVALUATE The kinetic energies of the two particles are given by $K_1 = \frac{1}{2}m_1v_1^2$ and $K_2 = \frac{1}{2}m_2v_2^2$. Using our equation for v_{2f} gives

$$K_2 = \frac{1}{2}m_2\left(\frac{2m_1v_1}{m_1 + m_2}\right)^2 = \frac{2m_2m_1^2v_1^2}{(m_1 + m_2)^2}$$

We want to compare this with K_1:

$$\frac{K_2}{K_1} = K_2\left(\frac{1}{K_1}\right) = \left(\frac{2m_2m_1^2v_1^2}{(m_1 + m_2)^2}\right)\left(\frac{1}{\frac{1}{2}m_1v_1^2}\right) = \frac{4m_1m_2}{(m_1 + m_2)^2} \quad (9.16)$$

(continued)

In this case $m_1 = 1$ u and $m_2 = 2$ u, so we have $K_2/K_1 = 8/9 \simeq 0.89$. Thus 89% of the incident energy is transferred in a single collision, leaving the neutron with 11% of its initial energy.

ASSESS Let's take a look at Equation 9.16 in the context of our three special cases. We numbered this equation because it's a general result for the fractional energy transfer in any one-dimensional elastic collision. In case 1, $m_1 \ll m_2$, so we neglect m_1 compared with m_2 in the denominator; then our energy ratio is approximately $4m_1/m_2$. This becomes zero in the extreme limit where m_1's mass is negligible—consistent with our case 1 where the massive object didn't move at all. In case 2, $m_1 = m_2$, and Equation 9.16 becomes $4m^2/(2m)^2 = 1$, where m is the mass of both objects. That too agrees with our earlier analysis: The incident object stops and transfers all its energy to the struck object. Finally, in case 3, $m_1 \gg m_2$, so we neglect m_2 in the denominator. Now the energy ratio becomes $4m_2/m_1$. As in case 1, this approaches zero as the mass ratio gets extremely large. So the maximum energy transfer occurs with two equal masses, and tails off toward zero if the mass ratio becomes extreme in either direction.

For the particles in this example, the mass ratio 1:2 is close enough to equality that the energy transfer is nearly 90% efficient. Problem 82 explores further this energy transfer. ■

GOT IT? 9.5 One ball is at rest on a level floor. A second ball collides elastically with the first, and the two move off separately but in the same direction. What can you conclude about the masses of the two balls?

Elastic Collisions in Two Dimensions

Analyzing an elastic collision in two dimensions requires the full vector statement of momentum conservation (Equation 9.12), along with the statement of energy conservation (Equation 9.13). But these equations alone don't provide enough information to solve a problem. In a collision between reasonably simple macroscopic objects, that information may be provided by the so-called **impact parameter**, a measure of how much the collision differs from being head-on (Fig. 9.18). More typically—especially with atomic and nuclear interactions—the necessary information must be supplied by measurements done after the collision. Knowing the direction of motion of one particle after collision, for example, provides enough information to analyze a collision if the masses and initial velocities are also known.

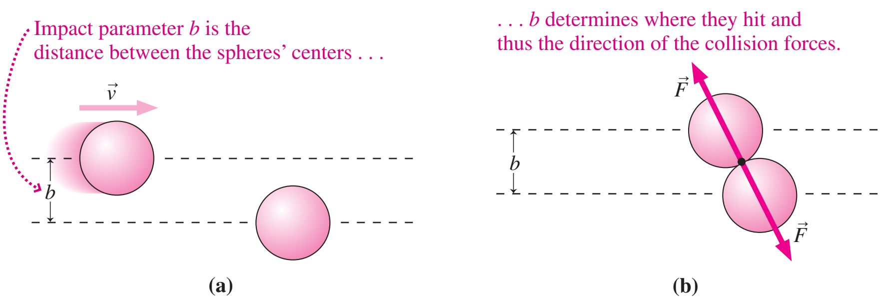


FIGURE 9.18 The impact parameter b determines the directions of the collision forces.

EXAMPLE 9.11 A Two-Dimensional Elastic Collision: Croquet

A croquet ball strikes a stationary ball of equal mass. The collision is elastic, and the incident ball goes off at 30° to its original direction. In what direction does the other ball move?

INTERPRET We've got an elastic collision, so both momentum and kinetic energy are conserved. The system consists of the two croquet balls. We aren't given a lot of information, but since we're asked only for a direction, the magnitudes of the velocities won't matter. Thus we've got what we need to know about the initial velocities, and we've got one other piece of information, so we have enough to solve the problem.

DEVELOP Figure 9.19 shows the situation, in which we're after the unknown angle θ. Since the collision is elastic, Equations 9.12 (momentum conservation) and 9.13 (energy conservation) both apply. The masses are equal, so they cancel from both equations. With $v_{2i} = 0$, we then have $\vec{v}_{1i} = \vec{v}_{1f} + \vec{v}_{2f}$ for momentum conservation and $v_{1i}^2 = v_{1f}^2 + v_{2f}^2$ for energy conservation. The rest will be algebra.

EVALUATE Solving for one unknown in terms of another is going to get messy here, with some velocities squared and some not. Here's a more clever approach: Rather than write the momentum equation in two components, let's take the dot product of each side with itself. That will bring in velocity-squared terms, letting us combine the momentum and energy equations. And the dot product includes an angle—which is what we're asked to find.

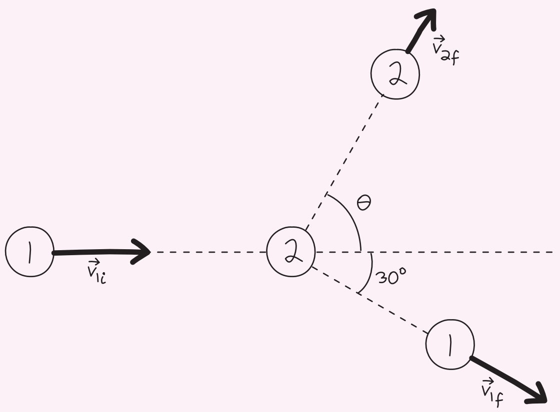


FIGURE 9.19 Our sketch of the collision between croquet balls of equal mass (Example 9.11).

The dot product is distributive and commutative, so here's what we get when we dot the momentum equation with itself:

$$\vec{v}_{1i}\cdot\vec{v}_{1i} = (\vec{v}_{1f}+\vec{v}_{2f})\cdot(\vec{v}_{1f}+\vec{v}_{2f})$$
$$= \vec{v}_{1f}\cdot\vec{v}_{1f}+\vec{v}_{2f}\cdot\vec{v}_{2f}+2\vec{v}_{1f}\cdot\vec{v}_{2f}$$

Recall that the dot product of two vectors is the product of their magnitudes with the cosine of the angle between them: $\vec{A}\cdot\vec{B} = AB\cos\theta$. Since the angle between a vector and itself is zero, the dot product of a vector with itself is the square of its magnitude: $\vec{A}\cdot\vec{A} = A^2\cos(0) = A^2$. So our equation becomes

$$v_{1i}^2 = v_{1f}^2 + v_{2f}^2 + 2v_{1f}v_{2f}\cos(\theta + 30°)$$

where the argument of the cosine follows because, as Fig. 9.19 shows, the angle between $\vec{v}_{1f}$ and $\vec{v}_{2f}$ is $\theta + 30°$. We now subtract the energy equation from this new equation to get

$$2v_{1f}v_{2f}\cos(\theta + 30°) = 0$$

But neither of the final speeds is zero, so this equation requires that $\cos(\theta + 30°) = 0$. Thus $\theta + 30° = 90°$, and our answer follows: $\theta = 60°$.

ASSESS This result seems reasonable, although we don't have a lot to go on because we haven't calculated the final speeds. But it's intriguing that the two balls go off at right angles to each other. Is this a coincidence? No: It happens in any two-dimensional elastic collision between objects of equal mass, when one is initially at rest. ■

The Center-of-Mass Frame

Two-dimensional collisions take a particularly simple form in a frame of reference moving with the center of mass of the colliding particles, since the total momentum in such a frame must be zero. That remains true after a collision, which involves only *internal* forces that don't affect the center of mass. Therefore, both the initial and final momenta form pairs of oppositely directed vectors of equal magnitude, as shown in Fig. 9.20. In an elastic collision, energy conservation requires further that the incident and final momenta have the same values, so a single number—the angle θ in Fig 9.20—completely describes the collision.

It's often easier to analyze a collision by transforming to the center-of-mass frame, doing the analysis, and then transforming the resulting momentum and velocity vectors back to the original or "lab" frame. High-energy physicists routinely make such transformations as they seek to understand the fundamental forces between elementary particles. Those forces are described most simply in the center-of-mass frame of colliding particles, but in some experiments—those where lighter particles slam into massive nuclei or stationary targets—the physicists and their particle accelerators are not in the center-of-mass frame.

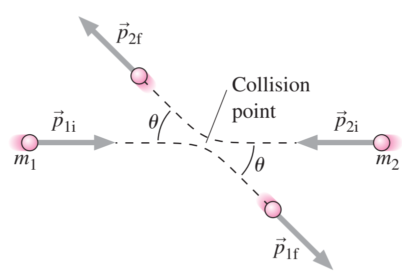


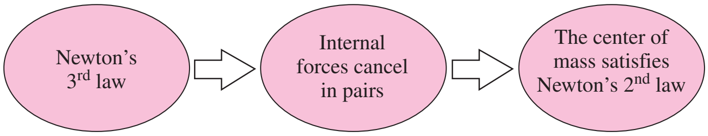
FIGURE 9.20 An elastic collision viewed in the center-of-mass frame, showing that the initial and final momentum vectors form pairs with equal magnitudes and opposite directions.

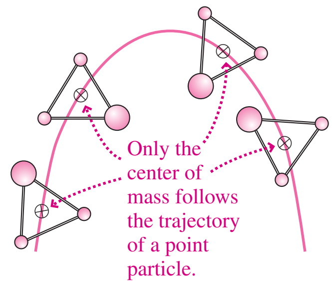
CHAPTER 9 SUMMARY

Big Picture

The big idea of this chapter is that systems consisting of many particles exhibit simple behaviors that don't depend on the complexities of their internal structure or motions. That, in turn, allows us to understand those internal details. In particular, a system responds to external forces as though it were a point particle located at the **center of mass**. If the net external force on a system is zero, then the center of mass does not accelerate and the system's total momentum is conserved. Conservation of momentum holds to a very good approximation during the brief, intense encounters called **collisions**, allowing us to relate particles' motions before and after colliding.

Newton's second and third laws are behind these big ideas. The third law, in particular, says that forces *internal* to a system cancel in pairs, and therefore they don't contribute to the net force on the system. That's what allows us to describe a system's overall motion without having to worry about what's going on internally.





Key Concepts and Equations

The **center of mass** position $\vec{r}_{cm}$ is a weighted average of the positions of a system's constituent particles:

$$\vec{r}_{cm} = \frac{\sum m_i \vec{r}_i}{M} \quad \text{or, with continuous matter,} \quad \vec{r}_{cm} = \frac{\int \vec{r}\, dm}{M}$$

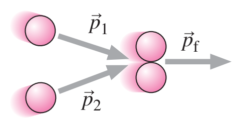
Here M is the system's total mass and the sum or integral is taken over the entire system. The center of mass obeys Newton's second law:

$$\vec{F}_{\text{net ext}} = M\vec{a}_{cm} = \frac{d\vec{P}}{dt}$$

where $\vec{F}_{\text{net ext}}$ is the net external force on the system, $\vec{a}_{cm}$ the acceleration of the center of mass, and $\vec{P}$ the system's total momentum.

A **collision** is a brief, intense interaction between particles involving large internal forces. External forces have little effect during a collision, so to a good approximation the total momentum of the interacting particles is conserved.

In a **totally inelastic collision**, the colliding objects stick together to form a composite; in that case momentum conservation entirely determines the outcome:



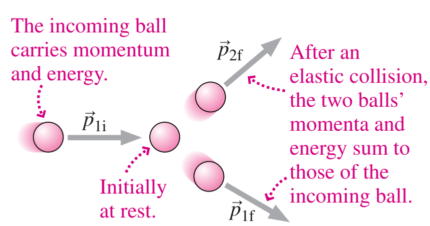
$$m_1\vec{v}_1 + m_2\vec{v}_2 = (m_1 + m_2)\vec{v}_f \qquad \text{(conservation of momentum, totally inelastic collision)}$$

An **elastic collision** conserves kinetic energy as well as momentum, and the colliding particles separate after the collision:

$$m_1\vec{v}_{1i} + m_2\vec{v}_{2i} = m_1\vec{v}_{1f} + m_2\vec{v}_{2f} \qquad \text{(conservation of momentum, elastic collision)}$$

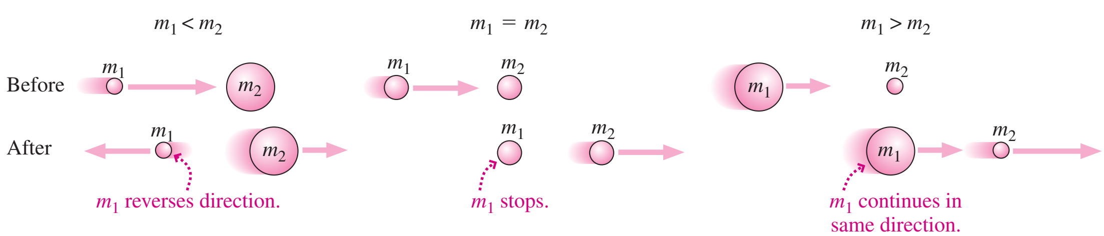
$$\tfrac{1}{2}m_1v_{1i}^2 + \tfrac{1}{2}m_2v_{2i}^2 = \tfrac{1}{2}m_1v_{1f}^2 + \tfrac{1}{2}m_2v_{2f}^2 \qquad \text{(conservation of energy, elastic collision)}$$

In the special case of a one-dimensional elastic collision, knowledge of the mass and initial velocities is sufficient to determine the outcome. To analyze elastic collisions in two dimensions requires an additional piece of information, such as the impact parameter or the direction of one of the particles after the collision.



Applications

One-dimensional collisions with one object initially at rest provide insights into the nature of collisions. There are three cases, depending on the relative masses:



Rockets provide an important technological application of momentum conservation. A rocket exhausts matter out the back at high velocity; momentum conservation then requires that the rocket gain momentum in the forward direction. Rocket propulsion requires no interaction with any external material, which is why rockets work in space.

For Thought and Discussion

1. Roughly where is your center of mass when you're standing?
2. Explain why a high jumper's center of mass need not clear the bar.
3. The center of mass of a solid sphere is clearly at its center. If the sphere is cut in half and the two halves are stacked as in Fig. 9.21, is the center of mass at the point where they touch? If not, roughly where is it? Explain.

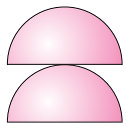


FIGURE 9.21 For Thought and Discussion 3

4. The momentum of a system of pool balls is the same before and after they are hit by the cue ball. Is it still the same after one of the balls strikes the edge of the table? Explain.
5. An hourglass is inverted and placed on a scale. Compare the scale readings (a) before sand begins to hit the bottom; (b) while sand is hitting the bottom; and (c) when all the sand is on the bottom.
6. Why are cars designed so that their front ends crumple during an accident?
7. Give three everyday examples of inelastic collisions.
8. Is it possible to have an inelastic collision in which all the kinetic energy of the colliding objects is lost? If so, give an example. If not, why not?
9. If you want to stop the neutrons in a reactor, why not use massive nuclei like lead?
10. A pitched baseball moves no faster than the pitcher's hand. But a batted ball can move much faster than the bat. What's the difference?
11. Two identical satellites are going in opposite directions in the same circular orbit when they collide head-on. Describe their subsequent motion if the collision is (a) elastic or (b) inelastic.

Exercises and Problems

Exercises

Section 9.1 Center of Mass

12. A 28-kg child sits at one end of a 3.5-m-long seesaw. Where should her 65-kg father sit so the center of mass will be at the center of the seesaw?
13. Two particles of equal mass m are at the vertices of the base of an equilateral triangle. The triangle's center of mass is midway between the base and the third vertex. What's the mass at the third vertex?
14. Rework Example 9.1 with the origin at the center of the barbell, showing that the physical location of the center of mass doesn't depend on your coordinate system.
15. Three equal masses lie at the corners of an equilateral triangle of side L. Find the center of mass.
16. How far from Earth's center is the center of mass of the Earth-Moon system? (*Hint*: Consult Appendix E.)

Section 9.2 Momentum

17. A popcorn kernel at rest in a hot pan bursts into two pieces, with masses 91 mg and 64 mg. The more massive piece moves horizontally at 47 cm/s. Describe the motion of the second piece.
18. A 60-kg skater, at rest on frictionless ice, tosses a 12-kg snowball with velocity $\vec{v} = 53.0\hat{\imath} + 14.0\hat{\jmath}$ m/s, where the x- and y-axes are in the horizontal plane. Find the skater's subsequent velocity.
19. A plutonium-239 nucleus at rest decays into a uranium-235 nucleus by emitting an alpha particle (^4He) with kinetic energy 5.15 MeV. Find the speed of the uranium nucleus.
20. A toboggan of mass 8.6 kg is moving horizontally at 23 km/h. As it passes under a tree, 15 kg of snow drop onto it. Find its subsequent speed.

Section 9.3 Kinetic Energy of a System

21. A 150-g trick baseball is thrown at 60 km/h. It explodes in flight into two pieces, with a 38-g piece continuing straight ahead at 85 km/h. How much energy do the pieces gain in the explosion?
22. An object with kinetic energy K explodes into two pieces, each of which moves with twice the speed of the original object. Compare the internal and center-of-mass energies after the explosion.

Section 9.4 Collisions

23. Two 140-kg satellites collide at an altitude where $g = 8.7\ \text{m/s}^2$, and the collision imparts an impulse of 1.8×10^5 N·s to each. If the collision lasts 120 ms, compare the collisional impulse to that imparted by gravity. Your result should show why you can neglect the external force of gravity.
24. BIO High-speed photos of a 220-μg flea jumping vertically show that the jump lasts 1.2 ms and involves an average vertical acceleration of $100g$. What (a) average force and (b) impulse does the ground exert on the flea during its jump? (c) What's the change in the flea's momentum during its jump?
25. You're working in mission control for an interplanetary space probe. A trajectory correction calls for a rocket firing that imparts an impulse of 5.64 N·s. If the rocket's average thrust is 135 mN, how long should the rocket fire?

Section 9.5 Totally Inelastic Collisions

26. In a railroad switchyard, a 56-ton freight car is sent at 7.0 mi/h toward a 31-ton car moving in the same direction at 2.6 mi/h. (a) What's the speed of the cars after they couple? (b) What fraction of the initial kinetic energy was lost in the collision?
27. In a totally inelastic collision between two equal masses, with one initially at rest, show that half the initial kinetic energy is lost.
28. A neutron (mass 1 u) strikes a deuteron (mass 2 u), and they combine to form a tritium nucleus. If the neutron's initial velocity was $28\hat{\imath} + 17\hat{\jmath}$ Mm/s and if the tritium leaves the reaction with velocity $12\hat{\imath} + 20\hat{\jmath}$ Mm/s, what was the deuteron's velocity?
29. Two identical trucks have mass 5500 kg when empty, and the maximum permissible load for each is 8000 kg. The first truck, carrying 3800 kg, is at rest. The second truck plows into it at 65 km/h, and the pair moves away at 40 km/h. As an expert witness, you're asked to determine whether the first truck was overloaded. What do you report?

Section 9.6 Elastic Collisions

30. An alpha particle (^4He) strikes a stationary gold nucleus (^{197}Au) head-on. What fraction of the alpha's kinetic energy is transferred to the gold? Assume a totally elastic collision.
31. Playing in the street, a child accidentally tosses a ball at 18 m/s toward the front of a car moving toward him at 14 m/s. What's the ball's speed after it rebounds elastically from the car?
32. A block of mass m undergoes a one-dimensional elastic collision with a block of mass M initially at rest. If both blocks have the same speed after colliding, how are their masses related?
33. A proton moving at 6.9 Mm/s collides elastically head-on with a second proton moving in the opposite direction at 11 Mm/s. Find their subsequent velocities.

34. A head-on, elastic collision between two particles with equal initial speed v leaves the more massive particle (m_1) at rest. Find (a) the ratio of the particle masses and (b) the final speed of the less massive particle.

Problems

35. Find the center of mass of a pentagon with five equal sides of length a, but with one triangle missing (Fig. 9.22). (*Hint:* See Example 9.3, and treat the pentagon as a group of triangles.)

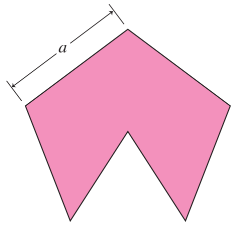


FIGURE 9.22 Problem 35

36. BIO Wildlife biologists fire 20-g rubber bullets to stop a rhinoceros charging at 0.81 m/s. The bullets strike the rhino and drop vertically to the ground. The biologists' gun fires 15 bullets each second, at 73 m/s, and it takes 34 s to stop the rhino. (a) What impulse does each bullet deliver? (b) What's the rhino's mass? Neglect forces between rhino and ground.
37. Consider a system of three equal-mass particles moving in a plane; their positions are given by $a_i\hat{\imath} + b_i\hat{\jmath}$, where a_i and b_i are functions of time with the units of position. Particle 1 has $a_1 = 3t^2 + 5$ and $b_1 = 0$; particle 2 has $a_2 = 7t + 2$ and $b_2 = 2$; particle 3 has $a_3 = 3t$ and $b_3 = 2t + 6$. Find the center-of-mass position, velocity, and acceleration of the system as functions of time.
38. You're with 19 other people on a boat at rest in frictionless water. The group's total mass is 1500 kg, and the boat's mass is 12,000 kg. The entire party walks the 6.5-m distance from bow to stern. How far does the boat move?
39. A hemispherical bowl is at rest on a frictionless counter. A mouse drops onto the bowl's rim from a cabinet directly overhead. The mouse climbs down inside the bowl to eat crumbs at the bottom. If the bowl moves along the counter a distance equal to one-tenth of its diameter, how does the mouse's mass compare with the bowl's mass?
40. BIO Physicians perform *needle biopsies* to sample tissue from internal organs. A spring-loaded gun shoots a hollow needle into the tissue; extracting the needle brings out the tissue core. A particular device uses 8.3-mg needles that take 90 ms to stop in the tissue, which exerts a stopping force of 41 mN. (a) Find the impulse imparted by the tissue. (b) How far into the tissue does the needle penetrate?
41. Find the center of mass of the uniform, solid cone of height h, base radius R, and constant density ρ shown in Fig. 9.23. (*Hint*: Integrate over disk-shaped mass elements of thickness dy, as shown in the figure.)

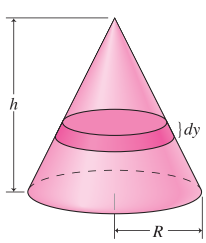


FIGURE 9.23 Problem 41

42. A firecracker, initially at rest, explodes into two fragments. The first, of mass 14 g, moves in the $+x$-direction at 48 m/s. The second moves at 32 m/s. Find the second fragment's mass and the direction of its motion.
43. An 11,000-kg freight car rests against a spring bumper at the end of a railroad track. The spring has constant $k = 0.32$ MN/m. The car is hit by a second car of 9400-kg mass moving at 8.5 m/s, and the two couple together. Find (a) the maximum compression of the spring and (b) the speed of the two cars when they rebound together from the spring.
44. On an icy road, a 1200-kg car moving at 50 km/h strikes a 4400-kg truck moving in the same direction at 35 km/h. The pair is soon hit from behind by a 1500-kg car speeding at 65 km/h, and all three vehicles stick together. Find the speed of the wreckage.
45. A car of mass M is initially at rest on a frictionless surface. A jet of water carrying mass at the rate dm/dt and moving horizontally at speed v_0 strikes the rear window of the car, which is at 45° to the horizontal; the water bounces off at the same relative speed with which it hit the window, as shown in Fig. 9.24. Find expressions for (a) the car's initial acceleration and (b) the maximum speed it reaches.

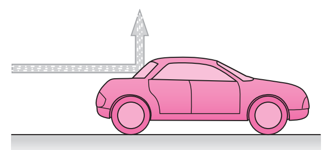


FIGURE 9.24 Problem 45

46. A 950-kg compact car is moving with velocity $\vec{v}_1 = 32\hat{\imath} + 17\hat{\jmath}$ m/s. It skids on a frictionless icy patch and collides with a 450-kg hay wagon with velocity $\vec{v}_2 = 12\hat{\imath} + 14\hat{\jmath}$ m/s. If the two stay together, what's their velocity?
47. Masses m and $3m$ approach at the same speed v and collide head-on. Show that mass $3m$ stops, while mass m rebounds at speed $2v$.
48. A ^{238}U nucleus is moving in the x-direction at 5.0×10^5 m/s when it decays into an alpha particle (^4He) and a ^{234}Th nucleus. The alpha moves at 1.4×10^7 m/s at 22° above the x-axis. Find the recoil velocity of the thorium.
49. A cylindrical concrete silo is 4.0 m in diameter and 30 m high. It consists of a 6000-kg concrete base and 38,000-kg cylindrical concrete walls. Locate the center of mass of the silo (a) when it's empty and (b) when it's two-thirds full of silage whose density is 800 kg/m^3. Neglect the thickness of the walls and base.
50. A 42-g firecracker is at rest at the origin when it explodes into three pieces. The first, with mass 12 g, moves along the x-axis at 35 m/s. The second, with mass 21 g, moves along the y-axis at 29 m/s. Find the velocity of the third piece.
51. A 60-kg astronaut floating in space simultaneously tosses away a 14-kg oxygen tank and a 5.8-kg camera. The tank moves in the x-direction at 1.6 m/s, and the astronaut recoils at 0.85 m/s in a direction 200° counterclockwise from the x-axis. Find the camera's velocity.
52. Assuming equal-mass pieces in Exercise 22, find the angles of the two velocities relative to the direction of motion before the explosion.
53. A 55-kg sprinter stands on the left end of a 240-kg cart moving leftward at 7.6 m/s. She runs to the right end and continues horizontally off the cart. What should be her speed relative to the cart so that once she's off the cart, she has no horizontal velocity relative to the ground?

54. You're a production engineer in a cookie factory, where mounds of dough drop vertically onto a conveyer belt at the rate of one 12-g mound every 2 seconds. You're asked to design a mechanism that will keep the conveyor belt moving at a constant 50 cm/s. What average force must the mechanism exert on the belt?
55. Mass m, moving at speed $2v$, approaches mass $4m$, moving at speed v. The two collide elastically head-on. Find expressions for their subsequent speeds.
56. Verify explicitly that kinetic energy is conserved in the collision of the preceding problem.
57. While standing on frictionless ice, you (mass 65.0 kg) toss a 4.50-kg rock with initial speed 12.0 m/s. If the rock is 15.2 m from you when it lands, (a) at what angle did you toss it? (b) How fast are you moving?
58. You're an accident investigator at a scene where a drunk driver in a 1600-kg car has plowed into a 1300-kg parked car with its brake set. You measure skid marks showing that the combined wreckage moved 25 m before stopping, and you determine a frictional coefficient of 0.77. What do you report for the drunk driver's speed just before the collision?
59. A fireworks rocket is launched vertically upward at 40 m/s. At the peak of its trajectory, it explodes into two equal-mass fragments. One reaches the ground 2.87 s after the explosion. When does the second reach the ground?
60. Two objects moving in opposite directions with the same speed v undergo a totally inelastic collision, and half the initial kinetic energy is lost. Find the ratio of their masses.
61. Explosive bolts separate a 950-kg communications satellite from its 640-kg booster rocket, imparting a 350-N·s impulse. At what relative speed do satellite and booster separate?
62. You're working in quality control for a model rocket manufacturer, testing a class-D rocket whose specifications call for an impulse between 10 and 20 N·s. The rocket's burn time is $\Delta t = 2.8$ s, and its thrust during that time is $F(t) = at(t - \Delta t)$, where $a = -4.6\ \text{N/s}^2$. Does the rocket meet its specs?
63. A 1200-kg Toyota and a 2200-kg Buick collide at right angles in an intersection. They skid together 22 m; the coefficient of friction is 0.91. Show that at least one car must have exceeded the 25 km/h speed limit at the intersection.
64. A 400-mg popcorn kernel is skittering across a nonstick frying pan at 8.2 cm/s when it pops and breaks into two equal-mass pieces. If one piece ends up at rest, how much energy was released in the popping?
65. Two identical objects with the same initial speed collide and stick together. If the composite object moves with half the initial speed of either object, what was the angle between the initial velocities?
66. A proton (mass 1 u) moving at 6.90 Mm/s collides elastically head-on with a second particle moving in the opposite direction at 2.80 Mm/s. After the collision, the proton is moving opposite its initial direction at 8.62 Mm/s. Find the mass and final velocity of the second particle.
67. Two objects, one initially at rest, undergo a one-dimensional elastic collision. If half the kinetic energy of the initially moving object is transferred to the other object, what is the ratio of their masses?
68. Blocks B and C have masses $2m$ and m, respectively, and are at rest on a frictionless surface. Block A, also of mass m, is heading at speed v toward block B as shown in Fig. 9.25. Determine the final velocity of each block after all subsequent collisions are over. Assume all collisions are elastic.

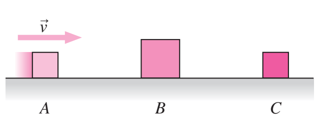


FIGURE 9.25 Problem 68

69. Derive Equation 9.15b.
70. An object collides elastically with an equal-mass object initially at rest. If the collision isn't head-on, show that the final velocity vectors are perpendicular.
71. A proton (mass 1 u) collides elastically with a stationary deuteron (mass 2 u). If the proton is deflected 37° from its original direction, what fraction of its kinetic energy does it transfer to the deuteron?
72. Two identical billiard balls are initially at rest when they're struck symmetrically by a third identical ball moving with velocity $\vec{v}_0 = v_0\hat{\imath}$ (Fig. 9.26). Find the velocities of all three balls after this elastic collision.

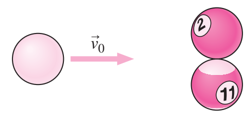


FIGURE 9.26 Problem 72

73. Find an expression for the impulse imparted by a force $F(t) = F_0 \sin(at)$ during the time $t = 0$ to $t = \pi/a$. Here a is a constant with units of s^{-1}.
74. A 32-u oxygen molecule (O_2) moving in the $+x$-direction at 580 m/s collides with an oxygen atom (mass 16 u) moving at 870 m/s at 27° to the x-axis. The particles stick together to form an ozone molecule. Find the ozone's velocity.
75. A 114-g Frisbee is lodged on a tree branch 7.65 m above the ground. To free it, you lob a 240-g dirt clod vertically upward. The dirt leaves your hand at a point 1.23 m above the ground, moving at 17.7 m/s. It sticks to the Frisbee. Find (a) the maximum height reached by the Frisbee-dirt combination and (b) the speed with which the combination hits the ground.
76. You set a small ball of mass m atop a large ball of mass $M \gg m$ and drop the pair from height h. Assuming the balls are perfectly elastic, show that the smaller ball rebounds to height $9h$.
77. A car moving at speed v undergoes a one-dimensional collision with an identical car initially at rest. The collision is neither elastic nor fully inelastic; 5/18 of the initial kinetic energy is lost. Find the velocities of the two cars after the collision.
78. A 200-g block is released from rest at a height of 25 cm on a frictionless 30° incline. It slides down the incline and then along a frictionless surface until it collides elastically with an 800-g block at rest 1.4 m from the bottom of the incline (Fig. 9.27). How much later do the two blocks collide again?

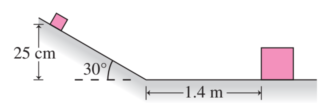


FIGURE 9.27 Problem 78

79. A 14-kg projectile is launched at 380 m/s at a 55° angle to the horizontal. At the peak of its trajectory it collides with a second projectile moving horizontally, in the opposite direction, at 140 m/s. The two stick together and land 9.6 km horizontally downrange from the first projectile's launch point. Find the mass of the second projectile.
80. During a crash test, a car moving at 50 km/h collides with a rigid barrier and comes to a complete stop in 200 ms. The collision force as a function of time is given by $F = at^4 + bt^3 + ct^2 + dt$, where $a = -8.86\ \text{GN/s}^4$, $b = 3.27\ \text{GN/s}^3$, $c = -362\ \text{MN/s}^2$, and $d = 12.5$ MN/s. Find (a) the total impulse imparted by the

collision, (b) the average collisional force, and (c) the car's mass.

81. Use numerical or graphical techniques to estimate the peak force of the collision in the preceding problem, and determine when it occurs.
82. A block of mass m_1 undergoes a one-dimensional elastic collision with an initially stationary block of mass m_2. Find an expression for the fraction of the initial kinetic energy transferred to the second block, and plot your result for mass ratios m_1/m_2 from 0 to 20.
83. Two objects of unequal mass, one initially at rest, undergo a one-dimensional elastic collision. For a given mass ratio, show that the fraction of the initial energy transferred to the initially stationary object doesn't depend on which object it is.
84. In Figure 9.6, the uniform semicircular wire has radius R. How far above the center of the semicircle is its center of mass?
85. Find the center of mass of a uniform slice of pizza with radius R and angular width θ.
86. In a ballistic pendulum demonstration gone bad, a 0.52-g pellet, fired horizontally with kinetic energy 3.25 J, passes straight through a 400-g Styrofoam pendulum block. If the pendulum rises a maximum height of 0.50 mm, how much kinetic energy did the pellet have after emerging from the Styrofoam?
87. An 80-kg astronaut has become detached from the safety line connecting her to the International Space Station. She's 200 m from the station, at rest relative to it, and has 4 min of air remaining. To get herself back, she tosses a 10-kg tool kit away from the station at 8.0 m/s. Will she make it back in time?
88. Astronomers detect extrasolar planets by measuring the slight movement of stars around the center of mass of the star-planet system. Considering just the Sun and Jupiter, determine the radius of the circular orbit the Sun makes about the Sun-Jupiter center of mass.
89. A thin rod extends from $x = 0$ to $x = L$. It carries a nonuniform mass per unit length $\mu = Mx^a/L^{1+a}$, where M is a constant with units of mass, and a is a non-negative dimensionless constant. Find expressions for (a) the rod's mass and (b) the location of its center of mass. (c) Are your results what you expect when $a = 0$?

Passage Problems

You're interested in the intersection of physics and sports, and you recognize that many sporting events involve collisions—bat and baseball, foot and football, hockey stick and puck, basketball and floor. Using strobe photography, you embark on a study of such collisions. Figure 9.28 is your strobe photo of a ball bouncing off the floor. The ball is launched from a point near the top left of the photo and your camera then captures it undergoing three subsequent collisions with the floor.

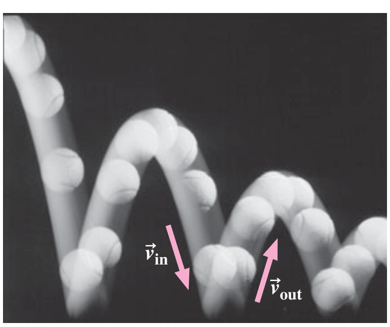


FIGURE 9.28 Passage Problems 90–93

90. The collisions between ball and floor are
    a. totally elastic.
    b. totally inelastic.
    c. neither totally elastic nor totally inelastic.
91. The fraction of the ball's mechanical energy that's lost in the second collision is
    a. about 10%.
    b. a little less than half.
    c. a little more than half.
    d. about 90%.
92. The component of the ball's velocity whose magnitude is most affected by the collisions is
    a. horizontal.
    b. vertical.
    c. Both are affected equally.
93. Compared with the time between bounces, the duration of each collision is
    a. a tiny fraction of the time between bounces.
    b. a significant fraction of the time between bounces.
    c. much longer than the time between bounces.

Answers to Chapter Questions

Answer to Chapter Opening Question

The skier's center of mass follows the simple path of a projectile because, as Newton's laws show, the skier's mass acts like it's all concentrated at this point.

Answers to GOT IT? Questions

9.1. The CM is the uppermost point A. You can see this by imagining horizontal strips through the loop; the higher the strip the more mass is included, so the CM must lie nearer the top of the loop. The bottommost point would be the CM for a complete circle.

9.2. Momentum is conserved, so the momentum both before and after the explosion is the same: $\vec{P} = m\vec{v} = (0.50\text{ kg})(60\hat{\jmath}\text{ m/s}) = 30\hat{\jmath}\text{ kg}\cdot\text{m/s}$.

9.3. Only (d). The individual skaters experience external forces from the ball, as does the ball from the skaters. A system consisting of the ball and one skater experiences external forces from the other skater. Only the system of all three has no net external force and therefore has conserved momentum.

9.4. All but (c) are collisions; (a) and (b) are nearly elastic; (d) and (e) are inelastic.

9.5. The ball initially at rest is less massive; otherwise, the incident ball would have reversed direction (or stopped if the masses were equal).

10 Rotational Motion

For a given blade mass, how should you engineer a wind turbine's blades so it's easiest for the wind to get the turbine rotating?

New Concepts, New Skills

By the end of this chapter you should be able to

- Describe the rotational motion of rigid bodies using the concepts of angular velocity and acceleration, rotational inertia, torque, and the rotational analog of Newton's law (10.1–10.3).
- Calculate the rotational inertias of objects with sufficient symmetry by summing or integrating (10.3).
- Solve problems that involve both linear and rotational motion (10.3).
- Calculate rotational kinetic energy, and explain its relation to torque and work (10.4).
- Describe rolling motion (10.5).

Connecting Your Knowledge

- The description of rotational motion here is directly analogous to Chapter 2's material on one-dimensional linear motion (2.1–2.4).
- This chapter's rotational analog of Newton's law builds on the one-dimensional applications of Newton's second law in Chapter 4 (4.2, 4.5).
- You should be comfortable with the concepts of work and kinetic energy introduced in Chapter 6 (6.1, 6.3).

You're sitting on a rotating planet. The wheels of your car rotate. Your favorite movie comes from a rotating DVD. A circular saw rotates to rip its way through a board. A dancer pirouettes, and a satellite spins about its axis. Even molecules rotate. Rotational motion is commonplace throughout the physical universe.

In principle, we could treat rotational motion by analyzing the motion of each particle in a rotating object. But that would be a hopeless task for all but the simplest objects. Instead, we'll describe rotational motion by analogy with linear motion as governed by Newton's laws.

This chapter parallels our study of one-dimensional motion in Chapters 2 and 4. In the next chapter we introduce a full vector description to treat multidimensional rotational motion.

10.1 Angular Velocity and Acceleration

You slip a DVD into a player, and it starts spinning. You could describe its motion by giving the speed and direction of each point on the disc. But it's much easier just to say that the disc is rotating at 800 revolutions per minute (rpm). As long as the disc is a

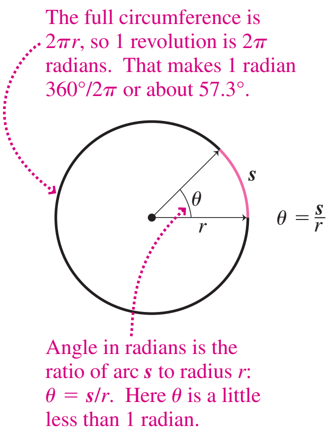


FIGURE 10.1 Radian measure of angles.

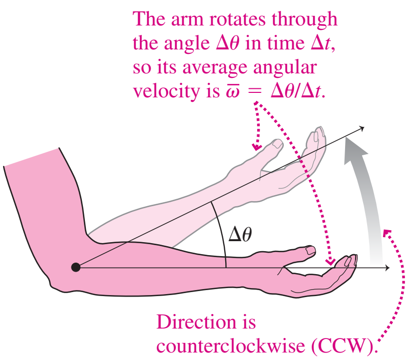


FIGURE 10.2 Average angular velocity.

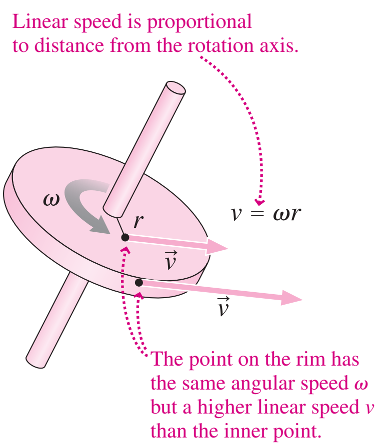


FIGURE 10.3 Linear and rotational speeds.

rigid body—one whose parts remain in fixed positions relative to one another—then that single statement suffices to describe the motion of the entire disc.

Angular Velocity

The rate at which a body rotates is its **angular velocity**—the rate at which the angular position of any point on the body changes. With our 800-rpm DVD, the unit of angle was one full revolution (360°, or 2π radians), and the unit of time was the minute. But we could equally well express angular speed in revolutions per second (rev/s), degrees per second (°/s), or radians per second (rad/s or simply s^{-1} since radians are dimensionless). Because of the mathematically simple nature of radian measure, we often use radians in calculations involving rotational motion (Fig. 10.1).

We use the Greek symbol ω (omega) for angular velocity and define **average angular velocity** $\overline{\omega}$ as

$$\overline{\omega} = \frac{\Delta\theta}{\Delta t} \qquad \text{(average angular velocity)} \qquad (10.1)$$

where $\Delta\theta$ is the **angular displacement**—that is, the change in angular position—occurring in the time Δt (Fig. 10.2). When angular velocity is changing, we define **instantaneous angular velocity** as the limit over arbitrarily short time intervals:

$$\omega = \lim_{\Delta t \to 0} \frac{\Delta\theta}{\Delta t} = \frac{d\theta}{dt} \qquad \text{(instantaneous angular velocity)} \qquad (10.2)$$

These definitions are analogous to those of average and instantaneous linear velocity introduced in Chapter 2. Just as we use the term *speed* for the magnitude of velocity, so we define **angular speed** as the magnitude of the angular velocity.

Velocity is a vector quantity, with magnitude and direction. Is angular velocity also a vector? Yes, but we'll wait until the next chapter for the full vector description of rotational motion. In this chapter, it's sufficient to know whether an object's rotation is clockwise (CW) or counterclockwise (CCW) about a fixed axis—as suggested by the curved arrow in Fig. 10.2. This restriction to a fixed axis is analogous to Chapter 2's restriction to one-dimensional motion.

Angular and Linear Speed

Individual points on a rotating object undergo circular motion. Each point has an instantaneous linear velocity $\vec{v}$ whose magnitude is the linear speed v. We now relate this linear speed v to the angular speed ω. The definition of angular measure in radians (Fig. 10.1) is $\theta = s/r$. Differentiating this expression with respect to time, we have

$$\frac{d\theta}{dt} = \frac{1}{r}\frac{ds}{dt}$$

because the radius r is constant. But $d\theta/dt$ is the angular velocity, and ds/dt is the linear speed v, so $\omega = v/r$, or

$$v = \omega r \qquad (10.3)$$

Thus the linear speed of any point on a rotating object is proportional both to the angular speed of the object and to the distance from that point to the axis of rotation (Fig. 10.3).

✓TIP Radian Measure

Equation 10.3 was derived using the definition of angle *in radians* and therefore holds for only angular speed measured in radians per unit time. If you're given other angular measures—degrees or revolutions, for example—you should convert to radians before using Equation 10.3.

EXAMPLE 10.1 **Angular Speed: A Wind Turbine**

A wind turbine's blades are 28 m long and rotate at 21 rpm. Find the angular speed of the blades in radians per second, and determine the linear speed at the tip of a blade.

INTEPRET This problem is about converting between two units of angular speed, revolutions per minute and radians per second, as well as finding linear speed given angular speed and radius.

DEVELOP We'll first convert the units to radians per second and then calculate the linear speed using Equation 10.3, $v = \omega r$.

EVALUATE One revolution is 2π rad, and 1 min is 60 s, so we have

$$\omega = 21 \text{ rpm} = \frac{(21 \text{ rev/min})(2\pi \text{ rad/rev})}{60 \text{ s/min}} = 2.2 \text{ rad/s}$$

The speed at the tip of a 28-m-long blade then follows from Equation 10.3: $v = \omega r = (2.2 \text{ rad/s})(28 \text{ m}) = 62$ m/s.

ASSESS With ω in radians per second, multiplying by length in meters gives correct velocity units of meters per second because radians are dimensionless.

Angular Acceleration

If the angular velocity of a rotating object changes, then the object undergoes **angular acceleration** α, defined analogously to linear acceleration:

$$\alpha = \lim_{\Delta t \to 0} \frac{\Delta \omega}{\Delta t} = \frac{d\omega}{dt} \quad \text{(angular acceleration)} \tag{10.4}$$

Taking the limit gives the instantaneous angular acceleration; if we don't take the limit, then we have an average over the time interval Δt. The SI units of angular acceleration are rad/s^2, although we sometimes use other units such as rpm/s or rev/s^2.

Angular acceleration has the same direction as angular velocity—CW or CCW—if the angular speed is increasing, and the opposite direction if it's decreasing. These situations are analogous to a car that's speeding up (acceleration and velocity in the same direction) or braking (acceleration opposite velocity).

When a rotating object undergoes angular acceleration, points on the object speed up or slow down. Therefore, they have **tangential acceleration** dv/dt directed parallel or antiparallel to their linear velocity (Fig. 10.4). We introduced this idea of tangential acceleration back in Chapter 3; here we can recast it in terms of the angular acceleration:

$$a_t = \frac{dv}{dt} = r\frac{d\omega}{dt} = r\alpha \quad \text{(tangential acceleration)} \tag{10.5}$$

Whether or not there's angular acceleration, points on a rotating object also have **radial acceleration** because they're in circular motion. Radial acceleration is given, as usual, by $a_r = v^2/r$; using $v = \omega r$ from Equation 10.3, we can recast this equation in angular terms as $a_r = \omega^2 r$.

Because angular velocity and acceleration are defined analogously to linear velocity and acceleration, all the relations among linear position, velocity, and acceleration automatically apply among angular position, angular velocity, and angular acceleration. If angular acceleration is constant, then all our constant-acceleration formulas of Chapter 2 apply when we make the substitutions θ for x, ω for v, and α for a. Table 10.1 summarizes this direct analogy between

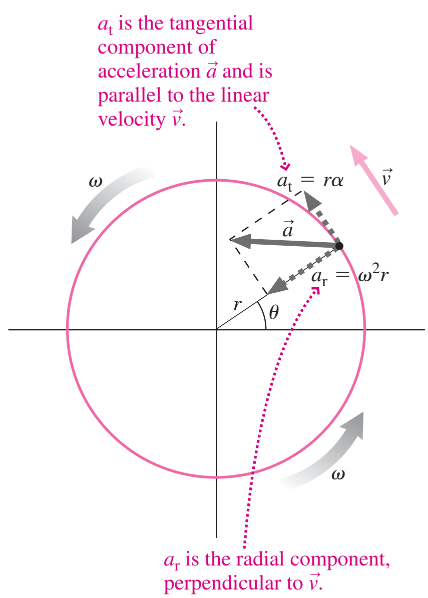


FIGURE 10.4 Radial and tangential acceleration.

Table 10.1 Angular and Linear Position, Velocity, and Acceleration

Linear Quantity	Angular Quantity
Position x	Angular position θ
Velocity $v = \dfrac{dx}{dt}$	Angular velocity $\omega = \dfrac{d\theta}{dt}$
Acceleration $a = \dfrac{dv}{dt} = \dfrac{d^2x}{dt^2}$	Angular acceleration $\alpha = \dfrac{d\omega}{dt} = \dfrac{d^2\theta}{dt^2}$

Equations for Constant Linear Acceleration		Equations for Constant Angular Acceleration	
$\bar{v} = \frac{1}{2}(v_0 + v)$	(2.8)	$\bar{\omega} = \frac{1}{2}(\omega_0 + \omega)$	(10.6)
$v = v_0 + at$	(2.7)	$\omega = \omega_0 + \alpha t$	(10.7)
$x = x_0 + v_0 t + \frac{1}{2}at^2$	(2.10)	$\theta = \theta_0 + \omega_0 t + \frac{1}{2}\alpha t^2$	(10.8)
$v^2 = v_0^2 + 2a(x - x_0)$	(2.11)	$\omega^2 = \omega_0^2 + 2\alpha(\theta - \theta_0)$	(10.9)

linear and rotational quantities. With Table 10.1, problems involving rotational motion are analogous to the one-dimensional linear problems you solved in Chapter 2.

EXAMPLE 10.2 Linear Analogies: Spin-down

When the wind dies, the turbine of Example 10.1 spins down with constant angular acceleration of magnitude $0.12\ \mathrm{rad/s^2}$. How many revolutions does the turbine make before coming to a stop?

INTERPRET The key to problems involving rotational motion is to identify the analogous situation for linear motion. This problem is analogous to asking how far a braking car travels before coming to a stop. We identify the number of rotations—the angular displacement—as the analog of the car's linear displacement. The given angular acceleration is analogous to the car's braking acceleration. The initial angular speed (2.2 rad/s, from Example 10.1) is analogous to the car's initial speed. And in both cases the final state we're interested in has zero speed—whether linear or angular.

DEVELOP Our plan is to develop the analogy further so we can find the angular displacement. The easiest way to solve the linear problem would be to use Equation 2.11, $v^2 = v_0^2 + 2a(x - x_0)$, with $v = 0$, v_0 the initial velocity, a the car's acceleration, and $\Delta x = x - x_0$ the distance we're solving for. In Table 10.1, Equation 10.9 is the analogous equation for rotational motion: $\omega^2 = \omega_0^2 + 2\alpha\Delta\theta$, where we've written $\theta - \theta_0 = \Delta\theta$ for the rotational displacement during the spin-down.

EVALUATE We solve for $\Delta\theta$:

$$\Delta\theta = \frac{\omega^2 - \omega_0^2}{2\alpha} = \frac{0 - (2.2\ \mathrm{rad/s})^2}{(2)(-0.12\ \mathrm{rad/s^2})} = 20\ \mathrm{rad} = 3.2\ \mathrm{revolutions}$$

where the last conversion follows because 1 revolution is 2π radians.

ASSESS The turbine blades are turning rather slowly—less than 1 revolution every second—so it's not surprising that a small angular acceleration can bring them to a halt in a short angular "distance." Note, too, how the units work out. Also, by taking ω as positive, we needed to treat α as negative because the angular acceleration is opposite the angular velocity when the rotation rate is slowing—just as the braking car's linear acceleration is opposite its velocity. ■

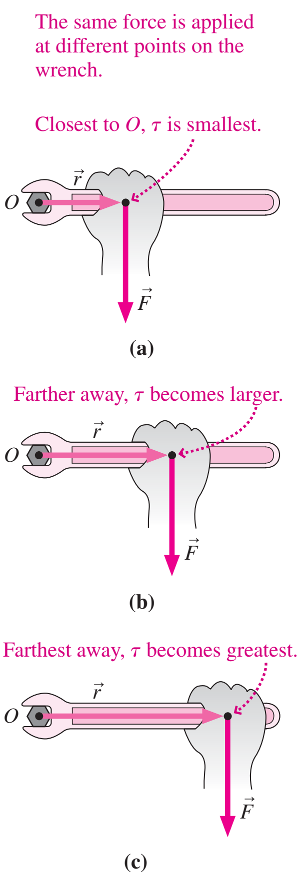


FIGURE 10.5 Torque increases with the distance r from the rotation axis O to the point where force is applied.

10.2 Torque

Newton's second law, $\vec{F} = m\vec{a}$, proved very powerful in our study of motion. Ultimately Newton's law governs all motion, but its application to every particle in a rotating object would be terribly cumbersome. Can we instead formulate an analogous law that deals with rotational quantities?

To develop such a law, we need rotational analogs of force, mass, and acceleration. Angular acceleration α is the analog of linear acceleration; in the next two sections we develop analogs for force and mass.

How can a small child balance her father on a seesaw? By sitting far from the seesaw's rotation axis; that way, her smaller weight at a greater distance from the pivot is as effective as her father's greater weight closer to the pivot. In general, the effectiveness of a force in bringing about changes in rotational motion—a quantity called **torque**—depends not only on the magnitude of the force but also on how far from the rotation axis it's applied (Fig. 10.5). The effectiveness of the force also depends on the *direction* in which it's applied, as Fig. 10.6 suggests. Based on these considerations, we define torque as the product of the distance r from the rotation axis and the component of force perpendicular to that axis. Torque is given the symbol τ (Greek tau, pronounced to rhyme with "how"). Then we can write

$$\tau = rF\sin\theta \qquad (10.10)$$

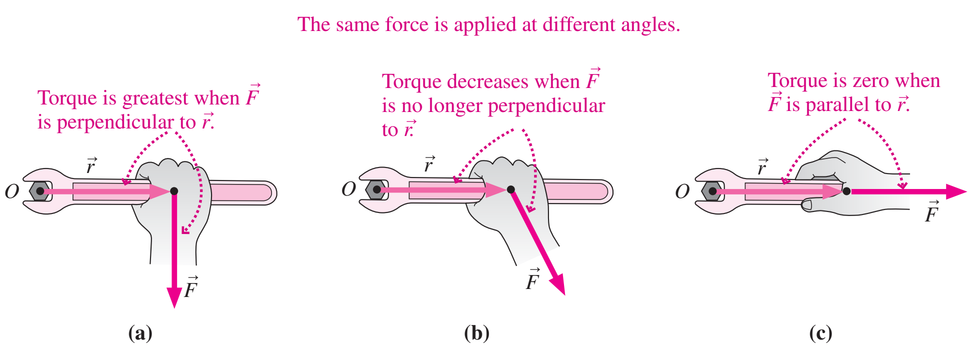


FIGURE 10.6 Torque is greatest with $\vec{F}$ and $\vec{r}$ at right angles, and diminishes to zero as they become colinear.

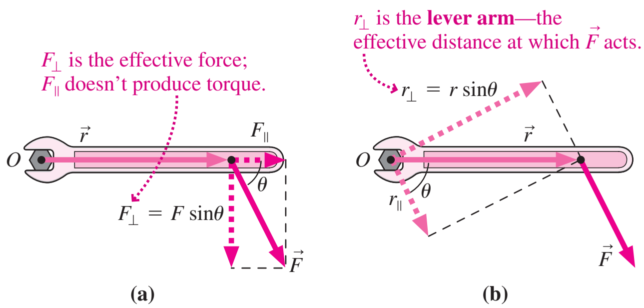


FIGURE 10.7 Two ways of thinking about torque. (a) $\tau = rF_\perp$; (b) $\tau = r_\perp F$. Both give $\tau = rF \sin\theta$.

where θ is the angle between the force vector and the vector $\vec{r}$ from the rotation axis to the force application point. Figure 10.7 shows two interpretations of Equation 10.10. Figure 10.7*b* also defines the so-called **lever arm**.

Torque, which you can think of as a "twisting force," plays the role of force in the rotational analog of Newton's second law. Equation 10.10 shows that torque is measured in newton-meters. Although this is the same unit as energy, torque is a different physical quantity, so we reserve the term *joule* $(=1\ \text{N}\cdot\text{m})$ for energy.

Does torque have direction? Yes, and we'll extend our notion of torque to provide a vector description in the next chapter. For now we'll specify the direction as either clockwise or counterclockwise.

EXAMPLE 10.3 Torque: Changing a Tire

You're tightening your car's wheel nuts after changing a flat tire. The instructions specify a tightening torque of 95 N·m so the nuts won't come loose. If your 45-cm-long wrench makes a 67° angle with the horizontal, with what force must you pull horizontally to produce the required torque?

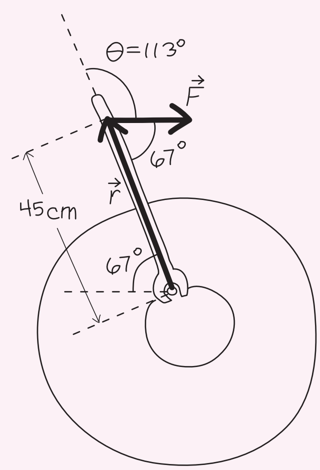


FIGURE 10.8 Our sketch of the wrench and wheel nut.

INTERPRET We need to find the force required to produce a specific torque, given the distance from the rotation axis and the angle the force makes with the wrench.

DEVELOP Figure 10.8 is our drawing, and we'll calculate the torque using Equation 10.10, $\tau = rF \sin\theta$. With the force applied horizontally, comparison of Figs. 10.7*a* and 10.8 shows that the angle θ in Equation 10.10 is $180° - 67° = 113°$.

EVALUATE We solve Equation 10.10 for the force F:

$$F = \frac{\tau}{r \sin\theta} = \frac{95\ \text{N}\cdot\text{m}}{(0.45\ \text{m})(\sin 113°)} = 230\ \text{N}$$

ASSESS Is a 230-N force reasonable? Yes: It's roughly the force needed to lift a 23-kg $(\sim 50\text{-lb})$ suitcase. Tightening torques, as in this example, are often specified for nuts and bolts in critical applications. Mechanics use specially designed "torque wrenches" that provide a direct indication of the applied torque. ■

✓TIP Specify the Axis

Torque depends on where the force is applied *relative to some rotation axis*. The same physical force results in different torques about different axes. Be sure the rotation axis is specified before you make a calculation involving torque.

GOT IT? 10.1 The forces in Figs. 10.5 and 10.6 all have the same magnitude. (a) Which of Figs. 10.5*a*, 10.5*b*, and 10.6*b* has the greatest torque? (b) Which of these has the least torque?

10.3 Rotational Inertia and the Analog of Newton's Law

Torque and angular acceleration are the rotational analogs of force and linear acceleration. To develop a rotational analog of Newton's law, we still need the rotational analog of mass.

The mass m in Newton's law is a measure of a body's inertia—of its resistance to changes in motion. So we want a quantity that describes resistance to changes in rotational motion. Figure 10.9 shows that it's easier to set an object rotating when its mass is concentrated near the rotation axis. Therefore, our rotational analog of inertia must depend not only on mass itself but also on the distribution of mass relative to the rotation axis.

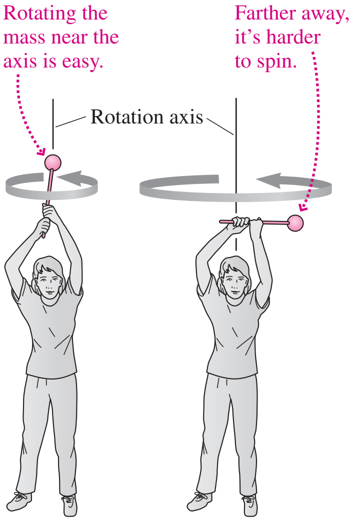


FIGURE 10.9 It's easier to set an object rotating if the mass is concentrated near the axis.

Suppose the object in Fig. 10.9 consists of an essentially massless rod of length R with a ball of mass m on the end. We allow the object to rotate about an axis through the free end of the rod and apply a force $\vec{F}$ to the ball, always at right angles to the rod (Fig. 10.10). The ball undergoes a tangential acceleration given by Newton's law: $F = ma_t$. (There's also a tension force in the rod, but because it acts along the rod, it doesn't contribute to the torque or angular acceleration.) We can use Equation 10.5 to express the tangential acceleration in terms of the angular acceleration α and the distance R from the rotation axis: $F = ma_t = m\alpha R$. We can also express the force F in terms of its associated torque. Since the force is perpendicular to the rod, Equation 10.10 gives $\tau = RF$. Using our expression for F, we have

$$\tau = (mR^2)\alpha$$

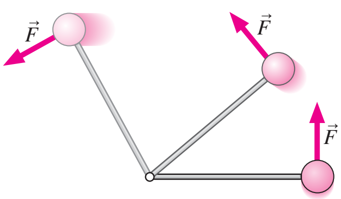


FIGURE 10.10 A force applied perpendicular to the rod results in angular acceleration.

Here we have Newton's law, $F = ma$, written in terms of rotational quantities. The torque—analogous to force—is the product of the angular acceleration and the quantity mR^2, which must therefore be the rotational analog of mass. We call this quantity the **rotational inertia** or **moment of inertia** and give it the symbol I. Rotational inertia is measured in kg·m^2 and accounts for both an object's mass and the distribution of that mass. Like torque, the value of the rotational inertia depends on the location of the rotation axis. Given the rotational inertia I, our rotational analog of Newton's law becomes

$$\tau = I\alpha \qquad \text{(rotational analog of Newton's 2}^{\text{nd}}\text{ law)} \qquad (10.11)$$

Although we derived Equation 10.11 for a single, localized mass, it applies to extended objects if we interpret τ as the net torque on the object and I as the sum of the rotational inertias of the individual mass elements making up the object.

Calculating the Rotational Inertia

When an object consists of a number of discrete mass points, its rotational inertia about an axis is the sum of the rotational inertias of the individual mass points:

$$I = \sum m_i r_i^2 \qquad \text{(rotational inertia)} \qquad (10.12)$$

Here m_i is the mass of the ith mass point, and r_i is its distance from the rotation axis.

EXAMPLE 10.4 Rotational Inertia: A Sum

A dumbbell-shaped object consists of two equal masses $m = 0.64$ kg on the ends of a massless rod of length $L = 85$ cm. Calculate its rotational inertia about an axis one-fourth of the way from one end of the rod and perpendicular to it.

INTERPRET Here we have two discrete masses, so this problem is asking us to calculate the rotational inertia by summing over the individual masses.

DEVELOP Figure 10.11 is our sketch. We'll use Equation 10.12, $I = \sum m_i r_i^2$, to sum the two individual rotational inertias.

EVALUATE $I = \sum m_i r_i^2 = m\left(\frac{1}{4}L\right)^2 + m\left(\frac{3}{4}L\right)^2 = \frac{5}{8}mL^2$
$= \frac{5}{8}(0.64\text{ kg})(0.85\text{ m})^2 = 0.29\text{ kg·m}^2$

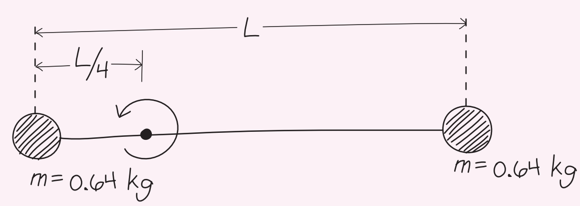


FIGURE 10.11 Our sketch for Example 10.4, showing rotation about an axis perpendicular to the page.

ASSESS Make sense? Even though there are two masses, our answer is less than the rotational inertia mL^2 of a single mass rotated about a rod of length L. That's because distance from the rotation axis is *squared*, so it contributes more in determining rotational inertia than does mass. ■

GOT IT? 10.2 Would the rotational inertia of the two-mass dumbbell in Example 10.4 (a) increase, (b) decrease, or (c) stay the same if the rotation axis were at the center of the rod? If it were at one end?

With continuous distributions of matter, we consider a large number of very small mass elements dm throughout the object, and sum the individual rotational inertias $r^2\,dm$ over the entire object (Fig. 10.12). In the limit of an arbitrarily large number of infinitesimally small mass elements, that sum becomes an integral:

$$I = \int r^2\,dm \qquad \left(\begin{array}{c}\text{rotational inertia,}\\ \text{continuous matter}\end{array}\right) \qquad (10.13)$$

where the limits of integration cover the entire object.

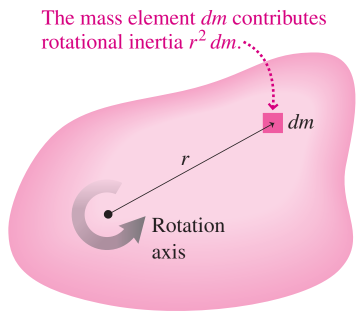


FIGURE 10.12 Rotational inertia can be found by integrating the rotational inertias $r^2\,dm$ of the mass elements making up an object.

EXAMPLE 10.5 Rotational Inertia by Integration: A Rod

Find the rotational inertia of a uniform, narrow rod of mass M and length L about an axis through its center and perpendicular to the rod.

INTERPRET The rod is a continuous distribution of matter, so calculating the rotational inertia is going to involve integration. We identify the rotation axis as being in the center of the rod.

DEVELOP Figure 10.13 shows the rod and rotation axis; we added a coordinate system with x-axis along the rod and the origin at the rotation axis. With a continuous distribution, Equation 10.13, $I = \int r^2\,dm$, applies. To develop a solution plan, we need to set up the integral in Equation 10.13. That equation may seem confusing because the integral contains both the geometric variable r and the mass element dm. How are they related? At this point you might want to review Tactics 9.1 (page 137); we'll follow its steps here. (1) We're first supposed to find a suitable mass element; here, with a one-dimensional rod, that can be a short section of the rod. We marked a typical mass element in Fig. 10.13. (2) This step is straightforward in this one-dimensional case; the length of the mass element is dx, signifying an infinitesimally short piece of the rod. (3) Now we form ratios to relate dx and the mass element dm. The total mass of the rod is M, and its total length is L. With the mass distributed uniformly, that means dx is the same fraction of L that dm is of M, or $dx/L = dm/M$. (4) We solve for the mass element: $dm = (M/L)\,dx$.

We're almost done. But the integral in Equation 10.13 contains r, and we've related dm and dx. No problem: On the one-dimensional rod, distances from the rotation axis are just the coordinates x. So r becomes x in our integral, and we have

$$I = \int r^2\,dm = \int_{-L/2}^{L/2} x^2 \frac{M}{L}\,dx$$

We chose the limits to include the entire rod; with the origin at the center, it runs from $-L/2$ to $L/2$.

EVALUATE The constants M and L come outside the integral, so we have

$$I = \int_{-L/2}^{L/2} x^2 \frac{M}{L}\,dx = \frac{M}{L}\int_{-L/2}^{L/2} x^2\,dx = \frac{M}{L}\frac{x^3}{3}\Bigg|_{-L/2}^{L/2} = \tfrac{1}{12}ML^2 \quad (10.14)$$

ASSESS Make sense? In Example 10.4 we found $I = \frac{5}{8}mL^2$ for a rod with two masses m on the ends. If you thought about GOT IT? 10.2, you probably realized that the rotational inertia would be $\frac{1}{2}mL^2$ for rotation about the rod's center. The total mass for that one was $M = 2m$, so in terms of total mass the rotational inertia about the center would be $I = \frac{1}{4}ML^2$—a lot larger than what we've found for the continuous rod. That's because much of the continuous rod's mass is close to the rotation axis, so it contributes less to the rotational inertia.

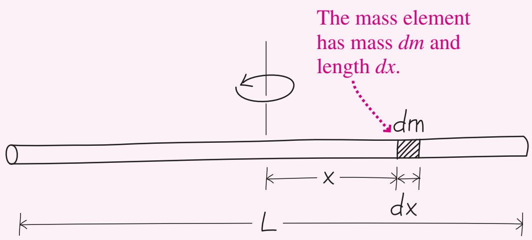


FIGURE 10.13 Our sketch of the uniform rod of Example 10.5.

EXAMPLE 10.6 Rotational Inertia by Integration: A Ring

Find the rotational inertia of a thin ring of radius R and mass M about the ring's axis.

INTERPRET This example is similar to Example 10.5, but the geometry has changed from a rod to a ring.

DEVELOP Figure 10.14 shows the ring with a mass element dm. All the mass elements in the ring are the same distance R from the rotation axis, so r in Equation 10.13 is the constant R, and the equation becomes

$$I = \int R^2\,dm = R^2 \int dm$$

where the integration is over the ring.

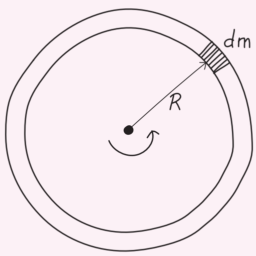


FIGURE 10.14 Our sketch of a thin ring, showing one mass element *dm*.

EVALUATE Because the sum of the mass elements over the ring is the total mass M, we find

$$I = MR^2 \qquad \text{(thin ring)} \qquad (10.15)$$

ASSESS The rotational inertia of the ring is the same as if all the mass were concentrated in one place a distance R from the rotation axis; the angular distribution of the mass about the axis doesn't matter. Notice, too, that it doesn't matter whether the ring is narrow like a loop of wire or long like a section of hollow pipe, as long as it's thin enough that all of it is essentially equidistant from the rotation axis (Fig. 10.15).

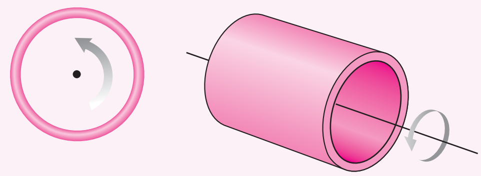


FIGURE 10.15 The rotational inertia is MR^2 for any thin ring, whether it's narrow like a wire loop or long like a pipe. ■

EXAMPLE 10.7 Rotational Inertia by Integration: A Disk

A disk of radius R and mass M has uniform density. Find the rotational inertia of the disk about an axis through its center and perpendicular to the disk.

INTERPRET Again we need to find the rotational inertia for a piece of continuous matter, this time a disk.

DEVELOP Because the disk is continuous, we need to integrate using Equation 10.13, $I = \int r^2\,dm$. We'll condense the strategy we applied in Example 10.5. The result of Example 10.6 suggests dividing the disk into rings, as shown in Fig. 10.16. Equation 10.15, with $M \rightarrow dm$, shows that a ring of radius r and mass dm contributes $r^2\,dm$ to the rotational inertia of the disk. Then the total inertia will be $I = \int_0^R r^2\,dm$, where we chose the limits to pick up contributions from all the mass elements on the disk. Again we need to relate r and dm. Think of "unwinding" the ring shown in Fig. 10.16; it becomes essentially a rectangle whose area dA is its circumference multiplied by its width: $dA = 2\pi r\,dr$. Next, we form ratios. The ring area dA is to the total disk area πR^2 as the ring mass dm is to the total mass M: $2\pi r\,dr/\pi R^2 = dm/M$. Solving for dm gives $dm = (2Mr/R^2)\,dr$.

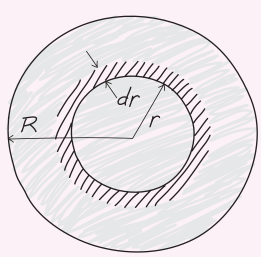


FIGURE 10.16 A disk may be divided into ring-shaped mass elements of mass *dm*, radius *r*, and width *dr*.

EVALUATE We now evaluate the integral:

$$I = \int_0^R r^2\,dm = \int_0^R r^2\left(\frac{2Mr}{R^2}\right)dr$$
$$= \frac{2M}{R^2}\int_0^R r^3\,dr = \frac{2M}{R^2}\,\frac{r^4}{4}\bigg|_0^R = \tfrac{1}{2}MR^2 \qquad \text{(disk)} \qquad (10.16)$$

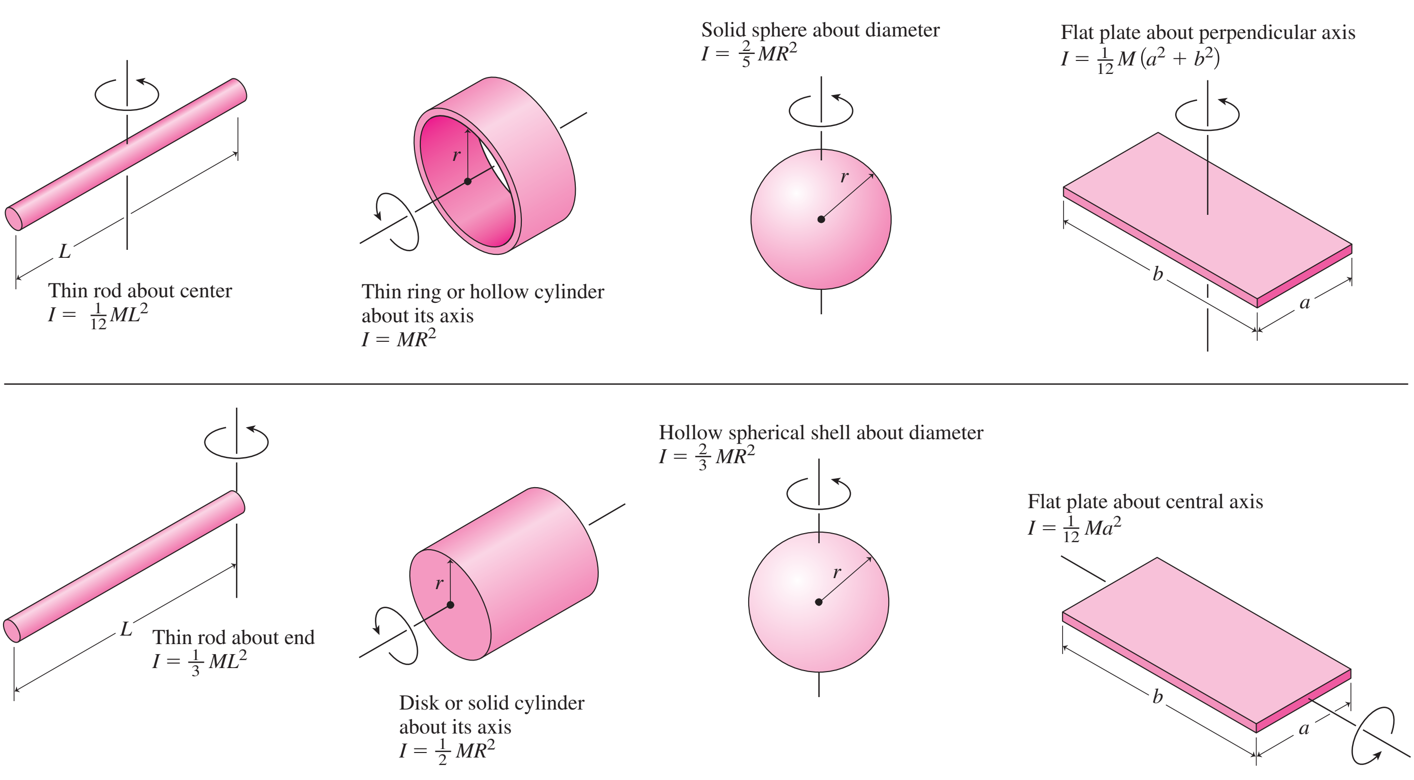
ASSESS Again, this result makes sense. In the disk, some of the mass is closer to the rotation axis, so the rotational inertia should be less than the value MR^2 for the ring.

✓TIP Constants and Variables

Note the different roles of R and r here. R represents a fixed quantity—the actual radius of the disk—and it's a constant that can go outside the integral. In contrast, r is the *variable of integration*, and it changes as we range from the disk's center to its edge, adding up all the infinitesimal mass elements. Because r is a variable over the region of integration, we can't take it outside the integral.

■

Table 10.2 Rotational Inertias



The rotational inertias of other shapes about various axes are found by integration as in these examples. Table 10.2 lists results for some common shapes. Note that more than one rotational inertia is listed for some shapes, since the rotational inertia depends on the rotation axis.

If we know the rotational inertia I_{cm} about an axis through the center of mass of a body, a useful relation called the **parallel-axis theorem** allows us to calculate the rotational inertia I through any parallel axis. The parallel-axis theorem states that

$$I = I_{cm} + Md^2 \tag{10.17}$$

where d is the distance from the center-of-mass axis to the parallel axis and M is the total mass of the object. Figure 10.17 shows the meaning of the parallel-axis theorem, which you can prove in Problem 78.

GOT IT? 10.3 Explain why the rotational inertia of the solid sphere in Table 10.2 is less than that of the spherical shell with the same radius and the same mass.

Rotational Dynamics

Knowing a body's rotational inertia, we can use the rotational analog of Newton's second law (Equation 10.11) to determine its behavior, just as we used Newton's law itself to analyze linear motion. Like the force in Newton's law, the torque in Equation 10.11 is the *net* external torque—the sum of all external torques acting on the body.

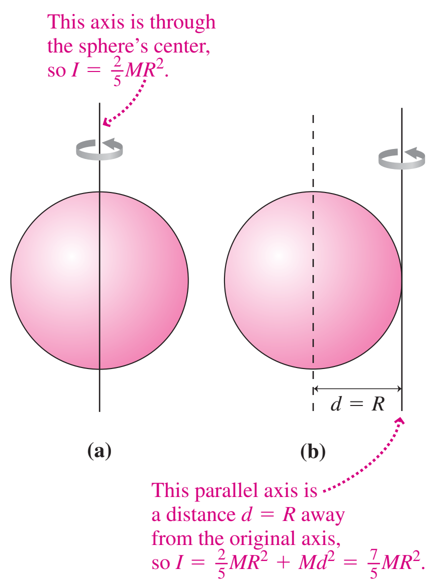


FIGURE 10.17 Meaning of the parallel-axis theorem.

EXAMPLE 10.8 Rotational Dynamics: De-Spinning a Satellite

A cylindrical satellite is 1.4 m in diameter, with its 940-kg mass distributed uniformly. The satellite is spinning at 10 rpm but must be stopped so that astronauts can make repairs. Two small gas jets, each with 20-N thrust, are mounted on opposite sides of the satellite and fire tangent to the satellite's rim. How long must the jets be fired in order to stop the satellite's rotation?

INTERPRET This is ultimately a problem about angular acceleration, but we're given the forces the jets exert. So it becomes a problem about calculating torque and then acceleration—that is, a problem in rotational dynamics using the rotational analog of Newton's law.

DEVELOP Figure 10.18 shows the situation. We're asked about the time, which we can get from the angular acceleration and initial angular speed. We can find the acceleration using the rotational analog of Newton's law, Equation 10.11, if we know both torque and rotational inertia. So here's our plan: (1) Find the satellite's rotational inertia from Table 10.2, treating it as a solid cylinder. (2) Find the torque due to the jets using Equation 10.10, $\tau = rF \sin\theta$. (3) Use the rotational analog of Newton's law—Equation 10.11, $\tau = I\alpha$—to find the angular acceleration. (4) Use the change in angular speed to get the time.

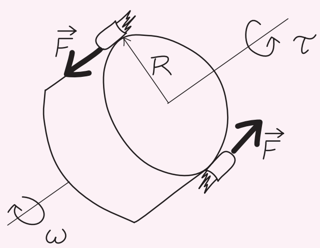


FIGURE 10.18 Torque from the jets stops the satellite's rotation.

EVALUATE Following our plan, (1) the rotational inertia from Table 10.2 is $I = \frac{1}{2}MR^2$. (2) With the jets tangent to the satellite, $\sin\theta$ in Equation 10.10 is 1, so each jet contributes a torque of magnitude RF, where R is the satellite radius and F the jet thrust force. With two jets, the net torque then has magnitude $\tau = 2RF$. (3) Equation 10.11 gives $\alpha = \tau/I = (2RF)/\left(\frac{1}{2}MR^2\right) = 4F/MR$. (4) We want this torque to drop the angular speed from $\omega_0 = 10$ rpm to zero, so the magnitude of the speed change is

$$\begin{aligned}\Delta\omega &= 10 \text{ rev/min} = (10 \text{ rev/min})(2\pi \text{ rad/rev})/(60 \text{ s/min}) \\ &= 1.05 \text{ rad/s}\end{aligned}$$

Since angular acceleration is $\alpha = \Delta\omega/\Delta t$, our final answer is

$$\begin{aligned}\Delta t &= \frac{\Delta\omega}{\alpha} = \frac{MR\,\Delta\omega}{4F} \\ &= \frac{(940\text{ kg})(0.70\text{ m})(1.05\text{ rad/s})}{(4)(20\text{ N})} = 8.6\text{ s}\end{aligned}$$

ASSESS Make sense? Yes: The thrust F appears in the denominator, showing that a larger force and hence torque will bring the satellite more rapidly to a halt. Larger M and R contribute to a larger rotational inertia, thus lengthening the stopping time—although a larger R also means a larger torque, an effect that reduces the R dependence from the R^2 that appears in the expression for rotational inertia. ■

A single problem can involve both rotational and linear motion with more than one object. The strategy for dealing with such problems is similar to the multiple-object strategy we developed in Chapter 5, where we identified the objects whose motions we were interested in, drew a free-body diagram for each, and then applied Newton's law separately to each object. We used the physical connections among the objects to relate quantities appearing in the separate Newton's law equations. Here we do the same thing, except that when an object is rotating, we use Equation 10.11, the rotational analog of Newton's law. Often the physical connection will entail relations between the force on an object in linear motion and the torque on a rotating object, as well as between the objects' linear and rotational accelerations.

EXAMPLE 10.9 Rotational and Linear Dynamics: Into the Well

A solid cylinder of mass M and radius R is mounted on a frictionless horizontal axle over a well, as shown in Fig. 10.19. A rope of negligible mass is wrapped around the cylinder and supports a bucket of mass m. Find an expression for the bucket's acceleration as it falls down the well shaft.

INTERPRET If it weren't connected to the cylinder, the bucket would fall with acceleration g. But the rope exerts an upward tension force $\vec{T}$ on the bucket, reducing its acceleration and at the same time exerting a torque on the cylinder. So we have a problem involving both rotational and linear dynamics. We identify the bucket and the cylinder as the objects of interest; the bucket is in linear motion while the cylinder rotates. The connection between them is the rope.

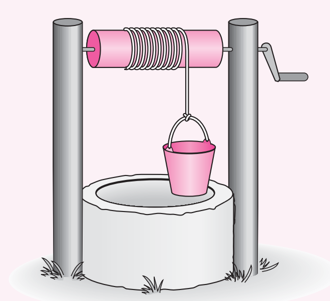


FIGURE 10.19 Example 10.9.

DEVELOP Figure 10.20 shows the free-body diagrams for the two objects; note that both involve the rope tension, $\vec{T}$. We chose the downward direction as positive in the bucket diagram and the clockwise direction as positive in the cylinder diagram. Now we're ready to write Newton's second law and its analog—Equation 10.11, $\tau = I\alpha$—for the two objects. Our plan is to formulate both equations and solve using the connection between them—physically the rope and mathematically the magnitude of the rope tension. We have to express the torque on the cylinder in terms of the tension force, using Equation 10.10, $\tau = rF \sin\theta$. We also need to relate the cylinder's angular acceleration to the bucket's linear acceleration, using Equation 10.5, $a_t = r\alpha$.

EVALUATE With the downward direction positive, Newton's second law for the bucket reads $F_{net} = mg - T = ma$. For the cylinder we have the rotational analog of Newton's second law: $\tau = I\alpha$. But here the torque is due to the rope tension, which exerts a force T at right angles to a line from the rotation axis and so produces torque RT. Then the Newton's law analog becomes $RT = I\alpha$. As the rope unwinds, the tangential acceleration of the cylinder's edge must be equal to the bucket's linear acceleration; thus, using Equation 10.5, we have $\alpha = a/R$, and the cylinder equation becomes $RT = Ia/R$ or $T = Ia/R^2$. But the cylinder's rotational inertia, from Table 10.2, is $I = \frac{1}{2}MR^2$, so $T = \frac{1}{2}Ma$. Using this result in the bucket equation gives $ma = mg - T = mg - \frac{1}{2}Ma$; solving for a, we then have

$$a = \frac{mg}{m + \frac{1}{2}M}$$

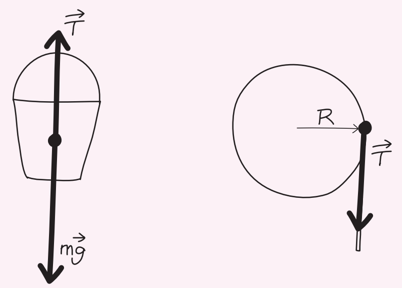
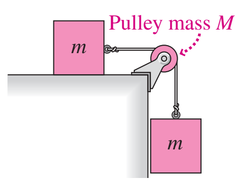


FIGURE 10.20 Our free-body diagrams for the bucket and cylinder.

ASSESS Make sense? If $M = 0$, there would be no rotational inertia and we would have $a = g$. Of course: With no torque needed to accelerate the cylinder, there would be no rope tension and the bucket would fall freely. But as the cylinder's mass M increases, the bucket's deceleration drops as greater torque and thus rope tension are needed to give the cylinder its rotational acceleration. You may be surprised to see that the cylinder radius doesn't appear in our answer. That, too, makes sense: The rotational inertia scales as R^2, but both the torque and the tangential acceleration scale with R. Since the cylinder's tangential acceleration is the same as the bucket's acceleration, the increases in torque and tangential acceleration cancel the effect of a greater rotational inertia. ■

GOT IT? 10.4 The figure shows two identical masses m connected by a string that passes over a frictionless pulley whose mass M is *not* negligible. One mass rests on a frictionless table; the other hangs vertically, as shown. Is the magnitude of the tension force in the vertical section of the string (a) greater than, (b) equal to, or (c) less than that in the horizontal section? Explain.



10.4 Rotational Energy

A rotating object clearly has kinetic energy. We define an object's **rotational kinetic energy** as the sum of the kinetic energies of all its mass elements, taken with respect to the rotation axis. Figure 10.21 shows that an individual mass element dm a distance r from the rotation axis has kinetic energy given by $dK = \frac{1}{2}(dm)(v^2) = \frac{1}{2}(dm)(\omega r)^2$. The rotational kinetic energy is given by summing—that is, integrating—over the entire object:

$$K_{rot} = \int dK = \int \tfrac{1}{2}(dm)(\omega r)^2 = \tfrac{1}{2}\omega^2 \int r^2\, dm$$

where we've taken ω^2 outside the integral because it's the same for every mass element in the rigid, rotating object. The remaining integral is just the rotational inertia I, so we have

$$K_{rot} = \tfrac{1}{2}I\omega^2 \quad \text{(rotational kinetic energy)} \qquad (10.18)$$

This formula should come as no surprise: Since I and ω are the rotational analogs of mass and speed, Equation 10.18 is the rotational equivalent of $K = \frac{1}{2}mv^2$.

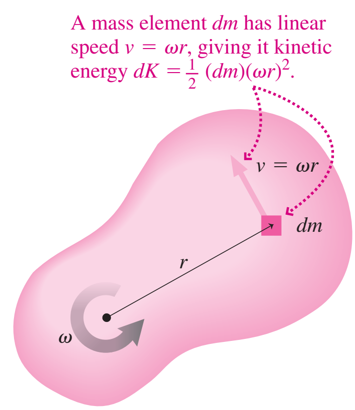


FIGURE 10.21 Kinetic energy of a mass element.

EXAMPLE 10.10 Rotational Energy: Flywheel Storage

A flywheel has a 135-kg solid cylindrical rotor with radius 30 cm and spins at 31,000 rpm. How much energy does it store?

INTERPRET We're being asked about kinetic energy stored in a rotating cylinder.

DEVELOP Equation 10.18, $K_{rot} = \frac{1}{2}I\omega^2$, gives the rotational energy. To use it, we need the rotational inertia from Table 10.2, and we need to convert the rotation rate in revolutions per minute to angular speed ω in radians per second.

EVALUATE Table 10.2 gives the rotational inertia, $I = \frac{1}{2}MR^2 = \left(\frac{1}{2}\right)(135\,\text{kg})(0.30\,\text{m})^2 = 6.1\ \text{kg}\cdot\text{m}^2$, and 31,000 rpm is equivalent to $(31{,}000\ \text{rev/min})(2\pi\ \text{rad/rev})/(60\ \text{s/min}) = 3246\ \text{rad/s}$. Then Equation 10.18 gives

$$K_{rot} = \tfrac{1}{2}I\omega^2 = \left(\tfrac{1}{2}\right)(6.1\ \text{kg}\cdot\text{m}^2)(3246\ \text{rad/s})^2 = 32\ \text{MJ}$$

ASSESS That 32 MJ is roughly the energy contained in a liter of gasoline. The advantages of the flywheel over a fuel or a chemical battery are more concentrated energy storage and greater efficiency at getting energy into and out of storage; see the Application below. ■

✓TIP When to Use Radians

We derived Equation 10.18, $K = \frac{1}{2}I\omega^2$, using Equation 10.3, $v = \omega r$. Since that equation works only with radian measure, the same is true of Equation 10.18.

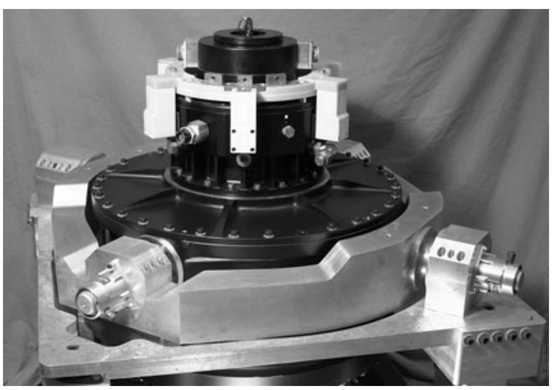
Energy and Work in Rotational Motion

In Section 6.3 we proved the work-energy theorem, which states that the change in an object's linear kinetic energy is equal to the net work done on the object. There the work was the product (or the integral, for a changing force) of the net force and the distance the object moves. Not surprisingly, there's an analogous relation for rotational motion: The change in an object's rotational kinetic energy is equal to the net work done on the object. Now the work is, analogously, the product (or the integral) of the torque and the angular displacement:

$$W = \int_{\theta_i}^{\theta_f} \tau\, d\theta = \Delta K_{rot} = \tfrac{1}{2}I\omega_f^2 - \tfrac{1}{2}I\omega_i^2 \quad \left(\begin{matrix}\text{work-energy theorem,}\\ \text{rotational motion}\end{matrix}\right) \qquad (10.19)$$

Here the subscripts refer to the initial and final states.

APPLICATION Flywheel Energy Storage



Flywheels provide an attractive alternative to batteries in applications requiring short bursts of power. Examples include acceleration and hill climbing in hybrid vehicles, industrial lifting equipment and amusement park rides, power management on the electric grid, and uninterruptible power supplies. Flywheel-based hybrid vehicles would achieve high efficiency by storing mechanical energy in the flywheel during braking, rather than dissipating it as heat in conventional brakes or even storing it in a chemical battery as in today's hybrids.

Equation 10.18 shows that the stored energy can be substantial, provided the flywheel has significant rotational inertia and angular speed—the latter being especially important because the energy scales as the *square* of the angular speed. Modern flywheels can supply tens of kilowatts of power for as long as a minute; unlike batteries, their output isn't reduced in cold weather. They achieve rotation rates of 30,000 rpm and more using advanced carbon composite materials that can withstand the forces needed to maintain the radial acceleration of magnitude $\omega^2 r$. Advanced flywheels spin in vacuum, using magnetic bearings to minimize friction. Some even use superconducting materials, which eliminate electrical losses that we'll examine in Chapter 26. The photo shows a high-speed flywheel used in a prototype hybrid bus operating in Austin, Texas. The flywheel helps the bus achieve 30% fuel savings.

EXAMPLE 10.11 Work and Rotational Energy: Balancing a Tire

An automobile wheel with tire has rotational inertia $2.7\ \text{kg}\cdot\text{m}^2$. What constant torque does a tire-balancing machine need to apply in order to spin this tire up from rest to 700 rpm in 25 revolutions?

INTERPRET The wheel's rotational kinetic energy changes as it spins up, so the machine must be doing work by applying a torque. Therefore, the concept behind this problem is the work-energy theorem for rotational motion.

DEVELOP The work-energy theorem of Equation 10.19 relates the work to the change in rotational kinetic energy:

$$W = \int_{\theta_i}^{\theta_f} \tau\, d\theta = \Delta K_{rot} = \tfrac{1}{2}I\omega_f^2 - \tfrac{1}{2}I\omega_i^2.$$

We're given the initial and final angular velocities, although we have to convert them to radians per second. With constant torque, the integral in Equation 10.19 becomes the product $\tau\,\Delta\theta$, so we can solve for the torque.

EVALUATE The initial angular speed ω_i is zero, and the final speed $\omega_f = (700\text{ rev/min})(2\pi\text{ rad/rev})/(60\text{ s/min}) = 73.3\text{ rad/s}$. The angular displacement $\Delta\theta$ is $(25\text{ rev})(2\pi\text{ rad/rev}) = 157\text{ rad}$. Then Equation 10.19 becomes $W = \tau\,\Delta\theta = \frac{1}{2}I\omega_f^2$, which gives

$$\tau = \frac{\frac{1}{2}I\omega_f^2}{\Delta\theta} = \frac{\left(\frac{1}{2}\right)(2.7\text{ kg}\cdot\text{m}^2)(73.3\text{ rad/s})^2}{157\text{ rad}} = 46\text{ N}\cdot\text{m}$$

ASSESS If this torque results from a force applied at the rim of a typical 40-cm-radius tire, then the magnitude of the force would be just over 100 N, about the weight of a 10-kg mass and thus a reasonable value. ■

10.5 Rolling Motion

A rolling object exhibits both rotational motion and translational motion—the motion of the whole object from place to place. How much kinetic energy is associated with each?

In Section 9.3, we found that the kinetic energy of a composite object comprises two terms: the kinetic energy of the center of mass and the internal kinetic energy relative to the center of mass: $K = K_{cm} + K_{internal}$. A wheel of mass M moving with speed v has center-of-mass kinetic energy $K_{cm} = \frac{1}{2}Mv^2$. In the center-of-mass frame, the wheel is simply rotating with angular speed ω about the center of mass, so its internal kinetic energy is $K_{internal} = \frac{1}{2}I_{cm}\omega^2$, where the rotational inertia is taken about the center of mass. We now sum K_{cm} and $K_{internal}$ to get the total kinetic energy:

$$K_{total} = \tfrac{1}{2}Mv^2 + \tfrac{1}{2}I_{cm}\omega^2 \qquad (10.20)$$

When a wheel is *rolling*—moving without slipping against the ground—its translational speed v and angular speed ω about its center of mass are related. Imagine a wheel that rolls half a revolution and therefore moves horizontally half its circumference (Fig. 10.22). Then the wheel's angular speed is the angular displacement $\Delta\theta$, here half a revolution, or π radians, divided by the time Δt: $\omega = \pi/\Delta t$. Its translational speed is the actual distance the wheel travels divided by the same time interval. But we've just argued that the wheel travels half a circumference, or πR, where R is its radius. So its translational speed is $v = \pi R/\Delta t$. Comparing our expressions for v and ω, we see that

$$v = \omega R \qquad \text{(rolling motion)} \qquad (10.21)$$

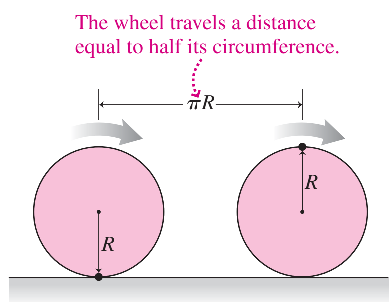


FIGURE 10.22 A rolling wheel turns through half a revolution.

Equation 10.21 looks deceptively like Equation 10.3. But it says more. In Equation 10.3, $v = \omega r$, v is the linear speed of a point a distance r from the center of a rotating object. In Equation 10.21, v is the translational speed of the whole object and R is its radius. The two equations look similar because, as our argument leading to Equation 10.21 shows, an object that rolls without slipping moves with respect to the ground at the same rate that a point on its rim moves in the center-of-mass frame.

Our description of rolling motion leads to a point you may at first find absurd: In a rolling wheel, the point in contact with the ground is, instantaneously, at rest! Figure 10.23 shows how this surprising situation comes about.

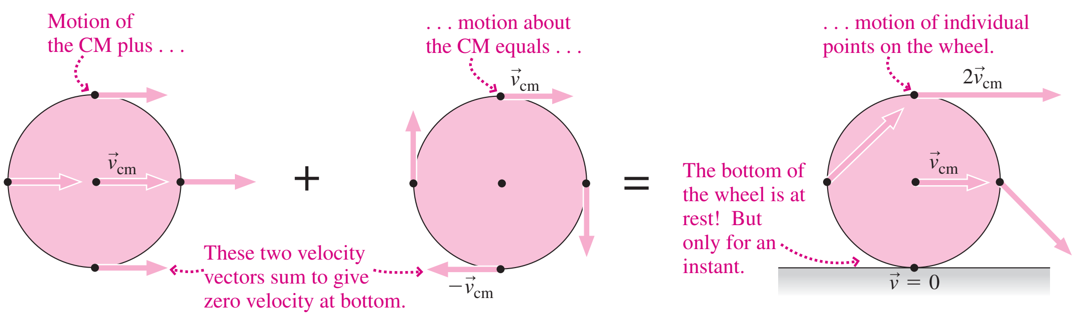


FIGURE 10.23 Motion of a rolling wheel, decomposed into translation of the entire wheel plus rotation about the center of mass.

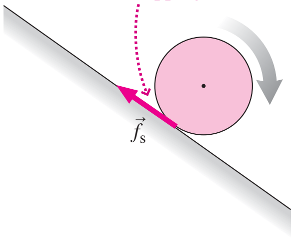


FIGURE 10.24 Rolling down a slope.

Why would an object roll without slipping? The answer is friction. On an icy slope, a wheel just slides down without rolling. Normally, though, the force of static friction keeps it from sliding. Instead, it rolls (Fig. 10.24). Because the contact point is at rest, the frictional force does no work and therefore mechanical energy is conserved. This lets us use the conservation-of-energy principle to analyze rolling objects.

EXAMPLE 10.12 Energy Conservation: Rolling Downhill

A solid ball of mass M and radius R starts from rest and rolls down a hill. Its center of mass drops a total distance h. Find the ball's speed at the bottom of the hill.

INTERPRET This is similar to conservation-of-energy problems from Chapter 7, but now we identify two types of kinetic energy: translational and rotational. The ball starts on the slope with some gravitational potential energy, which ends up as kinetic energy at the bottom. The frictional force that keeps the ball from slipping does no work, so we can apply conservation of energy.

DEVELOP Figure 10.25 shows the situation, including bar graphs showing the distribution of energy in the ball's initial and final states. We've determined that conservation of energy holds, so $K_0 + U_0 = K + U$. Here $K_0 = 0$ and, if we take the zero of potential energy at the bottom, then $U_0 = Mgh$ and $U = 0$. Finally, K consists of both translational and rotational kinetic energy as expressed in Equation 10.20, $K_{\text{total}} = \frac{1}{2}Mv^2 + \frac{1}{2}I\omega^2$. Our plan is to use this expression in the conservation-of-energy statement and solve for v. It looks like there's an extra variable, ω, that we don't know. But the ball isn't slipping, so Equation 10.21 holds and gives $\omega = v/R$. Then conservation of energy becomes

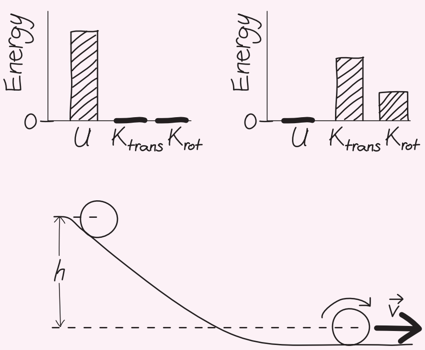


FIGURE 10.25 How fast is the ball moving at the bottom of the hill?

$$Mgh = \tfrac{1}{2}Mv^2 + \tfrac{1}{2}I\omega^2 = \tfrac{1}{2}Mv^2 + \tfrac{1}{2}\left(\tfrac{2}{5}MR^2\right)\left(\frac{v}{R}\right)^2 = \tfrac{7}{10}Mv^2$$

where we found the rotational inertia of a solid sphere, $\frac{2}{5}MR^2$, from Table 10.2.

EVALUATE Solving for v gives our answer:

$$v = \sqrt{\frac{10}{7}gh}$$

ASSESS This result is less than the speed $v = \sqrt{2gh}$ for an object that slides down a frictionless incline. Make sense? Yes: Some of the energy the rolling object gains goes into rotation, leaving less for translational motion. As often happens with gravitational problems, mass doesn't matter. Neither does radius: That factor $\frac{7}{10}$ results from the distribution of mass that gives the sphere its particular rotational inertia and would be the same for all spheres regardless of radius or mass. ■

CONCEPTUAL EXAMPLE 10.1 A Rolling Race

A solid ball and a hollow ball roll without slipping down a ramp. Which reaches the bottom first?

EVALUATE Example 10.12 shows that when a ball rolls down a slope, some of its potential energy gets converted into rotational kinetic energy—leaving less for translational kinetic energy. As a result, it moves more slowly, and therefore takes more time, than an object that slides without rolling. Here we want to compare two rolling objects—the solid ball treated in Example 10.12 and a hollow one. With its mass concentrated at its surface, far from the rotation axis, the hollow ball has greater rotational inertia. Thus more of its energy goes into rotation, meaning its translational speed is lower, so it reaches the bottom later.

ASSESS Make sense? Yes: Energy is conserved for both balls, but for the hollow ball more of that energy is in rotation and less in translation. As Example 10.12 shows, neither the mass nor the radius of a ball affects its speed; all that matters is its mass distribution and hence its rotational inertia.

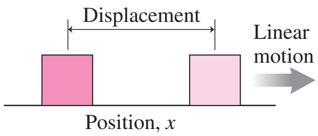
MAKING THE CONNECTION Compare the final speeds of the two balls in this example.

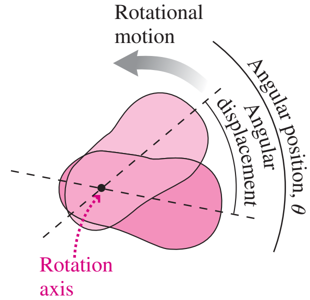
EVALUATE Example 10.12 gives $\sqrt{10gh/7}$ for the speed of the solid ball after it's rolled down a vertical drop h. Substituting the hollow ball's rotational inertia, $I = \frac{2}{3}MR^2$ from Table 10.2, in the calculation of Example 10.12 gives $v = \sqrt{6gh/5}$. So the solid ball is faster by a factor $\sqrt{10/7}/\sqrt{6/5} \simeq 1.1$.

CHAPTER 10 SUMMARY

Big Picture

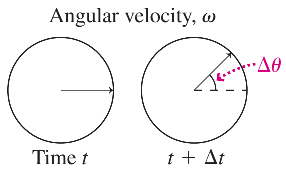
The big idea of this chapter is rotational motion, quantified as the rate of change of angular position of any point on a rotating object. All the quantities used to describe linear motion have analogs in rotational motion. The analogs of force, mass, and acceleration are, respectively, torque, rotational inertia, and angular acceleration—and together they obey the rotational analog of Newton's second law.



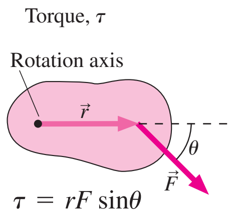


Key Concepts and Equations

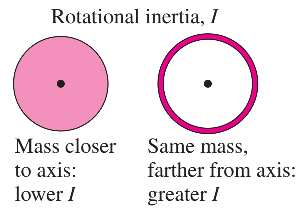
The defining relations for rotational quantities are analogous to those for linear quantities, as is the statement of Newton's second law for rotational motion. Key concepts include angular velocity and acceleration, torque, and rotational inertia.



$$\bar{\omega} = \frac{\Delta\theta}{\Delta t}$$



$$\tau = rF\sin\theta$$



$$I = \sum m_i r_i^2 \longrightarrow \int r^2\, dm$$

Discrete masses — Continuous matter

This table summarizes the analogies between linear and rotational quantities, along with quantitative relations that link rotational and linear quantities. Many of these relations require that angles be measured in radians, and most require explicit specification of a rotation axis.

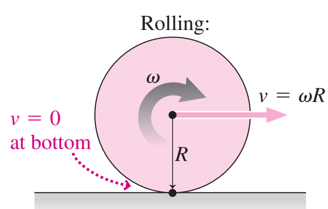
Linear Quantity or Equation	Angular Quantity or Equation	Relation Between Linear and Angular Quantities
Position x	Angular position θ	
Speed $v = dx/dt$	Angular speed $\omega = d\theta/dt$	$v = \omega r$
Acceleration a	Angular acceleration α	$a_t = \alpha r$
Mass m	Rotational inertia I	$I = \int r^2\, dm$
Force F	Torque τ	$\tau = rF\sin\theta$
Kinetic energy $K_{\text{trans}} = \frac{1}{2}mv^2$	Kinetic energy $K_{\text{rot}} = \frac{1}{2}I\omega^2$	
Newton's second law (constant mass or rotational inertia):		
$F = ma$	$\tau = I\alpha$	

Applications

Constant angular acceleration: When angular acceleration is constant, equations analogous to those of Chapter 2 apply.

Equations for Constant Linear Acceleration		Equations for Constant Angular Acceleration	
$\bar{v} = \frac{1}{2}(v_0 + v)$	(2.8)	$\bar{\omega} = \frac{1}{2}(\omega_0 + \omega)$	(10.6)
$v = v_0 + at$	(2.7)	$\omega = \omega_0 + \alpha t$	(10.7)
$x = x_0 + v_0 t + \frac{1}{2}at^2$	(2.10)	$\theta = \theta_0 + \omega_0 t + \frac{1}{2}\alpha t^2$	(10.8)
$v^2 = v_0^2 + 2a(x - x_0)$	(2.11)	$\omega^2 = \omega_0^2 + 2\alpha(\theta - \theta_0)$	(10.9)

Rolling motion: When an object of radius R rolls without slipping, the point in contact with the ground is instantaneously at rest. In this case the object's translational and rotational speeds are related by $v = \omega R$. The object's kinetic energy is shared among translational kinetic energy $\frac{1}{2}Mv^2$ and rotational kinetic energy $\frac{1}{2}I\omega^2$, with the division between these forms dependent on the rotational inertia.



For Thought and Discussion

1. Do all points on a rigid, rotating object have the same angular velocity? Linear speed? Radial acceleration?
2. A point on the rim of a rotating wheel has nonzero acceleration, since it's moving in a circular path. Does it necessarily follow that the wheel is undergoing angular acceleration?
3. Why doesn't it make sense to talk about a body's rotational inertia unless you specify a rotation axis?
4. Two forces act on an object, but the net force is zero. Must the net torque be zero? If so, why? If not, give a counterexample.
5. Is it possible to apply a counterclockwise torque to an object that's rotating clockwise? If so, how will the object's motion change? If not, why not?
6. A solid sphere and a hollow sphere of the same mass and radius are rolling along level ground. If they have the same total kinetic energy, which is moving faster?
7. A solid cylinder and a hollow cylinder of the same mass and radius are rolling along level ground at the same speed. Which has more kinetic energy?
8. A circular saw takes a long time to stop rotating after the power is turned off. Without the saw blade mounted, the motor stops much more quickly. Why?
9. A solid sphere and a solid cube have the same mass, and the side of the cube is equal to the diameter of the sphere. The cube's rotation axis is perpendicular to two of its faces. Which has greater rotational inertia about an axis through the center of mass?
10. BIO The lower part of a horse's leg contains essentially no muscle. How does this help the horse to run fast? Explain in terms of rotational inertia.
11. Given a fixed amount of a material, what shape should you make a flywheel so it will store the most energy at a given angular speed?
12. A ball starts from rest and rolls without slipping down a slope, then starts up a frictionless slope (Fig. 10.26). Compare its maximum height on the frictionless slope with its starting height on the first slope.

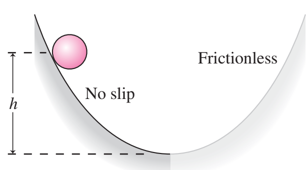


FIGURE 10.26 For Thought and Discussion 12, Problem 64

Exercises and Problems

Exercises

Section 10.1 Angular Velocity and Acceleration

13. Determine the angular speed, in rad/s, of (a) Earth about its axis; (b) the minute hand of a clock; (c) the hour hand of a clock; and (d) an eggbeater turning at 300 rpm.
14. What's the linear speed of a point (a) on Earth's equator and (b) at your latitude?
15. Express each of the following in radians per second: (a) 720 rpm; (b) 50°/h; (c) 1000 rev/s; (d) 1 rev/year (Earth's angular speed in its orbit).
16. A 25-cm-diameter circular saw blade spins at 3500 rpm. How fast would you have to push a straight hand saw to have the teeth move through the wood at the same rate as the circular saw teeth?
17. A compact disc's rotation varies from about 200 rpm to 500 rpm. If the disc plays for 74 min, what's its average angular acceleration in (a) rpm/s and (b) rad/s^2?
18. During startup, a power plant's turbine accelerates from rest at 0.52 rad/s^2. (a) How long does it take to reach its 3600-rpm operating speed? (b) How many revolutions does it make during this time?
19. A merry-go-round starts from rest and accelerates with angular acceleration 0.010 rad/s^2 for 14 s. (a) How many revolutions does it make during this time? (b) What's its average angular speed?

Section 10.2 Torque

20. A 320-N frictional force acts on the rim of a 1.0-m-diameter wheel to oppose its rotational motion. Find the torque about the wheel's central axis.
21. A 110-N·m torque is needed to start a revolving door rotating. If a child can push with a maximum force of 90 N, how far from the door's rotation axis must she apply this force?
22. A car tune-up manual calls for tightening the spark plugs to a torque of 35.0 N·m. To achieve this torque, with what force must you pull on the end of a 24.0-cm-long wrench if you pull (a) at a right angle to the wrench shaft and (b) at 110° to the wrench shaft?
23. A 55-g mouse runs out to the end of the 17-cm-long minute hand of a grandfather clock when the clock reads 10 past the hour. What torque does the mouse's weight exert about the rotation axis of the clock hand?
24. You have your bicycle upside-down for repairs. The front wheel is free to rotate and is perfectly balanced except for the 25-g valve stem. If the valve stem is 32 cm from the rotation axis and at 24° below the horizontal, what's the resulting torque about the wheel's axis?

Section 10.3 Rotational Inertia and the Analog of Newton's Law

25. Four equal masses m are located at the corners of a square of side L, connected by essentially massless rods. Find the rotational inertia of this system about an axis (a) that coincides with one side and (b) that bisects two opposite sides.
26. The shaft connecting a power plant's turbine and electric generator is a solid cylinder of mass 6.8 Mg and diameter 85 cm. Find its rotational inertia.
27. The chamber of a rock-tumbling machine is a hollow cylinder with mass 65 g and radius 7.1 cm. The chamber is closed by end caps in the form of uniform circular disks, each of mass 22 g. Find (a) the rotational inertia of the chamber about its central axis and (b) the torque needed to give the chamber an angular acceleration of 3.4 rad/s^2.
28. A wheel's diameter is 92 cm, and its rotational inertia is 7.8 kg·m^2. (a) What's the minimum mass it could have? (b) How could it have more mass?
29. Three equal masses m are located at the vertices of an equilateral triangle of side L, connected by rods of negligible mass. Find expressions for the rotational inertia of this object (a) about an axis through the center of the triangle and perpendicular to its plane and (b) about an axis that passes through one vertex and the midpoint of the opposite side.

30. (a) Estimate Earth's rotational inertia, assuming it to be a uniform solid sphere. (b) What torque applied to Earth would cause the length of a day to change by 1 second every century?
31. A neutron star is an extremely dense, rapidly spinning object that results from the collapse of a massive star at the end of its life. A neutron star of 1.8 times the Sun's mass has an approximately uniform density of 1×10^{18} kg/m^3. (a) What's its rotational inertia? (b) The neutron star's spin rate slowly decreases as a result of torque associated with magnetic forces. If the spin-down rate is 5×10^{-5} rad/s^2, what's the magnetic torque?
32. A 108-g Frisbee is 24 cm in diameter and has half its mass spread uniformly in the disk and the other half concentrated in the rim. (a) What's the Frisbee's rotational inertia? (b) With a quarter-turn flick of the wrist, a student sets the Frisbee rotating at 550 rpm. What's the magnitude of the torque, assumed constant, that the student applied?
33. At the MIT Magnet Laboratory, energy is stored in huge solid flywheels of mass 7.7×10^4 kg and radius 2.4 m. The flywheels ride on shafts 41 cm in diameter. If a frictional force of 34 kN acts tangentially on the shaft, how long will it take the flywheel to come to a stop from its usual 360-rpm rotation rate?

Section 10.4 Rotational Energy

34. A 25-cm-diameter circular saw blade has mass 0.85 kg, distributed uniformly in a disk. (a) What's its rotational kinetic energy at 3500 rpm? (b) What average power must be applied to bring the blade from rest to 3500 rpm in 3.2 s?
35. Humankind uses energy at the rate of about 15 TW. If we found a way to extract this energy from Earth's rotation, how long would it take before the length of the day increased by 1 minute?
36. A 150-g baseball is pitched at 33 m/s spinning at 42 rad/s. You can treat the baseball as a uniform solid sphere of radius 3.7 cm. What fraction of its kinetic energy is rotational?
37. (a) Find the energy stored in the flywheel of Exercise 33 when it's rotating at 360 rpm. (b) The wheel is attached to an electric generator and the rotation rate drops from 360 rpm to 300 rpm in 3.0 s. What's the average power output?

Section 10.5 Rolling Motion

38. A solid 2.4-kg sphere is rolling at 5.0 m/s. Find (a) its translational kinetic energy and (b) its rotational kinetic energy.
39. What fraction of a solid disk's kinetic energy is rotational if it's rolling without slipping?
40. A rolling ball has total kinetic energy 100 J, 40 J of which is rotational energy. Is the ball solid or hollow?

Problems

41. A wheel turns through 2.0 revolutions while accelerating from rest at 18 rpm/s. (a) What's its final angular speed? (b) How long does it take?
42. You're an engineer designing kitchen appliances, and you're working on a two-speed food blender, with 3600 rpm and 1800 rpm settings. Specs call for the blender to make no more than 60 revolutions while it's switching from high to low speed. If it takes 1.4 s to make the transition, does it meet its specs?
43. BIO An eagle with 2.1-m wingspan flaps its wings 20 times per minute, each stroke extending from 45° above the horizontal to 45° below. Downward and upward strokes take the same time. On a given downstroke, what's (a) the average angular velocity of the wing and (b) the average tangential velocity of the wingtip?
44. A compact disc (CD) player varies the rotation rate of the disc in order to keep the part of the disc from which information is being read moving at a constant linear speed of 1.30 m/s. Compare the rotation rates of a 12.0-cm-diameter CD when information is being read (a) from its outer edge and (b) from a point 3.75 cm from the center. Give your answers in rad/s and rpm.
45. You rev your car's engine and watch the tachometer climb steadily from 1200 rpm to 5500 rpm in 2.7 s. What are (a) the engine's angular acceleration and (b) the tangential acceleration of a point on the edge of the engine's 3.5-cm-diameter crankshaft? (c) How many revolutions does the engine make during this time?
46. A circular saw spins at 5800 rpm, and its electronic brake is supposed to stop it in less than 2 s. As a quality-control specialist, you're testing saws with a device that counts the number of blade revolutions. A particular saw turns 75 revolutions while stopping. Does it meet its specs?
47. BIO Full-circle rotation is common in mechanical systems, but less evident in biology. Yet many single-celled organisms are propelled by spinning, tail-like *flagella*. The flagellum of the bacterium *E. coli* spins at some 600 rad/s, propelling the bacterium at speeds around 25 μm/s. How many revolutions does *E. coli*'s flagellum make as the bacterium crosses a microscope's 150-μm-wide field of view?
48. A pulley 12 cm in diameter is free to rotate about a horizontal axle. A 220-g mass and a 470-g mass are tied to either end of a massless string, and the string is hung over the pulley. Assuming the string doesn't slip, what torque must be applied to keep the pulley from rotating?
49. A square frame is made from four thin rods, each of length L and mass m. Calculate its rotational inertia about the three axes shown in Fig. 10.27.

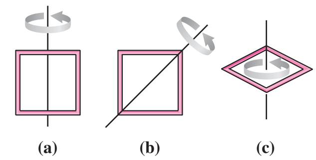


FIGURE 10.27 Problem 49

50. Use integration to show that the rotational inertia of a thick ring of mass M and inner and outer radii R_1 and R_2 is given by $\frac{1}{2}M(R_1^2 + R_2^2)$. (*Hint*: See Example 10.7.)
51. A uniform rectangular flat plate has mass M and dimensions a by b. Use the parallel-axis theorem in conjunction with Table 10.2 to show that its rotational inertia about the side of length b is $\frac{1}{3}Ma^2$.
52. Each propeller on a King Air twin-engine airplane consists of three blades, each of mass 10 kg and length 125 cm. The blades may be treated approximately as uniform, thin rods. (a) What's the propeller's rotational inertia? (b) If the plane's engine develops a torque of 2.7 kN·m, how long will it take to spin up the propeller from 1400 rpm to 1900 rpm?
53. BIO The cellular motor driving the flagellum in *E. coli* (see Problem 47) exerts a typical torque of 400 pN·nm on the flagellum. If this torque results from a force applied tangentially to the outside of the 12-nm-radius flagellum, what's the magnitude of that force?
54. Verify by direct integration Table 10.2's entry for the rotational inertia of a flat plate about a central axis. (*Hint*: Divide the plate into strips parallel to the axis.)
55. You're an astronaut in the first crew of a new space station. The station is shaped like a wheel 22 m in diameter, with essentially all its 5×10^5-kg mass at the rim. When the crew arrives, it will be

set rotating at a rate that requires an object at the rim to have radial acceleration g, thereby simulating Earth's surface gravity. This will be accomplished using two small rockets, each with 100-N thrust, mounted on the station's rim. Your job is to determine how long to fire the rockets and the number of revolutions the station will make during the firing.

56. A motor is connected to a solid cylindrical drum with diameter 1.2 m and mass 51 kg. A massless rope is attached to the drum and tied at the other end to a 38-kg weight, so the rope will wind onto the drum as it turns. What torque must the motor apply if the weight is to be lifted with acceleration 1.1 m/s^2?
57. A 2.4-kg block rests on a slope and is attached by a string of negligible mass to a solid drum of mass 0.85 kg and radius 5.0 cm, as shown in Fig. 10.28. When released, the block accelerates down the slope at 1.6 m/s^2. Find the coefficient of friction between block and slope.

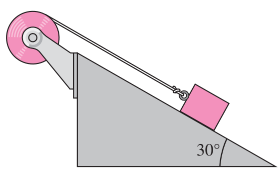


FIGURE 10.28 Problem 57

58. You've got your bicycle upside-down for repairs, with its 66-cm-diameter wheel spinning freely at 230 rpm. The wheel's mass is 1.9 kg, concentrated mostly at the rim. You hold a wrench against the tire for 3.1 s, applying a 2.7-N normal force. If the coefficient of friction between wrench and tire is 0.46, what's the final angular speed of the wheel?
59. A potter's wheel is a stone disk 90 cm in diameter with mass 120 kg. If the potter's foot pushes at the outer edge of the initially stationary wheel with a 75-N force for one-eighth of a revolution, what will be the final speed?
60. A ship's anchor weighs 5000 N. Its cable passes over a roller of negligible mass and is wound around a hollow cylindrical drum of mass 380 kg and radius 1.1 m, mounted on a frictionless axle. The anchor is released and drops 16 m to the water. Use energy considerations to determine the drum's rotation rate when the anchor hits the water. Neglect the cable's mass.
61. Starting from rest, a hollow ball rolls down a ramp inclined at angle θ to the horizontal. Find an expression for its speed after it's gone a distance d along the incline.
62. A hollow ball rolls along a horizontal surface at 3.7 m/s when it encounters an upward incline. If it rolls without slipping up the incline, what maximum height will it reach?
63. As an automotive engineer, you're charged with improving the fuel economy of your company's vehicles. You realize that the rotational kinetic energy of a car's wheels is a significant factor in fuel consumption, and you set out to lower it. For a typical car, the wheels' rotational energy is 40% of their translational kinetic energy. You propose a redesigned wheel with the same radius but 10% lower rotational inertia and 20% less mass. What do you report for the decrease in the wheel's total kinetic energy at a given speed?
64. A solid ball of mass M and radius R starts at rest at height h above the bottom of the path in Fig. 10.26. It rolls without slipping down the left side. The right side of the path, starting at the bottom, is frictionless. To what height does the ball rise on the right?
65. A disk of radius R has an initial mass M. Then a hole of radius $R/4$ is drilled, with its edge at the disk center (Fig. 10.29). Find the new rotational inertia about the central axis. (*Hint*: Find the rotational inertia of the missing piece, and subtract it from that of the whole disk. You'll find the parallel-axis theorem helpful.)

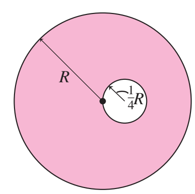


FIGURE 10.29 Problems 65 and 70

66. A 50-kg mass is tied to a massless rope wrapped around a solid cylindrical drum, mounted on a frictionless horizontal axle. When the mass is released, it falls with acceleration $a = 3.7$ m/s^2. Find (a) the rope tension and (b) the drum's mass.
67. Each wheel of a 320-kg motorcycle is 52 cm in diameter and has rotational inertia 2.1 kg·m^2. The cycle and its 75-kg rider are coasting at 85 km/h on a flat road when they encounter a hill. If the cycle rolls up the hill with no applied power and no significant internal friction, what vertical height will it reach?
68. A solid marble starts from rest and rolls without slipping on the loop-the-loop track in Fig. 10.30. Find the minimum starting height from which the marble will remain on the track through the loop. Assume the marble's radius is small compared with R.

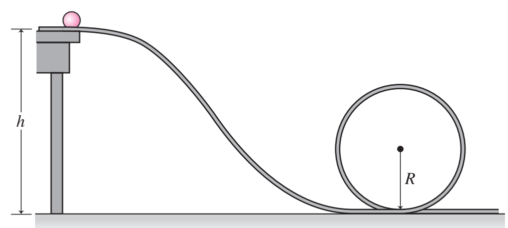


FIGURE 10.30 Problem 68

69. A disk of radius R and thickness w has a mass density that increases from the center outward, given by $\rho = \rho_0 r/R$, where r is the distance from the disk axis. Calculate (a) the disk's total mass M and (b) its rotational inertia about its axis in terms of M and R. Compare with the results for a solid disk of uniform density and for a ring.
70. The disk in Fig. 10.29 is rotating freely about a frictionless horizontal axle. Since the disk is unbalanced, its angular speed varies as it rotates. If the maximum angular speed is ω_{max}, find an expression for the minimum speed. (*Hint*: How does potential energy change as the wheel rotates?)
71. Consider the rotational inertia of a thin, flat object about an axis perpendicular to the plane of the object. Show that this is equal to the sum of the rotational inertias about two perpendicular axes in the plane of the object, passing through the given axis. (This is called the *perpendicular-axis theorem*.)
72. Use the perpendicular-axis theorem in Problem 71 (a) to verify the entry in Table 10.2 for a flat rectangular plate about a perpendicular axis and (b) to find the rotational inertia of a uniform thin disk of radius R and mass M about an axis along a diameter.
73. Calculate the rotational inertia of a uniform right circular cone of mass M, height h, and base radius R about its axis.

74. Show that the rotational inertia of a uniform solid spheroid about its axis of revolution is $\frac{2}{5}MR^2$, where M is its mass and R is the semi-axis perpendicular to the rotation axis. Why does this result look the same for both a prolate or oblate spheroid and a sphere?
75. A thin rod of length L and mass M is free to pivot about one end. If it makes an angle θ with the horizontal, find the torque due to gravity about the pivot. (*Hint*: Integrate the torques on the mass elements composing the rod.)
76. The local historical society has asked your assistance in writing the interpretive material for a display featuring an old steam locomotive. You have information on the torque in a flywheel but need to know the force applied by means of an attached horizontal rod. The rod joins the wheel with a flexible connection 95 cm from the wheel's axis. The maximum torque the rod produces on the flywheel is 10.1 kN·m. What force does the rod apply?
77. You're skeptical about a new hybrid car that stores energy in a flywheel. The manufacturer claims the flywheel stores 12 MJ of energy and can supply 40 kW of power for 5 minutes. You dig deeper and find that the flywheel is a 39-cm-diameter ring with mass 48 kg that rotates at 30,000 rpm. Are the specs correct?
78. Figure 10.31 shows an object of mass M with one axis through its center of mass and a parallel axis through an arbitrary point A. Both axes are perpendicular to the page. The figure shows an arbitrary mass element dm and vectors connecting the center of mass, A, and dm. (a) Use the law of cosines (Appendix A) to show that $r^2 = r_{cm}^2 + h^2 - 2\vec{h} \cdot \vec{r}_{cm}$. (b) Use this result in $I = \int r^2\, dm$ to calculate the object's rotational inertia about the axis through A. Each term in your expression for r^2 leads to a separate integral. Identify one as the rotational inertia about the CM, another as the quantity Mh^2, and argue that the third is zero. Your result is a statement of the parallel-axis theorem (Equation 10.17).

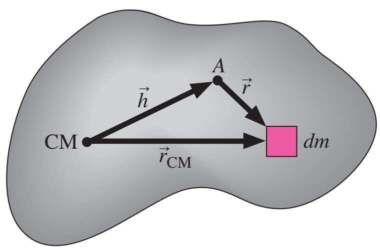


FIGURE 10.31 Problem 78

Passage Problems

Centrifuges are widely used in biology and medicine to separate cells and other particles from liquid suspensions. Figure 10.32 shows top and side views of two centrifuge designs. In both designs, the round holes are for tubes holding samples to be separated; the side views show two tubes in place. The total mass and radius of the rotating structure are the same for both, the sample-hole tubes are at the same radius, and the sample tubes are identical.

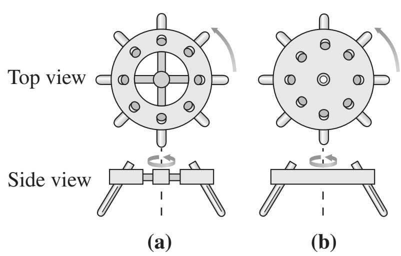


FIGURE 10.32 Two centrifuge designs, shown from the top and the side (Passage Problems 79–83)

79. Which design has greater rotational inertia?
    a. design A
    b. design B
    c. Both have the same rotational inertia.
80. If both centrifuges are made thicker in the vertical direction, without changing their masses or mass distribution, their rotational inertias will
    a. remain the same.
    b. increase.
    c. decrease.
81. If the sample tubes are made longer, the rotational inertia of the centrifuges with sample tubes inserted will
    a. remain the same.
    b. increase.
    c. decrease.
82. While the centrifuges are spinning, the net force on samples in the tubes is
    a. outward.
    b. inward.
    c. zero.
83. If a centrifuge's radius and mass are both doubled without otherwise changing the design, its rotational inertia will
    a. double.
    b. quadruple.
    c. increase by a factor of 8.
    d. increase by a factor of 16.

Answers to Chapter Questions

Answer to Chapter Opening Question

The blade mass should be concentrated toward the rotation axis, thus lowering the turbine's rotational inertia—the rotational analog of mass.

Answers to GOT IT? Questions

10.1. (a) 10.5*b*; (b) 10.5*a*.
10.2. (b) Rotational inertia would decrease with the axis at the center $(mL^2/2)$ and (a) increase (mL^2) with the axis at one end.
10.3. The mass of the shell is farther from the rotation axis.
10.4. (a). There must be a net torque acting to increase the pulley's clockwise angular velocity. The difference in the two tension forces provides that torque.

11 Rotational Vectors and Angular Momentum

Earth isn't quite round. How does this affect its rotation axis, and what's this got to do with ice ages? (The deviation from roundness is exaggerated in this photo.)

New Concepts, New Skills

By the end of this chapter you should be able to

- Specify the vector direction of angular velocity, angular acceleration, and torque (11.1, 11.2).
- Define the vector cross product and use it in expressing torque and angular momentum (11.2, 11.3).
- Solve quantitative problems involving colinear angular momentum vectors and the conservation of angular momentum (11.4).
- Describe qualitatively the precession of a rotation axis under the influence of a torque (11.5).

Connecting Your Knowledge

- This chapter generalizes concepts of rotational motion developed in Chapter 10. You should have a solid understanding of angular velocity, angular acceleration, and torque (10.1, 10.2).

Summer, fall, winter, spring: the cycle of the seasons is ultimately determined by the vector direction of Earth's angular velocity. The changing angular velocity of protons in living tissue produces MRI images that give physicians a noninvasive look inside the human body. Rising and rotating, moist, heated air forms itself into the ominous funnel of a tornado. You ride your bicycle, the rotating wheels helping stabilize what seems a precarious balance. These examples all involve rotational motion in which not only the magnitude but also the *direction* matters. They're best understood in terms of the rotational analog of Newton's law, which we introduce here in full vector form involving a rotational analog of momentum. The transition from Chapter 10 to Chapter 11 is analogous to the leap from Chapter 2's one-dimensional description of motion to the full vector description in Chapter 3. Here, as there, we'll find a new richness of phenomena involving motion.

11.1 Angular Velocity and Acceleration Vectors

So far we've ascribed direction to rotational motion using the terms "clockwise" and "counterclockwise." But that's not enough: To describe rotational motion fully we need to specify the direction of the rotation axis. We therefore define **angular velocity** $\vec{\omega}$ as a vector whose magnitude is the angular speed ω and whose direction is parallel to the rotation axis. There's an ambiguity in this definition, since there are two possible directions parallel to the axis. We resolve the ambiguity with the **right-hand rule**: If you curl the fingers of your right hand to follow the rotation, then your right thumb points in the direction of the angular velocity (Fig. 11.1). This refinement means that $\vec{\omega}$ not only gives the angular speed and the direction of the rotation axis but also distinguishes what we would have described previously as clockwise or counterclockwise rotation.

By analogy with the linear acceleration vector, we define angular acceleration as the rate of change of the angular velocity vector:

$$\vec{\alpha} = \lim_{\Delta t \to 0} \frac{\Delta \vec{\omega}}{\Delta t} = \frac{d\vec{\omega}}{dt} \quad \text{(angular acceleration vector)} \tag{11.1}$$

where, as with Equation 10.4, we get the average angular acceleration if we don't take the limit.

Equation 11.1 says that angular acceleration points in the direction of the *change* in angular velocity. If that change is only in magnitude, then $\vec{\omega}$ simply grows or shrinks, and $\vec{\alpha}$ is parallel or antiparallel to the rotation axis (Fig. 11.2*a, b*). But a change in *direction* is also a change in angular velocity. When the angular velocity $\vec{\omega}$ changes only in direction, then the angular acceleration vector is perpendicular to $\vec{\omega}$ (Fig. 11.2*c*). More generally, both the magnitude and direction of $\vec{\omega}$ may change; then $\vec{\alpha}$ is neither parallel nor perpendicular to $\vec{\omega}$. These cases are exactly analogous to the situations we treated in Chapter 3, where acceleration parallel to velocity changes only the speed, while acceleration perpendicular to velocity changes only the direction of motion.

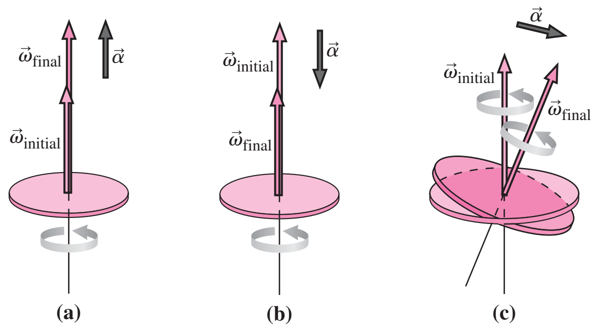


FIGURE 11.2 Angular acceleration can (a) increase or (b) decrease the magnitude of the angular velocity, or (c) change its direction.

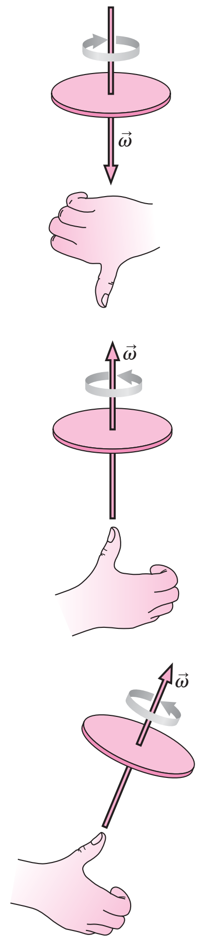


FIGURE 11.1 The right-hand rule gives the direction of the angular velocity vector.

11.2 Torque and the Vector Cross Product

Figure 11.3 shows a wheel, initially stationary, with a force applied at its rim. The torque associated with this force sets the wheel rotating in the direction shown; applying the right-hand rule, we see that angular velocity vector $\vec{\omega}$ points upward. Since the angular speed is increasing, the angular acceleration $\vec{\alpha}$ also points upward. So that our rotational analog of Newton's law—angular acceleration proportional to torque—will hold for directions as well as magnitudes, we'd like the torque to have an upward direction, too.

We already know the magnitude of the torque: From Equation 10.10, it's $\tau = rF\sin\theta$, where θ is the angle between the vectors $\vec{r}$ and $\vec{F}$ in Fig. 11.3. We define the direction of the torque as being perpendicular to both $\vec{r}$ and $\vec{F}$, as given by the right-hand rule shown in Fig. 11.4. You can verify that this rule gives an upward direction for the torque in Fig. 11.3.

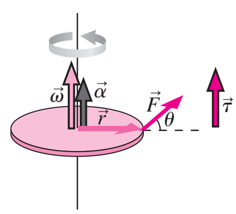


FIGURE 11.3 The torque vector is perpendicular to $\vec{r}$ and $\vec{F}$, and in the same direction as the angular acceleration. Here $\vec{F}$ lies in the plane of the disk.

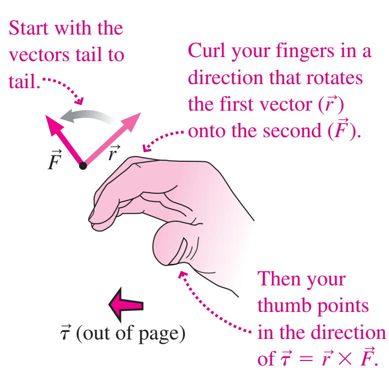


FIGURE 11.4 The right-hand rule for the direction of torque.

The Cross Product

The magnitude of the torque, $\tau = rF\sin\theta$, is determined by the magnitudes of the vectors $\vec{r}$ and $\vec{F}$ and the angle between them; the direction of the torque is determined by the vectors $\vec{r}$ and $\vec{F}$ through the right-hand rule. This operation—forming from two vectors $\vec{A}$ and $\vec{B}$ a third vector $\vec{C}$ of magnitude $C = AB\sin\theta$ and direction given by the right-hand rule—occurs frequently in physics and is called the **cross product**:

The cross product $\vec{C}$ of two vectors $\vec{A}$ and $\vec{B}$ is written

$$\vec{C} = \vec{A} \times \vec{B}$$

and is a vector with magnitude $AB\sin\theta$, where θ is the angle between $\vec{A}$ and $\vec{B}$, and where the direction of $\vec{C}$ is given by the right-hand rule of Fig. 11.4.

Torque is an instance of the cross product, and we can write the torque vector simply as

$$\vec{\tau} = \vec{r} \times \vec{F} \qquad \text{(torque vector)} \tag{11.2}$$

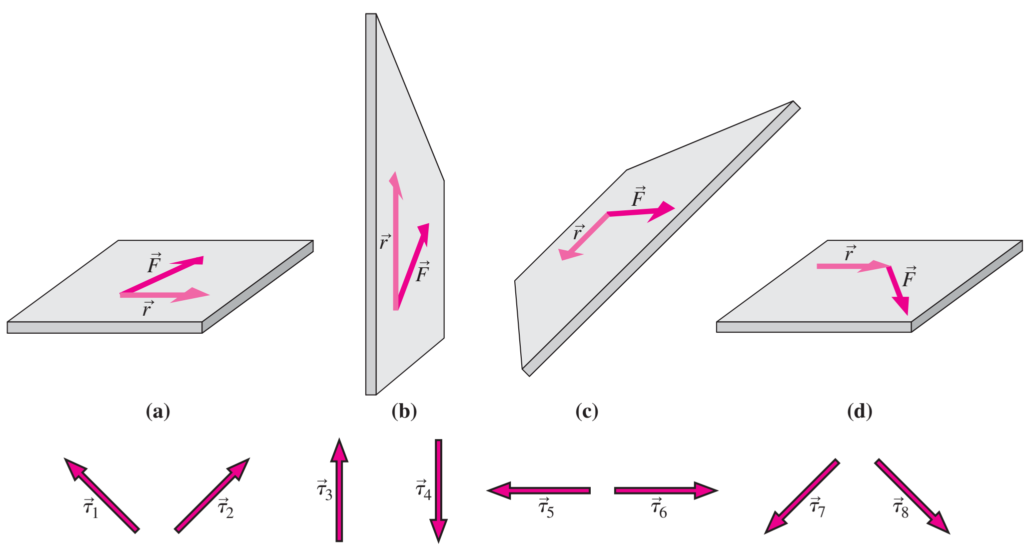
Both direction and magnitude are described succinctly in this equation.

TACTICS 11.1 Multiplying Vectors

The cross product $\vec{A} \times \vec{B}$ is the second way of multiplying vectors that you've encountered. The first was the scalar product $\vec{A} \cdot \vec{B} = AB\cos\theta$ introduced in Chapter 6 and also called the dot product. Both depend on the product of the vector magnitudes and on the angle between them. But where the dot product depends on the *cosine* of the angle and is therefore maximum when the two vectors are parallel, the cross product depends on the *sine* and is therefore maximum for perpendicular vectors. There's another crucial distinction between dot product and cross product: The dot product is a *scalar*—a single number, with no direction—while the cross product is a *vector*. That's why $AB\cos\theta$ completely specifies the dot product, but $AB\sin\theta$ gives only the magnitude of the cross product; it's also necessary to specify the direction via the right-hand rule.

The cross product obeys the usual distributive rule: $\vec{A} \times (\vec{B} + \vec{C}) = \vec{A} \times \vec{B} + \vec{A} \times \vec{C}$, but it's *not* commutative; in fact, as you can see by rotating $\vec{F}$ onto $\vec{r}$ instead of $\vec{r}$ onto $\vec{F}$ in Fig. 11.4, $\vec{B} \times \vec{A} = -\vec{A} \times \vec{B}$.

GOT IT? 11.1 The figure shows four pairs of force and radius vectors and eight torque vectors. Which numbered torque vector goes with each pair of force-radius vectors? Consider only direction, not magnitude.



11.3 Angular Momentum

We first used Newton's law in the form $\vec{F} = m\vec{a}$, but later found the form $\vec{F} = d\vec{p}/dt$ especially powerful. The same is true in rotational motion: To explore fully some surprising aspects of rotational dynamics, we need to define angular momentum and develop a relation between its rate of change and the applied torque. Once we've done that, we'll be able to answer questions like why a gyroscope doesn't fall over and how spinning protons yield MRI images of your body's innards.

Like other rotational quantities, angular momentum is specified with respect to a given point or axis. We begin with the **angular momentum** $\vec{L}$ of a single particle:

If a particle with linear momentum $\vec{p}$ is at position $\vec{r}$ with respect to some point, then its angular momentum $\vec{L}$ about that point is defined as

$$\vec{L} = \vec{r} \times \vec{p} \qquad \text{(angular momentum)} \qquad (11.3)$$

EXAMPLE 11.1 Calculating Angular Momentum: A Single Particle

A particle of mass m moves counterclockwise at speed v around a circle of radius r in the x-y plane. Find its angular momentum about the center of the circle, and express the answer in terms of its angular velocity.

INTERPRET We're given the motion of a particle—namely, uniform motion in a circle—and asked to find the corresponding angular momentum and its relation to angular velocity.

DEVELOP Figure 11.5 is our sketch, showing the particle in its circular path. We added an xyz coordinate system with the circular path in the x-y plane. Equation 11.3, $\vec{L} = \vec{r} \times \vec{p}$, gives the angular momentum in terms of the vector $\vec{r}$ and the linear momentum $\vec{p}$. We know that linear momentum is the product $m\vec{v}$, so we have everything we need to apply Equation 11.3. We'll then express our result in terms of angular velocity using $v = \omega r$.

EVALUATE Figure 11.5 shows that the linear momentum $m\vec{v}$ is perpendicular to $\vec{r}$, so $\sin\theta = 1$ in the cross product, and the magnitude of the angular momentum becomes $L = mvr$. Applying the right-hand rule shows that $\vec{L}$ points in the z-direction, so we can write $\vec{L} = mvr\hat{k}$. But $v = \omega r$, and the right-hand rule shows that $\vec{\omega}$, too, points in the z-direction. So we can write

$$\vec{L} = mvr\hat{k} = mr^2\omega\hat{k} = mr^2\vec{\omega}$$

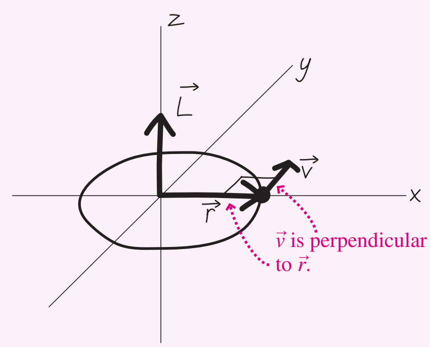


FIGURE 11.5 Finding the angular momentum $\vec{L}$ of a particle moving in a circle.

ASSESS Make sense? The faster the particle is going, the more linear momentum it has. But angular momentum depends on linear momentum and distance from the rotation axis, so at a given angular speed, the angular momentum scales as the *square* of the radius. ■

Angular momentum is the rotational analog of linear momentum $\vec{p} = m\vec{v}$. Since rotational inertia I is the analog of mass m, and angular velocity $\vec{\omega}$ is the analog of linear velocity $\vec{v}$, you might expect that we could write

$$\vec{L} = I\vec{\omega} \qquad (11.4)$$

The rotational inertia of a single particle is mr^2, so you can see that the result of Example 11.1 can indeed be written $\vec{L} = I\vec{\omega}$. Equation 11.4 also holds for symmetric objects like a wheel or sphere rotating about a fixed axis. But in more complicated cases, Equation 11.4 may not hold; surprisingly, $\vec{L}$ and $\vec{\omega}$ can even have different directions. We'll leave such cases for more advanced courses.

Torque and Angular Momentum

We're now ready to develop the full vector analog of Newton's law in the form $\vec{F} = d\vec{P}/dt$. Recall that $\vec{F}$ here is the *net* external force on a system, and $\vec{P}$ is the system's momentum—the vector sum of the momenta of its constituent particles. Can we write, by analogy, $\vec{\tau} = d\vec{L}/dt$? To see that we can, we write the angular momentum of a system as the sum of the angular momenta of its constituent particles:

$$\vec{L} = \sum \vec{L}_i = \sum (\vec{r}_i \times \vec{p}_i)$$

where the subscript i refers to the ith particle. Differentiating gives

$$\frac{d\vec{L}}{dt} = \sum \left(\vec{r}_i \times \frac{d\vec{p}_i}{dt} + \frac{d\vec{r}_i}{dt} \times \vec{p}_i \right)$$

where we've applied the product rule for differentiation, being careful to preserve the order of the cross product since it's not commutative. But $d\vec{r}_i/dt$ is the velocity of the ith particle, so the second term in the sum is the cross product of velocity $\vec{v}$ and momentum $\vec{p} = m\vec{v}$. Since these two vectors are parallel, their cross product is zero, and we're left with only the first term in the sum:

$$\frac{d\vec{L}}{dt} = \sum \left(\vec{r}_i \times \frac{d\vec{p}_i}{dt} \right) = \sum (\vec{r}_i \times \vec{F}_i)$$

where we've used Newton's law to write $d\vec{p}_i/dt = \vec{F}_i$. But $\vec{r}_i \times \vec{F}_i$ is the torque $\vec{\tau}_i$ on the ith particle, so

$$\frac{d\vec{L}}{dt} = \sum \vec{\tau}_i$$

The sum here includes both external and internal torques—the latter due to interactions among the particles of the system. Newton's third law assures us that internal *forces* cancel in pairs, but what about *torques*? They'll cancel, too, provided the internal forces act along lines joining pairs of particles. This condition is stronger than Newton's third law alone, and it usually but not always holds. When it does, the sum of torques reduces to the net *external* torque, and we have

$$\frac{d\vec{L}}{dt} = \vec{\tau} \qquad \left(\begin{array}{l}\text{rotational analog,}\\ \text{Newton's 2}^{\text{nd}}\text{ law}\end{array}\right) \tag{11.5}$$

where $\vec{\tau}$ is the net external torque. Thus our analogy between linear and rotational motion holds for momentum as well as for the other quantities we've discussed.

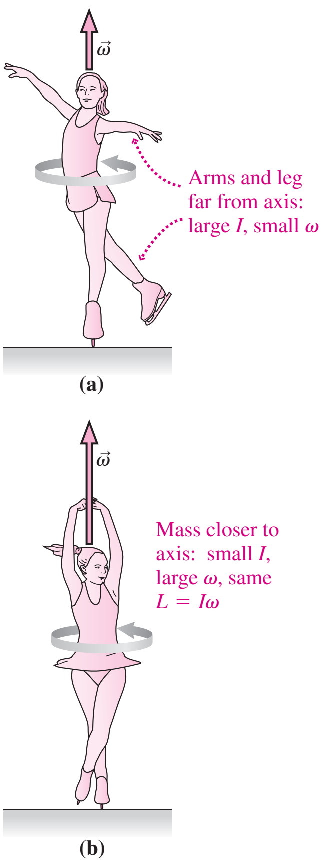


FIGURE 11.6 As the skater's rotational inertia decreases, her angular speed increases to conserve angular momentum.

11.4 Conservation of Angular Momentum

When there's no external torque on a system, Equation 11.5 tells us that angular momentum is constant. This statement—that the angular momentum of an isolated system cannot change—is of fundamental importance in physics, and applies to systems ranging from subatomic particles to galaxies. Because a composite system can change its configuration—and hence its rotational inertia I—conservation of angular momentum requires that angular speed increase if I decreases, and vice versa. The classic example is a figure skater who starts spinning relatively slowly with arms and leg extended and then pulls in her limbs to spin rapidly (Fig. 11.6). A more dramatic example is the collapse of a star at the end of its lifetime, explored in the next example.

EXAMPLE 11.2 Conservation of Angular Momentum: Pulsars

A star rotates once every 45 days. At the end of its life, it undergoes a supernova explosion, hurling much of its mass into the interstellar medium. But the inner core of the star, whose radius is initially 20 Mm, collapses into a neutron star only 6 km in radius. As it rotates, the neutron star emits regular pulses of radio waves, making it a *pulsar*. Calculate the rotation rate, which is the same as the pulse rate that radio astronomers detect. Consider the core to be a uniform sphere, and assume that no external torques act during the collapse.

INTERPRET Here we're given the radius and rotation rate of the stellar core before collapse and asked for the rotation rate afterward. That kind of "before and after" question often calls for the application of a conservation law. In this case there's no external torque, so it's angular momentum that's conserved.

DEVELOP The magnitude of the angular momentum is $I\omega$, so our plan is to write this expression before and after collapse, and then equate the two to find the new rotation rate: $I_1\omega_1 = I_2\omega_2$. We need to use Table 10.2's expression for the rotational inertia of a solid sphere: $I = \frac{2}{5}MR^2$.

EVALUATE Given I, our statement of angular momentum conservation becomes $\frac{2}{5}MR_1^2\omega_1 = \frac{2}{5}MR_2^2\omega_2$, or

$$\omega_2 = \omega_1\left(\frac{R_1}{R_2}\right)^2 = \left(\frac{1\text{ rev}}{45\text{ day}}\right)\left(\frac{2\times10^7\text{ m}}{6\times10^3\text{ m}}\right)^2 = 2.5\times10^5\text{ rev/day}$$

ASSESS Our answer is huge, about 3 revolutions per second. But that makes sense. This neutron star is a fantastic thing—an object with more mass than the entire Sun, crammed into a diameter of about 8 miles. It's because of that dramatic reduction in radius—and thus in rotational inertia—that the pulsar's rotation rate is so high. Note that in a case like this, where ω appears on both sides of the equation, it isn't necessary to convert to radian measure. ■

CONCEPTUAL EXAMPLE 11.1 On the Playground

A merry-go-round is rotating freely when a boy runs radially inward, straight toward the merry-go-round's center, and leaps on. Later, a girl runs tangent to the merry-go-round's edge, in the same direction the edge is moving, and also leaps on. Does the merry-go-round's angular speed increase, decrease, or stay the same in each case?

EVALUATE Because the merry-go-round is rotating freely, the only torques are those exerted by the children as they leap on. If we consider a system consisting of the merry-go-round and both children, then those torques are internal, and the system's angular momentum is conserved. In Fig. 11.7 we've sketched the situation, before either child leaps onto the merry-go-round and after both are on board.

The boy, running radially, carries no angular momentum (his linear momentum and the radius vector are in the same direction, making $\vec{L}$ zero), so you might think he doesn't change the merry-go-round's angular speed. Yet he does, because he adds mass and therefore rotational inertia. At the same time he doesn't change the angular momentum, so with I increased, ω must therefore drop.

Running in the same direction as the merry-go-round's tangential velocity, the girl clearly adds angular momentum to the system—an addition that would tend to increase the angular speed. But she also adds mass, and thus increases the rotational inertia—which, as in the boy's case, tends to decrease angular speed. So which wins out? That depends on her speed. Without knowing that, we can't tell whether the merry-go-round speeds up or slows down.

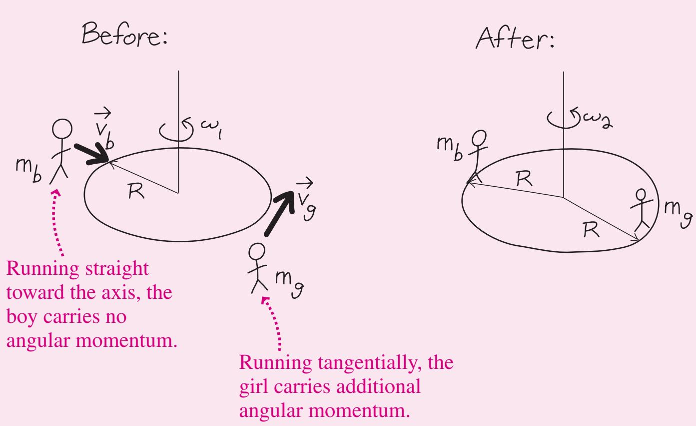


FIGURE 11.7 Our diagrams for Conceptual Example 11.1

ASSESS The angular momentum the girl adds is the product of her linear momentum mv and the merry-go-round's radius R, while she increases the rotational inertia by mR^2. With small m and large v, she could add a lot of angular momentum without increasing the rotational inertia significantly. That would increase the merry-go-round's rotation rate. But with a large m and small v—giving the same additional angular momentum—the increase in rotational inertia would more than offset the angular momentum added, and the merry-go-round would slow down. We can't answer the question about the merry-go-round's angular speed without knowing the numbers. Problem 45 explores a similar situation in more detail.

MAKING THE CONNECTION Take the merry-go-round's radius to be $R = 1.3$ m, its rotational inertia $I = 240\text{ kg}\cdot\text{m}^2$, and its initial angular speed $\omega_{\text{initial}} = 11$ rpm. The boy's and girl's masses are, respectively, 28 kg and 32 kg, and they run, respectively, at 2.5 m/s and 3.7 m/s. Find the merry-go-round's angular speed ω_{final} after both children are on board.

EVALUATE Following the conceptual example, take the system to include the merry-go-round and the two children. Before the children leap on, both the merry-go-round itself and the girl carry angular momentum; afterward, with children and merry-go-round rotating with a common angular speed, they all do. Thus conservation of angular momentum reads

$$I\omega_{\text{initial}} + m_g v_g R = I\omega_{\text{final}} + m_b R^2\omega_{\text{final}} + m_g R^2\omega_{\text{final}}$$

Solving with the given numbers yields $\omega_{\text{final}} = 12$ rpm. That's not much change, so the girl's effect must have been a speed increase, but only a little more than enough to overcome the boy's slowing effect. Note that the boy's speed didn't matter, since it didn't contribute to angular momentum or rotational inertia. And be careful with units: You've got to express all angular momenta in the same units. That means converting angular speeds to radians per second or expressing the girl's angular momentum $m_g v_g R$ in unusual units, $\text{kg}\cdot\text{m}^2\cdot\text{rpm}$.

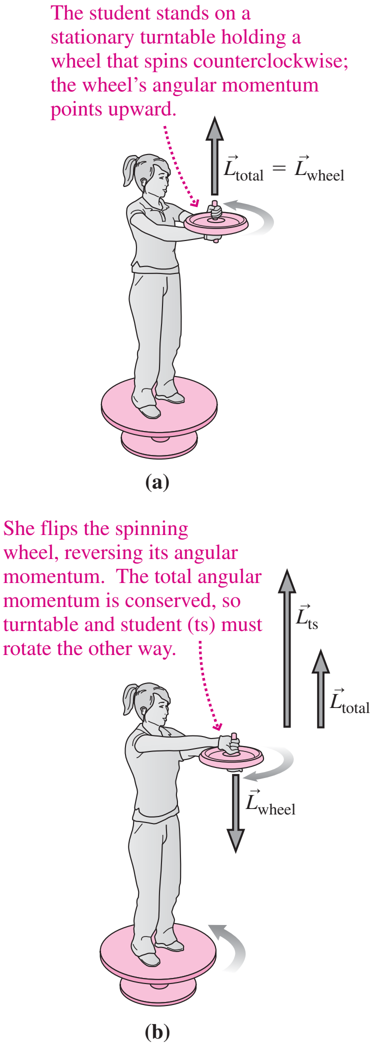


FIGURE 11.8 A demonstration of angular momentum conservation.

✓TIP Angular Momentum in Straight-line Motion

You don't have to be rotating to have angular momentum. The girl in Conceptual Example 11.1 was running in a straight line, yet she had nonzero angular momentum with respect to the merry-go-round's rotation axis. Problem 34 explores this point further.

In a popular demonstration, a student stands on a stationary turntable holding a wheel rotating about a vertical axis. The student flips the wheel upside down, and the turntable starts rotating. Figure 11.8 shows how angular momentum conservation explains this behavior. Once again, though, mechanical energy isn't conserved. In this case the student does work, exerting forces that result in torques on her body and the turntable. The end result is a greater rotational kinetic energy than was initially present.

GOT IT? 11.2 Suppose you step onto a nonrotating turntable like the one shown in Fig. 11.8, holding a nonrotating wheel with its axis vertical. (a) If you spin the wheel counterclockwise as viewed from above, which way will you rotate? (b) If you now turn the spinning wheel upside down, will your rotation rate increase, decrease, or remain unchanged? What about your direction of rotation?

11.5 Gyroscopes and Precession

Angular momentum—a vector quantity with direction as well as magnitude—is conserved in the absence of external torques. For symmetric objects, angular momentum has the same direction as the rotation axis, so the axis can't change direction unless an external torque acts. This is the principle behind the gyroscope—a spinning object whose rotation axis remains fixed in space. The faster a gyroscope spins, the larger its angular momentum and thus the harder it is to change its orientation. Gyroscopes are widely used for navigation, where their direction-holding capability provides an alternative to the magnetic compass. More sophisticated gyroscope systems guide missiles and submarines, stabilize cruise ships in heavy seas, and track the orientation of wireless computer mice. Space telescopes start and stop gyroscopic wheels oriented along three perpendicular axes; to conserve angular momentum, the entire telescope reorients itself to point toward a desired astronomical object. This approach avoids rocket exhaust that would foul the telescope's superb viewing, and ensures that there's no fuel to run out. Instead, solar-generated electricity operates the wheels' drive motors.

Precession

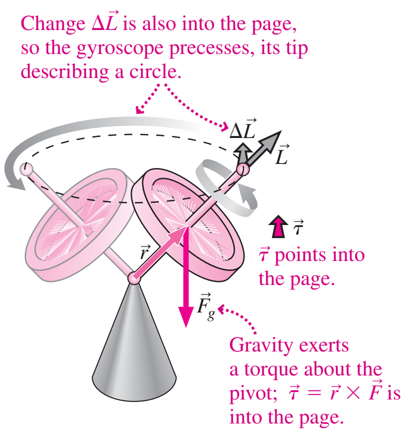


FIGURE 11.9 Why doesn't the spinning gyroscope fall over?

If an object does experience a net external torque, then, according to the rotational analog of Newton's law (Equation 11.5, $d\vec{L}/dt = \vec{\tau}$), its angular momentum must change. For rapidly rotating objects, the result is the surprising phenomenon of **precession**—a continual change in the direction of the rotation axis, which traces out a circle. You may have seen a toy gyroscope or top precess instead of simply falling over as you might expect. Figure 11.9 shows that precession comes about because the direction of the angular momentum change is the same as the direction of the torque.

Precession on the atomic scale helps explain the medical imaging technique MRI (*m*agnetic *r*esonance *i*maging). Protons in the body's abundant hydrogen precess because of torque resulting from a strong magnetic field. The MRI imager detects signals emitted at the precession frequency. By spatially varying the magnetic field, the device localizes the precessing protons and thus constructs high-resolution images of the body's interior.

On a much larger scale, Earth itself precesses. Because of its rotation, the planet bulges slightly at the equator. Solar gravity exerts a torque on the equatorial bulge, causing

Earth's rotation axis to precess with a period of about 26,000 years (Fig. 11.10). The axis now points toward Polaris, which for that reason we call the North Star, but it won't always do so. This precession, in connection with deviations in Earth's orbit from a perfect circle, results in subtle climatic changes that are believed partly responsible for the onset of ice ages.

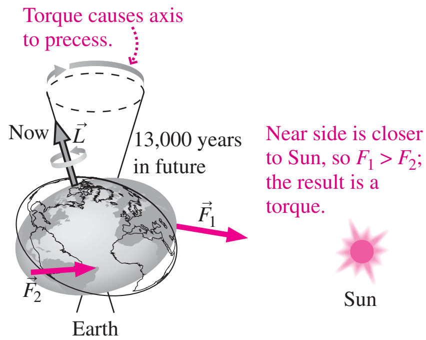
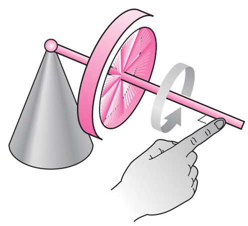


FIGURE 11.10 Earth's precession. The equatorial bulge is highly exaggerated.

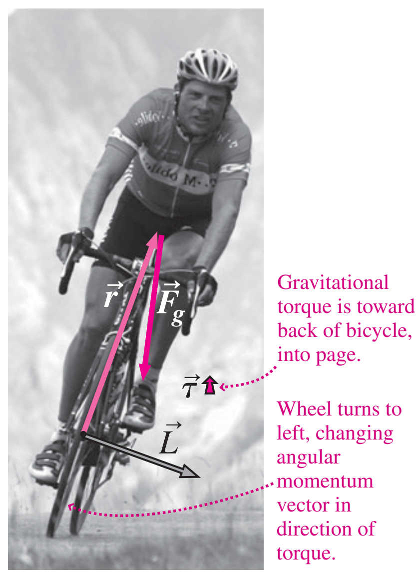
GOT IT? 11.3 You push horizontally at right angles to the shaft of a spinning gyroscope, as shown in the figure. Does the shaft move (a) upward, (b) downward, (c) in the direction of your push, or (d) opposite the direction of your push?



APPLICATION **Bicycling**

The rotational analog of Newton's second law helps explain why bicycles don't tip over. The photo shows why. If the bicycle is perfectly vertical, the gravitational force exerts no torque. But if it tips to the rider's left, as in the photo, then there's a torque $\vec{\tau} = \vec{r} \times \vec{F}_g$ toward the rear. A stationary bicycle, with no angular momentum, would respond by tipping further left, rotating about a front-to-back axis and gaining angular momentum toward the rear. That's just as Newton requires: a change in angular momentum in the direction of the torque. But a moving bicycle already has angular momentum $\vec{L}$ of its rotating wheels; as the photo shows, it points generally to the rider's left. A rearward change in the angular momentum then requires just a slight turn of the front wheel to the left. The rider subconsciously makes that turn, at once satisfying Newton and helping to keep the bicycle stable.

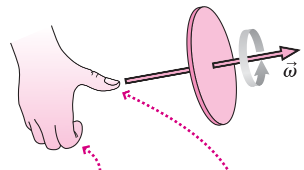
The physics of cycling is a complicated subject, and the role of angular momentum described here is only one of several effects that contribute to bicycle stability.



CHAPTER 11 SUMMARY

Big Picture

The big idea of this chapter is that rotational quantities can be described as vectors, with the vector direction at right angles to the plane in which the action—motion, acceleration, or effects associated with torque—is occurring. The direction is given by the right-hand rule. A new concept, angular momentum, is the rotational analog of linear momentum. The rotational analog of Newton's law equates the net torque on a system with the rate of change of its angular momentum. In the absence of a net torque, angular momentum is conserved.



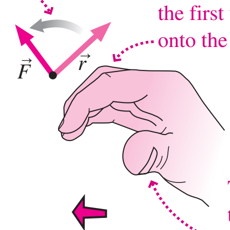
Key Concepts and Equations

The **vector cross product** is a way of multiplying two vectors $\vec{A}$ and $\vec{B}$ to produce a third vector $\vec{C}$ of magnitude $C = AB \sin\theta$ and direction at right angles to the other two, as given by the right-hand rule. It's written as

$$\vec{C} = \vec{A} \times \vec{B}$$

Torque is defined as the cross product of the radius vector $\vec{r}$ from a given axis to the point where a force $\vec{F}$ is applied:

$$\vec{\tau} = \vec{r} \times \vec{F}$$



Angular momentum $\vec{L}$ is the rotational analog of linear momentum $\vec{p}$. It's always defined with respect to a particular axis. For a point particle at position $\vec{r}$ with respect to the axis, moving with linear momentum $\vec{p} = m\vec{v}$, the angular momentum is defined as

$$\vec{L} = \vec{r} \times \vec{p}$$

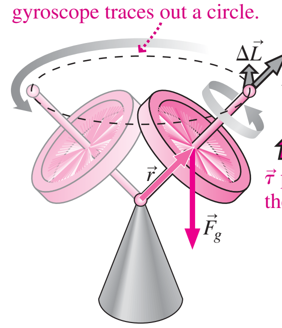
For a symmetric object with rotational inertia I rotating with angular velocity $\vec{\omega}$, angular momentum becomes $\vec{L} = I\vec{\omega}$.
In terms of angular momentum, the rotational analog of Newton's law states that the rate of change of angular momentum is equal to the net external torque:

$$\frac{d\vec{L}}{dt} = \vec{\tau}$$

If the external torque on a system is zero, then its angular momentum cannot change.

Applications

Conservation of angular momentum explains the action of gyroscopes—spinning objects whose rotation axis remains fixed in the absence of a net external torque. If an external torque is applied, the rotation axis undergoes a circular motion known as **precession**. Precession occurs in systems ranging from subatomic particles to tops and gyroscopes and on to planets.



For homework assigned on MasteringPhysics, go to www.masteringphysics.com | BIO Biology and/or medical-related problems | C Computer problems

For Thought and Discussion

1. Does Earth's angular velocity vector point north or south?
2. Figure 11.11 shows four forces acting on a body. In what directions are the associated torques about point O? About point P?

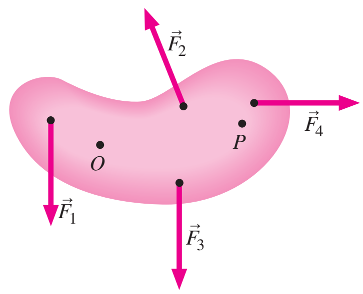


FIGURE 11.11 For Thought and Discussion 2

3. You stand with your right arm extended horizontally. What's the direction of the gravitational torque about your shoulder?
4. Although it contains no parentheses, the expression $\vec{A} \times \vec{B} \cdot \vec{C}$ is unambiguous. Why?
5. What's the angle between two vectors if their dot product is equal to the magnitude of their cross product?
6. Why does a tetherball move faster as it winds up its pole?
7. Why do helicopters have two rotors?
8. A group of polar bears is standing around the edge of a slowly rotating ice floe. If the bears all walk to the center, what happens to the rotation rate?
9. Tornadoes in the northern hemisphere rotate counterclockwise as viewed from above. A far-fetched idea suggests that driving on the right side of the road may increase the frequency of tornadoes. Does this idea have any merit? Explain in terms of the angular momentum imparted to the air as two cars pass.
10. Does a particle moving at constant speed in a straight line have angular momentum about a point on the line? About a point not on the line? In either case, is its angular momentum constant?
11. When you turn on a high-speed power tool such as a router, the tool tends to twist in your hands. Why?
12. Why is it easier to balance a basketball on your finger if it's spinning?

Exercises and Problems

Exercises

Section 11.1 Angular Velocity and Acceleration Vectors

13. A car is headed north at 70 km/h. Give the magnitude and direction of the angular velocity of its 62-cm-diameter wheels.
14. If the car of Exercise 13 makes a 90° left turn lasting 25 s, determine the average angular acceleration of the wheels.
15. A wheel is spinning at 45 rpm with its axis vertical. After 15 s, it's spinning at 60 rpm with its axis horizontal. Find (a) the magnitude of its average angular acceleration and (b) the angle the average angular acceleration vector makes with the horizontal.
16. A wheel is spinning about a horizontal axis with angular speed 140 rad/s and with its angular velocity pointing east. Find the magnitude and direction of its angular velocity after an angular acceleration of 35 rad/s^2, pointing 68° west of north, is applied for 5.0 s.

Section 11.2 Torque and the Vector Cross Product

17. A 12-N force is applied at the point $x = 3$ m, $y = 1$ m. Find the torque about the origin if the force points in (a) the x-direction, (b) the y-direction, and (c) the z-direction.
18. A force $\vec{F} = 1.3\hat{\imath} + 2.7\hat{\jmath}$ N is applied at the point $x = 3.0$ m, $y = 0$ m. Find the torque about (a) the origin and (b) the point $x = -1.3$ m, $y = 2.4$ m.
19. BIO When you hold your arm outstretched, it's supported primarily by the deltoid muscle. Figure 11.12 shows a case in which the deltoid exerts a 67-N force at 15° to the horizontal. If the force-application point is 18 cm horizontally from the shoulder joint, what torque does the deltoid exert about the shoulder?

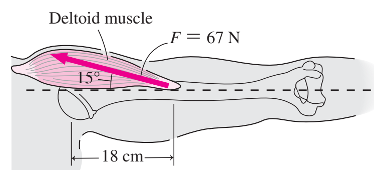


FIGURE 11.12 Exercise 19

Section 11.3 Angular Momentum

20. Express the units of angular momentum (a) using only the fundamental units kilogram, meter, and second; (b) in a form involving newtons; (c) in a form involving joules.
21. In the Olympic hammer throw, a contestant whirls a 7.3-kg steel ball on the end of a 1.2-m cable. If the contestant's arms reach an additional 90 cm from his rotation axis and if the ball's speed just prior to release is 27 m/s, what's the magnitude of the ball's angular momentum?
22. A gymnast of rotational inertia 62 kg·m^2 is tumbling head over heels with angular momentum 470 kg·m^2/s. What's her angular speed?
23. A 640-g hoop 90 cm in diameter is rotating at 170 rpm about its central axis. What's its angular momentum?
24. A 7.4-cm-diameter baseball has mass 145 g and is spinning at 2000 rpm. Treating the baseball as a uniform solid sphere, what's its angular momentum?

Section 11.4 Conservation of Angular Momentum

25. A potter's wheel with rotational inertia 6.40 kg·m^2 is spinning freely at 19.0 rpm. The potter drops a 2.70-kg lump of clay onto the wheel, where it sticks 46.0 cm from the rotation axis. What's the wheel's subsequent angular speed?
26. A 3.0-m-diameter merry-go-round with rotational inertia 120 kg·m^2 is spinning freely at 0.50 rev/s. Four 25-kg children sit suddenly on the edge of the merry-go-round. (a) Find the new angular speed, and (b) determine the total energy lost to friction between children and merry-go-round.
27. A uniform, spherical cloud of interstellar gas has mass 2.0×10^{30} kg, has radius 1.0×10^{13} m, and is rotating with period 1.4×10^{6} years. The cloud collapses to form a star 7.0×10^{8} m in radius. Find the star's rotation period.
28. A skater has rotational inertia 4.2 kg·m^2 with his fists held to his chest and 5.7 kg·m^2 with his arms outstretched. The skater is spinning at 3.0 rev/s while holding a 2.5-kg weight in each outstretched hand; the weights are 76 cm from his rotation axis. If he pulls his hands in to his chest, so they're essentially on his rotation axis, how fast will he be spinning?

Problems

29. You slip a wrench over a bolt. Taking the origin at the bolt, the other end of the wrench is at $x = 18$ cm, $y = 5.5$ cm. You apply a force $\vec{F} = 88\hat{\imath} - 23\hat{\jmath}$ N to the end of the wrench. What's the torque on the bolt?
30. Vector $\vec{A}$ points 30° counterclockwise from the x-axis. Vector $\vec{B}$ has twice the magnitude of $\vec{A}$. Their product $\vec{A} \times \vec{B}$ has magnitude A^2 and points in the negative z-direction. What's the direction of vector $\vec{B}$?
31. **BIO** A baseball player extends his arm straight up to catch a 145-g baseball moving horizontally at 42 m/s. It's 63 cm from the player's shoulder joint to the point the ball strikes his hand, and his arm remains stiff while it rotates about the shoulder during the catch. The player's hand recoils 5.00 cm horizontally while he stops the ball. What average torque does the player's arm exert on the ball?
32. Show that $\vec{A} \cdot (\vec{A} \times \vec{B}) = 0$ for any vectors $\vec{A}$ and $\vec{B}$.
33. A weightlifter's barbell consists of two 25-kg masses on the ends of a 15-kg rod 1.6 m long. The weightlifter holds the rod at its center and spins it at 10 rpm about an axis perpendicular to the rod. What's the magnitude of the barbell's angular momentum?
34. A particle of mass m moves in a straight line at constant speed v. Show that its angular momentum about a point located a perpendicular distance b from its line of motion is mvb regardless of where the particle is on the line.
35. **BIO** Biomechanical engineers have developed micromechanical devices for measuring blood flow as an alternative to dye injection following angioplasty to remove arterial plaque. One experimental device consists of a 300-μm-diameter, 2.0-μm-thick silicon rotor inserted into blood vessels. Moving blood spins the rotor, whose rotation rate provides a measure of blood flow. This device exhibited an 800-rpm rotation rate in tests with water flows at several m/s. Treating the rotor as a disk, what was its angular momentum at 800 rpm? (*Hint*: You'll need to find the density of silicon.)
36. Figure 11.13 shows the dimensions of a 880-g wooden baseball bat whose rotational inertia about its center of mass is $0.048\ \text{kg}\cdot\text{m}^2$. If the bat is swung so its far end moves at 50 m/s, find (a) its angular momentum about the pivot P and (b) the constant torque applied about P to achieve this angular momentum in 0.25 s. (*Hint:* Remember the parallel-axis theorem.)

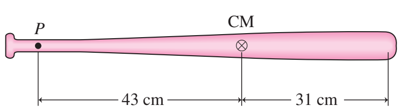


FIGURE 11.13 Problem 36

37. As an automotive engineer, you're charged with redesigning a car's wheels with the goal of decreasing each wheel's angular momentum by 30% for a given linear speed of the car. Other design considerations require that the wheel diameter go from 38 cm to 35 cm. If the old wheel had rotational inertia $0.32\ \text{kg}\cdot\text{m}^2$, what do you specify for the new rotational inertia?
38. A turntable of radius 25 cm and rotational inertia $0.0154\ \text{kg}\cdot\text{m}^2$ is spinning freely at 22.0 rpm about its central axis, with a 19.5-g mouse on its outer edge. The mouse walks from the edge to the center. Find (a) the new rotation speed and (b) the work done by the mouse.
39. A 17-kg dog is standing on the edge of a stationary, frictionless turntable of rotational inertia $95\ \text{kg}\cdot\text{m}^2$ and radius 1.81 m. The dog walks once around the turntable. What fraction of a full circle does the dog's motion make with respect to the ground?
40. A physics student is standing on an initially motionless, frictionless turntable with rotational inertia $0.31\ \text{kg}\cdot\text{m}^2$. She's holding a wheel with rotational inertia $0.22\ \text{kg}\cdot\text{m}^2$ spinning at 130 rpm about a vertical axis, as in Fig. 11.8. When she turns the wheel upside down, student and turntable begin rotating at 70 rpm. (a) Find the student's mass, considering her to be a 30-cm-diameter cylinder. (b) Neglecting the distance between the axes of the turntable and wheel, determine the work she did in turning the wheel upside down.
41. You're choreographing your school's annual ice show. You call for eight 60-kg skaters to join hands and skate side by side in a line extending 12 m. The skater at one end is to stop abruptly, so the line will rotate rigidly about that skater. For safety, you don't want the fastest skater to be moving at more than 8.0 m/s, and you don't want the force on that skater's hand to exceed 300 N. What do you determine is the greatest speed the skaters can have before they execute their rotational maneuver?
42. Find the angle between two vectors whose dot product is twice the magnitude of their cross product.
43. A circular bird feeder 19 cm in radius has rotational inertia $0.12\ \text{kg}\cdot\text{m}^2$. It's suspended by a thin wire and is spinning slowly at 5.6 rpm. A 140-g bird lands on the feeder's rim, coming in tangent to the rim at 1.1 m/s in a direction opposite the feeder's rotation. What's the rotation rate after the bird lands?
44. A force $\vec{F}$ applied at the point $x = 2.0$ m, $y = 0$ m produces a torque $4.6\hat{k}$ N·m about the origin. If the x-component of $\vec{F}$ is 3.1 N, what angle does it make with the x-axis?
45. A turntable has rotational inertia I and is rotating with angular speed ω about a frictionless vertical axis. A wad of clay with mass m is tossed onto the turntable and sticks a distance d from the rotation axis. The clay hits horizontally with its velocity $\vec{v}$ at right angles to the turntable's radius, and in the same direction as the turntable's rotation (Fig. 11.14). Find an expression for v that will result in (a) the turntable's angular speed dropping to half its initial value, (b) no change in the turntable's angular speed, and (c) the angular speed doubling.

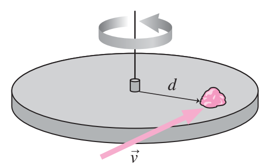


FIGURE 11.14 Problem 45

46. A uniform, solid, spherical asteroid with mass 1.2×10^{13} kg and radius 1.0 km is rotating with period 4.3 h. A meteoroid moving in the asteroid's equatorial plane crashes into the equator at 8.4 km/s. It hits at a 58° angle to the vertical and embeds itself at the surface. After the impact the asteroid's rotation period is 3.9 h. Find the meteoroid's mass.
47. About 99.9% of the solar system's total mass lies in the Sun. Using data from Appendix E, estimate what fraction of the solar system's angular momentum about its center is associated with the Sun. Where is most of the rest of the angular momentum?
48. You're a civil engineer for an advanced civilization on a solid spherical planet of uniform density. Running out of room for the expanding population, the government asks you to redesign your planet to give it more surface area. You recommend reshaping the planet, without adding any material or angular momentum, into a hollow shell whose thickness is one-fifth its outer radius. How much will your design increase the surface area, and how will it change the length of the day?
49. In Fig. 11.15, the lower disk, of mass 440 g and radius 3.5 cm, is rotating at 180 rpm on a frictionless shaft of negligible radius. The upper disk, of mass 270 g and radius 2.3 cm, is initially not rotating. It drops freely down onto the lower disk, and frictional forces bring the two disks to a common rotational speed. Find

(a) that common speed and (b) the fraction of the initial kinetic energy lost to friction.

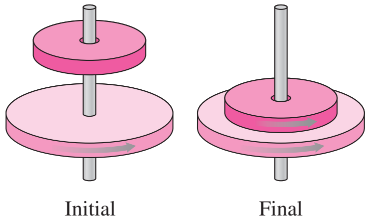


FIGURE 11.15 Problem 49

50. A massless spring with constant k is mounted on a turntable of rotational inertia I, as shown in Fig. 11.16. The turntable is on a frictionless vertical axle, though initially it's not rotating. The spring is compressed a distance Δx from its equilibrium, with a mass m placed against it. When the spring is released, the mass moves at right angles to a line through the turntable's center, at a distance b from the center, and slides without friction across the table and over the edge. Find expressions for (a) the linear speed of the mass and (b) the rotational speed of the turntable. (*Hint:* What's conserved?)

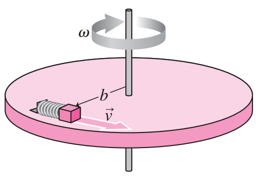


FIGURE 11.16 Problem 50

51. A solid ball of mass M and radius R is spinning with angular velocity ω_0 about a horizontal axis. It drops vertically onto a surface where the coefficient of kinetic friction with the ball is μ_k (Fig. 11.17). Find expressions for (a) the final angular velocity once it's achieved pure rolling motion and (b) the time it takes to achieve this motion.

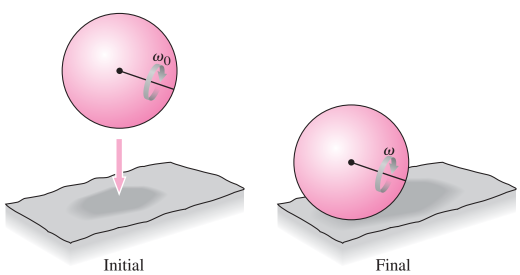


FIGURE 11.17 Problem 51

52. A time-dependent torque given by $\tau = a + b \sin ct$ is applied to an object that's initially stationary but is free to rotate. Here a, b, and c are constants. Find an expression for the object's angular momentum as a function of time, assuming the torque is first applied at $t = 0$.
53. Consider a rapidly spinning gyroscope whose axis is precessing uniformly in a horizontal circle of radius r, as shown in Fig. 11.9. Apply $\vec{\tau} = d\vec{L}/dt$ to show that the angular speed of precession about the vertical axis through the center of the circle is mgr/L.
54. When a star like our Sun exhausts its fuel, thermonuclear reactions in its core cease, and it collapses to become a *white dwarf*. Often the star will blow off its outer layers and lose some mass before it collapses. Suppose a star with the Sun's mass and radius is rotating with period 25 days and then it collapses to a white dwarf with 60% of the Sun's mass and a rotation period of 131 s. What's the radius of the white dwarf? Compare your answer with the radii of Sun and Earth.

Passage Problems

Figure 11.18 shows a demonstration gyroscope, consisting of a solid disk mounted on a shaft. The disk spins about the shaft on essentially frictionless bearings. The shaft is mounted on a stand so it's free to pivot both horizontally and vertically. A weight at the far end of the shaft balances the disk, so in the configuration shown there's no torque on the system. An arrowhead mounted on the disk end of the shaft indicates the direction of the disk's angular velocity.

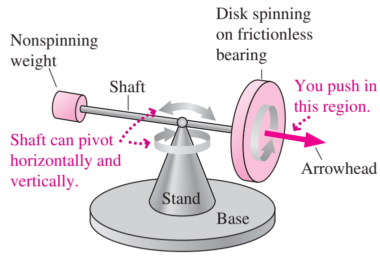


FIGURE 11.18 A gyroscope (Passage Problems 55–58)

55. If you push on the shaft between the arrowhead and the disk, pushing horizontally away from you (that is, into the page in Fig. 11.18), the arrowhead end of the shaft will move
    a. away from you (i.e., into the page).
    b. toward you (i.e., out of the page).
    c. downward.
    d. upward.
56. If you push on the shaft between the arrowhead and the disk, pushing directly upward on the bottom of the shaft, the arrowhead end of the shaft will move
    a. away from you (i.e., into the page).
    b. toward you (i.e., out of the page).
    c. downward.
    d. upward.
57. If an additional weight is hung on the left end of the shaft, the arrowhead will
    a. pivot upward until the weighted end of the shaft hits the base.
    b. pivot downward until the arrowhead hits the base.
    c. precess counterclockwise when viewed from above.
    d. precess clockwise when viewed from above.
58. If the system is precessing, and only the disk's rotation rate is increased, the precession rate will
    a. decrease.
    b. increase.
    c. stay the same.
    d. become zero.

Answers to Chapter Questions

Answer to Chapter Opening Question

The rotation axis precesses—changes orientation—over a 26,000-year cycle. This alters the relation between sunlight intensity and seasons, triggering ice ages.

Answers to GOT IT? Questions

11.1. (a) $\vec{\tau}_3$; (b) $\vec{\tau}_5$; (c) $\vec{\tau}_1$; (d) $\vec{\tau}_4$

11.2. (a) You'll go clockwise to keep total angular momentum at zero. (b) Total angular momentum remains unchanged at zero, so now you'll spin counterclockwise but at the same rate, assuming the wheel hasn't slowed.

11.3. (a)

12 Static Equilibrium

The Alamillo Bridge in Seville, Spain, is the work of architect Santiago Calatrava. What conditions must be met to ensure the stability of this dramatic structure?

New Concepts, New Skills

By the end of this chapter you should be able to

- Describe the two conditions necessary for a system to be in static equilibrium (12.1).
- Calculate the forces and torques necessary to ensure static equilibrium (12.3).
- Determine whether an equilibrium is stable or unstable (12.4).

Connecting Your Knowledge

- This chapter draws on Newton's second law as applied in Chapter 5, including the concept of force and the addition of force vectors to determine the net force (5.1–5.3).
- This chapter is also based on the concepts of torque and the rotational analog of Newton's second law (10.2, 10.3), and the idea of torque as a vector (11.2).
- The material here is a special case of what you've learned before, now with zero net force and torque in Newton's law and its rotational analog.

Architect Santiago Calatrava envisioned the boldly improbable bridge shown above. But it took engineers to make sure that the bridge would be stable in the face of what looks like an obvious tendency to topple to the left. The key to the engineers' success is **static equilibrium**—the condition in which a structure or system experiences neither a net force nor a net torque. Engineers use the principles of static equilibrium to design buildings, bridges, and aircraft. Scientists apply equilibrium principles at scales from molecular to astrophysical. Here we explore the conditions for static equilibrium required by the laws of physics.

12.1 Conditions for Equilibrium

A body is in **equilibrium** when the net external force and torque on it are both zero. In the special case when the body is also at rest, it's in **static equilibrium**. Systems in static equilibrium include not only engineered structures but also trees, molecules, and even your bones and muscles when you're at rest.

We can write the conditions for static equilibrium mathematically by setting the sums of all the external forces and torques both to zero:

$$\sum \vec{F}_i = \vec{0} \qquad (12.1)$$

and

$$\sum \vec{\tau}_i = \sum (\vec{r}_i \times \vec{F}_i) = \vec{0} \qquad (12.2)$$

Here the subscripts i label the forces $\vec{F}$ acting on an object, the positions $\vec{r}$ of the force-application points, and the associated torques $\vec{\tau}$.

In Chapters 10 and 11, we noted that torque depends on the choice of a rotation axis. Actually, the issue is not so much an axis but a single point—the point of origin of the vectors $\vec{r}$ that enter the expression $\vec{\tau} = \vec{r} \times \vec{F}$. In this chapter, where we have objects in equilibrium so they aren't rotating, we'll talk of this "pivot point" rather than a rotation axis. So the torque $\vec{\tau} = \vec{r} \times \vec{F}$ depends on the choice of pivot point. Then there seems to be an ambiguity in Equation 12.2, since we haven't specified a pivot point.

For an object to be in static equilibrium it can't rotate about *any* point, so Equation 12.2 must hold no matter what point we choose. Must we then check every possible point? Fortunately, no. If the first equilibrium condition holds—that is, if the net force on an object is zero—and if the net torque about *some* point is zero, then the net torque about *any other* point is also zero. Problem 53 leads you through the proof of this statement.

In solving equilibrium problems, we're thus free to choose any convenient point about which to evaluate the torques. An appropriate choice is often the application point of one of the forces; then $\vec{r} = \vec{0}$ for that force, and the associated torque $\vec{r} \times \vec{F}$ is zero. This leaves Equation 12.2 with one term fewer than it would otherwise have.

EXAMPLE 12.1 Choosing the Pivot: A Drawbridge

The raised span of the drawbridge shown in Fig. 12.1*a* has its 11,000-kg mass distributed uniformly over its 14-m length. Find the magnitude of the tension in the supporting cable.

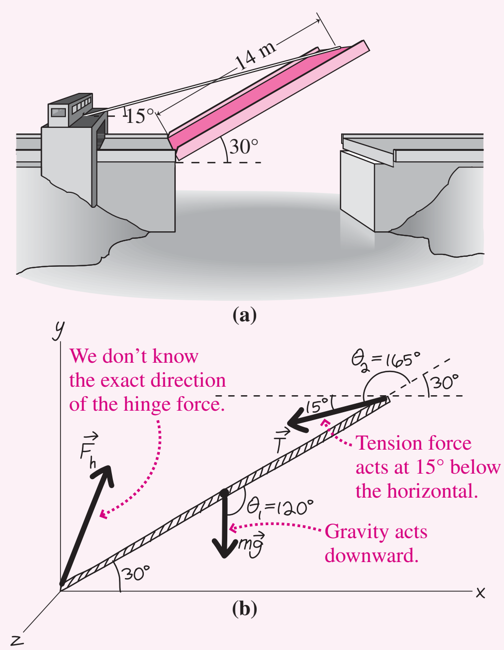


FIGURE 12.1 (a) A drawbridge. (b) Our sketch showing forces supporting the bridge.

INTERPRET Because the drawbridge is at rest, it's in static equilibrium.

DEVELOP Here we'll demonstrate how a sensible choice of the pivot point can make solving static-equilibrium problems easier. Figure 12.1*b* is a simplified diagram of the bridge, showing the three forces acting on it. These forces must satisfy both Equations 12.1 and 12.2, but we aren't asked about the hinge force $\vec{F}_h$, so it makes sense to choose the pivot at the hinge. We can then focus on Equation 12.2, $\sum \vec{\tau}_i = \vec{0}$, in which the only torques are due to gravity and tension. Gravity acts at the center of mass, half the bridge length L from the pivot (we'll prove this shortly). Therefore, it exerts a torque $\tau_g = -(L/2)mg \sin\theta_1$, where θ_1 is the angle between the gravitational force and a vector from the pivot. This torque is into the page, or in the negative z-direction—hence the negative sign. Similarly, the tension force, applied at the full length L, exerts a torque $\tau_T = LT \sin\theta_2$. Equation 12.2 then becomes

$$-\frac{L}{2}mg \sin\theta_1 + LT \sin\theta_2 = 0$$

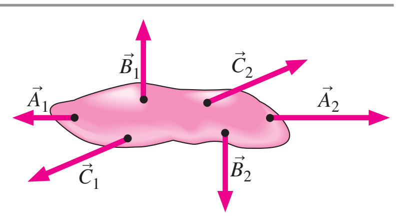
EVALUATE We solve for the tension T:

$$T = \frac{mg \sin\theta_1}{2 \sin\theta_2} = \frac{(11{,}000\ \text{kg})(9.8\ \text{m/s}^2)(\sin 120°)}{(2)(\sin 165°)} = 180\ \text{kN}$$

ASSESS This tension force is considerably larger than the approximately 110-kN weight of the bridge because the tension acts at a small angle to produce a torque that balances the torque due to gravity.

The point of this example is that a wise choice of the pivot point can eliminate a lot of work—in this case, allowing us to solve the problem using only Equation 12.2. If we had chosen a different pivot, then the force F_h would have appeared in the torque equation, and we would have had to eliminate it using the force equation, Equation 12.1 (see Exercise 15). ■

GOT IT? 12.1 The figure shows three pairs of forces acting on an object. Which pair, acting as the *only* forces on the object, results in static equilibrium? Explain why the others don't.



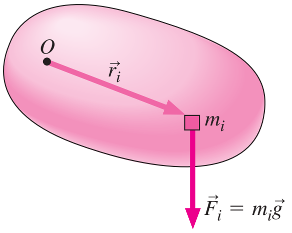


FIGURE 12.2 The gravitational force on the mass element m_i produces a torque about point O.

12.2 Center of Gravity

In Fig. 12.1*b* we drew the gravitational force acting at the center of mass of the bridge. That seems sensible, but is it correct? After all, gravity acts on all parts of an object. How do we know that the resulting torque is equivalent to the torque due to a single force acting at the center of mass? To see that it is, consider the gravitational forces on all parts of an object of mass M. The vector sum of those forces is $M\vec{g}$, but what about the torques? Figure 12.2 shows the ingredients we need to calculate the torque $\vec{\tau} = \vec{r} \times \vec{F}$ associated with one mass element; summing gives the total torque:

$$\vec{\tau} = \sum \vec{r}_i \times \vec{F}_i = \sum \vec{r}_i \times m_i\vec{g} = \left(\sum m_i\vec{r}_i\right) \times \vec{g}$$

We can rewrite this equation by multiplying the right-hand side by M/M, with M the total mass:

$$\vec{\tau} = \left(\frac{\sum m_i\vec{r}_i}{M}\right) \times M\vec{g}$$

The term in parentheses is the position of the center of mass (Section 9.1), and the right-hand term is the total weight. Therefore, the net torque on the body due to gravity is that of the gravitational force $M\vec{g}$ acting at the center of mass. In general, the point at which the gravitational force seems to act is called the **center of gravity**. We've just proven an important point: **The center of gravity coincides with the center of mass when the gravitational field is uniform.**

CONCEPTUAL EXAMPLE 12.1 Finding the Center of Gravity

Explain how you can find an object's center of gravity by suspending it from a string.

EVALUATE Suspend an object from a string and it will quickly come to equilibrium, as shown in Figs. 12.3*a, b*. In equilibrium there's no torque on the object and so, as Fig. 12.3*b* shows, its center of gravity (CG) must be directly below the suspension point. So far all we know is that the CG lies on a vertical line extending from the suspension point. But two intersecting lines determine a point, so all we have to do is suspend the object from a *different* point. In its new equilibrium, the CG again lies on a vertical line from the suspension point. Where the two lines meet is the center of gravity (Fig. 12.3*c*).

ASSESS Here's a quick, easy, and practical way to find the center of gravity—at least for two-dimensional objects.

MAKING THE CONNECTION Do the experiment! Determine the center of gravity of an isosceles triangle made from material of uniform density.

EVALUATE Cut a triangle of cardboard or wood and follow the procedure described here. You should get good agreement with Example 9.3: The triangle's CG (which is the same as its center of mass) lies two-thirds of the way from the apex to the base.

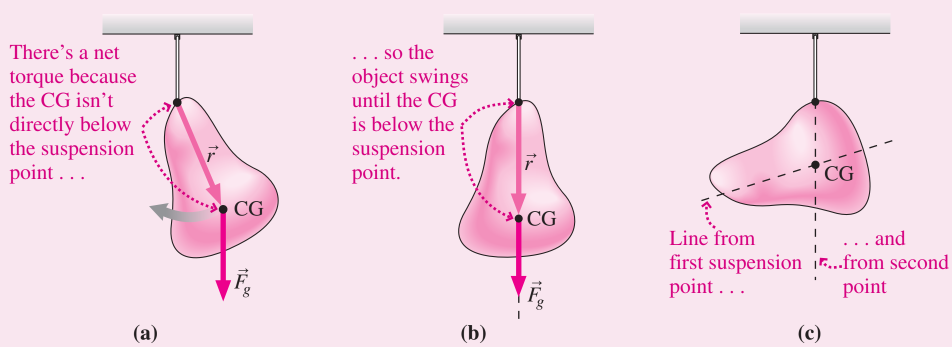
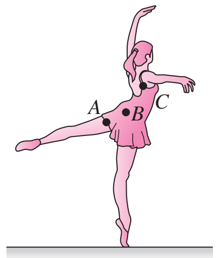


FIGURE 12.3 Finding the center of gravity.

GOT IT? 12.2 The dancer in the figure is balanced; that is, she's in static equilibrium. Which of the three lettered points could be her center of gravity?



12.3 Examples of Static Equilibrium

It's frequently the case that all the forces acting on a system lie in a plane, so Equation 12.1—the statement that there's no net force in static equilibrium—becomes two equations for the two force components in that plane. And with all the forces in a plane, the torques are all at right angles to that plane, so Equation 12.2—the statement that there's no net torque—becomes a single equation. We'll restrict ourselves to such cases in which the conditions for static equilibrium reduce to three scalar equations. Sometimes, as in Example 12.1, the torque equation alone will give what we're looking for, but usually that's not the case.

Solving static-equilibrium problems is much like solving Newton's law problems; after all, the equations for static equilibrium are Newton's law and its rotational analog, both with acceleration set to zero. Here we adapt our Newton's law strategy from Chapter 4 to problems of static equilibrium. The examples that follow illustrate the use of this strategy.

PROBLEM-SOLVING STRATEGY 12.1 Static-Equilibrium Problems

INTERPRET Interpret the problem to be sure it's about static equilibrium, and identify the object that you want to keep in equilibrium. Next, identify all the forces acting on the object.

DEVELOP Draw a diagram showing the forces acting on your object. Since you've got torques to calculate, it's important to show *where* each force is applied. So don't represent your object as a single dot but show it semirealistically with the force-application points. This is a static-equilibrium problem, so Equations 12.1, $\sum \vec{F}_i = \vec{0}$, and 12.2, $\sum \vec{\tau}_i = \vec{0}$, apply. Develop your solution by choosing a coordinate system that will help resolve the force vectors into components *and* choose its origin at an appropriate pivot point—usually the application point of one of the forces. In some problems the unknown is itself a force; in that case, draw a force vector that you think is appropriate and let the algebra take care of the signs and angles.

EVALUATE At this point the physics is done, and you're ready to evaluate your answer. Begin by writing the two components of Equation 12.1 in your coordinate system. Then evaluate the torques about your chosen origin, and write Equation 12.2 as a single scalar equation showing that the torques sum to zero. Now you've got three equations, and you're ready to solve. Since there are three equations, there will be three unknowns even if you're asked for only one final answer. You can use the equations to eliminate the unknowns you don't want.

ASSESS Assess your solution to see whether it makes sense. Are the numbers reasonable? Do the directions of forces and torques make sense in the context of static equilibrium? What happens in special cases—for example, when a force or mass goes to zero or gets very large, or for special values of angles among the various vectors?

EXAMPLE 12.2 Static Equilibrium: Ladder Safety

A ladder of mass m and length L is leaning against a wall, as shown in Fig. 12.4*a* (next page). The wall is frictionless, and the coefficient of static friction between ladder and ground is μ. Find an expression for the minimum angle ϕ at which the ladder can lean without slipping.

INTERPRET This problem is about static equilibrium, and the ladder is the object we want to keep in equilibrium. We identify four forces acting on the ladder: gravity, normal forces from the floor and wall, and static friction from the ground.

(continued)

DEVELOP Figure 12.4*b* shows the four forces and the unknown angle ϕ. We'll get the minimum angle when static friction is greatest: $f_s = \mu n_1$. Since we're dealing with static equilibrium, Equations 12.1 and 12.2 apply. In a horizontal/vertical coordinate system, Equation 12.1 has the two components:

$$\text{Force, } x: \qquad \mu n_1 - n_2 = 0$$

$$\text{Force, } y: \qquad n_1 - mg = 0$$

Now for the torques: If we choose the bottom of the ladder as the pivot, we eliminate two forces from the torque equation. That leaves only the gravitational torque and the torque due to the wall's normal force; both involve the unknown angle ϕ. The gravitational torque is into the page, or the negative z-direction, so it's given by $\tau_g = -(L/2)mg\sin(90° - \phi) = -(L/2)mg\cos\phi$. The torque due to the wall is out of the page: $\tau_w = Ln_2\sin(180° - \phi) = Ln_2\sin\phi$. We used two trig identities here: $\sin(90° - \phi) = \cos\phi$ and $\sin(180° - \phi) = \sin\phi$. Then Equation 12.2 becomes

$$\text{Torque:} \qquad Ln_2\sin\phi - \frac{L}{2}mg\cos\phi = 0$$

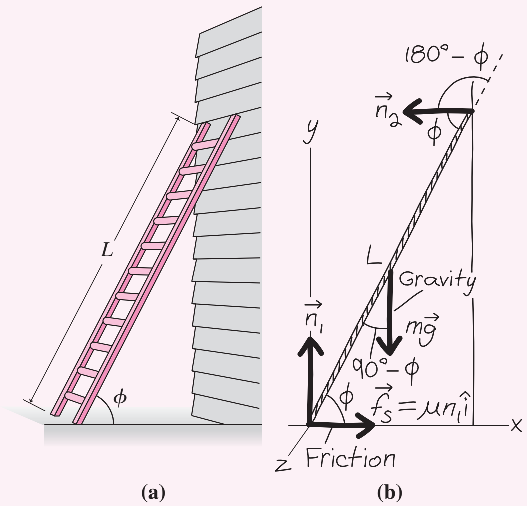


FIGURE 12.4 (a) At what angle will the ladder slip? (b) Our sketch.

EVALUATE We have three unknowns: n_1, n_2, and ϕ. The y-component of the force equation gives $n_1 = mg$, showing that the ground supports the ladder's weight. Using this result in the x-component of the force equation gives $n_2 = \mu mg$. Then the torque equation becomes $\mu mgL\sin\phi - (L/2)mg\cos\phi = 0$. The term mgL cancels, giving $\mu\sin\phi - \frac{1}{2}\cos\phi = 0$. We solve for the unknown angle ϕ by forming its tangent:

$$\tan\phi = \frac{\sin\phi}{\cos\phi} = \frac{1}{2\mu}$$

ASSESS Make sense? The larger the frictional coefficient, the more horizontal force holding the ladder in place, and the smaller the angle at which it can safely lean. On the other hand, a very small frictional coefficient makes for a very large tangent—meaning the angle approaches 90°. With no friction, you could stand the ladder only if it were strictly vertical. A word of caution: We worked this example with no one on the ladder. With the extra weight of a person, especially near the top, the minimum safe angle will be a lot larger. Problem 31 explores this situation. ■

EXAMPLE 12.3 Static Equilibrium: In the Body

Figure 12.5*a* shows a human arm holding a pumpkin, with masses and distances marked. Find the magnitudes of the biceps tension and the contact force at the elbow joint.

INTERPRET This problem is about static equilibrium, with the arm/pumpkin being the object in equilibrium. We identify four forces: the weights of the arm and the pumpkin, the biceps tension, and the contact force at the elbow.

DEVELOP Figure 12.5*b* shows the four forces, including the elbow contact force $\vec{F}_c$, whose exact direction we don't know. We can read the horizontal and vertical components of Equation 12.1, the force balance equation, from the diagram:

$$\text{Force, } x: \qquad F_{cx} - T\cos\theta = 0$$

$$\text{Force, } y: \qquad T\sin\theta - F_{cy} - mg - Mg = 0$$

Choosing the elbow as the pivot eliminates the contact force from the torque equation, giving

$$\text{Torque:} \qquad x_1T\sin\theta - x_2mg - x_3Mg = 0$$

where the x values are the coordinates of the three force-application points.

EVALUATE We begin by solving the torque equation for the biceps tension:

$$T = \frac{(x_2m + x_3M)g}{x_1\sin\theta} = 500\text{ N}$$

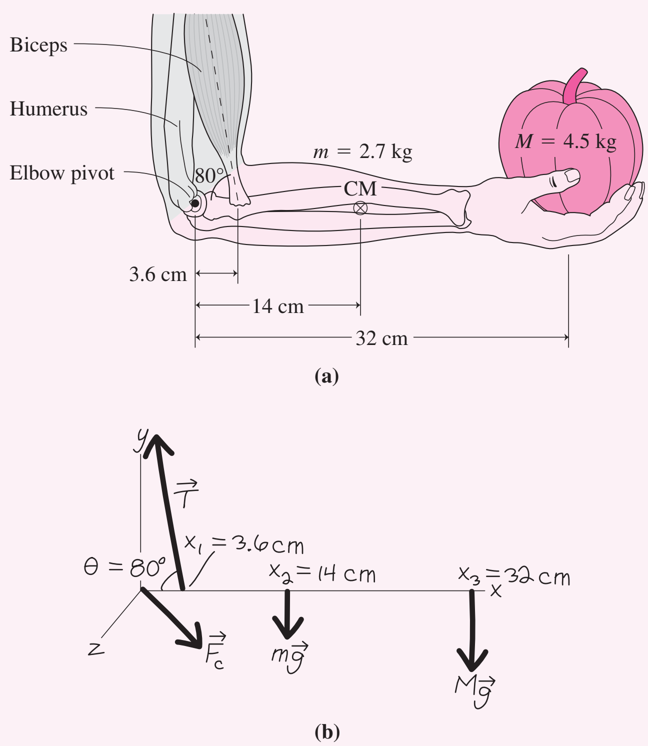


FIGURE 12.5 (a) Holding a pumpkin. (b) Our sketch.

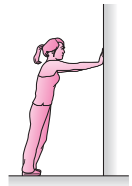
where we used the values in Fig. 12.5 to evaluate the numerical answer. The force equations then give the components of the elbow contact force:

$$F_{cx} = T\cos\theta = 87\text{ N} \quad \text{and} \quad F_{cy} = T\sin\theta - (m + M)g = 420\text{ N}$$

The magnitude of the elbow contact force then becomes $F_c = \sqrt{87^2 + 420^2}\text{ N} = 430\text{ N}$.

ASSESS These answers may seem huge—both the biceps tension and the elbow contact force are roughly ten times the weight of the pumpkin, on the order of 100 pounds. But that's because the biceps muscle is attached so close to the elbow; given this small lever arm, it takes a large force to balance the torque from the weight of pumpkin and arm. This example shows that the human body routinely experiences forces substantially greater than the weights of objects it's lifting. ■

GOT IT? 12.3 The figure shows a person in static equilibrium leaning against a wall. Which of the following must be true: (a) There must be a frictional force at the wall but not necessarily at the floor. (b) There must be a frictional force at the floor but not necessarily at the wall. (c) There must be frictional forces at both floor and wall.



12.4 Stability

If a body is disturbed from equilibrium, it generally experiences nonzero torques or forces that cause it to accelerate. Figure 12.6 shows two very different possibilities for the subsequent motion of two cones initially in equilibrium. Tip the cone on the left slightly, and a torque develops that brings it quickly back to equilibrium. Tip the cone on the right, and over it goes. The torque arising from even a slight displacement swings the cone permanently away from its original equilibrium. The former situation is an example of **stable equilibrium**, the latter of **unstable equilibrium**. Nearly all the equilibria we encounter in nature are stable, since a body in unstable equilibrium won't remain so. The slightest disturbance will set it in motion, bringing it to a very different equilibrium state.

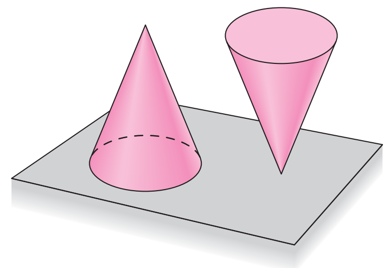


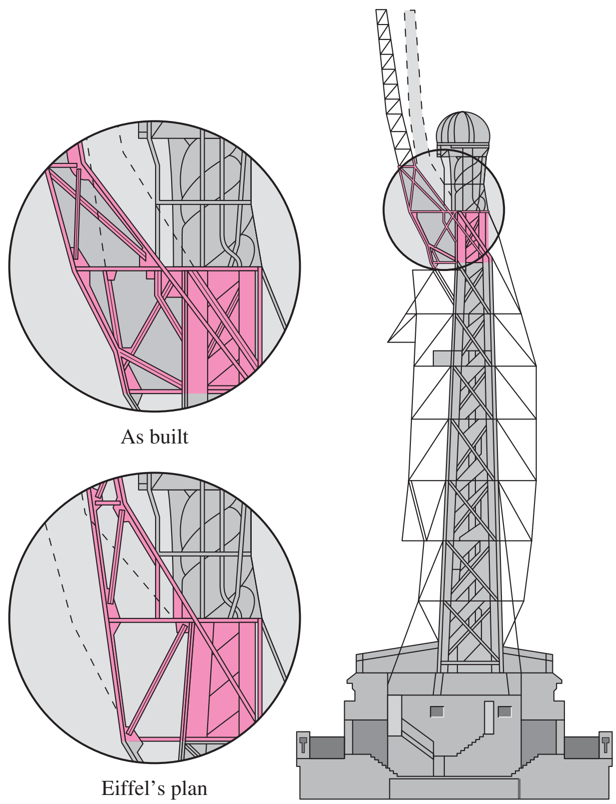
FIGURE 12.6 Stable (*left*) and unstable (*right*) equilibria.

APPLICATION **Restoring the Statue of Liberty**

The Statue of Liberty, France's famous gift to the United States, was shipped to New York in 300 pieces, assembled, and dedicated in 1886. Liberty was the artistic work of French sculptor Frédéric-Auguste Bartholdi, who suggested that his creation should last as long as Egypt's pyramids. But after only 100 years, Liberty was ready for a major renovation. Corrosive air pollution had taken its toll, along with a chemical reaction between the statue's iron frame and its copper skin. And an assembly error had resulted in excessive torques on the statue's structural members.

Sculptor Bartholdi was no engineer, and without the work of French engineer Gustave Eiffel—designer of the famous tower—the statue could not have maintained itself in static equilibrium. Eiffel designed an inner skeleton of iron to provide the forces necessary to counteract the forces and torques associated with the weights of the statue's components and also with the wind. But Liberty's head and upper arm were mounted contrary to Eiffel's plans, probably as a result of a conscious aesthetic decision. The figure shows how the incorrect arm mounting—a two-foot offset and a correspondingly greater angle—resulted in excessive torques about the shoulder.

Liberty underwent extensive renovations during its centennial year. For historical integrity, renovators chose not to correct the original assembly error. Instead, they reinforced the support structure so it would withstand better the excess forces and torques.



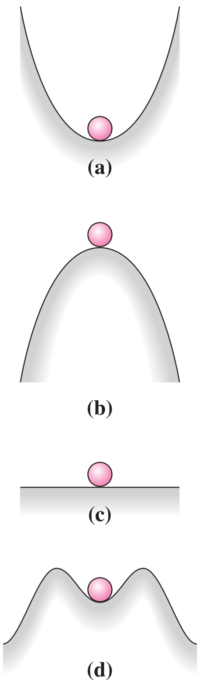


FIGURE 12.7 (a) Stable, (b) unstable, (c) neutrally stable, and (d) metastable equlilibria.

Figure 12.7 shows a ball in four different equilibrium situations. Clearly (a) is stable and (b) is unstable. Situation (c) is neither stable nor unstable; it's called **neutrally stable**. But what about (d)? For small disturbances, the ball will return to its original state, so the equilibrium is stable. But for larger disturbances—large enough to push the ball over the highest points on the hill—it's unstable. Such an equilibrium is **conditionally stable** or **metastable**.

A system disturbed from stable equilibrium may not return immediately. In Fig. 12.7*a*, for example, displacing the ball results in its rolling back and forth. Eventually friction dissipates its energy, and it comes to rest at equilibrium. Back-and-forth motion is common to many systems—from nuclei and atoms to skyscrapers and bridges—that are displaced from stable equilibrium. Such motion is the topic of the next chapter.

Stability is closely associated with potential energy. Because gravitational potential energy is directly proportional to height, the shapes of the hills and valleys in Fig. 12.7 are in fact potential-energy curves. In all cases of equilibrium, the ball is at a minimum or maximum of the potential-energy curve—at a place where the force (that is, the derivative of potential energy with respect to position) is zero. For the stable and metastable equilibria, the potential energy at equilibrium is a local minimum. A deviation from equilibrium requires that work be done against the force that tends to restore the ball to equilibrium. The unstable equilibrium, in contrast, occurs at a maximum in potential energy. Here, a deviation from equilibrium results in lower potential energy and in a force that accelerates the ball farther from equilibrium. For the neutrally stable equilibrium, there's no change in potential energy as the ball moves; consequently it experiences no force. Figure 12.8 gives another example of equilibria in the context of potential energy.

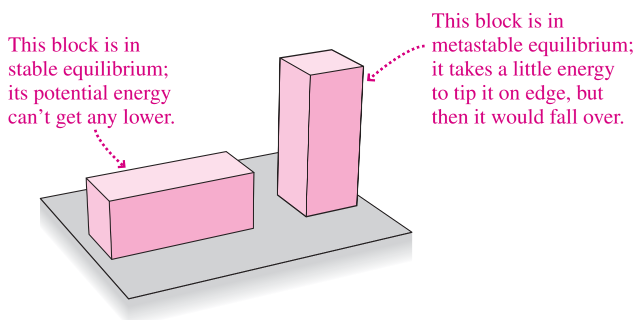


FIGURE 12.8 Identical blocks in stable and metastable equilibria.

We can sum up our understanding of equilibrium and potential energy in two simple mathematical statements. First, the force must be zero; that requires a local maximum or minimum in the potential energy:

$$\frac{dU}{dx} = 0 \qquad \text{(equilibrium condition)} \tag{12.3}$$

where U is the potential energy of a system and x is a variable describing the system's configuration. For the simple systems we've been considering, x measures the position or orientation of an object, but for more complicated systems, it could be another quantity such as the system's volume or even its composition. For a stable equilibrium, we require a local minimum, so the potential-energy curve is concave upward. (See Tactics 12.1 to review the relevant calculus.) Mathematically,

$$\frac{d^2U}{dx^2} > 0 \qquad \text{(stable equilibrium)} \tag{12.4}$$

This condition applies to metastable equilibria as well because they're *locally* stable. In contrast, unstable equilibrium occurs where the potential energy has a local maximum, or

$$\frac{d^2U}{dx^2} < 0 \qquad (\text{unstable equilibrium}) \tag{12.5}$$

The intermediate case $d^2U/dx^2 = 0$ corresponds to neutral stability.

TACTICS 12.1 **Finding Maxima and Minima**

1. Begin by sketching a plot of the function, which will give a visual check for your numerical answers.
2. Next take the function's first derivative and set it to zero. As Fig. 12.7 suggests, a hill (maximum) or valley (minimum) is level right at its top or bottom. So by setting the first derivative to zero, you're requiring that its slope be zero and therefore requiring the function to be at a maximum or minimum.
3. Find the sign of the function's second derivative at the points where you found the first derivative is zero. Your sketch should show this; where the curve is concave upward, as in Figs. 12.7*a* and *d*, the second derivative is positive and the point is a minimum. Where it's concave downward, as in Fig. 12.7*b*, d^2U/dt^2 is negative and you've got a maximum. If it wasn't obvious how to sketch the function, you can use calculus to determine the second derivative and then find its sign at the equilibrium points.
4. Check that the values you found for maxima and minima agree with your plot of the function.

EXAMPLE 12.4 Stability Analysis: Semiconductor Engineering

Physicists develop a new semiconductor device in which an electron's potential energy is given by $U(x) = ax^2 - bx^4$, where x is the electron's position in nm, U is its potential energy in aJ $(10^{-18}\,\text{J})$, and constants a and b are 8 aJ/nm^2 and 1 aJ/nm^4, respectively. Find the equilibrium positions for the electron, and describe their stability.

INTERPRET This problem is about stability in the context of a given potential-energy function. We're interested in the electron, and we're asked to find the values of x where it's in equilibrium and then examine their stability.

DEVELOP The potential-energy curve gives us insight into this problem, so we've drawn it by plotting the function $U(x)$ in Fig. 12.9. Equation 12.3, $dU/dx = 0$, determines the equilibria, while Equations 12.4, $d^2U/dx^2 > 0$, and 12.5, $d^2U/dx^2 < 0$, determine the stability. Our plan is first to find the equilibrium positions using Equation 12.3 and then to examine their stability.

EVALUATE Equation 12.3 states that equilibria occur where the potential energy has a maximum or minimum—that is, where its derivative is zero. Taking the derivative of U and setting it to zero gives

$$0 = \frac{dU}{dx} = 2ax - 4bx^3 = 2x(a - 2bx^2)$$

This equation has solutions when $x = 0$ and when $a = 2bx^2$ or $x = \pm\sqrt{a/2b} = \pm 2$ nm. We could take second derivatives to evaluate the stability, but the situation is evident from our plot: $x = 0$ lies at a local minimum of the potential-energy curve, so this equilibrium is metastable. The other two equilibria, at maxima of U, are unstable.

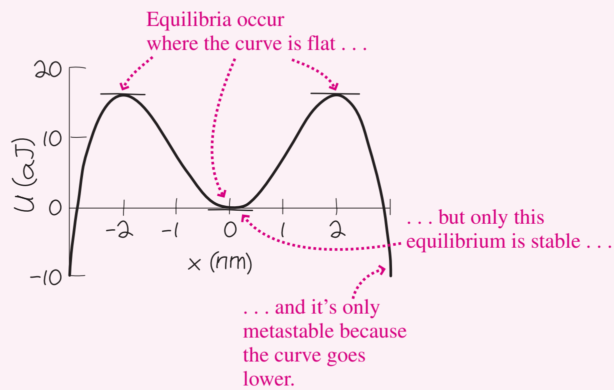


FIGURE 12.9 Our sketch of the potential-energy curve for Example 12.4.

ASSESS Do our numerical answers make sense? Yes: You can see that the potential-energy curve has zero slope at the points $x = -2$ nm, $x = 0$, and $x = 2$ nm, so we've found all the equilibria. Note that the equilibrium at $x = 0$ is only metastable; given enough energy, an electron disturbed from this position could make it all the way over the peaks and never return to $x = 0$. ■

Stability considerations apply to the overall arrangements of matter. A mixture of hydrogen and oxygen, for example, is in metastable equilibrium at room temperature. Lighting a match puts some atoms over the maxima in their potential-energy curves, at which point they rearrange into a state of lower potential energy—the state we call H_2O. Similarly, a uranium nucleus is at a local minimum of its potential-energy curve, and a little excess energy can result in its splitting into two smaller nuclei whose total potential energy is much lower. That transition from a less stable to a more stable equilibrium describes the basic physics of nuclear fission.

Potential-energy curves for complex structures like molecules or skyscrapers can't be described fully with one-dimensional graphs. If potential energy varies in different ways when the structure is altered in different directions, then in order to determine stability we need to consider all possible ways potential energy might vary. For example, a snowball sitting on a mountain pass—or any other system with a saddle-shaped potential-energy curve—is stable against displacements in one direction but not another (Fig. 12.10). Stability analysis of complex physical systems, ranging from nuclei and molecules to bridges and buildings and machinery, and on to stars and galaxies, is an important part of contemporary work in engineering and science.

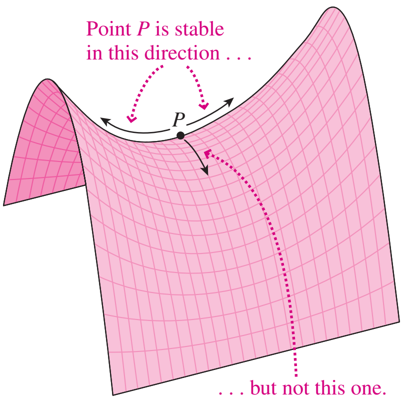
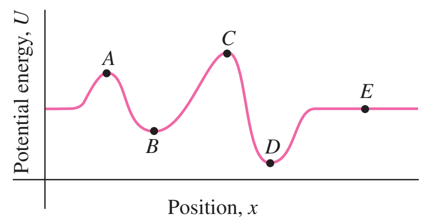


FIGURE 12.10 Equilibrium on a saddle-shaped potential-energy curve.

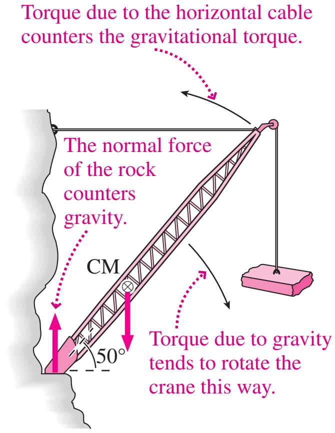
GOT IT? 12.4 Which of the labeled points in the figure are stable, metastable, unstable, or neutrally stable equilibria?



CHAPTER 12 SUMMARY

Big Picture

The big idea here is **static equilibrium**—the state in which a system at rest remains at rest because there's no net force to accelerate it and no net torque to start it rotating. An equilibrium is stable if a disturbance of the system results in its returning to the original equilibrium state.



Key Concepts and Equations

Static equilibrium requires that there be no net force and no net torque on a system; mathematically:

$$\sum \vec{F}_i = \vec{0}$$

and

$$\sum \vec{\tau}_i = \sum \vec{r}_i \times \vec{F}_i = \vec{0}$$

where the sums include all the forces applied to the system. Solving an equilibrium problem involves identifying all the forces $\vec{F}_i$ acting on the system, choosing an appropriate origin about which to evaluate the torques, and requiring that forces and torques sum to zero.

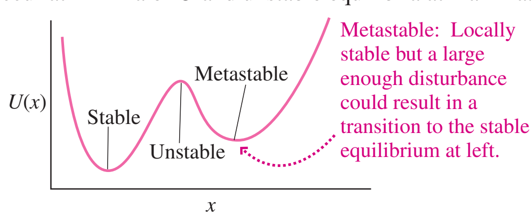
Equilibria occur where a system's potential energy $U(x)$ has a maximum or a minimum:

$$\frac{dU}{dx} = 0 \qquad \text{(equilibrium condition)}$$

$$\frac{d^2U}{dx^2} > 0 \qquad \text{(stable equilibrium)}$$

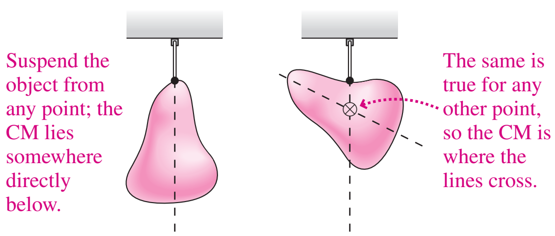
$$\frac{d^2U}{dx^2} < 0 \qquad \text{(unstable equilibrium)}$$

Stable equilibria occur at minima of U and unstable equilibria at maxima.

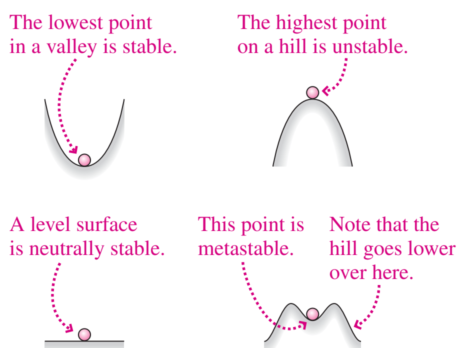


Applications

The **center of gravity** of a system is the point where the force of gravity appears to act. When the gravitational field is uniform over the system, the center of gravity coincides with the center of mass. This provides a handy way to locate the center of mass.



Four different types of equilibrium are **stable**, **unstable**, **neutrally stable**, and **metastable**.



MP For homework assigned on MasteringPhysics, go to www.masteringphysics.com | BIO Biology and/or medical-related problems C Computer problems

For Thought and Discussion

1. Give an example of an object on which the net force is zero, but that isn't in static equilibrium.
2. Give an example of an object on which the net torque about the center of gravity is zero, but that isn't in static equilibrium.
3. The best way to lift a heavy weight is to squat with your back vertical, rather than to lean over. Why?
4. Pregnant women often assume a posture with their shoulders held far back from their normal position. Why?
5. When you carry a bucket of water with one hand, you instinctively extend your opposite arm. Why?
6. Is a ladder more likely to slip when you stand near the top or the bottom? Explain.
7. How does a heavy keel help keep a boat from tipping over?
8. In addition to the wings, most airplanes have a smaller set of horizontal surfaces near the tail. Why? What does this suggest about the airplane's center of gravity?
9. Does choosing a pivot point in an equilibrium problem mean that something is necessarily going to rotate about that point?
10. If you take the pivot point at the application point of one force in a static-equilibrium problem, that force doesn't enter the torque equation. Does that make the force irrelevant to the problem? Explain.
11. You're hanging a heavy picture on a wall, using wire attached to the top corners of the picture. Is the wire more likely to break if you run it tightly between the corners or if you give it some slack? Explain.
12. A short dog and a tall person are standing on a slope. If the slope angle increases, which will fall over first? Why?
13. A stiltwalker is standing motionless on one stilt. What can you say about the location of the stiltwalker's center of mass?

Exercises and Problems

Exercises

Section 12.1 Conditions for Equilibrium

14. A body is subject to three forces: $\vec{F}_1 = 2\hat{\imath} + 2\hat{\jmath}$ N, applied at the point $x = 2$ m, $y = 0$ m; $\vec{F}_2 = -2\hat{\imath} - 3\hat{\jmath}$ N, applied at $x = -1$ m, $y = 0$ m; and $\vec{F}_3 = 1\hat{\jmath}$ N, applied at $x = -7$ m, $y = 1$ m. Show that (a) the net force and (b) the net torque about the origin are both zero.
15. To demonstrate that the choice of pivot point doesn't matter, show that the torques in Exercise 14 sum to zero when evaluated about the points (3 m, 2 m) and (−7 m, 1 m).
16. In Fig. 12.11 the forces shown all have the same magnitude F. For each case shown, is it possible to place a third force so as to meet both conditions for static equilibrium? If so, specify the force and a suitable application point. If not, why not?

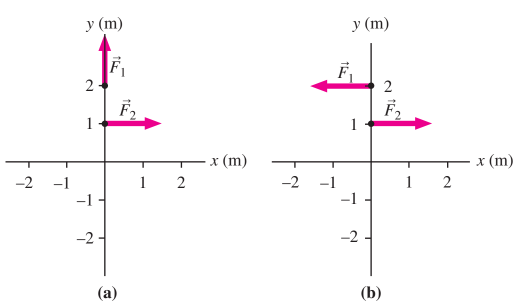


FIGURE 12.11 Exercise 16

Section 12.2 Center of Gravity

17. Figure 12.12a shows a thin, uniform square plate of mass m and side L. The plate is in a vertical plane. Find the magnitude of the gravitational torque on the plate about each of the three points shown.

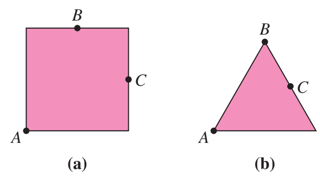


FIGURE 12.12 Exercises 17 and 18

18. Repeat the preceding problem for the equilateral triangle in Fig. 12.12b, which has side L.
19. A 23-m-long log of irregular cross section lies horizontally, supported by a wall at one end and a cable attached 4.0 m from the other end, as shown in Fig. 12.13. The log weighs 7.5 kN and the tension in the cable is 6.2 kN. Find the log's center of gravity.

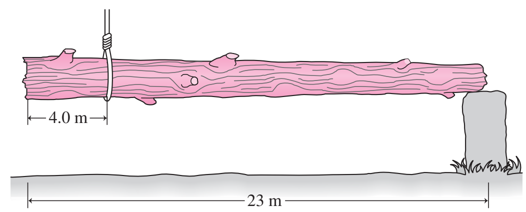


FIGURE 12.13 Exercise 19

Section 12.3 Examples of Static Equilibrium

20. A 60-kg uniform board 2.4 m long is supported by a pivot 80 cm from the left end and by a scale at the right end (Fig. 12.14). How far from the left end should a 40-kg child sit for the scale to read zero?

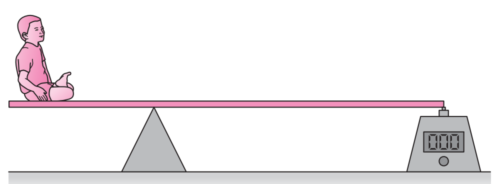


FIGURE 12.14 Exercises 20 and 21

21. Where should the child in Fig. 12.14 sit if the scale is to read (a) 100 N and (b) 300 N?
22. A 4.2-m-long beam is supported by a cable at its center. A 65-kg steelworker stands at one end of the beam. Where should a 190-kg bucket of concrete be suspended for the beam to be in static equilibrium?
23. Figure 12.15 shows how a scale with a capacity of only 250 N can be used to weigh a heavier person. The 3.4-kg board is 3.0 m

long and has uniform density. It's free to pivot about the end farthest from the scale. Assume that the beam remains essentially horizontal. What's the weight of a person standing 1.2 m from the pivot end if the scale reads 210 N?

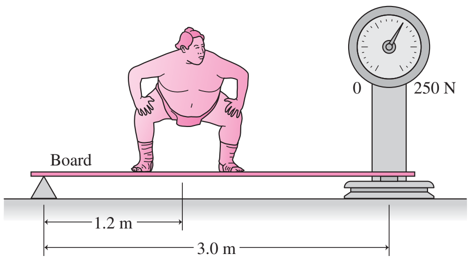


FIGURE 12.15 Exercise 23

Section 12.4 Stability

24. A portion of a roller-coaster track is described by $h = 0.94x - 0.010x^2$, where h and x are the height and horizontal position in meters. (a) Find a point where the roller-coaster car could be in static equilibrium on this track. (b) Is this equilibrium stable or unstable?
25. A particle's potential energy as a function of position is given by $U = 2x^3 - 2x^2 - 7x + 10$, with x in meters and U in joules. Find the positions of any stable and unstable equilibria.

Problems

26. You're a highway safety engineer, and you're asked to specify bolt sizes so the traffic signal in Fig. 12.16 won't fall over. The figure indicates the masses and positions of the structure's various parts. The structure is mounted with two bolts, located symmetrically about the vertical member's centerline, as shown. What tension force must the left-hand bolt be capable of withstanding?

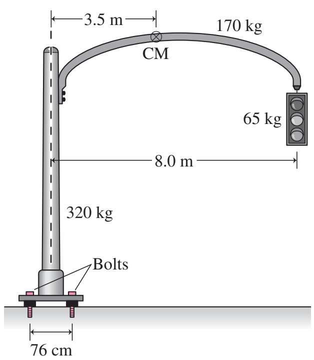


FIGURE 12.16 Problem 26

27. Figure 12.17a shows an outstretched arm with mass 4.2 kg. The arm is 56 cm long, and its center of gravity is 21 cm from the shoulder. The hand at the end of the arm holds a 6.0-kg mass. (a) Find the torque about the shoulder due to the weight of the arm and the 6.0-kg mass. (b) If the arm is held in equilibrium by the deltoid muscle, whose force on the arm acts below the horizontal at a point 18 cm from the shoulder joint (Fig. 12.17b), what's the force exerted by the muscle?

BIO

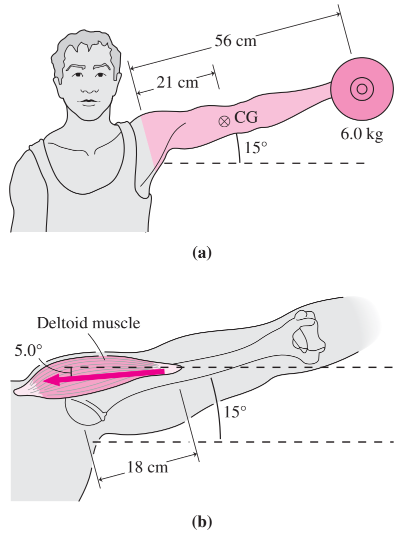


FIGURE 12.17 Problem 27

28. A uniform sphere of radius R is supported by a rope attached to a vertical wall, as shown in Fig. 12.18. The rope joins the sphere at a point where a continuation of the rope would intersect a horizontal line through the sphere's center a distance $\frac{1}{2}R$ beyond the center, as shown. What's the smallest possible value for the coefficient of friction between wall and sphere?

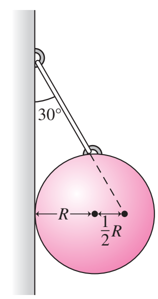


FIGURE 12.18 Problem 28

29. You work for a garden equipment company, and you're designing a new garden cart. Specifications to be listed include the horizontal force that must be applied to push the fully loaded cart (mass 55 kg, 60-cm-diameter wheels) up an abrupt 8.0-cm step, as shown in Fig. 12.19. What do you specify for the force?

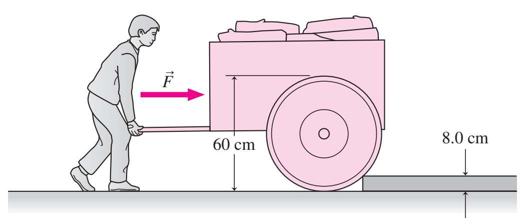


FIGURE 12.19 Problem 29

30. Figure 12.20 shows the foot and lower leg of a person standing on the ball of one foot. Three forces act to maintain this equilibrium: the tension force $\vec{T}$ in the Achilles tendon, the contact force

BIO

$\vec{F}_c$ at the ankle joint, and the normal force $\vec{n}$ of the ground that supports the person's weight. The person's mass is 70 kg, and the force-application points are as indicated in Fig. 12.20. Find the magnitude of (a) the tension in the Achilles tendon and (b) the contact force at the ankle joint.

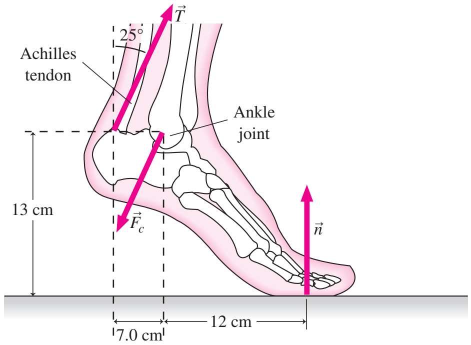


FIGURE 12.20 Problem 30

31. A uniform 5.0-kg ladder is leaning against a frictionless vertical wall, with which it makes a 15° angle. The coefficient of friction between ladder and ground is 0.26. Can a 65-kg person climb to the top of the ladder without it slipping? If not, how high can that person climb? If so, how massive a person would make the ladder slip?
32. The boom in the crane of Fig. 12.21 is free to pivot about point P and is supported by the cable attached halfway along its 18-m length. The cable passes over a pulley and is anchored at the back of the crane. The boom has mass 1700 kg distributed uniformly along its length, and the mass hanging from the boom is 2200 kg. The boom makes a 50° angle with the horizontal. Find the tension in the cable.

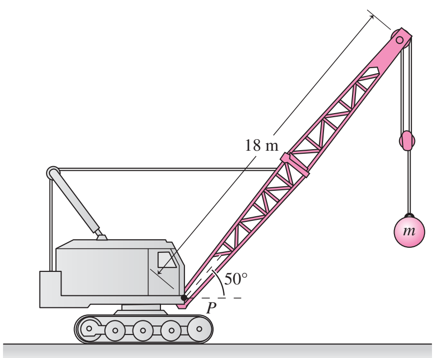


FIGURE 12.21 Problem 32

33. A uniform board of length L and weight W is suspended between two vertical walls by ropes of length $L/2$ each. When a weight w is placed on the left end of the board, it assumes the configuration shown in Fig. 12.22. Find the weight w in terms of the board weight W.

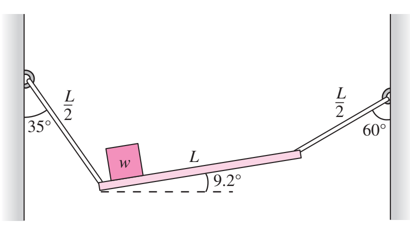


FIGURE 12.22 Problem 33

34. Figure 12.23 shows a 1250-kg car that has slipped over an embankment. People are trying to hold the car in place by pulling on a horizontal rope. The car's bottom is pivoted on the edge of the embankment, and its center of mass lies farther back, as shown. If the car makes a 34° angle with the horizontal, what force must the people apply to hold it in place?

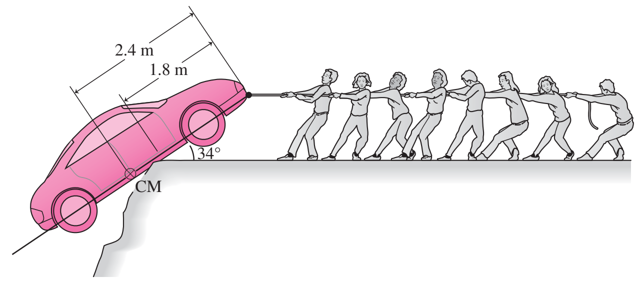


FIGURE 12.23 Problem 34

35. Repeat Example 12.2, now assuming that the coefficient of friction at the ground is μ_1 and at the wall is μ_2. Show that the minimum angle at which the ladder won't slip is now given by $\phi = \tan^{-1}[(1 - \mu_1\mu_2)/2\mu_1]$.
36. You are headwaiter at a new restaurant, and your boss asks you to hang a sign for her. You're to hang the sign, whose mass is 66 kg, in the configuration shown in Fig. 12.24. A uniform horizontal rod of mass 8.2 kg and length 2.3 m holds the sign. At one end the rod is attached to the wall by a pivot; at the other end it's supported by a cable that can withstand a maximum tension of 800 N. You're to determine the minimum height h above the pivot for anchoring the cable to the wall.

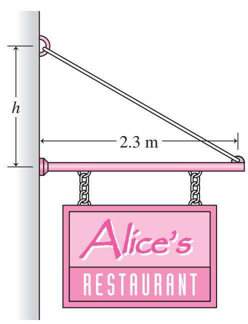


FIGURE 12.24 Problem 36

37. Climbers attempting to cross a stream place a 340-kg log against a vertical, frictionless ice cliff on the opposite side (Fig. 12.25). The log slopes up at 27° and its center of gravity is one-third of

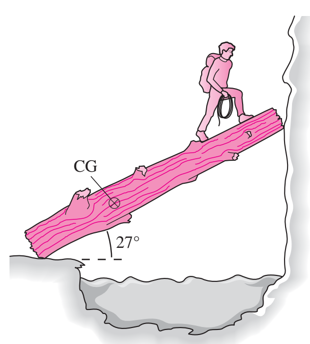


FIGURE 12.25 Problem 37

the way along its 6.3-m length. If the coefficient of friction between the left end of the log and the ground is 0.92, what's the maximum mass for a climber and pack to cross without the log slipping?

38. A crane in a marble quarry is mounted on the quarry's rock walls and is supporting a 2500-kg marble slab as shown in Fig. 12.26. The center of mass of the 830-kg boom is located one-third of the way from the pivot end of its 15-m length, as shown. Find the tension in the horizontal cable that supports the boom.

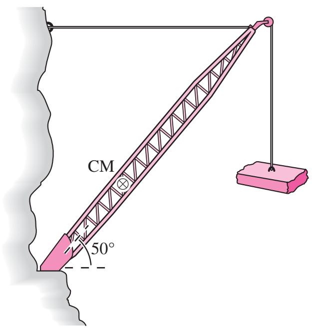


FIGURE 12.26 Problem 38

39. A uniform rectangular block is twice as long as it is wide. Letting θ be the angle that the long dimension makes with the horizontal, determine the angular positions of any static equilibria, and comment on their stability.

40. The potential energy as a function of position for a particle is given by

$$U(x) = U_0\left(\frac{x^3}{x_0^3} + a\frac{x^2}{x_0^2} + 4\frac{x}{x_0}\right)$$

where x_0 and a are constants. For what values of a will there be two static equilibria? Comment on the stability of these equilibria.

41. A cubical block rests on an inclined board with two sides parallel to the incline. The coefficient of static friction between block and board is 0.95. If the inclination angle of the board is increased, will the block first slide or first tip over?

42. A 160-kg highway sign of uniform density is 2.3 m wide and 1.4 m high. At one side it's secured to a pole with a single bolt, mounted a distance d from the top of the sign. The only other place where the sign contacts the pole is at its bottom corner. If the bolt can sustain a horizontal tension of 2.1 kN, what's the maximum permissible value for the distance d?

43. A 5.0-m-long ladder has mass 9.5 kg and is leaning against a frictionless wall, making a 66° angle with the horizontal. If the coefficient of friction between ladder and ground is 0.42, what's the mass of the heaviest person who can safely ascend to the top of the ladder? (The center of mass of the ladder is at its center.)

44. To what vertical height on the ladder in Problem 43 could a 95-kg person reach before the ladder starts to slip?

45. A uniform, solid cube of mass m and side s is in stable equilibrium when sitting on a level tabletop. How much energy is required to bring it to an unstable equilibrium where it's resting on its corner?

46. An isosceles triangular block of mass m and height h is in stable equilibrium, resting on its base on a horizontal surface. How much energy does it take to bring it to unstable equilibrium, resting on its apex?

47. You're investigating ladder safety for the Consumer Product Safety Commission. Your test case is a uniform ladder of mass m leaning against a frictionless vertical wall with which it makes an angle θ. The coefficient of static friction at the floor is μ. Your job is to find an expression for the maximum mass of a person who can climb to the top of the ladder without its slipping. With that result, you're to show that *anyone* can climb to the top if $\mu \geq \tan\theta$ but that *no one* can if $\mu < \frac{1}{2}\tan\theta$.

48. A 2.0-m-long rod has density λ in kilograms per meter of length described by $\lambda = a + bx$, where $a = 1.0$ kg/m, $b = 1.0$ kg/m^2, and x is the distance from the left end of the rod. The rod rests horizontally with each end supported by a scale. What do the two scales read?

49. What horizontal force applied at its highest point is necessary to keep a wheel of mass M from rolling down a slope inclined at angle θ to the horizontal?

50. A rectangular block twice as high as it is wide is resting on a board. The coefficient of static friction between board and incline is 0.63. If the board is tilted as shown in Fig. 12.27, will the block first tip over or first begin sliding?

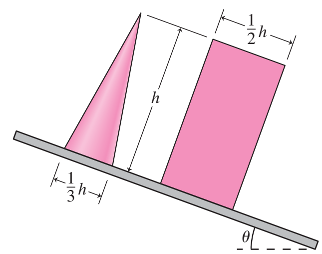


FIGURE 12.27 Problems 50, 51, and 52

51. What condition on the coefficient of friction in Problem 50 will cause the block to slide before it tips?

52. A uniform solid cone of height h and base diameter $\frac{1}{3}h$ sits on the board of Fig. 12.27. The coefficient of static friction between the cone and incline is 0.63. As the slope of the board is increased, will the cone first tip over or first begin sliding? (*Hint:* Start with an integration to find the center of mass.)

53. Prove the statement in Section 12.1 that the choice of pivot point doesn't matter when applying conditions for static equilibrium. Figure 12.28 shows an object on which the net force is assumed to be zero. The net torque about the point O is also zero. Show that the net torque about any other point P is also zero. To do so, write the net torque about P as $\vec{\tau}_P = \sum \vec{r}_{Pi} \times \vec{F}_i$, where the vectors $\vec{r}_P$ go from P to the force-application points, and the index i labels the different forces. In Fig. 12.28, note that $\vec{r}_{Pi} = \vec{r}_{Oi} + \vec{R}$, where $\vec{R}$ is a vector from P to O. Use this result in your expression for $\vec{\tau}_P$ and apply the distributive law to get two separate sums. Use the assumptions that $\vec{F}_{\text{net}} = \vec{0}$ and $\vec{\tau}_O = \vec{0}$ to argue that both terms are zero. This completes the proof.

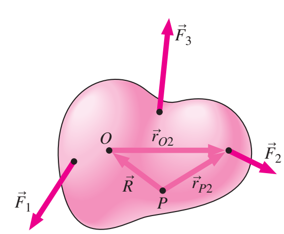


FIGURE 12.28 Problem 53

54. Three identical books of length L are stacked over the edge of a table as shown in Fig. 12.29. The top book overhangs the middle one by $L/2$, so it just barely avoids falling. The middle book

overhangs the bottom one by $L/4$. How much of the bottom book can overhang the edge of the table without the books falling?

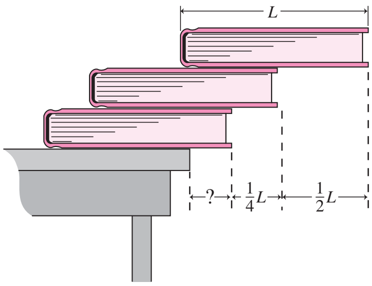


FIGURE 12.29 Problem 54

55. A uniform pole of mass M is at rest on an incline of angle θ secured by a horizontal rope as shown in Fig. 12.30. Find the minimum frictional coefficient that will keep the pole from slipping.

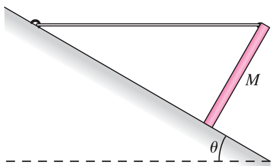


FIGURE 12.30 Problems 55 and 56

56. For what angle does the situation in Problem 55 require the greatest coefficient of friction?

57. Figure 12.31 shows a popular system for mounting bookshelves. An aluminum bracket is mounted on a vertical aluminum support by small tabs inserted into vertical slots. Contact between the bracket and support occurs only at the upper tab and at the bottom of the bracket, 4.5 cm below the upper tab. If each bracket in the shelf system supports 32 kg of books, with the center of gravity 12 cm out from the vertical support, what is the horizontal component of the force exerted on the upper bracket tab?

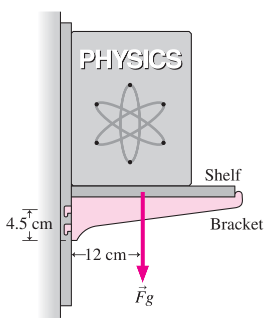


FIGURE 12.31 Problem 57

58. The *nuchal ligament* is a thick, cordlike structure that supports
BIO the head and neck in animals like horses. Figure 12.32 shows the nuchal ligament and its attachment points on a horse's skeleton, along with an approximation to the spine as a rigid rod. Centers of mass of head and neck are also shown. If the masses of head and neck are 29 kg and 68 kg, respectively, what's the tension in the nuchal ligament? (*Note*: Your answer will be an overestimate because muscles also provide support.)

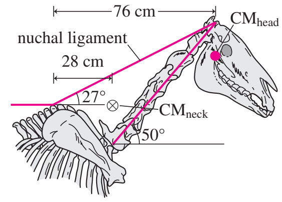


FIGURE 12.32 Problem 58

59. A 4.2-kg plant hangs from the bracket shown in Fig. 12.33. The bracket's mass is 0.85 kg, and its center of mass lies 9.0 cm from the wall. A single screw holds the bracket to the wall, as shown. Find the horizontal tension in the screw. (*Hint:* Imagine that the bracket is slightly loose and pivoting about its bottom end. Assume the wall is frictionless.)

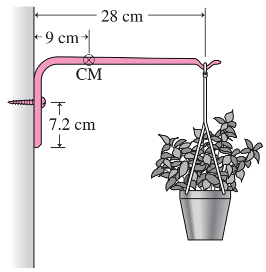


FIGURE 12.33 Problem 59

60. The wheel in Fig. 12.34 has mass M and is weighted with an additional mass m as shown. The coefficient of friction is sufficient to keep the wheel from sliding; however, it might still roll. Show that it won't roll only if $m > \dfrac{M \sin\theta}{1 - \sin\theta}$.

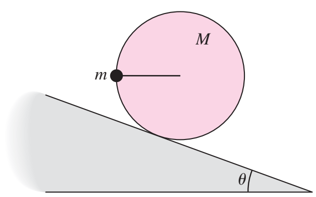


FIGURE 12.34 Problem 60

61. An interstellar spacecraft from an advanced civilization is hovering above Earth, as shown in Fig. 12.35. The ship consists of two pods of mass m separated by a rigid shaft of negligible mass

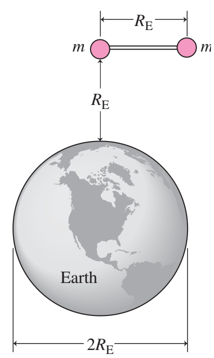


FIGURE 12.35 Problem 61

and one Earth radius (R_E) long. Find (a) the magnitude and direction of the net gravitational force on the ship and (b) the net torque about the center of mass. (c) Show that the ship's center of gravity is displaced approximately $0.083R_E$ from its center of mass.

62. You're called to testify in a product liability lawsuit. An infant sitting in the portable seat shown in Fig. 12.36 was injured when it fell to the floor. The manufacturer claims the child was too heavy for the seat; the parents claim the seat was defective. Tests show that the seat can safely hold a child if forces at *A* and *B* do not exceed 96.2 N and 229 N, respectively. The seat's mass is 2.0 kg, the injured child's is 10 kg, and the center of mass of child and seat was 16 cm from the table edge. In whose favor should the jury rule?

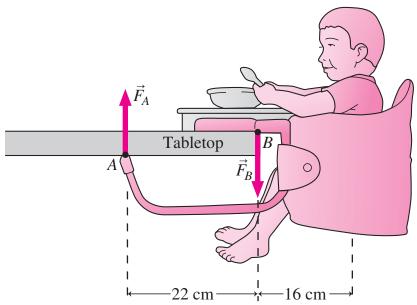


FIGURE 12.36 Problem 62

63. You're designing a vacation cabin at a ski resort. The cabin has a cathedral ceiling as shown in Fig. 12.37, and you estimate that each roof rafter needs to support up to 170 kg of snow and building materials. The horizontal tie beam near the roof peak can withstand a 7.5-kN force. You can neglect any horizontal force from the vertical walls, and treat contact forces as concentrated at the roof peak and the outside edge of the rafter/wall junctions. Will the tie beam hold? Will it be in tension or compression?

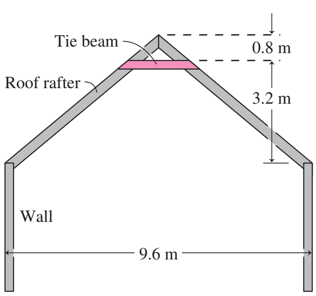


FIGURE 12.37 Problem 63

Passage Problems

You've been hired by your state's environmental agency to monitor carbon dioxide levels just above rivers, with the goal of understanding whether river water acts as a source or sink of CO_2. You've constructed the apparatus shown in Fig. 12.38, consisting of a boom mounted on a pivot, a vertical support, and a rope with pulley for raising and lowering the boom so its end can extend different distances over the river. In addition, there's a separate rope and pulley for dropping the sampling apparatus so it's just above the river.

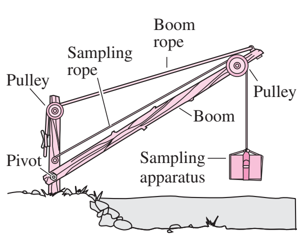


FIGURE 12.38 Passage Problems 64–67

64. When the boom rope is horizontal, it can't exert any vertical force. Therefore,
   a. it's impossible to hold the boom with the boom rope horizontal.
   b. the boom rope tension becomes infinite.
   c. the pivot supplies the necessary vertical force.
   d. the boom rope exerts no torque.

65. The tension in the boom rope will be greatest when
   a. the boom is horizontal.
   b. the boom rope is horizontal.
   c. the boom is vertical.
   d. in some orientation other than (a), (b), or (c).

66. If you secure the boom at a fixed angle and lower the sampling apparatus at constant speed, the boom rope tension will
   a. increase.
   b. decrease.
   c. remain the same.
   d. increase only if the sampling apparatus is more massive than the boom.

67. If you pull the boom rope with constant speed, the angle the boom makes with the horizontal will
   a. increase at a constant rate.
   b. increase at an increasing rate.
   c. increase at a decreasing rate.
   d. decrease.

Answers to Chapter Questions

Answer to Chapter Opening Question

Both the net force and the net torque on all parts of the bridge must be zero.

Answers to GOT IT? Questions

12.1. Pair *C*; pair *A* produces nonzero net force, and pair *B* produces nonzero net torque.

12.2. *B*; it's located directly over the point of contact with the floor, ensuring there's no gravitational torque.

12.3. (b); a frictional force at the floor is necessary to balance the normal force from the wall.

12.4. *D*: stable; *B*: metastable; *A* and *C*: unstable; *E*: neutrally stable.

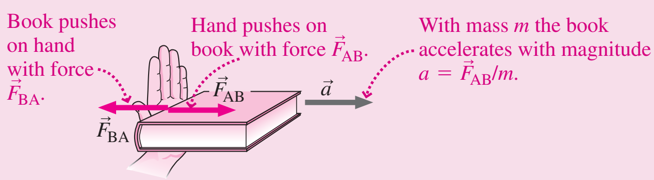
PART ONE SUMMARY

Mechanics

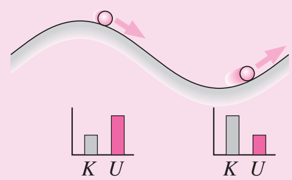
The big idea of Part One is Newton's realization that forces—pushes and pulls—don't cause motion but instead cause *changes* in motion. Newton's second law quantifies this idea. With momentum $\vec{p} = m\vec{v}$ as Newton's measure of "quantity of motion," the second law equates the net force on an object to the rate of change of its momentum: $\vec{F} = d\vec{p}/dt$ or, for constant mass, $\vec{F} = m\vec{a}$. The second law encompasses the first, the law of inertia: In the absence of a net force, an object continues in uniform motion, unchanging in speed or direction—a state that includes the special case of being at rest. Newton's third law rounds out the picture, providing a fully consistent description of motion with its statement that forces come in pairs: If object A exerts a force on B, then B exerts a force of equal magnitude but opposite direction on A.

From the concept of force and Newton's laws follow the essential ideas of work and energy, including kinetic and potential energy and the conservation of mechanical energy in the absence of nonconservative forces like friction. One important force is gravity, which Newton described through his law of universal gravitation and applied to explain the motions of the planets. Application of Newton's laws to systems of objects gives us the concept of center of mass and lets us describe the interactions of colliding objects. Finally, Newton's laws explain circular and rotational motion, the latter through the analogy between force and torque. That, in turn, gives us the tools needed to determine static equilibrium—the state in which an object at rest remains at rest, subject neither to a net force nor to a net torque.

Newton's laws provide a full description of motion.
Newton's 1st law: Force causes a change in motion.
Newton's 2nd law: $\vec{F} = d\vec{p}/dt$ or, for constant mass, $\vec{F} = m\vec{a}$
Newton's 3rd law: $\vec{F}_{AB} = -\vec{F}_{BA}$

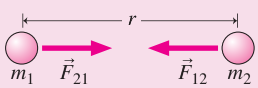


Work and **energy** are closely related concepts.
Work: $W = \vec{F} \cdot \Delta\vec{r}$ or, for a varying force, $W = \int \vec{F} \cdot d\vec{r}$
Work-energy theorem: $\Delta K = W$ with kinetic energy $K = \frac{1}{2}mv^2$
For conservative forces, work is stored as potential energy U. Then $K + U =$ constant.

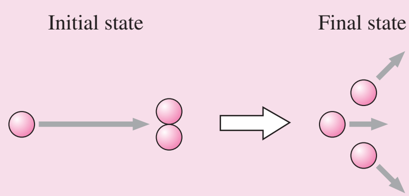


Universal gravitation describes the attractive force between all matter in the universe.

$$F = \frac{Gm_1m_2}{r^2}$$



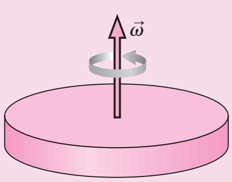
Momentum is conserved in a system that's not subject to external forces.



Initial momentum = Final momentum

$$\vec{P}_i = \Sigma m_i \vec{v}_i = m_1 \vec{v}_{1i} \quad \Rightarrow \quad \vec{P}_f = \Sigma m_f \vec{v}_f = m_1 \vec{v}_{1f} + m_2 \vec{v}_{2f} + m_3 \vec{v}_{3f}$$

Rotational motion is described by quantities analogous to those of linear motion.

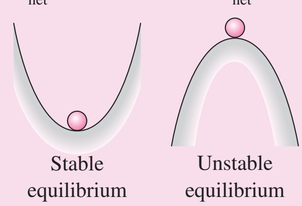


$\vec{v} \rightarrow \vec{\omega}$
$\vec{a} \rightarrow \vec{\alpha}$
$\vec{p} \rightarrow \vec{L}$
$\vec{F} \rightarrow \vec{\tau}$
$m \rightarrow I$

$$\vec{F} = m\vec{a} \rightarrow \vec{\tau} = I\vec{\alpha}$$
$$K = \tfrac{1}{2}mv^2 \rightarrow K = \tfrac{1}{2}I\omega^2$$

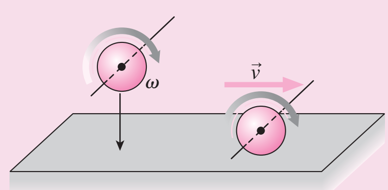
A system is in **static equilibrium** when the net force and the net torque on the system are both zero:

$\vec{F}_{net} = \vec{0}$ and $\vec{\tau}_{net} = \vec{0}$



Part One Challenge Problem

A solid ball of radius R is set spinning with angular speed ω about a horizontal axis. The ball is then lowered vertically with negligible speed until it just touches a horizontal surface and is released (see figure). If the coefficient of kinetic friction between the ball and the surface is μ, find (a) the linear speed of the ball once it achieves pure rolling motion, (b) the distance it travels before it achieves this motion, and (c) the fraction of the ball's initial rotational kinetic energy that's been lost to friction.



PART TWO

OVERVIEW

Oscillations, Waves, and Fluids

High-speed photo shows complex fluid behavior and spreading circular waves on water.

A tsunami crashes on shore, dissipating energy that has traveled across thousands of kilometers of open ocean. Near the epicenter of the earthquake that spawned the tsunami, a skyscraper sways in response, but suffers no damage thanks to a carefully engineered system that counters quake-induced vibrations. An electric guitar sounds loud during a rock concert, the sound waves following the vibrations of the guitar strings. Inside your watch, a tiny quartz crystal vibrates 32,768 times each second to keep near-perfect time. A radar-equipped police officer waits around the next turn in the highway ready to ticket your speeding car, while astrophysicists use the same principle to measure the expansion of the universe. A rafting party enters a narrow gorge, getting a wild ride as the river's speed increases. A plane cruises far overhead, supported by the force of air on its wings. All these examples involve the collective motion of many particles. In the next three chapters, we first explore the repetitive motion called oscillation and then show how oscillations in many-particle systems lead to wave motion. Finally, we apply the laws of motion to reveal the fascinating and sometimes surprising behavior of fluids like air and water.

13 Oscillatory Motion

Dancers from the Bandaloop Project perform on vertical surfaces, executing graceful slow-motion jumps. What determines the duration of these jumps?

New Concepts, New Skills

By the end of this chapter you should be able to

- Describe the relations between the period and frequency of oscillatory motion and between ordinary frequency and angular frequency (13.1, 13.2).
- State what simple harmonic motion is, and explain why it occurs universally in the physical world (13.2).
- Describe simple harmonic motion quantitatively (13.2–13.5).
- Explain damped harmonic motion and the phenomenon of resonance (13.6, 13.7).

Connecting Your Knowledge

- The material in this chapter draws on Newton's second law as applied in Chapter 5 and its rotational analog as introduced in Chapter 10 (5.1–5.3, 10.3).
- You should be familiar with the behavior of springs (4.6) and with the potential energy of a spring (7.2).
- You should have a solid understanding of the sine and cosine functions and be able to take derivatives of both these functions (Appendix A).

Displace a system from stable equilibrium, and forces or torques tend to restore that equilibrium. But, like the ball in Fig. 13.1, the system often overshoots its equilibrium and goes into **oscillatory motion** back and forth about equilibrium. In the absence of friction, this oscillation would continue forever; in reality, the system eventually settles into equilibrium.

Oscillatory motion occurs throughout the physical world. A uranium nucleus oscillates before it fissions. Water molecules oscillate to heat the food in a microwave oven. Carbon dioxide molecules in the atmosphere oscillate, absorbing energy and thus contributing to global warming. A watch—whether an old-fashioned mechanical one or a modern quartz timepiece—is a carefully engineered oscillating system. Buildings and bridges undergo oscillatory motion, sometimes with disastrous results. Even stars oscillate. And waves—from sound to ocean waves to seismic waves in the solid Earth—ultimately involve oscillatory motion.

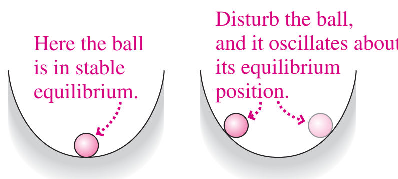


FIGURE 13.1 Disturbing a system results in oscillatory motion.

Oscillatory motion is universal because systems in stable equilibrium naturally tend to return toward equilibrium no matter how they're displaced. And it's not just the qualitative phenomenon of oscillation that's universal: Remarkably, the mathematical description of oscillatory motion is the same for systems ranging from atoms and molecules to cars and bridges and on to stars and galaxies.

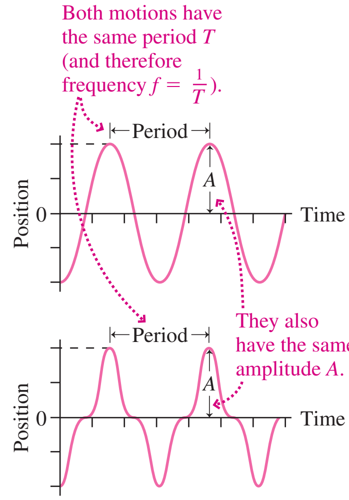


FIGURE 13.2 Position-time graphs for two oscillatory motions with the same amplitude A and period T (and therefore frequency).

13.1 Describing Oscillatory Motion

Figure 13.2 shows two quantities that characterize oscillatory motion: **Amplitude** is the maximum displacement from equilibrium, and **period** is the time it takes for the motion to repeat itself. Another way to express the time aspect is **frequency**, or number of oscillation cycles per unit time. Frequency f and period T are complementary ways of conveying the same information, and mathematically they're inverses:

$$f = \frac{1}{T} \tag{13.1}$$

The unit of frequency is the **hertz** (Hz), named after the German Heinrich Hertz (1857–1894), who was the first to produce and detect radio waves. One hertz is equal to one oscillation cycle per second.

EXAMPLE 13.1 Amplitude, Period, Frequency: An Oscillatory Distraction

Tired of homework, a student holds one end of a flexible plastic ruler against a desk and idly strikes the other end, setting it into oscillation (Fig. 13.3). The student notes that 28 complete cycles occur in 10 s and that the end of the ruler moves a total distance of 8.0 cm. What are the amplitude, period, and frequency of this oscillatory motion?

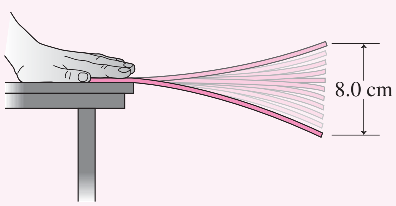


FIGURE 13.3 A ruler undergoing oscillatory motion.

INTERPRET We've got a case of oscillatory motion, and we're asked to describe it quantitatively in terms of amplitude, period, and frequency.

DEVELOP We can work from the definitions of these quantities: Amplitude is the maximum displacement from equilibrium, period is the time to complete a full oscillation, and frequency is the inverse of the period (Equation 13.1).

EVALUATE The ruler moves a total of 8.0 cm from one extreme to the other. Since the motion takes it to both sides of its equilibrium position, the amplitude is 4.0 cm. With 28 cycles in 10 s, the time per cycle, or the period, is

$$T = \frac{10\ \text{s}}{28} = 0.36\ \text{s}$$

The frequency is the inverse of the period: $f = 1/T = 1/0.36\ \text{s} = 2.8\ \text{Hz}$. We can also get this directly: 28 cycles/10 s $= 2.8$ Hz.

ASSESS Make sense? With a period that's less than 1 s, the frequency must be more than 1 cycle per second or 1 Hz. By the way, our definition of amplitude as the maximum displacement from equilibrium led to our 4.0-cm amplitude; the full 8.0 cm between extreme positions is called the **peak-to-peak amplitude**. ■

Amplitude and frequency don't provide all the details of oscillatory motion, since two quite different motions can have the same frequency and amplitude (Fig. 13.2). The differences reflect the restoring forces that return systems to equilibrium. Remarkably, though, restoring forces in many physical systems have the same mathematical form—a form we encountered before, when we introduced the force of an ideal spring in Chapter 4.

13.2 Simple Harmonic Motion

In many systems, the restoring force that develops when the system is displaced from equilibrium increases approximately in direct proportion to the displacement—meaning that if you displace the system twice as far from equilibrium, the force tending to restore equilibrium becomes twice as great. In the rest of this chapter, we therefore consider the case where the restoring force is directly proportional to the displacement. This is an approximation for most real systems, but often a very good approximation, especially for small displacements from equilibrium.

The type of motion that results from a restoring force proportional to displacement is called **simple harmonic motion** (SHM). Mathematically, we describe such a force by writing

$$F = -kx \qquad \text{(restoring force in SHM)} \tag{13.2}$$

where F is the force, x the displacement, and k a constant of proportionality between them. The minus sign in Equation 13.2 indicates a *restoring* force: If the object is displaced in one direction, the force is in the *opposite* direction, so it tends to restore the equilibrium.

We've seen Equation 13.2 before: It's the force exerted by an ideal spring of spring constant k. So a system consisting of a mass attached to a spring undergoes simple harmonic motion (Fig. 13.4). Many other systems—including atoms and molecules—can be modeled as miniature mass-spring systems.

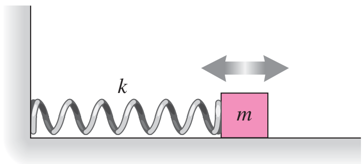


FIGURE 13.4 A mass attached to a spring undergoes simple harmonic motion.

How does a body in simple harmonic motion actually move? We can find out by applying Newton's second law, $F = ma$, to the mass-spring system of Fig. 13.4. Here the force on the mass m is $-kx$, so Newton's law becomes $-kx = ma$, where we take the x-axis along the direction of motion, with $x = 0$ at the equilibrium position. Now, the acceleration a is the second derivative of position, so we can write our Newton's law equation as

$$m\frac{d^2x}{dt^2} = -kx \qquad \text{(Newton's 2}^{\text{nd}}\text{ law for SHM)} \tag{13.3}$$

The solution to this equation is the position x as a function of time. What sort of function might it be? We expect periodic motion, so let's try periodic functions like sine and cosine. Suppose we pull the mass in Fig. 13.4 to the right and, at time $t = 0$, release it. Since it starts with a nonzero displacement, cosine is the appropriate function (recall that $\cos(0) = 1$, and $\sin(0) = 0$). We don't know the amplitude or frequency, so we'll try a form that has two unknown constants:

$$x(t) = A\cos\omega t \tag{13.4}$$

Because the cosine function itself varies between $+1$ and -1, A in Equation 13.4 is the amplitude—the greatest displacement from equilibrium (Fig. 13.5). What about ω? The cosine function undergoes a full cycle as its argument increases by 2π radians, or 360°, as shown in Fig. 13.5. In Equation 13.4, the argument of the cosine is ωt. Since the time for a full cycle is the period T, the argument ωt must go from 0 to 2π as the time t goes from 0 to T. So we have $\omega T = 2\pi$, or

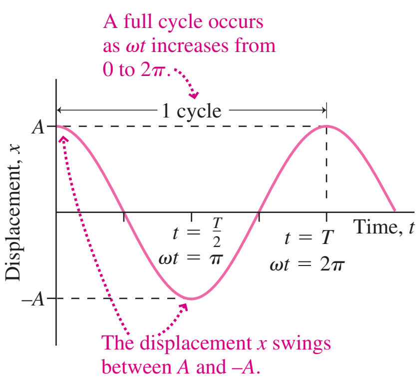


FIGURE 13.5 The function $A\cos\omega t$.

$$T = \frac{2\pi}{\omega} \tag{13.5}$$

The frequency of the motion is then

$$f = \frac{1}{T} = \frac{\omega}{2\pi} \tag{13.6}$$

Equation 13.6 shows that ω is a measure of the frequency, although it differs from the frequency f by the factor 2π. The quantity ω is called the **angular frequency**, and its units are radians per second or, since radians are dimensionless, simply inverse seconds (s^{-1}).

✓TIP Why Radians?

Here, as in Chapter 10, we use the angular quantity ω because it provides the simplest mathematical description of the motion. In fact, the relationship between angular frequency and frequency in hertz is the same as Chapter 10's relationship between angular speed in radians per second and in revolutions per second. We'll explore this similarity further in Section 13.4.

Writing the displacement x in the form 13.4 doesn't guarantee that we have a solution; we still need to see whether this form satisfies Equation 13.3. With $x(t)$ given by Equation 13.4, its first derivative is

$$\frac{dx}{dt} = \frac{d}{dt}(A\cos\omega t) = -A\omega\sin\omega t$$

where we've used the chain rule for differentiation (see Appendix A). Then the second derivative is

$$\frac{d^2x}{dt^2} = \frac{d}{dt}\left(\frac{dx}{dt}\right) = \frac{d}{dt}(-A\omega\sin\omega t) = -A\omega^2\cos\omega t$$

We can now try out our assumed solution for x (Equation 13.4) and its second derivative in Equation 13.3. Substituting $x(t)$ and d^2x/dt^2 in the appropriate places gives

$$m(-A\omega^2\cos\omega t) \stackrel{?}{=} -k(A\cos\omega t)$$

where the ? indicates that we're still trying to find out whether this is indeed an equality. If it is, the equality must hold *for all values of time t*. Why? Because Newton's law holds at all times, and we derived our questionable equality from Newton's law. Fortunately, the time-dependent term cos ωt appears on both sides of the equation, so we can cancel it. Also, the amplitude A and the minus sign cancel from the equation, leaving only $m\omega^2 = k$, or

$$\omega = \sqrt{\frac{k}{m}} \quad \text{(angular frequency, simple harmonic motion)} \qquad (13.7a)$$

Thus, Equation 13.4 *is* a solution of Equation 13.3, *provided* the angular frequency ω is given by Equation 13.7a.

Frequency and Period in Simple Harmonic Motion

We can recast Equation 13.7a in terms of the more familiar frequency f and period T using Equation 13.6, $f = \omega/2\pi$. This gives

$$f = \frac{\omega}{2\pi} = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \quad \text{and} \quad T = \frac{1}{f} = 2\pi\sqrt{\frac{m}{k}} \qquad (13.7b, c)$$

Do these relationships make sense? If we increase the mass m, it becomes harder to accelerate and we expect slower oscillations. This is reflected in Equations 13.7a and b, where m appears in the denominator. Increasing k, on the other hand, makes the spring stiffer and therefore results in greater force. That increases the oscillation frequency—as shown by the presence of k in the numerators of Equations 13.7a and b.

Physical systems display a wide range of m and k values and a correspondingly large range of oscillation frequencies. A molecule, with its small mass and its "springiness" provided by electric forces, may oscillate at 10^{14} Hz or more. A massive skyscraper, in contrast, typically oscillates at about 0.1 Hz.

Amplitude in Simple Harmonic Motion

The amplitude A canceled from our equations, so our analysis works for *any* value of A. This means that the oscillation frequency doesn't depend on amplitude. Independence of frequency and amplitude is a feature of simple harmonic motion, and arises because the restoring force is *directly proportional* to the displacement. When the restoring force does not have the simple form $F = -kx$, then frequency *does* depend on amplitude and the analysis of oscillatory motion becomes much more complicated. In many systems the relation $F = -kx$ breaks down if the displacement x gets too big; for this reason, simple harmonic motion usually occurs only for small oscillation amplitudes.

Phase

Equation 13.4 isn't the only solution to Equation 13.3; you can readily show that $x = A\sin\omega t$ works just as well. We chose the cosine because we took time $t = 0$ at the point of maximum displacement. Had we set $t = 0$ as the mass passed through its equilibrium point, sine would

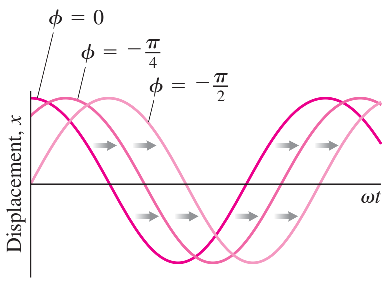


FIGURE 13.6 A negative phase constant shifts the curve to the right.

have been the appropriate function. More generally we can take the zero of time at some arbitrary point in the oscillation cycle. Then, as Fig. 13.6 shows, we can represent the motion by the form

$$x(t) = A\cos(\omega t + \phi) \qquad \text{(simple harmonic motion)} \tag{13.8}$$

where the **phase constant** ϕ has the effect of shifting the cosine curve to the left (for $\phi > 0$) or right ($\phi < 0$) but doesn't affect the frequency or amplitude.

Velocity and Acceleration in Simple Harmonic Motion

Equation 13.4 (or, more generally, Equation 13.8) gives the position of an object in simple harmonic motion as a function of time, so its first derivative must be the object's velocity:

$$v(t) = \frac{dx}{dt} = \frac{d}{dt}(A\cos\omega t) = -\omega A\sin\omega t \tag{13.9}$$

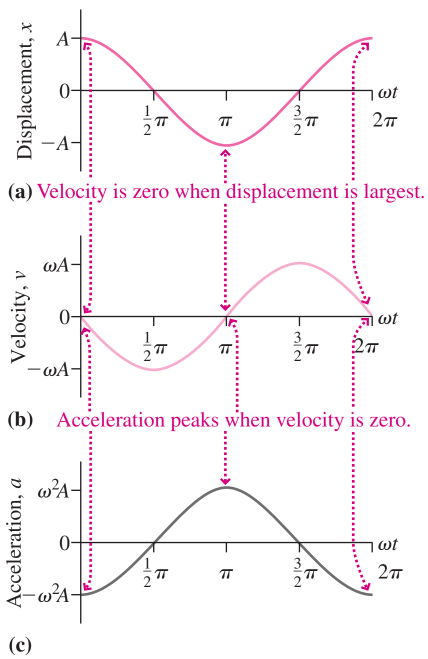


FIGURE 13.7 Displacement, velocity, and acceleration in simple harmonic motion.

Because the maximum value of the sine function is 1, this expression shows that the maximum velocity is ωA. This makes sense because a higher-frequency oscillation requires that the object traverse the distance A in a shorter time—so it must move faster. Equation 13.9 shows that the velocity $v(t)$ is a sine function when the displacement $x(t)$ is a cosine. Thus velocity is a maximum when displacement is zero, and vice versa; mathematically, we express this by saying that displacement and velocity differ in phase by $\frac{\pi}{2}$ radians or 90°. Does this make sense? Sure, because at the extremes of its motion, the object is instantaneously at rest as it reverses direction: maximum displacement, zero speed. And when it passes through its equilibrium position, the object is going fastest. Figures 13.7*a* and *b* show graphically the relationship between displacement and velocity in simple harmonic motion.

Just as velocity is the derivative of position, so acceleration is the derivative of velocity, or the second derivative of position:

$$a(t) = \frac{dv}{dt} = \frac{d}{dt}(-\omega A\sin\omega t) = -\omega^2 A\cos\omega t \tag{13.10}$$

Thus the maximum acceleration is $\omega^2 A$. Since acceleration is a cosine function if velocity is a sine, each reaches its maximum value when the other is zero (Fig. 13.7*b*, *c*).

GOT IT? 13.1 Two identical mass-spring systems are displaced different amounts from equilibrium and then released at different times. Of the amplitudes, frequencies, periods, and phase constants of the subsequent motions, which are the same for both systems and which are different?

EXAMPLE 13.2 Simple Harmonic Motion: A Tuned Mass Damper

The tuned mass damper in New York's Citicorp Tower (see Application on next page) consists of a 373-Mg concrete block that completes one oscillation in 6.80 s. The oscillation amplitude in a high wind is 110 cm. Determine the spring constant and the maximum speed and acceleration of the block.

INTERPRET This is a problem involving simple harmonic motion, with the concrete block and spring making up the oscillating system. We're given the period, mass, and amplitude.

DEVELOP Equation 13.7c, $T = 2\pi\sqrt{m/k}$, gives us the spring constant. Equations 13.9 and 13.10 show that the maximum speed and acceleration are $v_{max} = \omega A$ and $a_{max} = \omega^2 A$, and we can get the angular frequency ω from the period using Equation 13.5: $\omega = 2\pi/T$.

EVALUATE First we solve Equation 13.7c for the spring constant:

$$k = \frac{4\pi^2 m}{T^2} = \frac{(4\pi^2)(3.73\times10^5\text{ kg})}{(6.80\text{ s})^2} = 3.18\times10^5\text{ N/m}$$

The angular frequency is $\omega = 2\pi/T = 0.924\text{ s}^{-1}$. Then we have $v_{max} = \omega A = (0.924\text{ s}^{-1})(1.10\text{ m}) = 1.02$ m/s and $a_{max} = \omega^2 A = 0.939\text{ m/s}^2$.

ASSESS The large spring constant and relatively low velocity and acceleration make sense given the huge mass involved. Note that we had to convert the mass, given as 373 Mg (373×10^6 g), to kilograms before evaluating. ■

13.3 Applications of Simple Harmonic Motion

Simple harmonic motion occurs in any system where the tendency to return to equilibrium increases in direct proportion to the displacement from equilibrium. Analysis of such systems is like that of the mass-spring system we just considered but may involve different physical quantities.

The Vertical Mass-Spring System

A mass hanging vertically from a spring is subject to gravity as well as the spring force (Fig. 13.8). In equilibrium the spring stretches enough for its force to balance gravity: $mg - kx_1 = 0$, where x_1 is the new equilibrium position. Stretching the spring an additional amount Δx increases the spring force by $k\,\Delta x$, and this increased force tends to restore the equilibrium. So once again we have a restoring force that's directly proportional to displacement. And here, with the same spring constant k and mass m, our previous analysis still applies and we get simple harmonic motion with frequency $\omega = \sqrt{k/m}$. Thus gravity changes only the equilibrium position and doesn't affect the frequency.

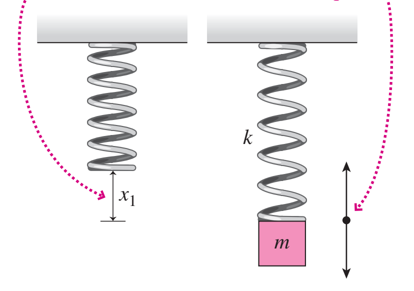
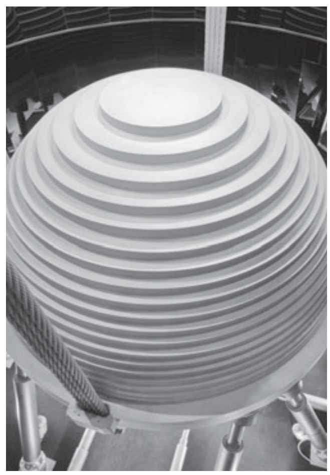


FIGURE 13.8 A vertical mass-spring system oscillates about a new equilibrium position x_1, with the same frequency $\omega = \sqrt{k/m}$.

APPLICATION Swaying Skyscrapers

Skyscrapers are tall, thin, flexible structures. High winds and earthquakes can set them oscillating, much like the ruler of Example 13.1. Wind-driven oscillations are uncomfortable to occupants of a building's upper floors, and earthquake-induced oscillations can be downright destructive.

Modern skyscrapers use so-called tuned mass dampers to counteract building oscillations. These devices are essentially large mass-spring systems mounted high in the building. They're engineered to oscillate with the same frequency as the building (hence the term "tuned") but 180° out of phase, thus reducing the amplitude of the building's own oscillation. The result is increased comfort for the building's occupants and improved safety for buildings in earthquake-prone regions. Tuned mass dampers also find applications in tall smokestacks, airport control towers, power-plant cooling towers, bridges, ski lifts, and even the new Grand Canyon skywalk. By suppressing vibrations, tuned mass dampers enable architects and engineers to design structures that don't need as much intrinsic stiffness, so they can be lighter and less expensive. The photos show the world's largest tuned mass damper and the building that houses it, Taiwan's Taipei 101 skyscraper. The damper helps the building survive earthquakes and typhoons. Example 13.2 explores another tuned mass damper.



The Torsional Oscillator

Figure 13.9 shows a disk suspended from a wire. Rotate the disk slightly, and a torque develops in the wire. Let go, and the disk oscillates by rotating back and forth. This is a **torsional oscillator**, and it's best described using the language of rotational motion. The **angular displacement** θ, **restoring torque** τ, and **torsional constant** κ relate the torque and displacement: $\tau = -\kappa\theta$, where again the minus sign indicates that the torque is opposite the displacement, tending to restore the system to equilibrium. The rotational analog of Newton's law, $\tau = I\alpha$, describes the system's behavior; here the rotational inertia I plays the role of mass. But the angular acceleration α is the second derivative of the angular position, so Newton's law becomes

$$I\frac{d^2\theta}{dt^2} = -\kappa\theta \qquad (13.11)$$

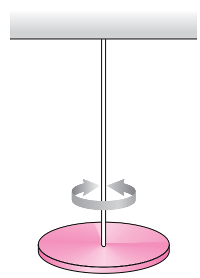


FIGURE 13.9 A torsional oscillator.

This is identical to Equation 13.3 for the linear oscillator, with I replacing m, θ replacing x, and κ replacing k. So we can immediately write $\theta(t) = A\cos\omega t$ for the angular displacement and, in analogy with Equation 13.7a,

$$\omega = \sqrt{\frac{\kappa}{I}} \qquad (13.12)$$

for the angular frequency.

Torsional oscillators constitute the timekeeping mechanism in mechanical watches, and they can provide accurate measures of rotational inertia.

The Pendulum

A **simple pendulum** consists of a point mass suspended from a massless string. Real systems approximate this ideal when a suspended object's size is negligible compared with the suspension length and its mass is much greater than that of the suspension. The dancers in the chapter's opening photo are essentially simple pendulums, as is the pendulum in a grandfather clock. Figure 13.10 shows a pendulum of mass m and length L displaced slightly from equilibrium. The gravitational force exerts a torque given by $\tau = -mgL\sin\theta$, where the minus sign indicates that the torque tends to rotate the pendulum back toward equilibrium. The rotational analog of Newton's law, $\tau = I\alpha$, then becomes

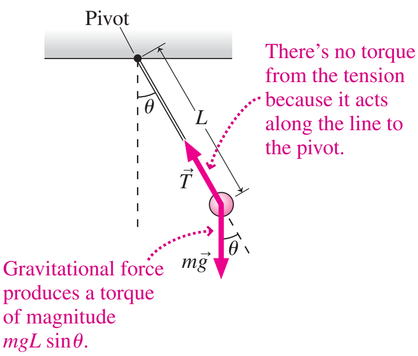


FIGURE 13.10 Forces on a pendulum.

$$I\frac{d^2\theta}{dt^2} = -mgL\sin\theta$$

where we've written the angular acceleration as the second derivative of the angular displacement. This looks like Equation 13.11 for the torsional oscillator—but not quite, since the torque involves $\sin\theta$ rather than θ itself. Thus the restoring torque is not *directly* proportional to the angular displacement, and the motion is therefore *not* simple harmonic.

If, however, the amplitude of the motion is small, then it *approximates* simple harmonic motion. Figure 13.11 shows that for small angles, $\sin\theta$ and θ are essentially equal. For a small-amplitude pendulum we can therefore replace $\sin\theta$ with θ to get

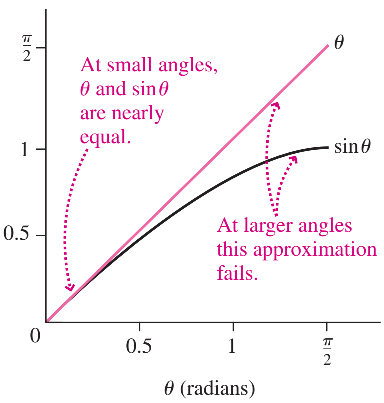


FIGURE 13.11 For θ much less than 1 radian, $\sin\theta$ and θ are nearly equal.

$$I\frac{d^2\theta}{dt^2} = -mgL\theta$$

This is essentially Equation 13.11, with mgL playing the role of κ. So the small-amplitude pendulum undergoes simple harmonic motion, with its angular frequency given by Equation 13.12 with $\kappa = mgL$:

$$\omega = \sqrt{\frac{mgL}{I}} \tag{13.13}$$

For a *simple* pendulum, the rotational inertia I is that of a point mass m a distance L from the rotation axis, or $I = mL^2$, as we found in Chapter 10. Then we have

$$\omega = \sqrt{\frac{mgL}{mL^2}} = \sqrt{\frac{g}{L}} \quad \text{(simple pendulum)} \tag{13.14}$$

or, from Equation 13.5,

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{L}{g}} \quad \text{(simple pendulum)} \tag{13.15}$$

These equations show that the frequency and period of a simple pendulum are independent of its mass, depending only on length and gravitational acceleration.

EXAMPLE 13.3 **A Pendulum: Rescuing Tarzan**

Tarzan stands on a branch as a leopard threatens. Fortunately, Jane is on a nearby branch of the same height, holding a 25-m-long vine attached directly above the point midway between her and Tarzan. She grasps the vine and steps off with negligible velocity. How soon does she reach Tarzan?

INTERPRET This is a problem about a pendulum, which we identify as consisting of Jane and the vine. The period of the pendulum is the time for a full swing back and forth, so the answer we're after—the time to reach Tarzan—is half the period.

DEVELOP We sketched the situation in Fig. 13.12. Equation 13.15, $T = 2\pi\sqrt{L/g}$, determines the period, so we can use this equation to find the half-period.

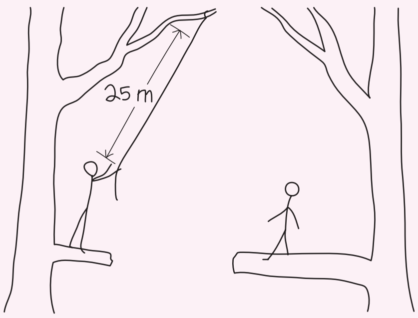


FIGURE 13.12 Our sketch for Example 13.3. Vine length is not to scale.

EVALUATE Equation 13.15 gives

$$\frac{1}{2}T = \left(\frac{1}{2}\right)(2\pi)\sqrt{\frac{L}{g}} = (\pi)\sqrt{\frac{25\text{ m}}{9.8\text{ m/s}^2}} = 5.0\text{ s}$$

ASSESS This seems a reasonable answer for a problem involving human-scale objects and many meters of vine. One caution: Jane's rescue will be successful only if the vine is strong enough—not only to support her weight but also to provide the acceleration that keeps her moving in a circular arc. You can explore that issue in Problem 56. ■

GOT IT? 13.2 What happens to the period of a pendulum if (a) its mass is doubled; (b) it's moved to a planet whose gravitational acceleration is one-fourth that of Earth; and (c) its length is quadrupled?

CONCEPTUAL EXAMPLE 13.1 The Nonlinear Pendulum

A pendulum consists of a weight on the end of a rigid rod of negligible mass, hanging vertically from a frictionless pivot at the opposite end of the rod. For small-amplitude disturbances from equilibrium, the system constitutes a simple pendulum. But for larger disturbances it becomes a *nonlinear pendulum*, so named because the restoring torque is no longer proportional to the displacement. Quantitative analysis of a nonlinear pendulum is difficult, but you can still understand it conceptually.

(a) As the pendulum's amplitude increases, how will its period change?
(b) If you start the pendulum by striking it when it's hanging vertically, will it undergo oscillatory motion no matter how hard it's hit?

EVALUATE (a) Before we made the small-amplitude approximation, we showed that a pendulum's restoring torque is, in general, proportional to $\sin\theta$. But Figure 13.11 shows that $\sin\theta$ doesn't increase as fast as θ itself. So for large-amplitude swings, the restoring torque is *less* than it would be in the small-amplitude approximation. This suggests the pendulum should return more slowly toward equilibrium—and thus its period should increase.

(b) When you strike the pendulum, you give it kinetic energy. If that energy is insufficient to invert it completely, then the pendulum will swing to one side, eventually stop, and return, undergoing back-and-forth oscillatory motion. But hit it hard enough, and it will go "over the top," reaching its fully inverted position with kinetic energy to spare. Round and round it goes, executing motion that's periodic and circular, but not oscillatory. This circular motion isn't uniform, because it moves more slowly at the top and faster at the bottom.

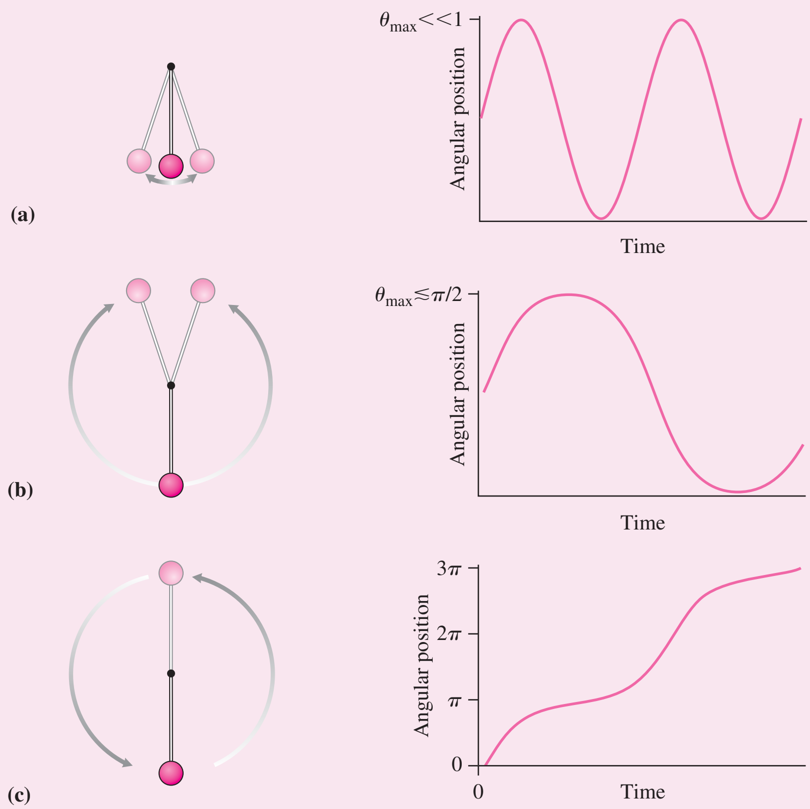


FIGURE 13.13 Conceptual Example 13.1: (a) Small-amplitude oscillations; (b) large-amplitude oscillations; (c) circular motion.

ASSESS Make sense? Yes: Consider a pendulum with just a little less energy than it takes to go "over the top." It will move very slowly near the top of its trajectory, so its period will be quite long. And its angular-position-versus-time curve will be flatter than the sine curve of a simple pendulum. Give it just a little more energy, and it goes into circular motion. Figure 13.13 illustrates all three situations.

MAKING THE CONNECTION If the pendulum has length L, what's the minimum speed that will get it "over the top," into periodic nonuniform circular motion?

EVALUATE Potential energy at the top is $U = mg(2L)$, so kinetic energy $K = \frac{1}{2}mv^2$ has to be at least this large. That gives $v > 2\sqrt{gL}$.

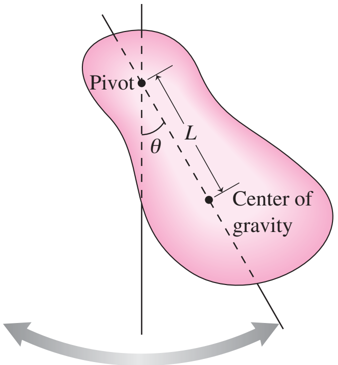


FIGURE 13.14 A physical pendulum.

The Physical Pendulum

A **physical pendulum** is an object of arbitrary shape that's free to swing (Fig. 13.14). It differs from a simple pendulum in that mass may be distributed over its entire length. Physical pendulums are everywhere: Examples include the legs of humans and other animals (see Example 13.4), a skier on a chair lift, a boxer's punching bag, a frying pan hanging from a rack, and a crane lifting any object of significant extent. In our analysis of the simple pendulum, we used the fact that mass was concentrated at the bottom only in the final step, when we wrote mL^2 for the rotational inertia. Our analysis before that step therefore applies to the physical pendulum as well.

In particular, a physical pendulum displaced slightly from equilibrium undergoes simple harmonic motion with frequency given by Equation 13.13. But how are we to interpret the length L in that equation? Because gravity—which provides the restoring torque for *any* pendulum—acts on an object's center of gravity, L must be the distance from the pivot to the center of gravity, as marked in Fig. 13.14.

EXAMPLE 13.4 A Physical Pendulum: Walking

When walking, the leg not in contact with the ground swings forward, acting like a physical pendulum. Approximating the leg as a uniform rod, find the period of this pendulum motion for a leg of length 90 cm.

INTERPRET This problem is about a physical pendulum, here identified as a uniform rod approximating the leg.

DEVELOP Figure 13.15 is our drawing, showing the leg as a rod pivoting at the hip. The center of mass of a uniform rod is at its center, so the effective length L is half the leg's length, or 45 cm. Equation 13.13, $\omega = \sqrt{mgL/I}$, determines the angular frequency, from which we can get the period using Equation 13.5, $T = 2\pi/\omega$. We also need the rotational inertia; from Table 10.2, that's $I = \frac{1}{3}M(2L)^2$, where we use $2L$ because Table 10.2's expression involves the *full* length of the rod.

EVALUATE Putting this all together, we evaluate to get the answer:

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{I}{mgL}} = 2\pi\sqrt{\frac{\frac{1}{3}m(2L)^2}{mgL}} = 2\pi\sqrt{\frac{4L}{3g}}$$

Using $L = 0.45$ m gives $T = 1.6$ s.

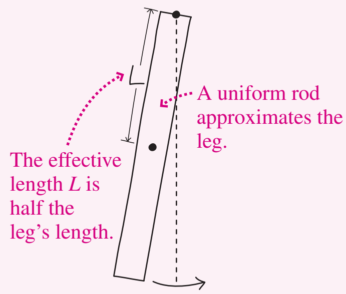


FIGURE 13.15 A human leg treated as a pendulum.

ASSESS The leg swings forward to complete a full stride in half a period, or 0.8 s. This seems a reasonable value for the pace in walking.

13.4 Circular and Harmonic Motion

Look down on the solar system, and you see Earth in circular motion about the Sun (Fig. 13.16*a*). But look in from the plane of Earth's orbit, and Earth appears to be moving back and forth (Fig. 13.16*b*). Figure 13.17 shows that this apparent back-and-forth motion is a single component of the actual circular motion, and that this component describes a sinusoidal function of time. Specifically, the position vector $\vec{r}$ for Earth or any other object in circular motion makes an angle that increases linearly with time: $\theta = \omega t$, where we measure θ with respect to the x-axis and take $t = 0$ when the object is on the x-axis. Then the two components $x = r\cos\theta$ and $y = r\sin\theta$ of the object's position become

$$x(t) = r\cos\omega t \quad \text{and} \quad y(t) = r\sin\omega t$$

These are the equations for two different simple harmonic motions, one in the x-direction and the other in the y-direction. Because one is a cosine and the other is a sine, they're out of phase by $\frac{\pi}{2}$ or 90°.

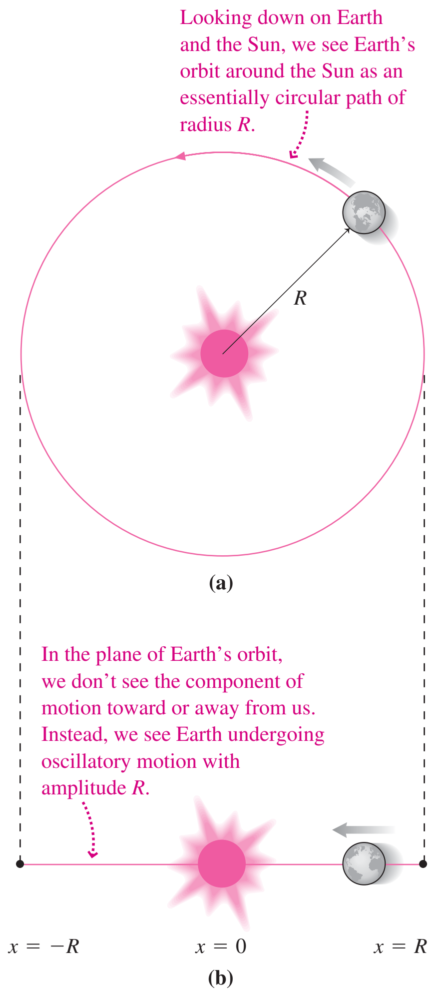


FIGURE 13.16 Two views of Earth's orbital motion.

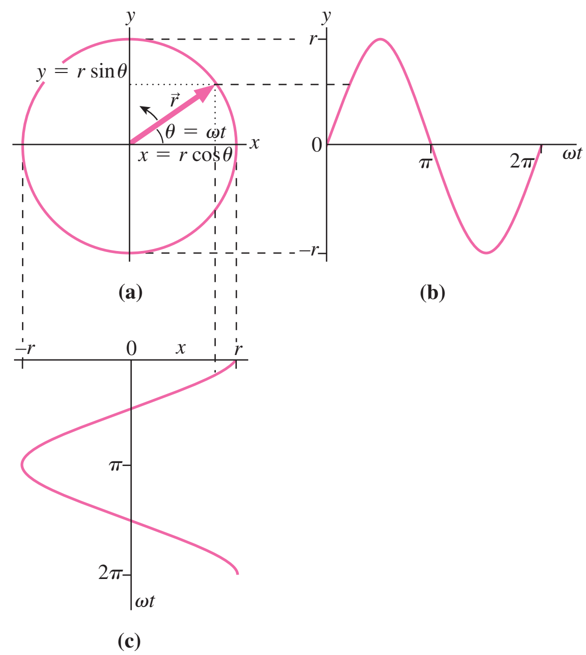


FIGURE 13.17 As the position vector $\vec{r}$ traces out a circle, its x- and y-components are sinusoidal functions of time.

So we can think of circular motion as resulting from perpendicular simple harmonic motions, with the same amplitude and frequency but 90° out of phase. This should help you to understand why we use the term "angular frequency" for simple harmonic motion even though there's no angle involved. The argument ωt in the description of simple harmonic motion is the same as the physical angle θ in the corresponding circular motion. The time for one cycle of simple harmonic motion is the same as the time for one revolution in the circular motion, so the values of T and therefore ω are exactly the same.

You can verify that two mutually perpendicular simple harmonic motions of the same amplitude and frequency sum vectorially to give circular motion (see Problem 53). If the amplitudes or frequencies aren't the same, then interesting complex motions occur, as shown in Fig. 13.18.

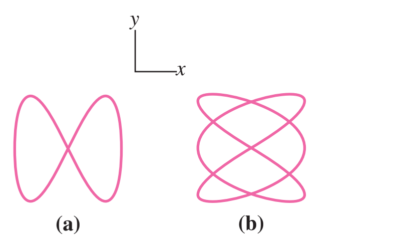


FIGURE 13.18 Complex paths resulting from different frequencies in different directions. Can you determine the frequency ratios?

GOT IT? 13.3 Figure 13.18 shows the paths traced in the horizontal plane by two pendulums swinging with different frequencies in two perpendicular directions. What's the ratio of x-direction frequency to y-direction frequency for the two paths shown?

13.5 Energy in Simple Harmonic Motion

Displace a mass-spring system from equilibrium, and you do work as you build up potential energy in the spring. Release the mass, and it accelerates toward equilibrium, gaining kinetic energy at the expense of potential energy. It passes through its

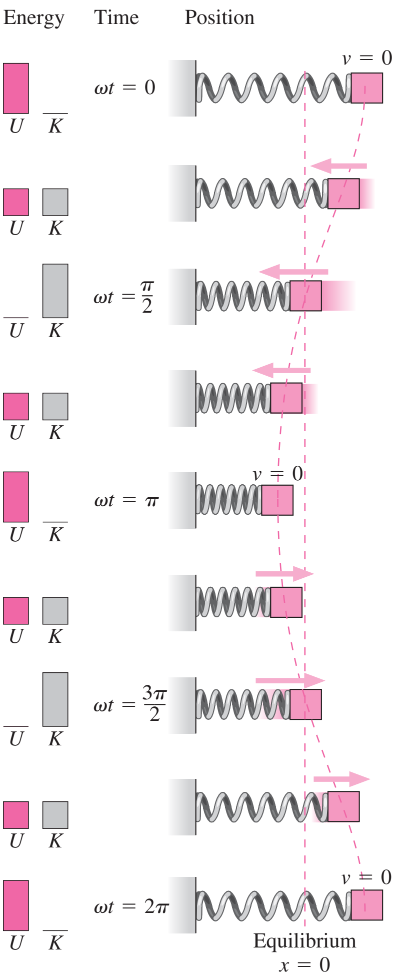


FIGURE 13.19 Kinetic and potential energy in simple harmonic motion.

equilibrium position with maximum kinetic energy and no potential energy, then slows and builds potential energy as it compresses the spring. If there's no energy loss, this process continues indefinitely. In oscillatory motion, energy is continuously transferred back and forth between its kinetic and potential forms (Fig. 13.19).

For a mass-spring system, the potential energy is given by Equation 7.4: $U = \frac{1}{2}kx^2$, where x is the displacement from equilibrium. Meanwhile, the kinetic energy is $K = \frac{1}{2}mv^2$. We can illustrate explicitly the interchange of kinetic and potential energy in simple harmonic motion by using x from Equation 13.4 and v from Equation 13.9 in the expressions for potential and kinetic energy. Then we have

$$U = \tfrac{1}{2}kx^2 = \tfrac{1}{2}k(A\cos\omega t)^2 = \tfrac{1}{2}kA^2\cos^2\omega t$$

and

$$K = \tfrac{1}{2}mv^2 = \tfrac{1}{2}m(-\omega A\sin\omega t)^2 = \tfrac{1}{2}m\omega^2A^2\sin^2\omega t = \tfrac{1}{2}kA^2\sin^2\omega t$$

where we used $\omega^2 = k/m$. Both energy expressions have the same maximum value—$\frac{1}{2}kA^2$—equal to the initial potential energy of the stretched spring. But the potential energy is a maximum when the kinetic energy is zero, and vice versa. What about the total energy? It's

$$E = U + K = \tfrac{1}{2}kA^2\cos^2\omega t + \tfrac{1}{2}kA^2\sin^2\omega t = \tfrac{1}{2}kA^2$$

where we used $\sin^2\omega t + \cos^2\omega t = 1$.

Our result is a statement of the conservation of mechanical energy—the principle we introduced in Chapter 7—applied to a simple harmonic oscillator. Although the kinetic and potential energies K and U both vary with time, their sum—the total energy E—does not (Fig. 13.20).

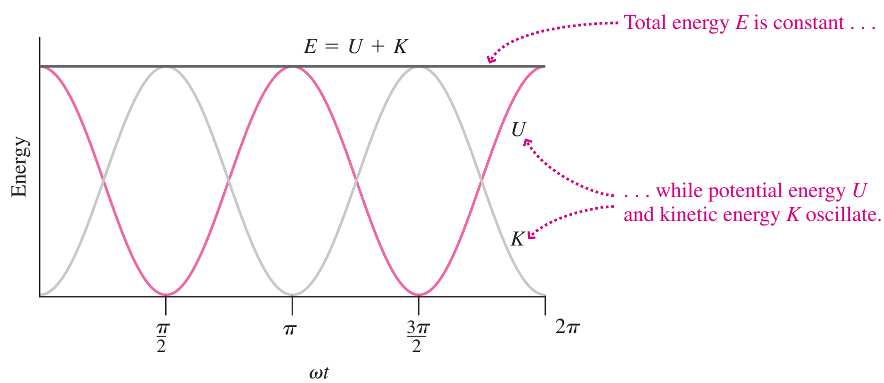


FIGURE 13.20 Energy of a simple harmonic oscillator.

EXAMPLE 13.5 Energy in Simple Harmonic Motion

A mass-spring system undergoes simple harmonic motion with angular frequency ω and amplitude A. Find its speed at the point where the kinetic and potential energies are equal.

INTERPRET This example involves the concept of energy conservation in simple harmonic motion. We're asked to find a speed, which is related to kinetic energy.

DEVELOP When the kinetic energy equals the potential energy, each must be half the total energy. What is that total? The speed is at its maximum, $v_{\max} = \omega A$ from Equation 13.9, when the energy is all kinetic. Thus the total energy is $E = \frac{1}{2}mv_{\max}^2 = \frac{1}{2}m\omega^2A^2$. The speed v we're after occurs when the kinetic energy has half this value, or $K = \frac{1}{2}mv^2 = \frac{1}{2}\left(\frac{1}{2}m\omega^2A^2\right) = \frac{1}{4}m\omega^2A^2$.

EVALUATE Solving for v gives our answer:

$$v = \frac{\omega A}{\sqrt{2}}$$

ASSESS Make sense? Yes. The speed at this point must obviously be less than the maximum speed, since half the energy is tied up as potential energy in the spring. And because kinetic energy depends on the *square* of the speed, it's lower not by a factor of 2 but by $\sqrt{2}$. ■

Potential-Energy Curves and Simple Harmonic Motion

We arrived at the expression $U = \frac{1}{2}kx^2$ for the potential energy of a spring by integrating the spring force, $-kx$, over distance. Since every simple harmonic oscillator has a restoring force or torque proportional to displacement, integration always results in a potential energy proportional to the *square* of the displacement—that is, in a parabolic potential-energy curve. Conversely, any system with a parabolic potential-energy curve exhibits simple harmonic motion. The simplest mathematical approximation to a smooth curve near a minimum is a parabola, and for that reason potential-energy curves for complex systems often approximate parabolas near their stable equilibrium points (Fig. 13.21). Small disturbances from these equilibria therefore result in simple harmonic motion, and that's why simple harmonic motion is so common throughout the physical world.

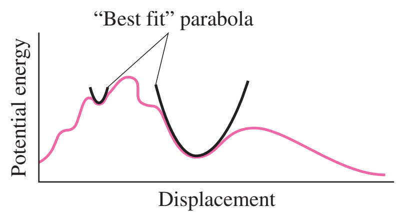


FIGURE 13.21 Near their minima, potential-energy curves often approximate parabolas; the result is simple harmonic motion.

GOT IT? 13.4 Two different mass-spring systems are oscillating with the same amplitude and frequency. If one has twice as much total energy as the other, how do (a) their masses and (b) their spring constants compare? (c) What about their maximum speeds?

13.6 Damped Harmonic Motion

In real oscillating systems, forces such as friction or air resistance normally dissipate the oscillation energy. This energy loss causes the oscillation amplitude to decrease, and the motion is said to be **damped**.

If dissipation is sufficiently weak that only a small fraction of the system's energy is lost in each oscillation cycle, then we expect that the system should behave essentially as in the undamped case, except for a gradual decrease in amplitude (Fig. 13.22).

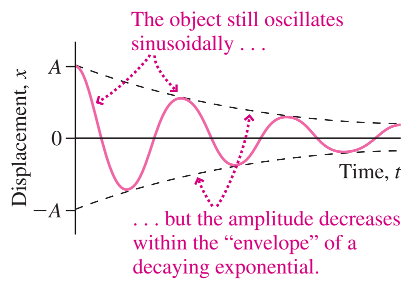


FIGURE 13.22 Weakly damped motion.

In many systems the damping force is approximately proportional to the velocity and in the opposite direction:

$$F_d = -bv = -b\frac{dx}{dt}$$

where b is a constant giving the strength of the damping. We can write Newton's law as before, now including the damping force along with the restoring force. For a mass-spring system, we have

$$m\frac{d^2x}{dt^2} = -kx - b\frac{dx}{dt} \qquad (13.16)$$

We won't solve this equation, but simply state its solution:

$$x(t) = Ae^{-bt/2m}\cos(\omega t + \phi) \qquad (13.17)$$

This equation describes sinusoidal motion whose amplitude decreases exponentially with time. How fast depends on the damping constant b and mass m: When $t = 2m/b$, the amplitude has dropped to $1/e$ of its original value. When the damping is so weak that only a small fraction of the total energy is lost in each cycle, the frequency ω in Equation 13.17 is essentially equal to the undamped frequency $\sqrt{k/m}$. But with stronger damping, the damping force slows the motion, and the frequency becomes lower. As long as oscillation occurs, the motion is said to be **underdamped** (Fig. 13.23*a*). For sufficiently strong damping, though, the effect of the damping force is as great as that of the spring force. Under this condition, called **critical damping**, the system returns to its equilibrium state without undergoing any oscillations (Fig. 13.23*b*). If the damping is made still stronger, the system becomes *overdamped*. The damping force now dominates, so the system returns more slowly to equilibrium (Fig. 13.23*c*).

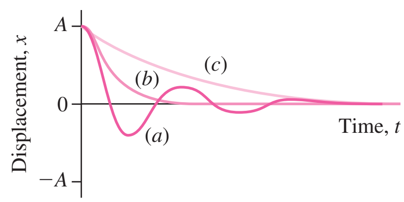


FIGURE 13.23 (a) Underdamped, (b) critically damped, and (c) overdamped oscillations.

Many physical systems, from atoms to the human leg, can be modeled as damped oscillators. Engineers often design systems with specific amounts of damping. Automobile shock absorbers, for example, coordinate with the springs to give critical damping. This results in rapid return to equilibrium while absorbing the energy imparted by road bumps.

EXAMPLE 13.6 Damped Simple Harmonic Motion: Bad Shocks

A car's suspension acts like a mass-spring system with $m = 1200$ kg and $k = 58$ kN/m. Its worn-out shock absorbers provide a damping constant $b = 230$ kg/s. After the car hits a pothole, how many oscillations will it make before the amplitude drops to half its initial value?

INTERPRET We interpret this problem as being about damped simple harmonic motion, and we identify the car as the oscillating system.

DEVELOP Our plan is to find out how long it takes the amplitude to decrease by half and then find the number of oscillation cycles in this time. Equation 13.17, $x(t) = Ae^{-bt/2m}\cos(\omega t + \phi)$, describes the motion, with the factor $e^{-bt/2m}$ giving the decrease in amplitude. At $t = 0$ this factor is 1, so we want to know when it's equal to one-half: $e^{-bt/2m} = \frac{1}{2}$.

EVALUATE Taking the natural logarithms of both sides gives $bt/2m = \ln 2$, where we used the facts that $\ln(x)$ and e^x are inverse functions and $\ln(1/x) = -\ln(x)$. Then

$$t = \frac{2m}{b}\ln 2 = \frac{(2)(1200\text{ kg})}{230\text{ kg/s}}\ln 2 = 7.23\text{ s}$$

is the time for the amplitude to drop to half its original value. For weak damping, the period is very close to the undamped period, which is

$$T = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{1200\text{ kg}}{58\times 10^3\text{ N/m}}} = 0.904\text{ s}$$

Then the number of cycles during the 7.23 s it takes the amplitude to drop in half is

$$\frac{7.23\text{ s}}{0.904\text{ s}} = 8$$

ASSESS That the number of oscillations is much greater than 1 tells us that the damping is weak, justifying our use of the undamped period. It also tells us that those are really bad shocks!

13.7 Driven Oscillations and Resonance

Pushing a child on a swing, you can build up a large amplitude by giving a relatively small push once each oscillation cycle. If your pushing were not in step with the swing's natural oscillatory motion, then the same force would have little effect.

When an external force acts on an oscillatory system, we say that the system is **driven**. Consider a mass-spring system, which you might drive as suggested in Fig. 13.24. Suppose the driving force is given by $F_0 \cos \omega_d t$, where ω_d is called the **driving frequency**. Then Newton's law is

$$m\frac{d^2x}{dt^2} = -kx - b\frac{dx}{dt} + F_0 \cos \omega_d t \qquad (13.18)$$

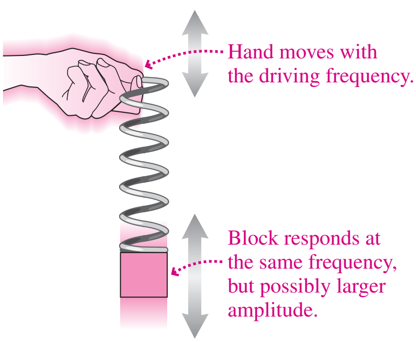


FIGURE 13.24 Driving a mass-spring system results in a large amplitude if the driving frequency is near the natural frequency $\sqrt{k/m}$.

where the first term on the right-hand side is the restoring force, the second the damping force, and the third the driving force. Since the system is being driven at the frequency ω_d, we expect it to undergo oscillatory motion at this frequency. So we guess that the solution to Equation 13.18 might have the form

$$x = A\cos(\omega_d t + \phi)$$

Substituting this expression and its derivatives into Equation 13.18 (see Problem 79) shows that the equation is satisfied if

$$A(\omega) = \frac{F_0}{m\sqrt{(\omega_d^2 - \omega_0^2)^2 + b^2\omega_d^2/m^2}} \qquad (13.19)$$

where ω_0 is the undamped **natural frequency** $\sqrt{k/m}$, as distinguished from the driving frequency ω_d.

Figure 13.25 shows **resonance curves**—plots of Equation 13.19 as a function of driving frequency—for three values of the damping constant. As long as the system is underdamped, the curve has a maximum at some nonzero frequency, and for weak damping, that maximum occurs at very nearly the natural frequency. The weaker the damping, the more sharply peaked is the resonance curve. Thus, in weakly damped systems, it's possible to build up large-amplitude oscillations with relatively small driving forces—a phenomenon known as **resonance**.

Most physical systems, from molecules to cars, and loudspeakers to buildings and bridges, exhibit one or more natural modes of oscillation. If these oscillations are weakly damped, then the buildup of large-amplitude oscillations through resonance can cause serious problems—sometimes even destroying the system (Fig. 13.26). Engineers designing complex structures spend a lot of their time exploring all possible oscillation modes and taking steps to avoid resonance. In an earthquake-prone area, for example, a building's natural frequencies would be designed to avoid the frequency of typical earthquake motions. A loudspeaker should be engineered so its natural frequency isn't in the range of sound it's intended to reproduce. Damping systems such as the shock absorbers of Example 13.6 or the tuned mass damper of Example 13.2 help limit resonant oscillations in cases where natural frequencies aren't easily altered.

Resonance is also important in microscopic systems. The resonant behavior of electrons in a special tube called a magnetron produces the microwaves that cook food in a microwave oven; the same resonant process heats ionized gases in some experiments designed to harness fusion energy. Carbon dioxide in Earth's atmosphere absorbs infrared radiation because CO_2 molecules—acting like miniature mass-spring systems—resonate at some of the frequencies of infrared radiation. The result is the greenhouse effect, which now threatens Earth with significant climatic change. The process called nuclear magnetic resonance (NMR) uses the resonant behavior of protons to probe the structure of matter and is the basis of magnetic resonance imaging (MRI) used in medicine. In NMR, the resonance involves the natural precession frequency of the protons due to magnetic torques; we described the classic model of this process in Chapter 11.

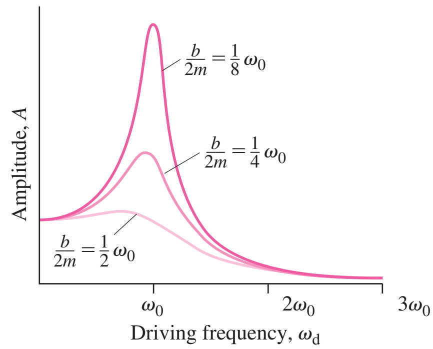


FIGURE 13.25 Resonance curves for three damping strengths; ω_0 is the undamped natural frequency $\sqrt{k/m}$.

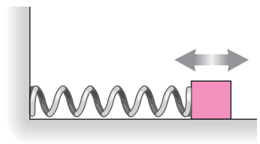
FIGURE 13.26 Collapse of the Tacoma Narrows Bridge—only four months after its opening in 1940—followed the resonant growth of large-amplitude oscillations.

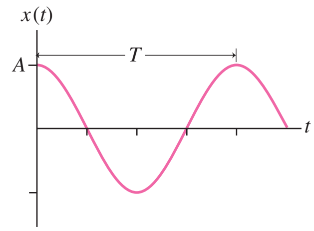
CHAPTER 13 SUMMARY

Big Picture

The big idea here is **simple harmonic motion** (SHM), oscillatory motion that is ubiquitous and that occurs whenever a disturbance from equilibrium results in a restoring force or torque that is directly proportional to the displacement. Position in SHM is a sinusoidal function of time:

$$x(t) = A\cos\omega t$$





Key Concepts and Equations

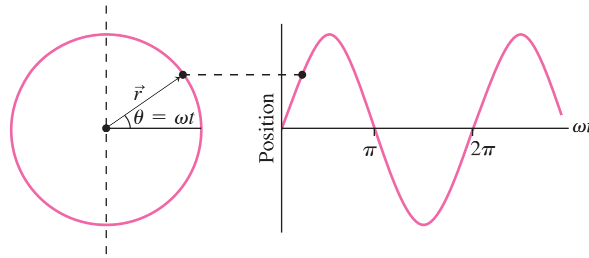
Period T is the time to complete one oscillation cycle; its inverse is **frequency**, or number of oscillations per unit time:

$$f = \frac{1}{T}$$

Another measure of frequency is **angular frequency** ω, given by

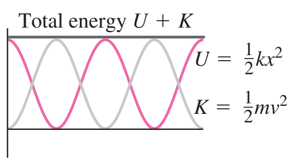
$$\omega = 2\pi f = \frac{2\pi}{T}$$

Angular frequency can be understood in terms of the close relationship between circular motion and simple harmonic motion.



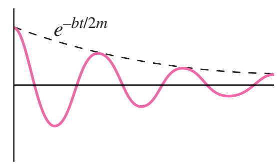
In the absence of friction and other dissipative forces, energy in SHM is conserved, although it's transformed back and forth between kinetic and potential forms:

$$E = \tfrac{1}{2}mv^2 + \tfrac{1}{2}kx^2 = \text{constant}$$



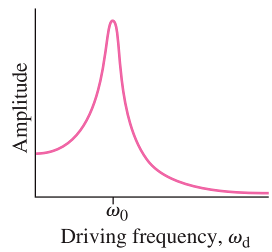
When dissipative forces act, the motion is **damped**. For small dissipative forces the oscillation amplitude decreases over time:

$$x(t) = Ae^{-bt/2m}\cos(\omega t + \phi)$$



If a system is driven at a frequency near its natural oscillation frequency ω_0, then large-amplitude oscillations can build; this is **resonance**. The amplitude A depends on the driving force F_0, the driving frequency ω_d, the natural frequency $\omega_0 = \sqrt{k/m}$, and the damping constant b:

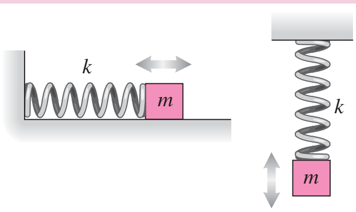
$$A(\omega) = \frac{F_0}{m\sqrt{(\omega_d^2 - \omega_0^2)^2 + b^2\omega_d^2/m^2}}$$



Applications

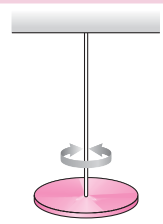
In mass-spring systems, the angular frequency is given by

$$\omega = \sqrt{\frac{k}{m}}$$



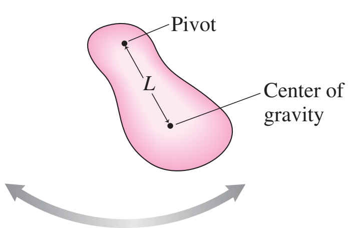
In systems involving rotational oscillations, the analogous relation involves the torsional constant and rotational inertia:

$$\omega = \sqrt{\frac{\kappa}{I}}$$



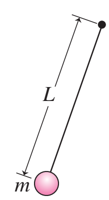
A special case is the **pendulum**, for which (with small-amplitude oscillations)

$$\omega = \sqrt{\frac{mgL}{I}}$$



In the case of a **simple pendulum**, the angular frequency reduces to

$$\omega = \sqrt{\frac{g}{L}}$$



For Thought and Discussion

1. Is a vertically bouncing ball an example of oscillatory motion? Of simple harmonic motion? Explain.
2. The vibration frequencies of molecules are much higher than those of macroscopic mechanical systems. Why?
3. What happens to the frequency of a simple harmonic oscillator when the spring constant is doubled? When the mass is doubled?
4. If the spring of a simple harmonic oscillator is cut in half, what happens to the frequency?
5. How does the frequency of a simple harmonic oscillator depend on its amplitude?
6. How would the frequency of a horizontal mass-spring system change if it were taken to the Moon? Of a vertical mass-spring system? Of a simple pendulum?
7. When in its cycle is the acceleration of an undamped simple harmonic oscillator zero? When is the velocity zero?
8. Explain how simple harmonic motion might be used to determine the masses of objects in an orbiting spacecraft.
9. One pendulum consists of a solid rod of mass m and length L, and another consists of a compact ball of the same mass m on the end of a massless string of the same length L. Which has the greater period? Why?
10. The x- and y-components of motion of a body are both simple harmonic with the same frequency and amplitude. What shape is the path of the body if the component motions are (a) in phase, (b) $\pi/2$ out of phase, and (c) $\pi/4$ out of phase?
11. Why is critical damping desirable in a car's suspension?
12. Explain why the frequency of a damped system is lower than that of the equivalent undamped system.
13. Opera singers have been known to break glasses with their voices. How?
14. What will happen to the period of a mass-spring system if it's placed in a jetliner accelerating down a runway? What will happen to the period of a pendulum in the same situation?
15. How can a system have more than one resonant frequency?

Exercises and Problems

Exercises

Section 13.1 Describing Oscillatory Motion

16. BIO A doctor counts 77 heartbeats in 1 minute. What are the corresponding period and frequency?
17. A violin string playing the note A oscillates at 440 Hz. What's its oscillation period?
18. The vibration frequency of a hydrogen chloride molecule is 8.66×10^{13} Hz. How long does it take the molecule to complete one oscillation?
19. Write expressions for simple harmonic motion (a) with amplitude 10 cm, frequency 5.0 Hz, and maximum displacement at $t = 0$, and (b) with amplitude 2.5 cm, angular frequency $5.0\ \mathrm{s}^{-1}$, and maximum velocity at $t = 0$.
20. The top of a skyscraper sways back and forth, completing 9 oscillation cycles in 1 minute. Find the period and frequency of its motion.
21. BIO A hummingbird's wings vibrate at about 45 Hz. What's the corresponding period?

Section 13.2 Simple Harmonic Motion

22. A 200-g mass is attached to a spring of constant $k = 5.6$ N/m and set into oscillation with amplitude $A = 25$ cm. Determine (a) the frequency in hertz, (b) the period, (c) the maximum velocity, and (d) the maximum force in the spring.
23. An automobile suspension has an effective spring constant of 26 kN/m, and the car's suspended mass is 1900 kg. In the absence of damping, with what frequency and period will the car undergo simple harmonic motion?
24. The quartz crystal in a watch executes simple harmonic motion at 32,768 Hz. (This is 2^{15} Hz, chosen so that 15 divisions by 2 give a signal at 1.00000 Hz.) If each face of the crystal undergoes a maximum displacement of 100 nm, find the maximum velocity and acceleration of the crystal faces.
25. A 50-g mass is attached to a spring and undergoes simple harmonic motion. Its maximum acceleration is 15 m/s^2 and its maximum speed is 3.5 m/s. Determine the (a) angular frequency, (b) spring constant, and (c) amplitude.
26. A particle undergoes simple harmonic motion with amplitude 25 cm and maximum speed 4.8 m/s. Find the (a) angular frequency, (b) period, and (c) maximum acceleration.
27. A particle undergoes simple harmonic motion with maximum speed 1.4 m/s and maximum acceleration 3.1 m/s^2. Find the (a) angular frequency, (b) period, and (c) amplitude.

Section 13.3 Applications of Simple Harmonic Motion

28. How long should you make a simple pendulum so its period is (a) 200 ms, (b) 5.0 s, and (c) 2.0 min?
29. At the heart of a grandfather clock is a simple pendulum 1.45 m long; the clock ticks each time the pendulum reaches its maximum displacement in either direction. What's the time interval between ticks?
30. A 640-g hollow ball 21 cm in diameter is suspended by a wire and is undergoing torsional oscillations at 0.78 Hz. Find the torsional constant of the wire.
31. A meter stick is suspended from one end and set swinging. Find the period of the resulting small-amplitude oscillations.

Section 13.4 Circular and Harmonic Motion

32. A wheel rotates at 600 rpm. Viewed from the edge, a point on the wheel appears to undergo simple harmonic motion. What are (a) the frequency in Hz and (b) the angular frequency for this SHM?
33. The x- and y-components of an object's motion are harmonic with frequency ratio 1.75:1. How many oscillations must each component undergo before the object returns to its initial position?

Section 13.5 Energy in Simple Harmonic Motion

34. A 450-g mass on a spring is oscillating at 1.2 Hz, with total energy 0.51 J. What's the oscillation amplitude?
35. A torsional oscillator of rotational inertia 1.6 kg·m^2 and torsional constant 3.4 N·m/rad has total energy 4.7 J. Find its maximum angular displacement and maximum angular speed.
36. You're riding in a friend's 1400-kg car with bad shock absorbers, bouncing down the highway at 20 m/s and executing vertical SHM with amplitude 18 cm and frequency 0.67 Hz. Concerned about fuel efficiency, your friend wonders what percentage of the car's kinetic energy is tied up in this oscillation. Make an estimate, neglecting the wheels' rotational energy and the fact that not all of the car's mass participates in the oscillation.

Sections 13.6 and 13.7 Damped Harmonic Motion and Resonance

37. The vibration of a piano string can be described by an equation analogous to Equation 13.17. If the quantity analogous to $b/2m$ in that equation has the value $2.8\ \mathrm{s}^{-1}$, how long will it take the amplitude to drop to half its original value?
38. A mass-spring system has $b/m = \omega_0/5$, where b is the damping constant and ω_0 the natural frequency. How does its amplitude at ω_0 compare with its amplitude when driven at frequencies 10% above and below ω_0?
39. A car's front suspension has a natural frequency of 0.45 Hz. The car's front shock absorbers are worn and no longer provide critical damping. The car is driving on a bumpy road with bumps 40 m apart. At a certain speed, the driver notices that the car begins to shake violently. What is this speed?

Problems

40. A simple model for carbon dioxide consists of three mass points (atoms) connected by two springs (electric forces), as shown in Fig. 13.27. One way this system can oscillate is if the carbon atom stays fixed and the two oxygens move symmetrically on either side of it. If the frequency of this oscillation is 4.0×10^{13} Hz, what's the effective spring constant? (*Note:* The mass of an oxygen atom is 16 u.)

FIGURE 13.27 Problem 40

41. Two identical mass-spring systems consist of 430-g masses on springs of constant $k = 2.2$ N/m. Both are displaced from equilibrium, and the first is released at time $t = 0$. How much later should the second be released so their oscillations differ in phase by $\pi/2$?
42. BIO The human eye and muscles that hold it can be modeled as a mass-spring system with typical values $m = 7.5$ g and $k = 2.5$ kN/m. What's the resonant frequency of this system? Shaking your head at this frequency blurs vision, as the eyeball undergoes resonant oscillations.
43. A mass m slides along a frictionless horizontal surface at speed v_0. It strikes a spring of constant k attached to a rigid wall, as shown in Fig. 13.28. After an elastic encounter with the spring, the mass heads back in the direction it came from. In terms of k, m, and v_0, determine (a) how long the mass is in contact with the spring and (b) the spring's maximum compression.

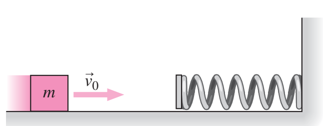


FIGURE 13.28 Problem 43

44. Show by substitution that $x(t) = A \sin \omega t$ is a solution to Equation 13.3.
45. A physics student, bored by a lecture on simple harmonic motion, idly picks up his pencil (mass 9.2 g, length 17 cm) by the tip with his frictionless fingers, and allows it to swing back and forth with small amplitude. If the pencil completes 6279 full cycles during the lecture, how long does the lecture last?
46. A pendulum of length L is mounted in a rocket. Find its period if the rocket is (a) at rest on its launch pad; (b) accelerating upward with acceleration $a = \frac{1}{2}g$; (c) accelerating downward with $a = \frac{1}{2}g$; and (d) in free fall.
47. BIO The protein dynein powers the flagella that propel some unicellular organisms. Biophysicists have found that dynein is intrinsically oscillatory, and that it exerts peak forces of about 1.0 pN when it attaches to structures called microtubules. The resulting oscillations have amplitude 15 nm. (a) If this system is modeled as a mass-spring system, what's the associated spring constant? (b) If the oscillation frequency is 70 Hz, what's the effective mass?
48. A mass is attached to a vertical spring, which then goes into oscillation. At the high point of the oscillation, the spring is in the original unstretched equilibrium position it had before the mass was attached; the low point is 5.8 cm below this. Find the oscillation period.
49. Derive the period of a simple pendulum by considering the horizontal displacement x and the force acting on the bob, rather than the angular displacement and torque.
50. A solid disk of radius R is suspended from a spring of spring constant k and torsional constant κ, as shown in Fig. 13.29. In terms of k and κ, what value of R will give the same period for the vertical and torsional oscillations of this system?

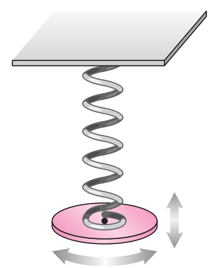


FIGURE 13.29 Problem 50

51. A thin steel beam 8.0 m long is suspended from a crane and is undergoing torsional oscillations. Two 75-kg steelworkers leap onto opposite ends of the beam, as shown in Fig. 13.30. If the frequency of torsional oscillations diminishes by 20%, what's the beam's mass?

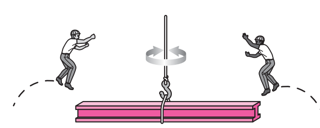


FIGURE 13.30 Problem 51

52. A cyclist turns her bicycle upside down to repair it. She then notices that the front wheel is executing a slow, small-amplitude, back-and-forth rotational motion with period 12 s. Treating the wheel as a thin ring of mass 600 g and radius 30 cm, whose only irregularity is the tire valve stem, determine the mass of the valve stem.
53. An object undergoes simple harmonic motion in two mutually perpendicular directions, its position given by $\vec{r} = A \sin \omega t \hat{\imath} + A \cos \omega t \hat{\jmath}$. (a) Show that the object remains a fixed distance from the origin (i.e., that its path is circular), and find that distance. (b) Find an expression for the object's velocity. (c) Show that the speed remains constant, and find its value. (d) Find the angular speed of the object in its circular path.

54. BIO The muscles that drive insect wings minimize the energy needed for flight by "choosing" to move at the natural oscillation frequency of the wings. Biologists study this phenomenon by clipping an insect's wings to reduce their mass. If the wing system is modeled as a simple harmonic oscillator, by what percent will the frequency change if the wing mass is decreased by 25%? Will it increase or decrease?
55. A pendulum consists of a 320-g solid ball 15.0 cm in diameter, suspended by an essentially massless string 80.0 cm long. Calculate the period of this pendulum, treating it first as a simple pendulum and then as a physical pendulum. What's the error in the simple-pendulum approximation? (*Hint:* Remember the parallel-axis theorem.)
56. If Jane and Tarzan are initially 8.0 m apart in Fig. 13.12, and Jane's mass is 60 kg, what's the maximum tension in the vine, and at what point does it occur?
57. BIO A *small mass measuring device* (SMMD) used for research on the biological effects of spaceflight consists of a small spring-mounted cage. Rats or other small subjects are introduced into the cage, which is set into oscillation. Calibration of a SMMD gives a linear function for the square of the oscillation period versus the subject's mass m in kg: $T^2 = 4.0\ \mathrm{s}^2 + (5.0\ \mathrm{s}^2/\mathrm{kg})m$. Find (a) the spring constant and (b) the mass of the cage alone.
58. A thin, uniform hoop of mass M and radius R is suspended from a horizontal rod and set oscillating with small amplitude, as shown in Fig. 13.31. Show that the period of the oscillations is $2\pi\sqrt{2R/g}$. (*Hint:* You may find the parallel-axis theorem useful.)

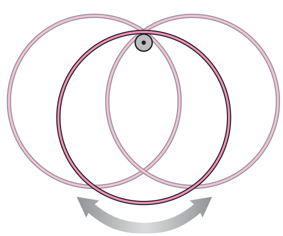


FIGURE 13.31 Problem 58

59. A mass m is mounted between two springs with constants k_1 and k_2, as shown in Figure 13.32. Show that the angular frequency of oscillation is $\omega = \sqrt{(k_1 + k_2)/m}$.

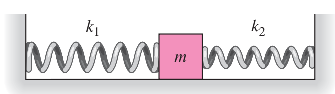


FIGURE 13.32 Problem 59

60. The equation for an ellipse is $(x^2/a^2) + (y^2/b^2) = 1$. Show that two-dimensional simple harmonic motion whose components have different amplitudes and are $\pi/2$ out of phase gives rise to elliptical motion. How are constants a and b related to the amplitudes?
61. Show that the potential energy of a simple pendulum is proportional to the square of the angular displacement in the small-amplitude limit.
62. The total energy of a mass-spring system is the sum of its kinetic and potential energy: $E = \frac{1}{2}mv^2 + \frac{1}{2}kx^2$. Assuming E remains constant, differentiate both sides of this expression with respect to time and show that Equation 13.3 results. (*Hint:* Remember that $v = dx/dt$.)
63. A solid cylinder of mass M and radius R is mounted on an axle through its center. The axle is attached to a horizontal spring of constant k, and the cylinder rolls back and forth without slipping (Fig. 13.33). Write the statement of energy conservation for this system, and differentiate it to obtain an equation analogous to Equation 13.3 (see Problem 62). Comparing your result with Equation 13.3, determine the angular frequency of the motion.

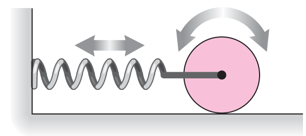


FIGURE 13.33 Problem 63

64. A mass m is free to slide on a frictionless track whose height y as a function of horizontal position x is $y = ax^2$, where a is a constant with units of inverse length. The mass is given an initial displacement from the bottom of the track and then released. Find an expression for the period of the resulting motion.
65. A 250-g mass is mounted on a spring of constant $k = 3.3$ N/m. The damping constant for this system is $b = 8.4\times10^{-3}$ kg/s. How many oscillations will the system undergo before the amplitude decays to $1/e$ of its original value?
66. A harmonic oscillator is underdamped if the damping constant b is less than $\sqrt{2}m\omega_0$, where ω_0 is the natural frequency of undamped motion. Show that for an underdamped oscillator, Equation 13.19 has a maximum at a driving frequency less than ω_0.
67. A massless spring with $k = 74$ N/m hangs from the ceiling. A 490-g mass is hooked onto the unstretched spring and allowed to drop. Find (a) the amplitude and (b) the period of the resulting motion.
68. A meter stick is suspended from a frictionless rod through a small hole at the 25-cm mark. Find the period of small-amplitude oscillations about the stick's equilibrium position.
69. A particle of mass m has potential energy given by $U = ax^2$, where a is a constant and x is the particle's position. Find an expression for the frequency of simple harmonic oscillations this particle undergoes.
70. Two balls each of unknown mass m are mounted on opposite ends of a 1.5-m-long rod of mass 850 g. The system is suspended from a wire attached to the center of the rod and set into torsional oscillations. If the wire has torsional constant 0.63 N·m/rad and the period of the oscillations is 5.6 s, what's the unknown mass m?
71. Two mass-spring systems with the same mass are undergoing oscillatory motion with the same amplitudes. System 1 has twice the frequency of system 2. How do (a) their energies and (b) their maximum accelerations compare?
72. Two mass-spring systems have the same mass and the same total energy. The amplitude of system 1 is twice that of system 2. How do (a) their frequencies and (b) their maximum accelerations compare?
73. A 500-g mass is suspended from a thread 45 cm long that can sustain a tension of 6.0 N before breaking. Find the maximum allowable amplitude for pendulum motion of this system.
74. A 500-g block on a frictionless, horizontal surface is attached to a rather limp spring with $k = 8.7$ N/m. A second block rests on the first, and the whole system executes simple harmonic motion with period 1.8 s. When the amplitude of the motion is increased to 35 cm, the upper block just begins to slip. What's the coefficient of static friction between the blocks?

75. Repeat Problem 64 for a small solid ball of mass M and radius R that rolls without slipping on the parabolic track.

76. You're working on the script of a movie whose plot involves a hole drilled straight through Earth's center and out the other side. You're asked to determine what will happen if a person falls into the hole. You find that the gravitational acceleration *inside* Earth points toward Earth's center, with magnitude given approximately by $g(r) = g_0(r/R_E)$, where g_0 is the surface value, r is the distance from Earth's center, and R_E is Earth's radius. What do you report for the person's motion, including equations and values for any relevant parameters?

77. A 1.2-kg block rests on a frictionless surface and is attached to a horizontal spring of constant $k = 23$ N/m (Fig. 13.34). The block oscillates with amplitude 10 cm and phase constant $\phi = -\pi/2$. A block of mass 0.80 kg moves from the right at 1.7 m/s and strikes the first block when the latter is at the rightmost point in its oscillation. The two blocks stick together. Determine the frequency, amplitude, and phase constant (relative to the *original* $t = 0$) of the resulting motion.

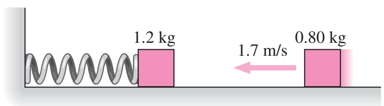


FIGURE 13.34 Problem 77

78. A disk of radius R is suspended from a pivot somewhere between its center and edge (Fig. 13.35). For what pivot point will the period of this physical pendulum be a minimum?

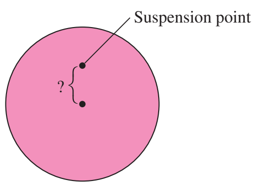


FIGURE 13.35 Problem 78

79. Show by direct substitution that $x = A\cos(\omega_d t + \phi)$ satisfies Equation 13.18 with A given by Equation 13.19.

80. You're a structural engineer working on a design for a steel beam, and you need to know its resonant frequency. You test the beam by clamping one end and deflecting the other so it bends, and you determine the associated potential energy. The table below gives the results:

Beam deflection x (cm)	Potential energy U (J)
−4.54	164
−3.49	141
−2.62	71.9
−1.22	9.15
−0.448	0.162
0	0
0.730	4.13
1.29	16.3
2.13	34.0
3.39	115
4.70	225

Using a spreadsheet or other software, plot U versus x^2, and find the best-fit line. Ignore any constant term or, if your software permits, constrain the line to go through $(0, 0)$ since you know $U = 0$ when $x = 0$. Then, assuming an effective beam mass of 3750 kg, find the resonant frequency.

81. Show that $x(t) = a\cos\omega t - b\sin\omega t$ represents simple harmonic motion, as in Equation 13.8, with $A = \sqrt{a^2 + b^2}$ and $\phi = \tan^{-1}(b/a)$.

82. You're working for the summer with an ornithologist who knows you've studied physics. She asks you for a noninvasive way to measure birds' masses. You propose using a bird feeder in the shape of a 50-cm-diameter disk of mass 340 g, suspended by a wire with torsional constant 5.00 N·m/rad (Fig. 13.36). Two birds land on opposite sides and the feeder goes into torsional oscillation at 2.6 Hz. Assuming the birds have the same mass, what is it?

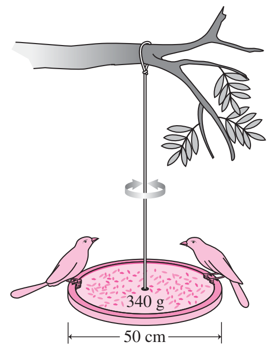


FIGURE 13.36 Problem 82

83. While waiting for your plane to take off, you suspend your keys from a thread and set the resulting pendulum oscillating. It completes exactly 90 cycles in 1 minute. You repeat the experiment as the plane accelerates down the runway, and now measure exactly 91 cycles in 1 minute. Find the plane's acceleration.

84. You're working for a playground equipment company, which wants to know the rotational inertia of its swing with a child on board; the combined mass is 20 kg. You observe the child twirling around in the swing, twisting the ropes as shown in Fig. 13.37. As a result, child and swing rise slightly, with the rise h in cm equal to the square of the number of full turns. When the child stops twisting, the swing begins torsional oscillations. You measure the period at 6.91 s. What do you report for the rotational inertia of the child-swing system?

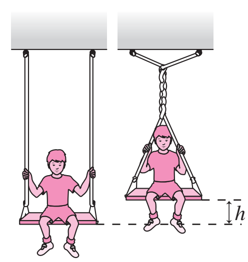


FIGURE 13.37 Problem 84

Passage Problems

Physicians and physiologists are interested in the long-term effects of apparent weightlessness on the human body. Among these effects are redistribution of body fluids to the upper body, loss of muscle tone, and overall mass loss. One method of measuring mass in the apparent weightlessness of an orbiting spacecraft is to strap the astronaut into a chair-like device mounted on springs (Fig. 13.38). This *body mass measuring device* (BMMD) is set oscillating in simple harmonic motion, and measurement of the oscillation period, along with the known spring constant and mass of the chair itself, then yields the astronaut's mass. When a 60-kg astronaut is strapped into the 20-kg chair, the time for three oscillation periods is measured to be 6.0 s.

FIGURE 13.38 Astronaut Tamara Jernigan uses a body mass measuring device in the Spacelab Life Sciences Module (Passage Problems 85–88)

85. If a 90-kg astronaut is "weighed" with this BMMD, the time for three periods will be
    a. 50% longer.
    b. shorter by less than 50%.
    c. longer by less than 50%.
    d. longer by more than 50%.
86. If the same device were used on Earth, the results for a given astronaut (assuming mass hasn't yet been lost in space) would be
    a. the same.
    b. greater than in an orbiting spacecraft.
    c. less than in an orbiting spacecraft.
    d. meaningless, because the device won't work on Earth.
87. If an astronaut's mass declines linearly with time while she's in orbit, the oscillation period of the BMMD will
    a. decrease at an ever-decreasing rate.
    b. decrease linearly with time.
    c. decrease at an ever-increasing rate.
    d. increase linearly with time.
88. The spring constant for the BMMD described here is
    a. 80 N/m.
    b. 80π N/m.
    c. 2 N/m.
    d. $80\pi^2$ N/m.
    e. none of the above.

Answers to Chapter Questions

Answer to Chapter Opening Question

The dancers undergo pendulum motion, whose period is determined entirely by their rope lengths and the acceleration of gravity.

Answers to GOT IT? Questions

13.1. Frequencies and periods are the same; amplitudes and phase constants are different because of the different initial displacements and times of release, respectively.
13.2. (a) No change; (b) doubles; (c) doubles.
13.3. (a) 1:2; (b) 3:2.
13.4. The more energetic oscillator has (a) twice the mass and (b) twice the spring constant. (c) Their maximum speeds are equal.

14 Wave Motion

Ocean waves travel thousands of kilometers across the open sea before breaking on shore. How much water moves with the waves?

New Concepts, New Skills

By the end of this chapter you should be able to

- Explain the nature of waves as traveling disturbances that transport energy but not matter (14.1).
- Describe waves mathematically in terms of amplitude, period, frequency, and wavelength (14.2).
- Understand the wave equation and use it to determine wave speeds (14.2).
- Evaluate wave energies and intensities (14.3).
- Describe specific wave types, including waves on strings and sound waves (14.3, 14.4).
- Explain interference and reflection and how these lead to standing waves (14.5–14.7).
- Describe the Doppler effect quantitatively and shock waves qualitatively (14.8).

Connecting Your Knowledge

- Waves are intimately associated with simple harmonic motion, so make sure you understand the terminology and quantitative descriptions of SHM in Chapter 13 (13.1, 13.2).

Humans and other animals communicate using sound waves. Light and related waves enable us to visualize our surroundings and provide virtually all our information about the universe beyond Earth. Our cell phones keep us connected via radio waves. Physicians probe our bodies with ultrasound waves. Radio waves connect our wireless laptops to the Internet and cook the food in our microwave ovens. Earthquakes trigger waves in the solid Earth and may generate dangerous tsunamis. **Wave motion** is an essential feature of our physical environment.

All these examples involve a disturbance that moves or **propagates** through space. The disturbance carries energy, but not matter. Air doesn't move from your mouth to a listener's ear, but sound energy does. Water doesn't move across the open ocean, but wave energy does. **A wave is a traveling disturbance that transports energy but not matter.**

14.1 Waves and Their Properties

In this chapter we'll deal with **mechanical waves**, which are disturbances of some material **medium**, such as air, water, a violin string, or Earth's interior. Visible and infrared light waves, radio waves, ultraviolet and X rays, in contrast, are **electromagnetic waves**. They share many properties with mechanical waves, but they don't require a material medium. We'll treat electromagnetic waves in Chapters 29–32.

Mechanical waves occur when a disturbance in one part of a medium is communicated to adjacent parts. Figure 14.1 shows a multiple mass-spring system that serves as a model for many types of mechanical waves. Disturb one mass, and it goes into simple harmonic motion. But because the masses are connected, that motion is communicated to the adjacent mass. As a result, both the disturbance and its associated energy propagate along the mass-spring system, disturbing successive masses as they go.

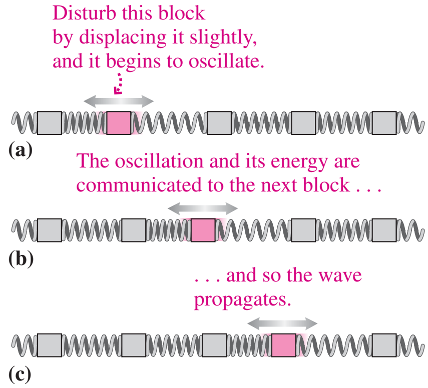


FIGURE 14.1 Wave propagation in a mass-spring system.

✓TIP Wave Motions

A wave moves energy from place to place, but not matter. However, that doesn't mean that the matter making up the wave medium doesn't move. It does, undergoing localized oscillatory motion as the wave passes. But once the wave is gone, the disturbed matter returns to its equilibrium state. Don't confuse this localized motion of the medium with the motion of the wave itself. Both occur, but only the latter carries energy from one place to another.

Longitudinal and Transverse Waves

In Fig. 14.1, we disturbed the system by displacing one block so its subsequent oscillations were back and forth along the structure—in the same direction as the wave propagation. The result is a **longitudinal wave**. Sound is a longitudinal wave, as we'll see in Section 14.4. We could equally well displace a mass at right angles, as in Fig. 14.2. Then we get a **transverse wave**, whose disturbance is at right angles to the wave propagation. Some waves include both longitudinal and transverse motions, as shown for a water wave in Fig. 14.3.

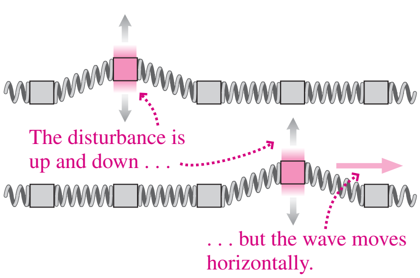


FIGURE 14.2 A transverse wave.

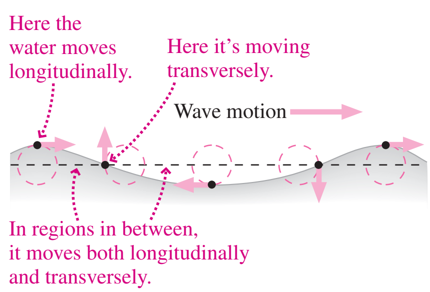


FIGURE 14.3 A water wave has both longitudinal and transverse components.

Amplitude and Waveform

The maximum value of a wave's disturbance is the wave **amplitude**. For a water wave, amplitude is the maximum height above the undisturbed water level; for a sound wave, it's the maximum excess air pressure; for the waves of Figs. 14.1 and 14.2, it's the maximum displacement of a mass.

Wave disturbances come in many shapes, called **waveforms** (Fig. 14.4). An isolated disturbance is a **pulse**, which occurs when the medium is disturbed only briefly. A **continuous wave** results from an ongoing periodic disturbance. Intermediate between these extremes is a **wave train**, resulting from a periodic disturbance lasting a finite time.

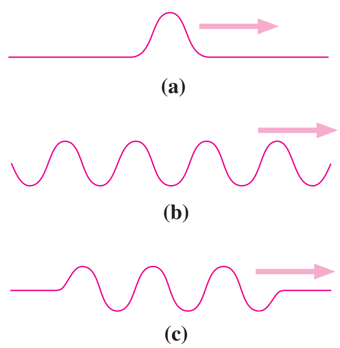


FIGURE 14.4 (a) A pulse, (b) a continuous wave, and (c) a wave train.

Wavelength, Period, and Frequency

A continuous wave repeats in both space and time. The **wavelength** λ is the *distance* over which the wave pattern repeats (Fig. 14.5). The wave **period** T is the *time* for one complete oscillation. The **frequency** f, or number of wave cycles per unit time, is the inverse of the period.

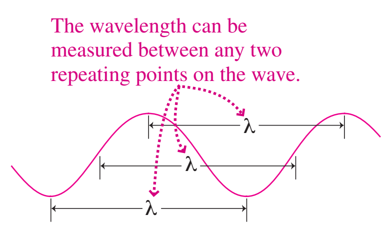


FIGURE 14.5 The wavelength λ is the distance over which the wave pattern repeats.

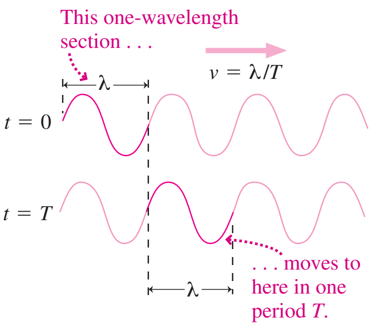


FIGURE 14.6 One full cycle passes a given point in one wave period T; the wave speed is therefore $v = \lambda/T$.

Wave Speed

A wave travels at a specific speed through its medium. The speed of sound in air is about 340 m/s. Small ripples on water move at about 20 cm/s, while earthquake waves travel at several kilometers per second. The physical properties of the medium ultimately determine the wave speed, as we'll see in Section 14.3.

Wave speed, wavelength, and period are related. In one wave period, a fixed observer sees one complete wavelength go by (Fig. 14.6). Thus, the wave moves one wavelength in one period, so its speed is

$$v = \frac{\lambda}{T} = \lambda f \qquad \text{(wave speed)} \tag{14.1}$$

where the second equality follows because period and frequency are inverses.

GOT IT? 14.1 A boat bobs up and down on a water wave, moving 2 m vertically in 1 s. A wave crest moves 10 m horizontally in 2 s. Is the wave speed (a) 2 m/s or (b) 5 m/s? Explain.

14.2 Wave Math

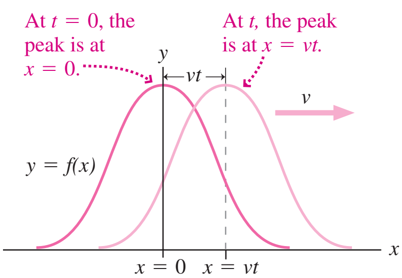


FIGURE 14.7 The wave pulse moves a distance vt in time t, but its shape stays the same.

Figure 14.7 shows "snapshots" of a wave pulse at time $t = 0$ and at some later time t. Initially the wave disturbance y is some function of position: $y = f(x)$. Later the pulse has moved to the right a distance vt, but its shape, described by the function f, is the same. We can represent this displaced pulse by replacing x with $x - vt$ as the argument of the function f. Then x has to be larger—by the amount vt—to give the same value of f as it did before. For example, this particular pulse peaks when the argument of f is zero. Initially, that occurred when x was zero. Replacing x by $x - vt$ ensures that the argument becomes zero when $x = vt$, putting the peak at this new position. As time increases, so does vt and therefore the value of x corresponding to the peak. Thus $f(x - vt)$ correctly represents the moving pulse.

Although we considered a single pulse, this argument applies to *any* function $f(x)$, including continuous waves: Replace the argument x with $x - vt$, and the function $f(x - vt)$ describes a wave moving in the positive x-direction with speed v. You can convince yourself that a function of the form $f(x + vt)$ describes a wave moving in the negative x-direction.

A particularly important case is a **simple harmonic wave**, for which a "snapshot" at time $t = 0$ shows a sinusoidal function. We'll choose coordinates so that $x = 0$ is at a maximum of the wave, making the function a cosine (Fig. 14.8*a*). Then $y(x, 0) = A\cos kx$, where A is the amplitude and k is a constant, called the **wave number**. We can find k because we know that the wave repeats in one wavelength λ. Since the period of the cosine function is 2π, we therefore want kx to be 2π when x equals λ. Then $k\lambda = 2\pi$, or

$$k = \frac{2\pi}{\lambda} \qquad \text{(wave number)} \tag{14.2}$$

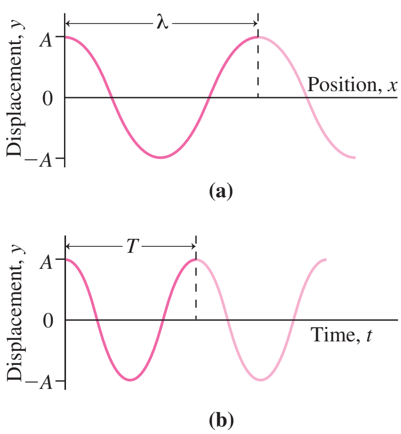


FIGURE 14.8 A sinusoidal wave (a) as a function of position at fixed time $t = 0$ and (b) as a function of time at fixed position $x = 0$.

To describe a wave moving with speed v, we replace x in the expression $A\cos kx$ with $x - vt$, giving $y(x, t) = A\cos[k(x - vt)]$. If we now sit at the point $x = 0$, we'll see an oscillation described by $y(0, t) = A\cos(-kvt) = A\cos(kvt)$, where the last step follows because $\cos(-x) = \cos x$. But we found that $k = 2\pi/\lambda$, and Equation 14.1 shows that $v = \lambda/T$, so the argument of the cosine function becomes $kvt = (2\pi/\lambda)(\lambda/T)t = 2\pi t/T$.

In Chapter 13, we introduced the **angular frequency** $\omega = 2\pi/T$ in describing simple harmonic motion; here the same quantity arises in describing wave motion. And no wonder: At a fixed point in space, the wave medium undergoes simple harmonic motion with angular frequency $\omega = 2\pi/T$ (Fig. 14.8*b*). Putting this all together, we can write a traveling sinusoidal wave in the form

$$y(x, t) = A\cos(kx \pm \omega t) \qquad \text{(sinusoidal wave)} \tag{14.3}$$

where we've written $\pm$ so we can describe a wave going in the positive x-direction ($-$ sign) or the negative x-direction ($+$ sign). The argument of the cosine is called the wave's **phase**. Note that k and ω are related to the more familiar wavelength λ and period T in the same way: $k = 2\pi/\lambda$ and $\omega = 2\pi/T$. Just as ω is a measure of frequency—oscillation cycles per unit *time*, with an extra factor of 2π—so is k a measure of **spatial frequency**—oscillation cycles per unit *distance*, again with that factor of 2π to make the math simpler. The relations between k, λ and ω, T allow us to rewrite the wave speed of Equation 14.1 in terms of k and ω:

$$v = \frac{\lambda}{T} = \frac{2\pi/k}{2\pi/\omega} = \frac{\omega}{k} \tag{14.4}$$

EXAMPLE 14.1 Describing a Wave: Surfing

A surfer paddles out beyond the breaking surf to where the waves are sinusoidal in shape, with crests 14 m apart. The surfer bobs a vertical distance 3.6 m from trough to crest, a process that takes 1.5 s. Find the wave speed, and describe the wave using Equation 14.3.

INTERPRET This is a problem about a simple harmonic wave—that is, a wave with sinusoidal shape.

DEVELOP We'll take $x = 0$ at the location of a wave crest when $t = 0$, so Equation 14.3, $y(x,t) = A\cos(kx \pm \omega t)$, applies. Let's take the positive x-direction toward shore, so we'll use the minus sign in Equation 14.3. In Fig. 14.9*a* we sketched a "snapshot" of the wave, showing the spatial information we're given. Figure 14.9*b* shows the temporal information.

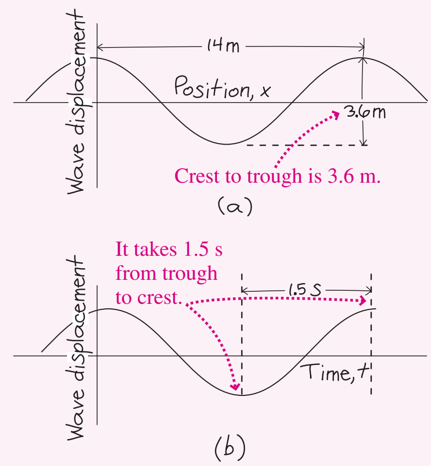


FIGURE 14.9 Our sketch of displacement versus (a) position and (b) time.

EVALUATE The 1.5-s trough-to-crest time in Fig. 14.9*b* is half the full crest-to-crest period T, so $T = 3.0$ s. The crest-to-crest distance in Fig. 14.9*a* is the wavelength λ, so $\lambda = 14$ m. Then Equation 14.1 gives

$$v = \frac{\lambda}{T} = \frac{14\text{ m}}{3.0\text{ s}} = 4.7\text{ m/s}$$

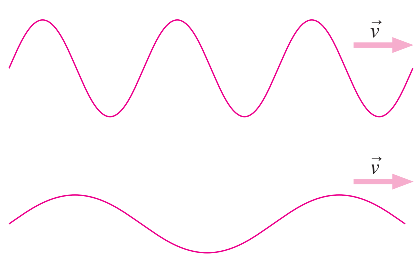
To describe the wave with Equation 14.3 we need the amplitude A, wave number k, and angular frequency ω. The amplitude is half the crest-to-trough displacement, or $A = 1.8$ m, as shown in Fig. 14.9*a*. The wave number k and angular frequency ω then follow from λ and T: $k = 2\pi/\lambda = 0.449\text{ m}^{-1}$ and $\omega = 2\pi/T = 2.09\text{ s}^{-1}$. Then the wave description is

$$y(x,t) = 1.8\cos(0.449x - 2.09t)$$

with y and x in meters and t in seconds.

ASSESS As a check on our answer, let's see whether our values of ω and k satisfy Equation 14.4: $v = \omega/k = 2.09\text{ s}^{-1}/0.449\text{ m}^{-1} = 4.7$ m/s. Thus the pairs λ, T and ω, k are equivalent ways to describe the same wave.

GOT IT? 14.2 The figure shows two waves propagating with the same speed. Which has the greater (a) amplitude, (b) wavelength, (c) period, (d) wave number, (e) frequency (f or ω)?



The Wave Equation

We argued our way to Equation 14.3 for a sinusoidal wave on mathematical grounds alone. Whether such waves are actually possible depends on the physical properties of the medium. Many media do, in fact, support waves as described by Equation 14.3. We'll explore one case in detail in the next section. More generally, physicists analyze the behavior of a medium in response to disturbances. Often the analysis results in an equation relating the space and time derivatives of the disturbed quantity:

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 y}{\partial t^2} \qquad \text{(wave equation)} \tag{14.5}$$

This is the **wave equation** for waves propagating in one dimension. Here y is the wave disturbance—the height of a water wave, the pressure in a sound wave, and so on. The quantity v is the wave speed, which usually appears as a combination of quantities related to properties of the medium, and thus allows physicists to deduce the wave speed. Because the wave disturbance is a function of the two variables x (spatial position) and t (time), the derivatives here are **partial derivatives**, designated with the symbol ∂ and indicating differentiation with respect to one variable while the other is held constant. Thus the wave equation is a **partial differential equation**. Solving such equations requires more advanced math courses, but you can show directly (Problem 71) that Equation 14.3 satisfies the wave equation, with wave speed $v = \omega/k$. More generally, *any* function of the form $f(x \pm vt)$ satisfies the wave equation, as you can show in Problem 72. You'll encounter the wave equation again in Chapter 29, when you study electromagnetic waves.

14.3 Waves on a String

Scientists and engineers generally explore wave possibilities in a medium by applying the laws of physics and deriving a wave equation similar to Equation 14.5. Such analysis reveals the wave speed and other wave properties. Here we'll take a simpler approach to one special case: transverse waves on a stretched string. Our results are directly applicable to musical instruments, climbing ropes, bridge cables, and other elongated structures.

Our string has mass per unit length μ in kilograms per meter, and it's stretched to a tension force F. Consider a wave pulse propagating to the right, as shown in Fig. 14.10*a*. We'll use Newton's law to analyze the string's motion and determine the speed of the pulse. It's easiest to do this in a frame of reference moving with the pulse; in that frame, the entire string moves *leftward* with the pulse speed v. At the pulse location, however, the string's motion deviates from horizontal as it rides up and down over the pulse (Fig. 14.10*b*).

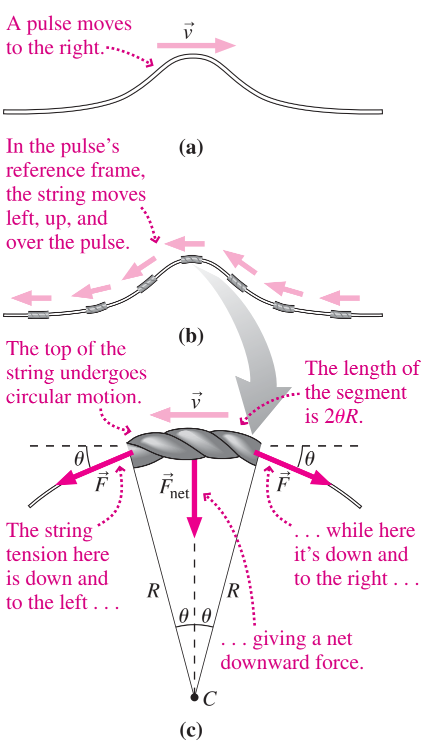


FIGURE 14.10 A wave pulse moving on a string.

Whatever the pulse shape, a small section at the top forms a circular arc of some radius R, as shown in Fig. 14.10*c*. Then the string right at the top of the pulse undergoes circular motion with speed v and radius R; if its mass is m, Newton's law requires that a force of magnitude mv^2/R act toward the center of curvature to keep the string on its circular path. This force is provided by the difference in the direction of the string tension between the two ends of the section; as Fig. 14.10*c* shows, the tension at each end contributes a downward component $F\sin\theta$. Then the net force on the segment has magnitude $2F\sin\theta$ and points toward the center of curvature.

Now we make an additional assumption: that the disturbance of the string is small, in the sense that the string remains almost horizontal even at the pulse. Then the angle θ is small, and we can apply the approximation $\sin\theta \simeq \theta$. Therefore, the net force on the string section becomes approximately $2F\theta$. Furthermore, the small-disturbance approximation means that the tension doesn't vary significantly from its undisturbed value, so F in this expression is essentially the same F we're using to characterize the tension throughout the string. Finally, our curved string section forms a circular arc whose length, from Fig. 14.10*c*, is $2\theta R$. Multiplying by the mass per unit length μ gives its mass: $m = 2\theta R\mu$. Now we can apply Newton's law, equating the net force $2F\theta$ to the mass times acceleration:

$$2F\theta = \frac{mv^2}{R} = \frac{2\theta R\mu v^2}{R} = 2\theta\mu v^2$$

Solving for the wave speed v then gives

$$v = \sqrt{\frac{F}{\mu}} \tag{14.6}$$

Does this make sense? The greater the tension F, the greater the string's acceleration, and the more rapidly the wave should propagate. The string's inertia, on the other hand, limits the acceleration, and therefore a greater mass per unit length should slow the wave. Equation 14.6, with F in the numerator and μ in the denominator, reflects both these trends.

We've made no assumptions here other than to assume that the disturbance is small. Therefore, Equation 14.6 applies to small-amplitude pulses, continuous waves, and wave trains of any shape.

EXAMPLE 14.2 **Wave Speed and Tension Force: Rock Climbing**

A 43-m-long rope of mass 5.0 kg joins two climbers. One climber strikes the rope, and 1.4 s later the second climber feels the effect. What's the rope tension?

INTERPRET We're asked for the rope tension. Although wave speed isn't mentioned explicitly, we just learned to relate wave speed and rope tension. Striking the rope produces a wave, which the second climber feels. We're given the time it takes that wave to propagate along the rope.

DEVELOP Equation 14.6, $v = \sqrt{F/\mu}$, gives the relations among rope tension, mass per unit length, and wave speed. Our plan is to solve for the rope tension, but first we need to find μ and v from the given information.

EVALUATE We're given the rope's mass m and length L, so its mass per unit length is $\mu = m/L$. We're given the time t for the wave to travel the rope length L, so the wave speed is $v = L/t$. Solving Equation 14.6 for F then gives

$$F = \mu v^2 = \left(\frac{m}{L}\right)\left(\frac{L}{t}\right)^2 = \frac{mL}{t^2} = \frac{(5.0\text{ kg})(43\text{ m})}{(1.4\text{ s})^2} = 110\text{ N}$$

ASSESS Is this number reasonable? A typical adult weights around 700 N, so the rope is supporting only a small fraction of the lower climber's weight—a reasonable situation.

Wave Power

Waves carry energy. For a wave on a string, the vertical component of the tension force does work that transfers energy along the string. Figure 14.11 shows that the vertical force on the string at the left side of the pulse is approximately $-F\theta$. As we showed in Chapter 6, power—the rate of doing work—is the product of force and velocity, so the power here is $P = -F\theta u$, where u is the vertical velocity of the string—*not* the wave speed. For a simple harmonic wave, the string velocity is the rate of change of its position $y(x, t) = A\cos(kx - \omega t)$:

$$u = \frac{dy}{dt} = A\omega\sin(kx - \omega t)$$

where we used the chain rule, differentiating cosine to $-$sine and then multiplying by the derivative, $-\omega$, of the cosine's argument $kx - \omega t$. As Fig. 14.11 shows, the tangent of the angle θ is the slope, dy/dx, of the string. For small angles, $\tan\theta \simeq \theta$, so $\theta \simeq dy/dx = -kA\sin(kx - \omega t)$. Putting these results for u and θ in our expression for power gives $P = -F\theta u = F\omega kA^2\sin^2(kx - \omega t)$. The sine term shows that the power fluctuates in space and time. Usually we're interested in the *average* power, $\overline{P} = \frac{1}{2}F\omega kA^2$, which follows because the average value of $\sin^2$ is $\frac{1}{2}$ (Fig. 14.12). We can give this a more physical meaning if we use Equations 14.4 and 14.6 to write $k = \omega/v$ and $F = \mu v^2$, with v the wave speed. Then we have

$$\overline{P} = \tfrac{1}{2}\mu\omega^2A^2v \qquad (14.7)$$

This equation gives the sensible result that wave power is directly proportional to the speed v at which energy moves along the wave.

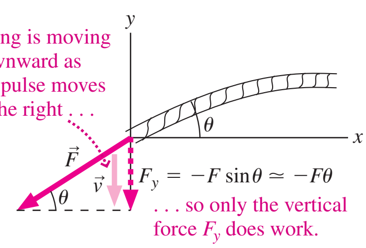


FIGURE 14.11 The vertical force component does work on the string; for small θ, $\sin\theta \simeq \theta$, so $F_y \simeq F\theta$.

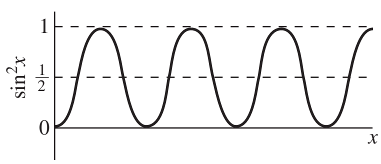


FIGURE 14.12 The function $\sin^2 x$ swings symmetrically between 0 and 1, so its average value is $\frac{1}{2}$.

Wave Intensity

Total power is useful in describing waves confined to narrow structures like strings for mechanical waves or optical fibers for electromagnetic waves. But for waves in three-dimensional media, like sound in air, it makes more sense to talk about the **intensity**, or the rate at which the wave carries energy across a unit area perpendicular to the wave propagation. Intensity is thus power per unit area, measured in watts per square meter (W/m^2).

Wavefronts are surfaces on which the wave phase is constant—for example, wave crests. A **plane wave** is one whose wavefronts are planes. Since the wave doesn't spread out, its intensity remains constant (Fig. 14.13*a*). But as waves propagate from a localized source, they spread and their intensity drops. **Spherical waves** originate from point sources and spherical wavefronts spread in all directions. Since the area of a sphere is $4\pi r^2$, the intensity of a spherical wave decreases as the inverse square of the distance from its source:

$$I = \frac{P}{A} = \frac{P}{4\pi r^2} \quad \text{(spherical wave)} \tag{14.8}$$

Note that energy isn't lost here; rather, the same energy is spread over ever-larger areas as the wave propagates (Fig. 14.13*b*). Table 14.1 lists some typical wave intensities.

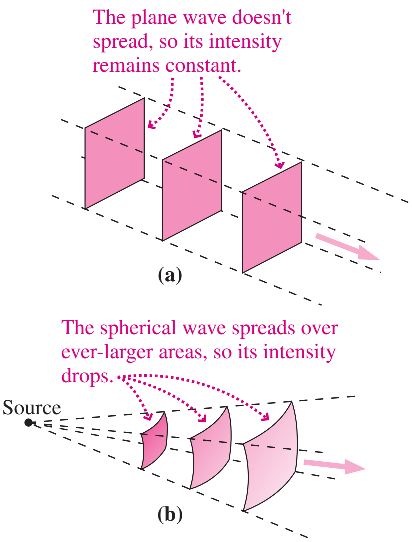


FIGURE 14.13 (a) Plane and (b) spherical waves.

Table 14.1 Wave Intensities

Wave	Intensity, W/m^2
Sound, 4 m from loud rock band	1
Sound, jet aircraft at 50 m	10
Sound, whisper at 1 m	10^{-10}
Light, sunlight at Earth's orbit	1368
Light, sunlight at Jupiter's orbit	50
Light, 1 m from typical camera flash	4000
Light, at target of laser fusion experiment	10^{18}
TV signal, 5 km from 50-kW transmitter	1.6×10^{-4}
Microwaves, inside microwave oven	6000
Earthquake wave, 5 km from Richter 7.0 quake	4×10^{4}

EXAMPLE 14.3 Evaluating Wave Intensity: A Reading Light

Your book is 1.9 m from a 75-watt lightbulb, and the light is barely adequate for reading. How far from a 40-W bulb would the book have to be to get the same intensity at the page?

INTERPRET This is a problem about wave intensity, and we identify the lightbulbs as sources of spherical waves.

DEVELOP Equation 14.8, $I = P/(4\pi r^2)$, gives the intensity. We want both bulbs to produce the same intensity, so we have $I = P_{75}/(4\pi r_{75}^2) = P_{40}/(4\pi r_{40}^2)$.

EVALUATE We then solve for the unknown distance r_{40}:

$$r_{40} = r_{75}\sqrt{\frac{P_{40}}{P_{75}}} = (1.9\text{ m})\sqrt{\frac{40\text{ W}}{75\text{ W}}} = 1.4\text{ m}$$

ASSESS Make sense? Although the 40-W bulb has only about half the power output, the decrease in distance isn't as great as you might expect because the intensity depends on the inverse *square* of the distance. ■

GOT IT? 14.3 Two identical stars are different distances from Earth, and the intensity of the light from the more distant star as received at Earth is only 1% that of the closer star. Is the more distant star (a) twice as far away, (b) 100 times as far away, (c) 10 times as far away, or (d) $\sqrt{10}$ times as far away?

14.4 Sound Waves

Sound waves are longitudinal mechanical waves that propagate through gases, liquids, and solids. Most familiar is sound in air. Here the wave disturbance comprises a small change in air pressure and density accompanied by a back-and-forth motion of the air (Fig. 14.14). The speed of sound in air and other gases depends on the background pressure P (force per unit area) and density ρ (mass per unit volume):

$$v = \sqrt{\frac{\gamma P}{\rho}} \tag{14.9}$$

where γ is a constant characteristic of the gas. For air and other diatomic gases, γ is $\frac{7}{5}$; for monatomic gases like helium, it's $\frac{5}{3}$. Sound propagates faster in liquids and solids because they're less compressible.

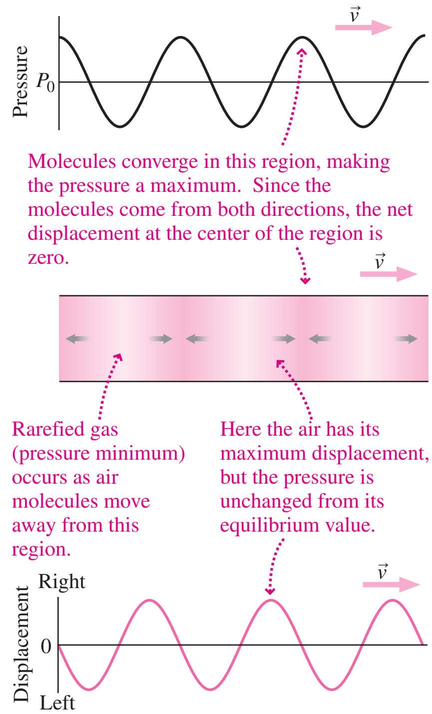


FIGURE 14.14 A sound wave consists of alternating regions of compression (higher density and pressure) and rarefaction (lower density and pressure) propagating through the air.

Sound and the Human Ear

The human ear responds to a wide range of sound intensities and frequencies, as shown in Fig. 14.15. Audible frequencies range from around 20 Hz to 20 kHz, although the upper limit drops with age. Figure 14.15 shows that the minimum intensity for audible sound increases at high and low frequencies; that's the reason for the "loudness" switch on your stereo system, which boosts lows and highs to make the sound richer at low volumes. Dolphins, bats, and other creatures can hear much higher frequencies than we humans; bats locate their prey with sound waves at frequencies approaching 100 kHz. Medical ultrasound frequencies extend to tens of MHz.

Decibels

Figure 14.15 shows that the human ear responds to an extremely broad range of sound intensities, covering some 12 orders of magnitude; that's why Fig. 14.15 has a logarithmic scale. We therefore quantify sound levels using a logarithmic unit called the **decibel** (dB). The **sound intensity level** β in decibels is defined by

$$\beta = 10 \log\left(\frac{I}{I_0}\right) \tag{14.10}$$

where I is the intensity in W/m^2 and $I_0 = 10^{-12}$ W/m^2 is a reference level chosen as the approximate threshold of hearing at 1 kHz. Since the logarithm of 10 is 1, an increase of 10 dB corresponds to a factor-of-10 increase in the intensity I. Your ears, however, don't respond linearly, and for intensity levels above about 40 dB, you perceive a 10-dB increase as making the sound roughly twice as loud.

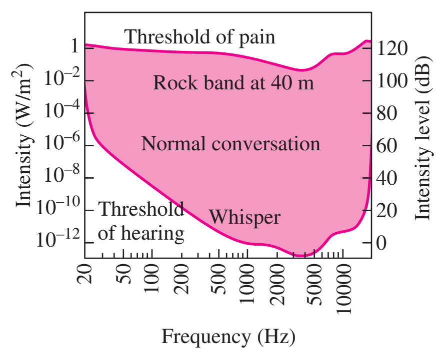


FIGURE 14.15 The human ear responds to sound whose intensity and frequency lie within the shaded region.

EXAMPLE 14.4 **Decibels: Turn Down the TV!**

Your sister is watching TV, the sound blasting at 75 dB. You yell to her to turn down the volume, and she lowers the intensity level to 60 dB. By what factor has the power dropped?

INTERPRET This problem is about the relation between power and sound intensity level as measured in decibels.

DEVELOP Equation 14.10, $\beta = 10 \log(I/I_0)$, relates the decibel level to the intensity, or power per unit area. At a fixed distance, the sound intensity is proportional to the power from the TV speaker, so in this example we can replace I by P in Equation 14.10.

EVALUATE Call the original 75-dB level β_1; then Equation 14.10 reads $\beta_1 = 10 \log(P_1/P_0) = 10 \log P_1 - 10 \log P_0$, where P_1 is the corresponding power and P_0 is the reference-level power. At the turned-down power P_2, the equation reads $\beta_2 = 10 \log P_2 - 10 \log P_0$. Subtracting our two equations gives

$$\beta_2 - \beta_1 = 10 \log P_2 - 10 \log P_1 = 10 \log\left(\frac{P_2}{P_1}\right)$$

Therefore, $\log(P_2/P_1) = (\beta_2 - \beta_1)/10 = (60 - 75)/10 = -1.5$. The answer we want is the ratio P_2/P_1, and because logarithms and exponentials are inverses, we have $P_2/P_1 = 10^{-1.5} = 0.032$.

(continued)

ASSESS Although we worked this problem using Equation 14.10, you can often do decibels in your head. Here the intensity level has dropped by 15 dB, corresponding to 1.5 orders of magnitude in actual intensity. So the intensity—and therefore the TV's power—has dropped by a factor of $10^{-1.5}$, or $1/(10\sqrt{10})$. Since $\sqrt{10}$ is about 3, that's about 1/30. Because you perceive each 10-dB change as a factor of about 2 in loudness, the reduced volume will sound somewhere between one-fourth and one-half as loud as before. ■

14.5 Interference

Figure 14.16 shows two wave trains approaching from opposite directions. Where they meet, experiment shows that the net displacement is the sum of the individual displacements. This is true for most waves, at least when the amplitude isn't too large. Waves whose displacements simply add are said to obey the **superposition principle**.

At the point shown in Fig. 14.16*b*, the wave crests coincide and so do the troughs. The resulting wave is, momentarily, twice as big. This is **constructive interference**—two waves superposing to produce a larger wave displacement. A little later, in Fig. 14.16*c*, the two waves cancel; this is **destructive interference**. Wave interference occurs throughout physics, from mechanical waves to light and even with the quantum-mechanical waves that describe matter at the atomic scale. Here we take a quick look at wave interference; we'll consider the interference of light waves in more detail in Chapter 32.

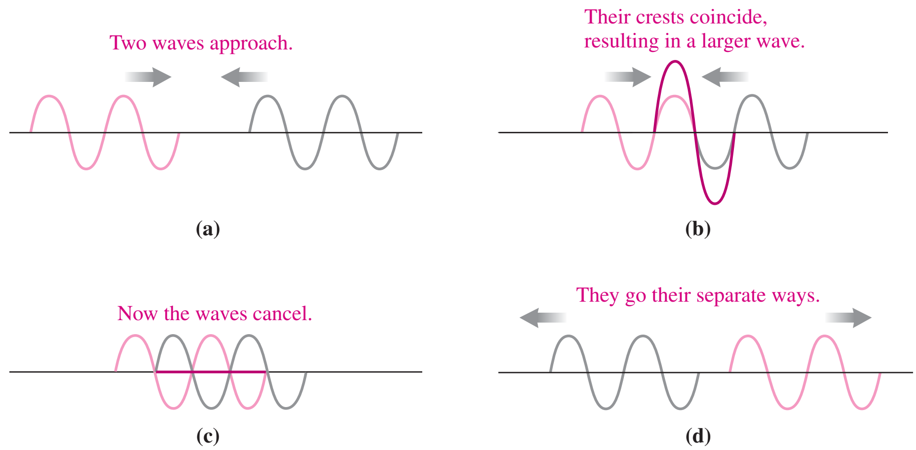


FIGURE 14.16 Wave superposition showing (b) constructive interference and (c) destructive interference.

Fourier Analysis

The superposition principle lets us build complex wave shapes by superposing simpler ones. The French mathematician Jean Baptiste Joseph Fourier (1768–1830) showed that *any* periodic wave can be written as a sum of simple harmonic waves, a process now known as **Fourier analysis**. Figure 14.17 shows a square wave—important, for example, as the "clock" signal that sets the speed of your computer—represented as a superposition of individual sine waves. Fourier analysis has applications ranging from music to structural engineering to communications, because it helps us understand how a complex wave behaves if we know how its harmonic components behave. The mix of Fourier components in the waveform from a musical instrument determines the exact sound we hear and accounts for the different sounds from different instruments even when they're playing the same note (Fig. 14.18).

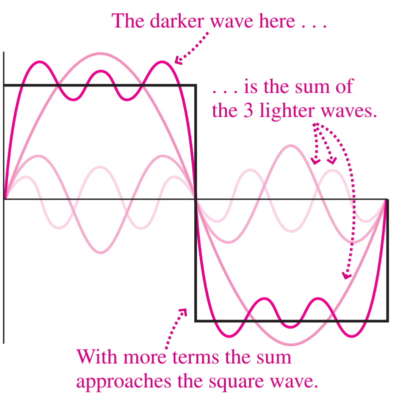
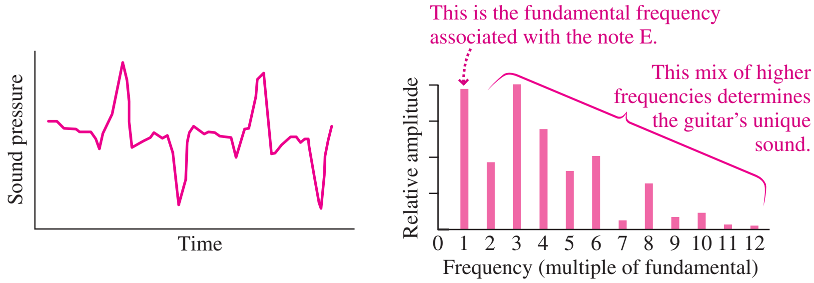


FIGURE 14.17 A square wave built up as a sum of simple harmonic waves. In this case the sum has the form $y(t) = A\sin(\omega t) + \frac{1}{3}A\sin(3\omega t) + \frac{1}{5}A\sin(5\omega t) + \cdots$. Only the first three terms are shown.

Dispersion

When wave speed is independent of wavelength, the simple harmonic components making up a complex waveform travel at the same speed. As a result, the waveform maintains its shape. But for some media, wave speed depends on wavelength. Then, individual harmonic



(a) **(b)**

FIGURE 14.18 (a) An electric guitar plays the note E, producing a complex waveform. (b) Fourier analysis shows the relative strengths of the individual sine waves whose sum produces the waveform.

waves travel at different speeds, and a complex waveform changes shape as it moves. This phenomenon is called **dispersion** and is illustrated in Fig. 14.19. Waves on the surface of deep water, for example, have speed given by

$$v = \sqrt{\frac{\lambda g}{2\pi}} \tag{14.11}$$

where λ is the wavelength and g the acceleration of gravity. Because v depends on λ, the waves are dispersive. Dispersion is also important in communications systems; for example, dispersion of the square wave pulses carrying digital data sets the maximum lengths for wires and optical fibers used in computer networks.

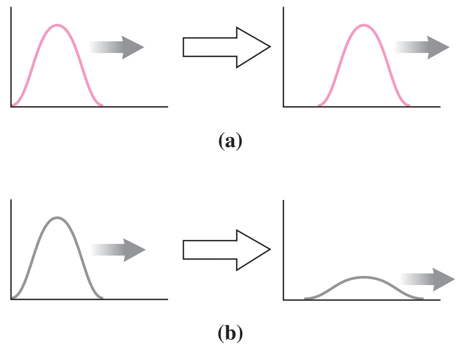


FIGURE 14.19 (a) A wave pulse in a nondispersive medium holds its shape as it propagates. (b) In a dispersive medium, the pulse shape changes.

CONCEPTUAL EXAMPLE 14.1 Storm Brewing!

It's a lovely, sunny day at the coast, but large waves, their crests far apart, are crashing on the beach. How do these waves tell of a storm at sea that may affect you later?

EVALUATE The phrase "crests far apart" is a clue: It says we're dealing with long-wavelength waves. Equation 14.11 shows that longer-wavelength waves on the ocean surface travel faster. Most ocean waves are generated by frictional forces between wind and water, so there must be strong winds somewhere out at sea. The longest wavelengths travel faster, so they reach shore well in advance of the storm.

ASSESS High-surf warnings often go up in advance of a storm, for the very reason elucidated in this example. Incidentally, wind isn't the only source of ocean waves; so are earthquakes. But the tsunamis they produce are shallow-water waves that don't obey Equation 14.11. You can explore tsunamis further in the Passage Problems.

MAKING THE CONNECTION A storm develops 600 km offshore and starts moving toward you at 40 km/h. Large waves with crests 250 m apart are your first hint of the storm. How long after you observe these waves will the storm hit?

EVALUATE At 40 km/h, it's going to take 15 hours for the storm to reach shore. Equation 14.11 gives 71 km/h for the wave speed when $\lambda = 250$ m. So the waves took 8.4 hours to reach shore. The storm is then 6.6 hours away.

Beats

When two waves of slightly different frequencies superpose, they interfere constructively at some points and destructively at others (Fig. 14.20*a*). Quantitatively, the combined wave is the sum of the two individual waves: $y(t) = A\cos\omega_1 t + A\cos\omega_2 t$. We can express this in a more enlightening form using the identity $\cos\alpha + \cos\beta = 2\cos\left[\frac{1}{2}(\alpha - \beta)\right]\cos\left[\frac{1}{2}(\alpha + \beta)\right]$ given in Appendix A. Then we have

$$y(t) = 2A\cos\left[\tfrac{1}{2}(\omega_1 - \omega_2)t\right]\cos\left[\tfrac{1}{2}(\omega_1 + \omega_2)t\right]$$

The second cosine factor represents a sinusoidal oscillation at the average of the two individual frequencies. The first term oscillates at a lower frequency—half the difference of the individual frequencies. If we think of the entire term $2A\cos\left[\frac{1}{2}(\omega_1 - \omega_2)t\right]$ as the "amplitude" of the higher-frequency oscillation, then this amplitude itself varies

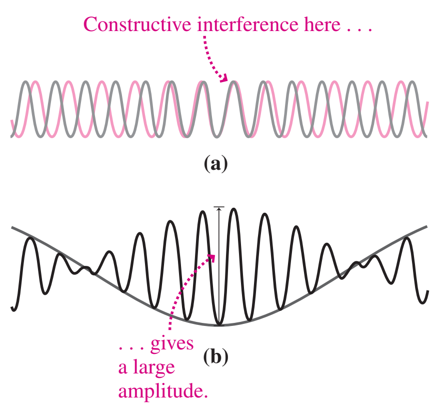


FIGURE 14.20 The origin of beats.

with time, as Fig. 14.20*b* shows. Note that there are *two* amplitude peaks for each cycle of the slow oscillation, so the frequency with which the amplitude varies is simply $\omega_1 - \omega_2$.

For sound waves, interference of two nearly equal frequencies produces intensity variations called **beats**; the closer the two frequencies, the longer the period between beats. Pilots, for example, synchronize airplane engines by reducing the beat frequency toward zero; musicians use the same trick to tune instruments. Beating of electromagnetic waves forms the basis for some very sensitive measurements.

Interference in Two Dimensions

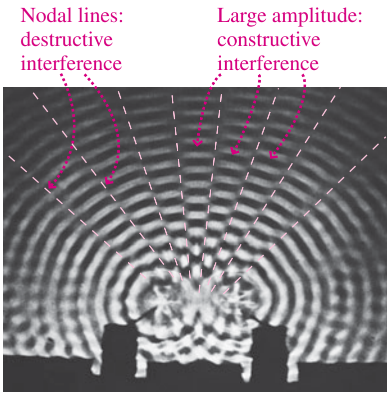


FIGURE 14.21 Water waves from two sources interfere to produce regions of low and high amplitude.

Waves propagating in two and three dimensions exhibit a rich variety of interference phenomena. Figure 14.21 shows one of the simplest and most important examples—the interference of waves from two point sources oscillating at the same frequency. Points on a perpendicular line midway between the sources are equidistant from both sources, and therefore waves arrive at this line in phase. Thus, they interfere constructively, producing a large amplitude. Some distance away, the waves arrive exactly half a period out of phase. They therefore interfere destructively, producing a **nodal line** where the wave amplitude is very small. Since waves travel half a wavelength in half a period, the nodal line occurs where the distances to the two sources differ by half a wavelength. Additional nodal lines occur where those distances differ by $1\frac{1}{2}$ wavelengths, $2\frac{1}{2}$ wavelengths, and so forth. In practice, two-source interference is observable only when the source separation is comparable to the wavelength. If it's much larger, then the regions of constructive and destructive interference are so close that they blur together.

Two-source interference also results when plane waves pass through two closely spaced apertures that act as sources of circular or spherical wavefronts. Such two-slit interference experiments are important in optics and modern physics, and are of historical interest because they were first used to demonstrate the wave nature of light.

EXAMPLE 14.5 Wave Interference in Two Dimensions: Calm Water

Ocean waves pass through two small openings, 20 m apart, in a breakwater. You're in a boat 75 m from the breakwater and initially midway between the openings, but the water is pretty rough. You row 33 m parallel to the breakwater and, for the first time, find yourself in relatively calm water. What's the wavelength of the waves?

INTERPRET This is a problem about wave interference. The water is rough at your initial location because constructive interference produces large-amplitude waves. You find calm water at the first nodal line, where destructive interference reduces the wave amplitude.

DEVELOP We sketched the situation in Fig. 14.22. We've seen that the first nodal line occurs when the path lengths from two sources differ by half a wavelength. So our plan is to calculate the wavelength by applying this fact to the distances *AP* and *BP*.

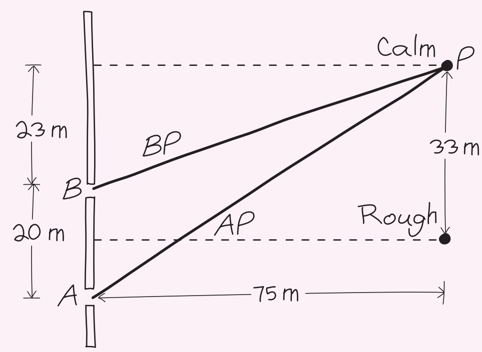


FIGURE 14.22 Calm water at *P* implies that paths *AP* and *BP* differ by half a wavelength.

EVALUATE Applying the Pythagorean theorem gives

$$AP = \sqrt{(75\text{ m})^2 + (43\text{ m})^2} = 86.5\text{ m}$$

$$BP = \sqrt{(75\text{ m})^2 + (23\text{ m})^2} = 78.4\text{ m}$$

The wavelength is twice the difference between these lengths, so

$$\lambda = 2(AP - BP) = 2(86.5\text{ m} - 78.4\text{ m}) = 16\text{ m}$$

ASSESS We expect two-source interference to be obvious when the source spacing is comparable to the wavelength. Here the 20-m spacing is indeed comparable to the 16-m wavelength, so our answer makes sense. ■

GOT IT? 14.4 Light shines through two small holes into a dark room, and a screen is mounted opposite the holes. The hole spacing is comparable to the wavelength of the light. Looking at the screen, will you see (a) two bright spots opposite the two holes or (b) a pattern of light and dark patches? Explain.

14.6 Reflection and Refraction

You shout in a mountain valley and hear echoes. You look in a mirror and see your reflection. A metal screen reflects microwaves to keep them in your oven. A physician's ultrasound probes your body, reflecting off internal structures. A bat uses reflected sound to home in on its prey. All these are examples of wave **reflection**.

You can see that wave reflection *must* occur when a wave hits a medium in which it can't propagate; otherwise, where would the wave energy go? Figures 14.23 and 14.24 detail the reflection process for waves on a stretched string, in the two cases where the string end is clamped at a rigid wall and free to move up and down. In the first case, the wave amplitude must remain zero at the end, so the incident and reflected pulses interfere destructively and the reflected wave is therefore inverted. In the second case, the displacement is a maximum at the free end, and the reflected wave is not inverted.

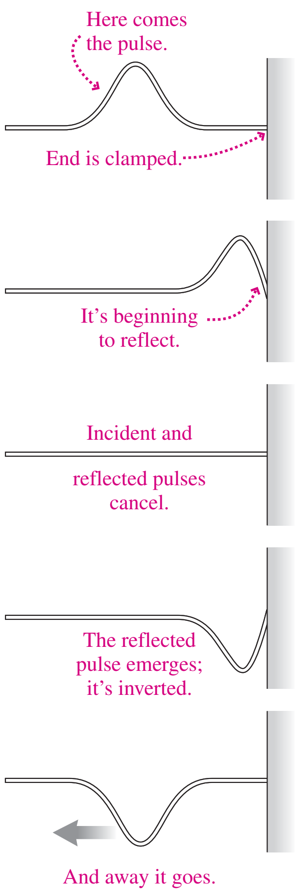


FIGURE 14.23 Reflection of a wave pulse at the rigidly clamped end of string.

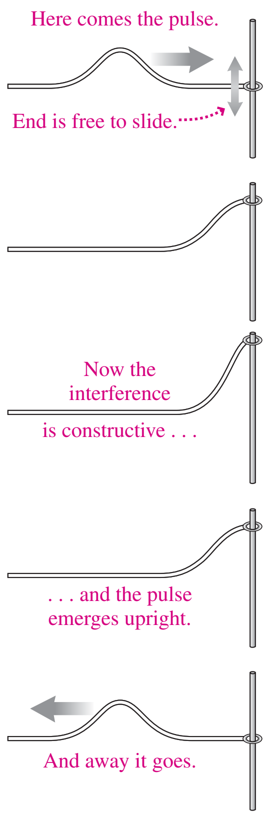


FIGURE 14.24 Reflection of a wave pulse at a free end.

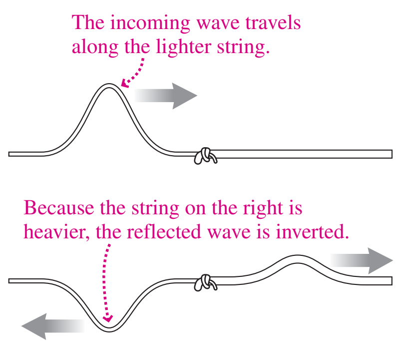


FIGURE 14.25 Partial reflection occurs at the junction between two strings.

Between the extremes of a rigid wall and a perfectly free end lies the case of one string connected to another with different mass per unit length. In this case, some wave energy is transmitted to the second string and some is reflected back along the first (Fig. 14.25).

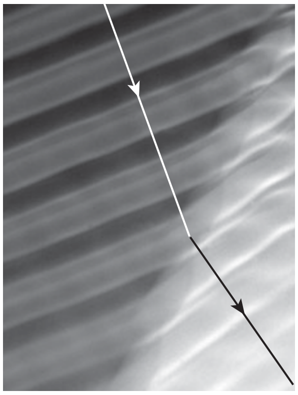


FIGURE 14.26 Waves in shallow water refract at the interface between two different water depths.

The phenomenon of partial reflection and transmission at a junction of strings has its analog in the behavior of all sorts of waves at interfaces between different media. For example, shallow-water waves are partially reflected if the water depth changes abruptly. Light incident on even the clearest glass undergoes partial reflection because of the difference in the light-transmitting capabilities of air and glass. Partial reflection of ultrasound waves at the interfaces of body tissues with different densities makes ultrasound a valuable medical diagnostic.

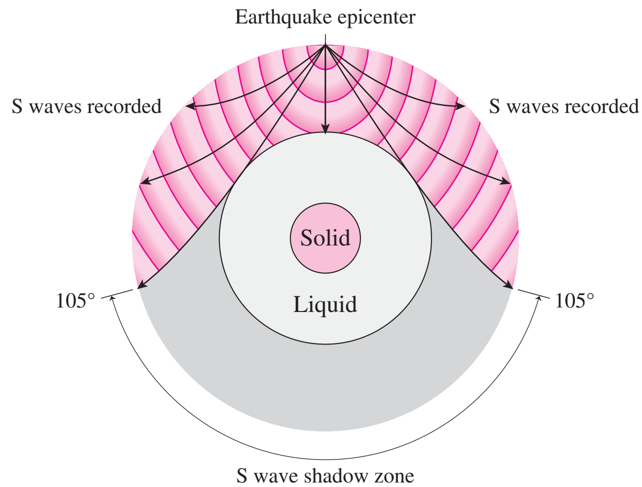
When waves strike an interface between two media at an oblique angle and are capable of propagating in the second medium, the phenomenon of **refraction** occurs. In refraction, the direction of wave propagation changes because of a difference in wave speed between the two media (Fig. 14.26). We'll discuss the mathematics of refraction in Chapter 30.

APPLICATION **Probing the Earth**

Waves propagating and reflecting inside the Earth help geologists deduce the planet's interior structure. That's because Earth's interior supports two types of waves. Longitudinal waves, also called P waves, propagate in both solids and liquids. Transverse, or S waves, propagate only in solids. Earthquakes generate S waves that propagate throughout the solid Earth. But as the figure suggests, they can't get through the liquid outer core, so they leave a "shadow" where seismographs don't record any S-wave activity. This effect is our clearest evidence that Earth has a liquid core.

P waves, however, do propagate through the liquid core. But they undergo partial reflections farther in—evidence for an abrupt change in core density. Careful analysis shows that wave speeds in the inner core are consistent with its being solid—giving our planet the solid–liquid–solid structure suggested in the figure.

Studies of Earth's large-scale structure generally use earthquake waves, although inner-core evidence also comes from underground nuclear explosions. At a smaller scale, explosive charges or machines that "thump" the ground produce waves whose reflections from rock layers down to a few kilometers depth help reveal oil and gas deposits.



14.7 Standing Waves

Imagine a string clamped tightly at both ends. Waves propagate back and forth by reflecting at the ends. But because the ends are clamped, the wave displacement at each end must always be zero. Only certain waves can satisfy this requirement; as Fig. 14.27 suggests, they're waves for which an integer number of half-wavelengths just fits the string's length L.

The waves in Fig. 14.27 are **standing waves**, so called because they essentially stand still, confined to the length of the string. At each point the string executes simple harmonic motion perpendicular to its undisturbed state. We can describe standing waves mathematically as arising from the superposition of two waves propagating in opposite directions and reflecting at the ends of the string. If we take the x-axis to coincide with the string, then we can write the string displacements in two such waves as $y_1(x, t) = A\cos(kx - \omega t)$ for the wave propagating in the $+x$-direction (recall Equation 14.3) and $y_2(x, t) = -A\cos(kx + \omega t)$ for the wave propagating in the $-x$-direction. (The minus sign in y_2 accounts for the phase change that occurs on reflection.) Their superposition is then

$$y(x, t) = y_1 + y_2 = A[\cos(kx - \omega t) - \cos(kx + \omega t)]$$

Appendix A lists a trig identity for the difference of two cosines:

$$\cos\alpha - \cos\beta = -2\sin\left[\tfrac{1}{2}(\alpha + \beta)\right]\sin\left[\tfrac{1}{2}(\alpha - \beta)\right]$$

Applying this identity with $\alpha = kx - \omega t$ and $\beta = kx + \omega t$ gives

$$y(x, t) = 2A \sin kx \sin \omega t \tag{14.12}$$

Equation 14.12 is the mathematical description of a standing wave, and it affirms our qualitative description that each point on the string simply oscillates up and down. Pick any point—that is, any fixed value of x—and Equation 14.12 does indeed describe simple harmonic motion in the y-direction, through the factor $\sin \omega t$. The amplitude of that motion depends on the point x you've chosen, and is given by the factor that multiplies $\sin \omega t$—namely, $2A \sin kx$.

Because the string is clamped at both ends, the amplitude at the ends must be zero. Our amplitude factor $2A \sin kx$ does give $y = 0$ in Equation 14.12 at $x = 0$, but what about at $x = L$? Here we'll get zero only if $\sin kL = 0$—and that requires kL to be a multiple of π. So we must have $kL = m\pi$, where m is any integer. But the wave number k is related to the wavelength λ by $k = 2\pi/\lambda$. Our condition $kL = m\pi$ can then be written

$$L = \frac{m\lambda}{2}, \quad m = 1, 2, 3, \ldots \tag{14.13}$$

This is just the condition we already guessed from Fig. 14.27—namely, that the string length L be an integer number of half-wavelengths.

Given a particular string length L, Equation 14.13 limits the allowed standing waves on the string to a discrete set of wavelengths. Those allowed waves are called **modes** or **harmonics**, and the integer m is the **mode number**. The $m = 1$ mode is the **fundamental** and is the longest-wavelength standing wave that can exist on the string. The higher modes are **overtones**.

Figure 14.27 shows that there are points where the string doesn't move at all. These are called **nodes**. Points where the amplitude of the wave displacement is a maximum, in contrast, are **antinodes**.

When a string is clamped rigidly at one end but is free at the other, its clamped end is a node but its free end is an antinode. Figure 14.28 shows that the string length must then be an odd multiple of a quarter-wavelength—a result that you can also get from Equation 14.12 by requiring $\sin kL = 1$ to give maximum amplitude at $x = L$.

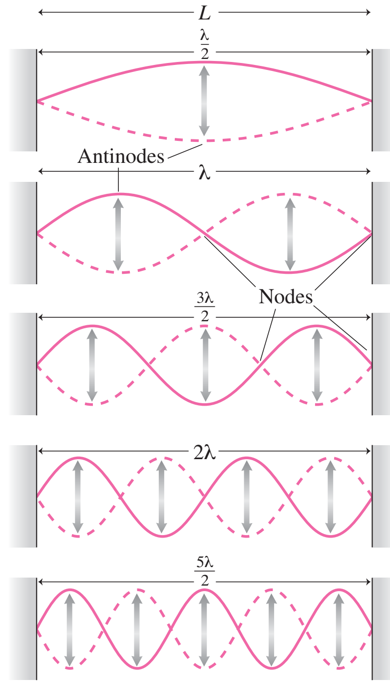


FIGURE 14.27 Standing waves on a string clamped at both ends; shown are the fundamental and four overtones.

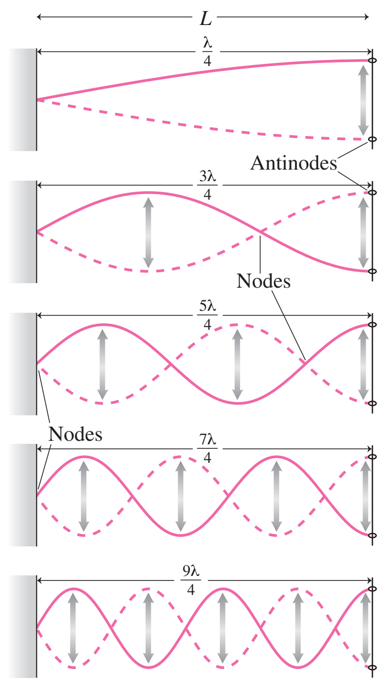


FIGURE 14.28 When one end of the string is fixed and the other free, the string can accommodate only an odd number of quarter-wavelengths.

Standing-Wave Resonance

We've discussed standing waves in terms of constraints on the wavelength λ rather than on the frequency f. But because waves on a string have a fixed speed v, and because $f\lambda = v$, Equation 14.13's discrete set of allowed wavelengths corresponds to a set of discrete frequencies. The lowest allowed frequency, the fundamental, corresponds to the longest wavelength; the overtones have higher frequencies.

Because a stretched string can oscillate in any of its allowed frequencies, the resonant behavior that we discussed in Chapter 13 can occur close to any of those frequencies. Buildings and other structures, in analogy with our simple string, support a variety of standing-wave modes. For example, a skyscraper is like the string of Fig. 14.28, with its base clamped to Earth but its top free to swing. Engineers must be sure to identify all possible modes of structures they design in order to avoid harmful resonances. The disastrous oscillations of the Tacoma Narrows Bridge shown in Fig. 13.26 are actually torsional standing waves.

Other Standing Waves

Standing waves are common phenomena. Water waves in confined spaces exhibit standing waves, and entire lakes can develop very slow oscillations corresponding to low-mode-number standing waves. Standing electromagnetic waves occur inside closed metal cavities; in microwave ovens the nodes of the standing-wave pattern would result in "cold" spots were not either the food or the source of microwaves kept in motion. Standing sound waves in the Sun help astrophysicists probe the solar interior. And even atomic structure can be understood in terms of standing waves associated with electrons.

FIGURE 14.29 Standing waves on a violin, imaged using holographic interference of laser light waves.

Musical Instruments

Our analysis of standing waves on strings applies directly to stringed musical instruments such as violins, guitars, and pianos. Standing-wave vibrations in the instrument strings are communicated to the air as sound waves, usually through the intermediary of a sounding box or electronic amplifiers. For instruments in the violin family, the body of the instrument itself undergoes standing-wave vibrations, excited by the vibration of the string, that establish each individual instrument's peculiar sound quality (Fig. 14.29). Similarly, the stretched membranes of drums exhibit a variety of standing-wave patterns representing the allowed modes on these two-dimensional surfaces.

Wind instruments generate standing sound waves in air columns, as suggested in Fig. 14.30. These must be open at one end to allow sound to escape; in many instruments the column is effectively open at both ends. An open end has its pressure fixed at atmospheric pressure; it is therefore a pressure node and thus, from Fig. 14.14, a displacement antinode. As a result, an instrument open at one end supports odd-integer multiples of a quarter-wavelength (Fig. 14.30*a*), in analogy with Fig. 14.28. An instrument open at both ends, on the other hand, supports integer multiples of a half-wavelength (Fig. 14.30*b*).

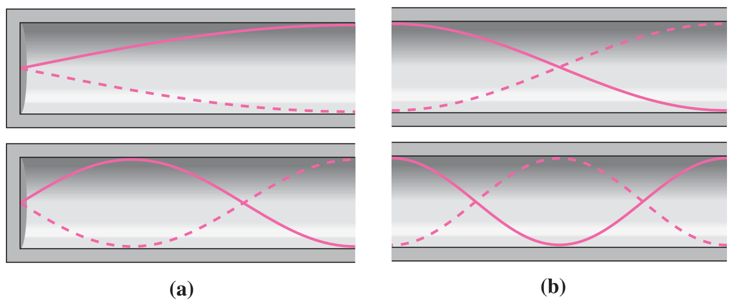


FIGURE 14.30 Standing waves in wind instruments: (a) open at one end and (b) open at both ends.

EXAMPLE 14.6 Standing-Wave Modes: The Double Bassoon

The double bassoon is the lowest-pitched instrument in a normal orchestra. The instrument is "folded" to achieve an effective air column 5.5 m long, and it acts like a pipe open at both ends. What's the frequency of the double bassoon's fundamental note? Assume the sound speed is 343 m/s.

INTERPRET This is a problem about standing-wave modes in a hollow pipe open at both ends.

DEVELOP Figure 14.30*b* applies to a pipe that's open at both ends. So our sketch of the fundamental mode in Fig. 14.31 looks like the upper of the two pictures in Fig. 14.30*b*. We can find the wavelength and then use Equation 14.1, $v = \lambda f$, to get the frequency.

EVALUATE The wavelength is twice the instrument's 5.5-m length, or 11 m. Then Equation 14.1 gives

$$f = \frac{v}{\lambda} = \frac{343\ \text{m/s}}{11\ \text{m}} = 31\ \text{Hz}$$

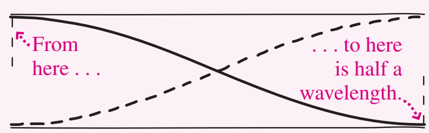


FIGURE 14.31 Sketch for Example 14.6.

ASSESS This frequency is the note B_0, which lies near the low-frequency limit of the human ear. Like most wind instruments, the bassoon has a number of holes that, when uncovered, alter the positions of the antinodes and therefore change the pitch. ■

GOT IT? 14.5 A string 1 m long is clamped tightly at one end and is free to slide up and down at the other. Which of the following are possible wavelengths for standing waves on this string: $\frac{4}{5}$ m, 1 m, $\frac{4}{3}$ m, $\frac{3}{2}$ m, 2 m, 3 m, 4 m, 5 m, 6 m, 7 m, 8 m?

14.8 The Doppler Effect and Shock Waves

The speed v of a wave is its speed relative to the medium through which it propagates. A point source at rest in the medium radiates waves uniformly in all directions (Fig. 14.32). But when the source moves, wave crests bunch up in the direction toward which the source is moving, resulting in a decreased wavelength (Fig. 14.33). In the opposite direction, wave crests spread out and the wavelength increases.

The wave speed is determined by the properties of the medium, so it doesn't change with source motion. Thus the equation $v = \lambda f$ still holds. This means that an observer in front of the moving source, where λ is smaller, experiences a higher wave frequency as more wave crests pass per unit time. Similarly, an observer behind the source experiences a lower frequency. This change in wavelength and frequency from a moving source is the **Doppler effect** or **Doppler shift**, after the Austrian physicist Christian Johann Doppler (1803–1853).

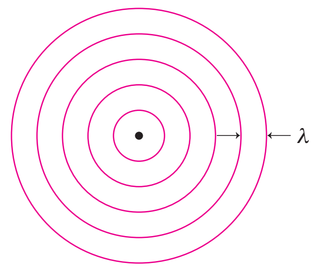


FIGURE 14.32 Circular waves from a source at rest with respect to the medium.

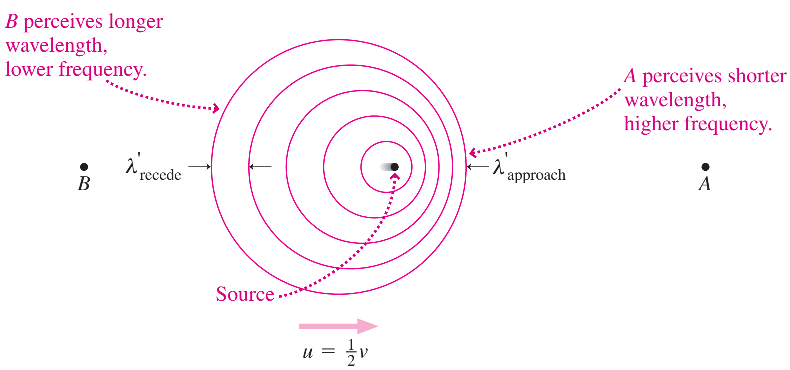


FIGURE 14.33 Origin of the Doppler effect, shown for a source moving with half the wave speed.

To analyze the Doppler effect, let λ be the wavelength measured when the source is stationary, and λ' the wavelength when the source is moving at speed u through a medium where the wave speed is v. At the source, the time between wave crests is the wave period T, and a wave crest moves one wavelength λ in this time. But during the same time T, the moving source covers a distance uT, after which it emits the next wave crest. So the distance between wave crests, as seen by an observer in front of the moving source, is $\lambda' = \lambda - uT$. Writing $T = \lambda/v$, we get

$$\lambda' = \lambda - u\frac{\lambda}{v} = \lambda\left(1 - \frac{u}{v}\right) \qquad \text{(source approaching)} \qquad (14.14a)$$

The situation is similar in the direction opposite the source motion, except that now the wavelength *increases* by the amount $\lambda u/v$, giving

$$\lambda' = \lambda\left(1 + \frac{u}{v}\right) \qquad \text{(source receding)} \qquad (14.14b)$$

We can recast these expressions in terms of frequency using the relations $\lambda = v/f$ and $\lambda' = v/f'$, where f' is the frequency of waves from the moving source as measured by an observer at rest in the medium. Substituting these relations in our expressions for λ' and then solving for f' gives

$$f' = \frac{f}{1 \pm u/v} \quad \text{(Doppler shift, moving source)} \tag{14.15}$$

for the Doppler-shifted frequency, where the + and − signs correspond to receding and approaching sources, respectively.

You've probably experienced the Doppler effect for sound when standing near a highway. A loud truck approaches with a high-pitched sound "aaaaaaaaaaa." As it passes, the pitch drops abruptly: "aaaaaaaaaeiooooooooooo," and stays low as the truck recedes. Practical uses of the Doppler effect are numerous. The Doppler shift in reflected ultrasound measures blood flow and fetal heartbeat. Police radar uses the Doppler shift of high-frequency radio waves reflected from moving cars. The Doppler shift of starlight reveals stellar motions, and Doppler-shifted light from distant galaxies is evidence that our entire universe is expanding.

EXAMPLE 14.7 Doppler Effect: The Wrong Note

A car speeds down the highway with its stereo blasting. An observer with perfect pitch is standing by the roadside and, as the car approaches, notices that a musical note that should be G $(f = 392\text{ Hz})$ sounds like A (440 Hz). How fast is the car moving?

INTERPRET This problem is about the Doppler effect in sound from a moving source.

DEVELOP Equation 14.15, $f' = f/(1 \pm u/v)$, relates the original and shifted frequencies to the source speed u, so our plan is to solve this equation for u. We'll use the minus sign because the source is approaching. We'll also need the sound speed v, which Example 14.6 gave as 343 m/s.

EVALUATE Solving Equation 14.15 for u gives

$$u = v\left(1 - \frac{f}{f'}\right) = (343\text{ m/s})\left(1 - \frac{392\text{ Hz}}{440\text{ Hz}}\right) = 37.4\text{ m/s}$$

ASSESS Our answer—some 134 km/h or 84 mi/h—seems reasonable for a speeding car, though not a particularly safe speed! And it's a little more than 10% of the sound speed, consistent with the roughly 10% change in the sound frequency. ■

Moving Observers

A Doppler shift in frequency, but not wavelength, also occurs when a moving observer approaches a stationary source—meaning a source at rest with respect to the wave medium. An observer moving toward a stationary source passes wave crests more often than would happen if the observer were at rest, and thus measures a shorter wave period and therefore a higher frequency. The result, as you can show in Problem 80, is a shifted frequency given by

$$f' = f\left(1 \pm \frac{u}{v}\right) \quad \text{(Doppler shift, moving observer)} \tag{14.16}$$

with the positive sign for an observer approaching the source and the negative sign for an observer receding. For observer velocities u small compared with the wave speed v, Equations 14.15 and 14.16 give essentially the same results.

Waves from a stationary source that reflect from a moving object undergo a Doppler shift *twice*. First, because the frequency as received at the reflecting object is shifted, according to Equation 14.16, due to the object's motion relative to the source. Then a stationary observer sees the reflected waves as coming from a moving source, so there's another shift, this time given by Equation 14.15. Police radar and other Doppler-based speed measurements make use of this double Doppler shift that occurs on reflection.

The Doppler Effect for Light

Although light and other electromagnetic waves do not require a material medium, they, too, are subject to the Doppler shift. Both Doppler formulas we derived here apply to electromagnetic waves, but only as approximations when the relative speed between source and observer is much lower than the speed of light.

The Doppler shift for electromagnetic waves is the same whether it's the source that moves or the observer. This reflects a profound fact at the root of Einstein's relativity: that "stationary" and "moving" are meaningful only as relative terms. Electromagnetic waves, unlike mechanical waves, do not require a medium—and therefore terms such as "stationary source" and "moving observer" are meaningless. All that matters is the relative motion between source and observer. We'll explore this point further in Chapter 33.

Shock Waves

Equation 14.14a suggests that wavelength goes to zero if a source approaches at exactly the wave speed. This happens because wave crests can't get away from the source, so they pile up just ahead of it to form a large-amplitude wave called a **shock wave** (Fig. 14.34). When the source moves faster than the wave speed, waves pile up on a cone whose half-angle is given by $\sin\theta = v/u$, as shown. The ratio u/v is called the **Mach number,** and the cone angle is the **Mach angle**.

Shock waves occur in a wide variety of physical situations. Sonic booms are shock waves from supersonic aircraft. The bow wave of a boat is a shock wave on the water surface. On a much larger scale, a huge shock wave forms in space as the solar wind—a high-speed flow of particles from the Sun—encounters Earth's magnetic field.

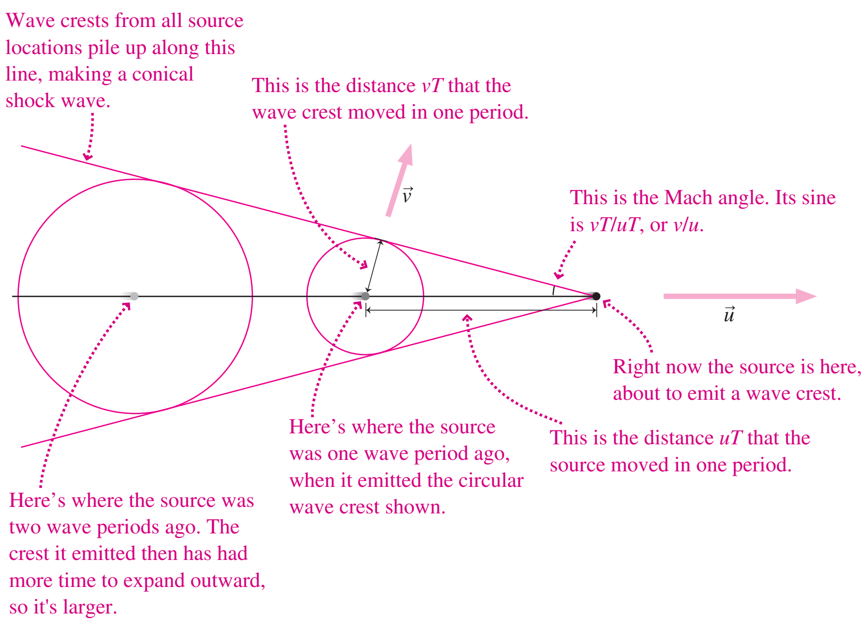
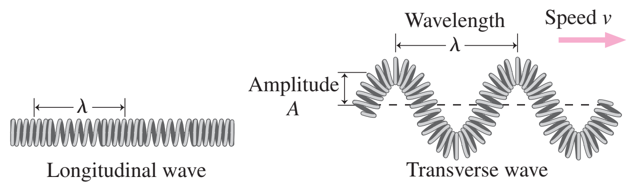


FIGURE 14.34 Shock waves form when the source speed u exceeds the wave speed v.

CHAPTER 14 SUMMARY

Big Picture

Waves are the big idea here. A wave is a propagating disturbance that carries energy but not matter. Waves are characterized by their amplitude, wavelength, and speed. They can be **longitudinal** or **transverse**.



Key Concepts and Equations

Wave **period** is the time for one complete wave cycle. Period and frequency are inverses, and wavelength λ, period T or frequency f, and wave speed v are all related:

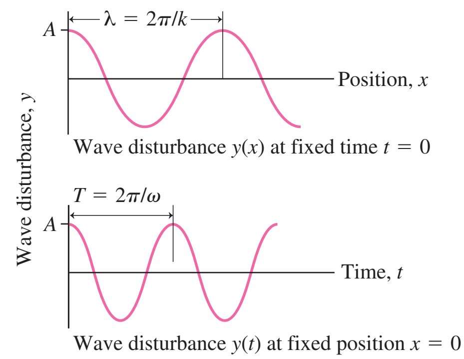
$$v = \frac{\lambda}{T} = \lambda f$$

A **simple harmonic wave** is sinusoidal in shape. The wave disturbance is a function of position and time and is most simply described in terms of its **wave number** k and **angular frequency** ω:

$$y(x, t) = A\cos(kx - \omega t)$$

They're related to wavelength and period by

$$k = \frac{2\pi}{\lambda} \qquad \text{and} \qquad \omega = \frac{2\pi}{T}$$



Wave **intensity** is the power per unit area carried by the wave: $I = P/A$. For a spherical wave that spreads in all directions from a localized source, intensity decreases as the inverse square of the distance from the source: $I = P/(4\pi r^2)$.

Applications

Wave speed is a characteristic of the medium.

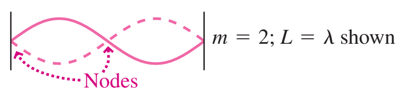
Transverse waves on strings: $v = \sqrt{\dfrac{F}{\mu}}$

Longitudinal sound waves in a gas: $v = \sqrt{\dfrac{\gamma P}{\rho}}$, about 343 m/s in air under standard conditions

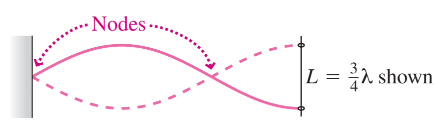
Surface waves in deep water: $v = \sqrt{\dfrac{\lambda g}{2\pi}}$

Standing waves on strings

Clamped at both ends, string length is an integer multiple of a half-wavelength: $L = m\lambda/2$



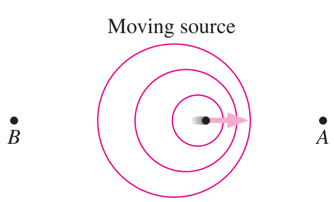
Clamped at one end, string length is an odd-integer multiple of a quarter-wavelength:



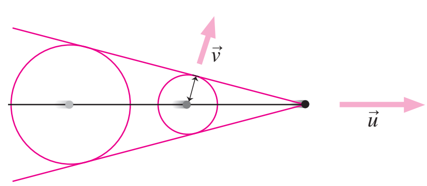
The **Doppler effect** is a frequency and/or wavelength shift due to the motion u of an observer or source relative to the medium with wave speed v.

Moving source: $f' = \dfrac{f}{(1 \pm u/v)}$, + for receding, − for approaching; λ also changes

Moving observer: $f' = f(1 \pm u/v)$, + for approaching, − for receding; no change in λ



Shock waves occur when a wave source (speed u) moves through a medium at greater than the wave speed (v).



For homework assigned on MasteringPhysics, go to www.masteringphysics.com | BIO *Biology and/or medical-related problems* C *Computer problems*

For Thought and Discussion

1. What distinguishes a wave from an oscillation?
2. Red light has a longer wavelength than blue light. Compare their frequencies.
3. Consider a light wave and a sound wave with the same wavelength. Which has the higher frequency?
4. In what sense is "the wave" passing through the crowd at a football game really a wave?
5. Must a wave be either transverse or longitudinal? Explain.
6. As a wave propagates on a stretched string, the string moves back and forth sideways. Is the string speed related to the wave speed? Explain.
7. If you doubled the tension in a string, what would happen to the speed of waves on the string?
8. A heavy cable is hanging vertically, its bottom end free. How will the speed of transverse waves near the top and bottom of the cable compare? Why?
9. The intensity of light from a localized source decreases as the inverse square of the distance from the source. Does this mean that the light loses energy as it propagates?
10. BIO Medical ultrasound uses frequencies on the order of 10^7 Hz, far above the range of the human ear. In what sense are these waves "sound"?
11. If you double the pressure of a gas while keeping its density the same, what happens to the sound speed?
12. Water is about a thousand times more dense than air, yet the speed of sound in water is greater than in air. How is this possible?
13. If you place a perfectly clear piece of glass in perfectly clear water, you can still see the glass. Why?
14. When a wave source moves relative to the medium, a stationary observer measures changes in both wavelength and frequency. But when the observer moves and the source is stationary, only the frequency changes. Why the difference?
15. Why can a boat easily produce a shock wave on the water surface, while only a very high-speed aircraft can produce a sonic boom?

Exercises and Problems

Exercises

Section 14.1 Waves and Their Properties

16. Ocean waves with 18-m wavelength travel at 5.3 m/s. What's the time interval between wave crests passing a boat moored at a fixed location?
17. Ripples in a shallow puddle propagate at 34 cm/s. If the wave frequency is 5.2 Hz, find (a) the period and (b) the wavelength.
18. An 88.7-MHz FM radio wave propagates at the speed of light. What's its wavelength?
19. Calculate the wavelengths of (a) a 1.0-MHz AM radio wave, (b) a channel 9 TV signal (190 MHz), (c) a police radar (10 GHz), (d) infrared radiation from a hot stove $(4\times10^{13}\text{ Hz})$, (e) green light $(6.0\times10^{14}\text{ Hz})$, and (f) 1.0×10^{18}-Hz X rays. All are electromagnetic waves that propagate at 3.0×10^{8} m/s.
20. A seismograph located 1200 km from an earthquake detects seismic waves 5.0 min after the quake occurs. The seismograph oscillates in step with the waves, at 3.1 Hz. Find the wavelength.
21. BIO Medical ultrasound waves travel at about 1500 m/s in soft tissue. Higher frequencies provide clearer images but don't penetrate to deeper organs. Find the wavelengths of (a) 8.0-MHz ultrasound used in fetal imaging and (b) 3.5-MHz ultrasound used to image an adult's kidneys.

Section 14.2 Wave Math

22. An ocean wave has period 4.1 s and wavelength 10.8 m. Find its (a) wave number and (b) angular frequency.
23. Find the (a) amplitude, (b) wavelength, (c) period, and (d) speed of a wave whose displacement is given by $y = 1.3\cos(0.69x + 0.31t)$, where x and y are in centimeters and t in seconds. (e) In which direction is the wave propagating?
24. BIO Ultrasound used in a medical imager has frequency 4.8 MHz and wavelength 0.31 mm. Find (a) the angular frequency, (b) the wave number, and (c) the wave speed.
25. A simple harmonic wave of wavelength 16 cm and amplitude 2.5 cm is propagating along a string in the negative x-direction at 35 cm/s. Find its (a) angular frequency and (b) wave number. (c) Write a mathematical expression describing the displacement y of this wave (in centimeters) as a function of position and time. Assume the displacement at $x = 0$ is a maximum when $t = 0$.
26. Analysis of waves in shallow water (depth much less than wavelength) yields the following wave equation:

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{gh}\frac{\partial^2 y}{\partial t^2}$$

where h is the water depth and g the gravitational acceleration. Give an expression for the wave speed.

Section 14.3 Waves on a String

27. The main cables supporting New York's George Washington Bridge have a mass per unit length of 4100 kg/m and are under 250-MN tension. At what speed would a transverse wave propagate on these cables?
28. A transverse wave 1.2 cm in amplitude propagates on a string; its frequency is 44 Hz. The string is under 21-N tension and has mass per unit length 15 g/m. Determine its speed.
29. A transverse wave with 3.0-cm amplitude and 75-cm wavelength propagates at 6.7 m/s on a stretched spring with mass per unit length 170 g/m. Find the spring tension.
30. A rope is stretched between supports 12 m apart; its tension is 35 N. If one end of the rope is tweaked, the resulting disturbance reaches the other end 0.45 s later. Find the total mass of the rope.
31. A rope with 280 g of mass per meter is under 550-N tension. Find the average power carried by a wave with frequency 3.3 Hz and amplitude 6.1 cm propagating on the rope.

Section 14.4 Sound Waves

32. Show that the quantity $\sqrt{P/\rho}$ from Equation 14.9 has the units of speed.
33. Find the sound speed in air under standard conditions with pressure 101 kN/m^2 and density 1.20 kg/m^3.
34. Timers in sprint races start their watches when they see smoke from the starting gun, not when they hear the sound. Why? How much error would be introduced by timing a 100-m race from the sound of the gun?
35. The factor γ for nitrogen dioxide (NO_2) is 1.29. Find the sound speed in NO_2 at 4.8×10^{4}-N/m^2 pressure and 0.35-kg/m^3 density.
36. A gas with density 1.0 kg/m^3 and pressure $8.0\times10^{4}\text{ N/m}^2$ has sound speed 365 m/s. Are the gas molecules monatomic or diatomic?
37. Divers in an underwater habitat breathe a special mixture of oxygen and neon to prevent the possibly fatal effects of nitrogen in ordinary air. With pressure $6.2\times10^{5}\text{ N/m}^2$ and density 4.5 kg/m^3, the effective γ value for the mixture is 1.61. Find the frequency in this mixture for a 50-cm-wavelength sound wave, and compare with its frequency in air under normal conditions.

Section 14.5 Interference

38. You're in an airplane whose two engines are running at 560 rpm and 570 rpm. How often do you hear a peak in the sound intensity?
39. What's the wavelength of the ocean waves in Example 14.5 if the calm water you encounter at 33 m is the *second* calm region on your voyage from the center line?

Section 14.7 Standing Waves

40. A 2.0-m-long string is clamped at both ends. (a) Find the longest-wavelength standing wave possible on this string. (b) If the wave speed is 56 m/s, what's the lowest standing-wave frequency?
41. When a stretched string is clamped at both ends, its fundamental frequency is 140 Hz. (a) What's the next higher frequency? If the same string, with the same tension, is now clamped at one end and free at the other, what are (b) the fundamental and (c) the next higher frequency?
42. A string is clamped at both ends and tensioned until its fundamental frequency is 85 Hz. If the string is then held rigidly at its midpoint, what's the lowest frequency at which it will vibrate?
43. A crude model of the human vocal tract treats it as a pipe closed at one end. Find the effective length of the vocal tract in a person whose fundamental tone is 620 Hz. Sound speed in air at body temperature is 354 m/s.

Section 14.8 The Doppler Effect and Shock Waves

44. A car horn emits 380-Hz sound. If the car moves at 17 m/s with its horn blasting, what frequency will a person standing in front of the car hear?
45. The stationary siren on a firehouse is blaring at 85 Hz. What's the frequency perceived by a firefighter racing toward the station at 120 km/h?
46. A fire truck's siren at rest wails at 1400 Hz; standing by the roadside as the truck approaches, you hear it at 1600 Hz. How fast is the truck going?
47. Red light emitted by hydrogen atoms at rest in the laboratory has wavelength 656 nm. Light emitted in the same process on a distant galaxy is received at Earth with wavelength 708 nm. Describe the galaxy's motion relative to Earth.

Problems

48. Figure 14.35 shows a simple harmonic wave at time $t = 0$ and later at $t = 2.6$ s. Write a mathematical description of this wave.

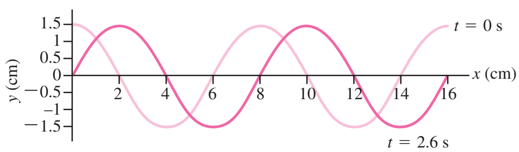


FIGURE 14.35 Problem 48

49. Transverse waves propagate at 18 m/s on a string under 14-N tension. What will be the wave speed if the tension is increased to 40 N?
50. A uniform cable hangs vertically under its own weight. Show that the speed of waves on the cable is given by $v = \sqrt{yg}$, where y is the distance from the bottom of the cable.
51. Figure 14.36 shows a wave train consisting of two sine wave cycles propagating along a string. Obtain an expression for the total energy in this wave train, in terms of the string tension F, the wave amplitude A, and the wavelength λ.

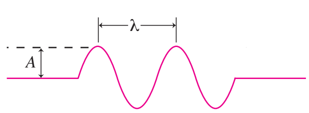


FIGURE 14.36 Problem 51

52. A loudspeaker emits energy at the rate of 50 W, spread in all directions. Find the intensity of sound 18 m from the speaker.
53. Light intensity 3.3 m from a lightbulb is 0.73 W/m^2. Find the bulb's power output, assuming it radiates equally in all directions.
54. Light emerges from a 5.0-mW laser in a beam 1.0 mm in diameter. The beam shines on a wall, producing a spot 3.6 cm in diameter. What is the beam's intensity (a) at the laser and (b) at the wall?
55. Two waves have the same angular frequency ω, wave number k and amplitude A, but they differ in phase: $y_1 = A\cos(kx - \omega t)$ and $y_2 = A\cos(kx - \omega t + \phi)$. Show that their superposition is also a simple harmonic wave, and determine its amplitude as a function of the phase difference ϕ.
56. A wave on a wire under 28-N tension is described by $y = 1.5\sin(0.10x - 560t)$, where x and y are in centimeters and t is in seconds. Find (a) the amplitude, (b) the wavelength, (c) the period, (d) the wave speed, and (e) the power carried by the wave.
57. A spring of mass m and spring constant k has an unstretched length L_0. Find an expression for the speed of transverse waves on this spring when it's been stretched to a length L.
58. When a 340-g spring is stretched to a total length of 40 cm, it supports transverse waves propagating at 4.5 m/s. When it's stretched to 60 cm, the waves propagate at 12 m/s. Find (a) the spring's unstretched length and (b) its spring constant.
59. At a point 15 m from a source of spherical sound waves, you measure the intensity 750 mW/m^2. How far do you need to walk, directly away from the source, until the intensity is 270 mW/m^2?
60. Figure 14.37 shows two observers 20 m apart on a line that connects them to a spherical light source. If the observer nearer the source measures a light intensity 50% greater than the other observer, how far is the nearer observer from the source?

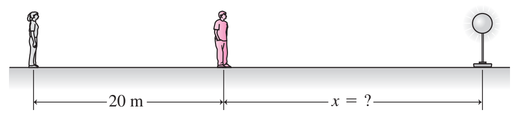


FIGURE 14.37 Problem 60

61. An ideal spring is stretched to a total length L_1. When that length is doubled, the speed of transverse waves on the spring triples. Find an expression for the unstretched length of the spring.
62. Show that the time it takes a wave to propagate up the cable in Problem 50 is $t = 2\sqrt{L/g}$, where L is the cable length.
63. You see an airplane 5.2 km straight overhead. Sound from the plane, however, seems to be coming from a point back along the plane's path at 35° to the vertical. What's the plane's speed, assuming an average sound speed of 330 ms?
64. What are the intensities in W/m^2 of sound with intensity levels of (a) 65 dB and (b) −5 dB?
65. Show that a doubling of sound intensity corresponds to approximately a 3-dB increase in the decibel level.
66. Sound intensity from a localized source decreases as the inverse square of the distance, according to Equation 14.8. If the distance from the source doubles, what happens to (a) the intensity and (b) the decibel level?
67. At 2.0 m from a localized sound source you measure the intensity level as 75 dB. How far away must you be for the perceived loudness to drop in half (i.e., to an intensity level of 65 dB)?
68. The A-string (440 Hz) on a piano is 38.9 cm long and is clamped tightly at both ends. If the string is under 667-N tension, what's its mass?

69. Show that the standing-wave condition of Equation 14.13 is equivalent to the requirement that the time it takes a wave to make a round trip from one end of the medium to the other and back be an integer multiple of the wave period.
70. You're designing an organ for a new concert hall; the lowest note is to be 22 Hz. The building's architects have asked you to minimize the lengths of the organ pipes. How long will the longest pipe be if it's (a) closed at one end and (b) open at both ends?
71. Show by differentiation and direct substitution that a wave described by Equation 14.3 satisfies the wave equation (Equation 14.5), with wave speed $v = \omega/k$.
72. Show by differentiation and direct substitution that *any* function of the form $y = f(x \pm vt)$ satisfies the wave equation (Equation 14.5).
73. You're standing roadside as a truck approaches, and you measure the dominant frequency in the truck noise to be 1100 Hz. As the truck passes, the frequency drops to 950 Hz. What's the truck's speed?
74. You're between two loudspeakers emitting 180-Hz tones. How fast would you have to move to perceive a beat frequency of 1.5 Hz between the two?
75. BIO You're a marine biologist concerned with the effect of sonic booms on plankton, and you need to estimate the altitude of a supersonic aircraft flying directly over you at 2.2 times the speed of sound. You hear its sonic boom 19 s later. Assuming a constant 340 m/s sound speed, find the plane's altitude.
76. A 1.5-m-long pipe has one end open. Among its possible standing-wave frequencies is 225 Hz; the next higher frequency is 375 Hz. Find (a) the fundamental frequency and (b) the sound speed.
77. A wave source recedes from you at 8.2 m/s, and the wavelength you measure is 20% greater than what you would measure if the source were at rest. What's the wave speed?
78. BIO Obstetricians use ultrasound to monitor fetal heartbeat. If 5.0-MHz ultrasound reflects off the moving heart wall with a 100-Hz frequency shift, what's the speed of the heart wall? (*Hint:* You have *two* shifts to consider.)
79. You're in traffic court, trying to argue your way out of a speeding ticket. You were stopped going 120 km/h in a 90-km/h zone. A technical expert testifies that the 70-GHz police radar signal underwent a 15.6-kHz frequency shift when it reflected off your car. You claim that corresponds to an impossible 240 km/h, so the police radar must be defective. How should the judge rule?
80. You move at speed u toward a wave source that's stationary with respect to the medium in which waves of wavelength λ propagate with speed v. Your speed relative to the wave crests is therefore $v + u$. Show that for you, the time between wave crests is $T' = \lambda/(v + u)$, and from this show that you perceive a frequency given by Equation 14.16, with the + sign.
81. You're a meteorologist specifying a new Doppler radar system that determines the velocity of distant raindrops by reflecting radar signals (which travel at the speed of light) off them and measuring the Doppler shift. You need a system that will measure speeds as low as 2.5 km/h. A vendor offers a 5.0-GHz radar that can detect a frequency shift of only 50 Hz. Is that sufficient?
82. C Use a computer to form the sum implied in the caption of Figure 14.17, taking $\omega = 1\ \text{s}^{-1}$ and using (a) the three terms shown and (b) 10 terms (note that only odd harmonics appear in the sum). Plot your result over one cycle (t from 0 to 2π) and compare with the square wave shown in the figure.
83. Your little sister and her friend build treehouses and stretch a rope between them for sending messages. They hang a 1.4-kg mass on one end of the rope that passes over a pulley. The other end is tied to the second treehouse. When your sister plucks the rope, a wave propagates at 18 m/s. The girls deem this too slow; they want to increase the wave speed to 30 m/s. Your sister asks, "What mass should I use?" What do you reply?

Passage Problems

Tsunamis are ocean waves generally produced when earthquakes suddenly displace the ocean floor, and with it a huge volume of water. Unlike ordinary waves on the ocean surface, a tsunami involves the entire water column, from surface to bottom. To a tsunami, the ocean is shallow—and that makes tsunamis *shallow-water waves*, whose speed is $v = \sqrt{gd}$, where d is the water depth and g the gravitational acceleration. Tsunamis can travel thousands of kilometers across an ocean to reach the shore with their initial energy nearly intact; when they do, they can cause massive damage and loss of life (Fig. 14.38).

FIGURE 14.38 People flee as the devastating tsunami of December 2004 strikes Thailand (Passage Problems 84–87).

84. As a tsunami approaches shore, it
    a. speeds up.
    b. slows down.
    c. maintains its speed.
85. For a tsunami to behave as a shallow-water wave, its wavelength
    a. must be comparable to or longer than the ocean depth.
    b. must be shorter than the ocean depth.
    c. can have any value.
86. A tsunami is traveling at 450 km/h when the ocean depth abruptly doubles. Its new speed is roughly
    a. 225 km/h.
    b. 320 km/h.
    c. 640 km/h.
    d. 900 km/h.
87. On the open ocean, a tsunami has relatively small amplitude—typically 1 m or less. As the tsunami approaches shore, its amplitude increases and its wavelength decreases. As a result,
    a. its total energy increases.
    b. the rate at which it carries energy shoreward increases.
    c. the wave frequency increases.
    d. none of these quantities changes.

Answers to Chapter Questions

Answer to Chapter Opening Question

None. The waves transport energy, but not matter.

Answers to GOT IT? Questions

14.1. (b) 5 m/s, because that's the speed of the wave crest. 2 m/s is the speed of the localized disturbance, not the wave speed.
14.2. (a) Upper wave; (b) lower; (c) lower; (d) upper; (e) upper (both f and ω).
14.3. (c).
14.4. (b), because of wave interference analogous to that shown in Fig. 14.21.
14.5. $\frac{4}{5}$ m, $\frac{4}{3}$ m, 4 m—one-fourth of each value fits into 1 m an odd number of times.

15 Fluid Motion

Why is only the "tip of the iceberg" above water?

New Concepts, New Skills

By the end of this chapter you should be able to

- Explain the relation between pressure and force (15.1).
- Calculate pressure as a function of depth in liquids (15.2).
- Explain why some objects float and others sink, and determine quantitatively the position of floating objects and the apparent weight of submerged objects (15.3).
- Express conservation of mass and energy for fluids through the continuity equation and Bernoulli's equation, and use the two to solve problems involving fluid dynamics (15.4, 15.5).

Connecting Your Knowledge

- Force (Chapter 4) and energy (Chapter 6) are the key concepts behind fluid motion.
- Kinetic energy and gravitational potential energy are both important in describing fluid motion (6.3, 7.2).

A tornado whirls across a darkened sky. A plane flies, supported by air pressure on its wings. Gas from a giant star forms a cosmic whirlpool before plunging into a black hole. Fluid in your car's brake system amplifies the force of your foot on the brake pedal. Your own body is sustained by air moving into and out of your lungs, and by the flow of blood throughout your tissues. All these examples involve fluid motion.

Fluid is matter that flows under the influence of external forces. Fluids include both liquids and gases. The intermolecular forces are weaker in fluids than in solids, and as a result the molecules move around readily. In a liquid, those forces are strong enough to keep the molecules in close contact, while in a gas they're almost negligible and the molecules are usually widely spaced. Mobility of the individual molecules means that a fluid spreads out to take the shape of its container.

15.1 Density and Pressure

If we could observe a fluid on the molecular scale, we would find large numbers of molecules in continuous motion, colliding frequently with each other and with the walls of their containers. This molecular behavior is governed by the laws of mechanics, and in principle we could study fluids by applying those laws to all the individual molecules. But even a drop of water contains about 10^{21} molecules; to calculate the motions of all those molecules would take the fastest computers many times the age of the universe!

Because the number of molecules is so large, we approximate a fluid by considering it to be continuous rather than composed of discrete particles. In this approximation, valid for fluid samples large compared with the distance between molecules, we describe the fluid by specifying macroscopic properties such as density and pressure.

Density

Density (symbol ρ, Greek rho) measures the mass per unit volume; its SI units are $\mathrm{kg/m^3}$. Water's density is normally about $1000\ \mathrm{kg/m^3}$; air's is about a factor of 1000 smaller. Because their molecules are essentially in contact, liquids are **incompressible**, meaning that their densities remain nearly constant. Gases, in contrast, are **compressible**: With relatively large intermolecular distances, their densities change readily.

Pressure

Pressure measures the normal force per unit area exerted by a fluid (Fig. 15.1):

$$p = \frac{F}{A} \quad \text{(pressure)} \tag{15.1}$$

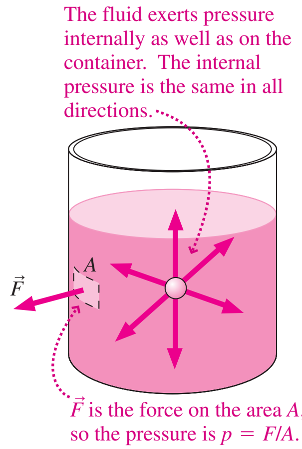


FIGURE 15.1 Pressure, the force per unit area, is exerted equally in all directions.

The SI pressure unit is $\mathrm{N/m^2}$, given the name **pascal** (Pa) after the French mathematician, scientist, and philosopher Blaise Pascal (1623–1662). Another commonly used pressure unit is the **atmosphere** (atm), defined as Earth's normal atmospheric pressure at sea level and equal to 101.3 kPa (14.7 pounds per square inch, or psi).

Pressure is a scalar quantity; at a given point in a fluid, pressure is exerted equally in all directions (Fig. 15.1), so it makes no sense to associate a direction with it. This property explains an aspect of pressure that you may find puzzling. Although the atmosphere bears down on your body with a pressure of 14.7 pounds on every square inch, you certainly don't feel that burden. That's because the force arising from this pressure is everywhere perpendicular to your body, and your body fluids respond by compressing until they're at the same pressure. If you've had your ears "pop" in a fast elevator or airplane, or when diving underwater, you know the pain that can develop when the pressure on your body is temporarily imbalanced.

15.2 Hydrostatic Equilibrium

For a fluid to remain at rest, the net force everywhere in the fluid must be zero; this condition is **hydrostatic equilibrium**. In the absence of any external forces, hydrostatic equilibrium requires that the pressure be constant throughout the fluid; otherwise, pressure differences would result in a net force, and the fluid would move in response. As Fig. 15.2 suggests, it's pressure *difference*, rather than pressure itself, that gives rise to forces within fluids.

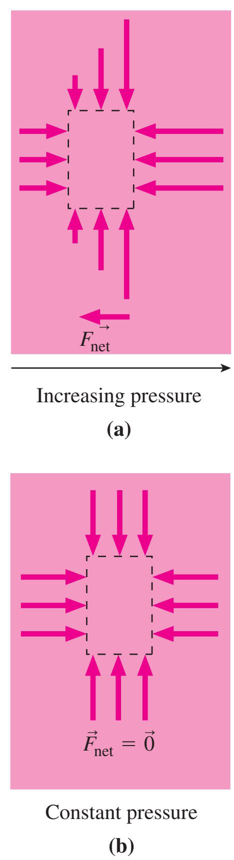


FIGURE 15.2 If pressure varies with position, then there's a net force on a volume of fluid.

Hydrostatic Equilibrium with Gravity

Hydrostatic equilibrium in the presence of gravity requires a pressure force to counteract the gravitational force. Since forces arise only from pressure differences, the fluid pressure must therefore vary with depth.

Figure 15.3 (next page) shows the forces on a fluid element of area A, thickness dh, and mass dm. A gravitational force acts downward on this fluid element; for it to be in equilibrium there must therefore be an upward pressure force—and that requires a greater pressure on the lower side. Suppose the pressures at the top and bottom are p and $p + dp$, respectively. Since pressure is force per unit area, the net pressure force is $dF_{\text{press}} = (p + dp)A - pA = A\,dp$. The gravitational force is $dF_g = -g\,dm$, where the minus sign designates the downward direction. But the mass dm is the density times the volume, so $dF_g = -g\,dm = -g\rho A\,dh$. Hydrostatic equilibrium requires that these forces sum to zero: $A\,dp - g\rho A\,dh = 0$, or

$$\frac{dp}{dh} = \rho g \quad \text{(hydrostatic equilibrium)} \tag{15.2}$$

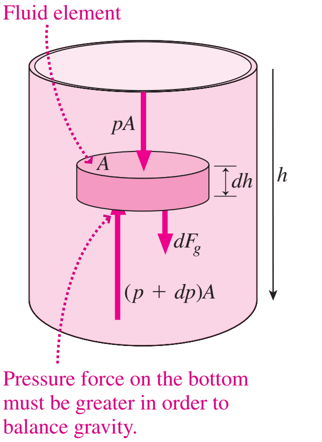


FIGURE 15.3 Forces on a fluid element in hydrostatic equilibrium.

This equation shows that dp/dh—the variation in pressure with depth h—is positive, confirming that pressure increases with depth. For a liquid, which is essentially incompressible, ρ is constant and Equation 15.2 shows that pressure increases linearly with depth:

$$p = p_0 + \rho g h \tag{15.3}$$

where p_0 is the pressure at the liquid surface.

Equation 15.2 applies to any fluid in a uniform gravitational field; Equation 15.3 follows from Equation 15.2 for the special case of a liquid. It's also possible to integrate Equation 15.2 to find the pressure in a gas that's subject to the gravitational force. Because the gas density isn't constant, this is a little more involved mathematically. Problem 68 explores the variation of pressure with height in Earth's atmosphere.

EXAMPLE 15.1 Calculating Pressure: Ocean Depths

(a) At what water depth is the pressure twice atmospheric pressure? (b) What is the pressure at the bottom of the 11-km-deep Marianas Trench, the deepest point in the ocean? Take atmospheric pressure as 100 kPa and the density of water as 1000 kg/m^3.

INTERPRET This problem is about hydrostatic equilibrium, with water the fluid.

DEVELOP We determine that Equation 15.3, $p = p_0 + \rho g h$, applies, with p_0 equal to the atmospheric pressure at the water surface. Then at twice atmospheric pressure, $p = 2p_0$, and we can solve for h to answer part (a). Because pressure increases linearly with depth, we can extrapolate our result for part (a) to find the answer to part (b).

EVALUATE Solving our equation for the depth h and substituting the given numbers in, we find for part (a):

$$h = \frac{p - p_0}{\rho g} = \frac{2.0\times10^5\ \text{Pa} - 1.0\times10^5\ \text{Pa}}{(1000\ \text{kg/m}^3)(9.8\ \text{m/s}^2)} = 10\ \text{m}$$

For part (b) we note that the pressure continues to increase by 100 kPa for every 10 m of depth. In the Marianas Trench, 11×10^3 m deep, the pressure increase is then

$$(11\times10^3\ \text{m})(100\ \text{kPa/10 m}) = 110\ \text{MPa}$$

ASSESS This is over a thousand times atmospheric pressure, or more than 8 tons per square inch! Creatures living at these depths are in pressure equilibrium with their surroundings. To bring them to the surface for study, scientists must maintain their natural pressure or they'll explode. A similar plight awaits scuba divers who hold their breath while ascending; air in the lungs expands, bursting the alveoli. ■

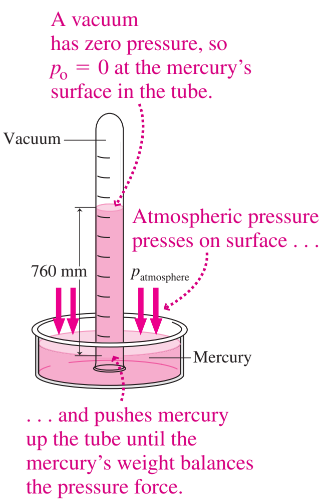


FIGURE 15.4 A mercury barometer.

Measuring Pressure

Figure 15.4 shows a **barometer**, in which air pressure acts on the open pool of mercury, pushing the liquid into the evacuated tube. Since $p_0 = 0$ in the vacuum at the top of the tube, Equation 15.3 becomes simply $p = \rho g h$, showing that the height h of the mercury is directly proportional to atmospheric pressure p. Standard atmospheric pressure of 101.3 kPa supports a mercury column 760 mm or 29.92 in. high. Pressure varies slightly with meteorological conditions, and weather forecasters regularly report atmospheric pressure in millimeters or inches of mercury. Mercury's high density makes for a reasonable-sized barometer. Example 15.1 shows that a water-filled barometer would need to be 10 m long!

A **manometer** is a U-shaped tube filled with liquid and used to measure pressure differences. A pressure difference between the two ends results in a height difference h between the liquid surfaces (Fig. 15.5, next page). Equation 15.3 shows that h is directly proportional to the pressure difference.

Barometers and manometers are the classic pressure-measuring instruments, and understanding them will help you grasp the meaning of pressure. But pressure-measuring devices today are usually electronic, using the pressure force to alter electrical properties and produce an electrical signal proportional to pressure.

The term **gauge pressure** describes the excess pressure above atmospheric. Inflation instructions for tires and sports equipment specify gauge pressure. A tire inflated to 200 kPa (about 30 psi) has an absolute pressure of about 300 kPa because of the additional 100-kPa atmospheric pressure.

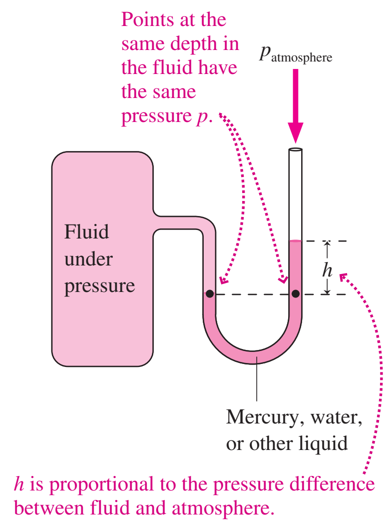


FIGURE 15.5 A manometer used to measure the pressure difference between a closed container and the atmosphere.

Pascal's Law

Equation 15.3 shows that an increase in surface pressure p_0 results in the same pressure increase throughout the fluid. More generally, a pressure increase anywhere is felt throughout the fluid—a fact known as **Pascal's law**. Pascal applied this principle in his invention of the hydraulic press. Today hydraulic systems, based on Pascal's law, control machinery ranging from automobile brakes to aircraft wings, bulldozers, cranes, and robots.

EXAMPLE 15.2 Applying Pascal's Law: A Hydraulic Lift

In the hydraulic lift of Fig. 15.6, a large piston supports a car; the total mass of car and piston is 3200 kg. What force must be applied to the smaller piston to support the car?

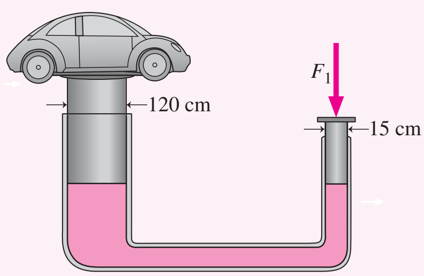


FIGURE 15.6 A hydraulic lift.

INTERPRET We interpret this as a problem involving Pascal's law. Whatever pressure results from the force on the smaller piston is transmitted through the fluid to the larger piston and thus supports the car.

DEVELOP We're given a drawing. Having determined that Pascal's law applies, and neglecting pressure variations with depth, we conclude that the pressure is the same throughout the system. Our plan, then, is to write expressions involving the pressures at both pistons and use the fact that they're equal to solve for the unknown force. We'll use the fact that the pressure on a piston is the applied force divided by the piston's area.

EVALUATE The small piston exerts a pressure $p = F_1/A_1 = F_1/\pi R_1^2$, where F_1 is the unknown force. The pressure at the large piston is the same and produces a force $F_2 = pA_2$. This force supports the weight mg of piston and car; therefore, we have

$$mg = pA_2 = p\pi R_2^2 = \frac{F_1}{\pi R_1^2}\pi R_2^2 = F_1\left(\frac{R_2}{R_1}\right)^2$$

Solving for F_1 gives our answer:

$$F_1 = mg\left(\frac{R_1}{R_2}\right)^2 = (3200\text{ kg})(9.8\text{ m/s}^2)\left(\frac{15\text{ cm}}{120\text{ cm}}\right)^2 = 490\text{ N}$$

We used the diameters from Fig. 15.3, rather than the radii, because their ratio is the same.

ASSESS How can a 490-N force—about 100 lb—support the car? Through the constant fluid pressure, this smaller force is effectively multiplied by the ratio of the piston areas. If we lifted the car farther, the work done in moving the small piston—the product of the force and the distance moved—would be equal to the work done on the large piston. Since the force on the large piston is greater, the distance moved is smaller, and energy is conserved. ■

15.3 Archimedes' Principle and Buoyancy

Why do some objects float while others sink? Figure 15.7*a* shows the upward pressure force on an arbitrary fluid volume balancing the downward gravitational force. Now imagine replacing the fluid volume with a solid object of identical shape (Fig. 15.7*b*). The remaining fluid hasn't changed, so it continues to exert an upward force on the object—a force whose magnitude equals the weight of the *original fluid volume*. This force is the **buoyancy force**, and in giving its magnitude we've stated **Archimedes' principle**: The buoyancy force on an object is equal to the weight of the fluid displaced by the object.

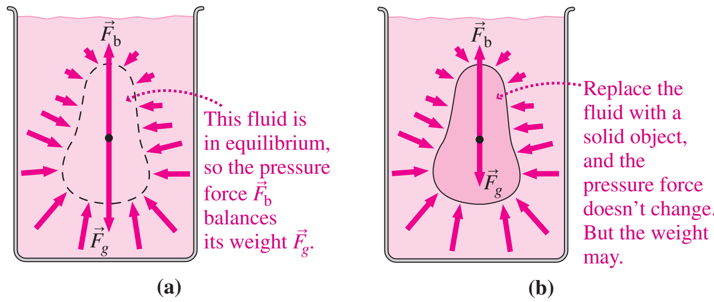


FIGURE 15.7 The buoyancy force $\vec{F}_b$ arises because pressure increases with depth.

If the submerged object weighs more than the displaced fluid, then the gravitational force exceeds the buoyancy force and the object sinks. If the object weighs less than the displaced fluid, buoyancy is greater and the object rises. Therefore, an object floats or sinks depending on whether its average density is greater or less than that of the fluid. In between is the case of **neutral buoyancy**, when an object's average density is the same as that of the fluid.

EXAMPLE 15.3 Finding the Buoyancy Force: Working Underwater

You're setting up a raft in a swimming area, and you need to move a 60-kg concrete block on the lake bottom. What's the apparent weight of the block as you lift it underwater? The density of concrete is 2200 kg/m^3.

INTERPRET We interpret this as a problem about buoyancy; the concrete will seem to weigh less underwater because of the upward buoyancy force. We identify the apparent weight as the force you'll need to apply to lift the block off the lake bottom.

DEVELOP Figure 15.8 is our sketch, showing gravity and the buoyancy force on the block; you'll need to apply a force equal but opposite to their sum. Archimedes' principle applies, giving a buoyancy force equal to the weight of water that occupies the same volume as the concrete block. So our plan is to find that force and compare it with the gravitational force on the block.

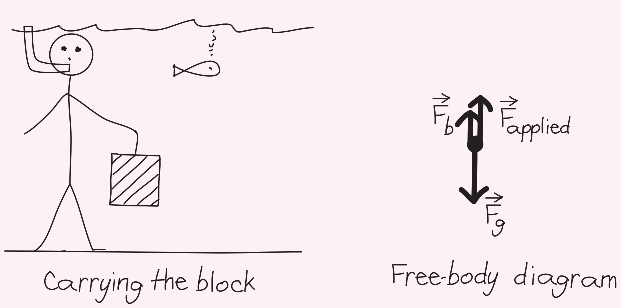


FIGURE 15.8 What's the apparent weight of the concrete block?

EVALUATE The concrete block's mass is m_c, so its weight is the gravitational force $F_g = m_c g$. Its volume is $V_c = m_c/\rho_c$, which also equals the volume of the displaced water: $V_w = V_c = m_c/\rho_c$. Archimedes' principle says that the weight of this displaced water is the magnitude of the buoyancy force, so $F_b = m_w g = V_w \rho_w g = m_c g(\rho_w/\rho_c)$. Then the upward buoyancy force and the downward gravitational force sum to give a downward force of magnitude:

$$F_g - F_b = m_c g - m_c g\left(\frac{\rho_w}{\rho_c}\right) = m_c g\left(1 - \frac{\rho_w}{\rho_c}\right)$$
$$= (60\text{ kg})(9.8\text{ m/s}^2)\left(1 - \frac{1}{2.2}\right) = 320\text{ N}$$

You have to apply an upward force of equal magnitude to lift the block off the bottom.

ASSESS This is about 70 lb—a lot more manageable than the block's weight *mg* of nearly 600 N or about 130 lb in air. Knowing the apparent weight of a submerged object would let us turn this problem around to determine its density. Archimedes purportedly used his principle in this way to find the density of the king's crown, and thus show that it was not pure gold. ■

Floating Objects

Archimedes' principle still holds for a floating object. But now the buoyancy force must balance the object's weight—which will happen if the fluid displaced by the submerged part of the object weighs the same as the object. This condition determines how high in the water the object floats, as the next example illustrates.

EXAMPLE 15.4 Floating Objects: The Tip of the Iceberg

The average density of a typical arctic iceberg is 0.86 that of seawater. What fraction of an iceberg's volume is submerged?

INTERPRET We interpret this problem also as being about buoyancy, but now we have a floating object with buoyancy balancing gravity. Only the submerged portion contributes to the buoyancy force, so the condition of force balance will enable us to find how much of the iceberg is submerged.

DEVELOP Figure 15.9 is our sketch, showing gravitational and buoyancy forces of equal magnitude. Archimedes' principle applies here, and states that the buoyancy force is equal to the weight of water displaced by the submerged portion of the iceberg. So our plan is to find the gravitational and buoyancy forces, and then equate their magnitudes to get the submerged volume. Since we're looking for volume, we'll write any masses as products of density and volume.

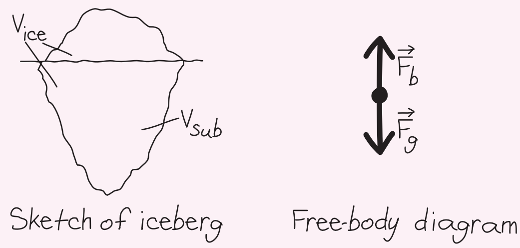


FIGURE 15.9 How much of the iceberg is submerged?

EVALUATE The iceberg's weight is $w_{ice} = m_{ice}g = \rho_{ice}V_{ice}g$, where V_{ice} is the volume of the *entire* iceberg. Only the submerged portion displaces water, so the volume of displaced water is V_{sub}, and the weight of the displaced water is therefore $w_{water} = m_{water}g = \rho_{water}V_{sub}g$. By Archimedes' principle, w_{water} is equal in magnitude to the buoyancy force, which balances gravity when the iceberg is in equilibrium. Equating the two gives $\rho_{water}V_{sub}g = \rho_{ice}V_{ice}g$, which we solve to get

$$\frac{V_{sub}}{V_{ice}} = \frac{\rho_{ice}}{\rho_{water}} = 0.86$$

ASSESS Our result means that 86% of the iceberg's volume is under water, leaving only 14% showing. Tip of the iceberg, indeed! Note that the volume ratio is just the density ratio ρ_{ice}/ρ_{water}, showing that the closer an object's density is to that of water, the lower it floats. ■

CONCEPTUAL EXAMPLE 15.1 The Shrinking Arctic

Arctic sea ice is melting rapidly as a result of global warming. Does this contribute to rising sea levels?

EVALUATE Your first answer might be "yes," but think again! Archimedes' principle tells us that the floating ice displaces a volume of water whose weight is equal to that of the *entire* ice—although only the submerged portion does the displacing. When the ice melts, it becomes water that, because it no longer sticks above the surface, displaces a volume equal to its entire weight. But since the weight hasn't changed, the amount of water displaced is the same. That means the water level is unchanged.

ASSESS Melting ice doesn't contribute to sea-level rise—as long as it's sea ice that melts. Land ice is a different story: Melting glaciers and "calving" of glaciers to form icebergs together cause about half of the observed sea-level rise. Thermal expansion, which we'll explore in Chapter 17, causes the rest.

MAKING THE CONNECTION The land-based Greenland ice cap occupies some 3 million km^3, while some 15,000 km^3 of ice are afloat in the Arctic Ocean. Compare the approximate rise in the world's oceans that would result from complete melting of these two ice volumes.

EVALUATE As this conceptual example shows, melting sea ice won't contribute to sea-level rise, but land-based ice will add water to the oceans. Its volume will be about 86% that of the ice (see Example 15.4), or about 2.6 million km^3. With oceans covering about 71% of Earth's surface, the meltwater will spread in a layer of thickness d, where $(0.71)(4\pi R_E^2)d = 2.6\times10^{15}$ m^3. Solving gives $d = 7$ m—enough to inundate most of today's coastal cities.

Center of Buoyancy

The buoyancy force acts not at the center of mass of a floating object, but at the center of mass of the water that would be there if the object weren't. This point is called the **center of buoyancy**, and for an object to float in stable equilibrium, the center of buoyancy must lie above the center of mass. Otherwise, a net torque results that tends to tip the object. The stability of watercraft depends critically on this condition (Fig. 15.10, next page).

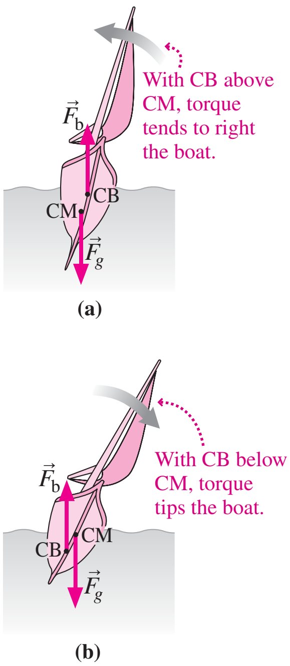


FIGURE 15.10 A boat's stability requires the center of buoyancy (CB) to be above the center of mass (CM).

15.4 Fluid Dynamics

We now turn our attention to moving fluids, described by the flow velocity at each point in the fluid and at each instant of time. We illustrate flow velocity by drawing continuous lines called **streamlines** that are everywhere tangent to the local flow direction (Fig. 15.11). Their spacing is a measure of flow speed, with closely spaced streamlines indicating higher speed. Small particles introduced into moving fluids follow streamlines and therefore give a visual indication of the flow velocity pattern.

In **steady flow**, the pattern of fluid motion remains the same at each point, even though individual fluid elements are in continuous motion. A river in steady flow always looks the same, even though you're not seeing the same water each time you look. At a given point, the water velocity is always the same. **Unsteady flow**, in contrast, involves fluid motion that changes with time. Blood flow in your arteries is unsteady; with each contraction of the heart ventricles, the pressure rises and the flow velocity increases. We'll restrict our quantitative description of fluid motion to steady flow.

Like all other motion in classical physics, fluid motion is governed by Newton's laws. It's possible to write Newton's second law in a form that involves explicitly the fluid velocity as a function of position and time. But the resulting equation is difficult to solve in any but the simplest cases. Instead of applying Newton's law directly, we'll approach fluid dynamics using energy conservation.

GOT IT? 15.1 The photo shows smoke particles tracing streamlines in a test of a car's aerodynamic properties. Is the flow speed greater (a) over the top or (b) at the back?

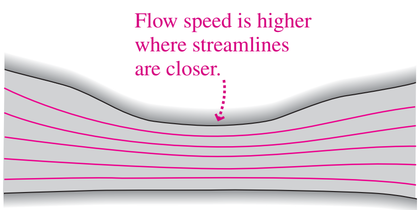


FIGURE 15.11 Streamlines represent flow velocity in a river.

Conservation of Mass: The Continuity Equation

In mechanics we had no trouble keeping track of the individual objects. But a fluid is continuous and deformable, so it's not easy to follow an individual fluid element as it moves. Yet fluid is conserved; as it moves, new fluid is neither created nor destroyed.

Consider a steady fluid flow represented by streamlines, as shown in Fig. 15.12*a*. We shaded a **flow tube**—a small tubelike region bounded on its sides by streamlines and on its ends by areas at right angles to the flow. The flow tube has a sufficiently small cross section that fluid velocity and other properties don't vary significantly over any cross section; however, fluid properties may vary along the flow tube. Although our flow tube has no physical boundaries, it nevertheless acts like a pipe because fluid flows *along*, not across, the streamlines. In steady flow, the rate at which fluid enters the tube at its left end must equal the rate at which it exits at the right.

Figure 15.12*b* shows a small fluid element just about to enter the flow tube, a process that will take some time Δt. Suppose the fluid is moving at speed v_1; since it takes time Δt to cross the tube end, its length is $v_1\,\Delta t$. With cross-sectional area A_1, length $v_1\,\Delta t$, and density ρ_1, the mass of the entering fluid is $m = \rho_1 A_1 v_1\,\Delta t$.

Another fluid element is shown just about to leave the tube. Suppose it has the *same* mass m as the entering fluid element. Then it must exit the tube in the *same* time Δt in order to keep the total mass in the tube constant. Its mass can be written as $m = \rho_2 A_2 v_2\,\Delta t$.

Equating our two expressions for m shows that $\rho_1 v_1 A_1 = \rho_2 v_2 A_2$. Since the endpoints of the tube are arbitrary, we conclude that the quantity $\rho v A$ must have the same value anywhere along the flow tube:

$$\rho v A = \text{constant along a flow tube} \quad \begin{pmatrix}\text{continuity equation,}\\ \text{any fluid}\end{pmatrix} \qquad (15.4)$$

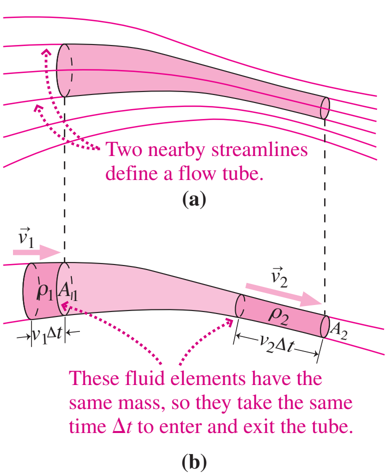


FIGURE 15.12 In steady flow, fluid enters and leaves a flow tube at the same rate.

Equation 15.4 is the **continuity equation**, which expresses the conservation of mass in steady fluid flow. The units of $\rho v A$ here are $(\text{kg/m}^3)(\text{m/s})(\text{m}^2)$, or simply kg/s. This quantity is therefore the **mass flow rate** or mass of fluid per unit time passing through the flow tube. Equation 15.4 says that the mass flow rate is constant in steady flow.

For a liquid, the density ρ is constant, and the continuity equation becomes simply

$$vA = \text{constant along a flow tube} \quad \left(\begin{array}{c}\text{continuity equation,}\\ \text{liquid}\end{array}\right) \tag{15.5}$$

Now the constant quantity is just vA, with units of $(\text{m/s})(\text{m}^2)$, or m^3/s. This is the **volume flow rate**. Equation 15.5 makes sense: Where a liquid's cross-sectional area is large, it flows slowly to transport a given volume of fluid per unit time. But in a constricted area, it must flow faster to carry the same volume. With a gas, obeying Equation 15.4 but not necessarily 15.5, the situation is slightly more ambiguous because density variations also play a role. For flow speeds below the speed of sound in a gas, it turns out that smaller area implies a higher flow speed just as for a liquid. But when the gas flow speed exceeds the sound speed, density changes become so great that flow speed actually decreases with smaller area.

EXAMPLE 15.5 Using the Continuity Equation: Ausable Chasm

The Ausable River in upstate New York is about 40 m wide. Under typical early summer conditions, it's 2.2 m deep and flows at 4.5 m/s. Just before it reaches Lake Champlain, the river enters Ausable Chasm, a deep gorge only 3.7 m wide. If the flow rate in the gorge is 6.0 m/s, how deep is the river at this point? Assume a rectangular cross section with uniform flow speed.

INTERPRET The concept behind this problem is mass conservation, embodied in the continuity equation for a liquid, Equation 15.5. Since the flow is uniform over the river's cross section, we can treat the entire river as a single flow tube.

DEVELOP Equation 15.5 says that the product vA is constant. For the river's rectangular cross section, the area A is the product of width w and depth d. Then Equation 15.5 becomes $v_1w_1d_1 = v_2w_2d_2$, where the subscripts indicate values upstream and in the gorge. Our plan is to solve for the depth d_2 in the gorge.

EVALUATE Solving gives

$$d_2 = \frac{v_1w_1d_1}{v_2w_2} = \frac{(4.5\text{ m/s})(40\text{ m})(2.2\text{ m})}{(6.0\text{ m/s})(3.7\text{ m})} = 18\text{ m}$$

ASSESS This is about 60 feet, quite a depth for a small river! But conservation of mass requires it. In the gorge, the river is much narrower but its flow speed is only a little higher, so it's got to be a lot deeper. ■

Conservation of Energy: Bernoulli's Equation

We now turn to conservation of fluid energy. Figure 15.13 shows the same fluid element as it enters and again as it leaves a flow tube. If it enters with speed v_1 and leaves with speed v_2, the change in its kinetic energy is

$$\Delta K = \tfrac{1}{2}m(v_2^2 - v_1^2)$$

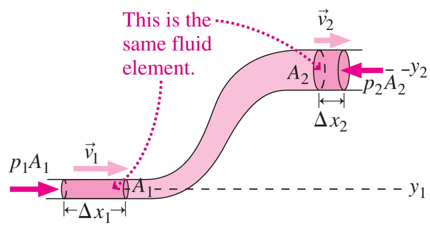


FIGURE 15.13 A flow tube showing the same fluid element entering and leaving. The work done by pressure and gravitational forces equals the change in kinetic energy of the fluid element.

The work-energy theorem (Equation 6.14) equates this change to the net work done on the fluid element. As the element enters the tube, it's subject to a pressure force p_1A_1 from the fluid to its left. This external force acts over the length Δx_1 of the fluid element as it enters, so it does work $W_1 = p_1A_1\,\Delta x_1$. Similarly, as it leaves the tube, the fluid element experiences a force p_2A_2 from the fluid to its right. Because this force is opposite the flow direction, it does negative work $W_2 = -p_2A_2\,\Delta x_2$. External forces from adjacent flow tubes act at right angles to the flow, so they do no work. Finally, the fluid element rises a distance $y_2 - y_1$ as it traverses the tube; therefore, gravity does negative work $W_g = -mg(y_2 - y_1)$. Summing these three contributions and applying the work-energy theorem, we have $W_1 + W_2 + W_g = \Delta K$, or $p_1A_1\,\Delta x_1 - p_2A_2\,\Delta x_2 - mg(y_2 - y_1) = \tfrac{1}{2}m(v_2^2 - v_1^2)$. The quantities $A_1\,\Delta x_1$ and $A_2\,\Delta x_2$ are the volumes of the fluid element as it enters and leaves the flow, respectively. If we restrict ourselves to incompressible fluids, then those volumes are equal. Dividing through by this common volume $V = A\,\Delta x$ and noting that $m/V = \rho$, we get $p_1 + \tfrac{1}{2}\rho v_1^2 + \rho g y_1 = p_2 + \tfrac{1}{2}\rho v_2^2 + \rho g y_2$, or

$$p + \tfrac{1}{2}\rho v^2 + \rho g y = \text{constant along a flow tube} \quad (\text{Bernoulli's equation}) \tag{15.6}$$

This is **Bernoulli's equation,** after the Swiss mathematician Daniel Bernoulli (1700–1782).

What do the terms in Bernoulli's equation mean? The quantity $\frac{1}{2}\rho v^2$ looks like kinetic energy $\frac{1}{2}mv^2$, except it has mass per unit volume ρ instead of mass m. It's therefore the kinetic energy per unit volume, or kinetic-energy density. Similarly, ρgy is the gravitational potential energy per unit volume. Pressure p, too, has the units of energy density and represents internal energy of the fluid. Bernoulli's equation therefore says that the total energy per unit volume of fluid is conserved as the fluid moves.

Bernoulli's equation in the form 15.6 applies to incompressible fluids. It neglects fluid friction, also called *viscosity*, that may dissipate fluid kinetic energy. It also neglects energy transfers associated with machinery such as turbines or pumps that may extract or add to the fluid's energy. Engineers often include those effects in Bernoulli's equation.

15.5 Applications of Fluid Dynamics

The laws of mass and energy conservation that we just derived for fluids allow us to analyze a wide variety of natural and technological phenomena. We'll usually need both the continuity equation and Bernoulli's equation, considering the values of the appropriate constant quantities at two points in a fluid flow. As you study the examples and applications that follow, remember that they're ultimately based in the same Newtonian principles we've been using to describe mechanical systems.

PROBLEM-SOLVING STRATEGY 15.1 Fluid Dynamics

The continuity equation and Bernoulli's equation are the keys to solving problems in fluid dynamics. Here's a strategy that will help you focus these two equations on a problem.

INTERPRET The form of Bernoulli's equation we derived applies only to incompressible fluids. So be sure you're dealing either with a liquid or with a gas flowing at speeds well below its sound speed.

DEVELOP

- Identify a flow tube. This may be a physical pipe or other structure, or a mathematical tube bounded by streamlines.
- Draw a sketch of the situation, showing the flow tube.
- Determine the point where you're interested in solving for some aspect of the flow, and another point where you know the quantities that go into the continuity equation and Bernoulli's equation. Note those quantities that you know at each point. Mark the two points on your sketch.
- Write the continuity equation and Bernoulli's equation, with the known quantities forming the terms on one side and the other side containing your unknown(s).

EVALUATE Evaluate by solving your equations for the unknown quantity or quantities. Often this will involve solving the continuity equation first and then using the result in Bernoulli's equation.

ASSESS Ask whether your result makes sense. Does flow speed increase at a constriction? Does pressure go up when flow speed drops, or vice versa? Are there any limitations that apply, or insights to be gained?

EXAMPLE 15.6 Bernoulli's Equation: Draining a Tank

A large, open tank is filled to height h with liquid of density ρ. Find the speed of liquid emerging from a small hole at the base of the tank.

INTERPRET We're dealing with a flow of water, an incompressible liquid. So we can apply our problem-solving strategy for fluid dynamics.

DEVELOP We take the tank to be a rather oddly shaped flow tube, and Fig. 15.14 is our sketch. We're interested in the water's velocity at the hole, so the hole is one of the points we'll use in the fluid equations. Since the hole is open to the atmosphere, the pressure at

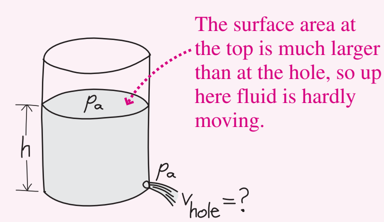


FIGURE 15.14 How fast does the liquid emerge from the tank?

the hole is atmospheric pressure p_a. The top surface is also open to the atmosphere, so here the pressure is also p_a. Now, because the hole is very small in relation to the tank, the water level drops only slowly. Therefore, we can make the approximation $v = 0$ at the top—and thus we know both p and v at the top. Although we didn't write a formal equation here, that approximation follows from the continuity equation because the ratio of hole to top surface area is so small. We also need the potential-energy terms in Bernoulli's equation. If we take $y = 0$ at the hole, then those terms are zero at the hole and ρgh at the top. Then Bernoulli's equation, $p + \frac{1}{2}\rho v^2 + \rho gy = \text{constant}$, becomes

$$p_a + \rho gh = p_a + \tfrac{1}{2}\rho v_{\text{hole}}^2$$

where the terms on the left are at the top surface and those on the right are at the hole. We've taken care of the continuity equation through our assumption of negligible flow speed at the top.

EVALUATE Atmospheric pressure cancels, and we solve for the unknown flow velocity at the hole:

$$v_{\text{hole}} = \sqrt{2gh}$$

ASSESS This is the same result we would get by dropping an object from a height h—and for the same reason: conservation of energy. Draining a gram of water from the hole is energetically equivalent to removing a gram of water from the top and dropping it. Just as the speed of a falling object is independent of its mass, so the speed of the liquid is independent of its density. As the liquid drains, of course, the height decreases and so does the flow rate. That's a calculus challenge you can try in Problem 67.

✓TIP Reasonable Approximations

Making reasonable approximations is often important in solving realistic problems. Look for opportunities to approximate a physical quantity, especially when other terms appear more significant. But always be sure that your approximations are reasonable. In this example, we reasoned that the fluid's speed at the top of the tank was negligible because it's proportional to the ratio of the hole to the top surface area, a very small value.

Venturi Flows and the Bernoulli Effect

A constriction in a pipe carrying incompressible fluid requires that the flow speed increase in order to maintain constant mass flow. Such a constriction is a **venturi**. Because of the increased speed, Bernoulli's equation requires the pressure to be lower in the venturi. The next example shows how this effect provides a measure of fluid flow.

EXAMPLE 15.7 Measuring Flow Speed: A Venturi Flowmeter

An incompressible fluid of density ρ flows through a horizontal pipe of cross-sectional area A_1. The pipe has a venturi constriction of area A_2, and a gauge measures the pressure difference Δp between the unconstricted pipe and the venturi. Find an expression for the flow speed in the unconstricted pipe.

INTERPRET This is a problem about incompressible fluid flow, so our strategy applies.

DEVELOP For a flow tube, we choose a section of pipe that includes the venturi. Figure 15.15 is a sketch showing some streamlines through this tube. We're interested in the flow velocity in the unconstricted pipe, so any point outside the venturi will do. The other point should be in the venturi. The continuity equation then reads $v_1A_1 = v_2A_2$, where the subscript 1 refers to the unconstricted pipe and 2 to the venturi. The pipe is horizontal, so the potential-energy term ρgh in Bernoulli's equation is the same on both sides, and it drops out. Bernoulli's equation then reads

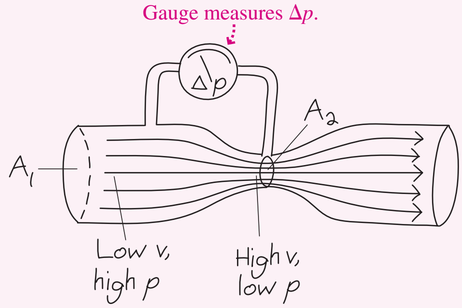


FIGURE 15.15 Our sketch of a venturi flowmeter.

$$p_1 + \tfrac{1}{2}\rho v_1^2 = p_2 + \tfrac{1}{2}\rho v_2^2$$

EVALUATE We can eliminate the velocity v_2 by solving the continuity equation: $v_2 = (A_1/A_2)v_1 = bv_1$, where we defined b as the ratio of the larger to smaller area. Using this result in Bernoulli's equation gives $p_1 + \frac{1}{2}\rho v_1^2 = p_2 + \frac{1}{2}\rho b^2 v_1^2$. In terms of the pressure difference $\Delta p = p_1 - p_2$, this becomes $\Delta p = \frac{1}{2}\rho b^2 v_1^2 - \frac{1}{2}\rho v_1^2 = \frac{1}{2}\rho v_1^2(b^2 - 1)$. We then solve for v_1 to get our answer:

$$v_1 = \sqrt{\frac{2\,\Delta p}{\rho(b^2 - 1)}}$$

ASSESS Make sense? The pressure difference results from the change in speed; no flow, no pressure difference. So it's reasonable that v increases with Δp. But a given pressure difference Δp is easier to get with a larger area ratio b, so flow speed depends inversely on b. Finally, the greater inertia of a denser fluid means a given pressure difference produces less acceleration, implying a lower initial speed; that's why ρ appears in the denominator.

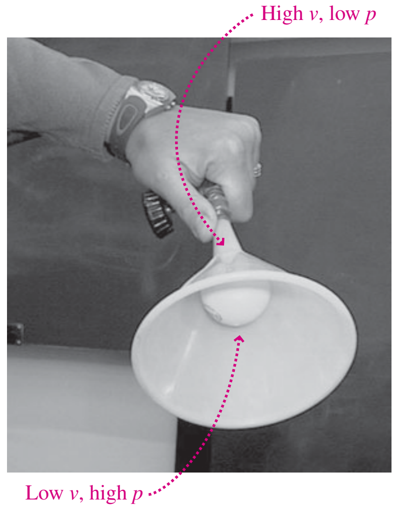
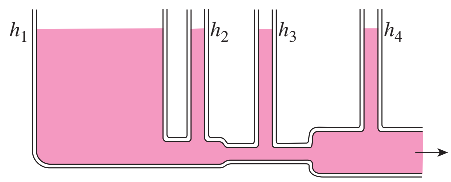


FIGURE 15.16 A ping-pong ball supported by downward-flowing air. High-velocity flow is inside the narrow part of the funnel.

The occurrence of lower pressure with higher flow speeds, and vice versa—the **Bernoulli effect**—has numerous manifestations. The dirt around a prairie dog's hole is mounded up in a way that forces wind to accelerate over the hole, resulting in lower pressure above the hole. Biologists speculate that prairie dogs have evolved this design to provide natural ventilation. The Bernoulli effect can be strikingly counterintuitive. Figure 15.16 shows a ping-pong ball suspended by *downward* airflow in an inverted funnel. Rapid divergence of the flow results in lower speed and therefore higher pressure below the ball.

GOT IT? 15.2 A large tank is filled with liquid to the level h_1 shown in the figure. It drains through a small pipe whose diameter varies; emerging from each section of pipe are vertical tubes open to the atmosphere. Although the picture shows the same liquid level in each pipe, they really won't be the same. Rank order the levels h_1 through h_4.



Flight and Lift

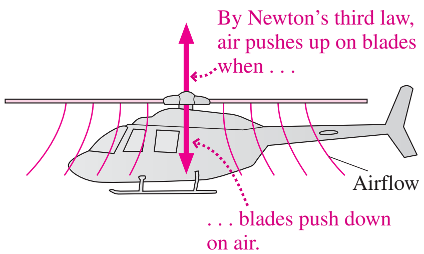


FIGURE 15.17 Newton's third law explains the helicopter's flight.

Airplanes, helicopters, and birds fly using forces resulting from their dynamic interaction with the air. Hydrofoil boats, water skis, and sailboards have analogous interactions with water. Projectiles such as baseballs, though not supported by the air, have their trajectories substantially modified by aerodynamic forces.

One of the simplest examples of aerodynamic **lift** is the helicopter. Its whirling blades are tilted so they force air downward as they move, just like a giant fan (Fig. 15.17). By Newton's third law, the air exerts an upward force on the blades, ultimately supporting the helicopter. An airplane wing works in the same way, except that it moves forward in a straight line instead of describing a circle. Wings are shaped to maximize the downward deflection of the air even with the wing horizontal, but in principle even a flat board would function as a wing if it were tilted to the oncoming air. Figure 15.18 shows the airflow around a wing. Note how the flow, initially horizontal, leaves the wing moving downward—a clear indication that the wing has exerted a downward force on the air. The third law requires a corresponding upward force, and that's what supports the plane.

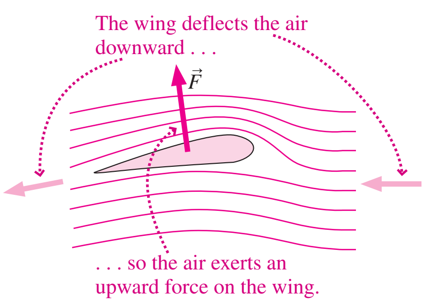


FIGURE 15.18 Flow past a wing.

Baseball's "curve ball" provides another example of aerodynamic lift. Figure 15.19*a* is a top view of the airflow around a baseball that's not spinning; the flow is symmetric and the air isn't deflected. But if the ball spins as shown in Fig. 15.19*b*, air is dragged around the ball and deflected. A corresponding third-law force then acts on the ball, curving its path.

Bernoulli's equation is frequently invoked to explain lift forces. It's true, as Figs. 15.18 and 15.19*b* suggest, that flow speeds are higher, and therefore—according to Bernoulli's equation—pressures are lower on top of a wing or spinning ball. Forces associated with that pressure difference provide the lift, so Bernoulli can help explain what's going on. But those pressure differences are manifestations of a simpler underlying phenomenon—namely, the paired forces of Newton's third law.

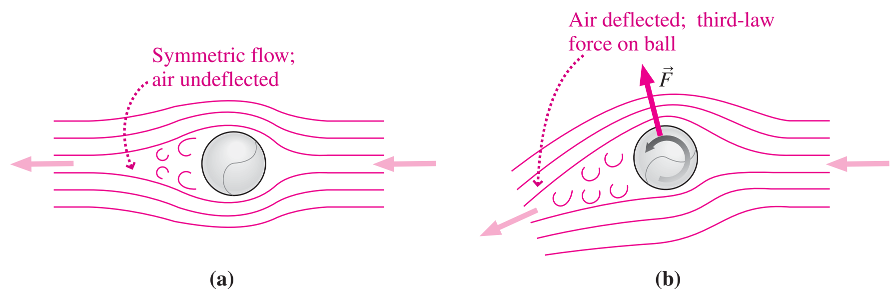


FIGURE 15.19 Top views of airflow around a baseball: (a) no spin; (b) spinning, thus a curve ball.

APPLICATION **Wind Energy**

Wind turbines extract kinetic energy from moving air. In a wind with speed v, Bernoulli's equation shows that the air has kinetic-energy density $\frac{1}{2}\rho v^2$. A chunk of air that passes through a wind turbine in time Δt has length $v\,\Delta t$ and volume $vA\,\Delta t$, where A is the area swept out by the blades. The kinetic energy in this volume is the energy density times the volume: $\Delta K = (\frac{1}{2}\rho v^2)(vA\,\Delta t) = \frac{1}{2}\rho v^3 A\,\Delta t$. Dividing by $A\,\Delta t$ gives the energy per time per unit area—that is, the power per unit area available from the wind:

$$\text{wind power per unit area} = \tfrac{1}{2}\rho v^3$$

Unfortunately, we can't extract *all* this energy because then the air would come to a complete stop behind the turbine, halting the flow. A careful analysis shows that the maximum rate for wind-energy extraction is $\frac{8}{27}\rho v^3$, about 59% of the wind's energy. Given air's density of 1.2 kg/m^3, this means a 10-m/s wind amounts to some 350 W/m^2. The factor v^3 shows that the available power increases rapidly at higher speeds. The best practical wind turbines can achieve about 80% of the theoretical maximum. Wind is the fastest-growing component of the world's energy supply, and in some European countries it provides as much as 20% of the electrical energy.

15.6 Viscosity and Turbulence

Moving fluid interacts with the surfaces it contacts, resulting in a kind of fluid friction called **viscosity**. Viscosity also results from the transfer of momentum among adjacent layers within a fluid. Viscosity is especially important right near fluid boundaries because viscous forces bring the fluid to a complete stop at the boundary (Fig. 15.20). This boundary effect produces drag forces on objects moving through fluids—but it's the same drag at the surfaces of airplane and ship propellers that exerts a force on the fluid. Without viscosity, propellers would spin uselessly and planes and ships would go nowhere.

Viscosity depends on fluid properties and dimensions. Honey is more viscous than water, but at the tiny scales of a human capillary or a bacterium wiggling its flagella for propulsion, water too can be extremely viscous. Viscosity is also important in stabilizing flows that would otherwise become **turbulent**, or chaotically unsteady. Turbulence results from the growth of waves that gain energy at the expense of the flow, turning a smooth flow into a chaotic mess (Fig. 15.21). Turbulence is still not fully understood and presents ongoing challenges to scientists and engineers.

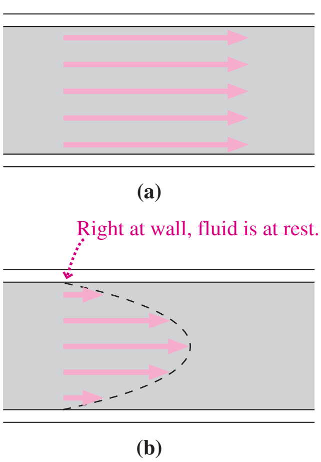


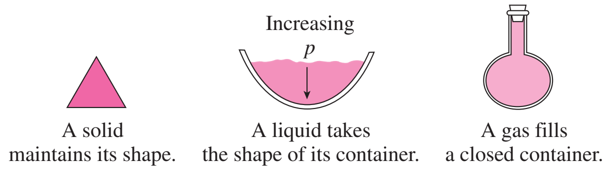
FIGURE 15.20 Velocity profiles in (a) inviscid and (b) viscous flow.

FIGURE 15.21 Smooth flow becomes turbulent.

CHAPTER 15 SUMMARY

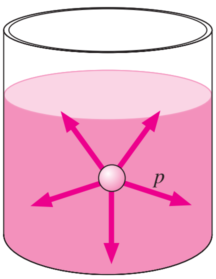
Big Picture

Fluid is matter that readily deforms and flows under the influence of forces. Pressure, density, and flow velocity characterize fluids. Liquids and slowly moving gases are **incompressible**, meaning their density is essentially constant. A fluid that isn't moving is in **hydrostatic equilibrium**. In the presence of gravity, equilibrium requires that fluid pressure increase with depth.

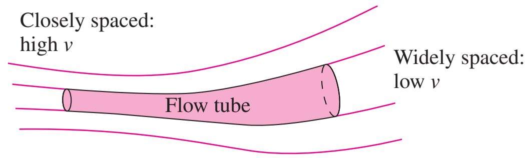


Key Concepts and Equations

Pressure is the force per unit area: $p = F/A$. The pressure in a fluid exerts itself equally in all directions.



Streamlines represent a moving fluid.



The **continuity equation** describes the conservation of mass along a flow tube:

$$\rho v A = \text{constant (any fluid)}$$

$$vA = \text{constant (incompressible fluid)}$$

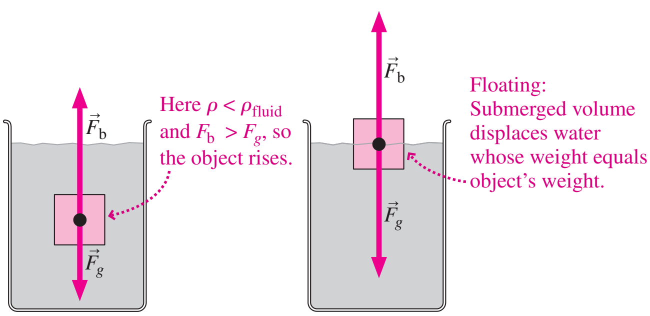
Bernoulli's equation describes the conservation of energy:

$$p + \tfrac{1}{2}\rho v^2 + \rho g y = \text{constant} \quad \text{(incompressible fluid, neglecting viscosity)}$$

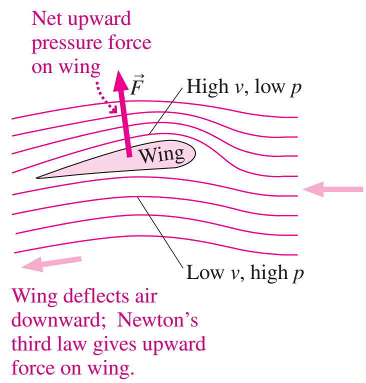
Viscosity, or fluid friction, is especially important when fluids interact with solid objects.

Applications

Archimedes' principle states that the **buoyancy force** $\vec{F}_b$ due to pressure on an object is equal to the weight of the displaced fluid. For an object less dense than a fluid, the buoyancy force exceeds gravity and the object floats; otherwise, it sinks or is in neutral buoyancy.



Bernoulli's principle helps explain lift forces, although ultimately these are based in Newton's third law.



For Thought and Discussion

1. Why do your ears "pop" when you drive up a mountain?
2. Commercial aircraft cabins are usually pressurized to the pressure of the atmosphere at about 2 km above sea level. Why don't you feel the lower pressure on your entire body?
3. Water pressure at the bottom of the ocean arises from the weight of the overlying water. Does this mean that the water exerts pressure only in the downward direction? Explain.
4. The three containers in Fig. 15.22 are filled to the same level and are open to the atmosphere. How do the pressures at the bottoms of the three containers compare?

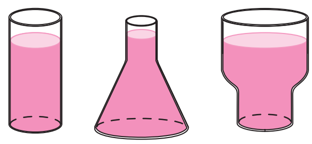


FIGURE 15.22 For Thought and Discussion 4

5. Why is it easier to float in the ocean than in fresh water?
6. Figure 15.23 shows a cork suspended from the bottom of a sealed container of water. The container is on a turntable rotating about a vertical axis, as shown. Explain the position of the cork.

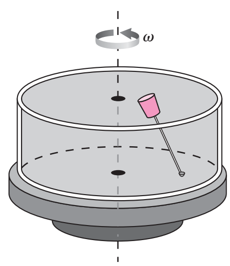


FIGURE 15.23 For Thought and Discussion 6

7. Meteorologists in the United States usually report barometer readings in inches. What are they talking about?
8. A mountain stream, frothy with entrained air bubbles, presents a serious hazard to hikers who fall into it, for they may sink in the stream where they would float in calm water. Why?
9. Why are dams thicker at the bottom than at the top?
10. It's not possible to breathe through a snorkel from a depth greater than a meter or so. Why not?
11. A helium-filled balloon stops rising long before it reaches the "top" of the atmosphere, but a cork released from the bottom of a lake rises all the way to the surface. Why the difference?
12. A barge filled with steel beams overturns in a lake, spilling its cargo. Does the water level in the lake rise, fall, or remain the same?
13. Why do airplanes take off into the wind?
14. Is the flow speed behind a wind turbine greater or less than in front? Is the pressure behind the turbine higher or lower than in front? Is this a violation of Bernoulli's principle? Explain.

Exercises and Problems

Exercises

Section 15.1 Density and Pressure

15. The density of molasses is 1600 kg/m^3. Find the mass of the molasses in a 0.75-L jar.
16. Atomic nuclei have densities around 10^{17} kg/m^3, while water's density is 10^3 kg/m^3. Roughly what fraction of water's volume is *not* empty space?
17. Compressed air with mass 8.8 kg is stored in a 0.050-m^3 cylinder. (a) What's the density of the compressed air? (b) What volume would the same gas occupy at a typical atmospheric density of 1.2 kg/m^2?
18. The pressure unit **torr** is defined as the pressure that will support a column of mercury 1 mm high. Meteorologists often give barometric pressure in **inches of mercury**, defined analogously. Express each of these in SI units. (*Hint:* Mercury's density is 1.36×10^4 kg/m^3.)
19. Measurement of small pressure differences—for example, between the interior of a chimney and the ambient atmosphere—is often given in inches of water, where 1 in. of water is the pressure that will support a 1-in.-high water column. Express this in SI units.
20. What's the weight of a column of air with cross-sectional area 1 m^2 extending from Earth's surface to the top of the atmosphere?
21. A 4300-kg circus elephant balances on one foot. If the foot is a circle 30 cm in diameter, what pressure does it exert on the ground?
22. You unbend a paper clip made from 1.5-mm-diameter wire and push the end against the wall. What force must you apply to give a pressure of 120 atm?

Section 15.2 Hydrostatic Equilibrium

23. What's the density of a fluid whose pressure increases at the rate of 100 kPa for every 6.0 m of depth?
24. A research submarine can withstand an external pressure of 50 MPa when its internal pressure is 100 kPa. How deep can it dive?
25. Scuba equipment provides a diver with air at the same pressure as the surrounding water. But at pressures higher than about 1 MPa, the nitrogen in air becomes dangerously narcotic. At what depth does nitrogen narcosis become a hazard?
26. A vertical tube open at the top contains 5.0 cm of oil with density 0.82 g/cm^3, floating on 5.0 cm of water. Find the gauge pressure at the bottom of the tube.
27. A child attempts to drink water through a 100-cm-long straw but finds that the water rises only 75 cm. By how much has the child reduced the pressure in her mouth below atmospheric pressure?
28. Barometric pressure in the eye of a hurricane is 0.91 atm (27.2 in. of mercury). How does the level of the ocean surface under the eye compare with the level under a distant fair-weather region where the pressure is 1.0 atm?

Section 15.3 Archimedes' Principle and Buoyancy

29. On land, the most massive concrete block you can carry is 25 kg. Given concrete's 2200-kg/m^3 density, how massive a block could you carry underwater?
30. A 5.4-g jewel has apparent weight 32 mN when submerged in water. Could the jewel be a diamond (density 3.51 g/cm^3)?
31. Styrofoam's density is 160 kg/m^3. What percent error is introduced by weighing a Styrofoam block in air (density 1.2 kg/m^3), which exerts an upward buoyancy force, rather than in vacuum?
32. A steel drum has volume 0.23 m^3 and mass 16 kg. Will it float in water when filled with (a) water or (b) gasoline (density 860 kg/m^3)?

Sections 15.4 and 15.5 Fluid Dynamics and Applications

33. Water flows through a 2.5-cm-diameter pipe at 1.8 m/s. If the pipe narrows to 2.0-cm diameter, what's the flow speed in the constriction?

34. Show that pressure has the units of energy density.
35. A typical mass flow rate for the Mississippi River is 1.8×10^7 kg/s. Find (a) the volume flow rate and (b) the flow speed in a region where the river is 2.0 km wide and averages 6.1 m deep.
36. A fire hose 10 cm in diameter delivers water at 15 kg/s. The hose terminates in a 2.5-cm-diameter nozzle. What are the flow speeds (a) in the hose and (b) at the nozzle?
37. BIO A typical human aorta, the main artery from the heart, is 1.8 cm in diameter and carries blood at 35 cm/s. Find the flow speed around a clot that reduces the flow area by 80%.

Problems

38. When a couple with total mass 120 kg lies on a waterbed, pressure in the bed increases by 4700 Pa. What surface area of the two bodies is in contact with the bed?
39. A fully loaded Volvo station wagon has mass 1950 kg. If each of its four tires is inflated to a gauge pressure of 230 kPa, what's the total tire area in contact with the road?
40. You're stuck in the exit row on a long flight, and you suddenly worry that your seatmate, who's next to the window, might pull the emergency window inward while you're in flight. The window measures 50 cm by 90 cm. Cabin pressure is 0.75 atm, and atmospheric pressure at the plane's altitude is 0.25 atm. Should you worry?
41. A vertical tube 1.0 cm in diameter and open at the top contains 5.0 g of oil (density 0.82 g/cm^3) floating on 5.0 g of water. Find the gauge pressure (a) at the oil-water interface and (b) at the bottom.
42. Dam breaks present a serious risk of widespread property damage and loss of life. You're asked to assess a 1500-m-wide dam holding back a lake 95 m deep. The dam was built to withstand a force of 100 GN, which is supposed to be at least 50% over the force it actually experiences. Should the dam be reinforced? (*Hint:* You'll need your calculus skills.)
43. A U-shaped tube open at both ends contains water and a quantity of oil occupying a 2.0-cm length of the tube, as shown in Figure 15.24. If the oil's density is 82% of water's, what's the height difference h?

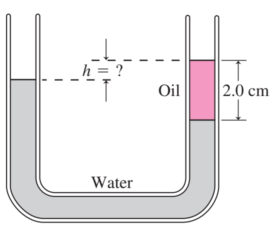


FIGURE 15.24 Problem 43

44. You're a robotics engineer designing a hydraulic system to move a robotic arm. The hydraulic cylinder that drives the arm has diameter 5.0 cm and can exert a maximum force of 5.6 kN. Hydraulic tubing comes rated in multiples of 1/2 MPa, and for safety, you're to specify tubing capable of withstanding 50% greater pressure than it will ever experience in use. What pressure rating do you specify?
45. A garage lift has a 45-cm-diameter piston supporting the load. Compressed air with maximum pressure 500 kPa is applied to a small piston at the other end of the hydraulic system. What's the maximum mass the lift can support?
46. Archimedes purportedly used his principle to verify that the king's crown was pure gold by weighing the crown submerged in water. Suppose the crown's actual weight was 25.0 N. What would be its apparent weight if it were made of (a) pure gold and (b) 75% gold and 25% silver, by volume? The densities of gold, silver, and water are 19.3 g/cm^3, 10.5 g/cm^3, and 1.00 g/cm^3, respectively.
47. You're testifying in a drunk-driving case for which a blood alcohol measurement is unavailable. The accused weighs 140 lb, and would be legally impaired after consuming 36 oz of beer. The accused was observed at a beach party where a keg with interior diameter 40 cm was floating in the lake to keep it cool. After the accused's drinking stint, the keg floated 1.2 cm higher than before. Beer's density is essentially that of water. Does your testimony help or hurt the accused's case?
48. A glass beaker measures 10 cm high by 4.0 cm in diameter. Empty, it floats in water with one-third of its height submerged. How many 15-g rocks can be placed in the beaker before it sinks?
49. A typical supertanker has mass 2.0×10^6 kg and carries twice that much oil. If 9.0 m of the ship is submerged when it's empty, what's the minimum water depth needed for it to navigate when full? Assume the sides of the ship are vertical.
50. A balloon contains gas of density ρ_g and is to lift a mass M, including the balloon but not the gas. Show that the minimum mass of gas required is $m_g = M\rho_g/(\rho_a - \rho_g)$, where ρ_a is the atmospheric density.
51. (a) How much helium (density 0.18 kg/m^3) is needed to lift a balloon carrying two people, if the total mass of people, basket, and balloon (but not gas) is 280 kg? (b) Repeat for a hot-air balloon whose air density is 10% less than that of the surrounding atmosphere.
52. A 55-kg swimmer climbs onto a Styrofoam block of density 160 kg/m^3. If the water level comes right to the top of the Styrofoam, what's the block's volume?
53. BIO If the blood pressure in the unobstructed artery of Exercise 37 is 16 kPa gauge (about 120 mm of mercury, the unit commonly reported by doctors), what will it be at the clot? (*Note:* Blood's density is 1.06 g/cm^3.)
54. You're a consultant for maple syrup producers. They tap maple trees and collect sap with plastic tubing that connects to a common pipe delivering sap to an evaporator. There it's boiled to produce thick, tasty syrup. The system can be modeled as a pipe with one end, of cross-sectional area A, exposed to atmospheric pressure. The pipe drops through a vertical distance h_1 while its area decreases to $A/2$, as shown in Fig. 15.25. A small vertical glass tube extends from the lower portion, as shown, and is open to atmospheric pressure. You're asked to provide a formula for the volume flow rate of the sap as a function of the height h_2 of sap in the tube.

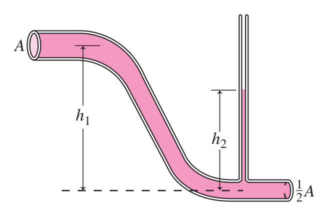


FIGURE 15.25 Problem 54

55. The water in a garden hose is at 140 kPa gauge pressure and is moving at negligible speed. The hose terminates in a sprinkler consisting of many small holes. Find the maximum height reached by the water emerging from the holes.
56. The venturi flowmeter shown in Fig. 15.26 is used to measure the flow rate of water in a solar collector system. The flowmeter is

inserted in a pipe with diameter 1.9 cm; at the venturi the diameter is 0.64 cm. The manometer tube contains oil with density 0.82 times that of water. If the difference in oil levels on the two sides of the manometer tube is 1.4 cm, what's the volume flow rate?

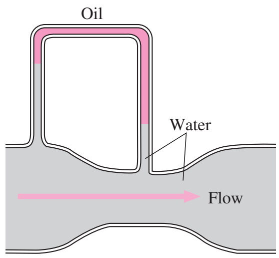


FIGURE 15.26 Problem 56

57. A 1.0-cm-diameter venturi flowmeter is inserted in a 2.0-cm-diameter pipe carrying water (density 1000 kg/m^3). Find (a) the flow speed in the pipe and (b) the volume flow rate if the pressure difference between venturi and unconstricted pipe is 17 kPa.
58. A spherical rubber balloon with mass 0.85 g and diameter 30 cm is filled with helium (density 0.18 kg/m^3). How many 1.0-g paper clips can you hang from the balloon before it loses buoyancy?
59. **BIO** Blood with density 1.06 g/cm^3 and 10-kPa gauge pressure flows through an artery at 30 cm/s. It encounters a plaque deposit where the pressure drops by 5%. What fraction of the artery's area is obstructed?
60. A venturi flowmeter in an oil pipeline has radius half that of the pipe. Oil flows in the unconstricted pipe at 1.9 m/s. If the pressure difference between unconstricted flow and venturi is 16 kPa, what's the oil's density?
61. A drinking straw 20 cm long and 3.0 mm in diameter stands vertically in a cup of juice 8.0 cm in diameter. A section of straw 6.5 cm long extends above the juice. A child sucks on the straw, and the juice level begins dropping at 2.0 mm/s. (a) By how much does the pressure in the child's mouth differ from atmospheric pressure? (b) What's the greatest height above the water surface from which the child could drink, assuming this same mouth pressure?
62. Water emerges from a faucet of diameter d_0 in steady, near-vertical flow with speed v_0. Show that the diameter of the falling water column is given by $d = d_0[v_0^2/(v_0^2 + 2gh)]^{1/4}$, where h is the distance below the faucet (Fig. 15.27).

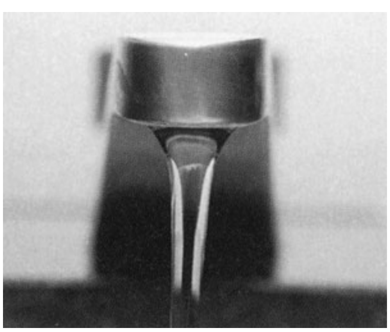


FIGURE 15.27 Problem 62

63. Assuming normal atmospheric pressure, how massive an object can a 5.0-cm-diameter suction cup support on a vertical wall, if the coefficient of friction between cup and wall is 0.72?
64. Figure 15.28 shows a simplified diagram of a Pitot tube, used for measuring aircraft speeds. The tube is mounted on the aircraft with opening A at right angles to the flow and opening B pointing into the flow. The gauge prevents airflow through the tube. Use Bernoulli's equation to show that the plane's speed relative to the air is $v = \sqrt{2\,\Delta p/\rho}$, where Δp is the pressure difference between the tubes and ρ is the density of air. (*Hint:* The flow must be stopped at B, but continues past A with its normal speed.)

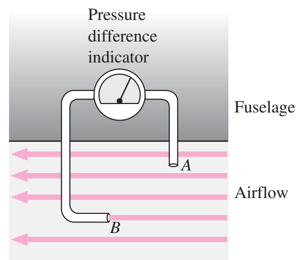


FIGURE 15.28 Problem 64

65. At a hearing on a proposed wind farm, a wind-energy advocate says an installation of 800 turbines, with blade diameter 95 m, could displace a 1-GW nuclear power plant. You're asked if that's really possible. How do you answer, given an average wind speed of 12 m/s and a turbine power output that averages 30% of the theoretical maximum?
66. A pencil is weighted so it floats vertically with length L submerged. It's pushed vertically downward without being totally submerged, then released. Show that it undergoes simple harmonic motion with period $T = 2\pi\sqrt{L/g}$.
67. A can of height h and cross-sectional area A_0 is initially full of water. A small hole of area $A_1 \ll A_0$ is cut in the bottom of the can. Find an expression for the time it takes all the water to drain from the can. (*Hint*: Call the water depth y, use the continuity equation, and integrate.)
68. Density and pressure in Earth's atmosphere are proportional: $\rho = p/h_0 g$, where $h_0 = 8.2$ km is a constant and g is the gravitational acceleration. (a) Integrate Equation 15.2 for this case to show that atmospheric pressure as a function of height h above the surface is given by $p = p_0 e^{-h/h_0}$, where p_0 is the surface pressure. (b) At what height will the pressure have dropped to half its surface value?
69. (a) Use the result of Problem 68 to express Earth's atmospheric density as a function of height. (b) Use your result from (a) to find the height below which half of Earth's atmospheric mass lies (this will require integration).
70. A circular pan of liquid with density ρ is centered on a horizontal turntable rotating with angular speed ω, as shown in Fig. 15.29. Atmospheric pressure is p_a. Find expressions for (a) the pressure at the bottom of the pan and (b) the height of the liquid surface as functions of the distance r from the axis, given that the height at the center is h_0.

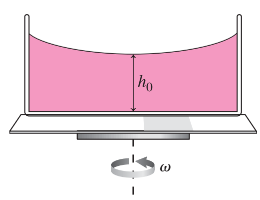


FIGURE 15.29 Problem 70

71. A solid sphere of radius R and mass M has density ρ that varies with distance r from the center: $\rho = \rho_0 e^{r/R-1}$. Find an expression for the central density ρ_0.
72. The difference in air pressure between the inside and outside of a ball is a constant Δp. Show by direct integration that the net pressure force on one hemisphere is $\pi R^2 \Delta p$, with R the ball's radius.
73. Find the torque that the water exerts about the bottom edge of the dam in Problem 42.
74. One vertical wall of an above-ground swimming pool is a regular trapezoid, with one base 10 m long on level ground and the other 20 m long at a height of 3 m above it. If the pool is filled to the top with water, what's the net fluid force on the wall? (*Hint*: Consider both the force exerted by the water on one side of the wall and the force exerted by the atmosphere on the other.)
75. You're a private investigator assisting a large food manufacturer in tracking down counterfeit salad dressing. The genuine dressing is by volume one part vinegar (density 1.0 g/cm^3) to three parts olive oil (density 0.92 g/cm^3). The counterfeit dressing is diluted with water (density 1.0 g/cm^3). You measure the density of a dressing sample and find it to be 0.97 g/cm^3. Has the dressing been altered?
76. A plumber comes to your ancient apartment building where you have a part-time job as caretaker. He's checking the hot-water heating system, and notes that the pressure in the basement is 18 psi. He asks, "How high is the building?" "Three stories, each about 11 feet," you reply. "OK, about 33 feet," he says, pausing to do some calculations in his head. "The pressure is fine," he declares. On what basis did he come to that conclusion?
77. Your class in naval architecture is working on the design for a ship with a V-shaped cross section, as shown in Fig. 15.30. The ship has total length L and keel-to-deck height h_0. When empty, the distance from water line to keel is h_1. You're asked for the maximum load the ship can carry below deck if water is not to come over the deck. Answer in terms of h_0, h_1, L, θ, and the water density ρ.

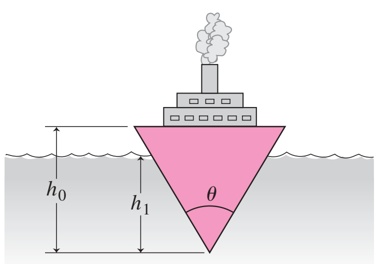


FIGURE 15.30 Problem 77

Passage Problems

Arterial stenosis is a constriction of an artery, often due to plaque buildup on the artery's inner walls. Serious medical conditions can result, depending on the affected artery. Stenosis of the carotid arteries that supply blood to the brain is a leading cause of stroke, while stenosis of the renal arteries can lead to kidney failure. Pulmonary artery stenosis results from birth defects, and can result in insufficient oxygen supply. Because the heart has to work harder to get blood through a constricted artery, stenosis can contribute to high blood pressure.

In answering the questions below, assume steady flow (which is true in arteries only on short timescales).

78. How does the volume flow rate of blood at a stenosis compare with the rate in the surrounding artery?
    a. lower
    b. the same
    c. higher
79. How does the blood flow speed at a stenosis compare with the speed in the surrounding artery?
    a. lower
    b. the same
    c. higher
80. Which of the following medical problems is more likely to occur?
    a. An artery might collapse because of lower blood pressure at the stenosis.
    b. An artery might burst because of higher blood pressure at the stenosis.
    c. Neither; pressure at the stenosis is the same as in the surrounding artery.
81. If the artery has circular cross section even at the stenosis, but the diameter at the stenosis is half that in the surrounding artery, the blood flow speed in the stenosis will be
    a. one-fourth that in the surrounding artery.
    b. one-half that in the surrounding artery.
    c. the same as in the surrounding artery.
    d. $\sqrt{2}$ times that in the surrounding artery.
    e. four times that in the surrounding artery.

Answers to Chapter Questions

Answer to Chapter Opening Question

Because the density of ice is only slightly less than that of water.

Answers to GOT IT? Questions

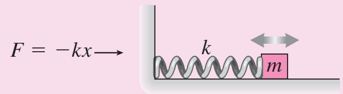
15.1. (a), over the top, where the streamlines are closer together.
15.2. $h_1 > h_4 > h_2 > h_3$, reflecting higher pressure with lower flow speed.

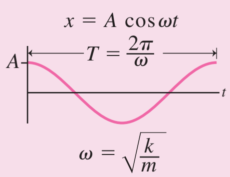
Oscillations, Waves, and Fluids

Part Two has extended Newtonian mechanics to systems that undergo oscillatory motion and wave motion or that involve the motion of fluids. Behind these more complex motions are the fundamental concepts of force, mass, and energy and their roles in characterizing motion.

Oscillatory motion describes the back-and-forth motion of a system disturbed from a stable equilibrium.

When the force or torque tending to restore equilibrium is directly proportional to the displacement, the result is simple harmonic motion.





A **wave** is a propagating disturbance that carries energy but not matter.

Simple harmonic waves are sinusoidal:

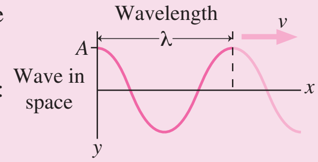
$$y(x,t) = A\cos(kx - \omega t)$$

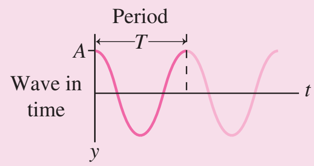
Angular frequency: $\omega = 2\pi f$

Wave number: $k = \frac{2\pi}{\lambda}$

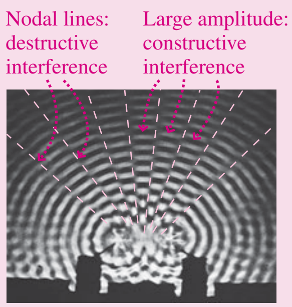
Wave period: $T = \frac{1}{f}$

Wave speed: $v = \frac{\omega}{k} = \frac{\lambda}{T} = f\lambda$

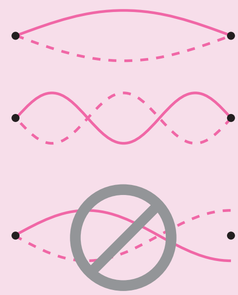




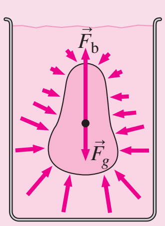
When waves overlap, the result is **interference**, which is constructive when the waves reinforce and destructive when they tend to cancel.



Standing waves occur when the medium has limited extent. Only certain wavelengths and frequencies are allowed, depending on the medium's length:



Fluids in **hydrostatic equilibrium** exhibit a depth-dependent pressure that results in an upward buoyancy force $\vec{F}_b$.



Archimedes' principle states that the buoyancy force is equal to the weight of the displaced fluid.

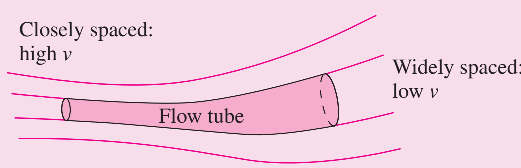
Moving fluids obey conservation of mass and, in the absence of fluid friction (viscosity), they also conserve energy.

In **fluid dynamics**, the continuity equation and Bernoulli's equation express these conservation laws. Both equations hold along a flow tube:

Continuity: $\rho vA = \text{constant}$

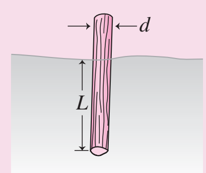
Bernoulli:

$$p + 1/2\rho v^2 + \rho gy = \text{constant}$$



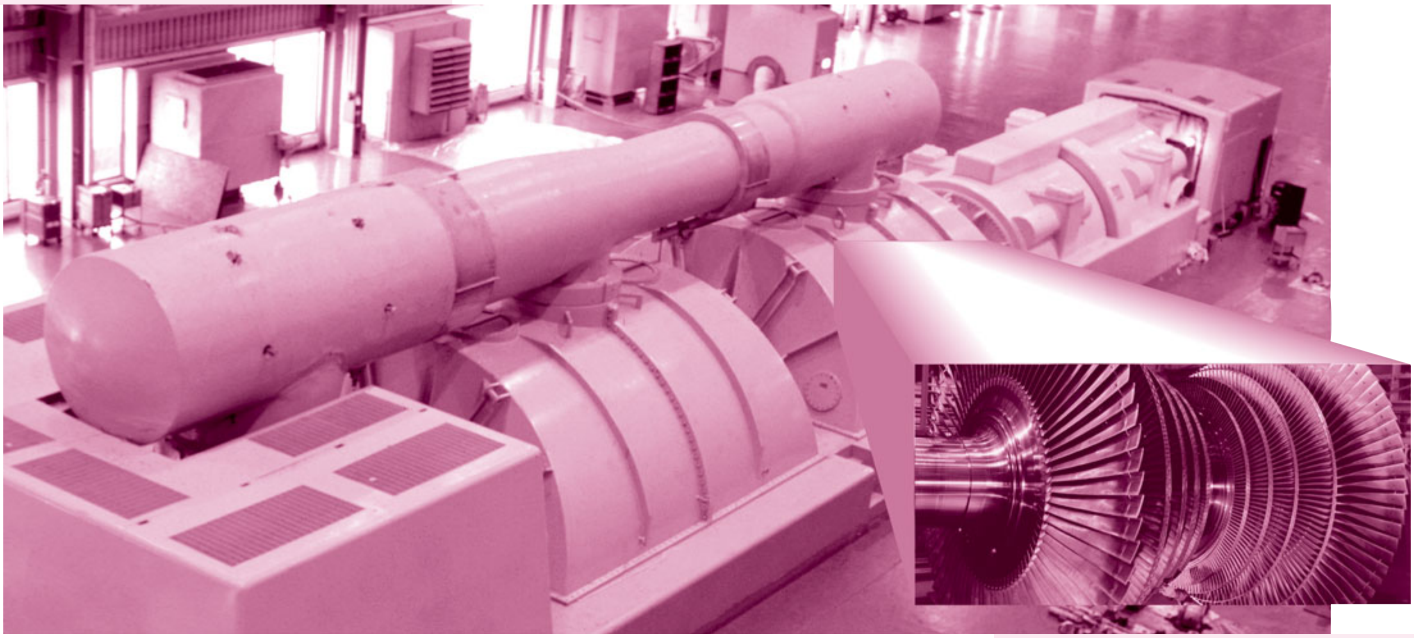
Part 2 Challenge Problem

A cylindrical log of total mass M and uniform diameter d has an uneven mass distribution that causes it to float in a vertical position, as shown in the figure. (a) Find an expression for the length L of the submerged portion of the log when it's floating in equilibrium, in terms of M, d, and the water density ρ. (b) If the log is displaced vertically from its equilibrium position and released, it will undergo simple harmonic motion. Find an expression for the period of this motion, neglecting viscosity and other frictional effects.



PART THREE OVERVIEW

Thermodynamics



This huge steam turbine converts the energy of high-pressure steam to mechanical energy and then electricity. Systems like this one produce nearly all the world's electrical energy, and their operation and efficiency are governed by the laws of thermodynamics.

Humanity consumes energy at the prodigious rate of some 10^{13} watts. Nearly all that energy comes from the combustion of fossil fuels—a process governed by the laws of thermodynamics. Engines that extract mechanical energy from the heat of burning fuels propel our cars, trucks, and airplanes, and produce most of our electricity. Despite the efforts of the cleverest engineers, the laws of thermodynamics set fundamental limitations on our ability to convert heat to mechanical energy. Many of the energy and environmental challenges humanity faces today are grounded in thermodynamics.

Many natural systems are also thermodynamic. Without the Sun's energy, radiated across a hundred million miles of empty space, Earth would be a lifeless, frozen rock. Heat flowing throughout Earth, its oceans, and its atmosphere governs processes ranging from continental drift to ocean currents to weather and climate. Concern over human-induced climate change is rooted in thermodynamic properties of the atmosphere as they affect thermal energy flows. On a grander scale, thermodynamic principles govern much of the energy that flows throughout the universe.

Thermodynamics—the study of heat and its connection to the all-important concept of energy—is the subject of the next four chapters.

16 Temperature and Heat

How does this infrared photo reveal heat loss from the house? And how can you tell that the car was recently driven?

New Concepts, New Skills

By the end of this chapter you should be able to

- Explain the meanings of temperature and heat and the difference between them (16.1).
- Work with different temperature scales (16.1).
- Calculate the equilibrium temperature of a mixture (16.2).
- Evaluate heat loss from buildings and other systems (16.3).
- Calculate radiant-heat transfer (16.3).
- Determine a system's temperature when it's in thermal-energy balance (16.4).

Connecting Your Knowledge

- The concept of energy is the ultimate basis for understanding heat and temperature (6.3, 7.1, 7.3).
- You should be comfortable expressing energies in different units (6.3).

Your own body gives you a good sense of "hot" and "cold." Questions about heat and temperature are ultimately about energy, and these concepts are crucial to understanding the energy flows that drive natural systems like Earth's climate and technologies such as engines, power plants, and refrigerators.

Properties like mass and kinetic energy apply equally to microscopic atoms and molecules and to cars and planets. But other properties, including temperature and pressure, apply only to macroscopic systems. It makes no sense to talk about the temperature or pressure of a single air molecule. **Thermodynamics** is the branch of physics that deals with these macroscopic properties. Ultimately, the thermodynamic behavior of matter follows from the motions of its constituent particles in response to the laws of mechanics. **Statistical mechanics** relates the macroscopic description of matter to the underlying microscopic processes. Historically, thermodynamics developed before the atomic theory of matter was fully established. The subsequent explanation of thermodynamics through statistical mechanics—the mechanics of atoms and molecules—was a triumph for physics.

16.1 Heat, Temperature, and Thermodynamic Equilibrium

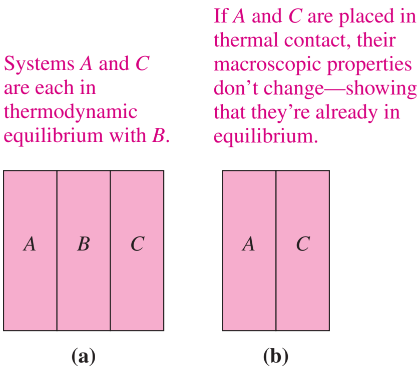


FIGURE 16.1 The zeroth law of thermodynamics.

Take a bottle of soda from the refrigerator, and eventually it reaches room temperature. At that point soda and room are in **thermodynamic equilibrium**, a state in which their macroscopic properties are no longer changing. To check for thermodynamic equilibrium we can consider any macroscopic property—length, volume, pressure, electrical resistance, whatever. If any macroscopic property changes when two systems are placed together, then they weren't originally in thermodynamic equilibrium. When changes cease, the systems have reached equilibrium.

The phrase "placed together" here has a definite meaning, stated more precisely as "placed in thermal contact." Two systems are in **thermal contact** if heating one of them results in macroscopic changes in the other. If that doesn't readily happen—for example, with a Styrofoam cup of coffee and its surroundings—then the systems are **thermally insulated**.

We're now ready to define temperature: **Two systems have the same temperature if they are in thermodynamic equilibrium**. Consider two systems A and C in thermal contact with a third system B but not with each other (Fig. 16.1a). Even though they're not in direct contact, A and C have the same temperature; that is, if you place A and C in thermal contact (Fig. 16.1b), no further changes occur. This fact—that two systems in equilibrium with a third system are therefore in equilibrium with each other—is so fundamental that it's called the **zeroth law of thermodynamics**.

A **thermometer** is a system with a conveniently observed macroscopic property that changes with temperature. It could be the length of a mercury column, gas pressure, electrical resistance, or the bending of a bimetal strip in a dial thermometer. Let the thermometer come to equilibrium with some system, and its temperature-dependent physical property provides a measure of temperature. The zeroth law assures consistency, in that two systems for which the thermometer gives the same reading must have the same temperature.

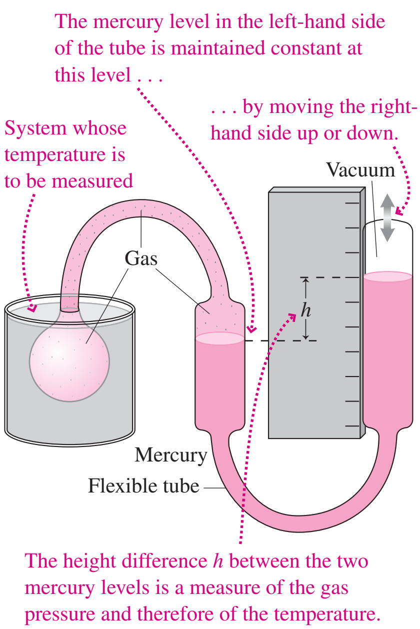


FIGURE 16.2 A constant-volume gas thermometer.

Gas Thermometers and the Kelvin Scale

Figure 16.2 shows a **constant-volume gas thermometer**, whose temperature indication is the pressure of gas held at constant volume. We define zero temperature as that temperature at which the gas pressure would become zero. A second known point is the so-called **triple point** of water—the unique temperature at which solid, liquid, and gaseous water can coexist in equilibrium (more on this in the next chapter). The SI temperature unit, the **kelvin** (symbol K; *not* "degrees kelvin" or °K), is defined by setting the triple point at 273.16 K. Other temperatures then follow from a linear relationship as shown in Fig. 16.3.

Since a gas can't have negative pressure, the zero of the kelvin scale is an absolute lower limit on temperature and is called **absolute zero**. We'll explore the meaning of absolute zero further in Chapter 19.

Gas thermometers are useful because they work over a wide temperature range. More important, as we'll see in the next chapter, all gases behave in essentially the same way in the limit of low pressure, so gas thermometry provides a reproducible temperature standard.

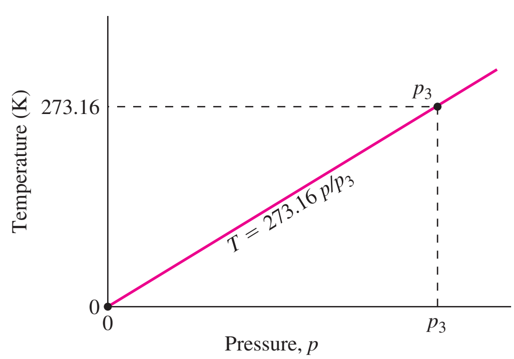


FIGURE 16.3 The Kelvin temperature scale defined using a gas thermometer; p_3 is the gas pressure at the triple point of water.

Temperature Scales

Other temperature scales include Celsius (°C), Fahrenheit (°F), and Rankine (°R) (Fig. 16.4). One Celsius degree represents the same temperature difference as one kelvin, but the zero of the Celsius scale occurs at 273.15 K, so

$$T_C = T - 273.15 \tag{16.1}$$

where T is the temperature in kelvins. On the Celsius scale the melting point of ice at standard atmospheric pressure is exactly 0°C, while the boiling point is 100°C. The triple point of water occurs at 0.01°C, which accounts for the 273.15 difference between the kelvin and Celsius scales. Equation 16.1 shows that absolute zero occurs at −273.15°C.

The Fahrenheit and Rankine scales, from the British unit system, are used primarily in the United States. Fahrenheit has water melting at 32°F and boiling at 212°F, so the relation between Fahrenheit and Celsius temperatures is

$$T_F = \tfrac{9}{5}T_C + 32 \tag{16.2}$$

A Rankine degree is the same size as a Fahrenheit degree, but the zero of the Rankine scale is at absolute zero (Fig. 16.4). Engineers in the United States often use Rankine.

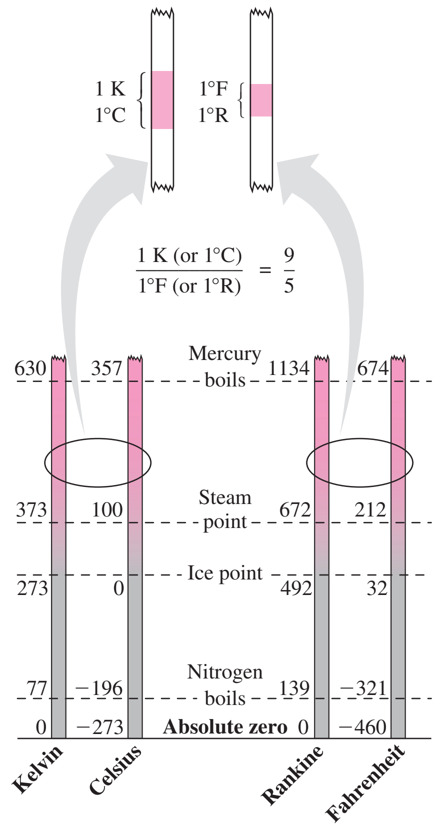


FIGURE 16.4 Relationships among four temperature scales.

Heat and Temperature

A match will burn your finger, but it doesn't provide much heat. This example shows our intuitive sense of temperature and heat: Heat measures an *amount* of "something," whereas temperature is the *intensity* of that "something."

Scientists once considered heat to be a material fluid, called **caloric**, that flowed from hot bodies to colder ones. But in the late 1700s, the American-born scientist Benjamin Thompson observed essentially limitless amounts of heat being produced in the boring of cannon, and he concluded that heat could not be a conserved fluid. Instead, Thompson suggested, heat was associated with mechanical work done by the boring tool. In the next half-century, a series of experiments confirmed the association between heat and energy. These culminated in the work of the British physicist James Joule (1818–1889), who quantified the relation between heat and energy. In so doing, Joule brought thermal phenomena under the powerful conservation-of-energy principle. In recognition of this major synthesis in physics, the SI energy unit bears Joule's name.

We rarely make statements about the amount of "heat" in an object; we're more concerned that the temperature be appropriate. Rather, we think of heat as something that gets transferred from one object to another, causing a temperature change. The scientific definition reflects this sense of heat as energy in transit: **Heat is energy being transferred from one object to another because of a temperature difference alone**. Strictly speaking, **heat** refers only to energy in transit. Following heat transfer, we say that the **internal energy** of the object has increased, not that it contains more heat. This distinction reflects the fact that processes other than heating—such as transfer of mechanical or electrical energy—can also change an object's temperature.

16.2 Heat Capacity and Specific Heat

Experimentally, we find that the heat Q transferred to an object and the resulting temperature change ΔT are proportional: $Q = C\,\Delta T$, where C is the **heat capacity** of the object. Since heat is a measure of energy transfer, the units of heat capacity are J/K. The heat capacity C applies to a specific object and depends on its mass and on the substance from which it's made. We characterize different substances in terms of their **specific heat** c, or heat capacity per unit mass. The heat capacity of an object is then the product of its mass and specific heat, so we can write

$$Q = mc\,\Delta T \tag{16.3}$$

The SI units of specific heat are J/kg·K. Table 16.1 lists specific heats of common materials.

Table 16.1 Specific Heats of Some Common Materials*

	Specific Heat, c	
Substance	**SI Units: J/kg·K**	**cal/g·°C, kcal/kg·°C, or Btu/lb·°F**
Aluminum	900	0.215
Concrete (varies with mix)	880	0.21
Copper	386	0.0923
Iron	447	0.107
Glass	753	0.18
Mercury	140	0.033
Steel	502	0.12
Stone (granite)	840	0.20
Water:		
Liquid	4184	1.00
Ice, −10°C	2050	0.49
Wood	1400	0.33

*Temperature range 0°C to 100°C except as noted.

Scientists first studied thermodynamic phenomena before they knew the relation between heat and energy, and they used other units for heat. The **calorie** (cal) was defined as the heat needed to raise the temperature of 1 g of water from 14.5°C to 15.5°C; consequently, the specific heat of water is 1 cal/g·°C. Several different definitions of the calorie exist today, based on different methods for establishing the heat-energy equivalence. In this book we use the so-called thermochemical calorie, defined as exactly 4.184 J. The "calorie" used in describing the energy content of foods is actually a kilocalorie. In the British system, still widely used in engineering in the United States, the unit of heat is the **British thermal unit** (Btu). One Btu is the amount of heat needed to raise the temperature of 1 lb of water from 63°F to 64°F, and is equal to 1054 J.

EXAMPLE 16.1 Specific Heat: Waiting to Shower

Your whole family has showered before you, dropping the temperature in the water heater to 18°C. If the heater holds 150 kg of water, how much energy will it take to bring it up to 50°C? If the energy is supplied by a 5.0-kW electric heating element, how long will that take?

INTERPRET Here we're interested in the energy it takes to raise the water temperature, so we interpret this problem as involving specific heat. For the second part, we're given the heater's power output and asked for the time, so we need to recall (Chapter 6) that power is energy per time.

DEVELOP Equation 16.3, $Q = mc\,\Delta T$, relates energy and temperature change via specific heat, so our plan is to calculate the required energy from this equation. We'll then use the relation between power and energy to find the time.

EVALUATE Equation 16.3 gives the energy:

$$Q = mc\,\Delta T = (150\text{ kg})(4184\text{ J/kg}\cdot\text{K})(50°\text{C} - 18°\text{C}) = 20\text{ MJ}$$

where we found the specific heat of water in Table 16.1. The heating element supplies energy at the rate of 5.0 kW or 5.0×10^3 J/s. At that rate the time needed to supply 20 MJ is

$$\Delta t = \frac{2.0\times10^7\text{ J}}{5.0\times10^3\text{ J/s}} = 4000\text{ s}$$

or a little over an hour.

ASSESS That's a long time to wait, but it's not an unreasonable answer!

✓TIP Is That C or K?

It doesn't matter when we're talking about temperature *differences*. That's why we could mix units, multiplying the specific heat in J/kg·K by the difference of Celsius temperatures.

■

Heat capacity and specific heat vary slightly with temperature, and they also depend on whether an object's pressure or its volume changes as it's heated. For solids and liquids, which don't expand much, that distinction isn't very important. But it makes a big difference whether a gas is confined or allowed to expand when heated. Consequently, gases have two different specific heats, depending on whether volume or pressure is constant. We'll deal with that issue in Chapter 18, where we explore the thermodynamic behavior of gases.

The Equilibrium Temperature

When objects at different temperatures are in thermal contact, heat flows from the hotter object to the cooler one until they reach thermodynamic equilibrium. If the objects are thermally insulated from their surroundings, then all the energy leaving the hotter object ends up in the cooler one. Mathematically, this statement reads

$$m_1 c_1 \, \Delta T_1 + m_2 c_2 \, \Delta T_2 = 0 \qquad (16.4)$$

For the hotter object, ΔT is negative, so the two terms in Equation 16.4 have opposite signs. One term represents the outflow of heat from the hotter object, the other inflow into the cooler one.

GOT IT? 16.1 A hot rock with mass 250 g is dropped into an equal mass of cool water. Which temperature changes more, that of the rock or the water? Explain.

EXAMPLE 16.2 Finding the Equilibrium Temperature: Cooling Down

An aluminum frying pan of mass 1.5 kg is at 180°C when it's plunged into a sink containing 8.0 kg of water at 20°C. Assuming that none of the water boils and that no heat is lost to the surroundings, find the equilibrium temperature of the water and pan.

INTERPRET Here we have two objects, initially at different temperatures, that come to thermal equilibrium. So this is a problem about the equilibrium temperature, with the system of interest comprising the pan and the water.

DEVELOP Equation 16.4, $m_1 c_1 \, \Delta T_1 + m_2 c_2 \, \Delta T_2 = 0$, applies. However, we're asked for the common equilibrium temperature T, so we write the temperature differences ΔT in terms of T and the initial temperatures T_p and T_w of pan and water. Equation 16.4 then becomes $m_p c_p (T - T_p) + m_w c_w (T - T_w) = 0$.

EVALUATE We now solve for the equilibrium temperature T:

$$T = \frac{m_p c_p T_p + m_w c_w T_w}{m_p c_p + m_w c_w}$$

Using the given values of m_p, T_p, m_w, and T_w, and taking c_p and c_w from Table 16.1, we get $T = 26°C$.

ASSESS The water has much greater mass and higher specific heat, so it makes sense that its 6°C temperature change is a lot less than the 154°C drop in the pan's temperature. ■

16.3 Heat Transfer

How is heat transferred? Engineers need to know so they can design heating and cooling systems. Scientists need to know so they can anticipate temperature changes, as in global warming. Here we'll consider three common heat-transfer mechanisms: conduction, convection, and radiation. In some situations, a single mechanism dominates; in other cases, we may need to take all three into account.

Conduction

Conduction is heat transfer through direct physical contact. It occurs as molecules in a hotter region collide with and transfer energy to those in an adjacent cooler region. **Thermal conductivity** (symbol k; SI unit W/m·K) characterizes this process. Common materials exhibit a broad range of thermal conductivities, from about 400 W/m·K for copper—a good conductor—to 0.029 W/m·K for Styrofoam, a good thermal insulator. Table 16.2 lists some thermal conductivities; they're given in both SI and British units because the latter are widely used in heat-loss calculations for buildings. The k values

Table 16.2 Thermal Conductivities*

	Thermal Conductivity, k	
Material	**SI Units: W/m·K**	**British Units: Btu·in./h·ft²·°F**
Air	0.026	0.18
Aluminum	237	1644
Concrete (varies with mix)	1	7
Copper	401	2780
Fiberglass	0.042	0.29
Glass	0.7–0.9	5–6
Goose down	0.043	0.30
Helium	0.14	0.97
Iron	80.4	558
Steel	46	319
Styrofoam	0.029	0.20
Water	0.61	4.2
Wood (pine)	0.11	0.78

*Temperature range 0°C to 100°C.

in Table 16.2 reflect physical properties of the materials. Metals, for example, are good thermal conductors because they contain free electrons that move quickly. Insulators like fiberglass and Styrofoam owe their insulating properties to a physical structure that traps small volumes of air or other gas.

Figure 16.5 shows a slab of thickness Δx and area A. One side is at temperature T and the other at $T + \Delta T$. The temperature difference ΔT drives a conductive heat flow through the slab; not surprisingly, we find that the heat flow is proportional to the temperature difference, the slab area, and the thermal conductivity k. The thicker the slab, on the other hand, the more resistance to heat flow, so the flow depends inversely on thickness. Therefore,

$$H = -kA\frac{\Delta T}{\Delta x} \quad \text{(conductive heat flow)} \tag{16.5}$$

where $H = dQ/dt$ is the rate of heat flow in watts, and where the minus sign shows that the flow is opposite the direction of increasing temperature—that is, from hotter to cooler.

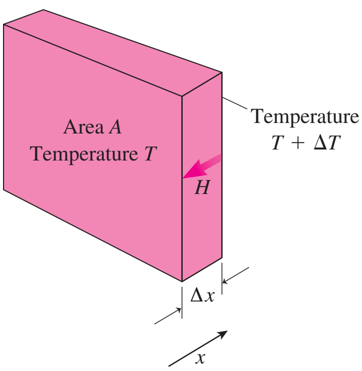


FIGURE 16.5 Heat flows from the hotter to the cooler face of the slab.

EXAMPLE 16.3 Conduction: Warming a Lake

A lake with a flat bottom and steep sides has surface area 1.5 km² and is 8.0 m deep. On a summer day, the surface water is at 30°C and the bottom water at 4.0°C. What's the rate of heat conduction through the lake?

INTERPRET This is a problem about heat conduction.

DEVELOP Our sketch, Fig. 16.6, shows that we can treat the lake like the slab shown in Fig. 16.5, provided we neglect heat flow out the sides. Then Equation 16.5, $H = -kA(\Delta T/\Delta x)$, will give the heat-flow rate.

EVALUATE Substituting numerical values, including water's thermal conductivity from Table 16.2, we get

$$\begin{aligned} H &= -kA\frac{\Delta T}{\Delta x} \\ &= -(0.61\ \text{W/m}\cdot\text{K})(1.5\times 10^6\ \text{m}^2)\frac{30°\text{C} - 4.0°\text{C}}{8.0\ \text{m}} = -3.0\ \text{MW} \end{aligned}$$

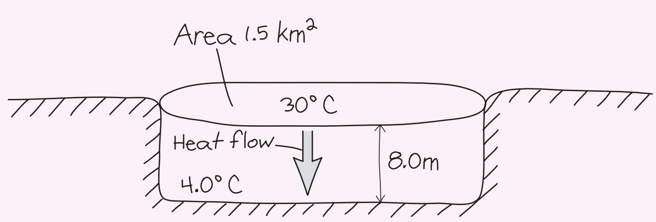


FIGURE 16.6 Our sketch for Example 16.3.

ASSESS This is a significant energy flow, but with direct sunlight averaging about 1 kW on every square meter, the lake's 1.5-km² surface area absorbs plenty of solar energy, and that's what maintains the temperature difference that drives the conductive heat flow. Figure 16.5 shows x increasing in the direction of increasing temperature, so the negative sign in our answer indicates that the flow is downward. ■

Equation 16.5 is strictly correct only when the temperature varies uniformly from one surface to the other. That's the case when two surfaces at different temperatures have the same area. With other geometries—as in the insulation surrounding a cylindrical pipe—we need to write $\Delta T/\Delta x$ as the derivative dT/dx and integrate to find the heat flow. Problems 70 and 75 explore this situation.

Often heat flows through several different materials. A building wall, for example, may contain wood, plaster, and fiberglass insulation. Figure 16.7 shows such a composite structure, with temperature T_1 on one side and T_3 on the other. The heat-flow rate H must be the same through both slabs so energy doesn't accumulate at the interface between the two. Then Equation 16.5 gives

$$H = -k_1 A \frac{T_2 - T_1}{\Delta x_1} = -k_2 A \frac{T_3 - T_2}{\Delta x_2}$$

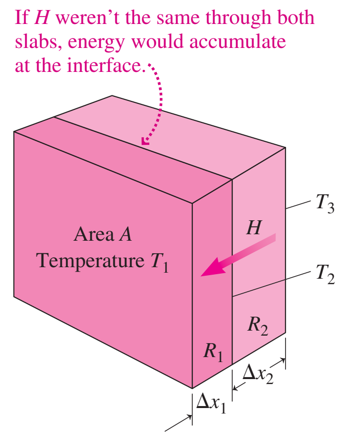


FIGURE 16.7 A composite slab.

where k_1 and k_2 are the thermal conductivities of the two materials, and T_2 is the temperature at the interface. We can express the heat-flow rate in terms of the surface temperatures T_1 and T_3 alone if we define the **thermal resistance** R of each slab:

$$R = \frac{\Delta x}{kA} \tag{16.6}$$

The SI units of R are K/W. Unlike the thermal conductivity k, which is a property of a *material*, R is a property of a *particular piece* of material, reflecting both its conductivity and its geometry. In terms of thermal resistance, our heat-flow equation becomes

$$H = -\frac{T_2 - T_1}{R_1} = -\frac{T_3 - T_2}{R_2}$$

so $R_1 H = T_1 - T_2$ and $R_2 H = T_2 - T_3$. Adding these two equations gives

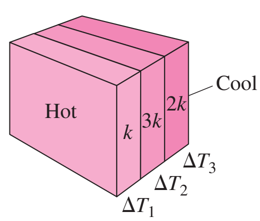
$$(R_1 + R_2)H = T_1 - T_2 + T_2 - T_3 = T_1 - T_3$$

or

$$H = \frac{T_1 - T_3}{R_1 + R_2} \tag{16.7}$$

Equation 16.7 shows that the composite slab acts like a single slab whose thermal resistance is the sum of the resistances of the two slabs that compose it. We could easily extend this treatment to show that the thermal resistances of three or more slabs add when the slabs are arranged so the same heat flows through all of them.

GOT IT? 16.2 The figure shows three slabs with the same thickness but different thermal conductivities: k, $3k$, and $2k$; the left side is hotter, as shown. Rank in order the three temperature differences ΔT.



Insulating properties of building materials are described by the $\mathcal{R}$**-factor**, which is the thermal resistance for a slab of unit area:

$$\mathcal{R} = RA = \frac{\Delta x}{k} \tag{16.8}$$

The SI units of $\mathcal{R}$ are $m^2 \cdot K/W$, and that's how you'll find it listed if you buy insulation in Europe or other SI-based regions. In the United States, $\mathcal{R}$ is in $ft^2 \cdot °F \cdot h/Btu$, although the units are almost never stated. This means that $\mathcal{R}$-19 fiberglass insulation loses $\frac{1}{19}$ Btu per hour for each square foot of insulation for each degree Fahrenheit temperature difference across the insulation (Fig. 16.8). The inverse of the $\mathcal{R}$-factor is the U value, often used in characterizing heat loss through windows.

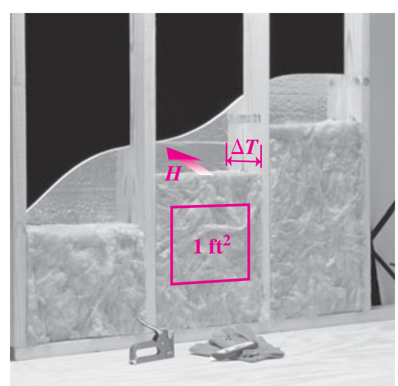


FIGURE 16.8 Each square foot of this $\mathcal{R}$-19 fiberglass insulation loses $\frac{1}{19}$ Btu per hour for every °F of temperature difference ΔT.

EXAMPLE 16.4 Calculating Heat Loss: The Cost of Oil

Figure 16.9 shows a house whose walls consist of plaster $(\mathcal{R} = 0.17)$, $\mathcal{R}$-11 fiberglass insulation, plywood $(\mathcal{R} = 0.65)$, and cedar shingles $(\mathcal{R} = 0.55)$. The roof has the same construction except it uses $\mathcal{R}$-30 fiberglass insulation. The average outdoor temperature in winter is 20°F, and the house is maintained at 70°F. The house's oil furnace produces 100,000 Btu for every gallon of oil, and oil costs \$2.20 per gallon. How much does it cost to heat the house for a month?

INTERPRET Although the problem asks for the monthly cost of oil, this isn't economics! We interpret this as a problem about heat loss and identify the walls and roof as systems for which we need to know the heat flow. This is a rare case of a problem stated in English units.

DEVELOP We're given the drawing in Fig. 16.9. We have the $\mathcal{R}$-factors; in English units, their inverses give the heat-loss rate on a square-foot basis. So our plan is to find the square footage of the walls and roof separately, calculate the total heat-loss rate, and then find the amount and cost of oil to compensate for a month's heat loss.

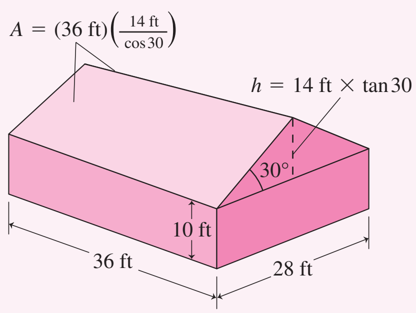


FIGURE 16.9 House for Example 16.4.

EVALUATE The $\mathcal{R}$-factors for the wall materials sum to give $\mathcal{R}_{\text{wall}} = 12.4$; similarly, $\mathcal{R}_{\text{roof}} = 31.4$. The perimeter of the house measures $2\times28\text{ ft} + 2\times36\text{ ft} = 128\text{ ft}$, so the 10-ft vertical walls have area 1280 ft^2. There are also the triangular gables. Since there are two of them, each with area $\frac{1}{2}bh$, they give another bh or $(28\text{ ft})(14\text{ ft}\tan 30°) = 226\text{ ft}^2$, so $A_{\text{wall}} = 1506\text{ ft}^2$. These $\mathcal{R}$-12.4 walls lose 1/12.4 Btu/h/ft^2/°F. With 1506 ft^2 and a temperature difference of 50°F, the total heat-loss rate through the walls is

$$H_{\text{wall}} = \left(\tfrac{1}{12.4}\text{Btu/h/ft}^2/°\text{F}\right)(1506\text{ ft}^2)(50°\text{F}) = 6073\text{ Btu/h}$$

The area of the pitched roof is larger than that of a flat roof by the factor 1/cos 30°, so the heat-loss rate through the roof is

$$H_{\text{roof}} = \left(\tfrac{1}{31.4}\text{Btu/h/ft}^2/°\text{F}\right)\frac{(36\text{ ft})(28\text{ ft})}{\cos 30°}(50°\text{F}) = 1853\text{ Btu/h}$$

The total heat-loss rate is then 7926 Btu/h. In a month, this results in a heat loss of $Q = (7926\text{ Btu/h})(30\text{ days/month})(24\text{ h/day}) = 5.7\text{ MBtu}$.

Now for the oil: With 10^5 Btu (0.1 MBtu) per gallon, we'll burn 57 gallons per month to produce that 5.7 MBtu. At \$2.20/gal, that will cost \$126.

ASSESS If you've paid for heat in a northern climate, you know that this figure is, if anything, low. That's because we neglected heat losses through windows, doors, and the floor, as well as cold-air infiltration. On the other hand, we also left out any solar energy gained through the windows on sunny days. Problem 67 provides a more realistic look at this house. ■

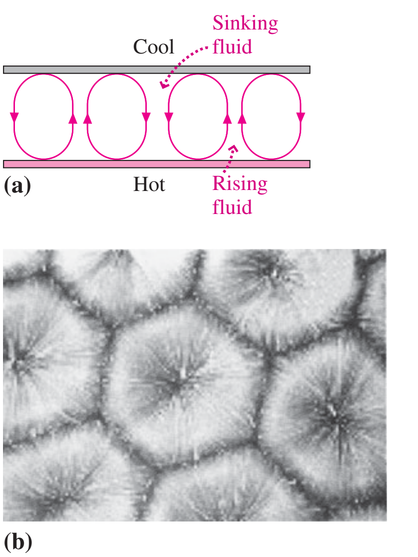


FIGURE 16.10 (a) Convection between two plates at different temperatures. (b) Top view of convection cells in a laboratory experiment. Fluid rises at the centers and sinks at the edges of the hexagonal cells.

Convection

Convection is heat transfer by fluid motion. It occurs as heated fluid becomes less dense and therefore rises. Figure 16.10*a* shows two plates at different temperatures, with fluid between them. Fluid heated by the lower plate rises and transfers heat to the upper plate. The cooled fluid sinks, and the process repeats. The pattern of rising and sinking fluid often acquires a striking regularity, as shown in Fig. 16.10*b*.

Convection is important in many technological and natural environments. When you heat water on a stove, convection carries heat through the water. Houses usually rely on convection from heat sources near floor level to circulate warm air throughout a room. Insulating materials trap air and thereby inhibit convection that would otherwise cause excessive heat loss. Convection associated with solar heating of Earth's surface drives the vast air movements that establish our overall climate. Violent convection, as in thunderstorms, is associated with localized temperature differences. On a much longer time scale, convection in Earth's mantle drives continental drift. Convection plays a crucial role in many astrophysical processes, including the generation of magnetic fields in stars and planets.

As with conduction, the convective heat-loss rate often is approximately proportional to the temperature difference. But the calculation of convective heat loss is complicated because of the associated fluid motion. The study of convection processes is an important research area in many fields of contemporary science and engineering.

Radiation

Turn a stove burner to "high" and it glows brightly; turn it to "low" and you can still sense its heat although it doesn't glow visibly. Either way, the burner loses energy by emitting electromagnetic waves, or **radiation**. The radiated power P increases rapidly with temperature, as described by the **Stefan-Boltzmann law**:

$$P = e\sigma AT^4 \quad \left(\begin{array}{c}\text{Stefan-Boltzmann law;}\\ \text{radiated power}\end{array}\right) \tag{16.9}$$

where A is the area of the emitting surface, T the temperature in kelvins, and σ the **Stefan-Boltzmann constant**, approximately $5.67\times10^{-8}\,\text{W/m}^2\cdot\text{K}^4$. The quantity e is the **emissivity**, a number from 0 to 1 that measures the material's effectiveness in emitting radiation. For radiation of a given wavelength, a material is equally good at emitting and absorbing radiation. A perfect emitter has $e = 1$ and is also a perfect absorber. Such an object would appear black at room temperature and is therefore called a **blackbody**. A shiny object, in contrast, reflects most of the radiation that hits it and is therefore also a poor emitter. Wood stoves are often painted black to increase their emissivity; Thermos bottles, on the other hand, have a shiny coating to reduce radiation.

Because of the strong T^4 temperature dependence, radiation is generally the dominant heat-loss mechanism at high temperatures but is less important at low temperatures. Radiation also dominates for objects in vacuum, since there's no material to carry conductive or convective heat flows; that makes Equation 16.9 crucial in understanding the climates of Earth and other planets.

Objects also absorb radiant energy from their surroundings, at a rate given by Equation 16.9 using the ambient temperature T_a, so the *net* radiated power becomes $P = e\sigma A(T^4 - T_a^4)$. For an object that's much hotter than its surroundings, the second term is negligible. But for an object that's only a little warmer, like a human body, it's significant.

It's not just the amount of radiation that changes with temperature; as our stove burner example suggests, it's also the wavelength. Objects at room temperature, for example, emit mostly invisible infrared radiation, while very hot objects like the Sun emit more visible light. We'll take a quantitative look at this relation in Chapter 34.

GOT IT? 16.3 Name the dominant form of heat transfer from (a) a red-hot stove burner with nothing on it, (b) a burner in direct contact with a pan of water, and (c) the bottom to the top of the water in the pan once it's begun to boil.

EXAMPLE 16.5 Calculating Radiation: The Sun's Temperature

The Sun radiates energy at the rate $P = 3.9\times10^{26}$ W, and its radius is 7.0×10^{8} m. Treating the Sun as a blackbody $(e = 1)$, find its surface temperature.

INTERPRET This is a problem about the radiation from a hot object.

DEVELOP The Stefan-Boltzmann law, Equation 16.9, gives the radiated power in terms of the temperature, emissivity, and surface area: $P = e\sigma AT^4$. Our plan is to solve this equation for T. For the Sun, radiation comes from the entire spherical surface of area $4\pi R^2$, as our sketch shows (Fig. 16.11).

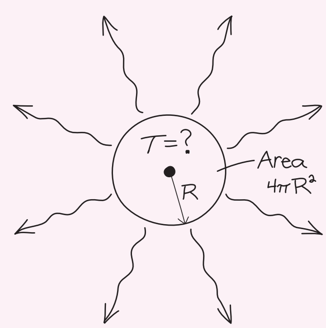


FIGURE 16.11 The Sun radiates from its spherical surface area $4\pi R^2$.

EVALUATE Using the Sun's spherical surface area and solving for T give

$$T = \left(\frac{P}{4\pi R^2\sigma}\right)^{1/4}$$

$$= \left[\frac{3.9\times10^{26}\,\text{W}}{4\pi(7.0\times10^{8}\,\text{m})^2(5.7\times10^{-8}\,\text{W/m}^2\cdot\text{K}^4)}\right]^{1/4} = 5.8\times10^{3}\,\text{K}$$

ASSESS Make sense? Yes: Our answer has the unit of temperature and agrees with observational measurements. Despite its bright glow, the Sun *is* essentially a blackbody, because it absorbs all radiation incident on it. But the Sun is so much hotter than its surroundings that we can neglect absorbed radiation in this calculation. ■

CONCEPTUAL EXAMPLE 16.1 Energy-Saving Windows

Why do double-pane windows reduce heat loss greatly compared with single-pane windows? Why is a window's $\mathcal{R}$-factor higher if the spacing between panes is small? And why do the best windows have "low-E" coatings?

EVALUATE Table 16.2 gives glass's thermal conductivity as around 0.8 W/m·K, while good insulators like air and Styrofoam have $k \sim 0.3$ W/m·K. That's why a layer of air between window panes greatly increases the window's $\mathcal{R}$-factor. But if the pane spacing is too great, convection currents develop between the sheets of glass, transferring heat from the warmer to the cooler surface; that's why narrower pane spacing is better. Finally, warm glass loses energy by radiation, and a thin coating of material with low emissivity ("low-E") reduces radiant heat loss.

ASSESS High-quality windows include all three features described here, so they suppress all three kinds of heat loss we've discussed. The best windows also use an inert gas—usually argon—between panes to reduce heat loss further.

MAKING THE CONNECTION Compare the $\mathcal{R}$-factor for a single-pane window made from 3.0-mm-thick glass with that of a double-pane window made from the same glass with a 5.0-mm air gap between panes.

EVALUATE Compute the $\mathcal{R}$-factors for the glass and air space, and you'll get about 0.004 $m^2 \cdot K/W$ for the single pane and, adding two layers of glass and the air space, 0.2 $m^2 \cdot K/W$ for the double-pane window. That's a factor of 50 improvement! In English units our answers translate into $\mathcal{R}$-factors of 0.02 and 1.1—although again they're lower than for actual windows because they neglect "dead air" layers and the other improvements discussed above.

16.4 Thermal-Energy Balance

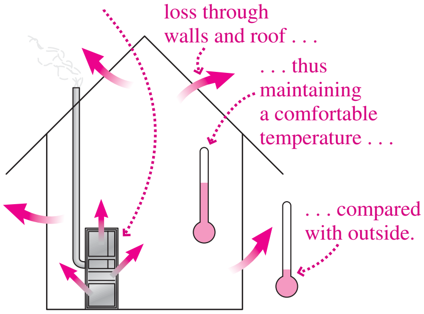


FIGURE 16.12 A house in thermal-energy balance.

You keep your house at a comfortable temperature in winter by balancing heat loss with energy from your heating system (Fig. 16.12). This state of **thermal-energy balance** occurs throughout science and engineering. Understanding thermal-energy balance enables engineers to specify a building's heat sources and helps scientists predict Earth's future climate.

Engineered systems actively control the thermal-energy balance to achieve a desired temperature. But even without active control, systems with a fixed rate of energy input naturally tend toward energy balance. That's because all heat-loss mechanisms give increased loss with increasing temperature. If the rate of energy input to a system is greater than the loss rate, then the system gains energy and its temperature increases—and so, therefore, does the loss rate. Eventually the two come to balance at some fixed temperature. If the loss exceeds the gain, the system cools until again it's in balance. Problems involving thermal-energy balance are similar regardless of the energy-loss mechanism or whether the application is to a technological or a natural system.

PROBLEM-SOLVING STRATEGY 16.1 Thermal-Energy Balance

INTERPRET Interpret the problem to be sure it deals with heat gains and losses. Identify the system of interest, the source(s) of energy input to the system, and the significant heat-loss mechanism(s).

DEVELOP Determine which equation(s) govern the heat loss; these will necessarily involve the system's temperature. Your plan is then to equate the rate of energy loss with the rate of energy input.

EVALUATE Write an equation that expresses equality of energy loss and input. Then evaluate by solving for the quantity the problem asks for—often the system's temperature.

ASSESS If your answer is a temperature, does it seem reasonable? Is the temperature of a heated system higher than that of its surroundings?

EXAMPLE 16.6 Thermal-Energy Balance: Hot Water

A poorly insulated electric water heater loses heat by conduction at the rate of 120 W for each Celsius degree difference between the water and its surroundings. It's heated by a 2.5-kW electric heating element and is located in a basement kept at 15°C. What's the water temperature if the heating element operates continuously?

INTERPRET The concept here is energy balance, and we identify the system of interest as the water. Its energy input comes from the heating element at the rate of 2.5 kW. The heat loss is by conduction.

DEVELOP Figure 16.13 is a sketch suggesting energy balance in the heater. We're given the conductive heat loss of 120 W/°C, meaning that the total heat-loss rate is $H = (120\ \text{W/°C})(\Delta T)$. We then equate the heat-loss rate to the energy-input rate: $(120\ \text{W/°C})(\Delta T) = 2.5\ \text{kW}$.

EVALUATE Solving for ΔT gives

$$\Delta T = \frac{2.5\ \text{kW}}{120\ \text{W/°C}} = 21\text{°C}$$

With the basement at 15°C, the water temperature is then 36°C.

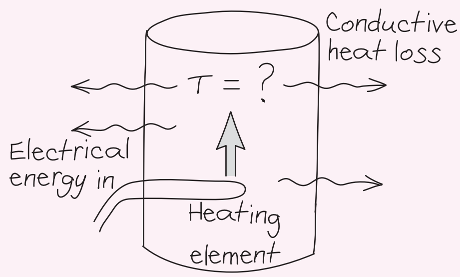


FIGURE 16.13 Balance between the heat supplied by the electric element and the conductive loss determines the water temperature.

ASSESS Is this answer reasonable? Not if you want a hot shower; our answer is 1°C below body temperature! But we're told the insulation is bad, so it's time for a new water heater! ■

EXAMPLE 16.7 Thermal-Energy Balance: A Solar Greenhouse

A solar greenhouse has 300 ft^2 of opaque $\mathcal{R}$-30 walls and 250 ft^2 of $\mathcal{R}$-1.8 double-pane glass that admits solar energy at the average rate of 40 Btu/h/ft^2. Find the greenhouse temperature on a day when the outdoor temperature is 15°F.

INTERPRET Again the concept is energy balance, now with the greenhouse as the system of interest. We're given $\mathcal{R}$-factors, suggesting that the energy loss is by conduction through walls and glazing. The energy input is sunlight.

DEVELOP As we saw in Example 16.4, the $\mathcal{R}$-factor determines a heat-loss rate that is related directly to area and temperature difference and inversely to the $\mathcal{R}$-factor. So we have

$$H_\text{w} = \frac{A_\text{w}\,\Delta T}{\mathcal{R}_\text{w}} = \left(\frac{300}{30}\right)\Delta T = (10\ \text{Btu/h/°F})\,\Delta T$$

for the heat loss through the walls and

$$H_\text{g} = \frac{A_\text{g}\,\Delta T}{\mathcal{R}_\text{g}} = \left(\frac{250}{1.8}\right)\Delta T = (139\ \text{Btu/h/°F})\,\Delta T$$

for the heat loss through the glass, giving a total heat loss $H = (149\ \text{Btu/h/°F})\Delta T$. Meanwhile, the energy input through the entire 250 ft^2 of glass is $(40\ \text{Btu/h/ft}^2)(250\ \text{ft}^2) = 1.0\times10^4$ Btu/h. Our plan is to equate energy input and loss and then solve for ΔT.

EVALUATE Equating loss and gain gives

$$(149\ \text{Btu/h/°F})\Delta T = 1.0\times10^4\ \text{Btu/h}.$$

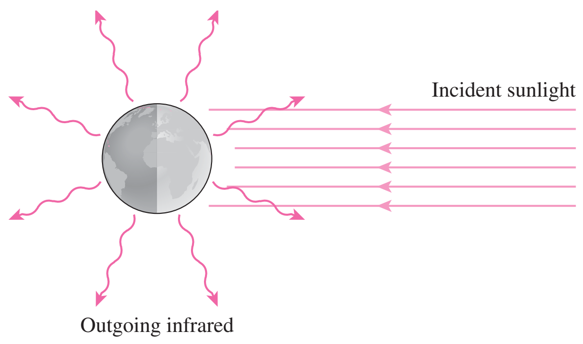
We then solve for ΔT:

$$\Delta T = \frac{1.0\times10^4\ \text{Btu/h}}{149\ \text{Btu/h/°F}} = 67\text{°F}$$

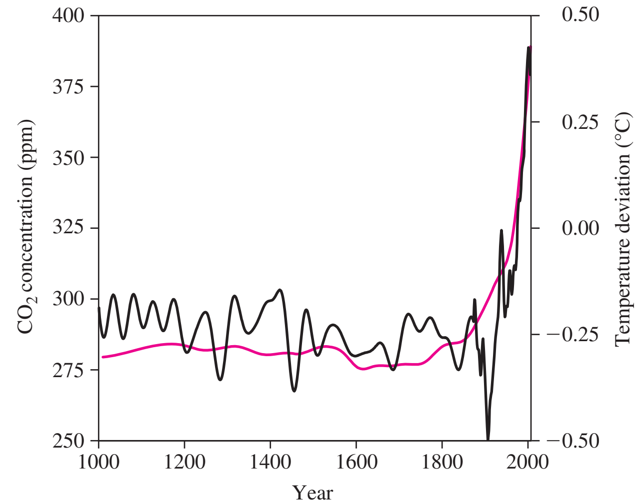
So when it's 15°F outside, the greenhouse is at a tropical 82°F.

ASSESS This seems a reasonable greenhouse temperature. Our calculation assumes that solar input remains constant; in a real greenhouse the temperature would fluctuate as the Sun's angle changes and clouds pass over. We could minimize these fluctuations by giving the greenhouse a large heat capacity, perhaps by incorporating a massive concrete slab or concrete walls. ■

APPLICATION The Greenhouse Effect and Global Warming



Earth's energy balance.



Atmospheric CO_2 concentration (green) and global temperature (black) from 1000–2010 A.D. Temperature is given as a deviation from the average for 1961–1990. Data through 1875 are reconstructed based on tree rings and other proxies; data from 1876 on are from thermometer records. The industrial era began around 1750.

Earth absorbs energy from the Sun at a rate $S = 960\ \text{W/m}^2$ averaged over the planet's cross-sectional area πR_E^2. It therefore warms and, in thermal-energy balance, radiates energy at the same rate. Earth is much cooler than the Sun, so this outgoing radiation is invisible infrared; furthermore, it's radiated from the planet's entire surface area, $4\pi R_E^2$. Assuming emissivity $e = 1$ in the infrared, energy balance using Equation 16.9 gives $\pi R_E^2 S = \sigma 4\pi R_E^2 T^4$. Solving yields $T = 255\ \text{K} = -18°\text{C}$ or 0°F. Is this reasonable? It's certainly in the right ballpark—not so hot as to boil the oceans or so cold as to freeze the atmosphere. But 0°F seems a bit cold for a global average temperature. And it is: Earth's average temperature is around 15°C or 59°F. Why the discrepancy?

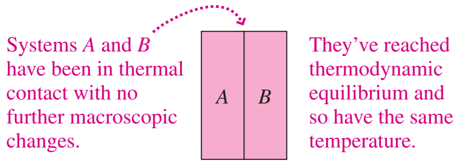
The answer lies with Earth's atmosphere. The dominant atmospheric gases, nitrogen and oxygen, are largely transparent to both incoming sunlight and outgoing infrared. But others—the so-called **greenhouse gases**, especially water vapor and carbon dioxide—let sunlight pass through but impede outgoing infrared. As a result, Earth's surface temperature has to be higher to get the same total radiation to space. This is the **natural greenhouse effect**, and it explains the 33°C temperature difference between our naïve calculation and Earth's actual surface temperature. Neighbor planets confirm this reasoning. Mars, with very little atmosphere, exhibits almost no greenhouse warming. Venus, whose atmosphere is 100 times denser than Earth's and largely CO_2, has a "runaway" greenhouse effect that keeps its surface hotter than an oven.

As the graph shows, we humans have increased atmospheric carbon dioxide some 40% since the start of the industrial era, to levels the planet has not seen for millions of years. Combustion of fossil fuels is the dominant source of this CO_2, although processes like deforestation also contribute, as do other greenhouse gases such as methane. Basic physics then dictates that Earth's surface temperature should rise. How much and how fast depend on complex interactions among atmosphere, surface, oceans, and life, and on future greenhouse emissions. Nevertheless, a consensus among climate scientists suggests that Earth warmed by some 0.6°C during the 20th century, mostly attributable to human activities (see graph). Further warming in the range of 1.5°C–6°C is projected by 2100. Although this may seem modest, the rate of increase is far greater than most natural climate change. And the increase is expected to be greater over land and at high latitudes. Even a few degrees' increase in the global average will result in significant climate change and a rise in sea level.

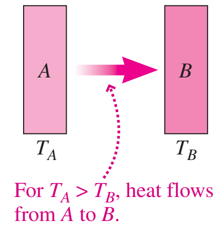
CHAPTER 16 SUMMARY

Big Picture

The big ideas here are **temperature** and **heat**. **Temperature** is a property common to systems in **thermodynamic equilibrium**. Temperature is quantified in SI units using the **kelvin scale**, defined in terms of gas-based thermometers.



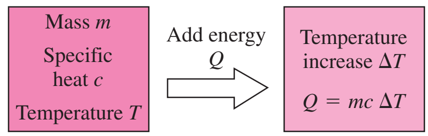
Heat is energy in transit as a result of a temperature difference.



Key Concepts and Equations

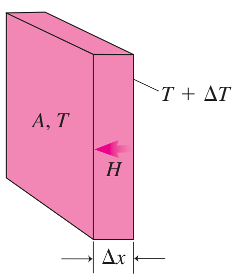
Heat capacity and **specific heat** quantify the energy Q required to raise an object's temperature by ΔT:

$$Q = mc\,\Delta T$$



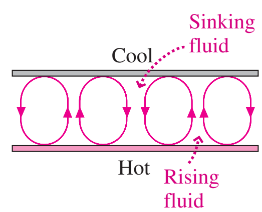
Three important heat-transfer mechanisms are:

Conduction

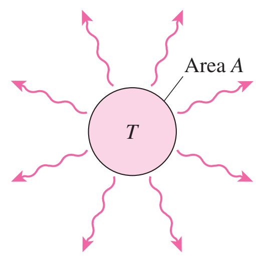


$$H = -kA\frac{\Delta T}{\Delta x} \quad \text{(conductive heat flow)}$$

Convection



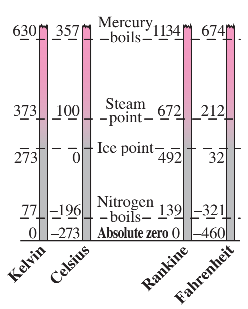
Radiation



$$P = e\sigma AT^4 \quad \left(\begin{array}{c}\text{Stefan-Boltzmann law;}\\ \text{radiated power}\end{array}\right)$$

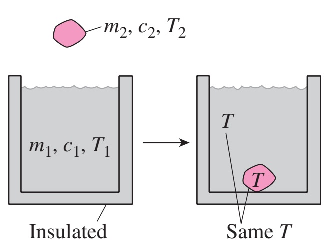
Applications

Temperature scales include Kelvin (K), Celsius (°C), Fahrenheit (°F), and Rankine (°R).

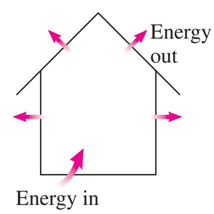


The Kelvin and Celsius scales are related by $T_C = T - 273.15$. The relation between Fahrenheit and Celsius scales is $T_F = \frac{9}{5}T_C + 32$.

Equilibrium temperature: Combining two systems at different temperatures results in a common equilibrium temperature given by $m_1c_1\,\Delta T_1 + m_2c_2\,\Delta T_2 = 0$.



Energy balance: A system experiencing both energy input and energy loss comes to energy balance at the temperature for which the energy-loss rate equals the rate of energy input.



For homework assigned on MasteringPhysics, go to www.masteringphysics.com | **BIO** *Biology and/or medical-related problems* **C** *Computer problems*

For Thought and Discussion

1. If system A is not in thermodynamic equilibrium with system B, and B is not in equilibrium with C, can you draw any conclusions about the temperatures of the three systems?
2. Does a thermometer measure its own temperature or the temperature of its surroundings? Explain.
3. Compare the relative sizes of the kelvin, the degree Celsius, the degree Fahrenheit, and the degree Rankine.
4. If you put a thermometer in direct sunlight, what do you measure: the air temperature, the temperature of the Sun, or some other temperature?
5. Why does the temperature in a stone building usually vary less than in a wooden building?
6. Why do large bodies of water exert a temperature-moderating effect on their surroundings?
7. A Thermos bottle consists of an evacuated, double-wall glass liner, coated with a thin layer of aluminum. How does it keep liquids hot?
8. Stainless-steel cookware often has a layer of aluminum or copper embedded in the bottom. Why?
9. What method of energy transfer dominates in baking? In broiling?
10. After a calm, cold night, the temperature a few feet above ground often drops just as the Sun comes up. Explain in terms of convection.
11. Glass and fiberglass are made from the same material, yet have dramatically different thermal conductivities. Why?
12. To keep your hands warm while skiing, you should wear mittens instead of gloves. Why?
13. Since Earth is exposed to solar radiation, why doesn't Earth have the same temperature as the Sun?

Exercises and Problems

Exercises

Section 16.1 Heat, Temperature, and Thermodynamic Equilibrium

14. A Canadian meteorologist predicts an overnight low of −15°C. How would a U.S. meteorologist express that prediction?
15. Normal room temperature is 68°F. What's this in Celsius?
16. The outdoor temperature rises by 10°C. What's that rise in Fahrenheit?
17. At what temperature do the Fahrenheit and Celsius scales coincide?
18. The normal boiling point of nitrogen is 77.3 K. Express this in Celsius and Fahrenheit.
19. A sick child's temperature reads 39.1 on a Celsius thermometer. What's the temperature in Fahrenheit?

Section 16.2 Heat Capacity and Specific Heat

20. Find the heat capacity of a 55-tonne concrete slab.
21. Find the energy needed to raise the temperature of a 2.0-kg chunk of aluminum by 18°C.
22. What's the specific heat of a material if it takes 7.5 kJ to increase the temperature of a 1-kg sample by 3.0°C?
23. **BIO** The average human diet contains about 2000 kcal per day. If all this food energy is released rather than stored as fat, what's the approximate average power output of the human body?
24. **BIO** Walking at 3 km/h requires an energy expenditure rate of about 200 W. How far would you have to walk to "burn off" a 420-kcal hamburger?
25. You bring a 350-g wrench into the house from your car. The house is 15°C warmer than the car, and it takes 2.52 kJ to warm the wrench by this amount. Find (a) the heat capacity of the wrench and (b) the specific heat of the metal it's made from.
26. (a) How much heat does it take to bring a 3.4-kg iron skillet from 20°C to 130°C? (b) If the heat is supplied by a stove burner at the rate of 2.0 kW, how long will it take to heat the pan?

Section 16.3 Heat Transfer

27. Building heat loss in the United States is usually expressed in Btu/h. What's that in SI units?
28. Find the heat-loss rate through a slab of (a) wood and (b) Styrofoam, each 2.0 cm thick, if one surface is at 20°C and the other at 0°C.
29. The top of a steel wood stove measures 90 cm by 40 cm and is 0.45 cm thick. The fire maintains the inside surface of the stovetop at 310°C, while the outside surface is at 295°C. Find the heat conduction rate through the stovetop.
30. You're a builder who's advising a homeowner to have her foundation walls insulated with 2 inches of Styrofoam. To make your point, you tell her how thick the concrete walls (normally 8 inches) would have to be to have the same insulating value as 2 inches of Styrofoam. What's this thickness?
31. An 8.0 m by 12 m house is built on a concrete slab 23 cm thick. Find the heat-loss rate through the floor if the interior is at 20°C while the ground is at 10°C.
32. Find the $\mathcal{R}$-factor for a wall that loses 0.040 Btu each hour through each square foot for each °F temperature difference.
33. Compute the $\mathcal{R}$-factors for 1-inch thicknesses of air, concrete, fiberglass, glass, Styrofoam, and wood.
34. A horseshoe has surface area 50 cm^2, and a blacksmith heats it to a red-hot 810°C. At what rate does it radiate energy?

Section 16.4 Thermal-Energy Balance

35. An oven loses energy at the rate of 14 W per °C temperature difference between its interior and the 20°C temperature of the kitchen. What average power must be supplied to maintain the oven at 180°C?
36. You're having your home's heating system replaced, and the heating contractor has specified a new system that supplies energy at the maximum rate of 40 kW. You know that your house loses energy at the rate of 1.3 kW per °C temperature difference between interior and exterior, and the minimum winter temperature in your area is −15°C. You'd like to maintain 20°C (68°F) indoors. Should you go with the system your contractor recommends?
37. The filament of a 100-W lightbulb is at 3.0 kK. What's the filament's surface area?
38. **BIO** A typical human body has surface area 1.4 m^2 and skin temperature 33°C. If the body's emissivity is about 1, what's the net radiation from the body when the ambient temperature is 18°C?

Problems

39. A constant-volume gas thermometer is filled with air whose pressure is 101 kPa at the normal melting point of ice. What would its pressure be at (a) the normal boiling point of water (373 K), (b) the normal boiling point of oxygen (90.2 K), and (c) the normal boiling point of mercury (630 K)?

40. A constant-volume gas thermometer is at 55-kPa pressure at the triple point of water. By how much does its pressure change for each kelvin temperature change?
41. In Fig. 16.2's gas thermometer, the height h is 60.0 mm at the triple point of water. When the thermometer is immersed in boiling sulfur dioxide, the height drops to 57.8 mm. What is the boiling point of SO_2 in kelvins and in degrees Celsius?
42. If your mass is 60 kg, what's the minimum number of Calories (kcal) you would "burn off" climbing a 1700-m-high mountain? (*Note:* The actual metabolic energy used would be much greater.) BIO
43. Typical fats contain about 9 kcal per gram. If the energy in body fat could be utilized with 100% efficiency, how much mass would a runner lose in a 26.2-mile marathon while consuming 125 kcal/mile? BIO
44. A circular lake 1.0 km in diameter is 10 m deep (Fig. 16.14). Solar energy is incident on the lake at an average rate of 200 W/m^2. If the lake absorbs all this energy and does not exchange heat with its surroundings, how long will it take to warm from 10°C to 20°C?

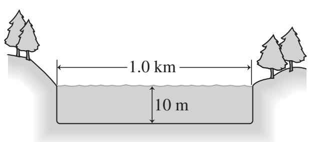


FIGURE 16.14 Problem 44

45. How much heat is required to raise an 800-g copper pan from 15°C to 90°C if (a) the pan is empty or contains (b) 1.0 kg of water and (c) 4.0 kg of mercury?
46. Initially, 100 g of water and 100 g of another substance listed in Table 16.1 are at 20°C. Heat is then transferred to each substance at the same rate for 1.0 min. At the end of that time, the water is at 32°C and the other substance at 76°C. (a) What's the other substance? (b) What's the heating rate?
47. You draw 330 mL of 10°C water from the tap and pop it into a 900-W microwave oven to heat for tea. How long should you microwave the water so it just reaches the boiling point?
48. Two neighbors return from Florida to find their houses at a frigid 35°F. Each house has a furnace that can supply 100,000 Btu/h. One house is made of stone and weighs 75 tons. The other is wood and weighs 15 tons. How long does it take each house to reach 65°F? Neglect heat loss, and assume the entire house mass reaches a uniform temperature.
49. You're arguing with your roommate about whether it's quicker to heat water on a stove burner or in a microwave. The burner supplies energy at the rate of 1.0 kW, the microwave at 625 W. You can heat water in the microwave in a paper cup of negligible heat capacity, but the stove requires a pan with heat capacity 1.4 kJ/K. How much water do you need before it becomes quicker to heat on the stovetop? Neglect energy loss to the surroundings.
50. When a nuclear power plant's reactor is shut down, radioactive decay continues to produce heat at about 10% of the reactor's normal power level of 3.0 GW. In a major accident, a pipe breaks and all the reactor cooling water is lost. The reactor is immediately shut down, the break is sealed, and 420 m^3 of 20°C water is injected into the reactor. If the water were not actively cooled, how long would it take to reach its normal boiling point?
51. A 1.2-kg iron tea kettle sits on a 2.0-kW stove burner. If it takes 5.4 min to bring the kettle and the water in it from 20°C to the boiling point, how much water is in the kettle?
52. The temperature of the eardrum provides a reliable measure of deep body temperature and is measured quickly with ear thermometers that sense infrared radiation. A thermometer that "views" 1 mm^2 of the eardrum requires 100 μJ of energy for a reliable reading at normal 37°C body temperature. How long does the measurement take? BIO
53. A 1500-kg car moving at 40 km/h is brought to a sudden stop. If all the car's energy is dissipated in heating its four 5.0-kg steel brake disks, by how much do the disk temperatures increase?
54. Your young niece complains that her cocoa, at 90°C, is too hot. You pour 2 oz of milk at 3°C into the 6 oz of cocoa. Assuming milk and cocoa have the same specific heat as water, what's the cocoa's new temperature?
55. A piece of copper at 300°C is dropped into 1.0 kg of water at 20°C. If the equilibrium temperature is 25°C, what's the mass of the copper?
56. While camping, you boil water to make spaghetti. Your pot contains 2.5 kg of water initially at 10°C. You stoke up the campfire, and as a result the water gains energy at an increasing rate: $P = a + bt$, where $a = 1.1$ kW, $b = 2.3$ W/s, and t is the time in s. To the nearest minute, how long will it take to bring the water to a boil?
57. A biology lab's walk-in cooler measures 3.0 m by 2.0 m by 2.3 m and is insulated with 8.0-cm-thick Styrofoam. If the surrounding building is at 20°C, at what average rate must the cooler's refrigeration unit remove heat in order to maintain 4.0°C in the cooler?
58. One end of an iron rod 40 cm long and 3.0 cm in diameter is in ice water, the other in boiling water (Fig. 16.15). The rod is well insulated so no heat is lost out the sides. Find the heat-flow rate along the rod.

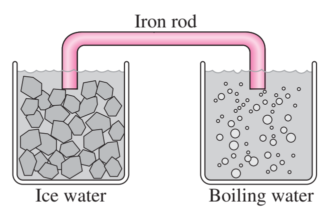


FIGURE 16.15 Problem 58

59. You arrive for a party on a night when it's 8°C outside. Your hosts meet you at the door and say the party may need to be cancelled, because the heating system has failed and they don't want to discomfort their guests. You say, "Not so fast!" A total of 36 people are expected, the average power output of a human body is 100 W, and the house loses 320 W/°C. Will the house remain comfortable?
60. An electric stove burner has surface area 325 cm^2 and emissivity $e = 1$. The burner consumes 1500 W and is at 900 K. If room temperature is 300 K, what fraction of the burner's heat loss is from radiation?
61. An electric current passes through a metal strip 0.50 cm by 5.0 cm by 0.10 mm, heating it at a rate of 50 W. The strip has emissivity $e = 1$ and its surroundings are at 300 K. What will be the strip's temperature if it's enclosed in (a) a vacuum bottle transparent to all radiation and (b) an insulating box with thermal resistance $R = 8.0$ K/W that blocks all radiation?
62. You're considering purchasing a new sleeping bag whose manufacturer claims will keep you warm to −10°F. The bag has down insulation with 4.0-cm loft (thickness). Your body produces heat

at the rate of 100 W and has area 1.5 m^2. Considering only conductive heat loss, will you be able to maintain normal body temperature in the bag at -10°F?

63. A blacksmith heats a 1.1-kg iron horseshoe to 550°C, then plunges it into a bucket containing 15 kg of water at 20°C. What's the equilibrium temperature?
64. What is the power output of a microwave oven that can heat 430 g of water from 20°C to the boiling point in 2.5 min? Neglect the container's heat capacity.
65. A cylindrical log 15 cm in diameter and 65 cm long is glowing red hot in a fireplace. The log's emissivity is essentially 1. If it's emitting radiation at the rate of 34 kW, what's its temperature?
66. A star whose surface temperature is 50 kK radiates 4.0×10^{27} W. If the star behaves like a blackbody, what's its radius?
67. Rework Example 16.4, now assuming the house has ten single-glazed windows, each measuring 2.5 ft by 5.0 ft. Four of the windows are on the south and admit solar energy at the average rate of 30 Btu/h·ft^2. *All* the windows lose heat; their $\mathcal{R}$-factor is 0.90. (a) Find the total heating cost for the month. (b) How much is the solar gain worth?
68. A black wood stove with surface area 4.6 m^2 is made from cast iron 4.0 mm thick. Its interior wall is at 650°C, while the exterior is at 647°C. (a) What's the rate of heat conduction through the stove wall? (b) What's the rate of heat loss by radiation from the stove? (c) Use the results of (a) and (b) to find how much heat the stove loses by a combination of conduction and convection in the surrounding air.
69. What's the average temperature on Pluto at its distance from the Sun? Treat Pluto as a blackbody.
70. In a cylindrical pipe where area isn't constant, Equation 16.5 takes the form $H = -kA(dT/dR)$. Use this equation to show that the heat-loss rate from a cylindrical pipe of radius R_1 and length L is

$$H = \frac{2\pi kL(T_1 - T_2)}{\ln(R_2/R_1)}$$

where the pipe is surrounded by insulation of outer radius R_2 and thermal conductivity k and where T_1 and T_2 are the temperatures at the pipe surface and the outer surface of the insulation, respectively. (*Hint:* Consider the heat flow through a thin section of pipe, with thickness dr, as shown in Fig. 16.16. Then integrate.)

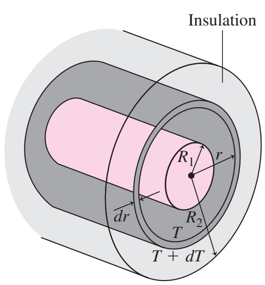


FIGURE 16.16 Problem 70

71. A friend who's skeptical about climate change argues that the roughly 0.75°C increase in Earth's temperature during the industrial era could be caused by an increase in the Sun's power output. The Sun's average power has, in fact, increased by about 0.04% during this time. Could your friend be right?
72. Your family is winterizing its lakefront camp, and you want at least $\mathcal{R}$-19 insulation in the walls. You've got some European-made insulation with $\mathcal{R}$-factor 3.5 m^2·K/W. Will it do?
73. Your niece from Problem 54 keeps her pet rabbit in a backyard hutch with thermal resistance 0.25 K/W. On a day when the outside temperature is -15°C, she's worried that the rabbit's water will freeze, so you put a 50-W heat lamp in the hutch. Will the bunny be able to drink its water? Neglect the heat due to the animal's metabolism.
74. At low temperatures a solid's specific heat is approximately proportional to the cube of the absolute temperature; for copper $c = 31(T/343\text{ K})^3$ J/g·K. Integrate Equation 16.3 in differential form $(dQ = mc\,dT)$ to find the heat required to bring a 40-g sample of copper from 10 K to 25 K.
75. Use the method outlined in Problem 70 to show that the steady heat-flow rate in the direction of the axis of a truncated cone with conductivity k, faces of radii R_1 and R_2, and length L is $H = \pi k R_1 R_2 (T_1 - T_2)/L$. Here, T_1 and T_2 are the temperatures on the two faces, and insulation prevents any heat flow out the sides (Fig. 16.17).

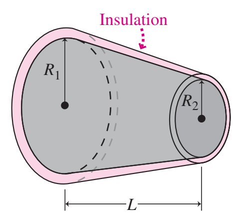


FIGURE 16.17 Problem 75

76. A house is at 20°C on a winter night when the outside temperature is a steady -15°C. The house's heat capacity is 6.5 MJ/K and its thermal resistance is 6.67 mK/W. If the furnace suddenly fails, how long will it take the house temperature to reach the freezing point? (*Hint:* Combine the differential forms of Equations 16.3 and 16.5 to show that the rate of temperature change is proportional to the temperature difference between the house and its surroundings. This relation is known as *Newton's law of cooling*.)
77. **C** A more realistic approach to the solar greenhouse of Example 16.7 considers the time dependence of the solar input. A function that approximates the solar input is $(40\text{ Btu/h/ft}^2)\sin^2(\pi t/24)$, where t is the time in hours, with $t = 0$ at midnight. Then the greenhouse is no longer in energy balance, but is described instead by the differential form of Equation 16.3 with Q the time-varying energy input. Use computer software or a calculator with differential-equation-solving capability to find the time-dependent temperature of the greenhouse, and determine the maximum and minimum temperatures. Assume the same numbers as in Example 16.7, along with a heat capacity $C = 1500$ Btu/°F for the greenhouse. You can assume any reasonable value for the initial temperature, and after a few days your greenhouse temperature should settle into a steady oscillation independent of the initial value.

Passage Problems

Fiberglass is a popular, economical, and fairly effective building insulation. It consists of fine glass fibers—often including recycled glass—formed loosely into rectangular slabs or rolled into blankets (Fig. 16.18). One side is often faced with heavy paper or aluminum foil. Fiberglass insulation comes in thicknesses compatible with common building materials—for example, 3.5 inch and 6 inch for wood-framed walls. Standard 6-inch fiberglass has an $\mathcal{R}$-factor of 19.

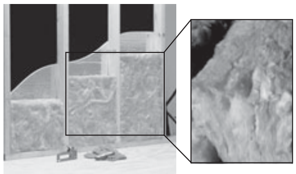


FIGURE 16.18 End view of a slab of fiberglass insulation (Passage Problems 78–81).

78. Fiberglass insulation owes its insulating quality primarily to
    a. the low thermal conductivity of glass.
    b. its ability to block cold air infiltration.
    c. the low thermal conductivity of air trapped between the glass fibers.
79. One purpose of foil facing on fiberglass insulation is to reduce heat loss by
    a. conduction.
    b. convection.
    c. radiation.
80. Fiberglass insulation for attics is available in 12-inch thickness. Its $\mathcal{R}$-factor is
    a. 38.
    b. 76.
    c. 29.
81. Since fiberglass insulation is readily compressible, you could squash two slabs initially 6 inches wide into a 6-inch wall space. This would
    a. double the overall $\mathcal{R}$-factor.
    b. increase the overall $\mathcal{R}$-factor but not double it.
    c. decrease the overall $\mathcal{R}$-factor.
    d. not change the overall $\mathcal{R}$-factor.

Answers to Chapter Questions

Answer to Chapter Opening Question

The photo is taken in infrared light, and the amount of infrared radiation increases rapidly with increasing temperature. The car's wheels are glowing with infrared, a result of frictional heating when the brakes were recently applied.

Answers to GOT IT? Questions

16.1. The rock's temperature changes more because its specific heat is lower.

16.2. $\Delta T_2 < \Delta T_3 < \Delta T_1$; since H and Δx are the same for each slab, the product $k\,\Delta T$ must be constant, so a higher conductivity means a lower ΔT.

16.3. (a) Radiation; (b) conduction; (c) convection.

17 The Thermal Behavior of Matter

What unusual property of water is evident in this photo?

New Concepts, New Skills

By the end of this chapter you should be able to

- Explain the relation between the statistical mechanics of molecules and the corresponding thermodynamic properties for simple gases. This includes especially the relation between molecular energy and temperature (17.1).
- Describe the three principal phases of matter, and calculate the energies required to change phase (17.2).
- Interpret phase diagrams to determine conditions under which the various phases exist (17.2).
- Calculate thermal-expansion effects in solids, liquids, and gases (17.3).
- Describe the unusual thermal-expansion properties of water (17.3).

Connecting Your Knowledge

- This chapter links thermodynamics with the statistical mechanics of atoms and molecules. Key concepts in this linkage are conservation of energy and conservation of momentum (7.3, 9.2).

Matter responds to heating in several ways. It may get hotter or it may melt. It may change size, shape, or pressure. This chapter explores the thermal behavior of matter. We start with a simple gaseous state, whose behavior follows from Newtonian mechanics at the molecular level. We then move to liquids and solids, whose behavior is still grounded in the molecular properties of matter, but whose description is more empirical.

17.1 Gases

Gases are simple because their molecules are far apart and only rarely interact. That makes gas behavior and its physical explanation particularly straightforward. Developing that explanation will clarify the relation between macroscopic properties—such as temperature and pressure—and the underlying microscopic properties of gas molecules.

The Ideal-Gas Law

The macroscopic state of a gas in thermodynamic equilibrium is determined by its temperature, pressure, and volume. Moreover, it turns out that all gases exhibit, to a very good approximation, the same relation among these three quantities.

A simple system for studying gas behavior consists of a gas-filled cylinder sealed by a movable piston (Fig. 17.1). This is not just a pedagogical abstraction: Practical devices including engines, pumps, and air compressors contain piston-cylinder systems, while lungs, balloons, gas bubbles, and many other natural systems are analogous to our piston-cylinder system.

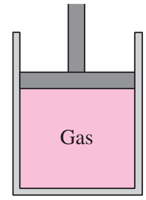


FIGURE 17.1 A piston-cylinder system.

If we maintain the system of Fig. 17.1 at constant temperature and move the piston to vary the gas volume, we find that the pressure varies inversely with the volume. If we increase the temperature while holding the volume fixed, the pressure rises in direct proportion to the temperature. If we double the amount of gas while holding temperature and volume constant, the pressure doubles. Putting all these results together, we can write

$$pV = NkT \quad (\text{ideal-gas law}) \tag{17.1}$$

with p, V, and T the pressure, volume, and temperature, respectively, and N the number of molecules in the gas. The constant $k = 1.38\times10^{-23}$ J/K is **Boltzmann's constant**, named for the Austrian physicist Ludwig Boltzmann (1844–1906), who was instrumental in developing the microscopic description of thermal phenomena. Equation 17.1 is the **ideal-gas law**. Most real gases obey this law to a very good approximation.

Because the number of molecules N in a typical gas sample is astronomically large, we often express the ideal-gas law in terms of the number of **moles** (mol) of gas molecules. One mole is an SI unit equal to Avogadro's number, $N_A = 6.022\times10^{23}$, of atoms or molecules. Formally, Avogadro's number is defined as the number of carbon-12 atoms in 12 grams of carbon-12.

If we have n moles of a gas, then $N = nN_A$ is the number of molecules, so the ideal-gas law becomes

$$pV = nN_AkT = nRT \tag{17.2}$$

where $R = N_Ak = 8.314$ J/K·mol is called the **universal gas constant**.

EXAMPLE 17.1 The Ideal-Gas Law: STP

What volume is occupied by 1.00 mol of an ideal gas at standard temperature and pressure (STP), where $T = 0°\text{C}$ and $p = 101.3$ kPa?

INTERPRET We're dealing with an ideal gas, and we're given the amount of gas, the temperature, and the pressure.

DEVELOP Because we're given the number of moles n, we'll use the ideal-gas law in the form of Equation 17.2, $pV = nRT$, to find the volume.

EVALUATE Solving for V gives

$$V = \frac{nRT}{p} = \frac{(1.00\text{ mol})(8.314\text{ J/K}\cdot\text{mol})(273\text{ K})}{1.01\times10^5\text{ Pa}}$$
$$= 22.4\times10^{-3}\text{ m}^3 = 22.4\text{ L}$$

where we expressed $T = 0°\text{C}$ as 273 K.

ASSESS This result may be familiar from earlier chemistry or physics courses: 1 mole of any ideal gas—no matter what its chemical composition—occupies 22.4 L at standard temperature and pressure. ■

The ideal-gas law is remarkably simple. Neither its form nor the constants k and R depend on the substance making up the gas or on the mass of the gas molecules. Yet most real gases follow the ideal-gas law very closely over a wide range of pressures. This nearly ideal behavior is what gives gas thermometers their high precision over a wide temperature range.

Kinetic Theory of the Ideal Gas

Why do gases obey such a simple relation among temperature, pressure, and volume? Here we answer that question with an analysis based ultimately on Newtonian mechanics.

We start with some simplifying assumptions:

1. The gas consists of many identical molecules, each with mass m but negligible size and no internal structure. This assumption is approximately true for real gases when the distance between molecules is large compared with their size. This allows us to

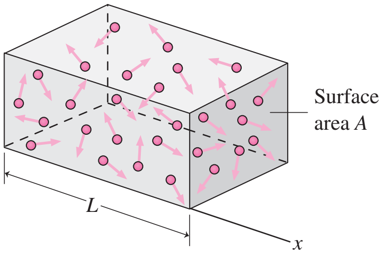


FIGURE 17.2 Gas molecules confined to a rectangular box.

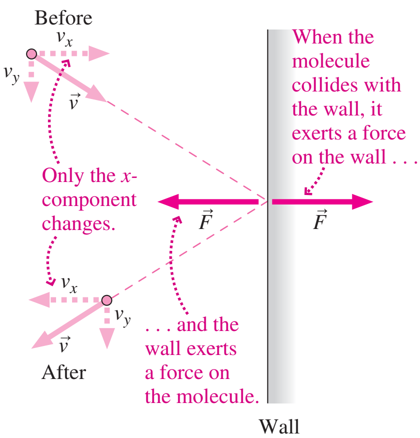


FIGURE 17.3 A molecule undergoes an elastic collision, reversing its x-component and transferring momentum $2mv_x$ to the wall.

neglect intermolecular collisions, an assumption that simplifies our analysis but isn't crucial to the ideal gas.

2. The molecules don't exert action-at-a-distance forces on each other. Thus there's no intermolecular potential energy, and therefore the molecules have only kinetic energy. This assumption is fundamental to an ideal gas.
3. The molecules move in random directions with a distribution of speeds that's independent of direction.
4. Collisions with the container walls are elastic, conserving the molecules' energy and momentum. Here's where we tie our gas model to Newtonian mechanics.

Consider N molecules confined to a rectangular box with length L (Fig. 17.2). Each molecule that collides with a wall exerts a force. There are so many molecules that individual collisions aren't evident; instead the wall experiences an essentially constant average force. The gas pressure p is a measure of this force on a unit area. We're going to find an expression for p and show that it takes the form of the ideal-gas law.

Figure 17.3 shows one molecule colliding with the right-hand wall. Since the collision is elastic, the y-component of the molecule's velocity is unchanged, while the x-component reverses sign. Thus the molecule undergoes a momentum change of magnitude $2mv_{xi}$, where i labels this particular molecule. After the molecule collides with the right-hand wall, nothing will change its x velocity until it hits the left-hand wall and its x velocity again reverses. So it will be back at the right-hand wall in the time $\Delta t_i = 2L/v_{xi}$ that it takes to go back and forth along the container.

Now each time our molecule collides with the right-hand wall, it delivers momentum $2mv_{xi}$ to the wall. Newton's second law says that force is the rate of change of momentum. So we can calculate the average force $\overline{F_i}$ due to one molecule by dividing the momentum delivered, $2mv_{xi}$, by the time, $2L/v_{xi}$, between collisions:

$$\overline{F_i} = \frac{2mv_{xi}}{2L/v_{xi}} = \frac{mv_{xi}^2}{L}$$

To get the total force on the wall, we sum over all N molecules with their different x velocities. Dividing by the wall area A then gives the pressure:

$$p = \frac{\overline{F}}{A} = \frac{\sum \overline{F_i}}{A} = \frac{\sum mv_{xi}^2/L}{A} = \frac{m\sum v_{xi}^2}{AL}$$

The last step follows because the box length L and molecular mass m are the same for all molecules, so they factor out of the sum. We can simplify by noting that the denominator AL is just the volume V. Let's also multiply by 1 in the form N/N, with N the number of molecules. Then we have

$$p = \frac{m\sum v_{xi}^2}{AL} = \frac{mN}{V}\frac{\sum v_{xi}^2}{N}$$

In the final expression here, the term $\sum v_{xi}^2/N$ is the average of the squares of all the x velocity components of all the molecules; we designate this quantity $\overline{v_x^2}$. So the pressure becomes

$$p = \frac{mN}{V}\overline{v_x^2}$$

We still haven't used assumption 3—that the molecules move in random directions with speeds independent of direction. If we grab a molecule at random, that means we're just as likely to find it moving in the x-direction, the y-direction, the z-direction, or any direction in between—and its speed, on average, won't depend on its direction of motion. So the average quantities $\overline{v_x^2}$, $\overline{v_y^2}$, and $\overline{v_z^2}$ must be equal. Since the three directions x, y, and z are perpendicular, the average of the molecular speeds squared is $\overline{v^2} = \overline{v_x^2} + \overline{v_y^2} + \overline{v_z^2}$. We've just argued that all three terms on the right are equal, so we can write $\overline{v^2} = 3\overline{v_x^2}$, or $\overline{v_x^2} = \frac{1}{3}\overline{v^2}$. Then our expression for pressure becomes

$$p = \frac{mN}{3V}\overline{v^2}$$

Multiplying through by V and by 1 in the form 2/2, we have

$$pV = \tfrac{2}{3}N\left(\tfrac{1}{2}m\overline{v^2}\right)$$

This looks a lot like the ideal-gas law (Equation 17.1), except that instead of kT we have $\tfrac{2}{3}\left(\tfrac{1}{2}m\overline{v^2}\right)$. Take a good look at the quantity in parentheses: You'll see that it's just the average kinetic energy of a gas molecule.

Think about what we've done here. We applied the fundamental laws of mechanics to an ideal gas and came up with an equation that looks like the experimentally verified ideal-gas law, except that it's expressed in terms of a microscopic quantity—molecular kinetic energy—rather than the macroscopic quantity temperature. Since our equation describes the behavior of an ideal gas, it *must be* the ideal-gas law. Comparing with the ideal-gas law in the form 17.1, we must therefore have

$$\tfrac{1}{2}m\overline{v^2} = \tfrac{3}{2}kT \quad \text{(temperature and molecular energy)} \tag{17.3}$$

Our derivation shows why, in terms of Newtonian mechanics, a gas obeying our four assumptions should obey the ideal-gas law. In Equation 17.3 we get an added bonus—a microscopic understanding of the meaning of temperature: **Temperature measures the average kinetic energy associated with random translational motion of the molecules.**

EXAMPLE 17.2 Molecular Energy and Speed: An Air Molecule

Find the average kinetic energy of a molecule in air at room temperature (20°C or 293 K), and determine the speed of a nitrogen molecule (N_2) with this energy.

INTERPRET This problem asks about the linkage between thermodynamic quantities and molecular energy. We just found that linkage: The temperature of a gas is a measure of the average kinetic energy of its molecules.

DEVELOP Equation 17.3, $\tfrac{1}{2}m\overline{v^2} = \tfrac{3}{2}kT$, quantifies the relation between temperature and molecular kinetic energy. Once we find the molecular kinetic energy, we'll need the molecular mass to determine the speed. We can get that using the atomic weight of nitrogen and the fact that an N_2 molecule contains two atoms.

EVALUATE We first evaluate the average molecular kinetic energy:

$$\overline{K} = \tfrac{1}{2}m\overline{v^2} = \tfrac{3}{2}kT = \tfrac{3}{2}(1.38\times10^{-23}\ \text{J/K})(293\ \text{K}) = 6.07\times10^{-21}\ \text{J}$$

We can solve for the corresponding speed if we know the molecular mass m. A nitrogen molecule consists of two atoms each with mass 14 u (see Appendix D), so its mass is

$$m = 2(14\ \text{u})(1.66\times10^{-27}\ \text{kg/u}) = 4.65\times10^{-26}\ \text{kg}$$

Since $\overline{K} = \tfrac{1}{2}m\overline{v^2}$, the speed corresponding to this kinetic energy is

$$v = \sqrt{\frac{2K}{m}} = \sqrt{\frac{2(6.07\times10^{-21}\ \text{J})}{4.65\times10^{-26}\ \text{kg}}} = 511\ \text{m/s}$$

ASSESS Make sense? Not surprisingly, the answer is the same order of magnitude as the speed of sound ($\sim$340 m/s) in air at room temperature. At the microscopic level, the speed of the individual molecules limits the rate at which information can be transmitted by disturbances—sound waves—propagating through the gas. ■

We call the speed calculated in Example 17.2 the **thermal speed**. In terms of temperature, Equation 17.3 shows

$$v_{\text{th}} = \sqrt{\frac{3kT}{m}} \tag{17.4}$$

GOT IT? 17.1 If you double the kelvin temperature of a gas, what happens to the thermal speed of the gas molecules?

The Distribution of Molecular Speeds

The thermal speed v_{th} is a typical molecular speed, but it doesn't tell us much about the distribution of speeds. Are molecular speeds limited to a narrow band about v_{th}? Or are lots of molecules moving much faster or much slower?

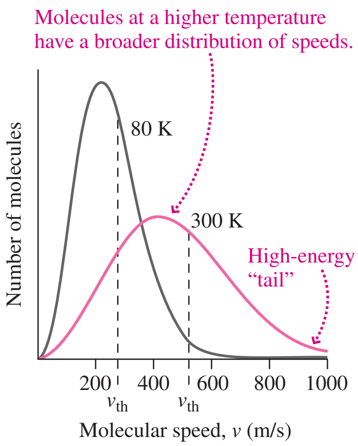


FIGURE 17.4 Maxwell-Boltzmann distribution of molecular speeds for nitrogen (N_2) at temperatures of 80 K and 300 K.

In the 1860s, the Scottish physicist James Clerk Maxwell showed that elastic collisions among molecules result in a speed distribution that peaks near the thermal speed but may extend considerably higher. Figure 17.4 plots this **Maxwell-Boltzmann distribution** for two different temperatures. Note that increasing temperature results in a higher thermal speed, as expected, but that it also broadens the distribution so there are more molecules at lower and higher speeds. The high-speed "tail" of the distribution is especially important to chemists because high-energy molecules participate most readily in chemical reactions. The rapid extension of the high-energy tail with increasing temperature shows why reaction rates are strongly temperature sensitive, and explains why foods keep much longer with even modest refrigeration. High-energy molecules are also the first to evaporate from a liquid, leaving slower, cooler molecules behind and thus explaining evaporative cooling. Without this effect, Earth's atmosphere would be much drier and it would rain far less frequently.

Real Gases

The ideal-gas law is a good approximation to the behavior of most real gases, but it's not perfect because our assumptions aren't entirely realistic. Two factors are especially important. First, real molecules take up space. This reduces the available volume, altering the ideal-gas law. Second, electrical effects that we'll explore in Chapter 20 result in a weak attractive force between nearby molecules. As they move apart, molecules do work against this **van der Waals force**, and their kinetic energy drops. This, too, results in a deviation from ideal-gas behavior. You can explore these effects further in the Passage Problems at the end of this chapter.

17.2 Phase Changes

Step out of a steamy shower, and you'll find the mirror fogged with water condensed on the cool glass. Climb a mountain in winter, and you'll be treated to the lovely spectacle of every branch and pine needle covered with a delicate coating of frost that's formed right from the air. Burn a rewritable CD or DVD, and you've stored information with a laser that melts tiny spots on the spinning disc. These examples involve **phase changes** between gas and liquid, gas and solid, and solid and liquid.

Heat and Phase Changes

Drop ice cubes into a drink and stir. What's the temperature of the drink? It's 0°C, and it stays at 0°C as long as any ice remains. The melting of a pure solid occurs at a fixed temperature. During the process, energy goes into breaking the molecular bonds that hold the material in its solid form. This increases the molecules' potential energy but not their kinetic energy. Since temperature is a measure of molecular kinetic energy, that means the temperature doesn't change either.

The energy per unit mass required to change phase is called a **heat of transformation** L; for the solid-liquid change it's the **heat of fusion** L_f, and for liquid-gas it's the **heat of vaporization** L_v. Less familiar is the **heat of sublimation** for the transition from solid directly to gas. These quantities have units of J/kg, so the energy required to change the phase of a mass m is

$$Q = Lm \quad \text{(heat of transformation)} \tag{17.5}$$

To reverse the change requires removing the same energy. Table 17.1 lists heats of transformation for some common materials. These quantities are typically quite large; water's heat of fusion, for example, is 334 kJ/kg or 80 cal/g—meaning it takes as much energy to melt 1 gram of ice as to heat the resulting water to 80°C.

Table 17.1 Heats of Transformation (at Atmospheric Pressure)

Substance	Melting Point (K)	L_f (kJ/kg)	Boiling Point (K)	L_v (kJ/kg)
Alcohol, ethyl	159	109	351	879
Copper	1357	205	2840	4726
Lead	601	24.7	2013	858
Mercury	234	11.3	630	296
Oxygen	54.8	13.8	90.2	213
Sulfur	388	53.6	718	306
Water	273	334	373	2257
Uranium dioxide	3120	259	3815	1533

CONCEPTUAL EXAMPLE 17.1 Water Phases

You put a block of ice initially at −20°C in a pan on a hot stove with a constant power output, and heat it until it has melted, boiled, and evaporated. Make a sketch of temperature versus time for this experiment.

EVALUATE As the ice starts heating, its temperature goes up, so our graph (Fig. 17.5) begins with an upward slope. At 0°C the ice starts melting, and while that's happening its temperature doesn't change, so the graph stays horizontal for a while. When the ice is all melted, the water starts to warm. Table 16.1 shows that liquid water's specific heat is about twice that of ice; given the same power input, that means the water heats more slowly than the ice. So our graph has a lower slope as the water goes from 0°C to the boiling point at 100°C. Then it starts turning to vapor, and stays at 100°C until it's all evaporated. Table 17.1 shows that water's heat of vaporization is much greater than its heat of fusion, so it takes much more time to boil the water away than it did to melt the ice. Our graph reflects that time difference.

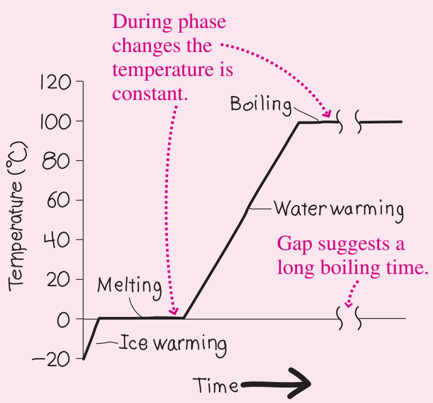


FIGURE 17.5 Temperature versus time for what's initially a block of ice at −20°C, supplied with energy at a constant rate. The process takes place at atmospheric pressure.

ASSESS Makes sense: It takes a lot longer to boil a pan dry than to bring it to a boil.

MAKING THE CONNECTION If you start with 0.95 kg of ice at −20°C and supply heat at the rate of 1.6 kW, how much time will it take until you're left with only water vapor?

EVALUATE Use Equation 16.3 for heating, with specific heats from Table 16.1. Use Equation 17.4 for phase changes, with heats of transformation from Table 17.1. The result is 2.9 MJ of heat required for the whole process; at 1.6 kW or 1.6 kJ/s, that takes 1.8 ks, or half an hour.

GOT IT? 17.2 You bring a pot of water to boil and then forget about it. Ten minutes later you come back to the kitchen to find the water still boiling. Is its temperature (a) less than, (b) greater than, or (c) equal to 100°C?

EXAMPLE 17.3 The Heat of Fusion: Meltdown!

A nuclear power plant's reactor vessel cracks, and all the cooling water drains out. Although nuclear fission stops, radioactive decay continues to heat the reactor's 2.5×10^5 kg of uranium-dioxide fuel at the rate of 120 MW. Once the melting point is reached, how much energy will it take to melt the fuel? How long will this take?

INTERPRET Since this problem is about melting, it must involve the heat of fusion. We identify the material as uranium dioxide (UO_2).

DEVELOP Our plan is to find UO_2's heat of fusion in Table 17.1 and then use Equation 17.5, $Q = Lm$, to calculate the energy required for melting. We're given the rate of energy generation by radioactive decay, and from that we'll be able to get the time.

EVALUATE Using UO_2's L_f value from Table 17.1 in Equation 17.5, we have

$$Q = L_f m = (259\text{ kJ/kg})(2.5\times10^5\text{ kg}) = 65\text{ GJ}$$

With a heating rate of 120 MW or 0.12 GJ/s, the time to melt the fuel is $(65\text{ GJ})/(0.12\text{ GJ/s}) = 542$ s.

ASSESS The time to meltdown is just under 10 minutes! Failsafe emergency cooling systems are essential to prevent nuclear meltdowns. ■

Often we're interested in the total energy needed to bring a material to its transition point and then to make the phase transition. Then we need to combine specific-heat considerations of Chapter 16 with the heats of transformation introduced here.

EXAMPLE 17.4 Heating and Phase Change: Enough Ice?

When 200 g of ice at $-10°C$ are added to 1.0 kg of water at 15°C, is there enough ice to cool the water to 0°C? If so, how much ice is left in the mixture?

INTERPRET This problem involves both a temperature rise and a phase change. We identify water as the substance involved.

DEVELOP Equation 16.3, $Q = mc\,\Delta T$, determines the energy for the temperature rise, and Equation 17.5, $Q = Lm$, determines the phase-change energy. But we don't know whether all the ice melts. So our plan is to find the energy that it *would* take to heat the ice to 0°C and then melt all of it; if *more* than that much is available in cooling the water to 0°C, we'll know that we end up with all water at $T > 0°C$. But if there isn't sufficient energy, then we'll have a mixture with both ice and water at 0°C, and we can use the energy extracted in cooling the water to find out how much ice melts.

EVALUATE We begin by evaluating the energy Q_1 to heat the ice and then melt it all, adding the energies from Equations 16.3 and 17.5 and then getting the specific heat and heat of fusion from Tables 16.1 and 17.1, respectively:

$$\begin{aligned} Q_1 &= m_{\text{ice}}c_{\text{ice}}\,\Delta T_{\text{ice}} + m_{\text{ice}}L_{\text{f}} \\ &= (0.20\text{ kg})(2.05\text{ kJ/kg}\cdot\text{K})(10\text{ K}) + (0.20\text{ kg})(334\text{ kJ/kg}) \\ &= 4.1\text{ kJ} + 66.8\text{ kJ} = 70.9\text{ kJ} \end{aligned}$$

Cooling the water to 0°C would extract energy Q_2 given by Equation 16.3:

$$Q_2 = m_{\text{water}}c_{\text{water}}\Delta T_{\text{water}} = (1.0\text{ kg})(4.184\text{ kJ/kg}\cdot\text{K})(15\text{ K}) = 62.8\text{ kJ}$$

This is far more than the 4.1 kJ needed to bring the ice to 0°C, but not quite the 70.9 kJ needed to leave it all melted. So there's enough ice to cool the water to 0°C, with some left over. How much? Our calculation of Q_1 shows that 4.1 kJ go into raising the ice temperature. Of the 62.8 kJ extracted from the water, the remaining 58.7 kJ go to melting ice. From Equation 17.5, the amount of ice melted is then

$$m_{\text{melted}} = \frac{Q}{L_{\text{f}}} = \frac{58.7\text{ kJ}}{334\text{ kJ/kg}} = 0.176\text{ kg} = 176\text{ g}$$

So we're left with 24 g of ice in 1176 g of water, all at 0°C.

ASSESS Make sense? Our 62.8 kJ was nearly enough to bring all the ice to the liquid phase, so it makes sense that only a small fraction of the ice remains.

Phase Diagrams

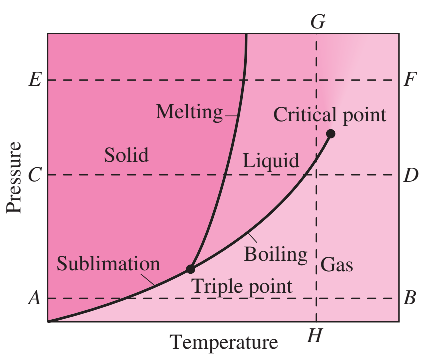


FIGURE 17.6 A phase diagram showing solid, liquid, and gas phases on a plot of pressure versus temperature.

Why can't mountaineers enjoy piping hot coffee? Because water's boiling point drops with the decreasing pressure at high altitudes. In general, the temperatures at which phase changes occur depend on pressure. A **phase diagram** shows the different phases on a plot of pressure versus temperature. Figure 17.6 is a phase diagram for a typical substance. Most phase diagrams are similar, although water's is slightly unusual for reasons we'll discuss in the next section.

The phase diagram divides pressure-temperature space into regions corresponding to solid, liquid, and gas phases. Lines separating these regions mark the phase transitions. Everyday experience suggests that heating takes a substance from solid, to liquid, to gas—as with water in Fig. 17.5. But Fig. 17.6 shows that this sequence doesn't always occur. At low pressure (line *AB* in Fig. 17.6) the substance goes directly from solid to gas. This is **sublimation**. We don't see this with water because normal atmospheric pressure is too high. For carbon dioxide, though, atmospheric pressure is low in the phase diagram, which is why "dry ice" turns directly into gaseous CO_2 without becoming liquid. At higher pressures (line *CD*) we get the familiar solid-liquid-gas sequence. Higher still (line *EF*), we're above the **critical point**, where the abrupt distinction between liquid and gas disappears. Instead, the substance starts out as a thick fluid whose properties change gradually from liquidlike to gaslike as it's heated.

We think of changing phase by applying heat, but Fig. 17.6 shows we can also change phase by changing pressure. Lowering pressure along line *GH*, for example, takes the substance from liquid to gas without any heat input. You may have seen a demonstration of water boiling vigorously at room temperature in a closed container pumped down to low pressure.

Don't let Fig. 17.6 fool you into thinking that phase transitions occur instantaneously. Those heats of transformation are large, and a substance moving, say, along line *CD* in response to heating will linger at each phase transition until all of it has changed phase; that's what the level portions of Fig. 17.5 showed.

The dividing curves in Fig. 17.6 show where two phases can coexist simultaneously, like ice floating in water at 0°C and atmospheric pressure. It's because phase changes occur along curves that terms like "melting point" and "boiling point" are meaningless unless pressure is specified. But there's one unique **triple point** where solid, liquid, and gas all coexist in equilibrium. Here temperature and pressure have unique, unambiguous values—which is why the 273.16-K triple point of water is used to define the kelvin scale.

17.3 Thermal Expansion

We've seen how heating causes changes in temperature and phase. But heating also results in pressure or volume changes. For a gas at constant pressure, for example, the ideal-gas law shows that volume increases in direct proportion to temperature. The volume and pressure relations for liquids and solids aren't so simple. Because their molecules are closely spaced, liquids and solids aren't very compressible, so thermal expansion is less pronounced.

We characterize the change in the volume with temperature using the **coefficient of volume expansion** β, defined as the fractional change in volume when a substance undergoes a small temperature change ΔT:

$$\beta = \frac{\Delta V/V}{\Delta T} \tag{17.6}$$

This equation assumes that β is independent of temperature; if it varies significantly, then we would need to define β in terms of the derivative dV/dT (Problem 68). Our definition of β also assumes constant pressure; we could entirely inhibit thermal expansion with appropriate pressure increases.

Often we want to know how one linear dimension of a solid changes with temperature. This is especially true with long structures, where the absolute change is greatest along the long dimension (Fig. 17.7). We then speak of the **coefficient of linear expansion** α, defined by

$$\alpha = \frac{\Delta L/L}{\Delta T} \tag{17.7}$$

The volume- and linear-expansion coefficients are related in a simple way: $\beta = 3\alpha$, as you can show in Problem 71. However, the linear-expansion coefficient α is really meaningful only with solids, because liquids and gases deform and don't expand proportionately in all directions. Table 17.2 lists the expansion coefficients for some common substances.

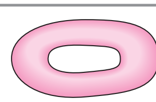
FIGURE 17.7 Thermal expansion distorted these tracks, causing a derailment. Expansion of long structures like this is best described using the coefficient of linear expansion.

Table 17.2 Expansion Coefficients*

Solids	α (K^{-1})	Liquids and Gases	β (K^{-1})
Aluminum	24×10^{-6}	Air	3.7×10^{-3}
Brass	19×10^{-6}	Alcohol, ethyl	75×10^{-5}
Copper	17×10^{-6}	Gasoline	95×10^{-5}
Glass (Pyrex)	3.2×10^{-6}	Mercury	18×10^{-5}
Ice	51×10^{-6}	Water, 1°C	-4.8×10^{-5}
Invar†	0.9×10^{-6}	Water, 20°C	20×10^{-5}
Steel	12×10^{-6}	Water, 50°C	50×10^{-5}

*At approximately room temperature unless noted.
†Invar, consisting of 64% iron and 36% nickel, is an alloy designed to minimize thermal expansion.

GOT IT? 17.3 The figure shows a donut-shaped object. If it's heated, will the hole get (a) larger or (b) smaller?



EXAMPLE 17.5 Thermal Expansion: Spilled Gasoline

A steel gas can holds 20 L at 10°C. It's filled to the brim with gas at 10°C. If the temperature now increases to 25°C, by how much does the can's volume increase? How much gas spills out?

INTERPRET This is a problem about thermal expansion. Since it involves volume, we identify the relevant quantity as the coefficient of volume expansion β.

DEVELOP Equation 17.6, $\beta = (\Delta V/V)/\Delta T$, determines the volume change. Our plan is to calculate the expanded volume of the tank and then of the gasoline. The difference will be the amount that spills out. Table 17.2 lists β for gasoline but α for steel; therefore, we'll use the equation $\beta = 3\alpha$ for the steel.

EVALUATE First we use Equation 17.6 to evaluate the volume change ΔV of the can. Using $\beta = 3\alpha$, we have

$$\Delta V_{\text{can}} = \beta V\,\Delta T = (3)(12\times10^{-6}\,\text{K}^{-1})(20\,\text{L})(15\,\text{K}) = 0.0108\,\text{L}$$

Similarly, for the gasoline,

$$\Delta V_{\text{gas}} = \beta V\,\Delta T = (95\times10^{-5}\,\text{K}^{-1})(20\,\text{L})(15\,\text{K}) = 0.285\,\text{L}$$

We therefore lose 0.275 L.

ASSESS Make sense? The thermal-expansion coefficient for gasoline is so much greater than for steel that the can's expansion is negligible and the gas has nowhere to go. By the way, that spill wastes nearly 10 MJ of energy!

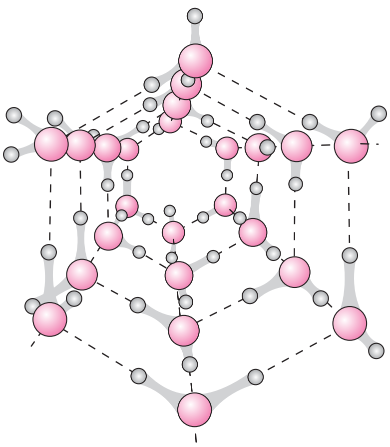


FIGURE 17.8 Water molecules in an ice crystal form an open structure, giving solid water a lower density than the liquid.

Thermal Expansion of Water

The entry for water at 1°C in Table 17.2 is remarkable, the negative expansion coefficient showing that water at this temperature actually *contracts* on heating. This unusual behavior occurs because ice has a relatively open crystal structure (Fig. 17.8) and therefore is less dense than liquid water. That's why ice floats. Immediately above the melting point, the intermolecular forces that bond H_2O molecules in ice still exert an influence, giving cold liquid water a lower density than at slightly higher temperatures. At 4°C water reaches its maximum density, and above this temperature the effect of molecular kinetic energy in keeping molecules apart wins out over intermolecular forces. From there on, water exhibits the more normal behavior of expansion with increasing temperature.

This unusual property of water near its melting point is reflected in its phase diagram, shown in Fig. 17.9. Note that the solid-liquid boundary extends leftward from the triple point, in contrast to the more typical behavior in Fig. 17.6. That means that ice at a fixed temperature will melt if the pressure is *increased*—an unusual property known as pressure melting.

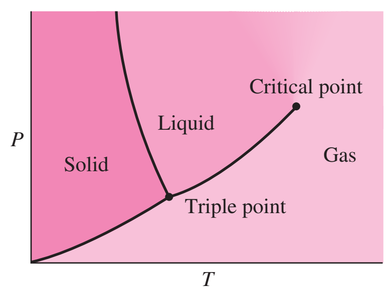


FIGURE 17.9 Phase diagram for water. Compare the solid-liquid boundary with that of Fig. 17.6.

APPLICATION Aquatic Life and Lake Turnover

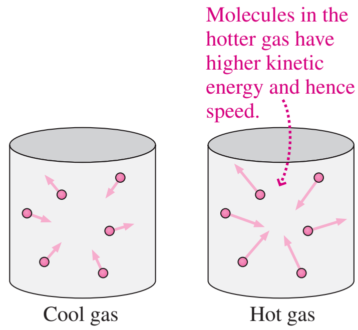
The anomalous behavior of water has important consequences for life. If ice didn't float, then ponds, lakes, and even oceans would freeze solid from the bottom up, making aquatic life impossible. What actually happens, instead, is that a thin layer of ice forms on the surface, insulating the water below and keeping it liquid; as a result, ice cover in temperate climates rarely exceeds a meter or so. Because water's density is greatest at 4°C, water at this temperature sinks to the bottom. At lake depths greater than a few meters, sunlight is inadequate to raise the temperature, which therefore remains year-round at 4°C.

Water's unusual density behavior also causes the twice-yearly turnover of lakes in temperate climates. In the summer, a lake's surface water is warm, but deep water remains at 4°C. In the winter, water just beneath the ice is at 0°C, while the bottom water is still at 4°C. Both situations are stable, with less dense and therefore more buoyant water at the surface. But in the spring, ice melts and the surface water warms. When that water reaches 4°C, there's no density variation and the lake water mixes freely. This is the spring overturning. A similar overturning occurs in the fall, as the surface water cools through 4°C. Turnover is important to aquatic life because it brings up nutrients that would otherwise be trapped in the deep water.

CHAPTER 17 SUMMARY

Big Picture

The big idea here is that matter responds to heating in a variety of ways in addition to changing temperature. Other responses include changes of phase and of volume and/or pressure. The ideal gas provides a particularly simple system for understanding volume and pressure changes. Analyzing ideal-gas behavior provides a link between the Newtonian mechanics of molecules and macroscopic thermodynamics, showing that temperature is a measure of the average molecular kinetic energy.

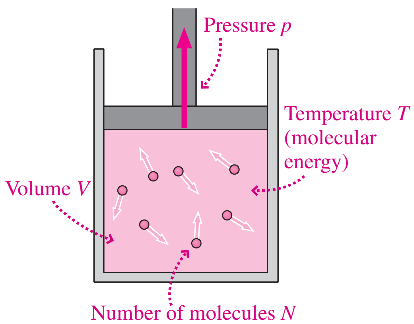


Key Concepts and Equations

The **ideal-gas law** relates pressure, volume, temperature, and the number of molecules in a gas:

$$pV = NkT \quad \text{(ideal-gas law)}$$

where **Boltzmann's constant** $k = 1.38\times10^{-23}$ J/K.



In terms of the number of moles n, the ideal-gas law is

$$pV = nN_AkT = nRT$$

where the **universal gas constant** $R = N_Ak = 8.314$ J/K·mol.

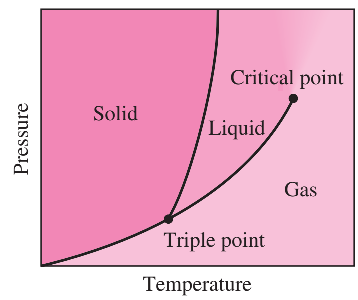
The temperature of an ideal gas is a measure of the gas molecules' average kinetic energy:

$$\tfrac{1}{2}m\overline{v^2} = \tfrac{3}{2}kT \quad \text{(temperature and molecular energy)}$$

Heats of transformation L describe the energy per unit mass needed to effect phase changes. The total energy required to change the phase of a mass m is given by

$$Q = Lm \quad \text{(heat of transformation)}$$

Phase diagrams plot solid, liquid, and gas phases against temperature and pressure, and reveal the **triple point**, where all three phases can coexist, and the **critical point**, where the liquid-gas distinction disappears.



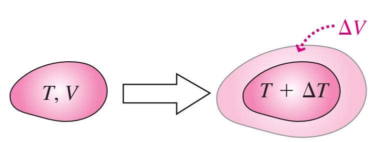
Applications

Thermal expansion is characterized by the **coefficient of volume expansion** and its linear counterpart. The volume-expansion coefficient relates the fractional volume change $\Delta V/V$ to the temperature change ΔT:

$$\beta = \frac{\Delta V/V}{\Delta T} \quad \text{(volume-expansion coefficient)}$$

while the **coefficient of linear expansion** relates the fractional length $\Delta L/L$ change to ΔT:

$$\alpha = \frac{\Delta L/L}{\Delta T} \quad \text{(linear-expansion coefficient)}$$



For Thought and Discussion

1. If the volume of an ideal gas is increased, must the pressure drop proportionately? Explain.
2. According to the ideal-gas law, what should be the volume of a gas at absolute zero? Why is this result absurd?
3. Why are you supposed to check tire pressure when your tires are cold?
4. The average *speed* of the molecules in a gas increases with increasing temperature. What about the average *velocity*?
5. Suppose you start running while holding a closed jar of air. Do you change the average speed of the air molecules? The average velocity? The temperature?
6. Two different gases are at the same temperature, and both have low enough densities that they behave like ideal gases. Do their molecules have the same thermal speeds? Explain.
7. Your roommate claims that ice and snow must be at 0°C. Is that true?
8. What's the temperature of water just under the ice layer of a frozen lake? At the bottom of a deep lake?
9. Ice and water have been together in a glass for a long time. Is the water hotter than the ice?
10. Which takes more heat: melting a gram of ice already at 0°C, or bringing the melted water to the boiling point?
11. Why does removing the plastic wrap from a package of frozen hamburger help it thaw faster?
12. Why do we use the triple point of water for thermometer calibration? Why not just use the melting point or boiling point?
13. How is it possible to have boiling water at a temperature other than 100°C?
14. How does a pressure cooker work?
15. Suppose mercury and glass had the same coefficient of volume expansion. Could you build a mercury thermometer?
16. A bimetallic strip consists of thin pieces of brass and steel bonded together (Fig. 17.10). What happens when the strip is heated? (*Hint:* Consult Table 17.2.)

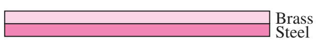


FIGURE 17.10 For Thought and Discussion 16

Exercises and Problems

Exercises

Section 17.1 Gases

17. Mars's atmospheric pressure is about 1% that of Earth, and its average temperature is around 215 K. Find the volume of 1 mol of the Martian atmosphere.
18. How many molecules are in an ideal-gas sample at 350 K that occupies 8.5 L when the pressure is 180 kPa?
19. What's the pressure of an ideal gas if 3.5 mol occupy 2.0 L at −150°C?
20. Your professor asks you to order a tank of argon gas for a lab experiment. You obtain a "type C" gas cylinder with interior volume 6.88 L. The supplier claims it contains 45 mol of argon. You measure its pressure to be 14 MPa at room temperature (20°C). Did you get what you paid for?
21. (a) If 2.0 mol of an ideal gas are initially at temperature 250 K and pressure 1.5 atm, what's the gas volume? (b) The pressure is now increased to 4.0 atm, and the gas volume drops to half its initial value. What's the new temperature?
22. A pressure of 10^{-10} Pa is readily achievable with laboratory vacuum apparatus. If the residual air in this "vacuum" is at 0°C, how many air molecules are in 1 L?
23. What's the thermal speed of hydrogen molecules at 800 K?
24. In which gas are the molecules moving faster: hydrogen at 75 K or sulfur dioxide at 350 K?

Section 17.2 Phase Changes

25. How much energy does it take to melt a 65-g ice cube?
26. It takes 200 J to melt an 8.0-g sample of one of the substances in Table 17.1. What's the substance?
27. If it takes 840 kJ to vaporize a sample of liquid oxygen, how large is the sample?
28. Carbon dioxide *sublimes* (changes from solid to gas) at 195 K. The heat of sublimation is 573 kJ/kg. How much heat must be extracted from 250 g of CO_2 gas at 195 K in order to solidify it?
29. Find the energy needed to convert 28 kg of liquid oxygen at its boiling point into gas.

Section 17.3 Thermal Expansion

30. A copper wire is 20 m long on a winter day when the temperature is −12°C. By how much does its length increase on a 26°C summer day?
31. You have exactly 1 L of ethyl alcohol at room temperature (20°C). You put it in a refrigerator at 2°C. What's its new volume?
32. A Pyrex glass marble is 1.00000 cm in diameter at 20°C. What will be its diameter at 85°C?
33. At 0°C, the hole in a steel washer is 9.52 mm in diameter. To what temperature must it be heated in order to fit over a 9.55-mm-diameter bolt?
34. Suppose a single piece of welded steel railroad track stretched 5000 km across the continental United States. If the track were free to expand, by how much would its length change if the entire track went from a cold winter temperature of −25°C to a hot summer day at 40°C?

Problems

35. The solar corona is a hot (2 MK) extended atmosphere surrounding the Sun's cooler visible surface. The coronal gas pressure is about 0.03 Pa. What's the coronal density in particles per cubic meter? Compare with Earth's atmosphere.
36. A helium balloon occupies 8.0 L at 20°C and 1.0-atm pressure. The balloon rises to an altitude where the air pressure is 0.65 atm and the temperature is −10°C. What is its volume when it reaches equilibrium at the new altitude?
37. A compressed air cylinder stands 100 cm tall and has internal diameter 20.0 cm. At room temperature, the pressure is 180 atm. (a) How many moles of air are in the cylinder? (b) What volume would this air occupy at 1.0 atm and room temperature?
38. You're a lawyer with an unusual case. A whipped-cream can burst at a wedding, damaging the groom's expensive tuxedo. The can warned against temperatures in excess of 50°C, and the manufacturer has evidence that it reached 60°C. You don't contest this, but you point out that the can was only half full when it burst, meaning that the gas propellant had more than twice the volume it would in a full can, and that some of the propellant had already been used. You argue that the real safety criterion is pressure, and that the can's maximum pressure wasn't exceeded. Who's right?

39. A 3000-mL flask is initially open in a room containing air at 1.00 atm and 20°C. The flask is then closed and immersed in boiling water. When the air in the flask has reached thermodynamic equilibrium, the flask is opened and air is allowed to escape. The flask is then closed and cooled back to 20°C. Find (a) the maximum pressure reached in the flask, (b) the number of moles that escape when air is released, and (c) the final pressure in the flask.
40. **BIO** The recommended treatment for frostbite is rapid heating in a water bath. Suppose a frostbitten hand with mass 120 g is immersed in water that conducts energy into the hand at the rate of 800 W. Treating the hand as essentially water, initially frozen solid, how long will it take for it to thaw and return to body temperature $(37°\mathrm{C})$?
41. A stove burner supplies heat to a pan at the rate of 1500 W. How long will it take to boil away 1.1 kg of water, once the water reaches its boiling point?
42. If a 1-megaton nuclear bomb were exploded deep in the Greenland ice cap, how much ice would it melt? Assume the ice is initially at about its freezing point, and consult Appendix C for the appropriate energy conversion.
43. You're winter camping and are melting snow for drinking water. The snow temperature is right around 0°C. You set a pot containing 5.0 kg of snow on your campfire, and you keep stoking up the fire. As a result, the snow gains energy at an increasing rate: $P = a + bt$, where $a = 1.1$ kW, $b = 2.3$ W/s, and t is the time in s. To the nearest minute, how long will it take to melt the snow?
44. At winter's end, Lake Superior's surface is frozen to a depth of 1.3 m; the ice density is 917 kg/m^3. (a) How much energy does it take to melt the ice? (b) If the ice disappears in 3 weeks, what's the average power supplied to melt it?
45. A refrigerator extracts energy from its contents at the rate of 95 W. How long will it take to freeze 750 g of water already at 0°C?
46. Climatologists have recently recognized that black carbon (soot) from burning fossil fuels and biomass contributes significantly to arctic warming. You're asked to determine whether this effect might cause ice to melt that would normally stay frozen year-round. Consider an ice layer 2.5 m thick that normally reflects 90% of the incident solar energy and absorbs the rest. Suppose black carbon darkens the ice so it now reflects only 50% of the incident solar energy. The arctic summertime solar input averages 300 W/m^2. You can assume 0°C for the initial ice temperature, and an ice density of 917 kg/m^3. What do you conclude?
47. Repeat Example 17.4 with an initial ice mass of 50 g.
48. How much energy does it take to melt 10 kg of ice initially at $-10°\mathrm{C}$?
49. Water is brought to its boiling point and then allowed to boil away completely. If the energy needed to raise the water to the boiling point is one-tenth of that needed to boil it away, what was the initial temperature?
50. During a nuclear accident, 420 m^3 of emergency cooling water at 20°C are injected into a reactor vessel where the reactor core is producing heat at the rate of 200 MW. If the water is allowed to boil at normal atmospheric pressure, how long will it take to boil the reactor dry?
51. What's the minimum amount of ice in Example 17.4 that will ensure a final temperature of 0°C?
52. A bowl contains 16 kg of punch (essentially water) at a warm 25°C. What's the minimum amount of ice at 0°C needed to cool the punch to 0°C?
53. A 50-g ice cube at $-10°\mathrm{C}$ is placed in an equal mass of water. What must the initial water temperature be if the final mixture still contains equal amounts of ice and water?
54. **BIO** Evaporation of sweat is the human body's cooling mechanism. At body temperature, it takes 2.4 MJ/kg to evaporate water. Marathon runners typically lose about 3 L of sweat each hour. How much energy gets lost to sweating during a 3-hour marathon?
55. What power is needed to melt 20 kg of ice in 6.0 min?
56. You put 300 g of water at 20°C into a 500-W microwave oven and accidentally set the time for 20 min instead of 2.0 min. How much is left at the end of 20 min?
57. If 4.5×10^5 kg of emergency cooling water at 10°C are dumped into a malfunctioning nuclear reactor whose core is producing energy at the rate of 200 MW, and if no circulation or cooling occurs, how long will it take for half the water to boil away?
58. Describe the composition and temperature of the equilibrium mixture after 1.0 kg of ice at $-40°\mathrm{C}$ is added to 1.0 kg of water at 5.0°C.
59. A glass marble 1.000 cm in diameter is to be dropped through a hole in a steel plate. At room temperature the hole diameter is 0.997 cm. By how much must the plate's temperature be raised so the marble will fit through the hole?
60. A 2000-mL graduated cylinder is filled with liquid at 350 K. When the liquid is cooled to 300 K, the cylinder is full to only the 1925-mL mark. Use Table 17.2 to identify the liquid.
61. A steel ball bearing is encased in a Pyrex glass cube 1.0 cm on a side. At 330 K, the ball bearing fits tightly inside the cube. At what temperature will it have a clearance of 1.0 μm all around?
62. Fuel systems of modern cars are designed so thermal expansion of gasoline doesn't result in wasteful and polluting fuel spills. As an engineer, you're asked to specify the size of an expansion tank that will handle this overflow. You know that gasoline comes from its underground tank at 10°C, and your tank must handle the expansion of a full 75-L gas tank when the gas reaches a hot summer day's temperature of 35°C. How large an expansion tank do you specify?
63. A rod of length L_0 is clamped rigidly at both ends. Its temperature increases by ΔT and in the ensuing expansion, it cracks to form two straight pieces, as shown in Fig. 17.11. Find an expression for the distance d shown in the figure, in terms of L_0, ΔT, and the linear expansion coefficient α.

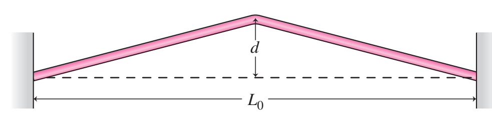


FIGURE 17.11 Problem 63

64. You're home from college on vacation, and there's a power failure. The power company says it will be 15 hours before it's repaired. Your parents send you out to buy ice to keep the 'fridge cold. You look up the thermal resistance of the refrigerator's walls; it's 0.12 K/W. If room temperature is 20°C, how much ice should you buy?
65. A solar-heated house stores energy in 5.0 tons of Glauber salt $(Na_2SO_4 \cdot 10H_2O)$, which melts at 90°F. The heat of fusion of Glauber salt is 104 Btu/lb and the specific heats of the solid and liquid are, respectively, 0.46 Btu/lb·°F and 0.68 Btu/lb·°F. After a week of sunny weather, the storage medium is all liquid at 95°F. Then comes a cloudy period during which the house loses heat at an average of 20,000 Btu/h. (a) How long is it before the temperature of the storage medium drops below 60°F? (b) How much of this time is spent at 90°F?

66. Show that the coefficient of volume expansion of an ideal gas at constant pressure is the reciprocal of its kelvin temperature.
67. Water's coefficient of volume expansion in the temperature range from 0°C to about 20°C is given approximately by $\beta = a + bT + cT^2$, where T is in Celsius and $a = -6.43\times 10^{-5}\,°\text{C}^{-1}$, $b = 1.70\times10^{-5}\,°\text{C}^{-2}$, and $c = -2.02\times10^{-7}\,°\text{C}^{-3}$. Show that water has its greatest density at approximately 4.0°C.
68. When the expansion coefficient varies with temperature, Equation 17.6 is written $\beta = (1/V)(dV/dT)$. If a sample of water occupies 1.00000 L at 0°C, find its volume at 12°C. (*Hint*: Use the information from Problem 67, and integrate the equation above.)
69. Ignoring air resistance, find the height from which to drop an ice cube at 0°C so it melts completely on impact. Assume no heat exchange with the environment.
70. The timekeeping of an old clock is regulated by a brass pendulum 20.0 cm long. If the clock is accurate at 20°C but is in a room at 18°C, how soon will the clock be off by 1 minute? Will it be fast or slow?
71. Prove the equation $\beta = 3\alpha$ (Section 17.3) by considering a cube of side s and therefore volume $V = s^3$ that undergoes a small temperature change dT and corresponding length and volume changes ds and dV.
72. You're on a team planning a mission to Venus to collect atmospheric samples for analysis. The design specs call for a 1-L sample container, while the scientists want at least 1 mol of gas. Venus's atmospheric pressure is 90 times that of Earth, and its average temperature is 730 K. Will the design work?

Passage Problems

A *pressure cooker* is a sealed pot that cooks food much faster than most other methods because increased pressure allows water to reach higher temperatures than the normal boiling point (Fig. 17.12). Pressure cookers afford many advantages: faster cooking, lower energy consumption, and less vitamin loss. The pressure-cooker principle is also used in autoclaves for sterilizing surgical instruments in hospitals and equipment in biology labs.

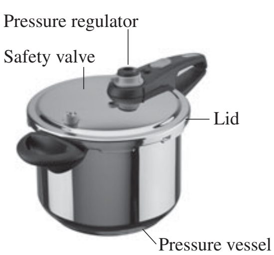


FIGURE 17.12 A pressure cooker (Passage Problems 73–76)

73. In water's phase diagram (Fig. 17.9), normal boiling occurs at a point on the line between the triple point and the critical point. In a pressure cooker, boiling occurs
    a. at a point in the diagram directly above where it normally occurs.
    b. higher up on the line between the triple and critical points.
    c. at a point directly to the right of where it normally occurs.
    d. beyond the critical point.
74. A typical pressure cooker operates at twice normal atmospheric pressure, raising water's boiling point to about 120°C. Compared with steam at 1 atm and the normal 100°C boiling point, the density of steam in a pressure cooker is
    a. double.
    b. somewhat more than double.
    c. somewhat less than double.
    d. quadruple.
75. Because some pathogens can survive 120°C temperatures, medical autoclaves typically operate at 3 atm pressure, where water boils at 134°C. Based on this information and that given in the preceding problem, you can conclude that
    a. Fig. 17.9's depiction of the liquid-gas interface for water is correct in being concave upward.
    b. Fig. 17.9's liquid-gas interface should actually be concave downward.
    c. autoclaves operate above the critical point.
    d. at its operating temperature, there can't be any liquid water in the autoclave.
76. A pressure cooker has a regulating mechanism that releases steam so as to maintain constant pressure. If that mechanism became clogged
    a. the pressure would nevertheless level off once water in the cooker began to boil.
    b. the pressure would continue to rise although the temperature would remain constant.
    c. both temperature and pressure would continue to rise.
    d. the density of the steam would decrease.

Answers to Chapter Questions

Answer to Chapter Opening Question

Water's solid phase is less dense than the liquid, which causes ice to float. Our world would be a very different place if ice were denser than water.

Answers to GOT IT? Questions

17.1. It increases by a factor of $\sqrt{2}$.
17.2. (c) Equal to 100°C.
17.3. (a) Larger, since all linear dimensions of the object expand equally.

18 Heat, Work, and the First Law of Thermodynamics

A jet engine converts the energy of burning fuel into mechanical energy. How does energy conservation apply in this process?

New Concepts, New Skills

By the end of this chapter you should be able to

- Explain how the first law of thermodynamics extends the conservation-of-energy principle to include thermal energy (18.1).
- Describe quantitatively the effects of basic thermodynamic processes—isothermal, adiabatic, isobaric, and constant-volume—on an ideal gas (18.2).
- Determine the specific heat of an ideal gas based on its molecular structure (18.3).

Connecting Your Knowledge

- This chapter joins the concept of work, introduced in Chapter 6 (6.1, 6.2), and the behavior of ideal gases as described in Chapter 17 (17.1).
- You should have a clear understanding of the conservation-of-energy principle introduced in Chapter 7 (7.3).

In Chapter 7 we introduced the powerful idea of energy conservation, limited to mechanical energy and conservative forces. Now we've learned that thermal processes involve energy—and that sets the stage for extending the conservation-of-energy principle. Here we'll explore this broader principle and see how it describes energy interchanges in thermodynamic systems ranging from engines to Earth's atmosphere.

18.1 The First Law of Thermodynamics

Figure 18.1 shows two ways to raise the temperature in a beaker of water: by heating with a flame and by stirring vigorously with a spoon. Using the flame involves heat—energy in transit because of the temperature difference between flame and water. But there's no temperature difference between spoon and water; here the energy transfer occurs because the spoon does mechanical work on the water. We already know that doing work can increase the kinetic or potential energy of a macroscopic object; here we see it, instead, changing the **internal energy** associated ultimately with individual molecules. The point is that both processes—heating and mechanical work—result in exactly the same final state—namely, water with a higher temperature and therefore greater internal energy. It's this common result that made possible Joule's quantitative identification of heat as a form of energy (Fig. 18.2).

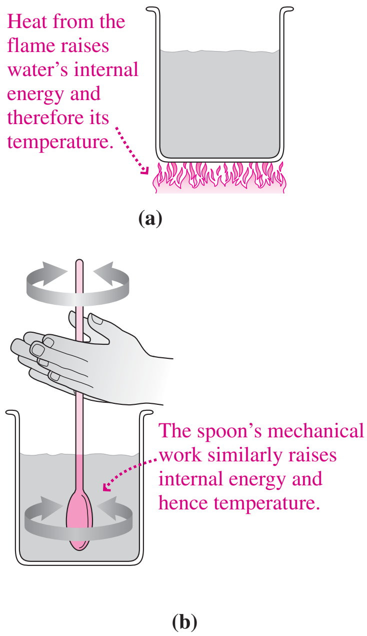


FIGURE 18.1 Two ways to raise temperature.

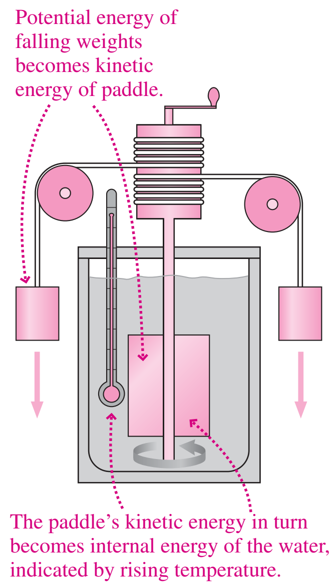


FIGURE 18.2 Joule's apparatus for determining what he called "the mechanical equivalent of heat."

Keep track of all the energy entering and leaving a system—both heat and work—and you'll find that the change in the system's internal energy depends only on the net energy transferred. In one sense this is hardly surprising; it just extends the idea of energy conservation to include heat. But in another way it's remarkable; it doesn't matter at all *how* the energy gets into the system—heat, work, or some combination of the two. This statement constitutes the **first law of thermodynamics**:

> **First law of thermodynamics** The change in the internal energy of a system depends only on the net heat transferred to the system and the net work done on the system, independent of the particular processes involved.

Mathematically, the first law is

$$\Delta U = Q + W \quad \text{(first law of thermodynamics)} \tag{18.1}$$

where ΔU is the change in a system's internal energy, Q the heat transferred to the system, and W the work done on the system.* The first law says that the change in a system's internal energy doesn't depend on how the energy gets transferred, but only on the net energy. Internal energy is therefore a **thermodynamic state variable**, meaning a quantity whose value doesn't depend on how a system got into its particular state. Temperature and pressure are also thermodynamic state variables; heat and work are not.

We're frequently concerned with *rates* of energy flow. Differentiating the first law with respect to time gives a statement about rates:

$$\frac{dU}{dt} = \frac{dQ}{dt} + \frac{dW}{dt} \tag{18.2}$$

where dU/dt is the rate of change of a system's internal energy, dQ/dt the rate of heat transfer to the system, and dW/dt the rate at which work is done on the system.

---

*Some books define W as the work done *by* the system, in which case there's a minus sign in the first law. This is because the law was first introduced in connection with engines, which take in heat and put out mechanical work.

EXAMPLE 18.1 **The First Law of Thermodynamics: Thermal Pollution**

The reactor in a nuclear power plant supplies energy at the rate of 3.0 GW, boiling water to produce steam that turns a turbine-generator. The spent steam is then condensed through thermal contact with water taken from a river. If the power plant produces electrical energy at the rate of 1.0 GW, at what rate is heat transferred to the river?

INTERPRET This problem is about heat and mechanical energy, which are related by the first law of thermodynamics. We identify the system as the entire power plant, comprising the nuclear reactor, including its fuel, and the turbine-generator. We identify U as the internal energy stored in the fuel, W as the mechanical work that ends up as electrical energy, and Q as the heat transferred to the river.

DEVELOP Since we're dealing here with *rates*, Equation 18.2, $dU/dt = dQ/dt + dW/dt$, applies. The reactor extracts internal energy from its fuel, so the rate dU/dt is negative. The power plant delivers electrical energy to the outside world, so it's *doing* work; since W in the first law is the work done *on* the system, dW/dt is therefore *negative*. Our plan is then to solve for dQ/dt, the rate of energy transfer to the river.

EVALUATE Solving, we have

$$\frac{dQ}{dt} = \frac{dU}{dt} - \frac{dW}{dt} = -3.0\text{ GW} - (-1.0\text{ GW}) = -2.0\text{ GW}$$

ASSESS Make sense? Since positive Q represents heat transferred *to* the system, the minus sign shows that heat is transferred *from* the power plant to the river at the rate of 2 GW. The numbers here are typical for large nuclear and coal-burning power plants, and show that about two-thirds of the energy extracted from the fuel is wasted in heating the environment. We'll see in the next chapter just why this waste occurs.

✓TIP Identify the System

The first law of thermodynamics deals with energy flows into and out of a system. It's up to you to define the system, and how you do so affects the meanings of the terms in the first law. In this Example we included the nuclear reactor, with the internal energy of its fuel, as part of the system. If we had considered only the turbine-generator, then we would have had 3 GW of heat coming in from the reactor and no change in internal energy. But the result would be the same: 1 GW going out as electricity and 2 GW of heat dumped into the river.

18.2 Thermodynamic Processes

Although the first law applies to *any* system, it's easiest to understand when applied to an ideal gas. The ideal-gas law relates the temperature, pressure, and volume of a given gas sample: $pV = nRT$. The thermodynamic state is completely determined by any two of the quantities p, V, or T. We'll find it convenient to represent different states as points on a ***pV* diagram**—a graph whose vertical and horizontal axes represent pressure and volume, respectively.

Reversible and Irreversible Processes

Imagine a gas sample immersed in a large reservoir of water and allowed to come to equilibrium (Fig. 18.3). If we then raise the reservoir temperature very slowly, both water and gas temperatures will rise essentially in unison, and the gas will remain in equilibrium. Such a slow change is called a **quasi-static process**. Because a system undergoing a quasi-static process is always in thermodynamic equilibrium, its evolution from one state to another is described by a continuous sequence of points—a curve—in its pV diagram (Fig. 18.4).

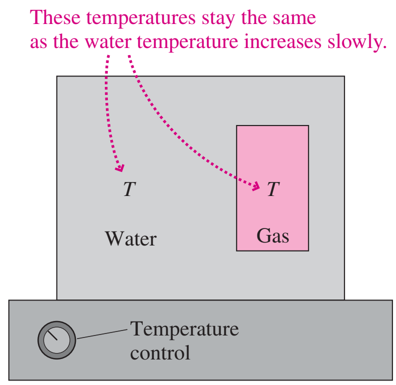


FIGURE 18.3 A quasi-static, or reversible, process keeps water and gas always in equilibrium.

We could reverse this heating process by slowly lowering the reservoir temperature; the gas would cool, reversing its path in the pV diagram. For that reason, a quasi-static process is also a **reversible process**. A process like suddenly plunging a cool gas sample into hot water is, in contrast, **irreversible**. During an irreversible process the system isn't in equilibrium, and thermodynamic variables like temperature and pressure don't have well-defined values. It therefore makes no sense to think of a path in the pV diagram. A process may be irreversible even though it returns a system to its original state. The distinction lies not in the end states but in the *process* that takes the system between states.

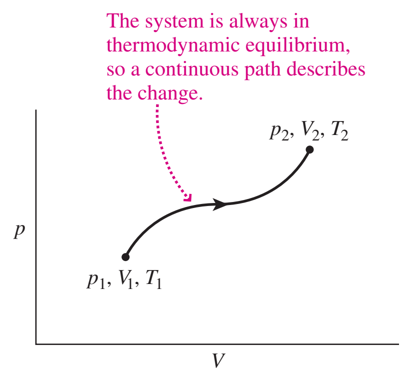


FIGURE 18.4 The pV diagram of a system undergoing quasi-static change.

There are many ways to change a system's thermodynamic state. Here we consider important special cases involving an ideal gas. These illustrate the physical principles behind a myriad of technological devices and natural phenomena, from the operation of a gasoline engine to the propagation of a sound wave to the oscillations of a star.

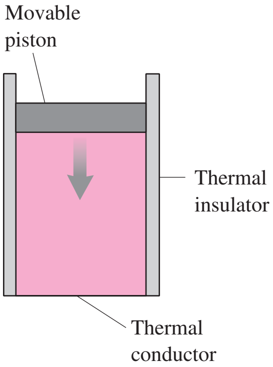


FIGURE 18.5 A gas-cylinder system with insulating walls and a conducting bottom.

Our system consists of an ideal gas confined to a cylinder sealed with a movable piston (Fig. 18.5). The piston and cylinder walls are perfectly insulating—they block all heat transfer—and the bottom is a perfect conductor of heat. We can change the thermodynamic state of the gas mechanically by moving the piston, or thermally by transferring heat through the bottom. We'll consider only reversible processes, which we can describe by paths in the pV diagram for the gas.

Work and Volume Changes

We begin by developing an expression for the work done on a gas that holds for all processes. If our piston-cylinder system has cross-sectional area A and gas pressure p, then $F_{gas} = pA$ is the force the gas exerts on the piston. If the piston moves a small distance Δx, the gas does work $\Delta W_{gas} = F_{gas}\,\Delta x = pA\,\Delta x = p\,\Delta V$, where $\Delta V = A\,\Delta x$ is the change in gas volume (Fig. 18.6*a*). Our expression for the first law of thermodynamics involves the work done *on* the gas; by Newton's third law, the piston exerts a force on the gas that's equal but opposite to F_{gas}, so the work done *on* the gas is $\Delta W = -F_{gas}\,\Delta x = -p\,\Delta V$. Pressure may vary with volume, so we find the total work done as the gas goes from volume V_1 to volume V_2 by replacing ΔV with the infinitesimal quantity dV and integrating:

$$W = \int dW = -\int_{V_1}^{V_2} p\,dV \quad \text{(work done on gas during volume change)} \tag{18.3}$$

Figure 18.6*b* shows that the work done on the gas is the negative of the area under the pV curve. That work is positive if the gas is compressed $(V_2 < V_1)$ and negative if it expands $(V_2 > V_1)$.

We'll now explore several basic thermodynamic processes, in each case holding one thermodynamic variable constant.

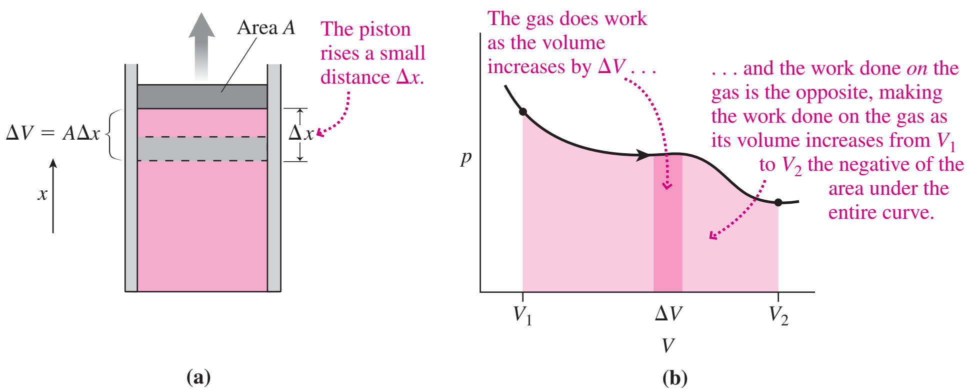


FIGURE 18.6 Work done *on* the gas as the piston rises is the negative of the area under the pV curve.

GOT IT? 18.1 Two identical gas-cylinder systems are taken from the same initial state to the same final state, but by different processes. Which are the same in both cases: (a) the work done on or by the gas; (b) the heat added or removed; or (c) the change in internal energy?

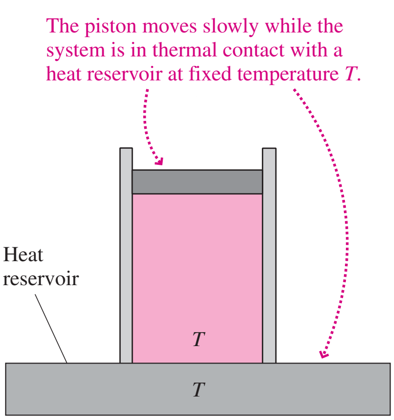


FIGURE 18.7 An isothermal process.

Isothermal Processes

An **isothermal process** occurs at constant temperature. Figure 18.7 shows one way to effect an isothermal process: Place a gas cylinder in thermal contact with a heat reservoir whose temperature is constant. Then move the piston to change the gas volume, slowly enough that the gas remains in equilibrium with the heat reservoir. The system moves from its initial state to its final state along a curve of constant temperature—an **isotherm**—in the pV diagram

(Fig. 18.8). The work done on the gas is given by Equation 18.3 and is the negative of the area under the isotherm.

To find that work, we relate pressure and volume through the ideal-gas law: $p = (nRT)/V$. Then Equation 18.3 becomes

$$W = -\int_{V_1}^{V_2} \frac{nRT}{V}\,dV$$

For an isothermal process, the temperature T is constant, giving

$$W = -nRT\int_{V_1}^{V_2}\frac{dV}{V} = -nRT\ \ln V\Big|_{V_1}^{V_2} = -nRT\ \ln\left(\frac{V_2}{V_1}\right)$$

The internal energy of an ideal gas consists only of the kinetic energy of its molecules, which, in turn, depends only on temperature. Thus, there's no change in the internal energy of an ideal gas during an isothermal process. The first law of thermodynamics then gives $\Delta U = 0 = Q + W$, so

$$Q = -W = nRT\ \ln\left(\frac{V_2}{V_1}\right) \quad \text{(isothermal process)} \tag{18.4}$$

Does this result $Q = -W$ make sense? Recall that Q is the heat transferred to the gas and W is the work done on it. So $-W$ is the work done *by* the gas, and our result shows that for a gas to do work without its temperature changing, it must absorb an equal amount of heat. Similarly, if work is done on the gas, it must transfer an equal amount of heat to its surroundings if it's to maintain a constant temperature.

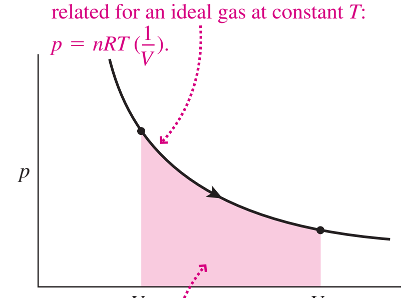


FIGURE 18.8 A pV diagram for an isothermal process.

EXAMPLE 18.2 An Isothermal Process: Bubbles!

A scuba diver is 25 m down, where the pressure is 3.5 atm or about 350 kPa. The air she exhales forms bubbles 8.0 mm in radius. How much work does each bubble do as it rises to the surface, assuming the bubbles remain at the uniform 300 K temperature of the water?

INTERPRET The constant 300 K temperature tells us we're dealing with an isothermal process.

DEVELOP Equation 18.4 determines the work: $-W = nRT\ln(V_2/V_1)$. Here $-W$ is just what we're after: the work done *by* the gas in the bubble. To use this equation, we need the quantity nRT and the volume ratio V_2/V_1. We know p and V (actually the radius, from which we can get V) at the 25-m depth, so we can use the ideal-gas law $pV = nRT$ to get nRT and also the bubble volume just before it reaches the surface. Then we'll have everything we need to apply Equation 18.4.

EVALUATE The ideal-gas law gives $nRT = pV = \frac{4}{3}\pi r^3 p$. The number of moles n doesn't change and R is a constant, so pV is itself constant in the isothermal process. That means $p_1V_1 = p_2V_2$, showing that the volume expands by a factor of 3.5 as the pressure drops from 3.5 atm to 1 atm at the surface—so $V_2/V_1 = 3.5$. Then Equation 18.4 gives

$$-W = nRT\ \ln\left(\frac{V_2}{V_1}\right) = \tfrac{4}{3}\pi r^3 p\ \ln 3.5$$

Using the 8-mm bubble radius and the 350-kPa pressure gives 0.94 J for the work. Note that we needed to use pressure in SI units here; to find the volume ratio, any units would do because V_2/V_1 followed from the pressure *ratio* p_1/p_2.

ASSESS Make sense? The work $-W$ done *by* the gas is positive because an expanding bubble pushes water outward and ultimately upward. It therefore raises the ocean's gravitational potential energy. When the bubble breaks, this excess potential energy becomes kinetic energy, appearing as small waves on the water surface. The bubble, in turn, gets its energy from heat that flows in to keep it at constant temperature. Energy is conserved! ■

Constant-Volume Processes and Specific Heat

A **constant-volume process** (also called isometric, isochoric, or isovolumic) occurs in a rigid closed container whose volume can't change. We could tightly clamp the piston in Fig. 18.5 for a constant-volume process. Because the piston doesn't move, the gas does no work, and the first law becomes simply $\Delta U = Q$. To express this result in terms of a temperature change ΔT, we introduce the **molar specific heat at constant volume** C_V, defined by

$$Q = nC_V\,\Delta T \quad \text{(constant-volume process)} \tag{18.5}$$

where n is the number of moles. This molar specific heat is like the specific heat defined in Chapter 16, except it's per mole rather than per unit mass. Using Equation 18.5 for Q in the statement $\Delta U = Q$ gives

$$\Delta U = nC_V\,\Delta T \quad \text{(any process)} \tag{18.6}$$

For an ideal gas, the internal energy is a function of temperature alone, so $\Delta U/\Delta T$ has the same value no matter what process the gas undergoes. Therefore, Equation 18.6, relating the temperature change ΔT and internal-energy change ΔU, applies not only to a constant-volume process but to *any* ideal-gas process. Why, then, have we been so careful to label C_V the specific heat *at constant volume*? Although Equation 18.6, $\Delta U = nC_V\,\Delta T$, holds for any process, it's only when there's no work that the first law lets us write $Q = \Delta U$, and therefore only for a constant-volume process that Equation 18.5 holds.

Isobaric Processes and Specific Heat

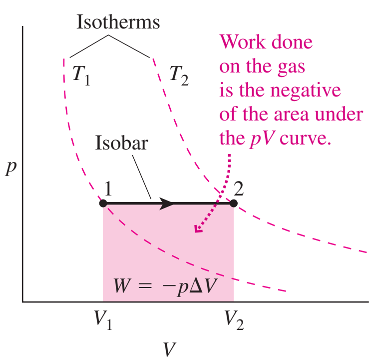


FIGURE 18.9 A pV diagram for an isobaric process; also shown are isotherms for the initial and final temperatures.

Isobaric means constant pressure. Processes occurring in systems exposed to the atmosphere are essentially isobaric. In a reversible isobaric process, a system moves along an isobar, or curve of constant pressure, in its pV diagram (Fig. 18.9). The work done on the gas as the volume changes from V_1 to V_2 is the negative of the area under the isobar, or

$$W = -p(V_2 - V_1) = -p\,\Delta V \tag{18.7}$$

a result we could obtain formally by integrating Equation 18.3.

Solving the first law (Equation 18.1) for Q and using our expression for work give $Q = \Delta U - W = \Delta U + p\,\Delta V$. For an ideal gas, we've just found that the change in internal energy is $\Delta U = nC_V\Delta T$ for *any* process. Therefore, $Q = nC_V\,\Delta T + p\,\Delta V$ for an ideal gas undergoing an isobaric process. We define the **molar specific heat at constant pressure** C_p as the heat required to raise 1 mol of gas by 1 K at constant pressure, or $Q = nC_p\,\Delta T$. Equating our two expressions for Q gives

$$nC_p\,\Delta T = nC_V\,\Delta T + p\,\Delta V \quad \text{(isobaric process)} \tag{18.8}$$

This is a useful form for calculating temperature changes in an isobaric process if we know both specific heats C_p and C_V. However, we really need only one of these specific heats because a simple relation holds between the two. The ideal-gas law, $pV = nRT$, allows us to write $p\,\Delta V = nR\,\Delta T$ for an isobaric process. Using this expression in Equation 18.8 gives $nC_p\,\Delta T = nC_V\,\Delta T + nR\,\Delta T$, so

$$C_p = C_V + R \quad \text{(molar specific heats)} \tag{18.9}$$

Does this make sense? Specific heat measures the heat needed to cause a given temperature change. In a constant-volume process, no work is done and all the heat goes into raising the internal energy and thus the temperature of an ideal gas. In a constant-pressure process, work *is* done and some of the added heat ends up as mechanical energy, leaving less available for raising the temperature. Therefore, a constant-pressure process requires *more* heat for a given temperature change. Thus the specific heat at constant pressure is greater than at constant volume, as reflected in Equation 18.9.

Why didn't we distinguish specific heats at constant volume and constant pressure earlier? Because we were concerned mostly with solids and liquids, whose coefficients of expansion are far lower than those of gases. As a result, much less work is done by a solid or liquid than by a gas. Since work is what gives rise to the difference between C_V and C_p, the distinction is less significant for solids and liquids. As a practical matter, measured specific heats are usually at constant pressure.

Adiabatic Processes

In an **adiabatic process**, no heat flows between a system and its environment. The way to achieve this is to surround the system with perfect thermal insulation. Even without insulation, processes that occur quickly are often approximately adiabatic because they're over before significant heat transfer has had time to occur. In a gasoline engine, for example, compression of the gasoline-air mixture and expansion of the combustion products are nearly adiabatic because they occur so rapidly that little heat flows through the cylinder walls.

Since the heat Q is zero in an adiabatic process, the first law becomes simply

$$\Delta U = W \quad (\text{adiabatic process}) \tag{18.10}$$

This says that if we do work on a system and there's no heat transfer, then the system must gain an equal amount of internal energy. Conversely, if the system does work on its environment, then it loses internal energy (Fig. 18.10).

As a gas expands adiabatically, its volume increases while its internal energy and temperature decrease. The ideal-gas law, $pV = nRT$, then requires that the pressure decrease as well—and by more than it would in an isothermal process where T remains constant. In a pV diagram, the path of an adiabatic process—called an **adiabat**—is therefore steeper than the isotherms (Fig. 18.11).

Tactics 18.1 details the math involved in finding the adiabatic path; the result is

$$pV^\gamma = \text{constant} \quad (\text{adiabatic process}) \tag{18.11a}$$

where $\gamma = C_p/C_V$ is the ratio of the specific heats. Because $C_p = C_V + R$, the ratio $\gamma = C_p/C_V$ is always greater than 1. As expected, an adiabatic process therefore results in a greater pressure change than would a comparable isothermal process, as reflected in the steeper adiabatic path in Fig. 18.11. Physically, the adiabatic path is steeper because the gas loses internal energy as it does work, so its temperature drops. Problem 65 shows how to rewrite Equation 18.11a in terms of temperature:

$$TV^{\gamma-1} = \text{constant} \quad (\text{adiabatic process}) \tag{18.11b}$$

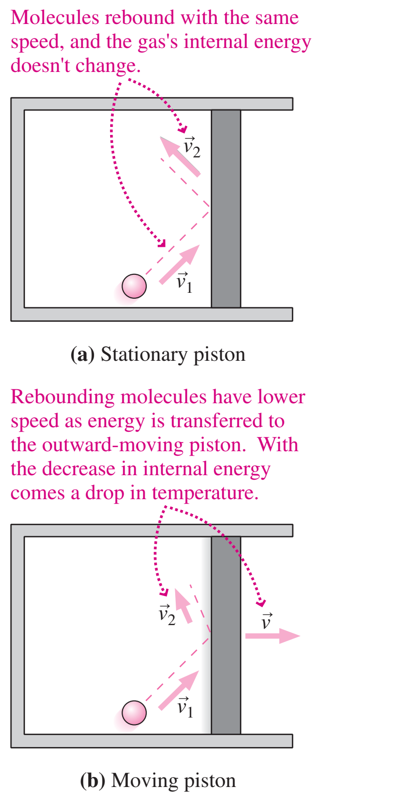


FIGURE 18.10 In an adiabatic expansion, a gas does work on the piston and its internal energy decreases. Part (b) shows microscopically how this occurs.

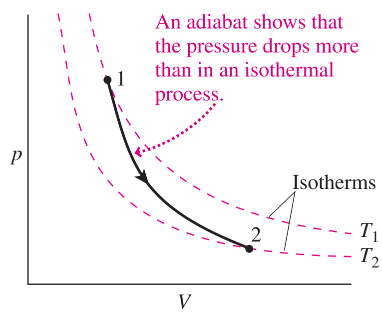


FIGURE 18.11 A pV curve for an adiabatic expansion (*dark curve*).

TACTICS 18.1 Deriving the Adiabatic Equation

Equation 18.6 gives the infinitesimal change in internal energy for *any* process: $dU = nC_V\,dT$. The corresponding work is $dW = p\,dV$ so, with $Q = 0$ in an adiabatic process, the first law becomes $nC_V\,dT = -p\,dV$. We can eliminate dT by differentiating the ideal-gas law, now letting *both* p and V change: $nR\,dT = d(pV) = p\,dV + V\,dp$. Solving for dT, substituting in our first-law statement, and multiplying through by R lead to $C_V V\,dp + (C_V + R)p\,dV = 0$. But $C_V + R = C_p$; substituting this and dividing through by $C_V pV$ give

$$\frac{dp}{p} + \frac{C_p}{C_V}\frac{dV}{V} = 0$$

Defining $\gamma \equiv C_p/C_V$ and integrating give

$$\ln p + \gamma \ln V = \ln(\text{constant})$$

where we've chosen to call the constant of integration ln(constant). Since $\gamma \ln V = \ln V^\gamma$, it follows by exponentiation that

$$pV^\gamma = \text{constant}$$

CONCEPTUAL EXAMPLE 18.1 Ideal-Gas Law Versus the Adiabatic Equation

The ideal-gas law says $pV = nRT$, but Equation 18.11a says $pV^\gamma =$ constant for an ideal gas undergoing an adiabatic process. Which is right?

EVALUATE The ideal-gas law is fundamental, so we know it's right. And we derived Equation 18.11a based on the behavior of an ideal gas. So *both* must be right. But how can that be, when one equation talks about pV and the other about pV^γ? The answer lies in the right-hand side of the ideal-gas law: nRT. For an adiabatic process, T isn't constant and therefore pV isn't constant—but pV^γ is.

ASSESS Compare the adiabatic process with an isothermal process. In that case, T is constant and we would write $pV =$ constant. Both processes obey the ideal-gas law, but the relation of p and V differs, so there's no contradiction.

MAKING THE CONNECTION Suppose you halve the volume of an ideal gas with $\gamma = 1.4$. What happens to the pressure if the process is (a) isothermal and (b) adiabatic?

EVALUATE For the isothermal process $pV =$ constant, so halving the volume doubles the pressure. For the adiabatic process it's pV^γ that's constant. Setting $p_1V_1^{\gamma} = p_2V_2^{\gamma}$ with $V_2 = V_1/2$ gives $p_2 = 2^\gamma p_1$, or in this case an increase by a factor of 2.64. The pressure increases more than in the isothermal case because the temperature goes up.

It's another exercise in calculus to integrate Equation 18.3 for the work done on the gas in an adiabatic process. You can do this in Problem 63; the result is

$$W = \frac{p_2V_2 - p_1V_1}{\gamma - 1} \tag{18.12}$$

EXAMPLE 18.3 An Adiabatic Process: Diesel Power

Fuel ignites in a diesel engine from the heat of compression as the piston moves toward the top of the cylinder; there's no spark plug as in a gasoline engine. Compression is fast enough that the process is essentially adiabatic. If the ignition temperature is 500°C, what compression ratio V_{max}/V_{min} is needed (Fig. 18.12)? Air's specific-heat ratio is $\gamma = 1.4$, and before compression the air is at 20°C.

INTERPRET We identify the thermodynamic process here as adiabatic compression.

DEVELOP The problem involves temperature and volume, so Equation 18.11b applies, giving $T_{min}V_{min}^{\gamma-1} = T_{max}V_{max}^{\gamma-1}$.

EVALUATE Solving for the compression ratio V_{max}/V_{min} gives

$$\frac{V_{max}}{V_{min}} = \left(\frac{T_{min}}{T_{max}}\right)^{1/(\gamma-1)} = \left(\frac{773\text{ K}}{293\text{ K}}\right)^{1/0.4} = 11$$

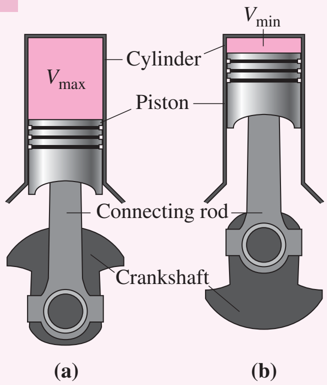


FIGURE 18.12 One cylinder of a diesel engine, shown with the piston (a) at the bottom of its stroke and (b) at the top. The compression ratio is V_{max}/V_{min}.

ASSESS Practical diesel engines have higher ratios to ensure reliable ignition. Their high compression makes diesels heavier than their gasoline counterparts, but also more fuel efficient. Problem 66 explores the diesel engine further. ■

APPLICATION Smog Alert!

The smog that blankets urban areas is an unfortunate manifestation of our prolific fossil-fueled energy consumption. Adiabatic processes in the atmosphere determine whether or not smog lingers over a city. Consider a volume of air that's heated, perhaps because it's over hot pavement that absorbs solar energy. The air becomes less dense, and its buoyancy makes it rise. As it ascends into regions of lower pressure, it expands, doing work against the surrounding atmosphere. Air is a poor heat conductor, so the process is essentially adiabatic. Therefore, the gas cools as it does work.

Now, temperature in the atmosphere normally decreases with altitude. So here's the crucial question: Does the rising air cool faster or slower than the surrounding atmosphere? If it cools more slowly, then it continues to be warmer, and it continues to rise. Any pollution is carried high into the atmosphere where it's dispersed. But if the decrease in air temperature with altitude isn't great, or in an **inversion** where it's actually warmer aloft, the rising air will soon reach equilibrium with its surroundings and won't rise any higher. The effect is to trap air and its entrained pollutants near the surface, as shown in this photo of Los Angeles. Smog alert!

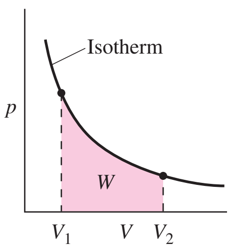
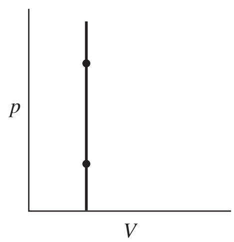
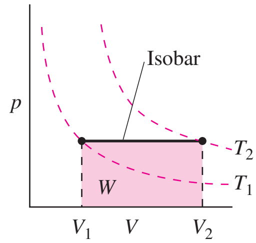
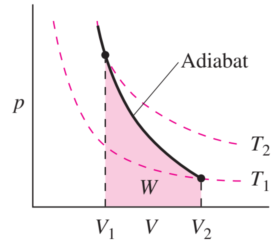
GOT IT? 18.2 Name the basic thermodynamic process involved when each of the following is done to a piston-cylinder system containing ideal gas, and tell also whether temperature, pressure, volume, and internal energy increase or decrease: (a) The piston is locked in place and a flame is applied to the bottom of the cylinder; (b) the cylinder is completely insulated and the piston is pushed downward; (c) the piston is exposed to atmospheric pressure and is free to move, while the cylinder is cooled by placing it on a block of ice.

Cyclic Processes

Many natural and technological systems undergo **cyclic processes**, in which the system returns periodically to the same thermodynamic state. Engineering examples include engines and refrigerators whose mechanical construction ensures cyclic behavior. Many natural oscillations, like those of a sound wave or a pulsating star, are essentially cyclic.

Cyclic processes often involve the four basic processes we've just explored, as summarized in Table 18.1. We've seen that the work done in any reversible process is just the area under the pV curve. A cyclic process returns to the same point in the pV diagram, so it involves both expansion and compression (Fig. 18.13). During compression, work is done on the gas; during expansion, the gas does work on its surroundings. The net work done on the gas is the difference between the two, shown in Figure 18.13 as the area enclosed by the cyclic path in the pV diagram.

Table 18.1 Ideal-Gas Processes

	ISOTHERMAL	CONSTANT-VOLUME	ISOBARIC	ADIABATIC
pV diagram				
Defining characteristic	T = constant	V = constant	p = constant	$Q = 0$
First law	$Q = -W$	$Q = \Delta U$	$Q = \Delta U - W$	$\Delta U = W$
Work done on gas	$W = -nRT \ln\left(\frac{V_2}{V_1}\right)$	$W = 0$	$W = -p\,(V_2 - V_1)$	$W = \frac{p_2V_2 - p_1V_1}{\gamma - 1}$
Other relationships	pV = constant	$Q = nC_V\Delta T$	$Q = nC_p\Delta T$ $C_p = C_V + R$	pV^γ = constant $TV^{\gamma-1}$ = constant

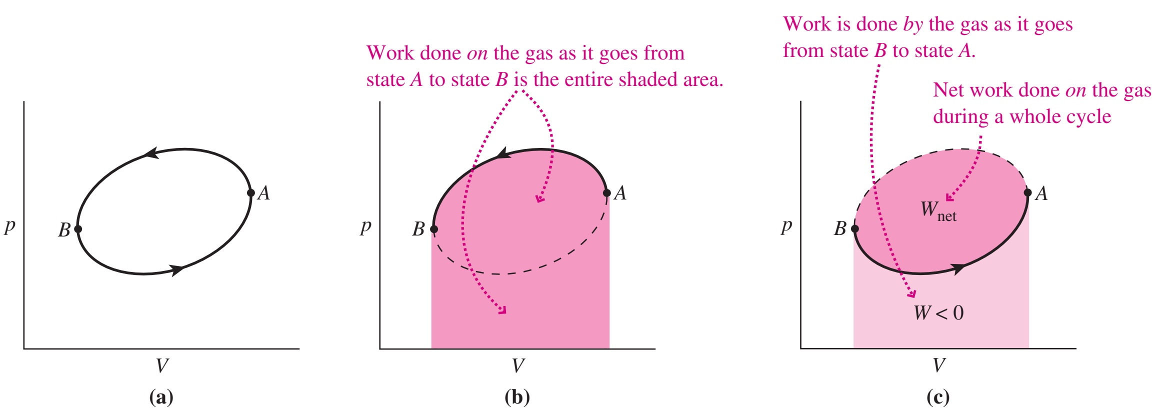


FIGURE 18.13 (a) A pV diagram for a cyclic process. (b), (c) Work done on the gas over one cycle is the area inside the closed path.

EXAMPLE 18.4 **A Cyclic Process: Finding the Work**

An ideal gas with $\gamma = 1.4$ occupies 4.0 L at 300 K and 100 kPa pressure. It's compressed adiabatically to one-fourth of its original volume, then cooled at constant volume back to 300 K, and finally allowed to expand isothermally to its original volume. How much work is done on the gas?

INTERPRET This problem involves a cyclic process, and we identify three separate thermodynamic processes that make up the cycle: adiabatic, constant-volume, and isothermal.

DEVELOP Here it helps to draw a pV diagram, shown in Fig. 18.14. Our plan is to use equations in Table 18.1 to determine the work for each of the basic processes and then combine them to get the net work. For the adiabatic process AB, Table 18.1 gives $W_{AB} = (p_B V_B - p_A V_A)/(\gamma - 1)$; for the constant-volume process BC, $W_{BC} = 0$; and for the isothermal process CA, the work is $W_{CA} = -nRT \ln(V_A/V_C)$.

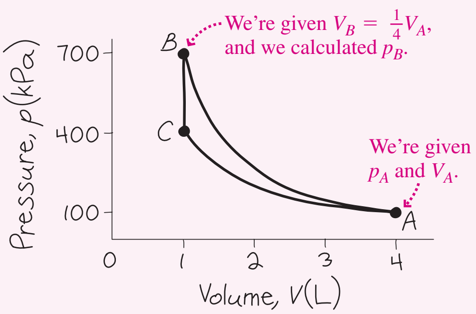


FIGURE 18.14 The cyclic process *ABCA* of Example 18.4 includes adiabatic (*AB*), constant-volume (*BC*), and isothermal (*CA*) sections.

EVALUATE For the adiabatic process AB we're given all quantities except p_B. This we can get from the adiabatic equation pV^γ = constant, or $p_B V_B{}^\gamma = p_A V_A{}^\gamma$. Solving gives $p_B = p_A(V_A/V_B)^\gamma = 696.4$ kPa, where we used the given information $p_A = 100$ kPa, $\gamma = 1.4$, and a compression to one-fourth the original volume $(V_A/V_B = 4)$. We now have enough information to find the work done over the adiabatic path:

$$W_{AB} = \frac{p_B V_B - p_A V_A}{\gamma - 1} = 741 \text{ J}$$

where, with pressures in kPa $(=10^3 \text{ Pa})$ and volumes in L $(=10^{-3} \text{ m}^3)$, the factors $10^{\pm 3}$ cancel and there's no need to convert. The work W_{AB} is positive because work is done *on* the gas when it's compressed.

In the expression $W_{CA} = -nRT \ln(V_A/V_C)$ for the isothermal work, we can evaluate the quantity nRT at *any* point on the isothermal curve because T is constant. The ideal-gas law says that $nRT = pV$, and we know both p and V at point A. So $nRT = p_A V_A = 400$ J, where again we could multiply $p_A = 100$ kPa by $V_A = 4.0$ L to get an answer in SI units. The isothermal work is then

$$W_{CA} = -nRT \ln\left(\frac{V_A}{V_C}\right) = -(400 \text{ J})(\ln 4) = -555 \text{ J}$$

This is negative because the gas does work in expanding from C to A.

Combining our results for all three segments gives the net work:

$$W_{ABCA} = W_{AB} + W_{BC} + W_{CA} = 741 \text{ J} + 0 \text{ J} - 555 \text{ J} = 186 \text{ J}$$

ASSESS Make sense? The final answer is positive because we've done net work *on* the gas; that's always the case in going counterclockwise around a cyclic path in a pV diagram. Since the system returns to its original state, its internal energy undergoes no net change. That means all the work that's done on it must be transferred to its surroundings as heat. Since no heat flows during the adiabatic process AB, and since the gas *absorbs* heat during the isothermal expansion CA, the only time it transfers heat to its surroundings is during the constant-volume cooling process BC. ■

18.3 Specific Heats of an Ideal Gas

We've found that the thermodynamic behavior of an ideal gas depends on the specific heats C_V and C_p. What are the values of those quantities?

Our ideal-gas model of Chapter 17 assumed the gas molecules were structureless point particles with only translational kinetic energy. The internal energy U of the gas is the sum of all those molecular kinetic energies. But the average kinetic energy is directly proportional to the temperature: $\frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$. If we have n moles of gas, the internal energy is then $U = nN_\text{A}\left(\frac{1}{2}m\overline{v^2}\right) = \frac{3}{2}nN_\text{A}kT$, where N_A is Avogadro's number. But $N_\text{A}k = R$, the gas constant, so $U = \frac{3}{2}nRT$. Solving Equation 18.6 for the molar specific heat then gives

$$C_V = \frac{1}{n}\frac{\Delta U}{\Delta T} = \frac{3}{2}R \tag{18.13}$$

For this gas of structureless particles, the adiabatic exponent γ is therefore

$$\gamma = \frac{C_p}{C_V} = \frac{C_V + R}{C_V} = \frac{\frac{5}{2}R}{\frac{3}{2}R} = \frac{5}{3} = 1.67$$

Some gases, notably the inert gases helium (He), neon (Ne), argon (Ar), and others in the last column of the periodic table, have adiabatic exponents and specific heats given by these equations. But others do not. At room temperature, for example, hydrogen (H_2), oxygen (O_2), and nitrogen (N_2) obey adiabatic laws with γ very nearly $\frac{7}{5}$ $(=1.4)$ and, correspondingly, specific heat $C_V = \frac{5}{2}R$. On the other hand, sulfur dioxide (SO_2) and nitrogen dioxide (NO_2) have specific-heat ratios close to 1.3 and therefore C_V of about $3.4R$.

What's going on here? A clue lies in the structure of individual gas molecules, reflected in their chemical formulas. The inert-gas molecules are **monatomic**, consisting of single atoms. To the extent that these atoms behave like structureless mass points, the only energy they can have is kinetic energy of translational motion. We can think of that kinetic energy as being a sum of *three* terms, each associated with motion in one of the three mutually perpendicular directions. We call each separate term in the energy of a system a **degree of freedom**, meaning a way that system can take on energy. So a monatomic molecule has three degrees of freedom.

In contrast, hydrogen, oxygen, and nitrogen molecules are **diatomic**, as shown in Fig. 18.15. Although a gas of such molecules should still obey the ideal-gas law $PV = nRT$, these molecules can have rotational as well as translational kinetic energy. Then the kinetic energy of a diatomic molecule consists of *five* terms, three for the three directions of translational motion and two for rotational motions about the two mutually perpendicular axes shown in Fig. 18.15. So a diatomic molecule has five degrees of freedom. We'll now see how this difference between *three* degrees of freedom for monatomic molecules and *five* for diatomic molecules accounts for the difference between their specific heats.

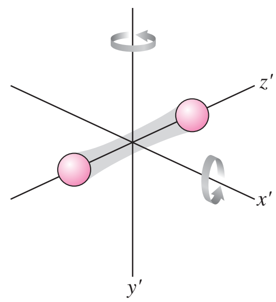


FIGURE 18.15 A diatomic molecule can have significant rotation about two perpendicular axes.

The Equipartition Theorem

We showed in Chapter 17 that the average kinetic energy associated with a gas molecule's motion in one direction is $\frac{1}{2}kT$. We then argued that all three directions are equally probable, making the molecular kinetic energy, on average, $\frac{3}{2}kT$. The argument from one direction to three is based on the assumption that random collisions will share energy equally among the possible motions. When a molecule can rotate as well as translate, energy should be shared also among possible rotational motions. The 19th-century Scottish physicist James Clerk Maxwell first proved this fact, which is known as the **equipartition theorem**:

> **Equipartition theorem** When a system is in thermodynamic equilibrium, the average energy per molecule is $\frac{1}{2}kT$ for each degree of freedom.

We've just seen that a diatomic molecule has five degrees of freedom: three translational and two rotational. The average energy of such a molecule is then $5\left(\frac{1}{2}kT\right) = \frac{5}{2}kT$, so the total internal energy in n moles of a diatomic gas is $U = nN_A\left(\frac{5}{2}kT\right) = \frac{5}{2}nRT$. Equation 18.6 then gives the molar specific heat at constant volume:

$$C_V = \frac{1}{n}\frac{\Delta U}{\Delta T} = \tfrac{5}{2}R \qquad \text{(diatomic molecule)}$$

Our result $C_p = C_V + R$ still holds, since it was derived from the first law of thermodynamics without regard to molecular structure, so $C_p = \frac{7}{2}R$ and $\gamma = C_p/C_V = \frac{7}{5} = 1.4$. These results describe the observed behavior of diatomic gases like hydrogen, oxygen, and nitrogen at room temperature.

A polyatomic molecule like NO_2 can rotate about any of three perpendicular axes (Fig. 18.16). It then has a total of six degrees of freedom, giving $U = 3nRT$ and corresponding specific heats $C_V = 3R$ and $C_p = C_V + R = 4R$. The adiabatic exponent is then $\gamma = \frac{4}{3} \simeq 1.33$, reasonably close to the experimental value $\gamma = 1.29$ for NO_2.

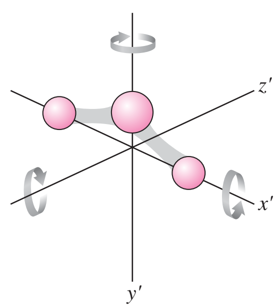


FIGURE 18.16 A triatomic molecule like NO_2 has three rotational degrees of freedom.

EXAMPLE 18.5 Specific Heat: A Gas Mixture

A gas mixture consists of 2.0 mol of oxygen (O_2) and 1.0 mol of argon (Ar). Find the volume specific heat of the mixture.

INTERPRET This problem is about specific heat and molecular structure. We identify the molecules involved as diatomic O_2 and monatomic Ar.

DEVELOP Equation 18.6, $\Delta U = nC_V\,\Delta T$, determines the volume specific heat, so we need to find how the internal energy U depends on temperature. Our plan is to use the equipartition theorem to get the energy per molecule for each gas, then find the total energy as a function of temperature, and from that the specific heat.

EVALUATE Being diatomic, O_2 has five degrees of freedom, so the equipartition theorem gives the average energy per molecule as $\frac{5}{2}kT$. Then the total energy in $n = 2$ moles of oxygen is $U_{O_2} = nN_A\left(\frac{5}{2}kT\right) = \frac{5}{2}nRT = 5.0RT$, where we used $N_Ak = R$. Monatomic Ar has three degrees of freedom, so the internal energy in our 1 mole of argon is, similarly, $U_{Ar} = \frac{3}{2}nRT = 1.5RT$. The total internal energy is then $U = 6.5RT$, so Equation 18.6 gives

$$C_V = \frac{1}{n}\frac{\Delta U}{\Delta T} = \frac{6.5R}{3.0\text{ mol}} = 2.2R$$

ASSESS Make sense? Our answer lies between the values $1.5R$ and $2.5R$ that we found for monatomic and diatomic gases, respectively. It's closer to $2.5R$ because there's more oxygen in the mixture. ■

Quantum Effects

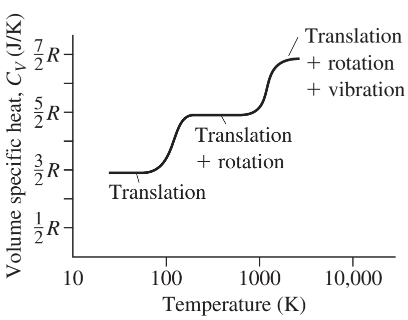


FIGURE 18.17 Volume specific heat of H_2 gas as a function of temperature. Below 20 K hydrogen is liquid, and above 3200 K it dissociates into individual atoms.

Relating molecular structure and gas behavior is a remarkable triumph for Newtonian physics. But hidden in our analysis is an assumption that Newtonian physics can't justify. Real atoms have size, so even monatomic molecules should rotate. Why not more degrees of freedom? The answer lies in quantum physics, which requires a certain minimum energy for a periodic motion such as rotation. At normal temperatures, the average thermal energy is too low to excite rotation of monatomic molecules, or of diatomic molecules about their long axis. So these molecules exhibit three and five degrees of freedom, respectively. That results in the volume specific heats $\frac{3}{2}R$ and $\frac{5}{2}R$ that we've seen. For diatomic molecules at higher temperatures, still another motion comes into play—the simple harmonic oscillation of the two atoms due to the springlike bond between them. That adds two more degrees of freedom, corresponding to the kinetic and potential energies of this oscillation, and the specific heat increases correspondingly. At very low temperatures, in contrast, there isn't enough thermal energy to excite any rotation in a diatomic gas, and it then exhibits the specific heat $C_V = \frac{3}{2}R$ that we normally associate with a monatomic gas. Figure 18.17 shows these effects for diatomic hydrogen (H_2).

Are you bothered by the strange restrictions quantum mechanics imposes on molecular rotation and vibration? You should be! Nothing in your experience suggests that a rotating object can't have any amount of energy you care to give it. But quantum mechanics deals with a realm much smaller than that of our daily experience. The quantization of energy is only one of many unusual things that occur in the quantum realm. We'll explore more quantum phenomena in Part 6.

CHAPTER 18 SUMMARY

Big Picture

The big idea here is conservation of energy, now expanded to include heat. The expanded statement of energy conservation is the **first law of thermodynamics**, which relates the change in a system's internal energy to the heat flowing into the system and the work done on the system. The first law can be used with the ideal-gas law to give a quantitative description of basic thermodynamic processes applied to ideal gases; these are described graphically using ***pV* diagrams**. The **equipartition theorem** states that in thermodynamic equilibrium, energy is shared equally among the possible energy modes of a system.

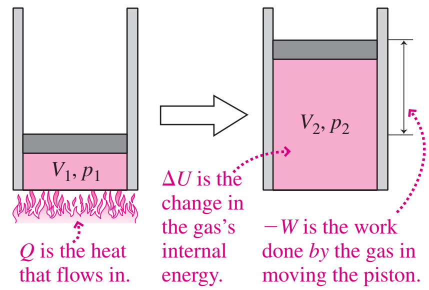
Key Concepts and Equations

Quantitatively, the first law of thermodynamics states

$$\Delta U = Q + W$$

Meaning of terms in the first law:

- ΔU is the change in a system's internal energy.
- Q is the heat transferred *to* the system.
  - Positive Q means a net heat input to the system.
  - Negative Q means heat leaves the system.
- W is the work done *on* the system.
  - Positive W means work is done on the system.
  - Negative W means the system does work on its surroundings.



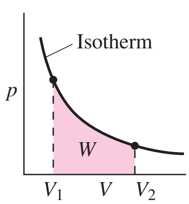
In general, the work done by a system is related to the changes in pressure and volume:

$$W = -\int_{V_1}^{V_2} p\,dV$$

Applications

Ideal-gas processes:

ISOTHERMAL

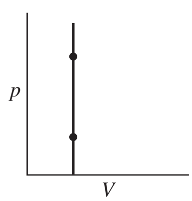


T = constant
$Q = -W$

$$W = -nRT\ln\left(\frac{V_2}{V_1}\right)$$

pV = constant

CONSTANT-VOLUME

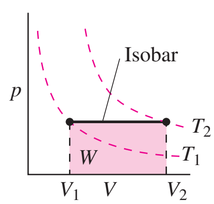


V = constant
$Q = \Delta U$

$W = 0$

$Q = nC_V\Delta T$

ISOBARIC

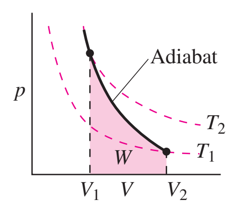


p = constant
$Q = \Delta U - W$

$W = -p(V_2 - V_1)$

$Q = nC_p\Delta T$
$C_p = C_V + R$

ADIABATIC

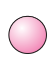


$Q = 0$
$\Delta U = W$

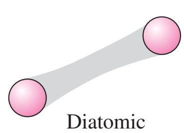
$$W = \frac{p_2V_2 - p_1V_1}{\gamma - 1}$$

pV^{γ} = constant
$TV^{\gamma-1}$ = constant

The specific heats of an ideal gas follow from the **degrees of freedom** of each molecule:



Monatomic
3 degrees of freedom
$C_V = \frac{3}{2}R$



Diatomic
5 degrees of freedom
$C_V = \frac{5}{2}R$

For homework assigned on MasteringPhysics, go to www.masteringphysics.com | **BIO** *Biology and/or medical-related problems* **C** *Computer problems*

For Thought and Discussion

1. The temperature of the water in a jar is raised by violently shaking the jar. Which of the terms Q and W in the first law of thermodynamics is involved in this case?
2. What's the difference between heat and internal energy?
3. Some water is tightly sealed in a perfectly insulated container. Is it possible to change the water temperature? Explain.
4. Are the initial and final equilibrium states of an irreversible process describable by points in a pV diagram? Explain.
5. Why can't an irreversible process be described by a path in a pV diagram?
6. Does the first law of thermodynamics apply to irreversible processes?
7. A quasi-static process begins and ends at the same temperature. Is the process necessarily isothermal?
8. Figure 18.18 shows two processes, A and B, that connect the same initial and final states, 1 and 2. For which process is more heat added to the system?

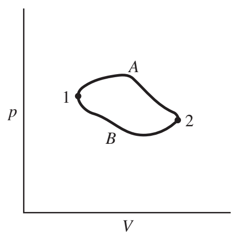


FIGURE 18.18 For Thought and Discussion 8

9. When you let air out of a tire, the air seems cool. Why? What kind of process is occurring?
10. Blow on the back of your hand with your mouth wide open. Your breath will feel hot. Now tighten your lips into a small opening and blow again. Now your breath feels cool. Why?
11. You boil water in an open pan. Of which of the four basic processes we considered is this an example?
12. Three identical gas-cylinder systems are compressed from the same initial state to final states that have the same volume, one isothermally, one adiabatically, and one isobarically. Which system has the most work done on it? The least?
13. Why is specific heat at constant pressure greater than at constant volume?
14. In what sense can a gas of diatomic molecules be considered an ideal gas, given that its molecules aren't point particles?

Exercises and Problems

Exercises

Section 18.1 The First Law of Thermodynamics

15. In a perfectly insulated container, 1.0 kg of water is stirred vigorously until its temperature rises by 7.0°C. How much work is done on the water?
16. In a closed but uninsulated container, 500 g of water are shaken violently until the temperature rises by 3.0°C. The mechanical work required in the process is 9.0 kJ. (a) How much heat is transferred during the shaking? (b) How much mechanical energy would have been required if the container had been perfectly insulated?
17. A 40-W heat source is applied to a gas sample for 25 s, during which time the gas expands and does 750 J of work on its surroundings. By how much does the internal energy of the gas change?
18. Find the rate of heat flow into a system whose internal energy is increasing at the rate of 45 W, given that the system is doing work at the rate of 165 W.
19. In a certain automobile engine, 17% of the total energy released in burning gasoline ends up as mechanical work. What's the engine's mechanical power output if its heat output is 68 kW?

Section 18.2 Thermodynamic Processes

20. An ideal gas expands from the state (p_1, V_1) to the state (p_2, V_2), where $p_2 = 2p_1$ and $V_2 = 2V_1$. The expansion proceeds along the straight diagonal path AB in Fig. 18.19. Find an expression for the work done *by* the gas during this process.

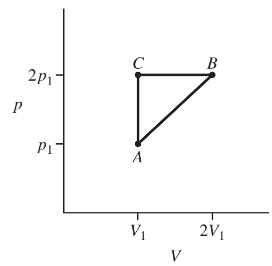


FIGURE 18.19 Exercises 20, 21 and Problem 69

21. Repeat Exercise 20 for a process that follows the path ACB in Fig. 18.19.
22. A balloon contains 0.30 mol of helium. It rises, while maintaining a constant 300-K temperature, to an altitude where its volume has expanded five times. Neglecting tension forces in the balloon, how much work is done *by* the helium during this isothermal expansion?
23. The balloon of Exercise 22 starts at 100 kPa pressure and rises to an altitude where $p = 75$ kPa, maintaining a constant 300 K temperature. (a) By what factor does its volume increase? (b) How much work does the gas in the balloon do?
24. How much work does it take to compress 2.5 mol of an ideal gas to half its original volume while maintaining a constant 300 K temperature?
25. By what factor must the volume of a gas with $\gamma = 1.4$ be changed in an adiabatic process if the kelvin temperature is to double?

Section 18.3 Specific Heats of an Ideal Gas

26. A gas mixture contains 2.5 mol of O_2 and 3.0 mol of Ar. What are this mixture's molar specific heats at constant volume and at constant pressure?
27. A mixture of monatomic and diatomic gases has specific-heat ratio $\gamma = 1.52$. What fraction of its molecules are monatomic?
28. What should be the approximate specific-heat ratio of a gas consisting of 50% NO_2 $(\gamma = 1.29)$, 30% O_2 $(\gamma = 1.40)$, and 20% Ar $(\gamma = 1.67)$?
29. By how much does the temperature of (a) an ideal monatomic gas and (b) an ideal diatomic gas (with molecular rotation but no vibration) change in an adiabatic process in which 2.5 kJ of work are done on each mole of gas?

Problems

30. An ideal gas expands to 10 times its original volume, maintaining a constant 440 K temperature. If the gas does 3.3 kJ of work on its surroundings, (a) how much heat does it absorb, and (b) how many moles of gas are there?

31. During cycling, the human body typically releases stored energy **BIO** from food at the rate of 500 W, and produces about 120 W of mechanical power. At what rate does the body produce heat during cycling?

32. A 0.25-mol sample of ideal gas initially occupies 3.5 L. If it takes 61 J of work to compress the gas isothermally to 3.0 L, what's the temperature?

33. As the heart beats, blood pressure in an artery varies from a high **BIO** of 125 mm of mercury to a low of 80 mm. These values are gauge pressures—that is, excesses over atmospheric pressure. An air bubble trapped in an artery has diameter 1.52 mm when blood pressure is at its minimum. (a) What will its diameter be at maximum pressure? (b) How much work does the blood (and ultimately the heart) do in compressing this bubble, assuming the air remains at the same 37.0°C temperature as the blood?

34. It takes 600 J to compress a gas isothermally to half its original volume. How much work would it take to compress it by a factor of 10 starting from its original volume?

35. A gas undergoes an adiabatic compression during which its volume drops to half its original value. If the gas pressure increases by a factor of 2.55, what's its specific-heat ratio γ?

36. A gas with $\gamma = 1.40$ is at 100 kPa pressure and occupies 5.00 L. (a) How much work does it take to compress the gas adiabatically to 2.50 L? (b) What's its final pressure?

37. A gas sample undergoes the cyclic process $ABCA$ shown in Fig. 18.20, where AB lies on an isotherm. The pressure at A is 60 kPa. Find (a) the pressure at B and (b) the net work done on the gas.

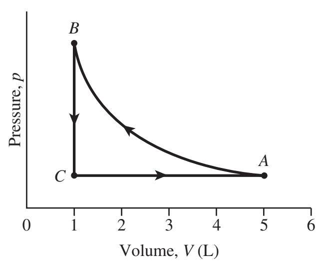


FIGURE 18.20 Problems 37 and 38

38. Repeat Problem 37 taking AB as an adiabat and using specific-heat ratio $\gamma = 1.4$.

39. A gasoline engine has compression ratio 8.5 (see Example 18.3 for the meaning of this term), and the fuel-air mixture compresses adiabatically with $\gamma = 1.4$. If the mixture enters the engine at 30°C, what will its temperature be at maximum compression?

40. By what factor must the volume of a gas with $\gamma = 1.4$ be changed in an adiabatic process if the pressure is to double?

41. Volvo's B5340 engine, used in the V70 series cars, has compression ratio 10.2, and the fuel-air mixture undergoes adiabatic compression with $\gamma = 1.4$. If air at 320 K and atmospheric pressure fills an engine cylinder at its maximum volume, what will be (a) the temperature and (b) the pressure at the point of maximum compression?

42. A gas expands isothermally from state A to state B, in the process absorbing 35 J of heat. It's then compressed isobarically to state C, where its volume equals that of state A. During the compression, 22 J of work are done on the gas. The gas is then heated at constant volume until it returns to state A. (a) Draw a pV diagram for this process. (b) How much work is done on or by the gas during the complete cycle? (c) How much heat is transferred to or from the gas as it goes from B to C to A?

43. A 2.0-mol sample of ideal gas with molar specific heat $C_V = \frac{5}{2}R$ is initially at 300 K and 100 kPa pressure. Determine the final temperature and the work done *by* the gas when 1.5 kJ of heat are added to the gas (a) isothermally, (b) at constant volume, and (c) isobarically.

44. Prove that the slope of an adiabat at a given point in a pV diagram is γ times the slope of the isotherm passing through the same point.

45. An ideal gas with $\gamma = 1.67$ starts at point A in Fig. 18.21, where its volume and pressure are 1.00 m^3 and 250 kPa, respectively. It undergoes an adiabatic expansion that triples its volume, ending at B. It's then heated at constant volume to C, and compressed isothermally back to A. Find (a) the pressure at B, (b) the pressure at C, and (c) the net work done on the gas.

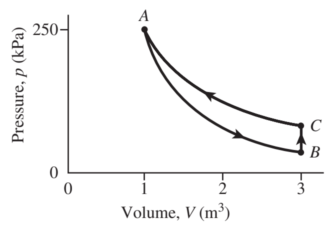


FIGURE 18.21 Problem 45

46. The gas of Example 18.4 starts at state A in Fig. 18.14 and is compressed adiabatically until its volume is 2.0 L. It's then cooled at constant pressure until it reaches 300 K, then allowed to expand isothermally back to state A. Find (a) the net work done on the gas and (b) the minimum volume of the gas.

47. The gas of Example 18.4 starts at state A in Fig. 18.14 and is heated at constant volume until its pressure has doubled. It's then compressed adiabatically until its volume is one-fourth its original value, then cooled at constant volume to 300 K, and finally allowed to expand isothermally to its original state. Find the net work done on the gas.

48. A 25-L sample of ideal gas with $\gamma = 1.67$ is at 250 K and 50 kPa. The gas is compressed isothermally to one-third of its original volume, then heated at constant volume until its state lies on the adiabatic curve that passes through its original state, and then allowed to expand adiabatically to that original state. Find the net work involved. Is net work done *on* or *by* the gas?

49. Show that the relation between pressure and temperature in an adiabatic process is $p^{1-\gamma}T^{\gamma} =$ constant.

50. A 25-L sample of ideal gas with $\gamma = 1.67$ is at 250 K and 50 kPa. The gas is compressed adiabatically until its pressure triples, then cooled at constant volume back to 250 K, and finally allowed to expand isothermally to its original state. (a) How much work is done on the gas? (b) What is the gas's minimum volume? (c) Sketch this cyclic process in a pV diagram.

51. You're the product safety officer for a company that makes cycling accessories. You're given a new design for a bicycle pump that includes a cylinder 30 cm long when the pump handle is all the way out. To keep the pump from getting too hot, you specify that the temperature rise should not exceed 50°C when the handle is pushed rapidly, with the outlet blocked, until the internal length of the cylinder is 17 cm. Assuming air initially at 20°C, does the pump meet your temperature-rise criterion?

52. Figure 18.22 shows data and a fit curve from an experimental **BIO** measurement of the pressure-volume curve for a human lung. Estimate the work involved in fully inflating the lung.

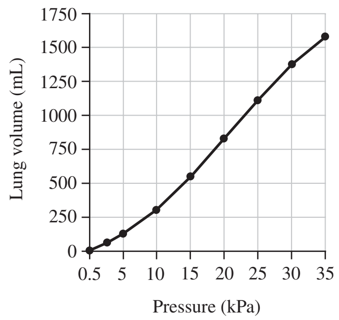


FIGURE 18.22 Problem 52

53. External forces compress 21 mol of ideal monatomic gas. During the process, the gas transfers 15 kJ of heat to its surroundings, yet its temperature rises by 160 K. How much work was done on the gas?
54. A gas with $\gamma = 7/5$ is at 273 K when it's compressed isothermally to one-third of its original volume and then further compressed adiabatically to one-fifth of its original volume. Find its final temperature.
55. An ideal gas with $\gamma = 1.3$ is initially at 273 K and 100 kPa. The gas is compressed adiabatically to 240-kPa pressure. Find its final temperature.
56. The curved path in Fig. 18.23 lies on the 350-K isotherm for an ideal gas with $\gamma = 1.4$. (a) Calculate the net work done on the gas as it goes around the cyclic path *ABCA*. (b) How much heat flows into or out of the gas on the segment *AB*?

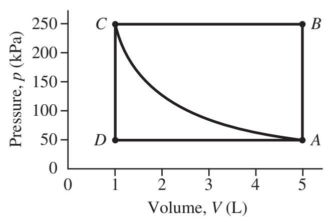


FIGURE 18.23 Problems 56 and 57

57. Repeat part (a) of Problem 56 for the path *ACDA* in Fig. 18.23. (b) How much heat flows into or out of the gas on the segment *CD*?
58. A gas mixture contains monatomic argon and diatomic oxygen. An adiabatic expansion that doubles its volume results in the pressure dropping to one-third of its original value. What fraction of the molecules are argon?
59. How much of a triatomic gas with $C_V = 3R$ would you have to add to 10 mol of monatomic gas to get a mixture whose thermodynamic behavior was like that of a diatomic gas?
60. An 8.5-kg rock at 0°C is dropped into a well-insulated vat containing a mixture of ice and water at 0°C. When equilibrium is reached, there are 6.3 g less ice. From what height was the rock dropped?
61. A piston-cylinder arrangement containing 0.30 mol of nitrogen at high pressure is in thermal equilibrium with an ice-water bath containing 200 g of ice. The pressure of the ambient air is 1.0 atm. The gas is allowed to expand isothermally until it's in pressure balance with its surroundings. After the process is complete, the bath contains 210 g of ice. What was the original gas pressure?
62. Experimental studies show that the pV curve for a frog's lung can **BIO** be approximated by $p = 10v^3 - 67v^2 + 220v$, with v in mL and p in Pa. Find the work done when such a lung inflates from zero to 4.5 mL volume.
63. Show that the application of Equation 18.3 to an adiabatic process results in Equation 18.12.
64. A horizontal piston-cylinder system containing n mol of ideal gas is surrounded by air at temperature T_0 and pressure p_0. If the piston is displaced slightly from equilibrium, show that it executes simple harmonic motion with angular frequency $\omega = Ap_0/\sqrt{MnRT_0}$, where A and M are the piston area and mass, respectively. Assume the gas temperature remains constant.
65. Use the ideal-gas law to eliminate pressure in Equation 18.11a, and show that the result can be written as Equation 18.11b.
66. Figure 18.24 shows the thermodynamic cycle of a diesel engine. As in Example 18.3, the compression ratio r is the ratio of maximum to minimum volume: $r = V_1/V_2$. In addition, the so-called *cutoff ratio* is defined by $r_c = V_3/V_2$. Find an expression for the engine's efficiency, in terms of the ratios r and r_c and the specific-heat ratio γ.

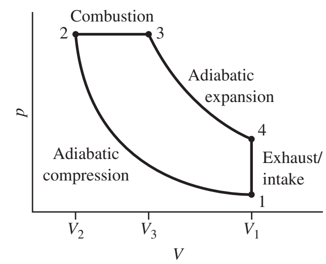


FIGURE 18.24 Problem 66

67. In a reversible process, a volume of air $V_0 = 17\ \text{m}^3$ at pressure $p_0 = 1$ atm is compressed such that the pressure and volume are related by $(p/p_0)^{-2} = V/V_0$. How much work is done by the gas in reaching a final pressure of 1.4 atm?
68. A real gas is more accurately described using the van der Waals equation: $[p + a(n/V)^2](V - nb) = nRT$, where a and b are constants. Find an expression, corresponding to Equation 18.4, for the work done by a van der Waals gas undergoing an isothermal expansion from V_1 to V_2.
69. Repeat Exercise 20 for an expansion along the path $p = p_1[1 + (V - V_1)^2/V_1^2]$.
70. The *adiabatic lapse rate* is the rate at which air cools as it rises and expands adiabatically in the atmosphere (see Application: Smog Alert, on page 302). Express dT in terms of dp for an adiabatic process, and use the hydrostatic equation (Equation 15.2) to express dp in terms of dy. Then, calculate the lapse rate dT/dy. Take air's average molecular weight to be 29 u and $\gamma = 1.4$, and remember that the altitude y is the negative of the depth h in Equation 15.2.
71. The nuclear power plant at which you're the public affairs manager has a backup gas-turbine system. The backup system produces electrical energy at the rate of 360 MW, while extracting energy from natural gas at the rate of 670 MW. The local town council has raised concern over waste thermal energy dumped into the environment. Their standards state the thermal waste power must not exceed 400 MW and that all power generation must be at least 50% efficient. Does the backup turbine meet this standard?

72. Your class on alternative habitats is designing an underwater habitat. A small diving bell will be lowered to the habitat. A hatch at the bottom of the bell is open, so water can enter to compress the air and thus keep the air pressure inside equal to the pressure of the surrounding water. The bell is lowered slowly enough that the inside air remains at the same temperature as the water. But the water temperature increases with depth in such a way that the air pressure and volume are related by $p = p_0\sqrt{V_0/V}$, where $V_0 = 17\ \text{m}^3$ and $p_0 = 1$ atm are the surface values. Suppose the diving bell's air volume cannot be less than $8.7\ \text{m}^3$ and the pressure must not exceed 1.5 atm when submerged. Are these criteria met?

Passage Problems

Warm winds called Chinooks (a Native-American term meaning "snow eaters") sometimes sweep across the plains just east of the Rocky Mountains. These winds carry air from high in the mountains down to the plains rapidly enough that the air has no time to exchange heat with its surroundings (Fig. 18.25). On a particular Chinook day, temperature and pressure high in the Colorado Rockies are 60 kPa and 260 K ($-13°$C), respectively; the plain below is at 90 kPa.

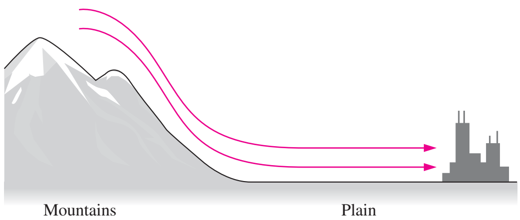


FIGURE 18.25 Chinooks (Passage Problems 73–76)

73. The process the air undergoes as it descends the mountains is
   a. isothermal.
   b. isovolumic.
   c. isobaric.
   d. adiabatic.

74. As the air descends, its internal energy
   a. increases.
   b. decreases.
   c. is unchanged.

75. As the air descends, its volume
   a. increases by 50%.
   b. increases by less than 50%.
   c. decreases by 50%.
   d. decreases by less than 50%.
   e. is unchanged.

76. When the air reaches the plain, its temperature is approximately
   a. 240 K.
   b. 260 K.
   c. 290 K.
   d. 390 K.

Answers to Chapter Questions

Answer to Chapter Opening Question

Energy is conserved, provided thermal energy is included. The engine produces both mechanical energy and thermal energy of its exhaust gases; together, they sum to the energy released in combustion.

Answers to GOT IT? Questions

18.1. (c). Only the internal energy is the same, since it's a thermodynamic state variable unique to a point in the pV diagram.

18.2. (a) Constant-volume; T and p increase, V doesn't change, U increases as heat flows into the gas. (b) Adiabatic; T and p increase, V decreases, U increases as work is done on the gas. (c) Isobaric; T decreases, p doesn't change, V decreases, U decreases as heat flows out of the gas.

19 The Second Law of Thermodynamics

Most of the energy extracted from fuel in power plants is discarded as waste heat. The large cooling tower shown here dumps this waste heat into the environment. Why is so much energy wasted?

New Concepts, New Skills

By the end of this chapter you should be able to

- Explain the second law of thermodynamics and the limitations it places on our ability to extract useful work from thermal-energy sources (19.2).
- Calculate the efficiencies of heat engines and refrigerators (19.2, 19.3).
- Describe the concepts of energy quality and entropy (19.4).
- Compute entropy changes for basic thermodynamic processes (19.4).

Connecting Your Knowledge

- This chapter builds on the first law of thermodynamics and the use of *pV* diagrams (18.1, 18.2).
- You should be familiar with the basic thermodynamic processes, particularly isothermal and adiabatic processes (18.2).

The first law of thermodynamics relates heat and other forms of energy. Much of our world depends on this relationship. Cars extract energy from the heat of burning gasoline. Most of our electricity originates in heat released by burning fuels or fissioning uranium. Our own bodies run on energy that begins as heat in the Sun's core. But the first law doesn't tell the whole story. Heat and mechanical energy aren't the same, and the difference makes the conversion of heat to work a more subtle task than the first law would imply.

19.1 Reversibility and Irreversibility

Figure 19.1 shows a movie of a bouncing ball. Play it backward and it still makes sense. Figure 19.2 shows a block sliding along a table, slowing because of friction—and warming in the process. Play this film backward and it makes no sense. You'll never see a block at rest suddenly start to move, cooling as it goes. Yet energy would be conserved if it did, so the first law of thermodynamics would be satisfied. Beat an egg, blending yolk and white. Reverse the beater, and you'll never see them separate again. Put cups of cold and hot water in contact; the hot water cools and the cold water warms. The opposite never occurs—although energy would still be conserved.

Why are these events **irreversible**? In each case we start with matter in an organized state. The molecules of the sliding block share a common motion. The yolk molecules are all in one place. The hot water has more energetic molecules. Of all possible states, these *organized* ones are rare. There are many more *disorganized* states—for example, all the possible arrangements of molecules in a scrambled egg. As a system evolves, chances are it will end up less organized, simply because there are far more such states available to it. It's very unlikely to assume spontaneously a more organized state.

A key word here is "spontaneous." We could restore organization—for example, by putting one cup of water in the refrigerator and the other in the microwave—but that requires a rather deliberate and energy-consuming process.

Irreversibility is a probabilistic notion. Events that *could* occur without violating the principles of Newtonian physics nevertheless *don't* occur because they're too improbable. As a practical consequence, harnessing the internal energy associated with random molecular motions is difficult because those motions won't spontaneously become organized. That makes much of the world's energy unavailable for doing useful work.

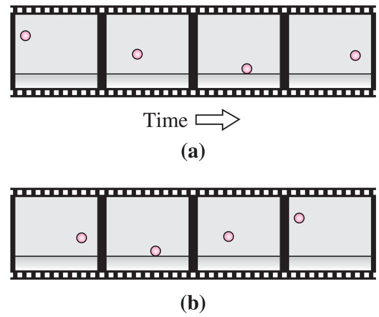


FIGURE 19.1 A movie of a bouncing ball makes sense whether it's shown forward or backward.

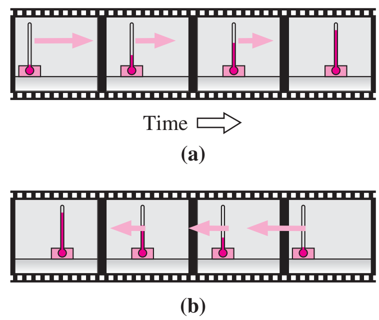


FIGURE 19.2 (a) A block warming (note thermometer) as friction dissipates its kinetic energy and it slows to a stop. (b) The reverse sequence would never happen, even though it doesn't violate energy conservation.

GOT IT? 19.1 Which of these processes is irreversible: (a) stirring sugar into coffee, (b) building a house, (c) demolishing a house with a wrecking ball, (d) demolishing a house by taking it apart piece by piece, (e) harnessing the energy of falling water to drive machinery, (f) harnessing the energy of falling water to heat a house?

19.2 The Second Law of Thermodynamics

Heat Engines

It's impossible to convert *all* the internal energy of a system to useful work. But **heat engines** extract *some* of that internal energy. Examples include gasoline and diesel engines, fossil-fueled and nuclear power plants, and jet aircraft engines.

Figure 19.3*a* is an energy-flow diagram for a "perfect" heat engine—one that extracts heat from a heat reservoir and converts it all to work. Such an engine would do exactly what we've just argued against: It would convert the random energy of thermal motion entirely to the ordered motion associated with mechanical work. In fact a perfect heat engine is impossible, for the same reason that we can't unscramble an egg or make a block accelerate spontaneously using its internal energy. This fact leads to one statement of the **second law of thermodynamics**:

Second law of thermodynamics (Kelvin-Planck statement) It is impossible to construct a heat engine operating in a cycle that extracts heat from a reservoir and delivers an equal amount of work.

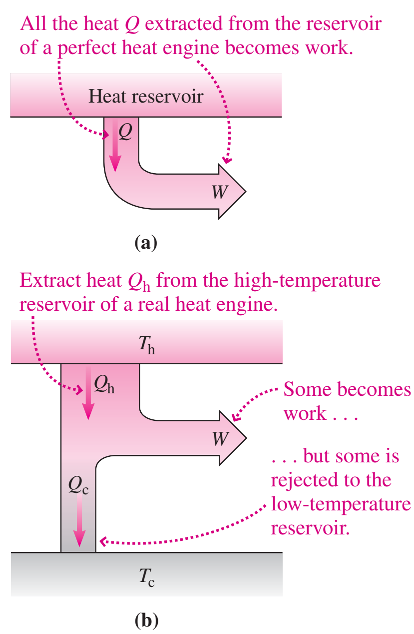


FIGURE 19.3 (a) Energy-flow diagram for a perfect heat engine. (b) A real engine delivers as work only a fraction of the energy extracted from the high-temperature reservoir.

The phrase "in a cycle" means that a practical engine goes through a repeated sequence of steps, as in the back-and-forth motions of the pistons in a gasoline engine.

A simple heat engine consists of a gas-cylinder system and a heat reservoir, the latter kept hot, perhaps, by burning a fuel. With the gas initially at high pressure, we place the cylinder in contact with the heat reservoir. The gas expands and does work W on the piston. In this isothermal process, the gas extracts heat $Q = W$ from the reservoir. Eventually the gas reaches pressure equilibrium and stops expanding. The piston must then be returned to its original position if it's to do more work.

If we just push the piston back, we'll have to do as much work as we got during the expansion, and our engine won't produce any net work. Instead we can cool the gas to reduce its volume, through contact with a cool reservoir. But then some energy leaves the system as heat rather than work, as shown conceptually in Fig. 19.3*b*. Our engine extracts heat from a source and delivers mechanical work, but over a full cycle the work delivered is less than the heat extracted. The remaining energy is rejected to the lower-temperature reservoir, usually the environment. That's why much of the energy released from fuels in car engines and power plants ends up as waste heat.

The second law of thermodynamics says we can't build a perfect heat engine. But how close can we come? We define the **efficiency** e of an engine as the ratio of the work W we get from it to what we have to supply—namely, the heat Q_h: $e = W/Q_h$. Since the process is cyclic, there's no net change in internal energy over one cycle. The first law of thermodynamics then shows that the work W done *by* the engine is the difference between the heat Q_h extracted from the high-temperature reservoir and the heat Q_c rejected to the cool reservoir:

$$e = \frac{W}{Q_h} = \frac{Q_h - Q_c}{Q_h} = 1 - \frac{Q_c}{Q_h} \qquad (19.1)$$

In this chapter we'll often use W for the work done *by* an engine; in the first law it's the work done *on* a system. That's why W here is equal to the net heat $Q_h - Q_c$.

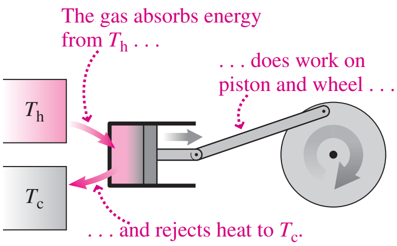


FIGURE 19.4 A simple heat engine.

Figure 19.4 shows a heat engine whose efficiency we can calculate. The engine consists of a cylinder containing an ideal gas, sealed by a movable piston. The piston is connected to a rod that turns a wheel. The engine gets its energy from a heat reservoir at a high temperature T_h, and it rejects heat to a cooler reservoir at temperature T_c. Figure 19.5 shows how the engine works in a cycle of four steps, starting with the piston in its leftmost position (state A in Fig. 19.5), where the gas volume is a minimum:

1. Isothermal expansion: The high-temperature reservoir is placed in thermal contact with the cylinder. The gas absorbs heat Q_h from the hot reservoir and expands isothermally along path AB. Since temperature remains constant, so does internal energy. The first law then shows that the engine does work $W = Q$ on the piston and wheel.
2. Adiabatic expansion: At B we remove the hot reservoir, so the gas can no longer exchange heat. Thus the expansion becomes adiabatic and follows path BC. We design the engine so the gas has cooled to T_c when the piston reaches its rightmost position (state C), the point of maximum gas volume.
3. Isothermal compression: At C we bring the cool reservoir into thermal contact with the cylinder. The wheel's inertia keeps it turning, so the piston does work on the gas, compressing it isothermally from state C to D. This work ends up as heat rejected to the cool reservoir.
4. Adiabatic compression: At D we remove the cool reservoir and the compression continues adiabatically until the gas temperature is once again at T_h and the engine is back at state A.

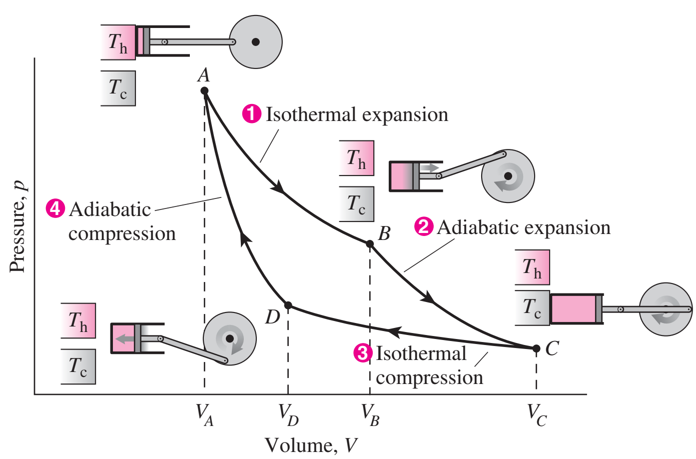


FIGURE 19.5 A pV diagram for the Carnot engine.

This cyclic process of two isothermal and two adiabatic steps is the **Carnot cycle** and the engine a **Carnot engine**, after the French engineer Sadi Carnot (1796–1832). The particular configuration of the engine isn't important, nor is the choice of an ideal gas as the engine's **working fluid**. What distinguishes the Carnot cycle from others is the sequence of thermodynamic processes and the fact that these processes are reversible. The Carnot engine is an example of a **reversible engine**—one in which thermodynamic equilibrium is maintained so that all steps could, in principle, be reversed.

What's the efficiency of a Carnot engine? To find out, we need the heats Q_h and Q_c absorbed and rejected during the isothermal parts of the cycle shown in Fig. 19.5. Equation 18.4 gives the heat Q_h absorbed during the isothermal expansion AB:

$$Q_h = nRT_h \ln\left(\frac{V_B}{V_A}\right)$$

and the heat Q_c rejected during the isothermal compression CD:

$$Q_c = -nRT_c \ln\left(\frac{V_D}{V_C}\right) = nRT_c \ln\left(\frac{V_C}{V_D}\right)$$

We put the minus sign here because the first law takes Q to be the heat *absorbed*, while Equation 19.1 for the engine efficiency requires that Q_c be the heat *rejected*. To calculate engine efficiency according to Equation 19.1, we need the ratio Q_c/Q_h:

$$\frac{Q_c}{Q_h} = \frac{T_c \ln(V_C/V_D)}{T_h \ln(V_B/V_A)} \tag{19.2}$$

This expression can be simplified by applying Equation 18.11b to the adiabatic processes BC and DA in the Carnot cycle: $T_hV_B^{\gamma-1} = T_cV_C^{\gamma-1}$ and $T_hV_A^{\gamma-1} = T_cV_D^{\gamma-1}$. Dividing the first of these two equations by the second gives

$$\left(\frac{V_B}{V_A}\right)^{\gamma-1} = \left(\frac{V_C}{V_D}\right)^{\gamma-1} \quad \text{or} \quad \frac{V_B}{V_A} = \frac{V_C}{V_D}$$

so Equation 19.2 becomes simply $Q_c/Q_h = T_c/T_h$. Using this result in Equation 19.1 then gives the efficiency of the Carnot engine:

$$e_{\text{Carnot}} = 1 - \frac{T_c}{T_h} \qquad \text{(Carnot engine efficiency)} \tag{19.3}$$

where the temperatures are measured on an absolute scale (Kelvin or Rankine). Equation 19.3 shows that the Carnot engine's efficiency depends only on the highest and lowest temperatures of its working fluid. In practice, the low temperature is usually that of the environment; then maximizing efficiency requires making the high temperature as high as possible. Real engines trade off efficiency with the ability of materials to withstand high temperature and pressure.

EXAMPLE 19.1 Calculating Efficiency: A Carnot Engine

A Carnot engine extracts 240 J from its high-temperature reservoir during each cycle, and rejects 100 J to the environment at 15°C. How much work does the engine do in one cycle? What's its efficiency? What's the temperature of the hot reservoir?

INTERPRET This problem is about a Carnot engine, which operates via the Carnot cycle.

DEVELOP Equation 19.3, $e_{\text{Carnot}} = 1 - (T_c/T_h)$, relates the two temperatures and the efficiency. Here $Q_h = 240\text{ J}$, $Q_c = 100\text{ J}$, and $T_c = 15°\text{C}$ or 288 K. The first law of thermodynamics relates work and heat flows. So our plan is to use the first law to find the work, then find the efficiency, and then use Equation 19.3 to find T_h.

EVALUATE Since there's no change in internal energy over one cycle, the first law requires that the work W done *by* the engine be equal to the net heat absorbed—namely, 240 J − 100 J. So $W = 140$ J. The efficiency is the ratio of work delivered to heat extracted, so $e = W/Q_h = 140\text{ J}/240\text{ J} = 58.3\%$. Knowing the efficiency, we solve Equation 19.3 for T_h:

$$T_h = \frac{T_c}{1-e} = \frac{288\text{ K}}{1-0.583} = 691\text{ K} = 418°\text{C}$$

ASSESS Make sense? The engine rejects somewhat less than half the 240 J as waste heat, so we should expect efficiency somewhat over 50%. And of course T_h must be greater than T_c, as our calculation confirms. ■

Engines, Refrigerators, and the Second Law

Why this emphasis on the Carnot engine? Because understanding this device will help answer the broader question of how much work we can hope to extract from thermal energy. That, in turn, will help us understand practical limitations on humankind's attempts to harness ever more energy and will lead to a deeper understanding of the second law of thermodynamics.

Why is Carnot's engine special? Couldn't we build a better engine with greater efficiency? The answer is no. The special role of the Carnot cycle is embodied in **Carnot's theorem**:

> **Carnot's theorem** All Carnot engines operating between temperatures T_h and T_c have the same efficiency, $e_{Carnot} = 1 - (T_c/T_h)$, and no other engine operating between the same two temperatures can have a greater efficiency.

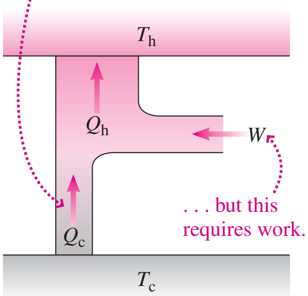


FIGURE 19.6 Energy-flow diagram for a real refrigerator.

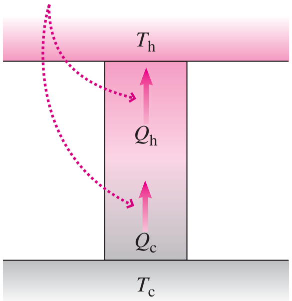


FIGURE 19.7 A perfect refrigerator is impossible.

To prove Carnot's theorem, we introduce the **refrigerator**. A refrigerator is the opposite of an engine: It extracts heat from a cool reservoir and rejects it to a hotter one, using work in the process (Fig. 19.6). A refrigerator forces heat to flow from cold to hot, but to do so it requires work. A household refrigerator cools its contents and warms the house (you can feel the heat coming out the back), but it uses electricity. That heat doesn't flow spontaneously from cold to hot leads to another statement of the second law of thermodynamics:

> **Second law of thermodynamics (Clausius statement)** It is impossible to construct a refrigerator operating in a cycle whose sole effect is to transfer heat from a cooler object to a hotter one.

The Clausius statement rules out a perfect refrigerator (Fig. 19.7).

Suppose the Clausius statement were false. Then we could build the device of Fig. 19.8*a*, consisting of a reversible Carnot engine and a perfect refrigerator. In each cycle the engine would extract, say, 100 J from the hot reservoir, put out 60 J of useful work, and reject 40 J to the cool reservoir. The perfect refrigerator could transfer the 40 J back to the hot reservoir. The net effect would be to extract 60 J from the hot reservoir and convert it entirely to work (Fig. 19.8*b*)—and we would have a perfect heat engine, in violation of the Kelvin-Planck statement of the second law. A similar argument (Problem 38) shows that if a perfect heat engine is possible, then so is a perfect refrigerator. So the Clausius and Kelvin-Planck statements of the second law are equivalent, in that if one is false, then so is the other.

Because the Carnot engine is reversible, we could run it backward and reverse its path in Fig. 19.5. The engine would extract heat from the cool reservoir, take in work, and reject heat to the hot reservoir. It would be a refrigerator. Although real refrigerators aren't designed exactly like engines, the two are, in principle, interchangeable.

We're now ready to prove Carnot's assertion that Equation 19.3 gives the maximum engine efficiency. Consider again the Carnot engine in Fig. 19.8*a*. It extracts 100 J of heat and delivers 60 J of work, so it's 60% efficient. Suppose we had another engine operating between the same two reservoirs, but with 70% efficiency. Since the Carnot engine is reversible, we can run it as a refrigerator. If we put the two together, we get the device of

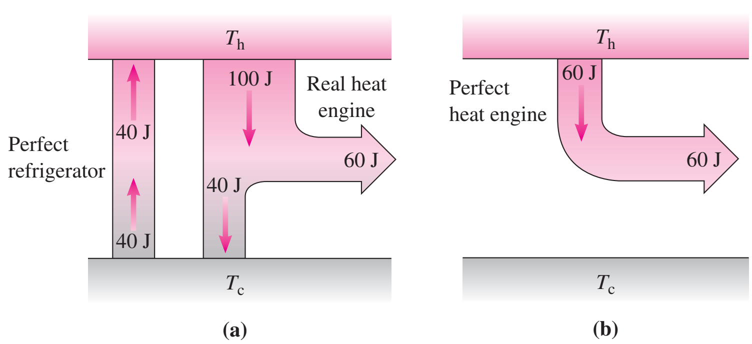


FIGURE 19.8 (a) A real heat engine combined with a perfect refrigerator is equivalent to (b) a perfect heat engine.

Fig. 19.9*a*. Its net effect is to extract 10 J from the cool reservoir and deliver 10 J of work—so it's a perfect heat engine, in violation of the second law (Fig. 19.9*b*). It's therefore impossible to make an engine that's more efficient than a Carnot engine, and thus Equation 19.3 gives the maximum possible efficiency for *any* heat engine operating between the same two fixed temperatures. For that reason the Carnot efficiency of Equation 19.3 is also called the **thermodynamic efficiency**.

Irreversible engines, because they involve processes that dissipate organized motion, are necessarily *less* efficient. So are reversible engines, if their heat exchange doesn't take place solely at the highest and lowest temperatures. The ordinary gasoline engine is a case in point; even if it could be made perfectly reversible, its efficiency would be less than that of a comparable Carnot engine (see Problem 53).

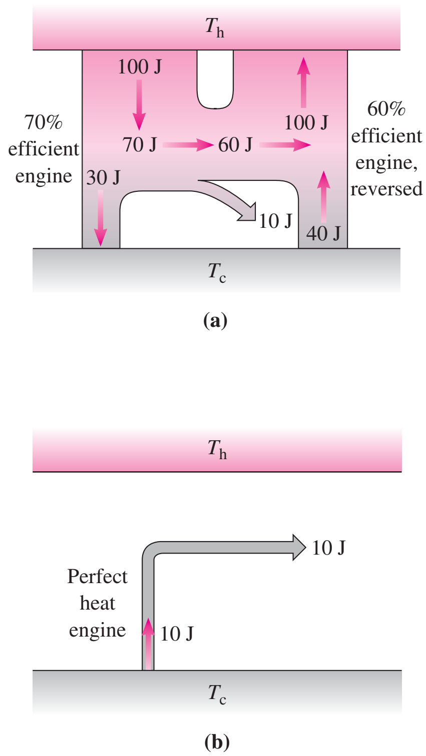


FIGURE 19.9 (a) A 60% efficient reversible engine run as a refrigerator, along with a hypothetical engine with 70% efficiency. (b) The combination is equivalent to a perfect heat engine.

19.3 Applications of the Second Law

The world abounds with thermal energy, but the second law of thermodynamics limits our ability to use that energy. Any device we construct that involves the interchange of heat and work is a heat engine or refrigerator, subject to the second law.

Limitations on Heat Engines

Most of our electricity is produced in large power plants that are heat engines powered by the fossil fuels coal, oil, or natural gas, or by nuclear fission. Figure 19.10 diagrams such a power plant. The working fluid is water, heated in a boiler and converted to steam at high pressure. The steam expands adiabatically to spin a fanlike turbine. The turbine turns a generator that converts mechanical work to electrical energy.

Steam leaving the turbine is still gaseous and is hotter than the water supplied to the boiler. Here's where the second law applies! Had the water returned from the turbine in its original state, we would have extracted as work all the energy acquired in the boiler, in violation of the second law. Therefore, we must run the steam through a **condenser**, where it contacts pipes carrying cool water, typically from a river, lake, or ocean. The condensed steam, now cool water, is fed back into the boiler to repeat the cycle.

The maximum steam temperature in a power plant is limited by the materials used in its construction. For a conventional fossil-fuel plant, current technology permits high temperatures of around 650 K. Potential damage to nuclear fuel rods limits the temperature in a nuclear plant to around 570 K. The average temperature of the cooling water is about 40°C (310 K), so the maximum possible efficiencies for these power plants, given by Equation 19.3, are

$$e_{\text{fossil}} = 1 - \frac{310\text{ K}}{650\text{ K}} = 52\% \quad \text{and} \quad e_{\text{nuclear}} = 1 - \frac{310\text{ K}}{570\text{ K}} = 46\%$$

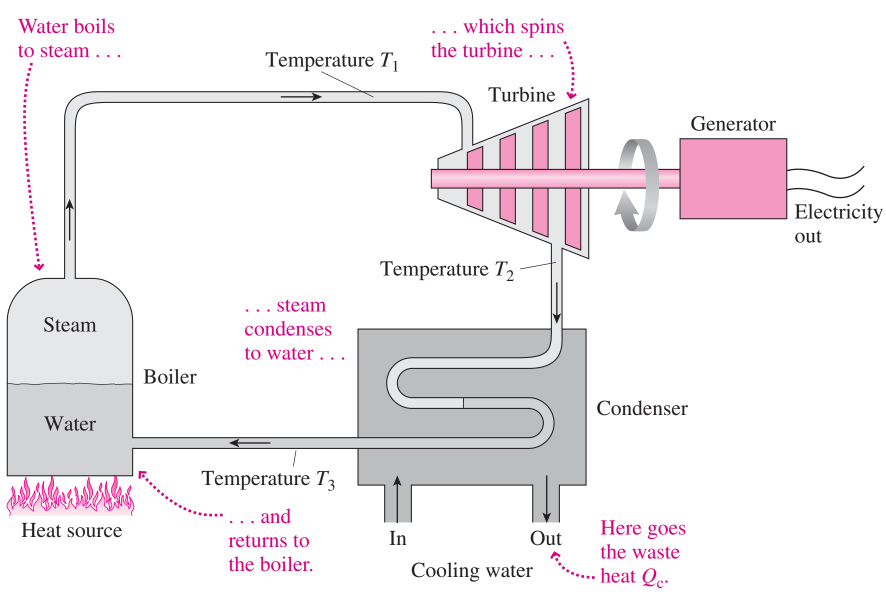


FIGURE 19.10 Schematic diagram of an electric power plant.

Temperature differences between steam and cooling water, mechanical friction, and energy needed for pumps and pollution-control devices all reduce efficiency further, to about 40% for fossil-fuel plants and 34% for nuclear plants. So roughly two-thirds of the fuel energy we use to make electricity ends up as waste heat.

A typical large power plant produces 1 GW of electricity, so another 2 GW of waste heat goes into the cooling water. The resulting temperature rise can cause serious ecological problems. The huge cooling towers you see at power plants reduce such "thermal pollution" by transferring much of the waste heat to the atmosphere (see this chapter's opening photo). Even so, a substantial fraction of all rainwater falling on the United States eventually finds its way through the condensers of power plants (see Problem 31).

EXAMPLE 19.2 Improving Efficiency: A Combined-Cycle Power Plant

The gas turbine in a combined-cycle power plant (see the Application below) operates at 1450°C. Its waste heat at 500°C is the input for a conventional steam cycle, with its average condenser temperature at 40°C. Find the thermodynamic efficiency of the combined cycle, and compare with the efficiencies of the individual components if they were operated independently.

INTERPRET This problem is about the thermodynamic efficiency of a combined-cycle power plant. As described in the Application, that means a plant using a high-temperature gas turbine whose waste heat becomes the energy input to a conventional steam turbine.

DEVELOP Figure 19.11 is a conceptual diagram of the combined-cycle plant, based on the Application. Equation 19.3, $e = 1 - (T_c/T_h)$, gives the thermodynamic efficiencies of each cycle and of the combination. We identify the 1450°C = 1723 K temperature as T_h in Equation 19.3 for the gas turbine. The intermediate temperature 500°C = 773 K serves as T_c for the gas turbine but as T_h for the steam cycle. Finally, the 40°C or 313-K condenser temperature is T_c for the steam cycle.

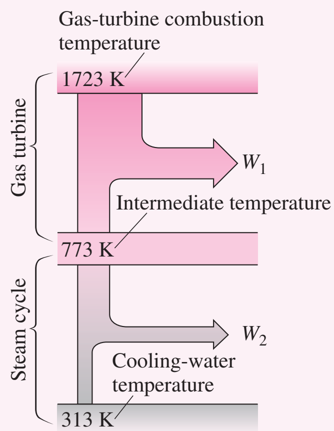


FIGURE 19.11 Conceptual diagram of a combined-cycle power plant.

EVALUATE To treat the entire plant as a single heat engine in Equation 19.3, we use the highest and lowest temperatures:

$$e_{\text{combined}} = 1 - \frac{T_c}{T_h} = 1 - \frac{313\text{ K}}{1723\text{ K}} = 0.82 = 82\%$$

Friction and other losses would reduce this figure substantially, but a combined-cycle plant operating at these temperatures could have a practical efficiency near 60%. The efficiencies of the individual components also follow from Equation 19.3:

$$e_{\text{gas turbine}} = 1 - \frac{773\text{ K}}{1723\text{ K}} = 55\% \quad \text{and} \quad e_{\text{steam}} = 1 - \frac{313\text{ K}}{773\text{ K}} = 60\%$$

ASSESS Make sense? Because of its extreme temperatures, the combined cycle gives an efficiency that's better than either of its parts! You can learn more about combined-cycle power plants in the Application below, and by working Problem 32. ■

APPLICATION Combined-Cycle Power Plants

Improving power-plant efficiency helps reduce air pollution and greenhouse-gas emissions, not to mention the cost of electricity. Modern *combined-cycle* power plants achieve efficiencies approaching 60% by combining a conventional steam system like that of Fig. 19.10 with a *gas turbine* similar to a jet aircraft engine. Gas turbines operate at high temperatures—between 1000 K and 2000 K—but they aren't very efficient because their exhaust temperature (T_c in Equation 19.3) is also high. In a combined-cycle plant, exhaust from a gas turbine drives a conventional steam cycle. The overall effect is the same as that of a single heat engine operating between the gas turbine's high combustion temperature and the low temperature of the environment (see Problem 32). The second law still limits the efficiency, but the high T_h and low T_c make for greater efficiency than in a conventional plant. The photo shows a gas-fired combined-cycle plant.

Gasoline and diesel engines provide another pervasive example of heat engines. A typical automobile engine has a theoretical maximum efficiency of around 50%, but irreversible thermodynamic processes make the actual efficiency much lower. Mechanical friction dissipates additional energy, with the end result that less than 20% of the fuel energy reaches the driving wheels. Problems 53 and 54 explore the gasoline engine.

We wouldn't be so concerned with efficiency if we didn't have to pay for fuel or worry about the environment. Engines with "free" fuel include solar-thermal power plants that concentrate sunlight to boil a fluid that drives a turbine and ocean thermal-energy conversion (OTEC) schemes that extract useful work from the modest temperature difference between tropical surface waters and the deep ocean. Neither provides significant energy today, but that could change as the world moves away from fossil fuels.

Refrigerators and Heat Pumps

A refrigerator works like an engine in reverse; it takes in mechanical work and transfers heat from its cooler interior to its warmer surroundings. An air conditioner is a refrigerator whose "interior" is the building being cooled. A close cousin is the **heat pump**, which transfers heat either way, cooling a building in the summer and warming it in the winter (Fig. 19.12). In warmer climates, heat pumps exchange energy between a building and the outside air; in cooler climates they use groundwater, typically at about 10°C year-round. Heat pumps require electricity, but they transfer more heat energy than they consume in electricity. That makes heat pumps potentially energy-saving devices for winter heating. However, some of that gain is offset by the inefficiency of the power plant producing the electricity.

An efficient refrigerator (or any other device, for that matter) should maximize what we want from the device compared with what we have to put in. The **coefficient of performance** (COP) quantifies this ratio:

$$\mathrm{COP} = \frac{\text{What we want}}{\text{What we put in}}$$

For a refrigerator or summertime heat pump, "what we want" is cooling, so the numerator is Q_c. For a wintertime heat pump, "what we want" is heating, so the numerator is Q_h. For either, "what we put in" is mechanical work, W, or its equivalent in electricity. Thus we have

$$\mathrm{COP}_{\text{refrigerator}} = \frac{Q_c}{W} = \frac{Q_c}{Q_h - Q_c} \qquad \mathrm{COP}_{\text{heat pump}} = \frac{Q_h}{W} = \frac{Q_h}{Q_h - Q_c}$$

In both cases the second equality follows from the first law of thermodynamics. In deriving the maximum efficiency of a heat engine, we found that $Q_c/Q_h = T_c/T_h$. Therefore the maximum possible COPs are

$$\mathrm{COP}_{\text{refrigerator}} = \frac{T_c}{T_h - T_c} \quad (19.4a) \qquad \mathrm{COP}_{\text{heat pump}} = \frac{T_h}{T_h - T_c} \quad (19.4b)$$

When the temperatures T_h and T_c are close, Equations 19.4 give high COPs—meaning the refrigerator or heat pump takes relatively little work to do its job. But as the difference increases, the COP drops and we have to supply more work. Incidentally, our COP expression works for engines as well, if we take "what we want" to be mechanical work W and "what we put in" to be the heat Q_h.

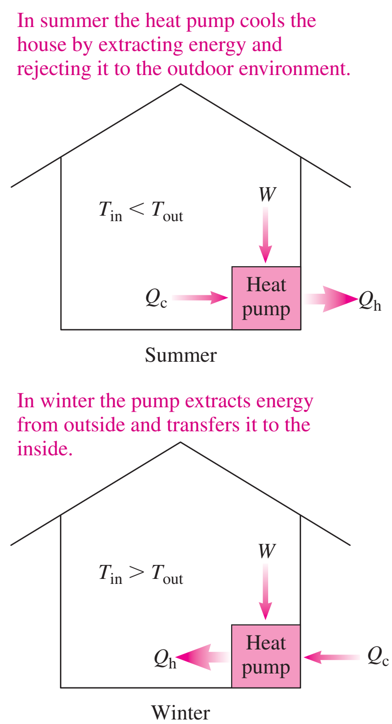


FIGURE 19.12 A heat pump.

EXAMPLE 19.3 The COP: A Home Freezer

A typical home freezer operates between a low of 0°F (−18°C or 255 K) and a high of 86°F (30°C or 303 K). What's its maximum possible COP? With this COP, how much electrical energy would it take to freeze 500 g of water initially at 0°C?

INTERPRET This problem is about a refrigerator—in this case a freezer. We identify T_h and T_c with the values 303 K and 255 K, respectively.

DEVELOP Equation 19.4a, $\mathrm{COP} = T_c/(T_h - T_c)$, will determine the COP. Then we'll use Equation 17.5, $Q = Lm$, to find the heat Q_c that the freezer must extract to freeze the water. From there we'll be able to use $\mathrm{COP} = Q_c/W$ to find the work—equivalently, the electrical energy—required.

(continued)

EVALUATE Equation 19.4a gives

$$\mathrm{COP} = \frac{T_c}{T_h - T_c} = \frac{255\ \mathrm{K}}{303\ \mathrm{K} - 255\ \mathrm{K}} = 5.31$$

From Equation 17.5 and Table 17.1, we find the heat that needs to be removed in freezing 500 g of ice: $Q_c = Lm = (334\ \mathrm{kJ/kg})(0.50\ \mathrm{kg}) = 167\ \mathrm{kJ}$. The COP is the ratio of the heat removed to the work or electrical energy required, so we have $W = Q_c/\mathrm{COP} = 167\ \mathrm{kJ}/5.31 = 31\ \mathrm{kJ}$.

ASSESS Make sense? A COP of 5.3 means that each unit of work transfers 5.3 units of heat from inside the freezer—so the electrical-energy requirement is modest. A practical freezer operating between these temperatures would have a lower COP and require more electrical energy. ■

GOT IT? 19.2 A clever engineer decides to increase the efficiency of a Carnot engine by cooling the low-temperature reservoir using a refrigerator with the maximum possible COP. Will the overall efficiency of this system (a) exceed, (b) be less than, or (c) equal that of the original engine alone?

19.4 Entropy and Energy Quality

If offered a joule of energy, would you rather it be in the form of mechanical work, heat at 1000 K, or heat at 300 K? Your answer might depend on what you want to do. To lift or accelerate a mass, you'd be smart to take your energy as work. But if you want to keep warm, heat at 300 K would be perfectly acceptable.

But which should you choose if you want to keep all your options open, making the energy available for the most possible uses? The second law of thermodynamics answers clearly: You should take the work. Why? Because you could use it directly as mechanical energy, or you could, through friction or other irreversible processes, use it to raise the temperature of something.

If you chose 300 K heat for your joule of energy, then you could supply a full joule only to objects cooler than 300 K. You couldn't do mechanical work unless you ran a heat engine. With its T_h only a little above the ambient temperature, your engine would be inefficient, and you could extract only a small fraction of a joule of mechanical energy. You would be better off with 1000-K heat since you could transfer it to anything cooler than 1000 K, or you could run a heat engine to produce up to 0.70 joule of mechanical energy.

CONCEPTUAL EXAMPLE 19.1 Energy Quality and Cogeneration

You need a new water heater, and you're trying to decide between gas and electric. The gas heater is 85% efficient, meaning 85% of the fuel energy goes into heating water. The electric heater is essentially 100% efficient. Thermodynamically, which heater makes the most sense?

EVALUATE Your electricity is energy of the highest quality. It probably comes from a thermal power plant, which typically discards as waste heat twice as much energy as it produces in electricity. The electric heater may be 100% efficient in your home, but when you consider the big picture, only about one-third of the fuel energy consumed at the power plant ends up heating your water. With 85% efficiency, the gas heater is the wiser choice.

ASSESS It makes sense to match energy sources to their end uses. Electricity is high-quality energy, so it's best for running motors, light sources, electronics, and other devices requiring high-quality energy. Turning it into low-grade heat is a thermodynamic folly! A really smart strategy is **cogeneration**, in which the waste heat from electric power generation is used to heat buildings. In Europe, whole communities are heated that way, and institutions in the United States are increasingly turning to cogeneration to reduce energy costs and carbon emissions.

MAKING THE CONNECTION If the electricity comes from a more efficient gas-fired power plant with $e = 48\%$, compare the gas consumption of your two heater choices.

EVALUATE The gas heater turns 1 unit of fuel energy into 0.85 unit of thermal energy in the water. The power plant turns 1 unit of fuel energy into 0.48 unit of electrical energy, which the electric heater converts to 0.48 unit of thermal energy. The electric heater is therefore responsible for $0.85/0.48 = 1.8$ times as much gas consumption.

Taking your energy in the form of work gives you the most options. Anything you can do with a joule of energy, you can do with the work. Heat is less versatile, with 300 K heat the least useful of the three. We're not talking here about the quantity of energy—we have exactly 1 joule in each case—but about **energy quality** (Fig. 19.13). We can readily convert an entire amount of energy from higher to lower quality, but the second law precludes going in the opposite direction with 100% efficiency.

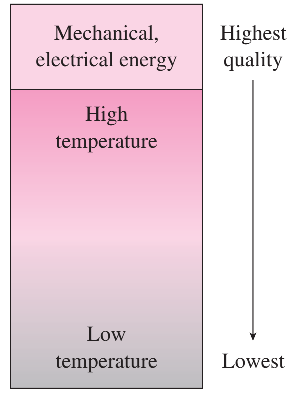


FIGURE 19.13 Energy quality measures the versatility of different energy forms.

Entropy

Mix hot and cold water, and you get lukewarm water. There's no energy loss, but you *have* lost something—namely, the ability to do useful work. In the initial state, we could have run a heat engine using the ΔT between the hot and cold water. In the final state, there's no temperature difference, so we couldn't run a heat engine. The *quantity* of energy hasn't changed, but its *quality* has decreased. **Entropy**, S, quantifies the loss of quality associated with energy transformations. In his Ninth Memoir, Clausius coined the term for its similarity to the word "energy" and its Greek root "troph," meaning *transformation*.

To motivate the definition of entropy, consider an ideal gas undergoing a Carnot cycle. Recall that a Carnot cycle consists of two isothermal and two adiabatic processes (Fig. 19.5). In deriving Equation 19.3 for the Carnot efficiency, we found that $Q_c/Q_h = T_c/T_h$, where Q_c was the heat *rejected* from the system to the low-temperature reservoir at T_c, and Q_h the heat *added* from the reservoir at T_h.

Let's focus on the ideal gas itself and define all heats as the heat *added* to the gas, so Q_c changes sign. The relationship between heats and temperatures can now be expressed as

$$\frac{Q_c}{T_c} + \frac{Q_h}{T_h} = 0 \quad \text{(Carnot cycle)}$$

We can generalize this result to *any* reversible cycle by approximating the cycle as a sequence of Carnot cycles, as shown in Fig. 19.14. For each segment, we have $\sum Q/T = 0$. As we increase the number of cycles, the volume change associated with each isothermal segment shrinks and the edges get less jagged. We can approximate the closed cycle ever closer by using more and more Carnot cycles. In the limit, the approximation becomes exact and the sum becomes an integral:

$$\oint \frac{dQ}{T} = 0 \quad \text{(any reversible cycle)} \tag{19.5}$$

where the circle indicates integration over a *closed* path.

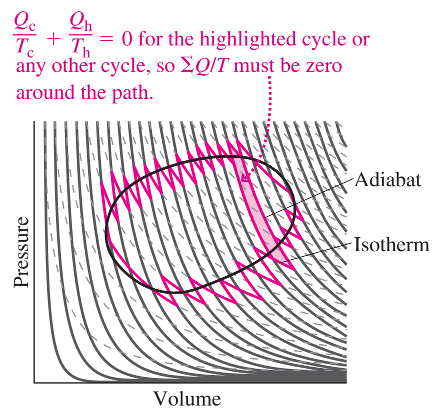


FIGURE 19.14 An arbitrary cycle approximated by isothermal (dashed curves) and adiabatic (solid curves) steps. Heat transfer occurs only during the isothermal steps.

Equation 19.5 holds for *any* closed path in the pV diagram—that is, for any *reversible cycle*. That means we can define the *entropy change*, ΔS, between an initial state 1 and a final state 2 as

$$\Delta S_{12} = \int_1^2 \frac{dQ}{T} \quad \text{(entropy change)} \tag{19.6}$$

Note that entropy has the units J/K, the same units as Boltzmann's constant k_B.

Take a system along a path from state 1 to state 2 in its pV diagram; Equation 19.6 gives the corresponding entropy change ΔS_{12}. Go back to state 1 by any other reversible path, and the resulting entropy change ΔS_{21} must be $-\Delta S_{12}$ so that there's no entropy change around the closed path (Fig. 19.15). Thus the entropy change of Equation 19.6 is independent of path; it depends only on the initial and final states. The only restriction is that we integrate over a reversible path. Like pressure and temperature, entropy is therefore a thermodynamic *state variable*—a quantity that characterizes a given state independently of how the system got into that state.

We restricted ourselves to reversible paths in Equation 19.6 since irreversible processes take a system out of thermodynamic equilibrium and therefore aren't described by paths in the pV diagram. But because entropy depends only on the initial and final states, we can calculate the entropy change in an *irreversible* process by using Equation 19.6 for a *reversible* process that goes between the same two states.

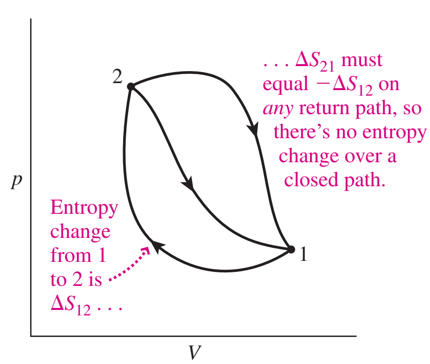


FIGURE 19.15 Entropy change is path-independent.

Note also that Equation 19.6 gives the entropy change of *just* the working fluid. The fluid—perhaps in an engine—is thermally connected to its surroundings, and if we're interested in the *total* entropy change resulting from the engine's operation, we'll need to add the entropy changes for its environment—in this case the hot and cold reservoirs.

Adiabatic Free Expansion

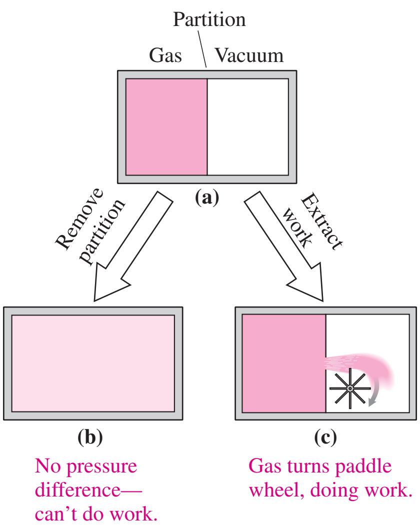


FIGURE 19.16 Two ways for a gas to expand into a vacuum.

In Fig. 19.16*a*, a partition confines an ideal gas to one side of a box; the other side is vacuum. Remove the partition, and the gas undergoes a **free expansion**, filling the box. Consider the box to be insulated, so there's no heat flow and the expansion is therefore adiabatic. But this expansion is *irreversible*, so it's significantly different from the adiabatic expansions we considered in Chapter 18. In our free expansion, the vacuum doesn't exert pressure to oppose the gas, so the gas does no work and therefore its internal energy doesn't change. Figure 19.16*c* shows how we could have used the expanding gas to turn a paddle wheel, extracting useful work. We can't do that with the uniform-pressure gas of Fig. 19.16*b*, so the free expansion results in the system's losing its ability to do work.

Let's determine the entropy change for this irreversible process. We do that by finding a reversible process that takes the gas between the same two states. Since the gas's internal energy doesn't change, neither does its temperature. So the corresponding reversible process is an isothermal expansion, for which Equation 18.4 gives the heat added: $Q = nRT\ln(V_2/V_1)$. With the temperature constant, the entropy change of Equation 19.6 becomes

$$\Delta S = \int \frac{dQ}{T} = \frac{1}{T}\int dQ = \frac{Q}{T} = nR\ln\left(\frac{V_2}{V_1}\right)$$

The final volume V_2 is larger than V_1, so entropy has *increased*. Although we computed this result for the reversible process, it holds for *any* process that takes the system between the same initial and final states—including our irreversible free expansion.

Entropy and the Availability of Work

Entropy increases during irreversible expansion—and energy quality decreases, in that the system loses its ability to do work. Had we let the gas in Fig. 19.16 undergo a reversible isothermal expansion instead of free expansion, it would have done work equal to the heat gained:

$$W = Q = nRT\ln\left(\frac{V_2}{V_1}\right)$$

After the irreversible free expansion, the gas can no longer do this work, even though its energy is unchanged. Comparing W with the entropy change ΔS we calculated above, we see that the energy that becomes unavailable to do work is $E_{\text{unavailable}} = T\,\Delta S$. This illustrates a more general relation between entropy and energy quality:

> During an irreversible process in which the entropy of a system increases by ΔS, energy $E = T_{\min}\Delta S$ becomes unavailable to do work, where $T_{\min}$ is the lowest temperature available to the system.

This statement shows that entropy provides our measure of energy quality. Given two systems with identical energy content, the one with the lower entropy contains the higher-quality energy. An entropy increase corresponds to a degradation in energy quality, as energy becomes unavailable to do work.

EXAMPLE 19.4 Increasing Entropy: The Loss of Energy Quality

A 2.0-L cylinder contains 5.0 mol of compressed gas at 300 K. If the cylinder is discharged into a 150-L vacuum chamber and its temperature remains 300 K, how much energy has become unavailable to do work?

INTERPRET This problem asks about the loss of energy quality during an irreversible and therefore entropy-increasing process—namely, an adiabatic free expansion.

DEVELOP Figure 19.17 is a sketch of the situation, similar to Fig. 19.16 except that here the gas volume changes more dramatically. In analyzing the free expansion of Fig. 19.16, we found $\Delta S = nR\ln(V_2/V_1)$. Our statement relating entropy and energy quality says that the energy made unavailable to do work is $T_{\min}\Delta S$. So our plan is to calculate ΔS and multiply by $T_{\min}$ to find that unavailable energy.

EVALUATE Because the temperature doesn't change, $T_{\min}$ is the 300 K temperature we're given, and we have

$$E_{\text{unavailable}} = T\,\Delta S = nRT\ln\left(\frac{V_2}{V_1}\right)$$
$$= (5.0\text{ mol})(8.314\text{ J/K}\cdot\text{mol})(300\text{ K})\ln\left(\frac{152\text{ L}}{2.0\text{ L}}\right) = 54\text{ kJ}$$

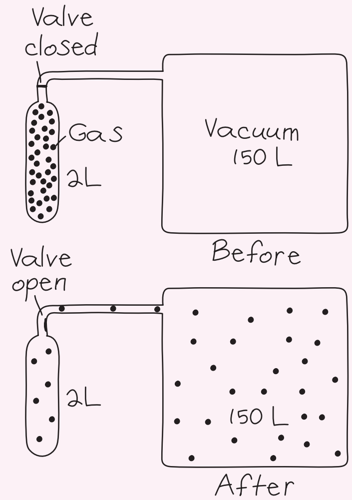


FIGURE 19.17 Our sketch for Example 19.4. Note that the final volume is 152 L.

ASSESS Make sense? Yes: This is the work we could have extracted from a reversible isothermal expansion. By letting the gas undergo an irreversible process, we gave up the possibility of extracting this work. ■

A Statistical Interpretation of Entropy

We began this chapter arguing that systems naturally evolve from ordered to disordered states. Entropy increase measures that loss of order, which is what makes energy unavailable to do work. Here we'll explore the meaning of entropy further, based on the partitioned box we used for adiabatic free expansion.

Suppose we have a gas with just two identical molecules. The left side of Fig. 19.18 shows that, with the partition removed, there are four possible **microstates**—specific arrangements of the individual molecules in the box. But say we only care about the number of molecules in each side of the box. Then two of these arrangements are indistinguishable, because they both have one molecule in each half of the box. Those two correspond to a single **macrostate**, specified by giving the number of molecules in each half of the box, without regard to which molecules they are. This is shown on the right in Fig. 19.18.

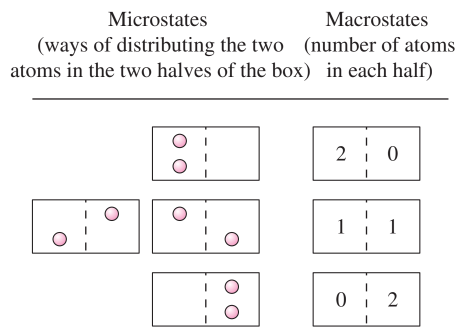


FIGURE 19.18 A gas of two molecules has four possible microstates and three macrostates.

With four available microstates, the probability of being in any one microstate is $\frac{1}{4}$. There's only one microstate with both molecules on the left, so the chances of being in the macrostate with two molecules on the left is also $\frac{1}{4}$; the same is true for the macrostate with two molecules on the right. But two of the possible microstates have one molecule on each side, so the probability for this macrostate is $\frac{1}{2}$.

Now consider a gas of four molecules. Figure 19.19 (next page) shows 16 possible microstates, corresponding to 5 macrostates. Again, the probability of finding the system in a given macrostate depends on the associated number of microstates; Fig. 19.19 enumerates these probabilities. Clearly, we're most likely to find the system in the macrostate with the molecules evenly divided; the states with all the molecules on one side are now quite improbable.

Raise the number of molecules to 100, and the number of microstates becomes huge—2^{100}, or more than 10^{30}. That makes the macrostates with all or nearly all the molecules on one side extremely improbable. The macrostate with half the molecules on each side

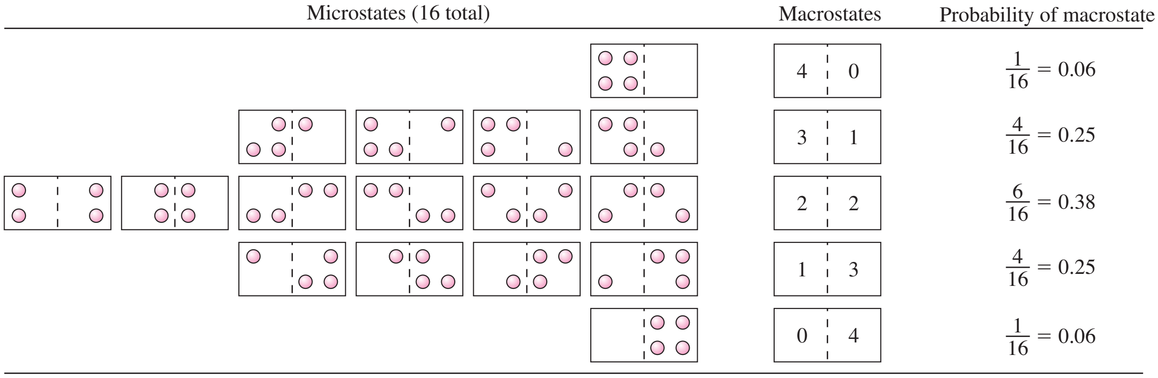


FIGURE 19.19 Microstates, macrostates, and probabilities for a gas of four molecules.

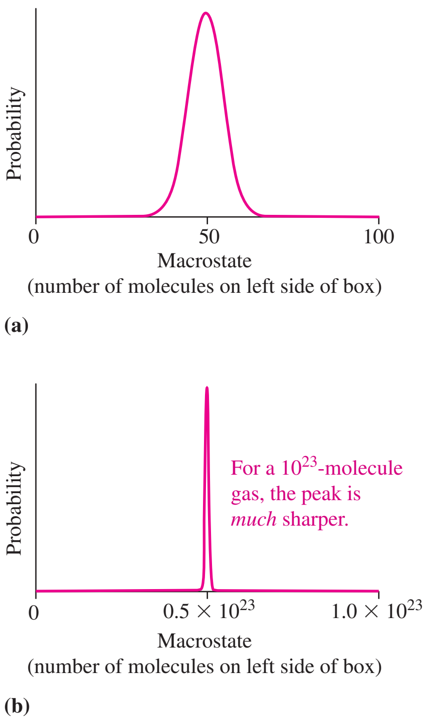


FIGURE 19.20 (a) Probability distributions for a gas of (a) 100 molecules and (b) 10^{23} molecules.

remains the most likely, although states with nearly equal divisions of molecules are also quite probable. Rather than enumerate these probabilities, we graph them (Fig. 19.20).

Typical gas samples have roughly 10^{23} molecules, and that makes macrostates with anything other than a nearly equal distribution of molecules extremely unlikely—as suggested by the spike-like probability distribution in Fig. 19.20. You could sit in your room for many times the age of the universe, and you'd never see all the air molecules spontaneously end up on one side of the room!

Entropy and the Second Law of Thermodynamics

The statistical improbability of more ordered states—in our example, those with significantly more molecules on one side of the box—is at the root of the second law of thermodynamics. Although we defined entropy in terms of heat flow and temperature (Equation 19.6), a more fundamental definition involves the probabilities of individual microstates. In that sense, entropy is indeed a measure of disorder.

Systems naturally evolve toward disordered or higher-entropy states simply because there are far more of these states available. So a general statement of the second law is:

Second law of thermodynamics The entropy of a closed system can never decrease.

At best, the entropy of a closed system remains constant—and that's only in an ideal, reversible process. If anything irreversible occurs—friction, or any deviation from thermodynamic equilibrium—then entropy increases. As it does, energy becomes unavailable to do work, and nothing within the closed system can restore that energy to its original quality. This new statement of the second law subsumes our previous statements about the impossibility of perfect heat engines and refrigerators, for their operation would require an entropy decrease.

We *can* decrease the entropy of a system that isn't closed—but only by supplying high-quality energy from outside. Running a refrigerator decreases the entropy of its contents, but this requires electrical energy to make heat flow from cold to hot. That high-quality electrical energy deteriorates into additional heat that's rejected to the refrigerator's environment. If we consider the entire system, not just the refrigerator's contents, the overall entropy has increased.

Any system whose entropy seems to decrease—that gets more rather than less organized—can't be closed. If we enlarge a system's boundaries to encompass the entire universe, then we have the ultimate statement of the second law:

Second law of thermodynamics The entropy of the universe can never decrease.

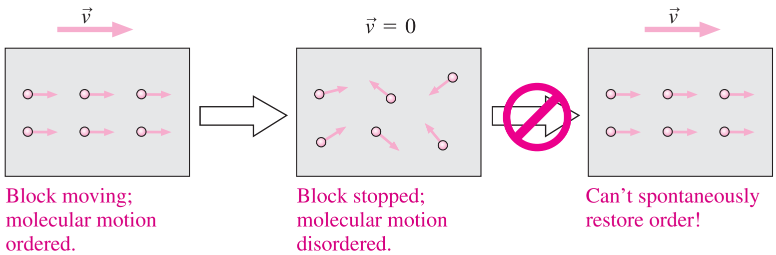
Examples include the growth of a living thing from the random mix of molecules in its environment, the construction of a skyscraper from materials that were originally dispersed about Earth, and the appearance of ordered symbols on a printed page from a bottle of ink. All these are entropy-decreasing processes in which matter goes from near chaos to a highly organized state—akin to separating yolk and white from a scrambled egg. But Earth isn't a closed system. It gets high-quality energy from the Sun, energy that's ultimately responsible for life. If we consider the Earth-Sun system, the entropy decrease associated with life and civilization is more than balanced by the entropy increase associated with the degradation of high-quality solar energy. We living things represent a remarkable phenomenon—the organization of matter in a universe governed by a tendency toward disorder. But we can't escape the second law of thermodynamics. Our highly organized selves and society, and the entropy decreases they represent, come into being only at the expense of greater entropy increases elsewhere.

GOT IT? 19.3 In each of the following processes, does the entropy of the named system alone increase, decrease, or stay the same? (a) A balloon deflates; (b) cells differentiate in a growing embryo, forming different physiological structures; (c) an animal dies, and its remains gradually decay; (d) an earthquake demolishes a building; (e) a plant utilizes sunlight, carbon dioxide, and water to manufacture sugar; (f) a power plant burns coal and produces electrical energy; (g) a car's friction-based brakes stop the car.

SUMMARY

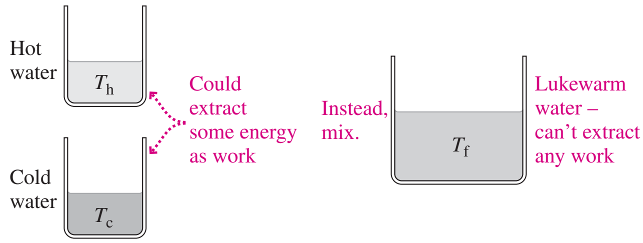
Big Picture

The big idea behind this chapter is the **second law of thermodynamics**—ultimately, the statement that systems tend naturally toward disorder, or states of higher **entropy**. The second law is manifest in the real world by forbidding the construction of perfect heat engines and perfect refrigerators—therefore preventing us from extracting as useful work all the energy that's contained in random thermal motions. Ultimately, the second law says that the entropy of any closed system, including the entire universe, cannot decrease.



Key Concepts and Equations

Entropy is a quantitative measure of energy quality and of disorder; the higher the entropy, the lower the energy quality and the greater the disorder. The highest-quality energy is mechanical or electrical energy, followed by the internal energy of systems at high temperature, and finally low-temperature internal energy. Whenever entropy increases, energy becomes unavailable to do work.

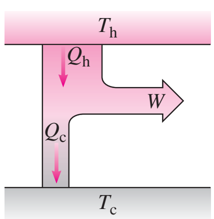


- $\Delta S = \int_1^2 \frac{dQ}{T}$ gives the entropy change as a system goes from state 1 to state 2.
- $E_{unavailable} = T_{min}\Delta S$ is the energy that becomes unavailable as a result of entropy increase ΔS.

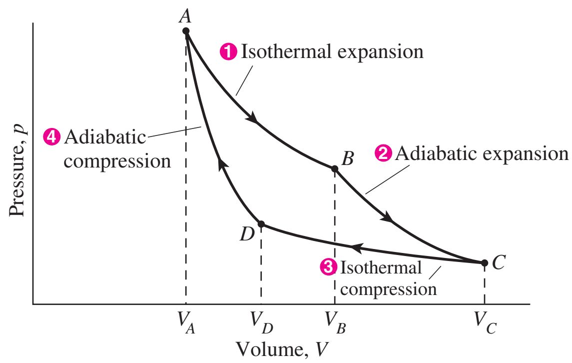
Applications

The second law sets the maximum possible efficiency of any heat engine as that of the **Carnot engine**, an engine that combines adiabatic and isothermal processes.

$$\underbrace{e = \frac{W}{Q_h}}_{\substack{\text{This defines}\\ \text{an engine's}\\ \text{efficiency.}}} \leq \underbrace{e_{max} = 1 - \frac{T_c}{T_h}}_{\substack{\text{This is the}\\ \text{maximum possible}\\ \text{efficiency.}}}$$



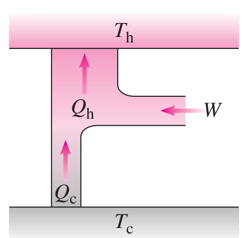
Energy-flow diagram for an engine



pV diagram for Carnot engine

Similarly, the second law limits the **coefficient of performance** of refrigerators and heat pumps:

$$\mathrm{COP}_{refrigerator} = \frac{T_c}{T_h - T_c} \qquad \mathrm{COP}_{heat\ pump} = \frac{T_h}{T_h - T_c}$$



MP For homework assigned on MasteringPhysics, go to www.masteringphysics.com | BIO Biology and/or medical-related problems C Computer problems

For Thought and Discussion

1. Could you cool the kitchen by leaving the refrigerator open? Explain.
2. Could you heat the kitchen by leaving the oven open? Explain.
3. Should a car get better mileage in the summer or the winter? Explain.
4. Is there a limit to the maximum temperature that can be achieved by focusing sunlight with a lens? If so, what is it?
5. Name some irreversible processes that occur in a real engine.
6. Your power company claims that electric heat is 100% efficient. Discuss.
7. A hydroelectric power plant, using the energy of falling water, can operate with efficiency arbitrarily close to 100%. Why?
8. A heat-pump manufacturer claims the device will heat your home using only energy already available in the ground. Is this true?
9. Why do refrigerators and heat pumps have different definitions of COP?
10. The heat Q added during adiabatic free expansion is zero. Why can't we then argue from Equation 19.6 that the entropy change is zero?
11. Energy is conserved, so why can't we recycle it as we do materials?
12. Why doesn't the evolution of human civilization violate the second law of thermodynamics?

Exercises and Problems

Exercises

Sections 19.2 and 19.3 The Second Law of Thermodynamics and Its Applications

13. What are the efficiencies of reversible heat engines operating between (a) the normal freezing and boiling points of water, (b) the 25°C temperature at the surface of a tropical ocean and deep water at 4°C, and (c) a 1000°C flame and room temperature?
14. A cosmic heat engine might operate between the Sun's 5600 K surface and the 2.7 K temperature of intergalactic space. What would be its maximum efficiency?
15. A reversible Carnot engine operating between helium's melting point and its 4.25 K boiling point has an efficiency of 77.7%. What's the melting point?
16. A Carnot engine absorbs 900 J of heat each cycle and provides 350 J of work. (a) What's its efficiency? (b) How much heat is rejected each cycle? (c) If the engine rejects heat at 10°C, what's its maximum temperature?
17. Find the COP of a reversible refrigerator operating between 0°C and 30°C.
18. How much work does a refrigerator with COP = 4.2 require to freeze 670 g of water already at its freezing point?
19. BIO The human body can be 25% efficient at converting chemical energy of fuel to mechanical work. Can the body be considered a heat engine, operating on the temperature difference between body temperature and the environment?

Section 19.4 Entropy and Energy Quality

20. Calculate the entropy change associated with melting 1.0 kg of ice at 0°C.
21. BIO You metabolize a 650-kcal burger at your 37°C body temperature. What's the associated entropy increase?
22. You heat 250 g of water from 10°C to 95°C. By how much does the entropy of the water increase?
23. Melting a block of lead already at its melting point results in an entropy increase of 900 J/K. What's the mass of the lead? (*Hint:* Consult Table 17.1.)
24. How much energy becomes unavailable for work in an isothermal process at 440 K, if the entropy increase is 25 J/K?
25. For a gas of 6 molecules confined to a box, find the probability that (a) all the molecules will be found on one side of the box and (b) half the molecules will be found on each side.

Problems

26. A Carnot engine extracts 890 J from a 550 K reservoir during each cycle and rejects 470 J to a cooler reservoir. It operates at 22 cycles per second. Find (a) the work done during each cycle, (b) its efficiency, (c) the temperature of the cool reservoir, and (d) its mechanical power output.
27. The maximum steam temperature in a nuclear power plant is 570 K. The plant rejects heat to a river whose temperature is 0°C in the winter and 25°C in the summer. What are the maximum possible efficiencies for the plant during these seasons?
28. You're engineering an energy-efficient house that will require an average of 4.6 kW to heat on cold winter days. You've designed a photovoltaic system for electric power, which will supply on average 2.0 kW. You propose to heat the house with an electrically operated groundwater-based heat pump. What should you specify as the minimum acceptable COP for the pump if the photovoltaic system supplies its energy?
29. A power plant's electrical output is 750 MW. Cooling water at 15°C flows through the plant at 2.8×10^4 kg/s, and its temperature rises by 8.5°C. Assuming that the plant's only energy loss is to the cooling water, which serves as its low-temperature reservoir, find (a) the rate of energy extraction from the fuel, (b) the plant's efficiency, and (c) its highest temperature.
30. A power plant extracts energy from steam at 250°C and delivers 800 MW of electric power. It discharges waste heat to a river at 30°C. The plant's overall efficiency is 28%. (a) How does this efficiency compare with the maximum possible at these temperatures? (b) Find the rate of waste-heat discharge to the river. (c) How many houses, each requiring 18 kW of heating power, could be heated with the waste heat from this plant?
31. The electric power output of all the thermal electric power plants in the United States is about 2×10^{11} W, and these plants operate at an average efficiency of around 33%. Find the rate at which all these plants use cooling water, assuming an average 5°C rise in cooling-water temperature. Compare with the 1.8×10^7 kg/s average flow at the mouth of the Mississippi River.
32. Consider a Carnot engine operating between temperatures T_h and T_i, where T_i is intermediate between T_h and the ambient temperature T_c (Fig. 19.21). It should be possible to operate a second engine between T_i and T_c. Show that the maximum overall

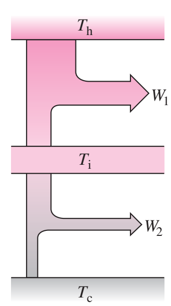


FIGURE 19.21 Problem 32

efficiency of such a two-stage engine is the same as that of a single engine operating between T_h and T_c (which is why combined-cycle power plants achieve high efficiencies).

33. An industrial freezer operates between 0°C and 32°C, consuming electrical energy at the rate of 12 kW. Assuming the freezer is perfectly reversible, (a) what's its COP? (b) How much water at 0°C can it freeze in 1 hour?
34. Use appropriate energy-flow diagrams to analyze the situation in Got It? 19.2; that is, show that using a refrigerator to cool the low-temperature reservoir can't increase the overall efficiency of a Carnot engine when the work input to the refrigerator is included.
35. It costs \$180 to heat a house with electricity in a typical winter month. (Electric heat simply converts all the incoming electrical energy to heat.) What would the monthly heating bill be after switching to an electrically powered heat-pump system with COP = 3.1?
36. A refrigerator maintains an interior temperature of 4°C while its exhaust temperature is 30°C. The refrigerator's insulation is imperfect, and heat leaks in at the rate of 340 W. Assuming the refrigerator is reversible, at what rate must it consume electrical energy to maintain a constant 4°C interior?
37. You operate a store that's heated by an oil furnace supplying 30 kWh of heat from each gallon of oil. You're considering switching to a heat-pump system. Oil costs \$1.75/gallon, and electricity costs 16.5¢/kWh. What's the minimum heat-pump COP that will reduce your heating costs?
38. Use energy-flow diagrams to show that the existence of a perfect heat engine would permit the construction of a perfect refrigerator, thus violating the Clausius statement of the second law.
39. A heat pump extracts energy from groundwater at 10°C and transfers it to water at 70°C to heat a building. Find (a) its COP and (b) its electric power consumption if it supplies heat at the rate of 20 kW. (c) Compare the pump's hourly operating cost with that of an oil furnace if electricity costs 15.5¢/kWh and oil costs \$2.60/gallon and releases about 30 kWh/gal when burned.
40. A reversible engine contains 0.20 mol of ideal monatomic gas, initially at 600 K and confined to 2.0 L. The gas undergoes the following cycle:
    - Isothermal expansion to 4.0 L
    - Isovolumic cooling to 300 K
    - Isothermal compression to 2.0 L
    - Isovolumic heating to 600 K

    (a) Calculate the net heat added during the cycle and the net work done. (b) Determine the engine's efficiency, defined as the ratio of the work done to the heat *absorbed* during the cycle.
41. (a) Determine the efficiency for the cycle shown in Fig. 19.22, using the definition given in the preceding problem. (b) Compare with the efficiency of a Carnot engine operating between the same temperature extremes. Why are the two efficiencies different?

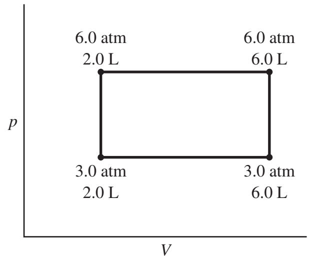


FIGURE 19.22 Problem 41

42. A 0.20-mol sample of an ideal gas goes through the Carnot cycle of Fig. 19.23. Calculate (a) the heat Q_h absorbed, (b) the heat Q_c rejected, and (c) the work done. (d) Use these quantities to determine the efficiency. (e) Find the maximum and minimum temperatures, and show explicitly that the efficiency as defined in Equation 19.1 is equal to the Carnot efficiency of Equation 19.3.

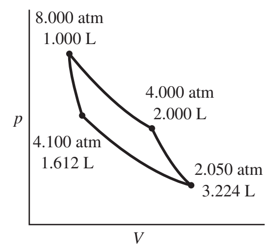


FIGURE 19.23 Problem 42

43. A shallow pond contains 94 Mg of water. In winter, it's entirely frozen. By how much does the entropy of the pond increase when the ice, already at 0°C, melts and then heats to its summer temperature of 15°C?
44. Estimate the rate of entropy increase associated with your body's normal metabolism. **BIO**
45. The temperature of n moles of ideal gas is changed from T_1 to T_2 at constant volume. Show that the corresponding entropy change is $\Delta S = nC_V \ln(T_2/T_1)$.
46. The temperature of n moles of ideal gas is changed from T_1 to T_2 with pressure held constant. Show that the corresponding entropy change is $\Delta S = nC_p \ln(T_2/T_1)$.
47. A 5.0-mol sample of an ideal diatomic gas is at 1.0 atm pressure and 300 K. Find the entropy change if the gas is heated to 500 K (a) at constant volume, (b) at constant pressure, and (c) adiabatically.
48. A 250-g sample of water at 80°C is mixed with 250 g of water at 10°C. Find the entropy changes for (a) the hot water, (b) the cool water, and (c) the system.
49. An ideal gas undergoes a process that takes it from pressure p_1 and volume V_1 to p_2 and V_2, such that $p_1V_1^\gamma = p_2V_2^\gamma$, where γ is the specific heat ratio. Find the entropy change if the process consists of constant-pressure and constant-volume segments. Why does your result make sense?
50. In an adiabatic free expansion, 8.7 mol of ideal gas at 288 K expand 10-fold in volume. How much energy becomes unavailable to do work?
51. Find the entropy change when a 2.4-kg aluminum pan at 155°C is plunged into 3.5 kg of water at 15°C.
52. An engine with mechanical power output 8.5 kW extracts heat from a source at 420 K and rejects it to a 1000-kg block of ice at its melting point. (a) What's its efficiency? (b) How long can it maintain this efficiency if the ice isn't replenished?
53. Gasoline engines operate approximately on the Otto cycle, consisting of two adiabatic and two constant-volume segments. Figure 19.24 shows the Otto cycle for a particular engine. (a) If

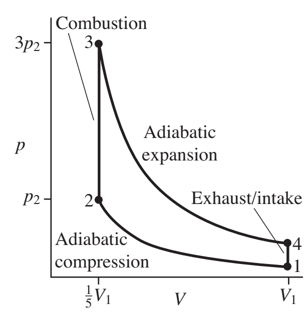


FIGURE 19.24 Problem 53

the gas in the engine has specific heat ratio γ, find the engine's efficiency, assuming all processes are reversible. (b) Find the maximum temperature in terms of the minimum temperature T_{min}. (c) How does the efficiency compare with that of a Carnot engine operating between the same temperature extremes?

54. The compression ratio r of an engine is the ratio of maximum to minimum gas volume. For the engine of the preceding problem, Fig. 19.24 shows that the compression ratio is 5. Find an expression for the engine's efficiency as a function of compression ratio, assuming that pressure continues to triple during the combustion phase.
55. The 54-MW wood-fired McNeil Generating Station in Burlington, Vermont, produces steam at 950°F to drive its turbines, and condensed steam returns to the boiler as 90°F water. (Note the temperatures in °F, used in U.S. engineering situations.) Find McNeil's maximum thermodynamic efficiency, and compare with its actual efficiency of 25%.
56. A 500-g copper block at 80°C is dropped into 1.0 kg of water at 10°C. Find (a) the final temperature and (b) the entropy change of the system.
57. An object's heat capacity is inversely proportional to its absolute temperature: $C = C_0(T_0/T)$, where C_0 and T_0 are constants. Find the entropy change when the object is heated from T_0 to T_1.
58. A Carnot engine extracts heat from a block of mass m and specific heat c initially at temperature T_{h0} but without a heat source to maintain that temperature. The engine rejects heat to a reservoir at constant temperature T_{c}. The engine is operated so its mechanical power output is proportional to the temperature difference $T_{\text{h}} - T_{\text{c}}$:

$$P = P_0 \frac{T_{\text{h}} - T_{\text{c}}}{T_{\text{h0}} - T_{\text{c}}}$$

where T_{h} is the instantaneous temperature of the hot block and P_0 is the initial power. (a) Find an expression for T_{h} as a function of time, and (b) determine how long it takes for the engine's power output to reach zero.

59. In an alternative universe, you've got the impossible: an infinite heat reservoir, containing infinite energy at temperature T_{h}. But you've only got a finite cool reservoir, with initial temperature T_{c0} and heat capacity C. Find an expression for the maximum work you can extract if you operate an engine between these two reservoirs.
60. You're the environmental protection officer for a 35% efficient nuclear power plant that produces 750 MW of electric power, situated on a river whose minimum flow rate is 110 m^3/s. State environmental regulations limit the rise in river temperature from your plant's cooling system to 5°C. Can you achieve this standard if you use river water for all your cooling, or will you need to install cooling towers that transfer some of your waste heat to the atmosphere?
61. Find an expression for the entropy gain when hot and cold water are irreversibly mixed. A corresponding reversible process you can use to calculate this change is to bring each water sample slowly to their common final temperature T_{f} and then mix them. Express your answer in terms of the initial temperatures T_{h} and T_{c}. Assume equal masses of hot and cold water, with constant specific heat c. What's the sign of your answer?
62. Problem 74 of Chapter 16 provided an approximate expression for the specific heat of copper at low absolute temperatures: $c = 31(T/343\text{ K})^3$ J/kg·K. Use this to find the entropy change when 40 g of copper are cooled from 25 K to 10 K. Why is the change negative?
63. The molar specific heat at constant pressure for a certain gas is given by $C_p = a + bT + cT^2$, where $a = 33.6$ J/mol·K, $b = 2.93\times10^{-3}$ J/mol·K^2, and $c = 2.13\times10^{-5}$ J/mol·K^3. Find the entropy change when 2 moles of this gas are heated from 20°C to 200°C.
64. Consider a gas containing an even number N of molecules, distributed among the two halves of a closed box. Find expressions for (a) the total number of microstates and (b) the number of microstates with half the molecules on each side of the box. (You can either work out a formula, or explore the term "combinations" in a math reference source.) (c) Use these results to find the ratio of the probability that all the molecules will be found on one side of the box to the probability that there will be equal numbers on both sides. (d) Evaluate for $N = 4$ and $N = 100$.

Passage Problems

Refrigerators remain among the greatest consumers of electrical energy in most homes, although mandated efficiency standards have decreased their energy consumption by some 80% in the past four decades. In the course of a day, one kitchen refrigerator removes 30 MJ of energy from its contents, in the process consuming 10 MJ of electrical energy. The electricity comes from a 40% efficient coal-fired power plant.

65. The electrical energy
    a. is used to run the light bulb inside the refrigerator.
    b. wouldn't be necessary if the refrigerator had enough insulation.
    c. retains its high-quality status after the refrigerator has used it.
    d. ends up as waste heat rejected to the kitchen environment.
66. The refrigerator's COP is
    a. $\frac{1}{3}$.
    b. 2.
    c. 3.
    d. 4.
67. The fuel energy consumed at the power plant to run this refrigerator for the day is
    a. 12 MJ.
    b. 25 MJ.
    c. 40 MJ.
    d. 75 MJ.
68. The total energy rejected to the surrounding kitchen during the course of the day is
    a. 10 MJ.
    b. 30 MJ.
    c. 40 MJ.
    d. 75 MJ.

Answers to Chapter Questions

Answer to Chapter Opening Question

The second law of thermodynamics prevents us from converting thermal energy to mechanical energy with 100% efficiency, and practical limits on temperature make it hard to achieve 50% efficiency in conventional power plants.

Answers to GOT IT? Questions

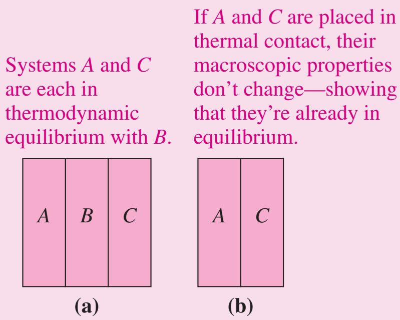
19.1. (a), (c), and (f).
19.2. (c); see Problem 34 for a proof.
19.3. (a) increase; (b) decrease; (c) increase; (d) increase; (e) decrease; (f) increase; (g) increase.

Thermodynamics

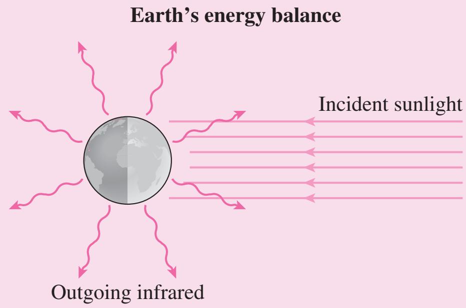
Thermodynamics is the study of heat, temperature, and related phenomena—and their relation to the all-important concept of energy. Thermodynamics provides a macroscopic description in terms of parameters like temperature and pressure.

This contrasts with **statistical mechanics**, which provides a microscopic description in terms of the properties and behavior of molecules.

Thermodynamic equilibrium occurs when two systems are brought into thermal contact and no further changes occur in any macroscopic properties. The **zeroth law of thermodynamics** says that two systems each in thermodynamic equilibrium with a third are also in thermodynamic equilibrium with each other. This law allows us to establish temperature scales and construct thermometers.



Heat is energy that's flowing because of a temperature difference. Important heat-transfer mechanisms include **conduction, convection,** and **radiation**. A system is in **thermal-energy balance** at a fixed temperature when its energy input balances heat transfer to its surroundings.

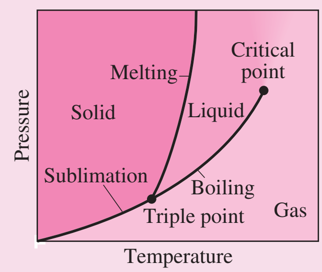


Ideal gases exhibit a simple relation among temperature, pressure, and volume:

$$pV = NkT = nRT$$

This is the **ideal gas law**, with $k = 1.38\times10^{-23}$ J/K and $R = 8.314$ J/K·mol.

Real substances undergo **phase changes** among liquid, solid, and gaseous phases. Substantial **heats of transformation** describe the energies involved in phase changes.



The **first law of thermodynamics** relates the change ΔU in a system's internal energy to the heat Q added *to* the system and the work W done *by* the system:

$$\Delta U = Q - W$$

For an ideal gas, **reversible thermodynamic processes** are described by curves in the pressure–volume diagram. Common processes include **isothermal** (constant temperature), **constant volume, constant pressure**, and **adiabatic** (no heat flow).

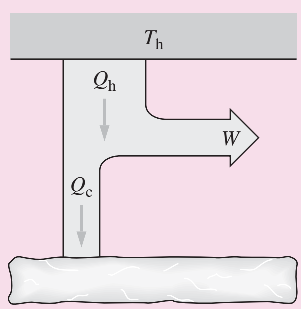
Entropy is a measure of disorder. The **second law of thermodynamics** states that the entropy of a closed system can never decrease. Applied to the heat engines that provide most of humankind's electrical and transportation energy, the second law shows that it's impossible to extract as useful work all the random internal energy of hot objects.

Maximum efficiency (Carnot):

$$e = \frac{W}{Q_h} = 1 - \frac{Q_c}{Q_h} = 1 - \frac{T_c}{T_h}$$

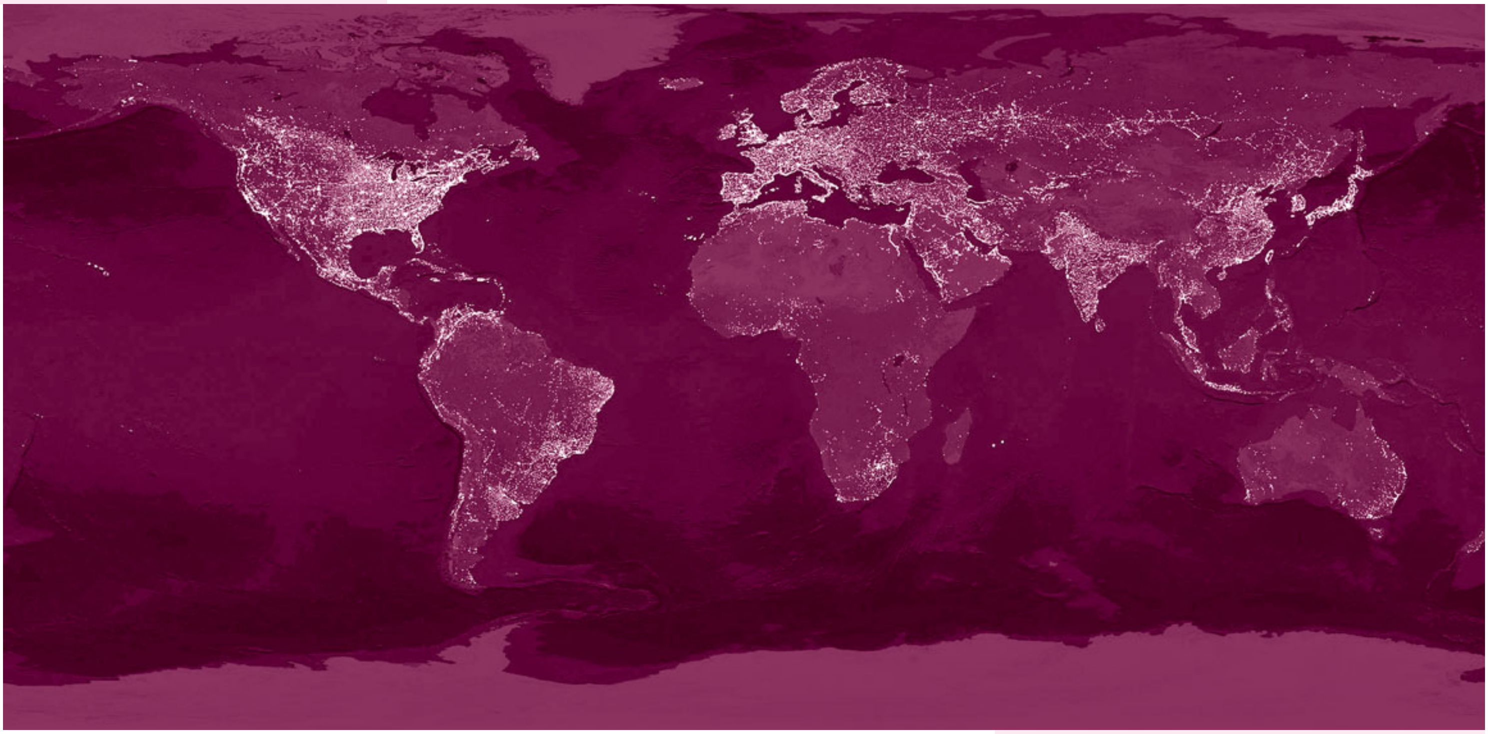
Part Three Challenge Problem

The ideal Carnot engine shown in the figure at right operates between a heat reservoir and a block of ice with mass M. An external energy source maintains the reservoir at a constant temperature T_h. At time $t = 0$, the ice is at its melting point T_0, but it's insulated from everything except the engine, so it's free to change state and temperature. The engine is operated in such a way that it extracts heat from the reservoir at a constant rate P_h. (a) Find an expression for the time t_1 at which the ice is all melted, in terms of the quantities given and any other appropriate thermodynamic parameters. (b) Find an expression for the mechanical power output of the engine as a function of time for times $t > t_1$. (c) Your expression in part (b) holds up only to some maximum time t_2. Why? Find an expression for t_2.



PART FOUR OVERVIEW

Electromagnetism



Electricity constitutes a significant portion of humankind's energy, as evidenced by this composite satellite image of Earth at night. Nearly all that electrical energy is produced by generators, devices that exploit an intimate relation between electricity and magnetism.

Electromagnetism is one of the fundamental forces, and it governs the behavior of matter from the atomic scale to the macroscopic world. Electromagnetic technology, from computer microchips to cell phones and on to large electric motors and generators, is essential to modern society. Even our bodies rely heavily on electromagnetism: Electric signals pace our heartbeat, electrochemical processes transmit nerve impulses, and the electric structure of cell membranes mediates the flow of materials into and out of the cell.

Four fundamental laws describe electricity and magnetism. Two deal separately with the two phenomena, while the others reveal profound connections that make electricity and magnetism aspects of a single phenomenon—electromagnetism. In this part you'll come to understand those fundamental laws, learn how electromagnetism determines the structure and behavior of nearly all matter, and explore the electromagnetic technologies that play so important a role in your life. Finally, you'll see how the laws of electromagnetism lead to electromagnetic waves and thus help us understand the nature of light.

20 Electric Charge, Force, and Field

What's the fundamental criterion for initiating a lightning strike?

New Concepts, New Skills

By the end of this chapter you should be able to

- Recognize the fundamentally electric nature of matter, with electric charge as an intrinsic property (20.1).
- Describe the behavior of the electric force using Coulomb's law (20.2).
- Calculate forces between electric charges (20.2).
- Explain the concept of electric field (20.3), and calculate the fields of discrete and continuous distributions of charge (20.4).
- Describe quantitatively how charges respond to electric fields (20.5).

Connecting Your Knowledge

- This chapter uses the concept of force and applies Newton's second law to electric forces (4.2).
- The idea of torque is used in discussing electric dipoles (10.2, 11.2).
- You'll also need to be familiar with integration from your calculus class and with the integration strategy we developed earlier (9.1).

What holds your body together? What keeps a skyscraper standing? What holds your car on the road as you round a turn? What governs the electronic circuitry in your computer or MP3 player, or provides the tension in your climbing rope? What enables a plant to make sugar from sunlight and simple chemicals? What underlies the awesome beauty of lightning? The answer, in all cases, is the **electric force**. With the exception of gravity, all the forces we've encountered in mechanics—including tension forces, normal forces, compression forces, and friction—are based on electric interactions; so are the forces responsible for all of chemistry and biology. The electric force, in turn, involves a fundamental property of matter—namely, electric charge.

20.1 Electric Charge

Electric charge is an intrinsic property of the electrons and protons that, along with uncharged neutrons, make up ordinary matter. What is electric charge? At the most fundamental level we don't know. We don't know what mass "really" is either, but we're familiar with it because we've spent our lives pushing objects around. Similarly, our knowledge of electric charge results from observing the behavior of charged objects.

Charge comes in two varieties, which Benjamin Franklin designated *positive* and *negative*. Those names are useful because the total charge on an object—the object's **net charge**—is the algebraic sum of its constituent charges. Like charges repel, and opposites attract, a fact that constitutes a qualitative description of the electric force.

Quantities of Charge

All electrons carry the same charge, and all protons carry the same charge. The proton's charge has *exactly* the same magnitude as the electron's, but with opposite sign. Given that electrons and protons differ substantially in other properties—like mass—this electric relation is remarkable. Exercise 13 shows how dramatically different our world would be if there were even a slight difference between the magnitudes of the electron and proton charges.

The magnitude of the electron or proton charge is the **elementary charge** e. Electric charge is **quantized**; that is, it comes only in discrete amounts. In a famous experiment in 1909, the American physicist R. A. Millikan measured the charge on small oil drops and found it was always a multiple of a basic value we now know as the elementary charge.

Elementary particle theories show that the fundamental charge is actually $\frac{1}{3}e$. Such "fractional charges" reside on quarks, the building blocks of protons, neutrons, and many other particles. Quarks always join to produce particles with integer multiples of the full elementary charge, and it seems impossible to isolate individual quarks.

The SI unit of charge is the **coulomb** (C), named for the French physicist Charles Augustin de Coulomb (1736–1806). Although the coulomb's formal definition is in terms of electric current, it's convenient to describe 1 coulomb as being about 6.25×10^{18} elementary charges, making the elementary charge approximately 1.60×10^{-19} C.

Charge Conservation

Electric charge is a conserved quantity, meaning that the net charge in a closed region remains constant. Charged particles may be created or annihilated, but always in pairs of equal and opposite charge. The net charge always remains the same.

20.2 Coulomb's Law

Rub a balloon; it gets charged and sticks to your clothing. Charge another balloon, and the two repel (Fig. 20.1). Socks cling to your clothes as they come from the dryer, and bits of Styrofoam cling annoyingly to your hands. Walk across a carpet, and you'll feel a shock when you touch a doorknob. All these are common examples where you're directly aware of electric charge.

Electricity would be unimportant if the only significant electric interactions were these obvious ones. In fact, the electric force dominates all interactions of everyday matter, from the motion of a car to the movement of a muscle. It's just that matter on a large scale is almost perfectly neutral, meaning it carries zero net charge. Therefore, electric effects aren't obvious. But at the molecular level, the electric nature of matter is immediately evident (Fig. 20.2).

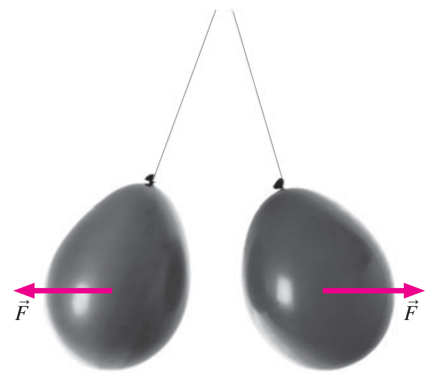


FIGURE 20.1 Two balloons carrying similar electric charges repel each other.

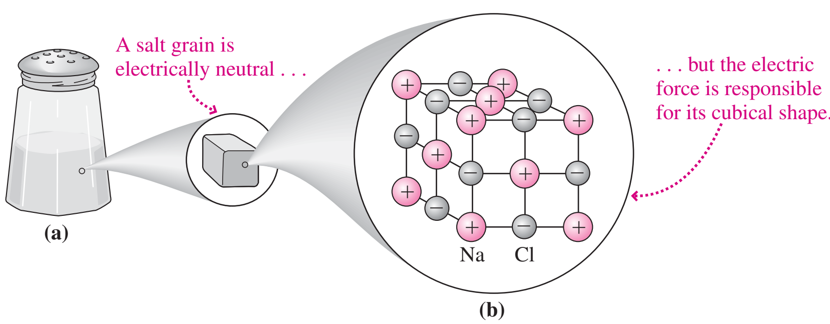


FIGURE 20.2 (a) A single salt grain is electrically neutral, so the electric force isn't obvious. (b) Actually, the electric force determines the structure of salt.

Attraction and repulsion of electric charges imply a force. Joseph Priestley and Charles Augustin de Coulomb investigated this force in the late 1700s. They found that the force between two charges acts along the line joining them, with the magnitude proportional to the product of the charges and inversely proportional to the square of the distance between them. **Coulomb's law** summarizes these results:

$$\vec{F}_{12} = \frac{kq_1q_2}{r^2}\hat{r} \quad \text{(Coulomb's law)} \tag{20.1}$$

where $\vec{F}_{12}$ is the force charge q_1 exerts on q_2 and r is the distance between the charges. The proportionality constant k has SI value $9.0\times10^9\ \text{N}\cdot\text{m}^2/\text{C}^2$. Force is a vector, and $\hat{r}$ is a unit vector that helps determine its direction. Figure 20.3 shows that $\hat{r}$ lies on a line passing through the two charges and points in the direction *from* q_1 *toward* q_2. Reverse the roles of q_1 and q_2, and you'll see that $\vec{F}_{21}$ has the same magnitude as $\vec{F}_{12}$ but the opposite direction; thus Coulomb's law obeys Newton's third law. Figure 20.3 also shows that the force is in the same direction as the unit vector when the charges have the same sign, but opposite the unit vector when the charges have different signs. Thus Coulomb's law accounts for the fact that like charges repel and opposites attract.

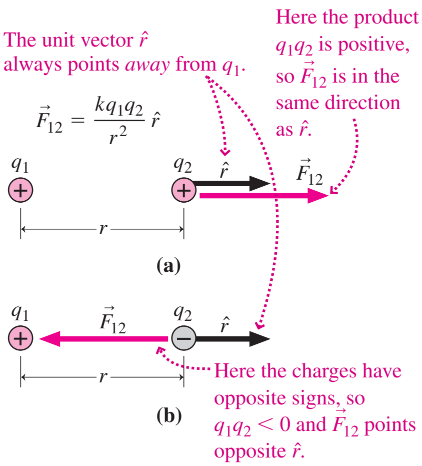


FIGURE 20.3 Quantities in Coulomb's law for calculating the force $\vec{F}_{12}$ that q_1 exerts on q_2.

PROBLEM-SOLVING STRATEGY 20.1 Coulomb's Law

The key to using Coulomb's law is to remember that force is a vector, and to realize that Coulomb's law in the form of Equation 20.1 gives both the magnitude and direction of the electric force. Dealing carefully with vector directions is especially important in situations with more than two charges.

INTERPRET First, make sure you're dealing with the electric force alone. Identify the charge or charges on which you want to calculate the force. Next, identify the charge or charges producing the force. These comprise the **source charge**.

DEVELOP Begin with a drawing that shows the charges, as in Fig. 20.4. If you're given charge coordinates, place the charges on the coordinate system; if not, choose a suitable coordinate system. For each source charge, determine the unit vector(s) in Equation 20.1. If the charges lie along or parallel to a coordinate axis, then the unit vector will be one of the unit vectors $\hat{\imath}$, $\hat{\jmath}$, or $\hat{k}$, perhaps with a minus sign. In Fig. 20.4, the force on q_3 due to q_1 is such a case. When the two charges don't lie on a coordinate axis, like q_1 and q_2 in Fig. 20.4, you can find the unit vector by noting that the displacement vector $\vec{r}_{12}$ points in the desired direction, from the source charge to the charge experiencing the force. Dividing $\vec{r}_{12}$ by its own magnitude then gives the unit vector in the direction of $\vec{r}_{12}$: $\hat{r} = \vec{r}_{12}/r_{12}$.

EVALUATE For each source charge, determine the electric force using Equation 20.1,

$$\vec{F}_{12} = (kq_1q_2/r^2)\hat{r}$$

with $\hat{r}$ the unit vector you've just found.

ASSESS As always, assess your answer to see that it makes sense. Is the direction of the force you found consistent with the signs and placements of the charges giving rise to the force?

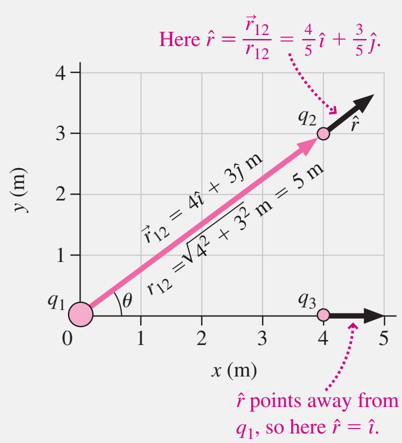


FIGURE 20.4 Finding unit vectors.

GOT IT? 20.1 Charge q_1 is located at $x = 1$ m, $y = 0$. What should you use for the unit vector $\hat{r}$ in Coulomb's law if you're calculating the force q_1 exerts on a charge q_2 located at (a) the origin and (b) the point $x = 0$, $y = 1$ m? Explain why you can answer without knowing the sign of either charge.

EXAMPLE 20.1 Finding the Force: Two Charges

A 1.0-μC charge is at $x = 1.0$ cm, and a -1.5-μC charge is at $x = 3.0$ cm. What force does the positive charge exert on the negative one? How would the force change if the distance between the charges tripled?

INTERPRET Following our strategy, we identify the -1.5-μC charge as the one on which we want to find the force and the 1-μC charge as the source charge.

DEVELOP We're given the coordinates $x_1 = 1.0$ cm and $x_2 = 3.0$ cm. Our drawing, Fig. 20.5, shows both charges at their positions on the x-axis. With the source charge q_1 to the left, the unit vector in the direction from q_1 toward q_2 is $\hat{\imath}$.

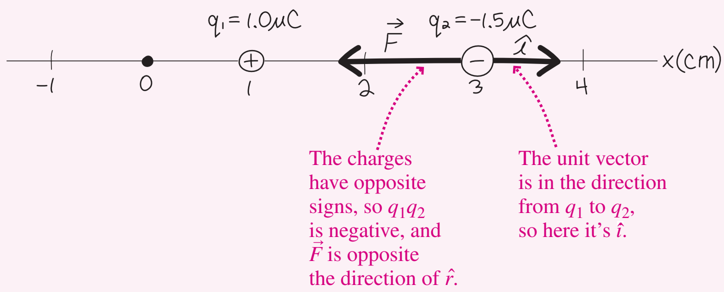


FIGURE 20.5 Sketch for Example 20.1.

EVALUATE Now we use Coulomb's law to evaluate the force:

$$\begin{aligned}\vec{F}_{12} &= \frac{kq_1q_2}{r^2}\hat{r}\\ &= \frac{(9.0\times10^9\ \mathrm{N\cdot m^2/C^2})(1.0\times10^{-6}\ \mathrm{C})(-1.5\times10^{-6}\ \mathrm{C})}{(0.020\ \mathrm{m})^2}\hat{\imath}\\ &= -34\hat{\imath}\ \mathrm{N}\end{aligned}$$

This force is for a separation of 2 cm; if that distance tripled, the force would drop by a factor of $1/3^2$, to $-3.8\hat{\imath}$ N.

ASSESS Make sense? Although the unit vector $\hat{\imath}$ points in the $+x$-direction, the charges have opposite signs and that makes the force direction opposite the unit vector, as shown in Fig. 20.5. In simpler terms, we've got two opposite charges, so they attract. That means the force exerted on a charge at $x = 3$ cm by an opposite charge at $x = 1$ cm had better be in the $-x$-direction. ■

CONCEPTUAL EXAMPLE 20.1 Gravity and the Electric Force

The electric force between elementary particles is far stronger than the gravitational force, yet gravity is much more obvious in everyday life. Why?

EVALUATE Gravity and the electric force obey similar inverse-square laws, and the magnitude of the force is proportional to the product of the masses or charges. There's a big difference, though: There's only one kind of mass, and gravity is always attractive, so large concentrations of mass—like a planet—result in strong gravitational forces. But charge comes in two varieties, and opposites attract, so large accumulations of matter tend to be electrically neutral, and their electrical interactions aren't obvious.

ASSESS Ironically, it's the very strength of the electric force that makes it less obvious in everyday life. Opposite charges bind strongly, making bulk matter electrically neutral and its electrical interactions subtle.

MAKING THE CONNECTION Compare the magnitudes of the electric and gravitational forces between an electron and a proton.

EVALUATE Equation 8.1 gives the gravitational force: $F_g = Gm_em_p/r^2$. Equation 20.1 gives the electric force: $|F_E| = ke^2/r^2$, where we wrote e^2 because the electron and proton charges have the same magnitude. We aren't given the distance, but that doesn't matter because both forces have the same inverse-square dependence. The ratio of the force magnitudes is huge: $F_E/F_g = ke^2/Gm_em_p = 2.3\times10^{39}$!

Point Charges and the Superposition Principle

Coulomb's law is strictly true only for **point charges**—charged objects of negligible size. Electrons and protons can usually be treated as point charges; so, approximately, can any two charged objects if their separation is large compared with their size. But often we're interested in the electric effects of **charge distributions**—arrangements of charge spread over space. Charge distributions are present in molecules, a memory cell in a computer

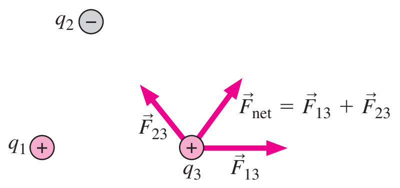


FIGURE 20.6 The superposition principle lets us add vectorially the forces from two or more charges.

chip, your heart, and a thundercloud. We need to combine the effects of two or more charges to find the electric effects of such charge distributions.

Figure 20.6 shows two charges q_1 and q_2 that constitute a simple charge distribution. We want to know the net force these exert on a third charge q_3. To find that net force, you might calculate the forces $\vec{F}_{13}$ and $\vec{F}_{23}$ from Equation 20.1, and then vectorially add them. And you'd be right: The force that q_1 exerts on q_3 is unaffected by the presence of q_2, and vice versa, so you can apply Coulomb's law separately to the pairs q_1q_3 and q_2q_3 and then combine the results. That may seem obvious, but nature needn't have been so simple.

The fact that electric forces add vectorially is called the **superposition principle**. Our confidence in this principle is ultimately based on experiments that show electric and indeed electromagnetic phenomena behave according to the principle. With superposition we can solve relatively complicated problems by breaking them into simpler parts. If the superposition principle didn't hold, the mathematical description of electromagnetism would be far more complicated.

Although the force that one point charge exerts on another decreases with the inverse square of the distance between them, the same is not necessarily true of the force resulting from a charge distribution. The next example provides a case in point.

EXAMPLE 20.2 Finding the Force: Raindrops

Charged raindrops are ultimately responsible for lightning, producing substantial electric charge within specific regions of a thundercloud. Suppose two drops with equal charge q are on the x-axis at $x = \pm a$. Find the electric force on a third drop with charge Q at an arbitrary point on the y-axis.

INTERPRET Coulomb's law and the superposition principle apply, and we identify Q as the charge for which we want the force. The two charges q are the source charges.

DEVELOP Figure 20.7 is our drawing, showing the charges, the individual force vectors, and their sum. The drawing shows that the distance r in Coulomb's law is the hypotenuse $\sqrt{a^2 + y^2}$. It's clear from symmetry that the net force is in the y-direction, so we need to find only the y-components of the unit vectors. The y-components are clearly the same for each, and the drawing shows that they're given by $\hat{r}_y = y/\sqrt{a^2 + y^2}$.

EVALUATE From Coulomb's law, the y-component of the force from each q is $F_y = (kqQ/r^2)\hat{r}_y$, and the net force on Q becomes

$$\vec{F} = 2\left(\frac{kqQ}{a^2 + y^2}\right)\left(\frac{y}{\sqrt{a^2 + y^2}}\right)\hat{\jmath} = \frac{2kqQy}{(a^2 + y^2)^{3/2}}\hat{\jmath}$$

The factor of 2 comes from the two charges q, which contribute equally to the net force.

ASSESS Make sense? Evaluating $\vec{F}$ at $y = 0$ gives zero force. Here, midway between the two charges, Q experiences equal but opposite forces and the net force is zero. At large distances $y \gg a$, on the other hand, we can neglect a^2 compared with y^2, and the force becomes $\vec{F} = k(2q)Q\hat{\jmath}/y^2$. This is just what we would expect from a single charge $2q$ a distance y from Q—showing that the system of two charges acts like a single charge $2q$ at distances that are large compared with the charge separation. In between our two extremes the behavior of force with distance is more complicated; in fact, its magnitude initially increases as Q moves away from the origin and then begins to decrease.

In drawing Fig. 20.7, we tacitly assumed that q and Q have the same signs. But our analysis holds even if they don't; then the product qQ is negative, and the forces actually point opposite the directions shown in Fig. 20.7. ■

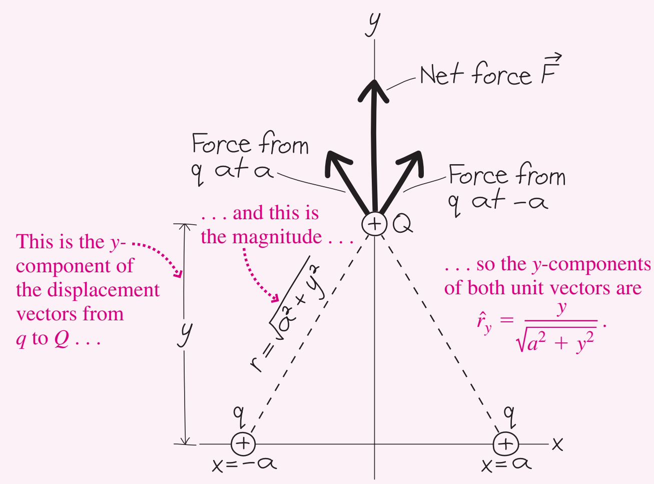


FIGURE 20.7 The force on Q is the vector sum of the forces from the individual charges.

20.3 The Electric Field

In Chapter 8 we defined the gravitational field at a point as the gravitational force per unit mass that an object at that point would experience. In that context, we can think of $\vec{g}$ as the *force per unit mass* that any object would experience due to Earth's gravity. So we can picture the gravitational field as a set of vectors giving the magnitude and direction of the gravitational force per unit mass at each point, as shown in Fig. 20.8*a*.

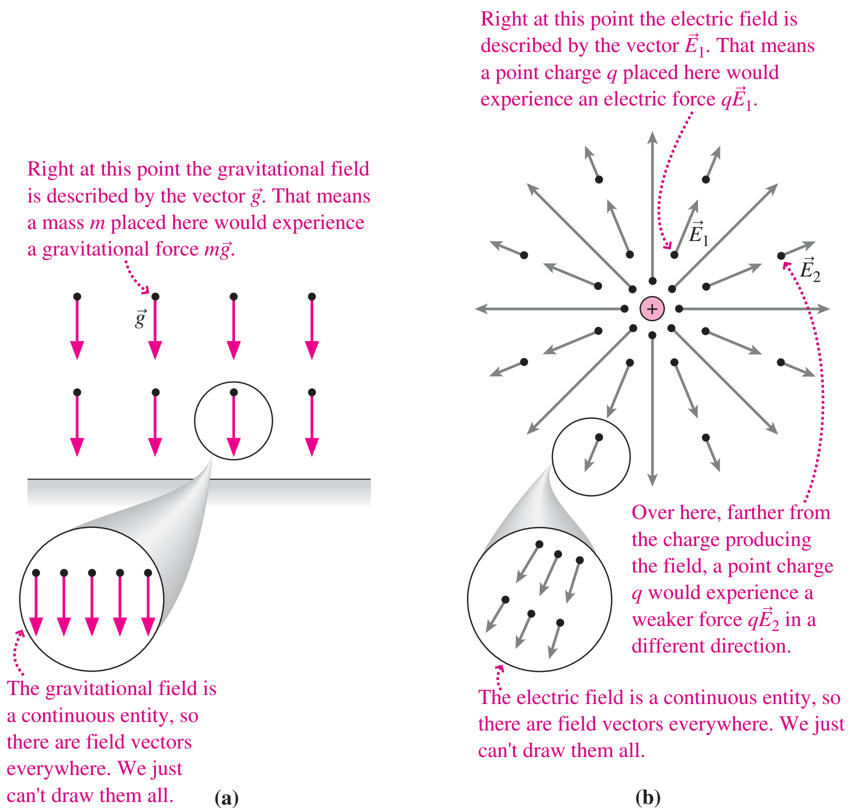


FIGURE 20.8 (a) Gravitational and (b) electric fields, here represented as sets of vectors.

We can do the same thing with the electric force, defining the **electric field** as the force per unit charge:

> The electric field at any point is the force per unit charge that a charge would experience at that point. Mathematically,
>
> $$\vec{E} = \frac{\vec{F}}{q} \qquad \text{(electric field)} \tag{20.2a}$$

The electric field exists at every point in space. When we represent the field by vectors, we can't draw one everywhere, but that doesn't mean there isn't a field at all points. Furthermore, we draw vectors as extended arrows, but each vector represents the field at only one point—namely, the tail end of the vector. Figure 20.8*b* illustrates this for the electric field of a point charge.

The field concept leads to a shift in our thinking about forces. Instead of the action-at-a-distance idea that Earth reaches across empty space to pull on the Moon, the field concept says that Earth creates a gravitational field and the Moon responds to the field at its location. Similarly, a charge creates an electric field throughout the space surrounding it. A second charge then responds to the field at its immediate location. Although the field reveals itself only through its effect on a charge, the field nevertheless exists at all points, whether or not charges are present. Right now you probably find the field concept a bit abstract, but as you advance in your study of electromagnetism you'll come to appreciate that fields are an essential feature of our universe, every bit as real as matter itself.

We can use Equation 20.2a as a prescription for measuring electric fields. Place a point charge at some point, measure the electric force it experiences, and divide by the charge to get the field. In practice, we need to be careful because the field generally arises from some distribution of source charges. If the charge we're using to probe the field—the **test charge**—is large, the field it creates may disturb the source charges, altering their configuration and thus the field they create. For that reason, it's important to use a very small test charge.

If we know the electric field $\vec{E}$ at a point, we can rearrange Equation 20.2a to find the force on any point charge q placed at that point:

$$\vec{F} = q\vec{E} \qquad \text{(electric force and field)} \tag{20.2b}$$

If the charge q is positive, then this force is in the same direction as the field, but if q is negative, then the force is opposite to the field direction.

Equations 20.2 show that the units of electric field are newtons per coulomb. Fields of hundreds to thousands of N/C are commonplace, while fields of 3 MN/C will tear electrons from air molecules.

EXAMPLE 20.3 Force and Field: Inside a Lightning Storm

A charged raindrop carrying 10 μC experiences an electric force of 0.30 N in the $+x$-direction. What's the electric field at its location? What would the force be on a -5.0-μC drop at the same point?

INTERPRET In this problem we distinguish between an electric force and an electric field. The electric field exists with or without the charged raindrop present, and the electric force arises when the charged raindrop is in the electric field.

DEVELOP Knowing the electric force and the charge on the raindrop, we can use Equation 20.2a, $\vec{E} = \vec{F}/q$, to get the electric field. Once we know the field, we can use Equation 20.2b, $\vec{F} = q\vec{E}$, to calculate the force that would act if a different charge were at the same point.

EVALUATE Equation 20.2a gives the electric field:

$$\vec{E} = \frac{\vec{F}}{q} = \frac{0.30\hat{\imath}\ \text{N}}{10\ \mu\text{C}} = 30\hat{\imath}\ \text{kN/C}$$

Acting on a -5.0-μC charge, this field would result in a force

$$\vec{F} = q\vec{E} = (-5.0\ \mu\text{C})(30\hat{\imath}\ \text{kN/C}) = -0.15\hat{\imath}\ \text{N/C}$$

ASSESS Make sense? The force on the second charge is opposite the direction of the field because now we've got a negative charge in the same field.

✓TIP The Field Is Independent of the Test Charge

Does the electric field in this example point in the $-x$-direction when the charge is negative? No. The field is independent of the particular charge experiencing that field. Here the electric field points in the $+x$-direction *no matter what charge* you put in the field. For a positive charge, the force $q\vec{E}$ points in the *same* direction as the field; for a negative charge, $q < 0$, the force is *opposite* the field.

■

The Field of a Point Charge

Once we know the field of a charge distribution, we can calculate its effect on other charges. The simplest charge distribution is a single point charge. Coulomb's law gives the force on a test charge q_{test} located a distance r from a point charge q: $\vec{F} = (kqq_{test}/r^2)\hat{r}$, where $\hat{r}$ is a unit vector pointing *away* from q. The electric field arising from q is the force per unit charge, or

$$\vec{E} = \frac{\vec{F}}{q_{test}} = \frac{kq}{r^2}\hat{r} \qquad \text{(field of a point charge)} \tag{20.3}$$

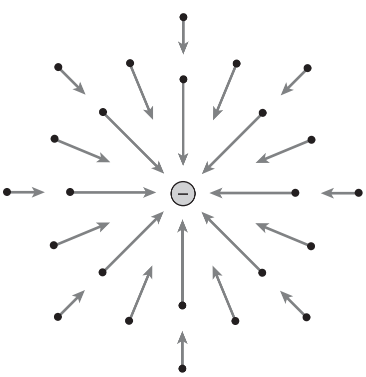


FIGURE 20.9 Field vectors for a negative point charge.

Since it's so closely related to Coulomb's law for the electric force, we also refer to Equation 20.3 as Coulomb's law. The equation contains no reference to the test charge q_{test} because the field of q exists independently of any other charge. Since $\hat{r}$ always points *away* from q, the direction of $\vec{E}$ is radially outward if q is positive and radially inward if q is negative. Figure 20.9 shows some field vectors for a negative point charge, analogous to those of the positive point charge in Fig. 20.8*b*.

20.4 Fields of Charge Distributions

Since the electric force obeys the superposition principle, so does the electric field. That means the field of a charge distribution is the vector sum of the fields of the individual point charges making up the distribution:

$$\vec{E} = \vec{E}_1 + \vec{E}_2 + \vec{E}_3 + \cdots = \sum_i \vec{E}_i = \sum_i \frac{kq_i}{r_i^2}\hat{r}_i \tag{20.4}$$

Here the $\vec{E}_i$'s are the fields of the point charges q_i located at distances r_i from the point where we're evaluating the field—called, appropriately, the **field point**. The $\hat{r}_i$'s are unit vectors pointing *from* each point charge *toward* the field point. In principle, Equation 20.4 gives the electric field of *any* charge distribution. In practice, the process of summing the individual field vectors is often complicated unless the charge distribution contains relatively few charges arranged in a symmetric way.

Finding electric fields using Equation 20.4 involves the same strategy we introduced for finding the electric force; the only difference is that there's no charge to experience the force. The first step then involves identifying the field point. We still need to find the appropriate unit vectors and form the vector sum in Equation 20.4.

EXAMPLE 20.4 Finding the Field: Two Protons

Two protons are 3.6 nm apart. Find the electric field at a point between them, 1.2 nm from one of the protons. Then find the force on an electron at this point.

INTERPRET We follow our electric-force strategy, except that instead of identifying the charge experiencing the force, we identify the field point as being 1.2 nm from one proton. The source charges are the two protons; they produce the field we're interested in.

DEVELOP Let's have the protons define the x-axis, as drawn in Fig. 20.10. Then the unit vector $\hat{r}_1$ from the left-hand proton toward the field point is $+\hat{\imath}$, while $\hat{r}_2$ from the right-hand proton toward P is $-\hat{\imath}$.

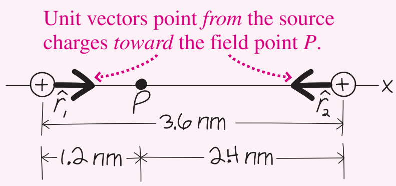


FIGURE 20.10 Finding the electric field at P.

EVALUATE We now evaluate the field at P using Equation 20.4:

$$\vec{E} = \vec{E}_1 + \vec{E}_2 = \frac{ke}{r_1^2}\hat{\imath} + \frac{ke}{r_2^2}(-\hat{\imath}) = ke\left(\frac{1}{r_1^2} - \frac{1}{r_2^2}\right)\hat{\imath}$$

We wrote e for q here because the protons' charge is the elementary charge.

Using $e = 1.6\times10^{-19}$ C, $r_1 = 1.2$ nm, and $r_2 = 2.4$ nm gives $\vec{E} = 750\hat{\imath}$ MN/C. An electron at P will therefore experience a force $\vec{F} = q\vec{E} = -e\vec{E} = -0.12\hat{\imath}$ nN.

ASSESS Make sense? The field points in the positive x-direction, reflecting the fact that P is closer to the left-hand proton with its stronger field at P. The force on the electron, on the other hand, is in the $-x$-direction; that's because the electron is negative (we used $q = -e$ for its charge), so the force it experiences is opposite the field. That field of almost 1 GN/C sounds huge—but that's not unusual at the microscopic scale, where we're close to individual elementary particles. ■

The Electric Dipole

One of the most important charge distributions is the **electric dipole**, consisting of two point charges of equal magnitude but opposite sign. Many molecules are essentially dipoles, so understanding the dipole helps explain molecular behavior (Fig. 20.11). During contraction the heart muscle becomes essentially a dipole, and physicians performing electrocardiography are measuring, among other things, the strength and orientation of that dipole. Technological devices, including radio and TV antennas, also make use of the dipole configuration.

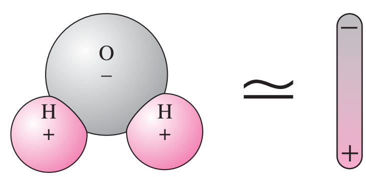


FIGURE 20.11 A water molecule behaves like an electric dipole. Its net charge is zero, but regions of positive and negative charge are separated.

EXAMPLE 20.5 The Electric Dipole: Modeling a Molecule

A molecule may be modeled approximately as a positive charge q at $x = a$ and a negative charge $-q$ at $x = -a$. Evaluate the electric field on the y-axis, and find an approximate expression valid at large distances $(y \gg a)$.

INTERPRET Here's another example where we'll use our strategy in applying Equation 20.4 to calculate the field of a charge distribution. We identify the field point as being anywhere on the y-axis and the source charges as being $\pm q$.

DEVELOP Figure 20.12 (next page) is our drawing. The individual unit vectors point from the two charges toward the field point, but the *negative* charge contributes a field *opposite* its unit vector; we've indicated the individual fields in Fig. 20.12. Here symmetry makes the

(continued)

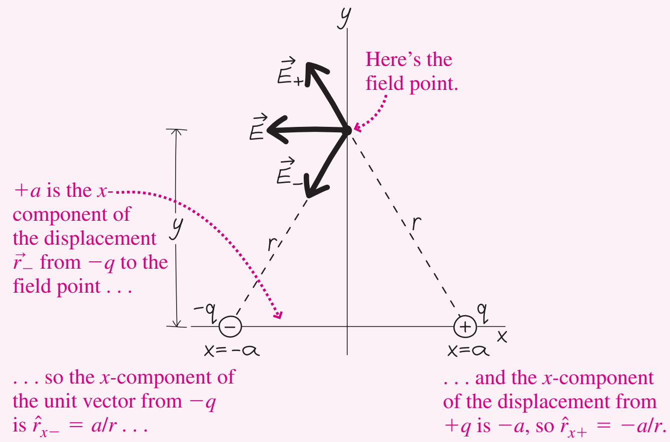


FIGURE 20.12 Finding the field of an electric dipole.

y-components cancel, giving a net field in the $-x$-direction. So we need only the x-components of the unit vectors, which Fig. 20.12 shows are $\hat{r}_{x-} = a/r$ for the negative charge at $-a$ and $\hat{r}_{x+} = -a/r$ for the positive charge at a.

EVALUATE We then evaluate the field using Equation 20.4:

$$\vec{E} = \frac{k(-q)}{r^2}\left(\frac{a}{r}\right)\hat{\imath} + \frac{kq}{r^2}\left(-\frac{a}{r}\right)\hat{\imath} = -\frac{2kqa}{(a^2+y^2)^{3/2}}\hat{\imath}$$

where in the last step we used $r = \sqrt{a^2+y^2}$. For $y \gg a$ we can neglect a^2 compared with y^2, giving

$$\vec{E} \simeq -\frac{2kqa}{y^3}\hat{\imath} \qquad (y \gg a)$$

ASSESS Make sense? The dipole has no net charge, so at large distances its field can't have the inverse-square drop-off of a point-charge field. Instead the dipole field falls faster, here as $1/y^3$.

✓TIP Approximations

Making approximations requires care. Here we're basically asking for the field when y is so large that a is negligible *compared with* y. So we neglect a^2 compared with y^2 when the two are summed, but we *don't* neglect a when it appears in the numerator.

Example 20.5 shows that the dipole field at large distances decreases as the inverse *cube* of distance. Physically, that's because the dipole has zero *net* charge. Its field arises entirely from the slight separation of two opposite charges. Because of this separation the dipole field isn't exactly zero, but it's weaker and more localized than the field of a point charge. Many complicated charge distributions exhibit the essential characteristic of a dipole—that is, they're neutral but consist of separated regions of positive and negative charge—and at large distances such distributions all have essentially the same field configuration.

At large distances the dipole's physical characteristics q and a enter the equation for the electric field only through the product qa. We could double q and halve a, and the dipole's electric field would remain unchanged. At large distances, therefore, a dipole's electric properties are characterized completely by its **electric dipole moment** p, defined as the product of the charge q and the separation d between the two charges making up the dipole:

$$p = qd \qquad \text{(dipole moment)} \tag{20.5}$$

In Example 20.5 the charge separation was $d = 2a$, so there the dipole moment was $p = 2aq$. In terms of the dipole moment, the field in Example 20.5 can then be written

$$\vec{E} = -\frac{kp}{y^3}\hat{\imath} \qquad \left(\begin{array}{c}\text{dipole field for } y \gg a,\\ \text{on perpendicular bisector}\end{array}\right) \tag{20.6a}$$

You can show in Problem 50 that the field on the dipole axis is given by

$$\vec{E} = \frac{2kp}{x^3}\hat{\imath} \qquad \left(\begin{array}{c}\text{dipole field}\\ \text{for } x \gg a \text{, on axis}\end{array}\right) \tag{20.6b}$$

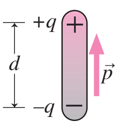


FIGURE 20.13 The dipole moment vector has magnitude $p = qd$ and points from the negative toward the positive charge.

Because the dipole isn't spherically symmetric, its field depends not only on distance but also on orientation; for instance, Equations 20.6 show that the field along the dipole axis at a given distance is twice as strong as along the bisector. So it's important to know the orientation of a dipole in space, and therefore we generalize our definition of the dipole moment to make it a vector of magnitude $p = qd$ in the direction from the negative toward the positive charge (Fig. 20.13).

GOT IT? 20.2 Far from a charge distribution, you measure an electric field strength of 800 N/C. What will the field strength be if you double your distance from the charge distribution, if the distribution consists of (a) a point charge or (b) a dipole?

Continuous Charge Distributions

Although any charge distribution ultimately consists of pointlike electrons and protons, it would be impossible to sum all the field vectors from the 10^{23} or so particles in a typical piece of matter. Instead, it's convenient to make the approximation that charge is spread continuously over the distribution. If the charge distribution extends throughout a volume, we describe it in terms of the **volume charge density** ρ, with units of C/m^3. For charge distributions spread over surfaces or lines, the corresponding quantities are the **surface charge density** σ (C/m^2) and the **line charge density** λ (C/m).

To calculate the field of a continuous charge distribution, we divide the charged region into very many small charge elements dq, each small enough that it's essentially a point charge. Each dq then produces an electric field $d\vec{E}$ given by Equation 20.3: $d\vec{E} = (k\,dq/r^2)\hat{r}$. We then form the vector sum of all the $d\vec{E}$'s (Fig. 20.14). In the limit of infinitely many infinitesimally small dq's and their corresponding $d\vec{E}$'s, that sum becomes an integral and we have

$$\vec{E} = \int d\vec{E} = \int \frac{k\,dq}{r^2}\hat{r} \quad \left(\begin{array}{c}\text{field of a continuous}\\ \text{charge distribution}\end{array}\right) \qquad (20.7)$$

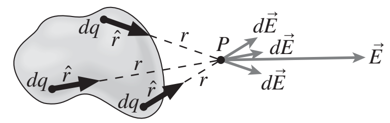


FIGURE 20.14 The electric field at P is the vector sum of the fields $d\vec{E}$ arising from the individual charge elements dq, each calculated using the appropriate distance r and unit vector $\hat{r}$.

The limits of this integral include the entire charge distribution.

Calculating the field of a continuous charge distribution involves the same strategy we've already used: We identify the field point and the source charges—although now the source is a continuous charge distribution. Summing the individual field contributions now presents us with an integral, and that means writing the unit vectors $\hat{r}$ and distances r in terms of coordinates over which we can integrate. Setting up the integral involves the same strategy we outlined in Chapter 9 to find the center of mass of a continuous distribution of matter.

EXAMPLE 20.6 Evaluating the Field: A Charged Ring

A ring of radius a carries a charge Q distributed evenly over the ring. Find an expression for the electric field at any point on the axis of the ring.

INTERPRET We identify the field point as lying anywhere on the ring's axis, and the source charge as the entire ring.

DEVELOP Let's take the x-axis to coincide with the ring axis, with the center of the ring at $x = 0$ (Fig. 20.15). The figure shows that the y-components of the field contributions from pairs of charge elements on opposite sides of the ring cancel; therefore, the net field points in the $+x$-direction (for $x > 0$) and we need only the x-components of the unit vectors. Those are the same for all unit vectors—namely, $\hat{r}_x = x/r$.

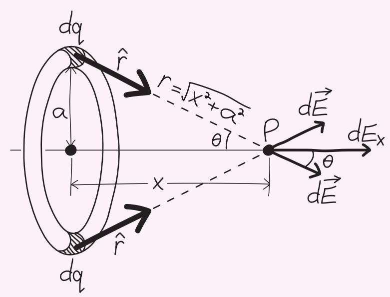


FIGURE 20.15 The electric field of a charged ring points along the ring axis, since field components perpendicular to the axis cancel in pairs.

EVALUATE We're now ready to set up the integral in Equation 20.7. Here each charge element contributes the same amount $dE_x = (k\,dq/r^2)\hat{r}_x = (k\,dq/r^2)(x/r)$ to the field. Figure 20.15 shows that $r = \sqrt{x^2 + a^2} = (x^2 + a^2)^{1/2}$, so the integral becomes

$$E = \int_{\text{ring}} dE_x = \int_{\text{ring}} \frac{kx\,dq}{(x^2 + a^2)^{3/2}} = \frac{kx}{(x^2 + a^2)^{3/2}}\int_{\text{ring}} dq$$

The last step follows because we have a fixed field point P, so its coordinate x is a constant for the integration. But the remaining integral is just the sum of all the charge elements on the ring—namely, the total charge Q. So our result becomes

$$E = \frac{kQx}{(x^2 + a^2)^{3/2}} \qquad \text{(on-axis field, charged ring)}$$

This is the magnitude; the direction is along the x-axis, away from the ring if Q is positive and toward it if Q is negative.

ASSESS Make sense? At $x = 0$ the field is zero. Of course: A charge placed at the ring center is pulled (or pushed) equally in all directions—no net force, so no electric field. But for $x \gg a$, we get $E = kQ/x^2$—just what we expect for a point charge Q. As always, a finite-size charge distribution looks like a point charge at large distances. ■

EXAMPLE 20.7 Line Charge: A Power Line's Field

A long, straight electric power line coincides with the x-axis and carries a uniform line charge density λ (unit: C/m). Find the electric field on the y-axis using the approximation that the wire is infinitely long.

INTERPRET We identify the field point as being a distance y from the wire, and the source charge as the whole wire.

DEVELOP Figure 20.16 is our drawing, showing a coordinate system with the field point P along the y-axis. We divide the wire into small charge elements dq and note that field contributions from two such elements dq on opposite sides of the y-axis contribute fields $d\vec{E}$ whose x-components cancel. Then we need only the y-component of each unit vector, and Figure 20.16 shows that's $\hat{r}_y = y/r$.

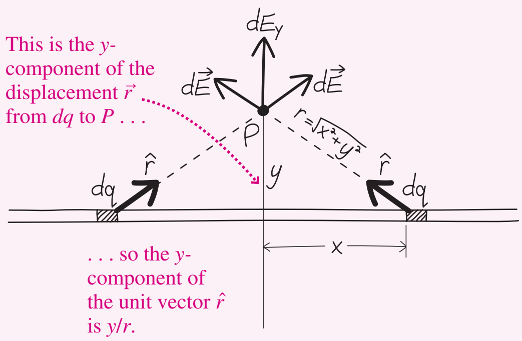


FIGURE 20.16 The field of a charged line is the vector sum of the fields $d\vec{E}$ from all the individual charge elements dq along the line.

EVALUATE We're now ready to set up the integral in Equation 20.7. As described in Chapter 9's integral strategy, we need to relate dq to a geometric variable so we can do the integral. Here our wire has charge density λ C/m, so if a charge element has length dx, then its charge is $dq = \lambda\, dx$. Putting all this together gives the y-component of the field from an arbitrary dq anywhere on the wire:

$$dE_y = \frac{k\,dq}{r^2}\hat{r}_y = \frac{k\lambda\,dx}{r^2}\frac{y}{r} = \frac{k\lambda y}{(x^2+y^2)^{3/2}}dx$$

where we used $r = \sqrt{x^2+y^2}$. Since the x-components cancel, we can sum—that is, integrate—the y-components to get the net field:

$$E = E_y = \int_{-\infty}^{+\infty}\frac{k\lambda y\,dx}{(x^2+y^2)^{3/2}} = k\lambda y\int_{-\infty}^{+\infty}\frac{dx}{(x^2+y^2)^{3/2}}$$
$$= k\lambda y\left[\frac{x}{y^2\sqrt{x^2+y^2}}\right]_{-\infty}^{+\infty} = k\lambda y\left[\frac{1}{y^2} - \left(-\frac{1}{y^2}\right)\right] = \frac{2k\lambda}{y}$$

Here we used the integral table in Appendix A and applied the limits $x = \pm\infty$. Our result is the field's magnitude; the direction is away from the line for positive λ and toward the line for negative λ.

ASSESS Make sense? For an infinite line there's nothing to favor one direction along the line over another, so the only way the field can point is radially, away from or toward the line (Fig. 20.17). And because the line is infinite, it never resembles a point no matter howfar away we are. As a result the field falls more slowly than the field of a point charge—in this case, as $1/y$. If we let r designate the radial distance from the line rather than the diagonal in Fig. 20.16, then the field decreases as $1/r$. An infinite line is impossible, but our result holds approximately for finite lines of charge as long as we're much closer to the line than its length, and not near an end. Far from a *finite* line, on the other hand, its field will resemble that of a point charge. ■

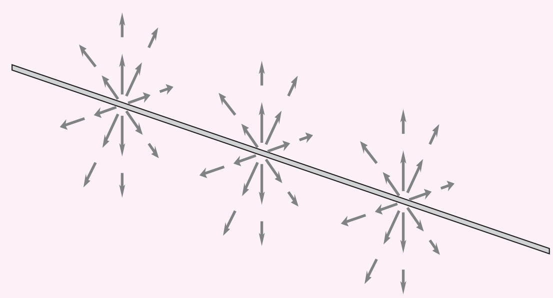


FIGURE 20.17 Field vectors for an infinite line of positive charge point radially outward, with magnitude decreasing inversely with distance.

20.5 Matter in Electric Fields

Electric fields give rise to forces on charged particles. Because matter consists of such particles, much of the behavior of matter is fundamentally determined by electric fields.

Point Charges in Electric Fields

The motion of a single charge in an electric field is governed by the definition of the electric field, $\vec{F} = q\vec{E}$, and Newton's law, $\vec{F} = m\vec{a}$. Combining these equations gives the acceleration of a particle with charge q and mass m in an electric field $\vec{E}$:

$$\vec{a} = \frac{q}{m}\vec{E} \tag{20.8}$$

This equation shows that it's the charge-to-mass ratio, q/m, that determines a particle's response to an electric field. Electrons, nearly 2000 times less massive than protons but carrying the same charge, are readily accelerated by electric fields. Many practical devices, from X-ray machines to fluorescent lights, use electrons accelerated in electric fields.

When the electric field is uniform, problems involving the motion of charged particles reduce to the constant-acceleration problems of Chapter 2. An ink-jet printer is one application; a pair of oppositely charged plates creates a uniform field that "steers" charged ink droplets to the right place on the page (Fig. 20.18).

When the field isn't uniform, it's generally more difficult to calculate particle trajectories. An important exception is a particle moving perpendicular to a field that points radially. Under appropriate conditions, the result is uniform circular motion (see Section 5.3), as shown in the next example.

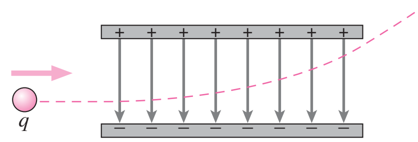


FIGURE 20.18 A pair of parallel charged plates creates a uniform electric field that deflects a charged particle. Can you tell the sign of the charge q?

EXAMPLE 20.8 Particle Motion: An Electrostatic Analyzer

Two oppositely charged curved metal plates establish an electric field given by $E = E_0(b/r)$, where E_0 and b are constants with the units of electric field and length, respectively. The field points toward the center of curvature, and r is the distance from the center. Find an expression for the speed v with which a proton entering vertically from below in Fig. 20.19 will leave the device moving horizontally.

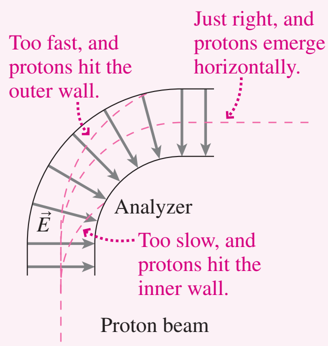


FIGURE 20.19 An electrostatic analyzer.

INTERPRET This problem is about charged-particle motion in an electric field that points radially. We're asked for the condition that will have a proton exiting the field region moving horizontally. Figure 20.19 shows that this requires its trajectory to be a circular arc.

DEVELOP Equation 20.8, $\vec{a} = (q/m)\vec{E}$, determines the acceleration of a charged particle in an electric field. Here we want uniform circular motion, so our plan is to write this equation with the given field and the acceleration v^2/r that we know applies in circular motion. Then we'll solve for v.

EVALUATE Under these conditions, Equation 20.8 becomes

$$a = \frac{v^2}{r} = \frac{eE}{m} = \frac{e}{m}E_0\frac{b}{r}$$

We then solve to get $v = \sqrt{eE_0b/m}$.

ASSESS Make sense? Strengthen the field by increasing E_0 or b, and the electric force becomes greater. For a given speed, that would result in more bending of the trajectory; to maintain the desired trajectory, we must therefore increase the speed. Note that the radius r canceled from our equations, showing that it doesn't matter where the protons enter the device. That's because the $1/r$ decrease in field strength matches the $1/r$ dependence of the acceleration. This device is called an electrostatic analyzer because it can sort charged particles by speed and charge-to-mass ratio. Spacecraft use such analyzers to characterize charged particles in interplanetary space. ■

GOT IT? 20.3 An electron, a proton, a deuteron (a neutron combined with a proton), a helium-3 nucleus (2 protons, 1 neutron), a helium-4 nucleus (2 protons, 2 neutrons), a carbon-13 nucleus (6 protons, 7 neutrons), and an oxygen-16 nucleus (8 protons, 8 neutrons) all find themselves in the same electric field. Rank in order their accelerations from lowest to highest under the assumption (only approximately correct) that the neutron and proton have the same mass and that the mass of a composite particle is the sum of the masses of its constituent neutrons and protons. Note any that have the same acceleration.

Dipoles in Electric Fields

Earlier in this chapter we calculated the field of an electric dipole, which consists of two opposite charges of equal magnitude. Here we study a dipole's response to electric fields. Since the dipole provides a model for many molecules, our results help explain molecular behavior.

Figure 20.20 shows a dipole with charges $\pm q$ separated a distance d, located in a uniform electric field. The dipole moment vector $\vec{p}$ has magnitude qd and points from the negative to the positive charge (recall Fig. 20.13). Since the field is uniform, it's the same at both ends of the dipole. Since the dipole charges are equal in magnitude but opposite in sign, they experience equal but opposite forces $\pm q\vec{E}$—and therefore there's no net force on the dipole.

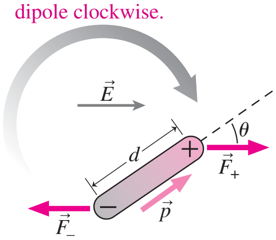


FIGURE 20.20 A dipole in a uniform electric field experiences a torque, but no net force.

However, Fig. 20.20 shows that the dipole does experience a torque that tends to align it with the field. In Chapter 11 we described torque as the cross product of the position vector with the force: $\vec{\tau} = \vec{r} \times \vec{F}$, where the magnitude of the torque vector is $rF\sin\theta$ and its direction is given by the right-hand rule. Figure 20.20 thus shows that the torque about the center of the dipole due to the force on the positive charge has magnitude $\tau_+ = rF\sin\theta = \left(\frac{1}{2}d\right)(qE)\sin\theta$. The torque associated with the negative charge has the same magnitude, and both torques are in the same direction since both tend to rotate the dipole clockwise. Thus the net torque has magnitude $\tau = qdE\sin\theta$; applying the right-hand rule shows that this torque is into the page. But qd is the magnitude of the dipole moment $\vec{p}$, and Fig. 20.20 shows that θ is the angle between the dipole moment vector and the electric field $\vec{E}$; therefore, we can write the torque vectorially as

$$\vec{\tau} = \vec{p} \times \vec{E} \qquad \text{(torque on a dipole)} \tag{20.9}$$

Because of this torque, it takes work to rotate a dipole in an electric field. If we start with the dipole oriented at right angles to the field $(\theta = \pi/2)$, then Equation 10.18 gives the work required to rotate it until it makes an angle θ with the field:

$$W = \int_{\pi/2}^{\theta} \tau\, d\theta = \int_{\pi/2}^{\theta} pE\sin\theta\, d\theta = pE\left[-\cos\theta\right]_{\pi/2}^{\theta} = -pE\cos\theta$$

This work ends up as stored potential energy U. Since the product of two vector magnitudes with the cosine of the angle between them defines the dot product, we can write the potential energy as

$$U = -\vec{p} \cdot \vec{E} \tag{20.10}$$

where $U = 0$ corresponds to the dipole at right angles to the field.

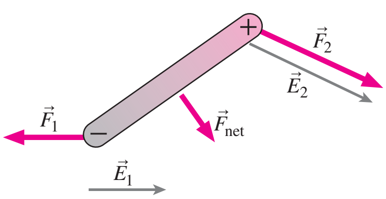


FIGURE 20.21 When the electric field differs in magnitude or direction at the two ends of the dipole, the dipole experiences a nonzero net force as well as a torque.

When the electric field isn't uniform, the charges at opposite ends of the dipole experience forces that differ in magnitude and/or aren't exactly opposite in direction. Then the dipole experiences a net force as well as a torque (Fig. 20.21). An important instance of this effect is the force on a dipole in the field of another dipole (Fig. 20.22). Because the dipole field falls off rapidly with distance and because the dipole responding to the field has closely spaced charges of equal magnitude but opposite sign, the dipole-dipole force is quite weak and falls extremely rapidly with distance. This weak force, which Fig. 20.22 shows to be attractive, is partly responsible for the van der Waals interaction between gas molecules that we mentioned in Chapter 17.

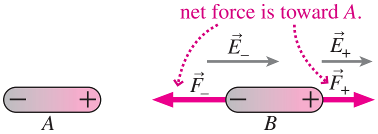


FIGURE 20.22 Dipole B aligns with the field of dipole A and then experiences a net force toward A.

Conductors, Insulators, and Dielectrics

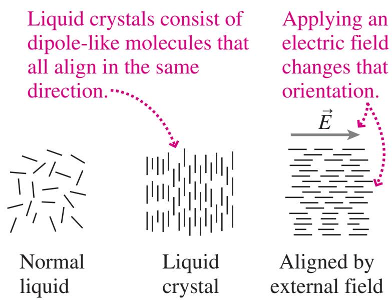
Bulk matter contains vast numbers of point charges—namely, electrons and protons. In some matter—notably metals, ionic solutions, and ionized gases—individual charges are free to move throughout the material. In these **conductors**, the application of an electric field results in the ordered motion of electric charge that we call **electric current**. We'll consider conductors and current in later chapters.

Materials in which charge is not free to move are **insulators**, since they don't support electric current. Insulators, however, still contain charges—it's just that their charges are bound into neutral molecules. Some molecules, like water, have intrinsic dipole moments

APPLICATION **Microwave Cooking and Liquid Crystals**

The torque on dipoles in electric fields forms the basis of two widespread contemporary technologies: the microwave oven and the liquid-crystal display (LCD).

A microwave oven works by generating an electric field whose direction changes several billion times per second. Water molecules, whose dipole moment is much greater than most others, attempt to align with the field. But the field is changing, so the molecules swing rapidly back and forth. As they jostle against each other, the energy they gain from the field is dissipated as heat that cooks the food.



Computer displays, TVs, digital cameras, cell phones, watches, and many other devices display visual images using liquid crystals. These unique materials combine the fluidity of a liquid with the order of a solid. The liquid crystal consists of long molecules whose chemical structure results in a dipole-like charge separation. In response to each others' electric fields, the molecules tend to align. As the figure shows, an external electric field can rotate the liquid-crystal dipoles, altering the material's optical properties. With optical components we'll study in Chapter 29, different sections of a liquid-crystal display can then be made to appear visible or invisible. Liquid-crystal displays consume very little power, but they generate no light of their own and therefore most have a built-in light source. The photo shows an iPhone—a popular device that sports a liquid-crystal display; also shown is a microphoto of the liquid crystals.

and therefore rotate in response to an applied electric field. Even if they don't have dipole moments, molecules may respond to an electric field by stretching and acquiring **induced dipole moments** (Fig. 20.23). In either case, the application of an electric field results in the alignment of molecular dipoles with the field (Fig. 20.24). The fields of the dipoles, pointing from their positive to their negative charges, then reduce the applied electric field within the material. We'll explore the consequences of this effect further in Chapter 23. Materials in which molecules either have intrinsic dipole moments or acquire induced moments are called **dielectrics**.

If the electric field applied to a dielectric becomes too great, individual charges are ripped free, and the material then acts like a conductor. Such **dielectric breakdown** can cause severe damage in electric equipment. On a larger scale, lightning results from dielectric breakdown in air.

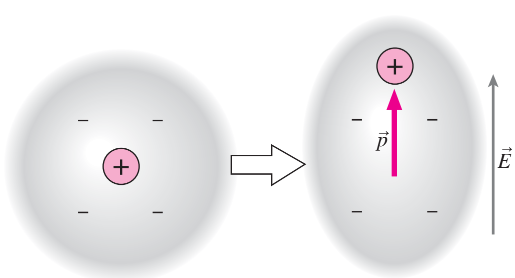


FIGURE 20.23 A molecule stretches in response to an electric field, acquiring a dipole moment.

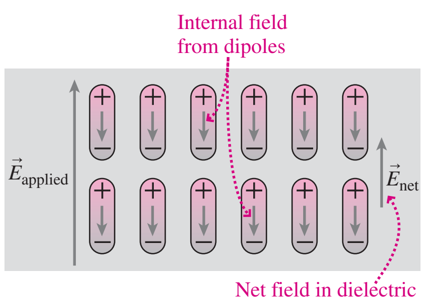


FIGURE 20.24 Alignment of molecular dipoles in a dielectric reduces the electric field within the dielectric.

CHAPTER 20 SUMMARY

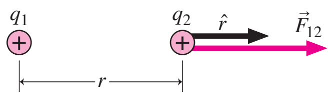
Big Picture

This chapter introduces several big ideas. First is **electric charge,** a fundamental property of matter that comes in positive and negative forms. Like charges repel and opposites attract; this is the **electric force.** It's convenient to define the **electric field** as the force per unit charge that a charge would experience if placed in the vicinity of other charges. Both force and field obey the **superposition principle**, meaning that the effects of several charges add vectorially.

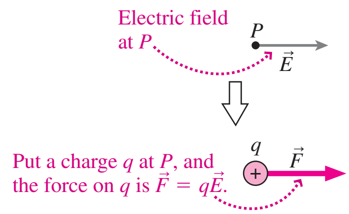
Key Concepts and Equations

Coulomb's law describes the electric force between point charges:

$$\vec{F}_{12} = \frac{kq_1q_2}{r^2}\hat{r}$$

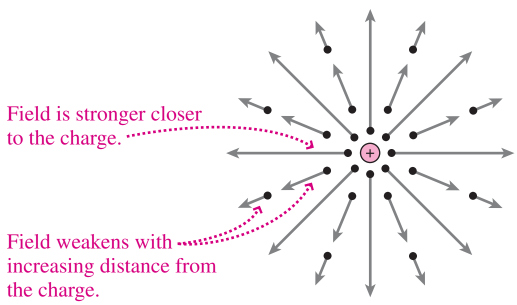


The electric field is the force per unit charge, $\vec{E} = \vec{F}/q$, and therefore the force a given charge q experiences in a field is $\vec{F} = q\vec{E}$.

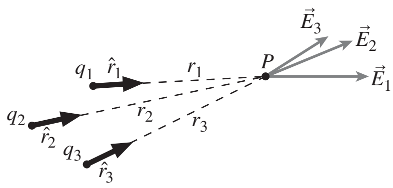


The field of a point charge follows from Coulomb's law:

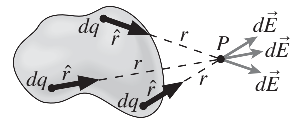
$$\vec{E} = \frac{kq}{r^2}\hat{r}$$



Fields of charge distributions are found by summing fields of individual point charges, or by integrating in the case of continuously distributed charge:



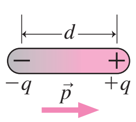
$$\vec{E}(P) = \vec{E}_1 + \vec{E}_2 + \vec{E}_3 = \sum_i \frac{kq}{r_i^2}\hat{r}_i$$



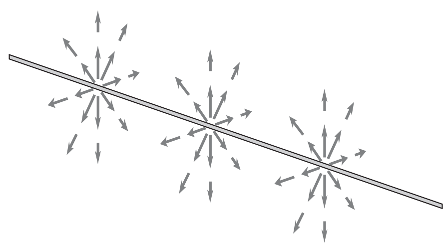
$$\vec{E}(P) = \int d\vec{E} = \int \frac{k\,dq}{r^2}\hat{r}$$

Applications

A **dipole** consists of equal but opposite charges $\pm q$ a distance d apart. For distances large compared with d, the dipole field drops as $1/r^3$, and the dipole is completely characterized by its **dipole moment** $p = qd$.

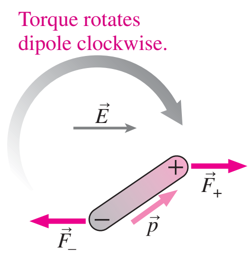


The field of an infinite line drops as $1/r$: $E = 2k\lambda/r$, with λ the charge per unit length. This is a good approximation to the field near an elongated structure like a wire.



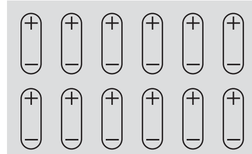
Point charges respond to electric fields with acceleration proportional to the charge-to-mass ratio q/m.

A dipole in an electric field experiences a torque that tends to align it with the field: $\vec{\tau} = \vec{p} \times \vec{E}$.



If the field is nonuniform, there's also a net force on the dipole.

Dielectrics are insulating materials whose molecules act like electric dipoles.



MP For homework assigned on MasteringPhysics, go to www.masteringphysics.com | BIO Biology and/or medical-related problems C Computer problems

For Thought and Discussion

1. Conceptual Example 20.1 shows that the gravitational force between an electron and a proton is about 10^{-40} times weaker than the electric force between them. Since matter consists largely of electrons and protons, why is the gravitational force important?
2. A free neutron is unstable and soon decays to other particles, one of them a proton. Must there be others? If so, what electric properties must they have?
3. Where in Fig. 20.5 could you put a third charge so it would experience no net force? Would it be in stable or unstable equilibrium?
4. Why should the test charge used to measure an electric field be small?
5. Equation 20.3 gives the electric field of a point charge. Does the direction of (a) $\hat{r}$ or (b) $\vec{E}$ depend on whether the charge is positive or negative?
6. Is the electric force on a charged particle always in the direction of the field? Explain.
7. Why does a dipole, which has no net charge, produce an electric field?
8. The ring in Example 20.6 carries total charge Q, and the point P is the same distance $r = \sqrt{x^2 + a^2}$ from all parts of the ring. So why isn't the electric field of the ring just kQ/r^2?
9. A spherical balloon is initially uncharged. If you spread positive charge uniformly over the balloon's surface, would it expand or contract? What would happen if you spread negative charge instead?
10. Under what circumstances is the path of a charged particle a parabola? A circle?
11. Why should there be a force between two dipoles, which each have zero net charge?
12. Dipoles A and B are both located in the field of a point charge Q, as shown in Fig. 20.25. Does either experience a net torque? A net force? If each dipole is released from rest, describe qualitatively its subsequent motion.

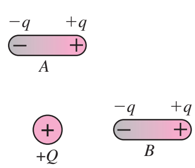


FIGURE 20.25 For Thought and Discussion 12

Exercises and Problems

Exercises

Section 20.1 Electric Charge

13. Suppose the electron and proton charges differed by one part in one billion. Estimate the net charge on your body, assuming it contains equal numbers of electrons and protons.
14. A typical lightning flash delivers about 25 C of negative charge from cloud to ground. How many electrons are involved?
15. Protons and neutrons are made from combinations of the two most common quarks, the u quark (charge $+\frac{2}{3}e$) and the d quark (charge $-\frac{1}{3}e$). How could three of these quarks combine to make (a) a proton and (b) a neutron?

Section 20.2 Coulomb's Law

16. The electron and proton in a hydrogen atom are 52.9 pm apart. Find the magnitude of the electric force between them.
17. An electron at Earth's surface experiences a gravitational force $m_e g$. How far away can a proton be and still produce the same force on the electron? (Your answer should show why gravity is unimportant on the molecular scale!)
18. You break a piece of Styrofoam packing material, and it releases lots of little spheres whose electric charge makes them stick annoyingly to you. If two of the spheres carry equal charges and repel with a force of 21 mN when they're 15 mm apart, what's the magnitude of the charge on each?
19. A charge q is at the point $x = 1$ m, $y = 0$ m. Write expressions for the unit vectors you would use in Coulomb's law if you were finding the force that q exerts on other charges located at (a) $x = 1$ m, $y = 1$ m; (b) the origin; and (c) $x = 2$ m, $y = 3$ m. You're not given the sign of q. Why doesn't this matter?
20. A proton is at the origin and an electron is at the point $x = 0.41$ nm, $y = 0.36$ nm. Find the electric force on the proton.

Section 20.3 The Electric Field

21. An electron experiences an electric force of 0.61 nN. What's the field strength at its location?
22. Find the magnitude of the electric force on a 2.0-μC charge in a 100-N/C electric field.
23. A 68-nC charge experiences a 150-mN force in a certain electric field. Find (a) the field strength and (b) the force that a 35-μC charge would experience in the same field.
24. BIO The electric field inside a cell membrane is 8.0 MN/C. What's the force on a singly charged ion in this field?
25. A -1.0-μC charge experiences a $10\hat{\imath}$-N electric force in a certain electric field. What force would a proton experience in the same field?
26. The electron in a hydrogen atom is 52.9 pm from the proton. What's the proton's electric field strength at this distance?

Section 20.4 Fields of Charge Distributions

27. In Fig. 20.26, point P is midway between the two charges. Find the electric field in the plane of the page (a) 5.0 cm to the left of P, (b) 5.0 cm directly above P, and (c) at P.

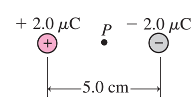


FIGURE 20.26 Exercise 27

28. The water molecule's dipole moment is 6.2×10^{-30} C·m. What would be the separation distance if the molecule consisted of charges $\pm e$? (The effective charge is actually less because H and O atoms share the electrons.)
29. The electric field 22 cm from a long wire carrying a uniform line charge density is 1.9 kN/C. What's the field strength 38 cm from the wire?
30. Find the line charge density on a long wire if the electric field 45 cm from the wire has magnitude 260 kN/C and points toward the wire.
31. Find the magnitude of the electric field due to a charged ring of radius a and total charge Q on the ring axis at distance a from the ring's center.

Section 20.5 Matter in Electric Fields

32. In his famous 1909 experiment that demonstrated quantization of electric charge, R. A. Millikan suspended small oil drops in an electric field. With field strength 20 MN/C, what mass drop can be suspended when the drop carries 10 elementary charges?
33. How strong an electric field is needed to accelerate electrons in an X-ray tube from rest to one-tenth the speed of light in a distance of 5.0 cm?
34. A proton moving to the right at 3.8×10^5 m/s enters a region where a 56-kN/C electric field points to the left. (a) How far will the proton get before it momentarily stops? (b) Describe its subsequent motion.
35. An electrostatic analyzer like that of Example 20.8 has $b =$ 7.5 cm. What value of E_0 will enable the device to select protons moving at 84 m/s?

Problems

36. A 2-g ping-pong ball rubbed against a wool jacket acquires a net positive charge of 1 μC. Estimate the fraction of the ball's electrons that have been removed.
37. Two charges, one twice as large as the other, are located 15 cm apart and experience a repulsive force of 95 N. What's the magnitude of the larger charge?
38. A proton is on the x-axis at $x = 1.6$ nm. An electron is on the y-axis at $y = 0.85$ nm. Find the net force the two exert on a helium nucleus (charge $+2e$) at the origin.
39. A 9.5-μC charge is at $x = 15$ cm, $y = 5.0$ cm and a -3.2-μC charge is at $x = 4.4$ cm, $y = 11$ cm. Find the force on the negative charge.
40. A charge $3q$ is at the origin, and a charge $-2q$ is on the positive x-axis at $x = a$. Where would you place a third charge so it would experience no net electric force?
41. You have two charges $+4q$ and one charge $-q$. How would you place them along a line so there's no net force on any of the three?
42. In Fig. 20.27, take $q_1 = 68\ \mu$C, $q_2 = -34\ \mu$C, and $q_3 = 15\ \mu$C. Find the electric force on q_3.

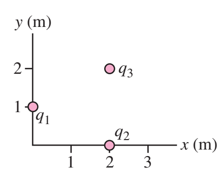


FIGURE 20.27 Problems 42 and 43

43. In Fig. 20.27, take $q_1 = 25\ \mu$C and $q_2 = 20\ \mu$C. If the force on q_1 points in the $-x$-direction, find (a) q_3 and (b) the magnitude of the force on q_1.
44. Three identical charges $+q$ and a fourth charge $-q$ form a square of side a. (a) Find the magnitude of the electric force on a charge Q placed at the square's center. (b) Describe the direction of this force.
45. A 65-μC point charge is at the origin. Find the electric field at the points (a) $x = 50$ cm, $y = 0$ cm; (b) $x = 50$ cm, $y = 50$ cm; and (c) $x = 25$ cm, $y = -75$ cm.
46. A 1.0-μC charge and a 2.0-μC charge are 10 cm apart. Find a point where the electric field is zero.
47. A proton is at the origin and an ion is at $x = 5.0$ nm. If the electric field is zero at $x = -5$ nm, what's the ion's charge?
48. (a) Find an expression for the electric field on the y-axis due to the two charges q in Fig. 20.7. (b) At what point is the field on the y-axis a maximum?
49. A dipole lies on the y-axis and consists of an electron at $y = 0.60$ nm and a proton at $y = -0.60$ nm. Find the electric field (a) midway between the two charges; (b) at the point $x = 2.0$ nm, $y = 0$ nm; and (c) at the point $x = -20$ nm, $y = 0$ nm.
50. Show that the field on the x-axis for the dipole of Example 20.5 is given by Equation 20.6b, for $x \gg a$.
51. You're 1.5 m from a charge distribution whose size is much less than 1 m. You measure an electric field strength of 282 N/C. You move to a distance of 2.0 m, and the field strength becomes 119 N/C. What's the net charge of the distribution? (*Hint:* Don't try to calculate the charge. Determine instead how the field decreases with distance, and from that infer the charge.)
52. Three identical charges q form an equilateral triangle of side a, with two charges on the x-axis and one on the positive y-axis. (a) Find an expression for the electric field at points on the y-axis above the uppermost charge. (b) Show that your result reduces to the field of a point charge $3q$ for $y \gg a$.
53. Two identical small metal spheres initially carry charges q_1 and q_2. When they're 1.0 m apart, they experience a 2.5-N attractive force. Then they're brought together so charge moves from one to the other until they have the same net charge. They're again placed 1.0 m apart, and now they repel with a 2.5-N force. What were the original charges q_1 and q_2?
54. Two 34-μC charges are attached to opposite ends of a spring with spring constant $k = 150$ N/m and equilibrium length 50 cm. By how much does the spring stretch?
55. A thin rod lies on the x-axis between $x = 0$ and $x = L$ and carries total charge Q distributed uniformly over its length. Show that the electric field strength for $x > L$ is given by $E = kQ/[x(x - L)]$.
56. An electron is moving in a circular path around a long, uniformly charged wire carrying 2.5 nC/m. What's the electron's speed?
57. You have a job examining patent applications. You're presented with the device in Fig. 20.28, which its inventor claims will separate isotopes of a particular element. Atoms are first stripped completely of their electrons, then accelerated from rest through an electric field chosen to give the desired isotope exactly the right speed to pass through the electrostatic analyzer (see Example 20.8). Will the device work?

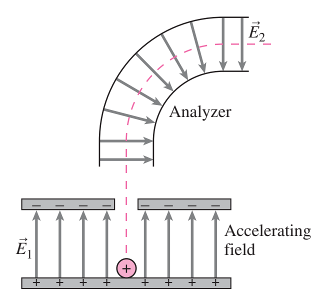


FIGURE 20.28 Problem 57

58. A 5.0-μm strand of DNA carries charge $+e$ per nm of length.
BIO Treating it as a charged line, what's the electric field strength 25 nm from the DNA, not near either end?
59. Find the line charge density on a long wire if a 6.8-μg particle carrying 2.1 nC describes a circular orbit about the wire with speed 280 m/s.

60. A dipole with dipole moment 1.5 nC·m is oriented at 30° to a 4.0-MN/C electric field. Find (a) the magnitude of the torque on the dipole and (b) the work required to rotate the dipole until it's antiparallel to the field.
61. A molecule has its dipole moment aligned with a 1.2-kN/C electric field. If it takes 3.1×10^{-27} J to reverse the molecule's orientation, what's its dipole moment?
62. Two identical dipoles, each of charge q and separation a, are a distance x apart, as shown in Fig. 20.29. (a) By considering forces between pairs of charges in the different dipoles, calculate the force between the dipoles and show that, in the limit $a \ll x$, it has magnitude $6kp^2/x^4$, where $p = qa$ is the dipole moment. (b) Is the force attractive or repulsive?

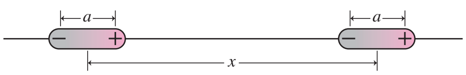


FIGURE 20.29 Problem 62

63. A dipole with charges $\pm q$ and separation $2a$ is located a distance x from a point charge $+Q$, oriented as shown in Fig. 20.30. Find expressions for the magnitude of (a) the net torque and (b) the net force on the dipole, both in the limit $x \gg a$. (c) What's the direction of the net force?

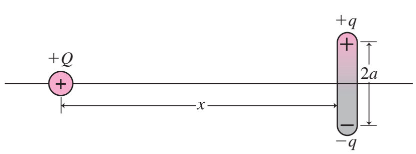


FIGURE 20.30 Problem 63

64. An electron is at the origin, and an ion with charge $+5e$ is at $x = 10$ nm. Find a point where the electric field is zero.
65. You're taking physical chemistry, and your professor is discussing molecular dipole moments. Water, he says, has a dipole moment of "1.85 debyes," while carbon monoxide's dipole moment is only "0.12 debye." Your physics professor wants these moments expressed in SI. She tells you that the atomic separation in these two covalent compounds is about the same, and asks what that indicates about the way shared charge is distributed. What do you tell her?
66. The electric field on the axis of a uniformly charged ring has magnitude 380 kN/C at a point 5.0 cm from the ring center. The magnitude 15 cm from the center is 160 kN/C; in both cases the field points away from the ring. Find (a) the ring's radius and (b) its charge.
67. An *electric quadrupole* consists of two oppositely directed dipoles in close proximity. (a) Calculate the field of the quadrupole shown in Fig. 20.31 for points to the right of $x = a$ and (b) show that for $x \gg a$ the quadrupole field falls off as $1/x^4$.

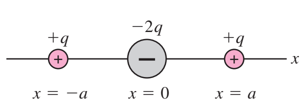


FIGURE 20.31 Problem 67

68. Show that the electric field at a point 45° from a dipole's axis is $1/\sqrt{5/8}$ times the field on the axis, assuming both points are the same distance from the dipole and that distance is large compared with the dipole spacing.
69. A straight wire 10 m long carries 25 μC distributed uniformly over its length. (a) What's the line charge density on the wire? Find the electric field strength (b) 15 cm from the wire axis, not near either end, and (c) 350 m from the wire. Make suitable approximations in both cases.
70. Figure 20.32 shows a thin rod of length L carrying charge Q distributed uniformly over its length. (a) What's the line charge density on the rod? (b) What must be the electric field direction on the rod's perpendicular bisector (taken to be the y-axis)? (c) Modify the calculation of Example 20.7 to find an expression for the electric field at a point P a distance y along the perpendicular bisector.

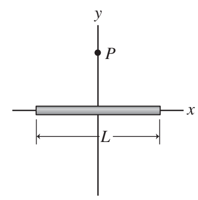


FIGURE 20.32 Problem 70

71. Figure 20.33 shows a thin, uniformly charged disk of radius R. Imagine the disk divided into rings of varying radii r, as suggested in the figure. (a) Show that the area of such a ring is very nearly $2\pi r\,dr$. (b) If the disk carries surface charge density σ, use the result of part (a) to write an expression for the charge dq on an infinitesimal ring. (c) Use the result of (b) along with the result of Example 20.6 to write the infinitesimal electric field dE of this ring at a point on the disk axis, taken to be the positive x-axis. (d) Integrate over all such rings to show that the net electric field on the disk axis has magnitude

$$E = 2\pi k\sigma\left(1 - \frac{1}{\sqrt{x^2 + R^2}}\right)$$

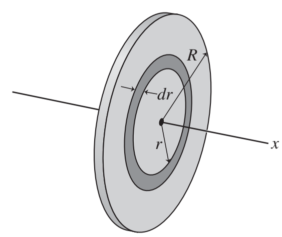


FIGURE 20.33 Problem 71

72. Use the result of Problem 71 to show that the field of an *infinite*, uniformly charged flat sheet is $2\pi k\sigma$, where σ is the surface charge density. (This result is independent of distance from the sheet.)
73. Use the binomial theorem to show that, for $x \gg R$, the result of Problem 71 reduces to the field of a point charge whose total charge is the charge density times the disk area.
74. A semicircular loop of radius a carries positive charge Q distributed uniformly. Find the electric field at the loop's center (point P in Fig. 20.34). (*Hint:* Divide the loop into charge elements dq as shown, write dq in terms of the angle $d\theta$, then integrate over θ.)

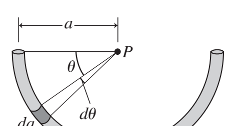


FIGURE 20.34 Problem 74

75. In Example 20.2, find the position where Q will experience the greatest force.
76. A thin rod carries charge Q distributed uniformly over its length L and is situated on the x-axis between $x = \pm L/2$. (a) Find the electric field at an arbitrary point (x, y). (You'll have to do separate integrals for the x- and y-components.) (b) Show that your result reduces to that of Problem 70 when $x = 0$ and to that of Problem 55 when $y = 0$ and $x > L/2$.
77. A thin rod extends along the x-axis from $x = 0$ to $x = L$ and carries line charge density $\lambda = \lambda_0(x/L)^2$, where λ_0 is a constant. Find the electric field at $x = -L$.
78. A rod of length $2L$ lies on the x-axis, centered at the origin, and carries line charge density $\lambda = \lambda_0(x/L)$, where λ_0 is a constant. (a) Find an expression for the electric field strength at points on the x-axis for $x > L$. (b) Show that for $x \gg L$ your result has the $1/x^3$ dependence of a dipole field, and determine the dipole moment of the rod.
79. You're working on the design of an ink-jet printer. Ink drops of mass m, speed v, and charge q will enter a region of uniform electric field E between two charged plates (Fig. 20.35). The drops enter midway between the plates, and the electric field deflects them toward the correct place on the page. Find an expression for the maximum electric field for which drops can still get through without hitting either plate.

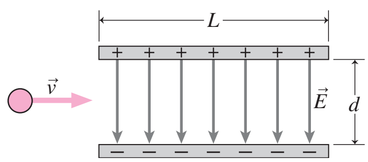


FIGURE 20.35 Problem 79

Passage Problems

BIO The human heart consists largely of elongated muscle cells, some 100 μm long and 15 μm in diameter. In its resting state, a cell contains two concentric layers of charge, which confine the electric field to the cell membrane (Fig. 20.36*a*). When the heart contracts, a wave of depolarization sweeps through, depleting charge and giving each cell a dipole moment (Fig. 20.36*b*). As a result, the entire organ acts like an electric dipole, producing an external field, which is indirectly detected by electrocardiography. Although the direction of the heart's dipole moment varies, Fig. 20.36*c* is typical. In answering the questions that follow, consider the heart in isolation—don't concern yourself with the effect of surrounding tissues on its electric field.

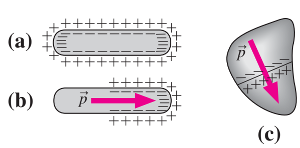


FIGURE 20.36 Heart cells (a) in the resting state and (b) partially depolarized, resulting in a dipole moment $\vec{p}$. (c) Typical orientation of the heart's dipole moment vector. Cells along the line are depolarizing.

80. At a distance r, far from the heart, the heart's electric field
    a. falls off as $1/r$.
    b. falls off as $1/r^2$.
    c. falls off as $1/r^3$.
    d. falls off as $1/r^4$.
81. At a given distance, far from the heart compared with its size, the electric field
    a. is weaker along an extension of the line shown in Fig. 20.36*c* than on a perpendicular line.
    b. is stronger along a an extension of the line shown in Fig. 20.36*c* than on a perpendicular line.
    c. has the same value at positions perpendicular and parallel to the line in Fig. 20.36*c*.
82. The difference between Figs. 20.36*a* and 20.36*b* that results in an external electric field in one case but not the other is that
    a. there's no net charge in Fig. 20.36*a* but there is a net charge in Fig. 20.36*b*.
    b. the total charge is greater in Fig. 20.36*a*.
    c. the charge is distributed in Fig. 20.36*b* so there's more negative charge to the left and more positive charge to the right.
83. At the instant shown in Fig. 20.36*c*, there's an electric field within the heart that points approximately
    a. in the direction of the dipole moment vector $\vec{p}$.
    b. opposite the dipole moment vector $\vec{p}$.
    c. perpendicular to the dipole moment vector $\vec{p}$.

Answers to Chapter Questions

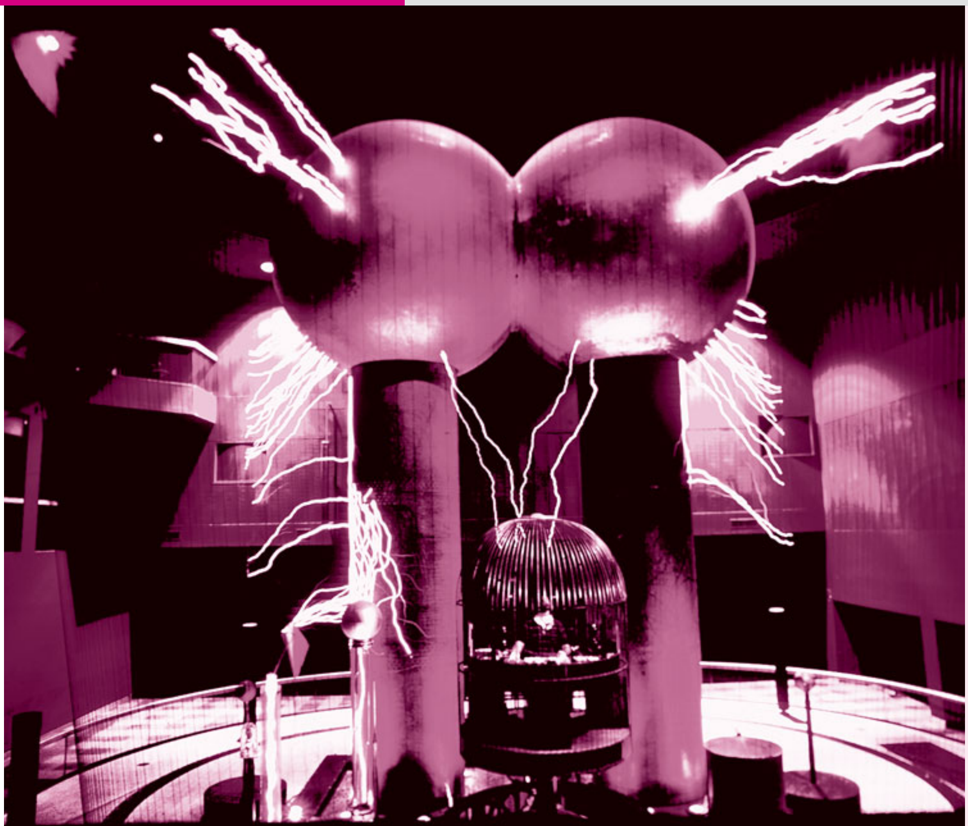
Answer to Chapter Opening Question

The electric field in the atmosphere must be strong enough to rip electrons from air molecules, making the air an electrical conductor. This typically happens when E exceeds about 3 MV/m.

Answers to GOT IT? Questions

20.1. (a) $-\hat{\imath}$; (b) $-\frac{\sqrt{2}}{2}\hat{\imath} + \frac{\sqrt{2}}{2}\hat{\jmath}$; the unit vector always points away from the source charge regardless of the sign.
20.2. (a) drops by $1/2^2$, to 200 N/C; (b) drops by $1/2^3$, to 100 N/C.
20.3. carbon-13, (oxygen-16, helium-4, deuteron—all the same), helium-3, proton, electron

21 Gauss's Law



Huge sparks jump to the operator's cage in the Hall of Electricity at the Boston Museum of Science, but the operator is unharmed. Why?

New Concepts, New Skills

By the end of this chapter you should be able to

- Represent electric fields using field lines and interpret field-line diagrams (21.1).
- Explain Gauss's law for electric fields and relate it to Coulomb's law (21.3).
- Calculate electric fields of symmetric charge distributions using Gauss's law (21.4).
- Give the electric fields of basic charge distributions, and estimate the fields of arbitrary distributions under appropriate approximations (21.5).
- Explain the concept of electrostatic equilibrium, and why Gauss's law requires excess electric charge to reside only on the surface of a conductor in equilibrium (21.6).

Connecting Your Knowledge

- This chapter expands on the concepts of electric force and, especially, electric field (20.2, 20.3).
- Integration becomes more central—although easier to apply—than in Chapter 20. To review integration, see (9.1) and Appendix A.2.

In this chapter we introduce an elegant way of describing electric fields that makes it much easier to calculate the fields of certain charge distributions. In the process we'll formulate one of the four fundamental laws of electromagnetism—a statement that embodies Coulomb's law but that gives deeper insights into the electric field.

21.1 Electric Field Lines

We can visualize electric fields by drawing **electric field lines**, continuous lines whose direction is everywhere the same as that of the electric field. To draw a field line, determine the field direction at some point. Move a small distance in the direction of the field, and evaluate the field direction at the new point. Extending the process in both directions from your starting point traces out an electric field line. You'll find that field lines begin on positive charges and either end on negative charges or extend to infinity. Drawing many field lines gives a picture of the overall field.

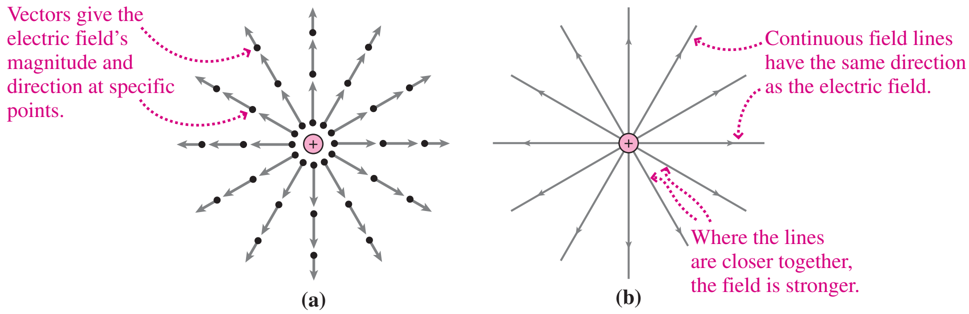


FIGURE 21.1 Vectors (a) and field lines (b) provide two ways to visualize the electric field.

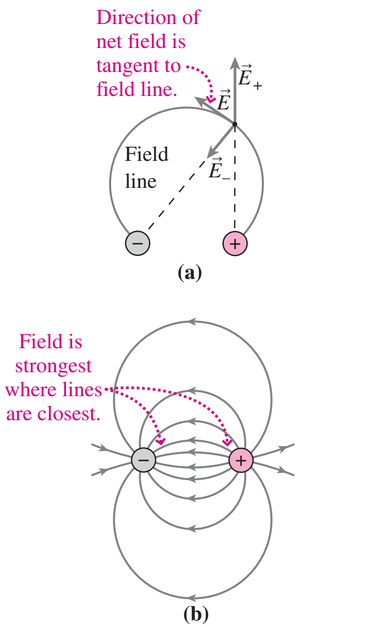


FIGURE 21.2 Field of an electric dipole. (a) At each point, the field-line direction is that of the *net* electric field, $\vec{E} = \vec{E}_+ + \vec{E}_-$. (b) Tracing several field lines shows the overall dipole field.

To explore the field of a positive point charge, as shown in Fig. 21.1*a*, start near the charge. The field points radially outward, so move a little way outward. The field is still radial. Repeat the process, and you'll trace a straight line that extends indefinitely. So the field lines of a positive point charge are straight lines that begin on the charge and extend radially to infinity (Fig. 21.1*b*).

In Fig. 21.1*b* the field lines spread apart as they extend farther from the point charge. Coulomb's law shows that the field weakens farther from the charge, so in Fig. 21.1*b* the field is stronger where the lines are closer and weaker where they're farther apart. This is generally true, and lets us infer the field's relative magnitude as well as direction from field-line pictures.

To trace the field lines of a charge distribution, follow the net field—the vector sum of the field contributions from all charges in the distribution. Usually the field direction varies, so the line is curved. Figure 21.2*a* shows the details for one field line of a dipole. Figure 21.2*b* shows a number of dipole field lines; here you can see that the field is strongest near the individual charges and in the region between them. The electric field exists everywhere, so there are really infinitely many field lines. We obviously can't draw them all. To make field-line pictures somewhat more precise, we associate a fixed number of field lines with a charge of a given magnitude. In Fig. 21.3, for example, eight field lines correspond to a charge of magnitude q. Study the figure, and you'll see how all the fields are consistent with this convention.

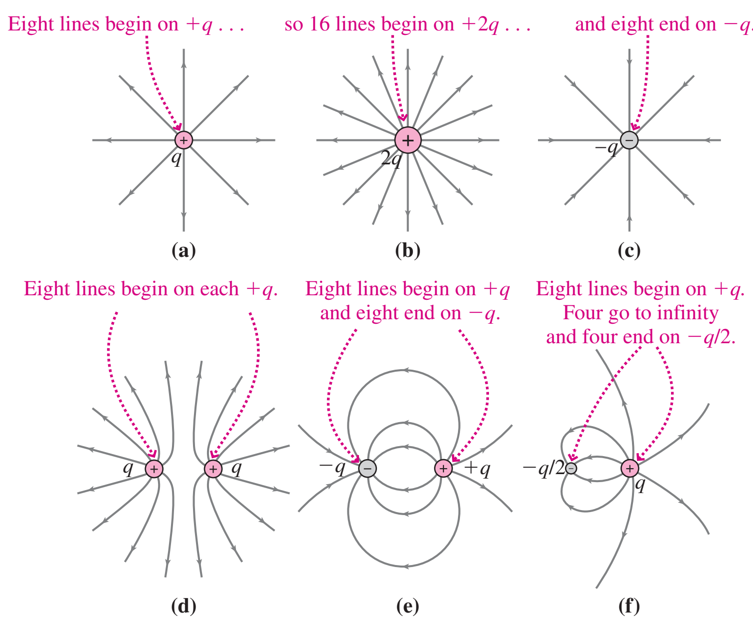


FIGURE 21.3 Field lines for six charge distributions, using the convention that eight lines correspond to a charge of magnitude q.

21.2 Electric Flux and Field

In Fig. 21.4 we've surrounded each charge distribution from Fig. 21.3 with several surfaces. Each surface is closed, so it's impossible to get from inside to outside without crossing the surface. (Figure 21.4 shows only the two-dimensional cross section of each surface.) How many field lines emerge from within each surface?

In Fig. 21.4*a* the answer for surfaces 1 and 2 is obvious: eight. For surface 3 one field line crosses twice going out and once going in; if we count the inward-going line as negative, then there's still a net of eight lines going out. *Any* closed surface surrounding $+q$ will have eight lines emerging from it, for the simple reason that eight lines begin on the charge and extend to infinity; to get there they all have to cross the closed surface.

What about surface 4? Two lines cross going outward and two inward, for a net of zero lines emerging. What's different is that surface 4 doesn't enclose the charge. You can convince yourself that *any* surface not enclosing the charge will have as many lines going in as out, for zero net field lines emerging.

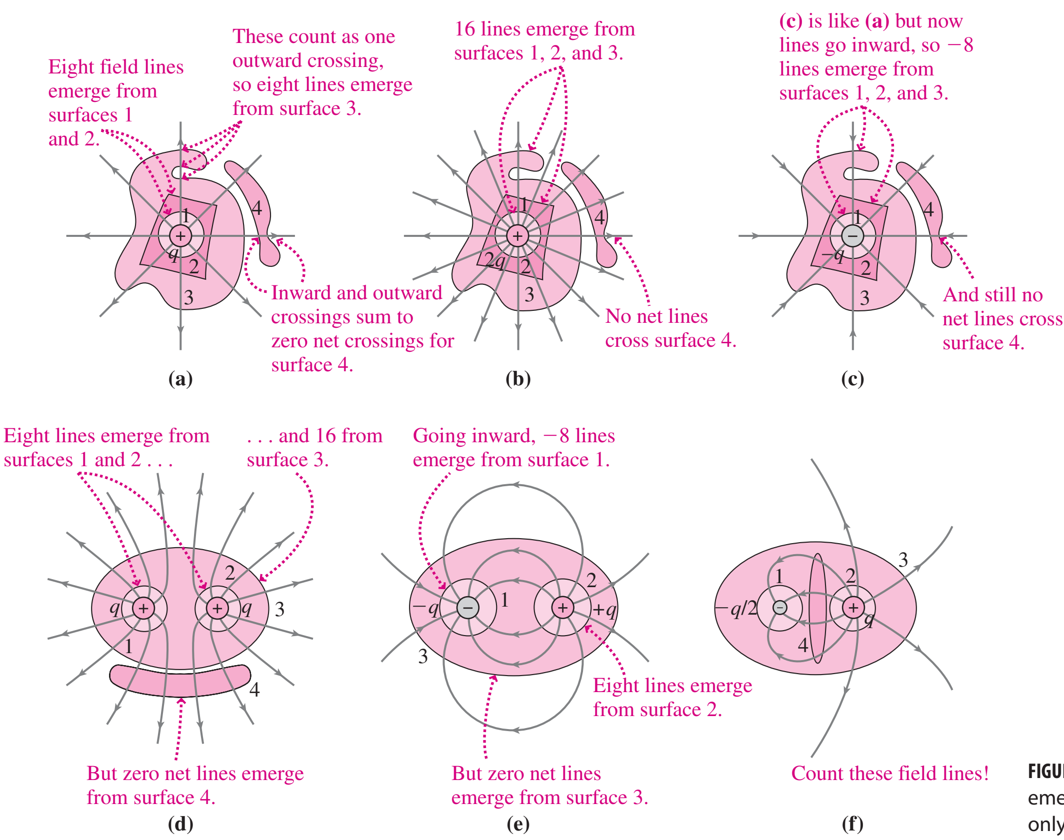


FIGURE 21.4 The number of field lines emerging from a closed surface depends only on the net charge enclosed.

Figure 21.4*b* is identical except that now 16 field lines emerge from any surface surrounding the charge: $+2q$ enclosed, so 16 field lines emerge. And Fig. 21.4*c* is similar, too, but with the negative charge, the field lines go inward and we count them as negative: $-q$ enclosed, so -8 field lines emerge. And, in all three cases, zero net lines emerge from surface 4, the one that doesn't enclose the charge: zero charge enclosed, so zero field lines emerging. The same is true even for surface 3 surrounding the dipole in Fig. 21.4*e*; here there are two charges within the surface, but the *net charge enclosed* is zero, and sure enough, there are zero net field lines emerging. Study the rest of Fig. 21.4 and you can convince yourself that in all cases **the number of field lines emerging from any closed surface is proportional to the net charge enclosed**.

This statement is very general. It doesn't matter what shape the surface is or whether the enclosed charge is a single point charge or a lot of charges carrying the same net charge. Nor does it matter how the charges are arranged, as long as they're *enclosed* by the surface in question. The presence of charges *outside* the surface doesn't alter our conclusion about the number of field lines emerging—although it may alter the shape of individual lines. We'll now rephrase our statement mathematically to obtain one of the four fundamental laws of electromagnetism. As we proceed, remember that the mathematics just reflects what's clear from Fig. 21.4: The number of field lines emerging from a closed surface depends only on the net charge enclosed.

Electric Flux

Figure 21.5 shows that the number of field lines crossing a flat surface depends on three factors: the field strength E, the surface area A, and the orientation of the surface relative to the field. We get the most field-line crossings when the surface is perpendicular to the field and none when it's parallel; Fig. 21.5*d* shows an intermediate case. If we define a vector $\vec{A}$ normal to the surface, then the number of field-line crossings is proportional to $\cos\theta$, where θ is the angle between the normal vector $\vec{A}$ and the field $\vec{E}$. Putting all this together, we find that the number of field lines crossing the surface is proportional to $EA\cos\theta$. This quantity has a definite value that captures the spirit of the more vague "number of field lines crossing a surface." We call $EA\cos\theta$ the **electric flux** Φ through the surface. If we make the magnitude of the surface normal vector $\vec{A}$ equal to the surface area A, then we can define flux using the vector dot product:

$$\Phi = \vec{E} \cdot \vec{A} \tag{21.1}$$

where the dot product, defined in Chapter 6, is the product of the two vector magnitudes with the cosine of the angle between them. Since the units of $\vec{E}$ are N/C, flux is measured in $\text{N}\cdot\text{m}^2/\text{C}$.

The surfaces in Fig. 21.5 are *open* surfaces, meaning it's possible to get from one side to the other without passing through the surface. For open surfaces there's an ambiguity in the sign of Φ, since we could have taken $\vec{A}$ in either of the two directions along the perpendicular to the surface. But for *closed* surfaces, we unambiguously define the direction of $\vec{A}$ as the direction of the outward-pointing normal to the surface.

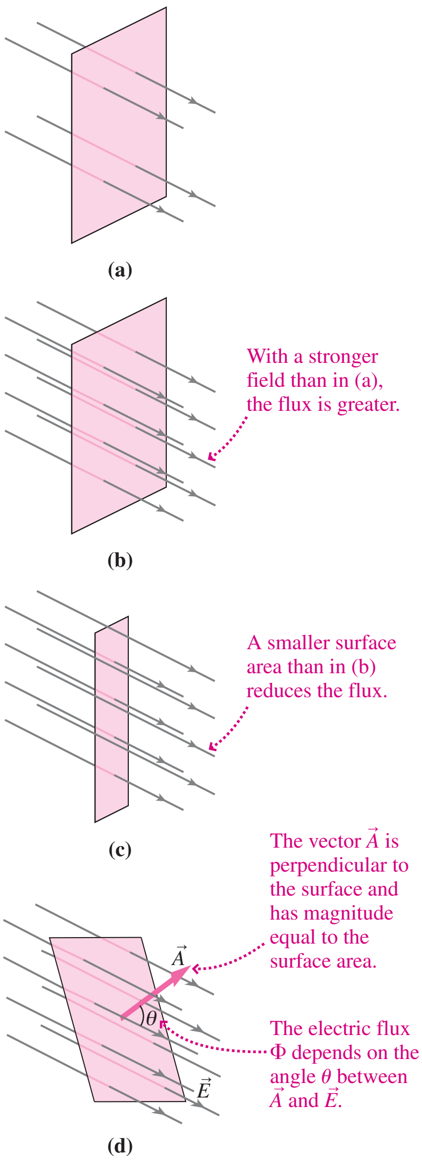


FIGURE 21.5 Electric flux through flat surfaces.

✓TIP The Flux Is Not the Field

The flux Φ and field $\vec{E}$ are related but distinct quantities. The field is a vector defined at each point in space. The flux is a scalar and a global property, depending on how the field behaves over an extended surface rather than at a single point; it's a quantification of the number of field lines crossing a surface.

What if a surface is curved and/or the field varies with position? Then we divide the surface into patches, each small enough that it's essentially flat and that the field is essentially uniform over each (Fig. 21.6). If a patch has area dA, then Equation 21.1 gives the flux through it: $d\Phi = \vec{E} \cdot d\vec{A}$, where the vector $d\vec{A}$ is normal to the patch. The total flux through the surface is then the sum over all the patches. If we make the patches arbitrarily small, that sum becomes an integral, and the flux is

$$\Phi = \int_{\text{surface}} \vec{E} \cdot d\vec{A} \tag{21.2}$$

The limits of the integral range over the entire surface, picking up contributions from all the patches $d\vec{A}$. Although the integral can be difficult to evaluate, we'll find it most useful in cases where it's almost trivial.

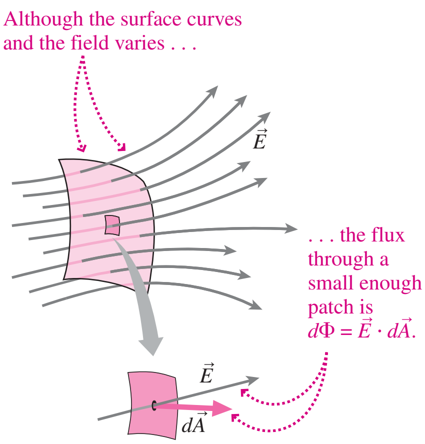
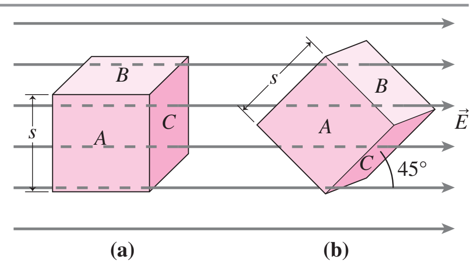


FIGURE 21.6 Finding the flux through a small area *dA*, so small it's essentially flat.

GOT IT? 21.1 The figure shows a cube of side s in a uniform electric field $\vec{E}$. What's the flux through each of the three cube faces A, B, and C with the cube oriented as in (a)? Repeat for the orientation in (b), with the cube rotated 45°.



21.3 Gauss's Law

We've seen that the number of field lines emerging from a closed surface is proportional to the charge enclosed. Now that we've developed electric flux to express more rigorously the notion "number of field lines," we can state that **the electric flux through any closed surface is proportional to the net charge enclosed by that surface.** Writing the same thing mathematically gives $\Phi = \oint \vec{E} \cdot d\vec{A} \propto q_{\text{enclosed}}$, where the circle indicates that the integral is over a *closed* surface.

To evaluate the proportionality between flux and charge, consider a positive point charge q and a spherical surface of radius r centered on the charge (Fig. 21.7). The flux through this surface is given by Equation 21.2:

$$\Phi = \oint \vec{E} \cdot d\vec{A} = \oint E\, dA \cos\theta$$

But Fig. 21.7 shows that the surface normal $d\vec{A}$ and the electric field $\vec{E}$ are parallel at any point on the sphere, so $\cos\theta = 1$. Since the electric field varies as $1/r^2$, its magnitude is the same everywhere at the fixed radius r of our sphere. Thus, we can take E outside the integral, giving

$$\Phi = \oint_{\text{sphere}} E\, dA = E \oint_{\text{sphere}} dA = E(4\pi r^2)$$

where the last step follows because $\oint dA$ is just the surface area of the sphere. Now, the electric field of a point charge is given by Equation 20.3: $E = kq/r^2$. So we have $\Phi = E(4\pi r^2) = (kq/r^2)(4\pi r^2) = 4\pi kq$. Since the point charge q is the only charge inside our spherical surface, the proportionality constant between flux and enclosed charge is $4\pi k$.

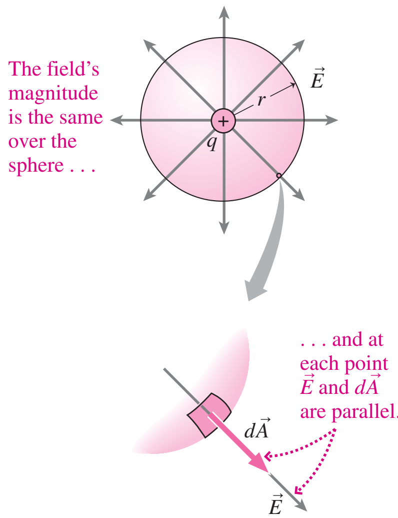


FIGURE 21.7 The electric field of a point charge, shown with a spherical surface centered on the charge.

Before proceeding, we introduce the so-called permittivity constant ϵ_0, defined as $\epsilon_0 = 1/4\pi k$, where k is the Coulomb constant. The value of ϵ_0 is $8.85\times10^{-12}\ \text{C}^2/\text{N}\cdot\text{m}^2$. There's no physics here, just a new constant that conveys the same information as k. That there are two redundant constants is a historical artifact, and we switch now from k to ϵ_0 because doing so makes subsequent formulas simpler. In terms of ϵ_0, the proportionality $4\pi k$ between flux and enclosed charge becomes $1/\epsilon_0$. So our statement that the flux through any closed surface is proportional to the net charge enclosed becomes

$$\oint \vec{E} \cdot d\vec{A} = \frac{q_{\text{enclosed}}}{\epsilon_0} \qquad \text{(Gauss's law)} \tag{21.3}$$

Here the integral is taken over *any closed surface*, and q_{enclosed} is the charge enclosed *by that surface*.

Equation 21.3 is **Gauss's law**, one of four fundamental relations that govern the behavior of electromagnetic fields throughout the universe. Whether you journey into a star in some remote galaxy, down among the strands of a DNA molecule, or into the microprocessor chip at the heart of your computer, you'll find that the flux of the electric field through any closed surface depends only on the enclosed charge. In nearly 200 years of experiments, no electric field has ever been found to violate Gauss's law.

Gauss's law, though clothed in the mathematical finery of a surface integral, is just a more rigorous way of saying what's clear in Fig. 21.4: The number of field lines emerging from a closed surface is proportional to the net charge enclosed.

Gauss and Coulomb

Gauss's law and Coulomb's law look completely different, but they're closely related. Figure 21.8 shows that their relationship involves the inverse-square law. Gauss's law tells us that the flux through the two surfaces in the figure is the same and is equal to q/ϵ_0. But why? Because, as our arguments leading to Gauss's law show, the flux through a spherical surface of radius r centered on a point charge is the product of the surface area $4\pi r^2$ and the electric field E at the surface. But Coulomb's law says that the electric field drops off as $1/r^2$. As r increases, the surface area grows as r^2, but the $1/r^2$ decrease in field strength just compensates, giving a constant value for the flux. If the inverse-square law (e.g., Coulomb's law) didn't hold, then the flux wouldn't be constant and Gauss's law wouldn't hold either.

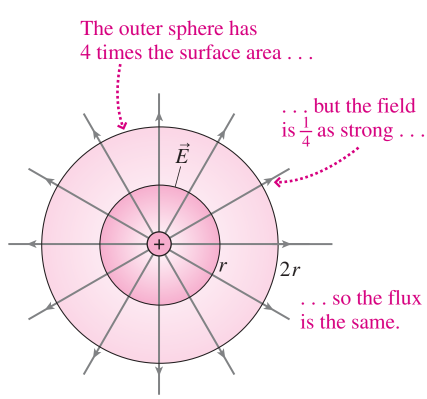


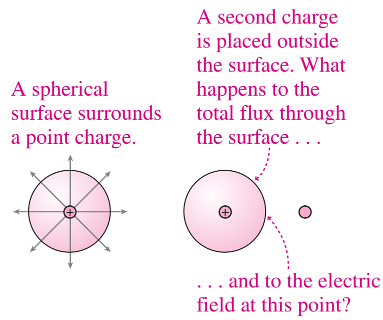
FIGURE 21.8 Gauss's law follows from the inverse-square aspect of Coulomb's law.

It's also the inverse-square law that makes electric field-line pictures useful for visualizing fields. Field lines begin or end only on charges; otherwise, they go off to infinity. As the field lines of a point charge spread in three dimensions, the number crossing any spherical surface (or any closed surface) remains the same. But larger spheres have larger surface areas, in proportion to r^2—and that means the density of field lines drops as $1/r^2$, accurately reflecting the field strength. Once again, the inverse-square law (Coulomb) and the relation between flux and enclosed charge (Gauss) are intimately connected. Incidentally, field-line pictures printed in a book or drawn on a blackboard generally can't be quantitatively correct because they don't show the spreading of field lines in all three dimensions.

We've been talking here only about isolated point charges, but we emphasize that Gauss's law applies to *all* electric fields, no matter how complicated the charge distributions that produce them. That's because the superposition principle allows us to add vectorially the electric fields described individually by Coulomb's law for the point-charge field. So our argument leading to Gauss's law still applies when the field $\vec{E}$ is a superposition of point-charge fields.

For static charge distributions, Gauss's and Coulomb's laws are completely equivalent. But with moving charges only Gauss's law remains exact. So Gauss's law is more fundamental, and we count it among the four basic laws of electromagnetism.

GOT IT? 21.2 A spherical surface surrounds an isolated positive charge, as shown. If a second charge is placed outside the surface, which of the following will be true of the total flux through the surface? (a) It doesn't change; (b) it increases; (c) it decreases; (d) it increases or decreases depending on the sign of the second charge. Repeat for the electric field on the surface at the point between the charges.



21.4 Using Gauss's Law

Gauss's law is a universal statement about electric fields; it's true for *any* surface enclosing *any* charge distribution. For charge distributions with sufficient symmetry—symmetric about a point, a line, or a plane—Gauss's law also provides a powerful alternative to Coulomb's law that makes electric-field calculations much easier. For such distributions it's possible to evaluate the flux integral on the left-hand side of Gauss's law (Equation 21.3) without actually knowing the field. We can then solve for E in terms of the enclosed charge. We begin with a general strategy for applying Gauss's law to symmetric charge distributions, followed by examples of the three symmetries.

PROBLEM-SOLVING STRATEGY 21.1 Gauss's Law

INTERPRET Check that your charge distribution has sufficient symmetry to use Gauss's law to find the electric field, and identify the symmetry. Is it spherical, line, or plane? If the charge distribution doesn't exhibit one of these symmetries, then Gauss's law—though always true—won't help you find the field.

DEVELOP Draw a diagram showing the charge distribution, and use the symmetry to infer the direction of the electric field. Then draw an appropriate **Gaussian surface**—an imaginary, closed surface that will let you evaluate the flux integral in Gauss's law. The field should have constant magnitude over the surface and should be perpendicular to the surface. Sketch some field lines; the symmetry should indicate their direction. With line and plane symmetry, there may be parts of the surface where the field is perpendicular and parts where it's parallel; we'll show in examples how to handle these situations. If you can't find a suitable Gaussian surface, there probably isn't sufficient symmetry to use Gauss's law to calculate the field.

EVALUATE

- Evaluate the flux $\Phi = \oint \vec{E} \cdot d\vec{A}$ over your Gaussian surface. Since you've found a surface to which the field is perpendicular, $\vec{E}$ and the surface normal vector $d\vec{A}$ are parallel, so $\cos\theta = 1$ and the dot product becomes $E\,dA$. With the field strength E constant over the surface, it can come out of the integral, leaving $\oint dA$. And that's the surface area A. So the flux will be EA. If $\vec{E}$ is parallel to some parts of the area—as happens in line and plane symmetry—then $\vec{E} \perp d\vec{A}$, so $\vec{E} \cdot d\vec{A} = 0$ and there's no contribution to the flux from those areas.
- Evaluate the *enclosed* charge. This may or may not be the same as the total charge, depending on whether the position at which you're evaluating the field lies outside or inside the charge distribution.
- Evaluate the field E by invoking Gauss's law, equating the flux to $q_{\text{enclosed}}/\epsilon_0$, and solving for E. This is the field magnitude; the direction should be evident from symmetry.

ASSESS Does your answer make sense? Does the field behave as you would expect given what you know of simpler charge distributions—point charges, line charges, or charged sheets, depending on the symmetry?

You'll quickly get the hang of this strategy because Gauss's law is useful for finding the field only in the three common symmetries. That means you need to evaluate the integral for the flux just once for a given symmetry.

Spherical Symmetry

A charge distribution has spherical symmetry when the charge density depends only on the radial distance r from the center of the distribution—also called the *point of symmetry*. A point charge is one example; so is the uniformly charged sphere that we'll treat in the next example. So, in fact, is a nonuniform spherical charge, provided the charge density varies solely with distance from the center. The only electric field consistent with spherical symmetry is a field that points in the radial direction, either away from or toward the point of symmetry. The next example shows the application of Gauss's law to spherical symmetry.

EXAMPLE 21.1 Gauss's Law: A Uniformly Charged Sphere

A charge Q is spread uniformly throughout a sphere of radius R. Find the electric field at all points, first outside and then inside the sphere.

INTERPRET The charge distribution has spherical symmetry, so we can use Gauss's law to find the field.

DEVELOP Figure 21.9 shows the spherical charge distribution. Given the symmetry, the field has to point radially, outward for a positive charge or inward for a negative one. We've shown some field lines on our drawing. So an appropriate Gaussian surface is itself a sphere, centered on the center of the charge distribution. Since we're asked for the field both outside and inside the charge distribution, we've drawn two such Gaussian surfaces in Fig. 21.9, one outside and one inside the distribution.

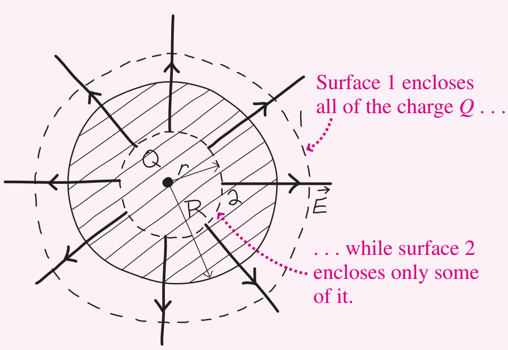


FIGURE 21.9 Finding the field of a uniformly charged sphere.

EVALUATE

- We begin with the flux integral $\Phi = \oint \vec{E} \cdot d\vec{A}$. Our strategy shows that with $\vec{E}$ perpendicular to the area, $\cos\theta = 1$ and the dot product becomes $E\,dA$. So the flux integral simplifies to $\oint E\,dA$. With the Gaussian surface centered on the point of symmetry, all points on the surface are the same distance from the center, so E has the same value over our surface. E comes outside the integral, leaving $\oint dA$, with the integral over the entire spherical surface. That's just the surface area, $4\pi r^2$, so the flux is $\Phi = 4\pi r^2 E$. Which surface were we using in doing this calculation? It doesn't matter; the result holds for *any* spherical surface centered on the symmetry point at the center of the charge distribution. So our expression $\Phi = 4\pi r^2 E$ is valid for both our Gaussian surfaces.
- Now we evaluate the *enclosed* charge. That *is* different for the two surfaces, so we'll first deal with surface 1, which lies outside the charge distribution. This Gaussian surface encloses the entire charge Q, so $q_{\text{enclosed}} = Q$ for $r > R$. The interior is a bit trickier. Here, Gaussian surface 2 has radius r, so it encloses volume $\frac{4}{3}\pi r^3$. The volume of the entire sphere is $\frac{4}{3}\pi R^3$, so Gaussian surface 2 encloses a fraction r^3/R^3 of the total volume. We're told that the sphere is uniformly charged, so that's also the fraction of the total charge Q enclosed by the Gaussian surface. So for $r < R$, the enclosed charge is $q_{\text{enclosed}} = Q(r^3/R^3)$.

(continued)

- Next we apply Gauss's law, evaluating the field by equating the flux $\Phi = 4\pi r^2 E$ to the enclosed charge divided by ϵ_0. When we're outside the sphere, $q_{\text{enclosed}} = Q$, so we have $4\pi r^2 E = Q/\epsilon_0$, or

$$E = \frac{Q}{4\pi\epsilon_0 r^2} \quad \text{(field outside spherical charge distribution)} \quad (21.4)$$

Inside the sphere, $q_{\text{enclosed}} = Q(r^3/R^3)$; equating $q_{\text{enclosed}}/\epsilon_0$ to the flux $\Phi = 4\pi r^2 E$ gives $4\pi r^2 E = Qr^3/\epsilon_0 R^3$. Solving for E, we get

$$E = \frac{Qr}{4\pi\epsilon_0 R^3} \quad \text{(field inside uniformly charged sphere)} \quad (21.5)$$

ASSESS Make sense? Consider first the field outside the spherical charge. Since $1/4\pi\epsilon_0$ is the Coulomb constant k, this field can be written as $E = kQ/r^2$—precisely the field of a point charge! Inside, in contrast, the field increases linearly with distance from the center, reflecting two opposite effects. First, we're getting farther from the center, which causes a $1/r^2$ decrease in the field. At the same time, the enclosed charge grows as the enclosed volume—that is, as r^3. The result is a linear increase in field strength. Figure 21.10 shows the combined results for the field both inside and outside the sphere. As a check on our results, the figure shows that our two expressions agree at $r = R$. ■

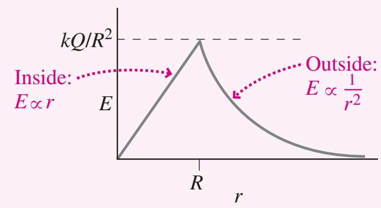


FIGURE 21.10 Field strength versus radial distance for a uniformly charged sphere of radius *R*.

Our result for the field outside a spherical charge distribution (Equation 21.4) is *not* an approximation but an exact result that holds right up to the sphere's surface. So a spherical charge distribution produces exactly the same field as a point charge located at its center—at least for points outside the charge. Imagine how hard it would have been to calculate this using the superposition principle! Yet somehow all the charge elements produce $d\vec{E}$ vectors that add to give the same field as a single point charge. That result doesn't even require the sphere to be charged uniformly, as long as the distribution of charge is spherically symmetric. Outside *any* spherically symmetric charge distribution, the electric field is exactly the same as that of a point charge at the center.

Incidentally, this result also holds for gravity because it, too, obeys an inverse-square law. That's why we can treat planets as though they're point masses located at their centers. And our result for the field inside the charged sphere (Equation 21.5) shows also why the gravitational acceleration g decreases approximately linearly as one descends into Earth's interior.

What if our charged sphere wasn't uniformly charged throughout its interior? The next example considers the extreme case when there's no charge in the interior.

EXAMPLE 21.2 Gauss's Law: A Hollow Spherical Shell

A thin, hollow spherical shell of radius R contains a total charge Q distributed uniformly over its surface. Find the electric field both inside and outside the shell.

INTERPRET This is a spherically symmetric charge distribution, so we can apply Gauss's law just as we did in Example 21.1. There we found that the field outside *any* spherically symmetric charge distribution is that of a point charge: $E = kQ/r^2$, pointing radially. So we need to find the field only inside the spherical shell.

DEVELOP Figure 21.11 is a sketch of the hollow shell. With spherical symmetry, the appropriate Gaussian surface is itself a concentric sphere, which we've drawn inside the shell.

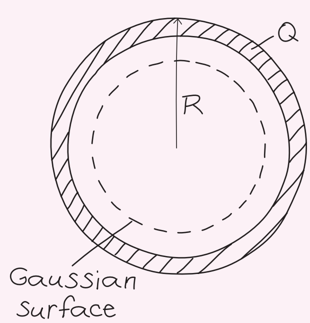


FIGURE 21.11 Sketch for Example 21.2.

EVALUATE

- The flux $\Phi = 4\pi r^2 E$ from Example 21.1 holds for *any* spherical Gaussian surface centered on a spherically symmetric charge distribution, so we've already got the flux.
- Next we need the enclosed charge. But there's no charge inside the hollow shell and therefore none inside our Gaussian surface. So $q_{\text{enclosed}} = 0$.
- Equating the flux $4\pi r^2 E$ to $q_{\text{enclosed}}/\epsilon_0$—zero in this case—gives $4\pi r^2 E = 0$. Therefore, $E = 0$ inside the shell.

ASSESS Make sense? The radius r of our Gaussian sphere is arbitrary as long as $r < R$, so we're *inside* the shell. Thus this example shows that the electric field is exactly zero *everywhere* inside the shell. Figure 21.12 shows how this remarkable result is a consequence of the inverse-square law. ■

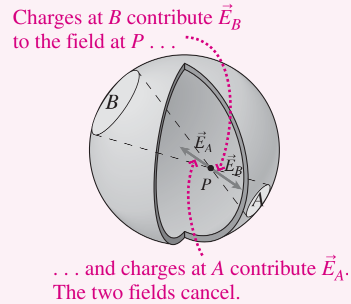


FIGURE 21.12 At any point P inside a charged shell, the field from the relatively few nearby charges at A is exactly canceled by the field from the more numerous but more distant charges at B. The result is zero field everywhere inside the shell.

GOT IT? 21.3 A spherical shell carries charge Q uniformly distributed over its surface. If the charge on the shell doubles, what happens to the electric field (a) inside and (b) outside the shell?

EXAMPLE 21.3 Gauss's Law: A Point Charge Within a Shell

A positive point charge $+q$ is at the center of a spherical shell of radius R carrying charge $-2q$, distributed uniformly over its surface. Find expressions for the field strength inside and outside the shell.

INTERPRET This problem is about a charge distribution with spherical symmetry.

DEVELOP We want to find the fields both inside and outside the distribution, so we show two spherical Gaussian surfaces in our sketch, Fig. 21.13. We've also drawn some field lines, which will help in assessing our answer.

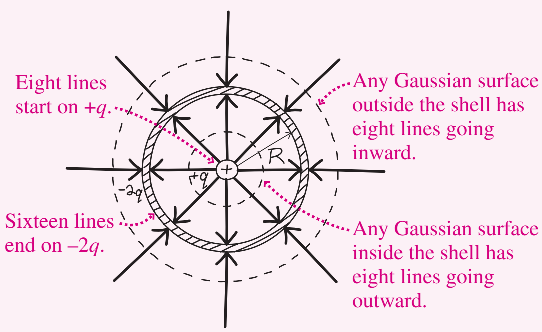


FIGURE 21.13 A shell carrying charge $-2q$ surrounds a point charge $+q$.

EVALUATE

- We already know that the flux through a spherical Gaussian surface is $\Phi = 4\pi r^2 E$ when we have spherical symmetry, so that's the flux for both surfaces.
- The outer surface encloses the charge $+q$ at the center and $-2q$ on the shell, so the net charge enclosed is $q_{\text{enclosed}} = -q$ for $r > R$. The inner surface encloses only the point charge $+q$, so $q_{\text{enclosed}} = +q$ for $r < R$.
- Now we find the field by equating the flux $\Phi = 4\pi r^2 E$ to $q_{\text{enclosed}}/\epsilon_0$. The result for $r > R$ is $E = -q/4\pi\epsilon_0 r^2$, where the minus sign appears because the enclosed charge is $-q$. For $r < R$ with enclosed charge $+q$, the result is $E = q/4\pi\epsilon_0 r^2$.

ASSESS Make sense? We've seen that the field outside a spherically symmetric charge distribution is that of an equivalent point charge at the center. Here the net charge is $-q$, and our result for $r > R$ is indeed the field of a point charge $-q$. Another way to see this is to apply superposition: The field outside due to the shell alone is that of a point charge $-2q$; that adds to the field of the central point charge $+q$ to give, again, the field of a point charge $-q$ outside the shell. In Example 21.2 we found that the shell produces no field in its interior, so here superposition leaves us with the field of the central point charge alone, just as our result shows. The field lines in Fig. 21.13 also show how our results make sense. ■

✓TIP Symmetry Matters!

We used the fact that there's no charge inside an empty spherical shell to conclude that the field inside the shell is zero. But be careful: That conclusion follows only when there's enough symmetry, as Fig. 21.14 makes obvious. Here, the Gaussian surface encloses zero net charge, so the flux through the surface is zero. But the electric field both on and inside the surface isn't zero.

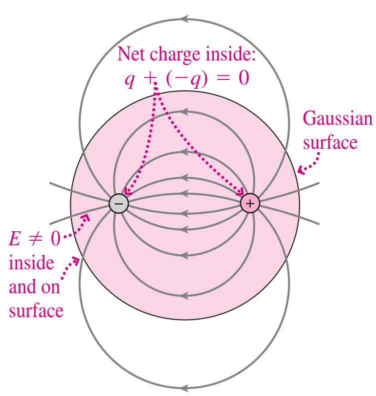


FIGURE 21.14 A spherical Gaussian surface surrounds a dipole. We have $q_{\text{enclosed}} = 0$, but $E \neq 0$ inside the surface because the dipole isn't spherically symmetric.

Line Symmetry

A charge distribution has line symmetry when its charge density depends only on the perpendicular distance r from a line, called the *symmetry axis*. Symmetry then requires that the field point radially and that the field magnitude depend only on distance from the axis. It also requires the charge distribution to be infinitely long, so there's no variation parallel to the line. That's impossible, of course, but nevertheless the infinite line is a reasonable approximation to elongated structures like wires. The next two examples explore the application of Gauss's law to line symmetry.

EXAMPLE 21.4 Gauss's Law: An Infinite Line of Charge

Use Gauss's law to find the electric field of an infinite line charge carrying charge density λ in coulombs per meter.

INTERPRET An infinite line has line symmetry, so we can apply Gauss's law to find the electric field.

DEVELOP Symmetry requires that the field point radially and its magnitude be the same at a given distance r from the line charge. So an appropriate Gaussian surface is a cylinder coaxial with the line, as we've drawn in Fig. 21.15.

EVALUATE

- We begin with the flux $\Phi = \oint \vec{E} \cdot d\vec{A}$. $\vec{E}$ is everywhere perpendicular to the curved part of the cylinder, so here the dot product $\vec{E} \cdot d\vec{A}$ becomes $E\,dA$, and this part of the flux integral simplifies to $\int_{\text{curved part}} E\,dA$. But the cylinder has ends, and Gauss's law requires that we consider the flux through those ends, too. No field lines emerge from the ends, so there's no flux. Mathematically, $\vec{E}$ is perpendicular to the normal vector $d\vec{A}$, so $\vec{E} \cdot d\vec{A} = 0$ on the ends. All we need is the flux through the curved part. Since the cylindrical Gaussian surface is concentric with the line of symmetry, all points on it are the same distance r from the line, and thus the magnitude E is the same over the curved part of the cylinder. So we can take E outside the integral, giving $E\int_{\text{curved part}} dA$. The integral is just the surface area of the curved part, which, if you unwound it, would be a rectangle of length L and width equal to the circumference $2\pi r$. So the surface area is $2\pi rL$, and the flux becomes $\Phi = 2\pi rLE$. This result depends only on the symmetry, so it applies in all problems involving line symmetry.
- Next we need the enclosed charge. The line carries λ C/m and our Gaussian cylinder encloses L meters of the line, so $q_{\text{enclosed}} = \lambda L$.
- Finally, we invoke Gauss's law by equating the flux Φ to $q_{\text{enclosed}}/\epsilon_0$. The result, $2\pi rLE = \lambda L/\epsilon_0$, solves to give

$$E = \frac{\lambda}{2\pi\epsilon_0 r} \qquad \text{(field of a line charge)} \qquad (21.6)$$

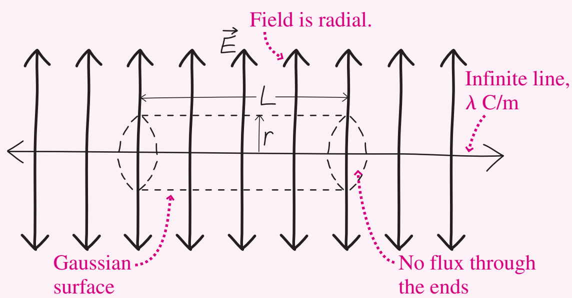


FIGURE 21.15 A cylindrical Gaussian surface surrounds a line charge.

ASSESS Make sense? We worked this same problem in Example 20.7 in a much more difficult calculation involving a complicated integral. Since $1/2\pi\epsilon_0 = 2k$, our result here is the same. But Gauss's law makes the problem much easier! ■

Although Example 21.4 involved an infinitesimally thin line of charge, you can see from Fig. 21.16 that our result must hold *outside* any charge with line symmetry. And, as we argued in Example 20.7, it's a good approximation for the field of any long, cylindrical structure as long as we're not too near its ends.

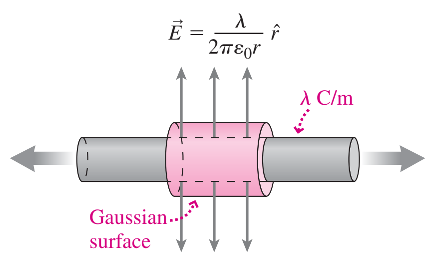


FIGURE 21.16 The arguments of Example 21.4 apply outside *any* cylindrical charge distribution.

EXAMPLE 21.5 Gauss's Law: A Hollow Pipe

A thin-walled pipe 3.0 m long and 2.0 cm in radius carries a net charge $q = 5.7\ \mu C$ distributed uniformly over its surface. Find the electric field both 1.0 cm and 3.0 cm from the pipe axis, far from either end.

INTERPRET Although the pipe has finite length, both distances are small compared with the length, so we can approximate the pipe as an infinitely long structure with line symmetry.

DEVELOP With line symmetry the appropriate Gaussian surface is a cylinder coaxial with the pipe. We've drawn two such cylinders in Fig. 21.17, one for each radius where we're asked to evaluate the field.

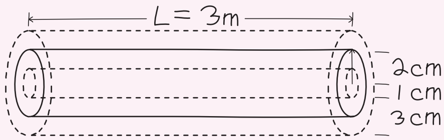


FIGURE 21.17 Gaussian surfaces for Example 21.5.

EVALUATE
- We showed in Example 21.4 that the flux integral in line symmetry gives $\Phi = 2\pi rLE$.
- Next we need the enclosed charge. At 3 cm we're outside the pipe, so the Gaussian surface with this radius encloses all the charge: $q_{\text{enclosed}} = 5.7\ \mu C$. The pipe is hollow, so at 1 cm the enclosed charge is zero.
- Equating the flux to $q_{\text{enclosed}}/\epsilon_0$ and solving for E give

$$E = \frac{q_{\text{enclosed}}}{2\pi\epsilon_0 rL} = \frac{5.7\ \mu C}{(2\pi\epsilon_0)(3.0\times 10^{-2}\ \text{m})(3.0\ \text{m})} = 1.1\ \text{MN/C}$$

for the field at 3 cm and $E = 0$ for the field inside the pipe.

ASSESS Make sense? *Inside* the pipe, there's no field, and for the same reason as inside a uniformly charged hollow sphere—namely, that fields from near and far parts of the pipe cancel due, ultimately, to the inverse-square law. Again, be careful: That result follows because of the symmetry, although we'll soon see that with *conducting* pipes and shells there's no interior field even without that symmetry. We argued earlier that the field *outside* any line-symmetric distribution should be given by Equation 21.6, $E = \lambda/2\pi\epsilon_0 r$. In our result, the quantity q_{enclosed}/L is the line charge density λ, so our result is indeed consistent with that equation. ■

Plane Symmetry

A charge distribution has plane symmetry when its charge density depends only on the perpendicular distance from a plane. The only electric-field direction consistent with this symmetry is perpendicular to the plane. As with line symmetry, true plane symmetry implies a charge distribution that's infinite in extent. That's impossible—but plane symmetry remains a good approximation when charge is spread uniformly over large, flat surfaces or slabs. The next example applies Gauss's law to plane symmetry.

EXAMPLE 21.6 Gauss's Law: A Sheet of Charge

An infinite sheet of charge carries uniform surface charge density σ in coulombs per square meter. Find the resulting electric field.

INTERPRET Since the sheet is infinite, we have plane symmetry with the sheet itself as the symmetry plane.

DEVELOP Figure 21.18 shows the Gaussian surface as a cylinder that straddles the charged sheet, extending equally on either side. Then the electric field is perpendicular to the ends and parallel to the sides; symmetry requires that it be uniform over the ends.

EVALUATE
- First we evaluate the flux through our Gaussian cylinder. The field is parallel to the sides, so there's no flux contribution here. Then the total flux is through the ends, each of which has area A. Since the field is uniform over each end and perpendicular to the ends, the flux through each end is EA. So the total flux through both ends is $\Phi = 2EA$. This result is independent of the details of the charge distribution, so it holds in all cases of plane symmetry.

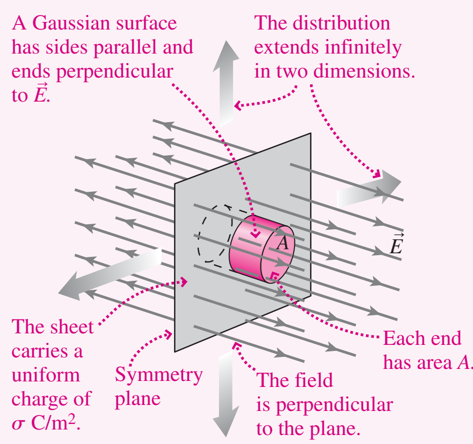


FIGURE 21.18 An infinite sheet of charge, with a Gaussian surface straddling the sheet.

- Next we need the enclosed charge. The Gaussian surface encloses an area of the sheet equal to the area A of its ends. With surface charge density σ on the sheet, the enclosed charge is $q_{\text{enclosed}} = \sigma A$.

(continued)

- Now we apply Gauss's law, equating the flux to $q_{\text{enclosed}}/\epsilon_0$. Thus $2EA = \sigma A/\epsilon_0$, so

$$E = \frac{\sigma}{2\epsilon_0} \qquad \text{(field of a charged sheet)} \qquad (21.7)$$

The direction of this field on either side of the sheet is outward from the sheet if it's positively charged and inward if negative.

ASSESS Make sense? With an infinite plane, symmetry requires that the field lines be perpendicular to the plane. So they don't spread out, and that means the field doesn't vary with distance—as Equation 21.7 shows because it doesn't involve the distance from the sheet. Although our result is exact only for a truly infinite sheet, it's a good approximation near any large, flat, uniformly charged surface as long as we're not close to an edge. ■

21.5 Fields of Arbitrary Charge Distributions

Although Gauss's law is always true, most charge distributions lack the symmetry needed to apply Gauss's law to find the field. The alternative, Coulomb's law, is hard to use in all but the simplest cases. But we can often learn a lot by considering the distributions whose fields we calculated here and in Chapter 20. Figure 21.19 summarizes four of these fields. For the last three, note the simple relation between the number of dimensions and the behavior of the field. The plane has two dimensions, and its field doesn't decrease with distance. The line has one dimension, and its field decreases as $1/r$. The point has no dimensions, and its field falls as $1/r^2$. In a sense, the dipole continues this progression. It consists of opposite point charges whose effects nearly cancel; no wonder its field decreases faster still, as $1/r^3$. In fact, there's a hierarchy of charge distributions whose fields decrease ever faster, as dipoles nearly cancel dipoles, and so on. Scientists and engineers use this hierarchy in modeling charged structures ranging from molecules to radio antennas.

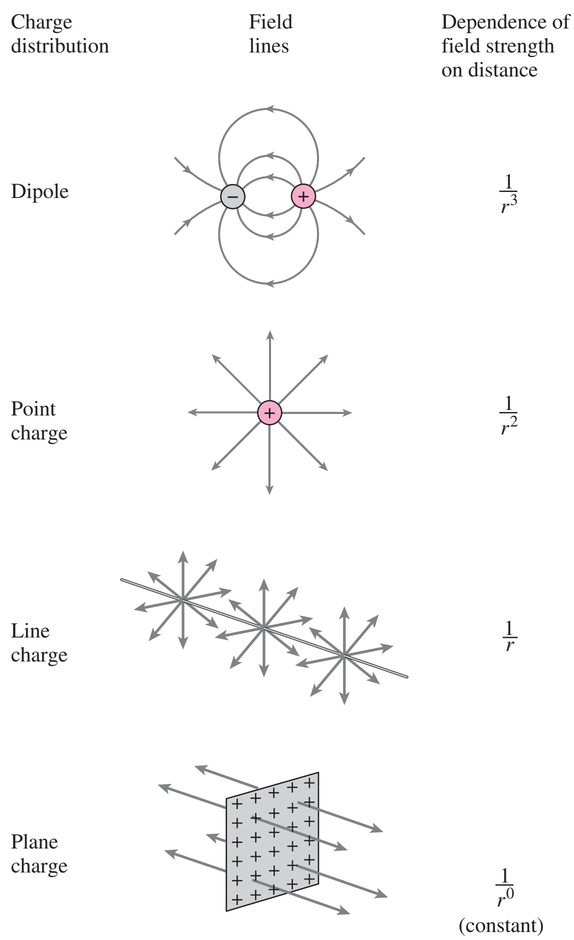


FIGURE 21.19 Fields of a dipole, a point charge, a charged line, and a charged plane.

CONCEPTUAL EXAMPLE 21.1 **A Charged Disk**

Sketch some electric field lines for a uniformly charged disk, starting at the disk and extending out to several disk diameters.

EVALUATE When we're near the disk and not close to its edge, the disk looks like a large, flat, charged plane. Its field is essentially the uniform field of an infinite plane charge, so we draw straight field lines emanating perpendicular to the disk. Far from *any* finite distribution carrying a nonzero net charge, the field approximates that of a point charge, so farther away we draw field lines going radially outward. Field lines begin only on charges—in this case on the charged disk—so we have to connect close-in and far-out lines. We don't know exactly how the field looks in the intermediate region neither close to nor far from the disk, so we connect them as best we can. Figure 21.20 is the result.

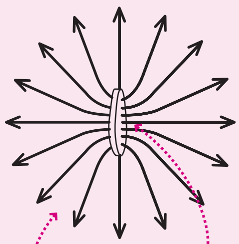


FIGURE 21.20 The field of a charged disk.

ASSESS Our sketch is a good approximation to the field of a charged disk. And it certainly obeys Gauss's law because the same number of field lines—namely, all 16 lines we chose to draw—cross any closed surface surrounding the entire disk.

MAKING THE CONNECTION Suppose the disk is 1.0 cm in diameter and carries charge 20 nC spread uniformly over its surface. Find the electric field strength (a) 1.0 mm from the disk surface and (b) 1.0 m from the disk.

EVALUATE (a) Close to the disk, assuming we're not near the edge, Equation 21.7 applies: $E = \sigma/2\epsilon_0 = 14$ MN/C, where we used the total charge and the disk area to get the surface charge density σ. (b) At 1 meter, the disk is so small it looks essentially like a point charge, so Equation 20.3 applies: $E = kq/r^2 = 180$ N/C.

21.6 Gauss's Law and Conductors

Electrostatic Equilibrium

We've defined conductors as materials that contain free charges, like the free electrons in metals. Figure 21.21 shows what happens when an electric field is applied to a conductor. Free charges respond to the electric force $q\vec{E}$ by moving—in the direction of the field if they're positive, opposite the field if negative. The resulting charge separation gives rise to an electric field within the material that's opposite to the applied field. As more charge moves, this internal field becomes stronger until its magnitude eventually equals that of the applied field. At that point free charges within the conductor experience zero net force, and the conductor is in **electrostatic equilibrium**. Although individual charges continue to move about in random thermal motion, there's no longer any net charge motion. Once equilibrium is reached, the internal and applied fields are equal but opposite, and therefore:

The electric field is zero inside a conductor in electrostatic equilibrium.

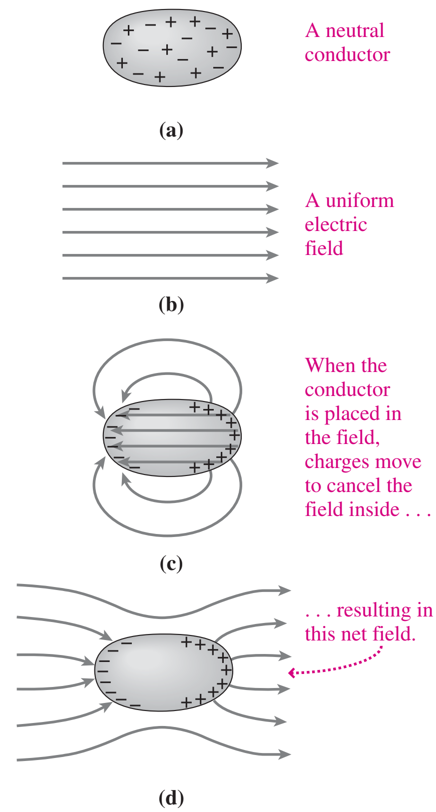


FIGURE 21.21 A conductor in a uniform electric field.

It could not be otherwise: Since a conductor contains free charges, the presence of any internal electric field would result in bulk charge motion, and we wouldn't have equilibrium. This result doesn't depend on the size or shape of the conductor, the magnitude or direction of the applied field, or even the nature of the material as long as it's a conductor. This is a macroscopic view; it considers only average fields within the material. At the atomic and molecular level, there are still strong electric fields near individual electrons and positive ions. But the *average* field, taken over larger distances, is zero inside a conductor in electrostatic equilibrium.

Charged Conductors

Although they contain free charges, conductors are normally electrically neutral because they include equal numbers of electrons and protons. But suppose we give a conductor a nonzero net charge, for example, by injecting excess electrons into its interior. There's a

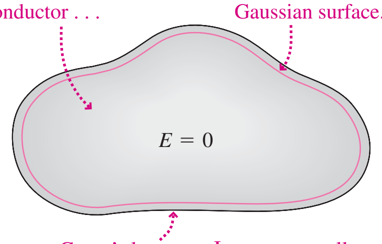


FIGURE 21.22 Gauss's law implies that any net charge resides on the surface of a conductor in electrostatic equilibrium.

mutual repulsion among the electrons and, because these are *excess* electrons, there's no compensating attraction from positive charges. We might expect, therefore, that the electrons will move as far apart as possible—namely, to the surface of the conductor.

We now use Gauss's law to prove that excess charge *must* be at the surface of a conductor in electrostatic equilibrium. Figure 21.22 shows a conducting material with a Gaussian surface drawn just below the material surface. In equilibrium there's no electric field inside the conductor, and thus the field is zero everywhere on the Gaussian surface. The flux, $\oint \vec{E} \cdot d\vec{A}$, through the Gaussian surface is therefore also zero. But Gauss's law says that the flux through a closed surface is proportional to the net charge enclosed, and therefore the net charge inside our Gaussian surface must be zero. This is true no matter where the Gaussian surface is as long as it's *inside* the conductor. We can move it arbitrarily close to the conductor surface and it still encloses no net charge. If there is a net charge on the conductor, it lies outside the Gaussian surface, and therefore we conclude: **If a conductor in electrostatic equilibrium carries a net charge, it must reside on the conductor surface**.

EXAMPLE 21.7 Gauss's Law: A Hollow Conductor

An irregularly shaped conductor has a hollow cavity. The conductor itself carries a net charge of 1 μC, and there's a 2-μC point charge inside the cavity. Find the net charge on the cavity wall and on the outer surface of the conductor, assuming electrostatic equilibrium.

INTERPRET This problem involves a conductor in electrostatic equilibrium, which means (1) there's no electric field inside the conducting material and (2) the net charge resides on the conductor surface—which in this case includes both the inner and outer surfaces.

DEVELOP We draw the situation in Fig. 21.23. Our plan is to apply Gauss's law to find the charges. We consider a Gaussian surface inside the conductor and enclosing the cavity, as shown.

EVALUATE Since there's no electric field inside the conductor, the flux through the Gaussian surface is zero, and therefore the net charge enclosed is also zero. But there's that $+2$-μC point charge in the cavity. For the Gaussian surface to enclose zero net charge, there must be $-2\ \mu$C somewhere else—and the only place it can be is on the cavity wall. However, the entire conductor carries $+1\ \mu$C. With $-2\ \mu$C on the inside wall, that leaves $+3\ \mu$C on the outer surface.

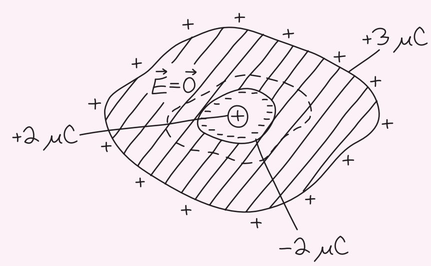


FIGURE 21.23 The Gaussian surface encloses zero net charge, so there must be $-2\ \mu$C on the cavity wall.

ASSESS Make sense? Yes: This distribution of charge is the only one that's consistent with both Gauss's law and the requirement $E = 0$ inside a conductor in electrostatic equilibrium. As another check, think about what this charge distribution must look like from far away—namely, a point charge with net charge of 3 μC. Since the fields of the cavity wall and the inner point charge don't penetrate the conductor, the only field lines that reach out beyond the conductor are those from the charge on its outer surface. So that charge must be 3 μC, as we've found. ■

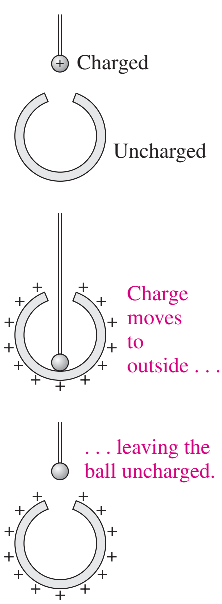


FIGURE 21.24 Experimental test of Gauss's law.

GOT IT? 21.4 A conductor carries a net charge $+Q$. There's a cavity inside the conductor that contains a point charge $-Q$. In electrostatic equilibrium, is the charge on the outer surface of the conductor (a) $-2Q$, (b) $-Q$, (c) 0, (d) Q, or (e) $2Q$?

Experimental Tests of Gauss's Law

That net charge moves to a conductor surface provides a sensitive test of Gauss's law and thus—through the arguments of pp. 355–356—a test of the inverse-square law for the electric force. Figure 21.24 shows a charged conducting ball touched to the inside of a hollow, initially neutral conductor. As required by Gauss's law, charge flows to the outer surface of the hollow conductor, leaving the ball uncharged. Measuring zero charge on the ball confirms Gauss's law, and thus the inverse-square law; such experiments show that the exponent in $1/r^2$ is indeed 2 to some 16 decimal places!

The Field at a Conductor Surface

There can't be an electric field *within* a conductor in electrostatic equilibrium, but there *may* be a field right at the conductor surface (Fig. 21.25*a*). Such a field must be perpendicular to the surface; otherwise, charge would move along the surface in response to the field's parallel component, and we wouldn't have equilibrium.

We can compute the strength of this surface field by considering a small Gaussian surface that straddles the conductor surface, as shown in Fig. 21.25*b*. There's no flux through the sides, and because the field is zero inside the conductor, there's no flux through the inner end either. So the only flux is through the outer end, with area A. Since the end is perpendicular to the field, the flux is EA. The Gaussian surface encloses charge σA, where σ is the surface charge density. Gauss's law equates the flux with $q_{\text{enclosed}}/\epsilon_0$, so we have $EA = \sigma A/\epsilon_0$, or

$$E = \frac{\sigma}{\epsilon_0} \qquad \text{(field at conductor surface)} \tag{21.8}$$

This result shows that large fields develop where the charge density on a conductor is high. Engineers who design electrical devices must avoid high charge densities whose associated fields lead to sparks, arcing, and breakdown of electric insulation.

Equation 21.8 gives a field that depends only on the local charge density. Does that mean the field at a conductor surface arises only from the local charge? No! As always, the field is the vector sum of contributions from all charges. Remarkably, Gauss's law requires that charges on a conductor arrange themselves in such a way that the field at any point on the conductor surface depends only on the surface charge density right at that point—even though that field arises from *all* the charges on the surface (as well as from charges elsewhere if there are any)!

Consider a thin, flat, isolated, conducting sheet that has charge density σ on one of its two faces (Fig. 21.26*a*). Equation 21.8 shows that the field at the surface of this plate is σ/ϵ_0. But if the plate is large and flat, we can approximate it as an infinite sheet of charge—for which we found earlier (Equation 21.7) that the field should be $\sigma/2\epsilon_0$. Is there a contradiction here? No! If the plate is isolated, then symmetry requires that the charge spread itself evenly over *both* faces. If one face has charge density σ, so must the other—so we really have *two* charge sheets, each with density σ (Fig 21.26*b*). Each gives a field of magnitude $\sigma/2\epsilon_0$, and *outside* the conductor those fields superpose to give the net field σ/ϵ_0 (Fig. 21.26*c*). *Inside* the conductor their directions are opposite, and the result is zero field inside the conductor. Applying Equation 21.8 skips these details. But because Equation 21.8 was derived on the assumption that the field inside the conductor is zero, it

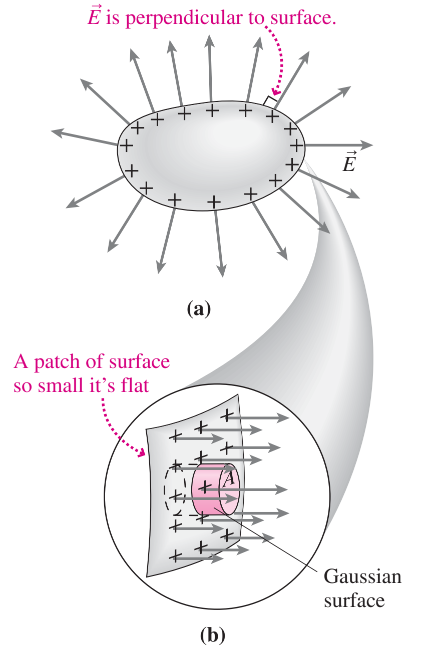


FIGURE 21.25 (a) The electric field at the surface of a charged conductor is perpendicular to the conductor surface. (b) A Gaussian surface straddles the conductor surface.

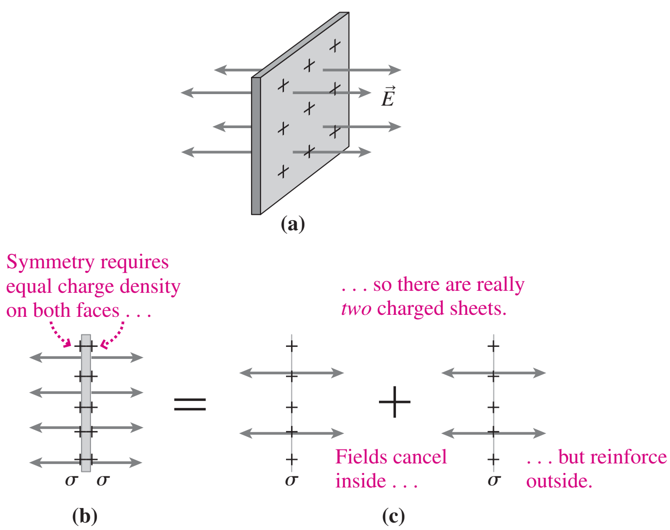


FIGURE 21.26 (a) An isolated, charged conducting plate. Its field points outward from both faces. (b) Edge-on view of the plate. (c) The field anywhere is the sum of the fields of the two faces, each treated as a single charged sheet.

APPLICATION Shielding and Lightning Safety

We've seen that charge moves to the outside of a conductor surface, leaving the interior free of charge and electric field—even if the interior is hollow. This is the basis of electric shielding, in which a conducting enclosure keeps out external electric fields. A common example is the *coaxial cable* that delivers TV signals from your cable company; coaxial cables also connect electronic instruments in scientific and medical research. A coaxial cable consists of an inner wire surrounded by a cylindrical conducting shield in which charge moves to block external electric fields that could cause interference. In another application of shielding, researchers doing experiments with very weak electric signals often construct entire rooms with conducting walls to minimize interference.

Shielding is also the reason a car is a relatively safe place in a thunderstorm. A lightning strike dumps charge on the car's metal body, but the charge distributes itself on the outside so as to prevent any electric fields from developing inside the car (see photo). That, in turn, prevents harmful currents from flowing through the cars' occupants. The operator's cage in this chapter's opening photo has the same effect, harmlessly deflecting charge from the artificial lightning and keeping the interior free of electric fields.

Strictly speaking, charge resides on the outside of a conductor only in equilibrium. But electrons in metals respond so quickly that equilibrium results almost instantaneously—meaning that metallic shielding is effective even against the rapidly varying electric fields of high-frequency radio, TV, and microwave signals.

"knows" about charges everywhere on the conductor—and in this case that means on the second face.

Equation 21.8 also applies to a pair of oppositely charged conducting plates (Fig. 21.27); the result, for the field between the plates, is σ/ϵ_0, where σ is the surface charge density on either plate. Why not $2\sigma/\epsilon_0$? Again, Equation 21.8 gives the field at a conductor surface—and it takes into account other charges that may be present. Here each plate's charge attracts the other's opposite charge, all to the *inner* face. Each plate is thus a single charge layer, giving a field $\sigma/2\epsilon_0$, and between the plates the fields sum to Equation 21.8's result, σ/ϵ_0. Beyond the plates the fields sum to zero—a result that also follows from Equation 21.8 because now there's zero surface charge on the outer faces.

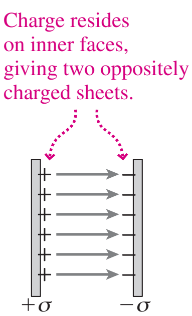
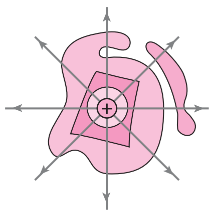


FIGURE 21.27 Edge-on view of two parallel conducting plates carrying opposite charges.

CHAPTER 21 SUMMARY

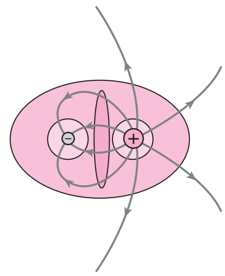
Big Picture

The big idea here is **Gauss's law**—a universal statement about electric fields that's closely related to Coulomb's inverse-square law but expressed in terms of the global behavior of the field over any closed surface. Using the **electric field-line** picture, Gauss's law says that the number of field lines emerging from a closed surface depends only on the net charge enclosed; more rigorously, it says that the **electric flux** through the surface is proportional to the enclosed charge.



Eight lines pass through any closed surface surrounding q.

Point charge q



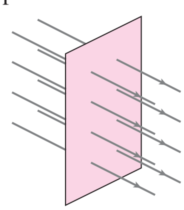
The number of lines through a surface depends on the net charge enclosed.

Point charges $+q$, $-q/2$

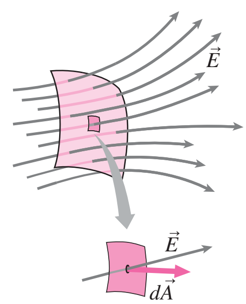
Key Concepts and Equations

Electric flux Φ describes the amount of electric field crossing an area.

$\Phi = EA$ for a flat area perpendicular to a uniform field



$\Phi = \int \vec{E} \cdot d\vec{A}$, in general



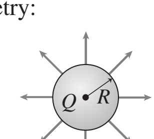
In terms of flux, Gauss's law reads $\oint \vec{E} \cdot d\vec{A} = q_{\text{enclosed}}/\epsilon_0$.

Here $\epsilon_0 = 1/4\pi k = 8.85 \times 10^{-12}\ \text{C}^2/\text{N}\cdot\text{m}^2$ is another way of expressing the Coulomb constant $k = 9.0 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2$.

Applications

Gauss's law gives the fields of symmetric charge distributions:

Spherical symmetry:



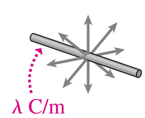
Outside: $E = \dfrac{Q}{4\pi\epsilon_0 r^2} = \dfrac{kQ}{r^2}$

Inside uniformly charged sphere:

$$E = \frac{kQr}{R^3}$$

Inside hollow sphere: $E = 0$

Line symmetry:

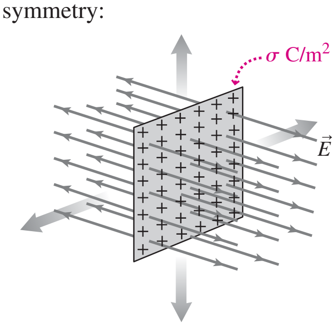


Outside: $E = \dfrac{\lambda}{2\pi\epsilon_0 r}$

Inside hollow pipe:

$$E = 0$$

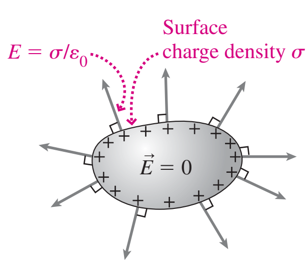
Plane symmetry:



Outside charged slab: $E = \dfrac{\sigma}{2\epsilon_0}$

Gauss's law and conductors:

- The field is zero inside a conductor in electrostatic equilibrium.
- Any net charge resides on a conductor's surface.
- The field at the surface is perpendicular to the surface and has magnitude σ/ϵ_0.



For Thought and Discussion

1. Can electric field lines ever cross? Why or why not?
2. The electric flux through a closed surface is zero. Must the electric field be zero on that surface? If not, give an example.
3. If the flux of the gravitational field through a closed surface is zero, what can you conclude about the region interior to the surface?
4. Under what conditions can the electric flux through a surface be written as EA, where A is the surface area?
5. Eight field lines emerge from a closed surface surrounding an isolated point charge. Would the number of field lines change if a second identical charge were brought to a point just *outside* the surface? If not, would anything change? Explain.
6. If a charged particle were released from rest on a curved field line, would its subsequent motion follow the field line? Explain.
7. In Gauss's law, $\oint \vec{E} \cdot d\vec{A} = q/\epsilon_0$, does the field $\vec{E}$ necessarily arise only from charges within the closed surface?
8. In a certain region the electric field points to the right and its magnitude increases as you move to the right, as shown in Fig. 21.28. Does the region contain net positive charge, net negative charge, or zero net charge?

FIGURE 21.28 For Thought and Discussion 8. Left side marks the beginning of the field lines, which extend indefinitely to the right.

9. A point charge is located a fixed distance outside of a uniformly charged sphere. If the sphere shrinks in size without losing any charge, what happens to the force on the point charge?
10. The field of an infinite charged line decreases as $1/r$. Why isn't this a violation of the inverse-square law?
11. Why can't you use Gauss's law to determine the field of a uniformly charged cube? Why couldn't you use a cubical Gaussian surface?
12. You're sitting inside an uncharged, hollow spherical shell. Suddenly someone dumps a billion coulombs of charge on the shell, distributed uniformly. What happens to the electric field at your location?
13. Does Gauss's law apply to a spherical Gaussian surface not centered on a point charge, as shown in Fig. 21.29? Would this be a useful surface to use in calculating the electric field?

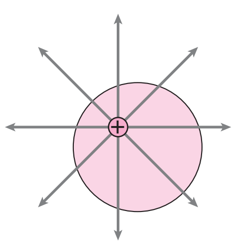


FIGURE 21.29 For Thought and Discussion 13

14. An insulating sphere carries charge spread uniformly throughout its volume. A conducting sphere has the same radius and net charge, but of course the charge is spread over its surface only. Compare the electric fields outside these two charge distributions.
15. Why must the electric field be zero inside a conductor in electrostatic equilibrium?
16. The electric field of a flat sheet of charge is $\sigma/2\epsilon_0$. Yet the field of a flat conducting sheet—even a thin one, like a piece of aluminum foil—is σ/ϵ_0. Explain this apparent discrepancy.

Exercises and Problems

Exercises

Section 21.1 Electric Field Lines

17. In Fig. 21.30, the magnitude of the middle charge is $3\ \mu C$. What's the net charge shown?

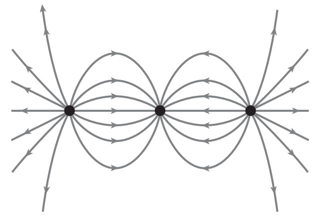


FIGURE 21.30 Exercise 17

18. Charges $+2q$ and $-q$ are near each other. Sketch some field lines for this charge distribution, using eight lines for a charge of magnitude q.
19. The net charge shown in Fig. 21.31 is $+Q$. Identify each of the charges A, B, and C shown.

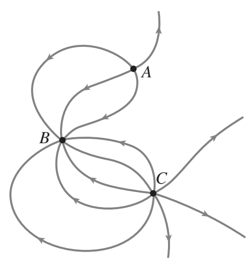


FIGURE 21.31 Exercise 19

Section 21.2 Electric Flux and Field

20. A flat surface with area $2.0\ m^2$ is in a uniform 850-N/C electric field. Find the electric flux through the surface when it's (a) at right angles to the field, (b) at 45° to the field, and (c) parallel to the field.
21. What's the electric field strength in a region where the flux through a $1.0\ cm \times 1.0\ cm$ flat surface is $65\ N{\cdot}m^2/C$, if the field is uniform and the surface is at right angles to the field?
22. A flat surface with area $0.14\ m^2$ lies in the x-y plane, in a uniform electric field $\vec{E} = 5.1\hat{\imath} + 2.1\hat{\jmath} + 3.5\hat{k}$ kN/C. Find the flux through the surface.
23. The electric field on the surface of a 10-cm-diameter sphere is perpendicular to the sphere and has magnitude 47 kN/C. What's the electric flux through the sphere?

Section 21.3 Gauss's Law

24. A sock comes out of the dryer with a trillion (10^{12}) excess electrons. What's the electric flux through a surface surrounding the sock?

25. What's the electric flux through the closed surfaces marked (a), (b), (c), and (d) in Fig. 21.32?

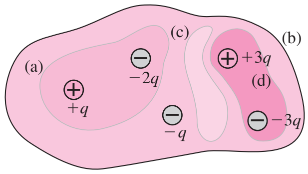


FIGURE 21.32 Exercise 25

26. A 6.8-μC charge and a -4.7-μC charge are inside an uncharged sphere. What's the electric flux through the sphere?
27. A 2.6-μC charge is at the center of a cube 7.5 cm on each side. What's the electric flux through one face of the cube? (*Hint:* Think about symmetry, and don't do an integral.)

Section 21.4 Using Gauss's Law

28. The electric field at the surface of a 5.0-cm-radius uniformly charged sphere is 90 kN/C. What's the field strength 10 cm from the surface?
29. A solid sphere 25 cm in radius carries 14 μC, distributed uniformly throughout its volume. Find the electric field strength (a) 15 cm, (b) 25 cm, and (c) 50 cm from its center.
30. A 10-nC point charge is located at the center of a thin spherical shell of radius 8.0 cm carrying -20 nC distributed uniformly over its surface. Find the magnitude and direction of the electric field (a) 2.0 cm, (b) 6.0 cm, and (c) 15 cm from the point charge.
31. The electric field strength outside a charge distribution and 18 cm from its center has magnitude 55 kN/C. At 23 cm the field strength is 43 kN/C. Does the distribution have spherical or line symmetry?
32. An electron close to a large, flat sheet of charge is repelled from the sheet with a 1.8-pN force. Find the surface charge density on the sheet.
33. Find the field produced by a uniformly charged sheet carrying 87 pC/m^2.
34. What surface charge density on an infinite sheet will produce a 1.4-kN/C electric field?

Section 21.5 Fields of Arbitrary Charge Distributions

35. A rod 50 cm long and 1.0 cm in radius carries a 2.0-μC charge distributed uniformly over its length. Find the approximate magnitude of the electric field (a) 4.0 mm from the rod surface, not near either end, and (b) 23 m from the rod.
36. What's the approximate field strength 1 cm above a sheet of paper carrying uniform surface charge density $\sigma = 45$ nC/m^2?
37. The disk in Fig. 21.20 has area 0.14 m^2 and is uniformly charged to 5.0 μC. Find the approximate field strength (a) 1 mm from the disk, not near the edge, and (b) 2.5 m from the disk.

Section 21.6 Gauss's Law and Conductors

38. What is the electric field strength just outside the surface of a conducting sphere carrying surface charge density 1.4 μC/m^2?
39. A net charge of 5.0 μC is applied on one side of a solid metal sphere 2.0 cm in diameter. Once electrostatic equilibrium is reached, and assuming no other conductors or charges nearby, what are (a) the volume charge density inside the sphere and (b) the surface charge density on the sphere?
40. A positive point charge q lies at the center of a spherical conducting shell carrying net charge $\frac{3}{2}q$. Sketch the field lines both inside and outside the shell, using eight field lines to represent a charge of magnitude q.
41. A total charge of 18 μC is applied to a thin, square metal plate 75 cm on a side. Find the electric field strength near the plate's surface.

Problems

42. What's the flux through the hemispherical open surface of radius R in a uniform field of magnitude E shown in Fig. 21.33? (*Hint:* Don't do a messy integral!)

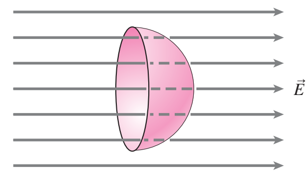


FIGURE 21.33 Problem 42

43. An electric field is given by $\vec{E} = E_0(y/a)\hat{k}$, where E_0 and a are constants. Find the flux through the square in the x-y plane bounded by the points $(0, 0)$, $(0, a)$, (a, a), $(a, 0)$.
44. The electric field in a certain region is given by $\vec{E} = ax\hat{\imath}$, where $a = 40$ N/C·m and x is in meters. Find the volume charge density in the region. (*Hint:* Apply Gauss's law to a cube 1 m on a side.)
45. BIO A study shows that mammalian red blood cells (RBCs) carry electric charge resulting from 4.4 million (rabbit cells) to 15 million (human cells) excess electrons spread over their surfaces. Approximating rabbit and human RBCs as spheres with radii 30 μm and 36 μm, respectively, find the electric field strengths at the cells' surfaces.
46. Positive charge is spread uniformly over the surface of a spherical balloon 70 cm in radius, resulting in an electric field of 26 kN/C at the balloon's surface. Find the field strength (a) 50 cm from the balloon's center and (b) 190 cm from the center. (c) What's the net charge on the balloon?
47. A solid sphere 2.0 cm in radius carries a uniform volume charge density. The electric field 1.0 cm from the sphere's center has magnitude 39 kN/C. (a) At what other distance does the field have this magnitude? (b) What's the net charge on the sphere?
48. A point charge of $-2Q$ is at the center of a spherical shell of radius R carrying charge Q spread uniformly over its surface. Find the electric field at (a) $r = \frac{1}{2}R$ and (b) $r = 2R$. (c) How would your answers change if the charge on the shell were doubled?
49. A friend is working on a biology experiment and needs to create an electric field of magnitude 430 N/C at 10 cm from the central portion of a large nonconducting square plate 4.5 m on each side. She needs to know how much charge to put on the plate. What do you tell her?
50. A spherical shell of radius 15 cm carries 4.8 μC distributed uniformly over its surface. At the center of the shell is a point charge. (a) If the electric field at the sphere's surface is 750 kN/C and points outward, what are (a) the point charge and (b) the field just inside the shell?
51. A spherical shell 30 cm in diameter carries 85 μC distributed uniformly over its surface. A 1.0-μC point charge is located at the shell's center. Find the electric field strength (a) 5.0 cm from the center and (b) 45 cm from the center. (c) How would your answers change if the charge on the shell were doubled?
52. A thick, spherical shell of inner radius a and outer radius b carries a uniform volume charge density ρ. Find an expression for the electric field strength in the region $a < r < b$, and show that your result is consistent with Equation 21.5 when $a = 0$.

53. A long, thin wire carrying 5.6 nC/m runs down the center of a long, thin-walled, pipe with radius 1.0 cm carrying -4.2 nC/m spread uniformly over its surface. Find the electric field (a) 0.50 cm from the wire and (b) 1.5 cm from the wire.
54. An infinitely long rod of radius R carries a uniform volume charge density ρ. Show that the electric field strengths outside and inside the rod are given, respectively, by $E = \rho R^2/2\epsilon_0 r$ and $E = \rho r/2\epsilon_0$, where r is the distance from the rod axis. (Although an infinite rod is an impossibility, your answer is a good approximation for the field of a finite rod whose length is much greater than its diameter.)
55. A long, solid rod 4.5 cm in radius carries a uniform volume charge density. If the electric field strength at the surface of the rod (not near either end) is 16 kN/C, what's the volume charge density?
56. If you "painted" positive charge on the floor, what surface charge density would be necessary to suspend a 15 μC, 5.0-g particle above the floor?
57. A charged slab extends infinitely in two dimensions and has thickness d in the third dimension, as shown in Fig. 21.34. The slab carries a uniform volume charge density ρ. Find expressions for the electric field strength (a) inside and (b) outside the slab, as functions of the distance x from the center plane. (Although the infinite slab is impossible, your answer is a good approximation to the field of a finite slab whose width is much greater than its thickness.)

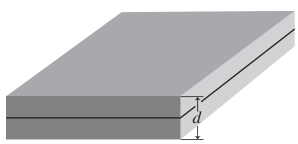


FIGURE 21.34 Problem 57

58. A solid sphere 10 cm in radius carries a 40-μC charge distributed uniformly throughout its volume. It's surrounded by a concentric shell 20 cm in radius, also uniformly charged with 40 μC. Find the electric field (a) 5.0 cm, (b) 15 cm, and (c) 30 cm from the center.
59. A nonconducting square plate 75 cm on a side carries a uniform surface charge density. The electric field strength 1 cm from the plate, not near an edge, is 45 kN/C. What's the approximate field strength 15 m from the plate?
60. A 250-nC point charge is placed at the center of an uncharged spherical conducting shell 20 cm in radius. Find (a) the surface charge density on the outer surface of the shell and (b) the electric field strength at the shell's outer surface.
61. An irregular conductor containing an irregular, empty cavity carries a net charge Q. (a) Show that the electric field inside the cavity must be zero. (b) If you put a point charge inside the cavity, what value must it have in order to make the charge density on the outer surface of the conductor everywhere zero?
62. You're an engineer for a cable TV company that delivers signals over coaxial cables consisting of an inner wire and a concentric cylindrical outer conductor. A new colleague in your department is worried that electric fields from charge on the outer conductor will interfere with other electrical signals. Formulate an argument to convince your colleague that, as long as the conductors carry equal but opposite charges, any electric field associated with the cable can't extend beyond the outer conductor.
63. A point charge $-q$ is at the center of a spherical shell carrying charge $+2q$. That shell, in turn, is concentric with a larger shell carrying $-\frac{3}{2}q$. Draw a cross section of this structure, and sketch the electric field lines using the convention that eight lines correspond to a charge of magnitude q.
64. A point charge q is at the center of a spherical shell of radius R carrying charge $2q$ spread uniformly over its surface. Write expressions for the electric field strength at (a) $\frac{1}{2}R$ and (b) $2R$.
65. The volume charge density inside a solid sphere of radius a is $\rho = \rho_0 r/a$, where ρ_0 is a constant. Find (a) the total charge and (b) the electric field strength within the sphere, as a function of distance r from the center.
66. Figure 21.35 shows a rectangular box with sides $2a$ and length L surrounding a line carrying uniform line charge density λ. The line passes directly through the center of the box faces. Integrate the field of the line charge over strips of width dx as shown to find the electric flux through one face of the box. Multiply by 4 to get the total flux, and show that your result is consistent with Gauss's law.

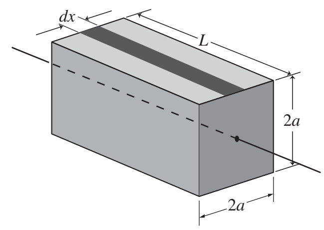


FIGURE 21.35 Problem 66

67. The charge density within a uniformly charged sphere of radius R is $\rho = \rho_0 - ar^2$, where ρ_0 and a are constants and r is the distance from the center. Find an expression for a such that the electric field outside the sphere is zero.
68. Calculate the electric fields in Example 21.2 directly, using the superposition principle and integration. Consider the shell to be composed of charge elements that are coaxial rings, whose axes pass through the field point, which is a distance r from the center. (*Hint*: Consult Example 20.6. You'll have to evaluate the cases $r < R$ and $r > R$ separately.)
69. A solid sphere of radius R carries volume charge density $\rho = \rho_0 e^{r/R}$, where ρ_0 is a constant and r is the distance from the center. Find an expression for the electric field strength at the sphere's surface.
70. Problem 76 of Chapter 13 explored what happened to a person falling into a hole extending all the way through Earth's center and out the other side, assuming that $g(r) = g_0(r/R_E)$ for points inside Earth $(r < R_E)$. Prove this assumption, treating Earth as a uniform sphere and using the gravitational version of Gauss's law: $\oint \vec{g} \cdot d\vec{A} = -4\pi G M_{\text{enclosed}}$.

Passage Problems

Coaxial cables are widely used with audio-visual technology, electronic instrumentation, and radio broadcasting, because they minimize interference with or from signals traveling on the cable. Coaxial cables consist of a wire inner conductor surrounded by a thin cylindrical conducting shield, usually of braided copper (Fig. 21.36). Flexible

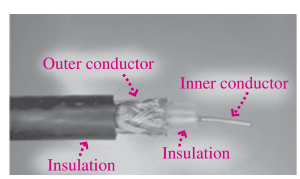


FIGURE 21.36 A coaxial cable (Passage Problems 71–74)

insulation separates the conductors. A straight length of coaxial cable can be approximated as an infinitely long wire surrounded by a cylindrical shell. Normally the two conductors carry charges of equal magnitude but opposite sign. (Charge actually varies with time and position as signals travel down the cable, but for these problems consider the charge to be fixed and spread uniformly.)

71. For a coaxial cable in electrostatic equilibrium carrying equal but opposite charges on its two conductors, there's a nonzero electric field
    a. only in the space between the wire and shield.
    b. in the space between wire and shield, and outside the shield.
    c. inside the metal conducting wire and shield, as well as between the wires and outside the shield.
    d. only outside the shield.

72. A coaxial cable carries equal but opposite charges on its two conductors. In electrostatic equilibrium, charge on the shield
    a. lies entirely on its outer surface.
    b. is divided evenly between inner and outer surfaces.
    c. lies entirely on its inner surface.
    d. distributes itself differently depending on the magnitude of the charge.

73. How does the electric field *between* the conductors in a coaxial cable in electrostatic equilibrium depend on the radial distance r from the cable's axis?
    a. it's constant
    b. as $1/r$
    c. as $1/r^2$
    d. as $1/r^3$

74. A coaxial cable in electrostatic equilibrium carries charge $-Q$ on its inner conductor and $+Q$ on its shield. If the charge on the shield *only* is doubled,
    a. the magnitude of the electric field between the conductors will double.
    b. the magnitude of the electric field outside the shield will double.
    c. the magnitude of the electric field at the outer surface of the shield will become twice the magnitude of the field at the shield's inner surface.
    d. the magnitude of the electric field at the outer surface of the shield will equal the magnitude of the field at the shield's inner surface.

Answers to Chapter Questions

Answer to Chapter Opening Question

Gauss's law requires that electric charge remain on the outside of the metal cage, arranging itself so there's no electric field within the cage.

Answers to GOT IT? Questions

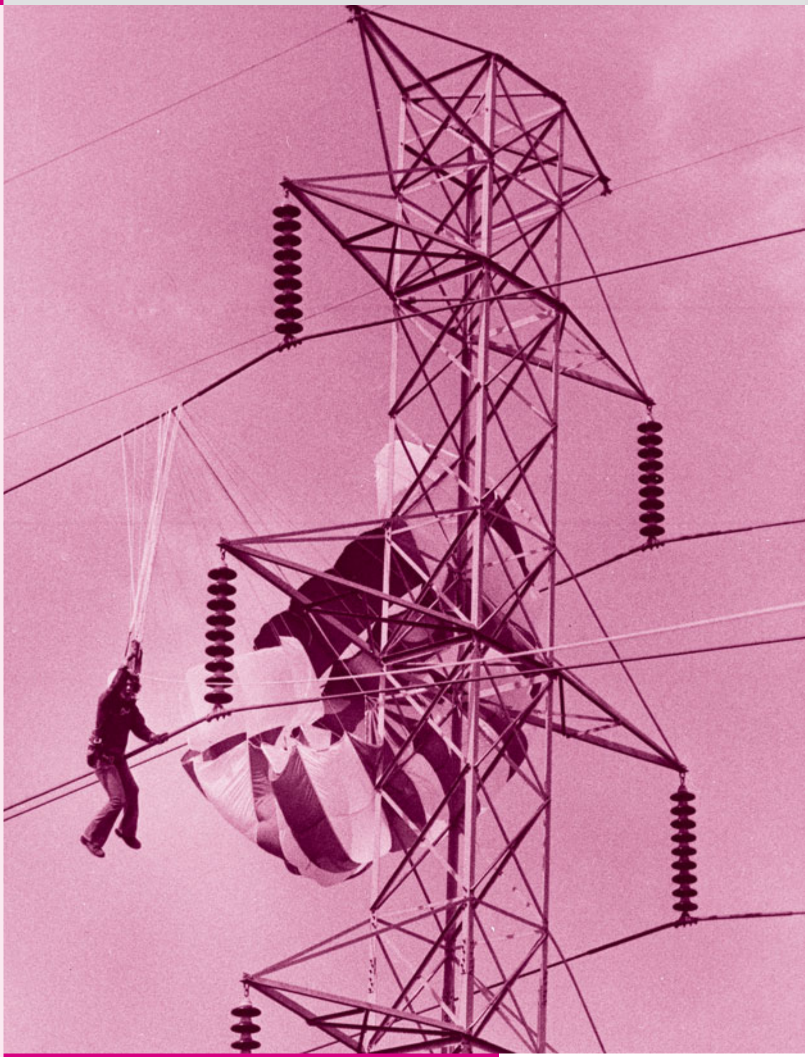
21.1. (a) $\Phi_A = 0, \Phi_B = 0, \Phi_C = s^2E$;
(b) $\Phi_A = 0, \Phi_B = \Phi_C = s^2E\cos 45° = s^2E/\sqrt{2}$

21.2. Flux: (a); doesn't change. Field: (d); increases if charges are opposite, decreases if same.

21.3. (a) The field stays zero; (b) the field kQ/r^2 doubles.

21.4. (c).

22 Electric Potential



This parasailer landed on a 138,000-volt power line. Why wasn't he electrocuted?

New Concepts, New Skills

By the end of this chapter you should be able to

- Explain the meaning of electric potential difference, including the familiar term "volt" (22.1).
- Calculate the potential difference between two points in simple electric fields using a line integral (22.1).
- Describe the potential in the field of a point charge, and use superposition to calculate potential differences by summing or integrating over point charges (22.2).
- Explain the concept of equipotentials, and determine electric fields from potential differences (22.3).
- Describe qualitatively how charge distributes itself on conductors (22.4).

Connecting Your Knowledge

- This chapter merges the concepts of work (6.1, 6.2), conservative forces (7.1), and potential energy (7.2) with that of the electric field (20.3).
- We build on results from Chapters 20 and 21, including the fields of a dipole (20.4), a spherical charge distribution (21.4), a line charge (20.4, 21.4), and a charged sheet (21.4).
- We further our understanding of charged conductors (21.6).
- We use Tactics 9.1 for handling integrals (9.1).

Like gravity, the electric force is conservative. That means the work done in moving a charge against an electric force results in stored potential energy. It's convenient to consider the energy per unit charge, a measure that defines the concept of electric potential. Here we'll see how potential provides a simpler approach to calculating electric fields and helps characterize everyday devices like batteries.

22.1 Electric Potential Difference

In Chapter 7 we defined the potential-energy difference ΔU_{AB} as the negative of the work W_{AB} done by a conservative force $\vec{F}$ on an object moved from point A to point B (Equation 7.2):

$$\Delta U_{AB} = U_B - U_A = -W_{AB} = -\int_A^B \vec{F} \cdot d\hat{r}$$

where $d\hat{r}$ is an element of the path from A to B, and ΔU_{AB} is independent of the path taken from A to B. When the force doesn't vary, we can calculate the work more easily using Equation 6.5, $W = \vec{F} \cdot \Delta\vec{r}$.

Consider a positive charge q moved between points A and B a distance Δr apart in a uniform electric field $\vec{E}$, as shown in Fig. 22.1. Since the field is uniform, a constant electric force $\vec{F} = q\vec{E}$ acts on the charge, so we use Equation 6.5 to evaluate the work done by the field and the resulting potential-energy change:

$$\Delta U_{AB} = -W_{AB} = -q\vec{E}\cdot\Delta\vec{r} = -qE\,\Delta r\cos 180° = qE\,\Delta r$$

where the factor $\cos 180° = -1$ appears because $\vec{E}$ and $\Delta\vec{r}$ have opposite directions. Make sense? Pushing a positive charge from A to B against the electric field is like pushing a car up a hill: Potential energy increases in both cases. Let go of the charge, and the field accelerates it back, just as gravity would accelerate the car back down the hill.

Had we moved a charge $2q$ in Fig. 22.1, the potential-energy change ΔU would have been twice as great; a charge $\frac{1}{2}q$ would have cut ΔU in half. Since ΔU is proportional to charge, it's convenient to consider the *potential-energy change per unit charge* involved in moving a charge between two points. Mathematically, we write $\vec{F} = q\vec{E}$ in our general expression for ΔU_{AB} and divide by q. The result defines the **electric potential difference** ΔV:

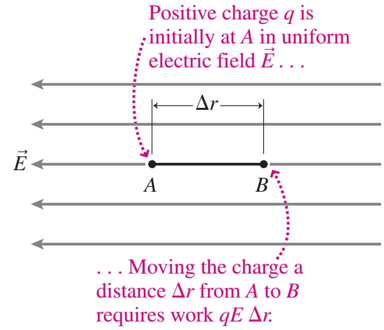


FIGURE 22.1 Work done in moving a charge q against an electric field $\vec{E}$.

> The electric potential difference from point A to point B is the potential-energy change per unit charge in moving a charge from A to B:
>
> $$\Delta V_{AB} = \frac{\Delta U_{AB}}{q} = -\int_A^B \vec{E}\cdot d\vec{r} \qquad \text{(electric potential difference)} \qquad (22.1a)$$

Here Δ and the subscripts AB show explicitly that we're talking about a *change* or *difference* from one point to another. We'll sometimes use just the symbol V for potential difference, in cases where the starting point A is understood. Note that potential difference, although computed from vectors, is itself a scalar quantity.

The switch from potential energy to electric potential is analogous to Chapter 20's introduction of the electric field as the electric force *per unit charge*; similarly, the electric potential difference is the change in potential energy *per unit charge*. The reason is the same: We want to express electric properties in terms that don't involve specific charges. Table 22.1 summarizes the relations among force and field, potential energy and electric potential.

In the special case of a uniform field, Equation 22.1a reduces to

$$\Delta V_{AB} = -\vec{E}\cdot\Delta\vec{r} \qquad \text{(uniform field)} \qquad (22.1b)$$

where $\Delta\vec{r}$ is a vector from A to B. Figure 22.1 shows the special case when the field $\vec{E}$ and path $\Delta\vec{r}$ are in opposite directions; here, Equation 22.1b gives $\Delta V_{AB} = E\,\Delta r$.

Potential difference can be positive or negative, depending on whether the path goes against or with the field. Moving a positive charge through a positive potential difference is like going uphill: Potential energy increases. Moving a positive charge through a negative potential difference is like going downhill: Potential energy decreases. The converse is true for a negative charge; even though the potential difference remains the same, the force is opposite and so the potential energy reverses sign.

We emphasize that potential difference is a property *of two points*; it doesn't depend on the path between those points. In Fig. 22.1, considering a straight path from A to B made the calculation of potential difference easy, but we would have found the same result—albeit with much more effort—using any path (Fig. 22.2).

Table 22.1 Force and Field, Potential Energy and Electric Potential

Quantity	**Symbol/Equation**	**Units**
Force	$\vec{F}$	N
Electric field	$\vec{E} = \dfrac{\vec{F}}{q}$	N/C or V/m
Potential-energy difference	$\Delta U = -\int_A^B \vec{F}\cdot d\vec{r}$	J
Electric potential difference	$\Delta V = \dfrac{\Delta U}{q}$ or $\Delta V = -\int_A^B \vec{E}\cdot d\vec{r}$	J/C or V

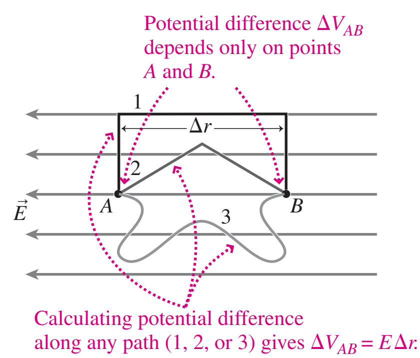


FIGURE 22.2 Potential difference is path independent.

GOT IT? 22.1 What would happen to the potential difference V_{AB} in Fig. 22.1 if (a) the electric field strength were doubled; (b) the distance Δr were doubled; (c) the points were moved so the path lay at right angles to the field; and (d) the positions of A and B were interchanged?

The Volt and the Electronvolt

Potential difference measures work or energy per unit charge, so its units are joules per coulomb, as shown in Table 22.1. Potential difference is important enough that this unit has a special name, the **volt** (V). To say that a car has a 12-V battery, for example, means the battery does 12 J of work on every coulomb of charge that moves between its terminals. Multiplying the first equality in Equation 22.1a by q shows that the change in potential energy of a charge q as it moves through a potential difference ΔV is $\Delta U = q\,\Delta V$. If a charge q "falls" freely through a potential difference ΔV, it therefore gains kinetic energy given by $|q\,\Delta V|$.

We often use the term **voltage** to mean potential difference, especially in electric circuits. The two are subtly different, however, when changing magnetic fields are present; more on this in Chapter 27. Table 22.2 lists some typical potential differences in technological and natural systems.

Table 22.2 Typical Potential Differences

Between human arm and leg due to heart's electrical activity	1 mV
Across biological cell membrane	80 mV
Between terminals of flashlight battery	1.5 V
Car battery	12 V
Electric outlet (depends on country)	100–240 V
Between long-distance electric transmission line and ground	365 kV
Between base of thunderstorm cloud and ground	100 MV

✓TIP Potential Difference Involves Two Points

Potential difference is the energy per unit charge involved in moving *between those points*. This is ultimately a practical matter; if you forget it, you won't be able to hook up a voltmeter properly or connect jumper cables safely to your car battery! This chapter's opening photo provides a dramatic illustration of this point.

Sometimes we say "the potential (or voltage) at point P." This is *always* a shorthand way of talking, and we *must* have in mind some other point. What we mean is the potential difference going from that other point to P.

In molecular, atomic, and nuclear systems it's often convenient to measure energy in **electronvolts** (eV), defined as **the energy gained by a particle carrying one elementary charge when it moves through a potential difference of 1 volt**. Since one elementary charge is 1.6×10^{-19} C, 1 eV is 1.6×10^{-19} J. Energy in eV is particularly easy to calculate when charge is given in units of the elementary charge e; then, with ΔV in volts, $q\,\Delta V$ gives the energy in eV. However, the eV is *not* an SI unit and should be converted to joules before calculating other quantities, like velocity.

GOT IT? 22.2 (a) A proton (charge e), (b) an alpha particle (charge $2e$), and (c) a singly ionized oxygen atom each move through a 10-V potential difference. What's the work in eV done on each?

EXAMPLE 22.1 Potential Difference, Work, and Energy: X Rays

In an X-ray tube, a uniform electric field of 300 kN/C extends over a distance of 10 cm, from an electron source to a target; the field points from the target to the source. Find the potential difference between source and target and the energy gained by an electron as it accelerates from source to target (where its abrupt deceleration produces X rays). Express the energy in both electronvolts and joules.

INTERPRET This problem requires first calculating the potential difference from the field and then the energy from the potential difference.

DEVELOP Figure 22.3 is our drawing with point A the source and point B the target. Equation 22.1b, $\Delta V_{AB} = -\vec{E}\cdot\Delta\vec{r}$, determines the potential difference in this uniform field. Given the potential difference, or energy per unit charge, we can find the energy gain from the magnitude of the product $q\,\Delta V$.

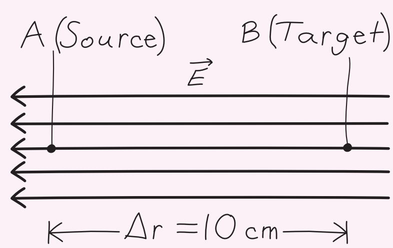


FIGURE 22.3 Sketch for Example 22.1.

EVALUATE With the field and path in opposite directions, $\cos\theta = -1$ in the dot product, so Equation 22.1b gives

$$\Delta V_{AB} = E\,\Delta r = (300\text{ kN/C})(0.10\text{ m}) = 30\text{ kV}$$

Although this difference is positive, a *negative* electron moves "downhill" from source to target, and thus gains kinetic energy as work gets done on it. With the charge measured in elementary charges, the product $|q\,\Delta V|$ gives this energy directly in electronvolts: $(1\text{ elementary charge } e)(30\text{ kV}) = 30$ keV. With 1.6×10^{-19} J/eV, this is 4.8 fJ.

ASSESS Make sense? An electronvolt is a lot smaller than a joule, so the SI answer (in fJ $= 10^{-15}$ J) is numerically much smaller. ■

EXAMPLE 22.2 Potential of a Charged Sheet

An isolated, infinite charged sheet carries uniform surface charge density σ. Find an expression for the potential difference from the sheet to a point a perpendicular distance x from the sheet.

INTERPRET This is a question about calculating the potential difference from the field.

DEVELOP The result of Example 21.6 gives the field of a charged sheet: It's uniform, with magnitude $E = \sigma/2\epsilon_0$ and direction perpendicular to the sheet. We've drawn the sheet and a few of its field lines in Fig. 22.4. Since the field is uniform, Equation 21.1b, $\Delta V_{AB} = -\vec{E}\cdot\Delta\vec{r}$, determines the potential difference.

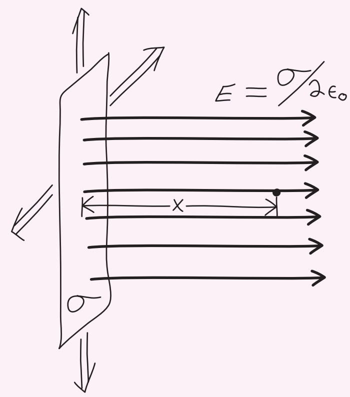


FIGURE 22.4 Sketch for Example 22.2. The field also extends to the left from the sheet, but we haven't drawn that.

EVALUATE Moving away from the sheet means going in the direction of the field (assuming positive σ), so $\cos\theta = 1$ in the dot product, and we evaluate to get

$$V_{0x} = -Ex = -\frac{\sigma x}{2\epsilon_0}$$

Here we've used x for the displacement Δr and V_{0x} for the potential difference because we're measuring from the sheet $(x = 0)$ to the point x.

ASSESS Make sense? Our result shows that the potential difference in a *uniform* field varies *linearly* with distance. Moving a positive charge away from the sheet is like going "downhill," in this case with a constant slope. Give the sheet a negative charge $(\sigma < 0)$ and the potential difference changes sign; now moving a positive charge away from the sheet is going "uphill." (And moving a negative charge away is "downhill"—since like charges repel.) ■

Curved Paths and Nonuniform Fields

If the electric field isn't uniform or the path isn't straight, then we need to use the integral in Equation 22.1a to find the potential difference. With that equation we're dividing the path into segments $d\vec{r}$, each so short that it's essentially straight with a uniform field over its length (Fig. 22.5). Then Equation 22.1b gives the potential difference $dV = -\vec{E}\cdot d\vec{r}$ across the segment, and in integrating we're summing infinitely many infinitesimal dV values to get the potential difference between two points A and B. We'll see some examples in the next section.

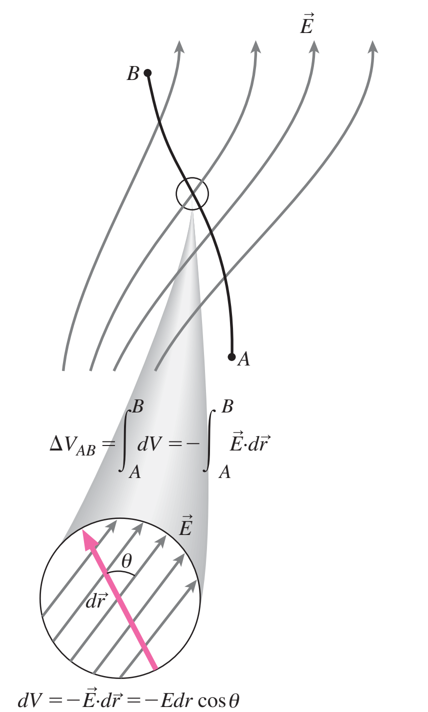
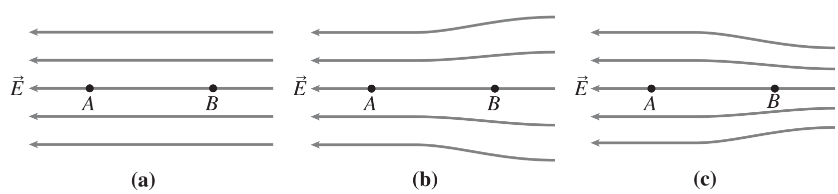


FIGURE 22.5 The integral in Equation 22.1a is the sum of infinitely many infinitesimally small potential differences dV.

GOT IT? 22.3 The figure shows three straight paths AB of the same length, each in a different electric field. The field at A is the same in each. Rank the potential differences ΔV_{AB}.



22.2 Calculating Potential Difference

Here we use Equation 22.1a to calculate the potential differences for several charge distributions. Most important is the point charge, which then provides an easy way to find potential differences for more complicated charge distributions.

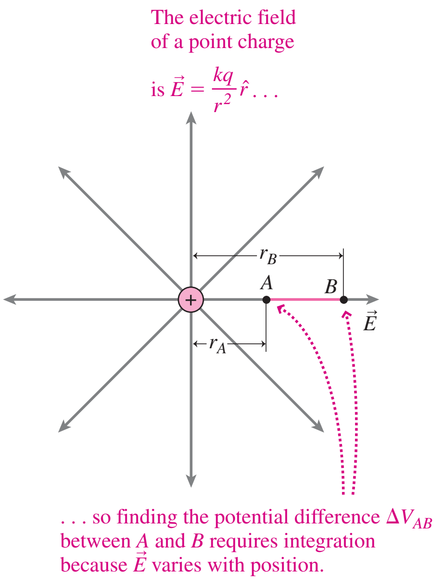


FIGURE 22.6 Potential difference in the field of a point charge.

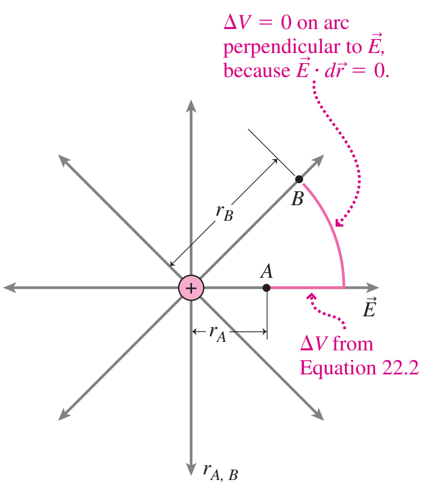


FIGURE 22.7 Potential difference is path independent, so ΔV_{AB} here still follows from Equation 22.2.

The Potential of a Point Charge

Equation 20.3 gives the electric field of a point charge: $\vec{E} = (kq/r^2)\hat{r}$. Let's find the potential difference between two points A and B at distances r_A and r_B from a positive point charge, as shown in Fig. 22.6. We can't just multiply the distance $r_B - r_A$ by E because the field varies with position. Instead we integrate, using Equation 22.1a:

$$\Delta V_{AB} = V(B) - V(A) = -\int_{r_A}^{r_B} \vec{E} \cdot d\vec{r} = -\int_{r_A}^{r_B} \frac{kq}{r^2}\hat{r} \cdot d\vec{r}$$

As we move from A to B, the path elements are increments dr in the radial direction, so we write $d\vec{r} = \hat{r}\,dr$. Then the potential difference is

$$\Delta V_{AB} = -\int_{r_A}^{r_B} \frac{kq}{r^2}\hat{r} \cdot \hat{r}\,dr = -kq\int_{r_A}^{r_B} r^{-2}\,dr$$

since the dot product of the unit vector $\hat{r}$ with itself is 1. Evaluating the integral gives

$$\Delta V_{AB} = -kq\left[-\frac{1}{r}\right]_{r_A}^{r_B} = kq\left(\frac{1}{r_B} - \frac{1}{r_A}\right) \qquad (22.2)$$

Make sense? For $r_B > r_A$, the potential difference is negative, showing that a positive test charge at r_A would "fall" toward r_B. Going the other way would require that work be done on a positive charge, as it's pushed "up" against the repulsive force of the charge q. Our result holds as well for $q < 0$, in which case the sign of the potential difference changes.

Although we derived Equation 22.2 for two points on the same radial line, Fig. 22.7 shows that the result holds for *any* two points in the field of a charge q. It doesn't matter which point is at the greater distance either; if $r_B < r_A$, Equation 22.2 still gives the correct potential difference, which then becomes positive, showing that we do work to move a positive test charge *toward* a positive q.

The Zero of Potential

Only potential differences have physical significance. But it's often convenient to define a point of zero potential; we can then speak of the potential V at some other point P, meaning the potential difference between our zero point and P. In this context the expression ΔV_{AB} can be written $V(B) - V(A)$. The choice for the zero of potential is usually based on mathematical or physical convenience. In electric power systems, Earth, called "ground," is usually taken as the zero of potential; in automobile electric systems, the car's metal frame is a convenient zero point.

When we deal with isolated charges, it's convenient to take the zero of potential at infinity. Then $r_A \to \infty$ in Equation 22.2 and $1/r_A$ becomes zero. We'll omit the subscript on r_B because it can be any point; then Equation 22.2 becomes

$$V_{\infty r} = V(r) = \frac{kq}{r} \quad \text{(point-charge potential)} \qquad (22.3)$$

When we call this expression $V(r)$ "the potential of a point charge," we really mean that $V(r)$ is the potential difference going from a point very far from a charge q to a point a distance r from the charge—an interpretation that's consistent with our definition of potential difference as depending on *two* points. Because the field outside any spherically symmetric charge distribution is that of a point charge, Equation 22.3 also gives the potential outside a spherically symmetric charge distribution.

Does it bother you that potential difference can be finite over an infinite distance? The reason lies in the inverse-square dependence of the field, which drops so rapidly that the work done in moving a charge from infinity to the vicinity of a point charge remains finite. We found an analogous result in Chapter 8, where it took only a finite amount of energy to escape completely from a planet's gravitational attraction. As long as a charge distribution is finite in size—so its field at large distances falls at least as fast as $1/r^2$—it makes sense to take the zero of potential at infinity.

GOT IT? 22.4 You measure a potential difference of 50 V between two points a distance 10 cm apart in the field of a point charge. If you move closer to the charge and measure the potential difference over another 10-cm interval, will it be (a) greater, (b) less, or (c) the same?

EXAMPLE 22.3 Potential and Work: At the Science Museum

The Hall of Electricity at the Boston Museum of Science contains a large Van de Graaff generator, a device that builds up charge on a metal sphere (see Chapter 21's opening photo). The sphere has radius $R = 2.30$ m and develops a charge $Q = 640\ \mu\text{C}$. Considering this to be a single isolated sphere, find (a) the potential at its surface, (b) the work needed to bring a proton from infinity to the sphere's surface, and (c) the potential difference between the sphere's surface and a point $2R$ from its center.

INTERPRET This problem is about potential differences in the field of a spherically symmetric charge distribution. In Chapter 21 we found that the field outside such a distribution is identical to that of a point charge. The term "potential" is meaningless unless we're talking about two points, so here, with a point-charge field, we interpret the question as asking us to take the zero of potential at infinity.

DEVELOP Because the field outside the spherical charge distribution is the same as that of a point charge, Equation 22.3, $V(r) = kQ/r$, determines the potential for $r \geq R$. We've sketched this $1/r$ potential curve in Fig. 22.8. Because the zero of potential is at infinity, we can multiply the potential at the surface by the proton's charge to get the work required to bring a proton from infinity. Finally, we can evaluate the potential difference ΔV_{R2R} from the potentials at R and $2R$.

EVALUATE (a) Equation 22.3 gives

$$V(R) = \frac{kQ}{R} = 2.50\ \text{MV}$$

using Q and R given for the museum's device. (b) This 2.5-MV result is the potential difference between infinity and the sphere's surface. Then the work needed to move a proton—1 elementary charge e—from infinity is 2.5 MeV or 4.0×10^{-13} J. (c) We find the potential difference from the surface to $2R$ by subtracting the potentials at the two points:

$$\Delta V_{R2R} = V(2R) - V(R) = \frac{kQ}{2R} - \frac{kQ}{R} = -\frac{kQ}{2R} = -1.25\ \text{MV}$$

ASSESS Make sense? The potential difference ΔV_{R2R} is negative because we're moving away from the positively charged sphere. Our result also shows that fully half the potential difference between the sphere and infinity occurs within one radius of the sphere's surface—a consequence of the rapid $1/r^2$ decrease in the field. ■

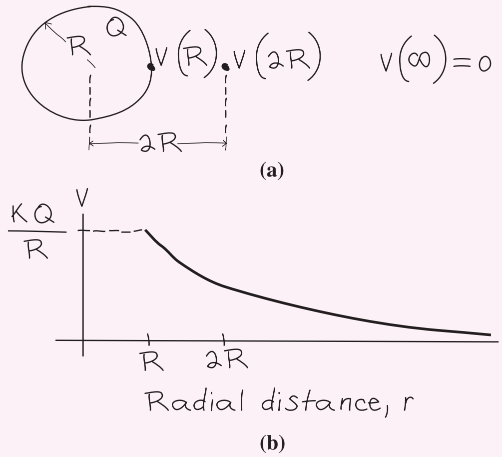


FIGURE 22.8 Sketch for Example 22.3.

EXAMPLE 22.4 Potential Difference: A High-Voltage Power Line

A long, straight power-line wire has radius 1.0 cm and carries line charge density $\lambda = 2.6\ \mu\text{C/m}$. Assuming no other charges are present, what's the potential difference between the wire and the ground, 22 m below?

INTERPRET We can interpret the long, straight wire as essentially an infinitely long charge distribution with line symmetry.

DEVELOP In Chapter 21 we found that the field outside any line-symmetric distribution is that of a line charge, $\vec{E} = (\lambda/2\pi\epsilon_0 r)\hat{r}$, so this equation determines the power line's field. We haven't been given any explicit expression for potential differences in this field, so because the field varies with position, our plan is to apply Equation 22.1a, $\Delta V = -\int \vec{E}\cdot d\vec{r}$. We've drawn the situation in Fig. 22.9.

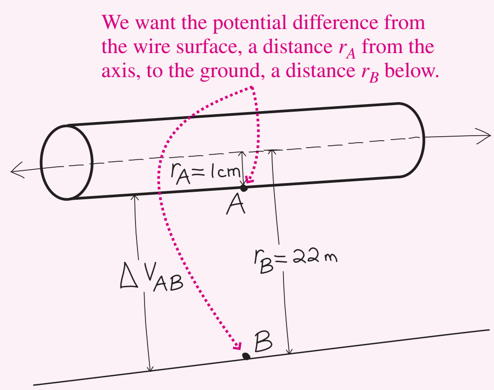


FIGURE 22.9 A long, straight power line approximated as an infinite charged rod whose field is that of a line charge.

(continued)

EVALUATE We evaluate the integral in Equation 22.1a over a straight path perpendicular to the wire, from its surface at r_A to the ground at r_B:

$$\begin{aligned}\Delta V_{AB} &= -\int_{r_A}^{r_B} \vec{E}\cdot d\vec{r} = -\int_{r_A}^{r_B} \frac{\lambda}{2\pi\epsilon_0 r}\hat{r}\cdot\hat{r}\,dr \\ &= -\frac{\lambda}{2\pi\epsilon_0}\int_{r_A}^{r_B}\frac{dr}{r} = -\frac{\lambda}{2\pi\epsilon_0}\ln r\Big|_{r_A}^{r_B} \\ &= \frac{\lambda}{2\pi\epsilon_0}\ln\left(\frac{r_A}{r_B}\right)\end{aligned} \qquad (22.4)$$

where the last step follows because $\ln x - \ln y = \ln(x/y)$. The numbers of this example give $\Delta V = -360\,\text{kV}$, a value typical of long-distance electric power transmission lines.

ASSESS Make sense? Our result is negative because the path AB goes away from a positive charge. (Mathematically, $r_A < r_B$ so the logarithm is negative.) The symbolic form of our answer shows that we can't let r_B go to infinity. Physically, that's because we're assuming the charge distribution has infinite extent, so it never resembles a point charge no matter how far away we get; mathematically, it's because the field falls off more slowly than a point-charge field—namely, as $1/r$. In practice, our answer here should be modified to account for the presence of other wires and of charges drawn to the ground surface. ■

Finding Potential Differences Using Superposition

When we don't know the field of a charge distribution, or when the field is too complicated to integrate easily, we can find the potential using superposition. This often provides an easier approach to calculating the field, as we'll see in Sec. 22.3.

Consider a charge q brought from infinity to a point P in the vicinity of some other charges. The superposition principle states that the electric field of a charge distribution is the sum of the fields of the individual charges that make up the distribution. Therefore, the work per unit charge—that is, the potential difference—between infinity and P is the sum of the potential differences associated with the individual point charges. Mathematically, we find $V(P)$ by summing Equation 22.3 over the individual point charges q_i:

$$V(P) = \sum_i \frac{kq_i}{r_i} \qquad (22.5)$$

where the r_i's are the distances from each of the charges to the point P. Equation 22.5 has one enormous advantage over its counterpart for the electric field, Equation 20.4. Electric potential is a *scalar*, so the sum in Equation 22.5 is a scalar sum, with no angles, vector components, or unit vectors.

EXAMPLE 22.5 Discrete Charges: The Dipole Potential

An electric dipole consists of point charges $\pm q$ a distance $2a$ apart. Find the potential at an arbitrary point P, and approximate for the case where the distance to P is large compared with the charge separation.

INTERPRET We have two point charges, so this problem is based on the point-charge potential, and therefore we'll take the zero of potential at infinity.

DEVELOP Figure 22.10 is our drawing, showing the distances from the two charges to a point P. Our plan is to apply superposition, summing the potentials of the individual point charges at P as determined in Equation 22.5, $V(P) = \sum(kq/r)$.

EVALUATE Applying Equation 22.5 gives

$$V(P) = \frac{kq}{r_1} + \frac{k(-q)}{r_2} = \frac{kq(r_2 - r_1)}{r_1 r_2}$$

This is an exact result valid for any P. We're also asked for an approximation for large distances. If r is the distance to the dipole center, as shown in Fig. 22.10, then for $r \gg a$, the quantities r_1, r_2, and r are nearly the same and the term $r_1 r_2$ is very nearly r^2. We have to be a little more careful with the term $r_1 - r_2$ because here we're comparing nearly equal quantities. Figure 22.10 shows that this

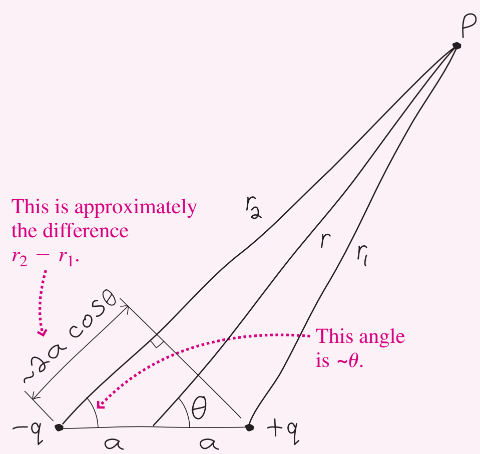


FIGURE 22.10 Finding the dipole potential.

term—the difference between the distances from the two charges to P—is approximately $2a\cos\theta$. So the dipole potential for $r \gg a$ becomes

$$V(r, \theta) = \frac{k(2aq)\cos\theta}{r^2} = \frac{kp\cos\theta}{r^2} \qquad \text{(dipole potential)} \qquad (22.6)$$

with $p = 2aq$ the dipole moment.

ASSESS Make sense? The dipole *potential* drops as $1/r^2$; earlier, we found the dipole *field* dropping as $1/r^3$. The difference of one power in r occurs because the potential results from integrating the field over distance. The same is true for the point charge, whose *field* drops as $1/r^2$ while its *potential* drops as $1/r$. Note also that Equation 22.6 gives $V = 0$ when $\theta = 90°$. There, on the dipole's perpendicular bisector, a charge brought from infinity is always moving at right angles to the dipole field (recall Fig. 21.2), so no work is involved (Fig. 22.11). ■

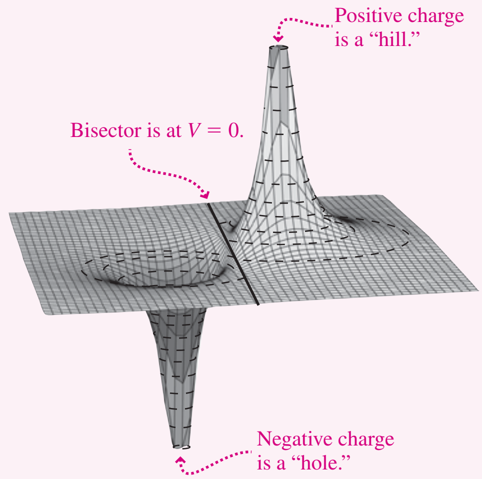
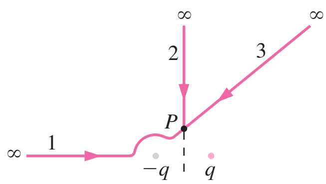


FIGURE 22.11 3-D plot of the dipole potential in the plane of Fig. 22.10.

GOT IT? 22.5 The figure shows three paths from infinity to a point P on a dipole's perpendicular bisector. Compare the work done in moving a charge to P on each of the paths.



Continuous Charge Distributions

We can calculate the potential of a continuous charge distribution by considering it to be made up of infinitely many infinitesimal charge elements dq. Each acts like a point charge and therefore contributes to the potential at some point P an amount dV given by $dV = k\,dq/r$, where the zero of potential is at infinity. The potential at P is the sum—that is, the integral—of the contributions dV from all the charge elements:

$$V = \int dV = \int \frac{k\,dq}{r} \quad \left(\begin{array}{c}\text{potential of a continuous}\\ \text{charge distribution}\end{array}\right) \qquad (22.7)$$

where the integration is over the entire charge distribution.

EXAMPLE 22.6 Potential of a Continuous Distribution: A Charged Ring

A total charge Q is distributed uniformly around a thin ring of radius a. Find the potential on the ring's axis.

INTERPRET We interpret the ring as a continuous charge distribution.

DEVELOP Equation 22.7, $V = \int k\,dq/r$, gives the potential for continuous charge distributions. Figure 22.12 is our drawing, showing an x-axis coincident with the ring axis, with $x = 0$ at the ring center. Charge elements dq in this case are small segments of the ring, and Fig. 22.12 shows that the distance $r = \sqrt{x^2 + a^2}$ is the same for all charge elements.

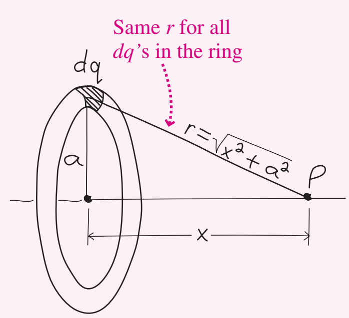


FIGURE 22.12 A charged ring.

(continued)

EVALUATE Equation 22.7 becomes

$$V(x) = \int \frac{k\,dq}{r} = \frac{k}{r}\int dq = \frac{kQ}{r} = \frac{kQ}{\sqrt{x^2 + a^2}} \qquad (22.8)$$

The integration here simplified because r is the same for all charge elements within the ring and so comes outside the integral; the remaining integral, $\int dq$, is the total charge Q. Example 22.7 will present a more typical—and more challenging—integral.

ASSESS Make sense? At large distances $(x \gg a)$, a^2 is negligible and our result becomes $V(x) = kQ/x$. This is the potential of a point charge Q—just as we'd expect when we're so far from the ring that its size isn't significant. At the ring's center, on the other hand, $V(0) = kQ/a$. Here we're a distance a from all parts of the ring, and since potential is a *scalar*, the different directions don't matter. The result is therefore the same as being a distance a from a point charge Q. ■

EXAMPLE 22.7 Potential of a Continuous Distribution: A Charged Disk

A charged disk of radius a carries a charge Q distributed uniformly over its surface. Find the potential at a point P on the disk axis, a distance x from the disk.

INTERPRET This problem, too, involves a continuous charge distribution.

DEVELOP Equation 22.7, $V = \int k\,dq/r$, determines the potential. But now all parts of the charge distribution aren't the same distance from P, so we have to set up the integral using the procedure first outlined in Chapter 9 and used more recently in Chapter 20 when we calculated the fields of continuous charge distributions. We've drawn the disk and its axis in Fig. 22.13. The preceding example suggests that we divide the disk into ring-shaped charge elements, as we've drawn in the figure. Each ring contributes a potential dV given by Equation 22.8: $dV = k\,dq/\sqrt{x^2 + r^2}$, where r is the ring radius. We get the potential of the entire disk by integrating over all the rings that make up the disk:

$$V(x) = \int_{\text{disk}} dV = \int_{r=0}^{r=a} \frac{k\,dq}{\sqrt{x^2 + r^2}}$$

Before we can evaluate this integral, we need to relate the charge element dq and the geometric variable r. Here the relation involves area: The ratio of ring area to disk area is the same as the ratio of dq to the total charge Q. "Unwinding" a ring gives a rectangle of length $2\pi r$ and width dr, so the ring area is $2\pi r\,dr$. The disk area is πa^2, so $dq/Q = 2\pi r\,dr/\pi a^2$, giving $dq = (2Q/a^2)r\,dr$.

EVALUATE Using this result in our integral for the potential $V(x)$ gives

$$V(x) = \int_0^a \frac{2kQ}{a^2} \frac{r\,dr}{\sqrt{x^2 + r^2}} = \frac{kQ}{a^2}\int_0^a \frac{2r\,dr}{\sqrt{x^2 + r^2}}$$

Now, $2r\,dr = d(r^2) = d(x^2 + r^2)$ since x is a constant with respect to the integration. The integral therefore has the form $\int u^{-1/2}\,du$, where $u = x^2 + r^2$, and the result is $2u^{1/2}$ or

$$V(x) = \frac{2kQ}{a^2}\sqrt{x^2 + r^2}\,\Bigg|_{r=0}^{r=a} = \frac{2kQ}{a^2}\left(\sqrt{x^2 + a^2} - |x|\right)$$

ASSESS Make sense? Figure 22.14 shows that it does. Close to the disk, the potential resembles that of an infinite sheet, changing linearly with distance (recall Example 22.2); far away, it has the $1/r$ behavior of a point-charge potential. It's only at intermediate distances—on the order of the disk radius a—that we really need our exact expression. ■

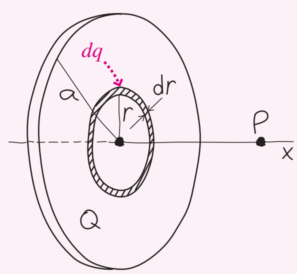


FIGURE 22.13 A charged disk, showing a ring-shaped charge element dq of radius r and width dr.

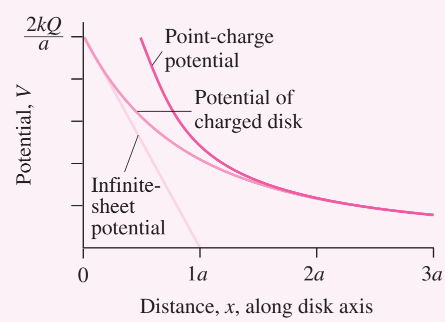


FIGURE 22.14 Charged-disk potential approaches that of an infinite sheet for points close to the disk, and that of a point charge far from the disk.

22.3 Potential Difference and the Electric Field

It takes no work to move a charge at right angles to an electric field, so there's no potential difference between two points on a surface perpendicular to the field. Such surfaces are called **equipotentials**. Equipotentials are like contour lines on a topographic map (Fig. 22.15). A contour is a line of constant elevation, so it takes no work to move along it. Where contours are close, elevation changes rapidly. Similarly, closely spaced equipotentials indicate a large potential difference between nearby points. That means there must be a strong electric field present. Figure 22.15 might just as well represent electric potential, with closely spaced equipotentials—steep slopes on the "potential hill"—representing strong electric fields. Similarly, the equipotentials for a dipole (Fig. 22.16) describe the steep "hill" of the positive charge and the "hole" of the negative charge that we showed in Fig. 22.11. There is one difference, though: Equipotentials are surfaces in three dimensions, and when we draw them as contour lines, we're showing only the surfaces' intersections with a plane.

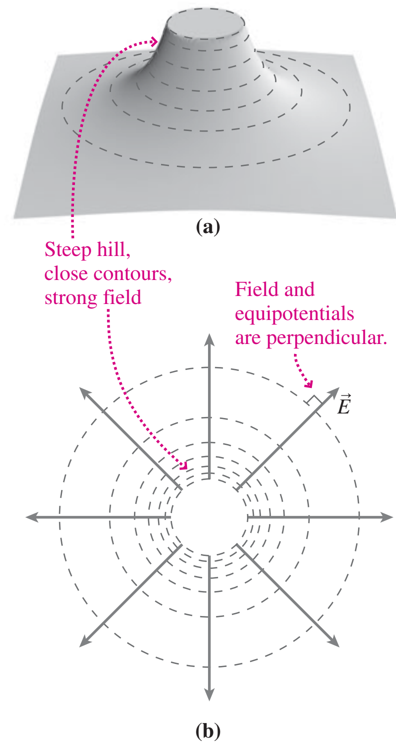
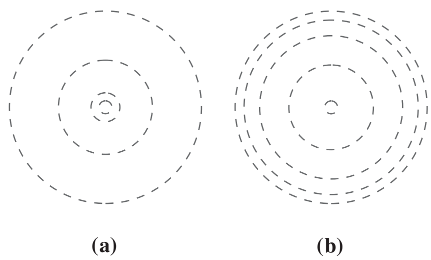


FIGURE 22.15 A flat-topped hill (a) and its contour map (b) represent equipotentials (dashed curves) for a charged spherical shell, in a plane through the shell's center.

GOT IT? 22.6 The figure shows cross sections through two equipotential surfaces. In both diagrams the potential difference between adjacent equipotentials is the same. Which could represent the field of a point charge? Explain.



Calculating Field from Potential

Given electric field lines, we can construct equipotentials. Conversely, given equipotentials, we can reconstruct the field by sketching field lines at right angles to the equipotentials. Specifying the potential at each point thus conveys all the information needed to determine the field.

We can quantify the relation between potential and field by considering the potential difference dV between two nearby points. Suppose they're separated by a small displacement dx in the x-direction. Then Equation 22.1b becomes $dV = -E_x\,dx$, where we handled the dot product by considering only the component of $\vec{E}$ along the displacement. Dividing through by dx shows that we can write the electric-field component in the x-direction as $E_x = -dV/dx$. We can write similar expressions for the y- and z-components. When a function depends on more than one variable, as the potential generally does, we write derivatives with the partial derivative symbol ∂ instead of d to indicate the rate of change with respect to only one variable. Thus we have $E_x = -\partial V/\partial x$, $E_y = -\partial V/\partial y$, and $E_z = -\partial V/\partial z$. Putting together these three components lets us write the entire electric-field vector:

$$\vec{E} = -\left(\frac{\partial V}{\partial x}\hat{\imath} + \frac{\partial V}{\partial y}\hat{\jmath} + \frac{\partial V}{\partial z}\hat{k}\right) \qquad (22.9)$$

Equation 22.9 confirms that the electric field is strong where the potential changes rapidly. The minus sign here is the same as in Equation 22.1: It says that if we move in the direction of *increasing* potential, then we're moving *against* the electric field. Equation 22.9 also shows that the units of electric field, N/C, can be written as volts per meter—a unit widely used in both science and engineering.

Because potential is a scalar, it's often easier to calculate the potential and then use Equation 22.9 to get the field. The next example shows how.

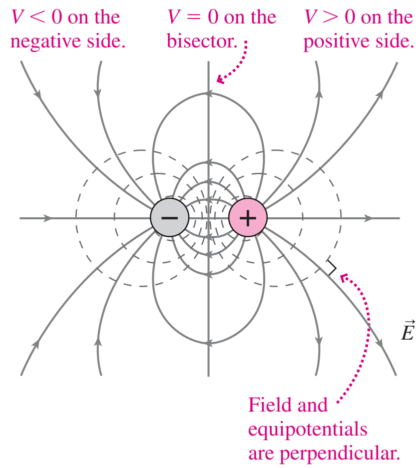


FIGURE 22.16 Equipotentials (dashed curves) and field lines for a dipole, in a plane containing the dipole. You should convince yourself that the equipotentials provide a contour map of Fig. 22.11.

EXAMPLE 22.8 Field from Potential: A Charged Disk

Use the result of Example 22.7 to find the electric field on the axis of a charged disk.

INTERPRET Example 22.7 gives the potential of a charged disk, so this problem is about calculating electric field from potential.

DEVELOP Example 22.7 gives the potential on the axis of a charged disk: $V(x) = (2kQ/a^2)\left(\sqrt{x^2 + a^2} - |x|\right)$. Equation 22.9 shows that we can get the electric field by differentiating the potential with respect to all three coordinates. But here the potential depends only on x, so $\vec{E}$ has only an x-component—a fact that should also be evident from symmetry. Our plan is to apply Equation 22.9, differentiating $V(x)$ to get the field component E_x.

EVALUATE We apply Equation 22.9 to $V(x)$ to get

$$E_x = -\frac{dV}{dx} = -\frac{d}{dx}\left[\frac{2kQ}{a^2}\left(\sqrt{x^2 + a^2} - |x|\right)\right] = \frac{2kQ}{a^2}\left(\pm 1 - \frac{x}{\sqrt{x^2 + a^2}}\right)$$

where the $+/-$ signs apply for $x > 0$ and $x < 0$, respectively. Because V depends only on x, we wrote the total derivative dV/dx rather than the partial derivative.

ASSESS Make sense? In Fig. 22.14 we showed that the field of a disk ought to look like that of a charged sheet close up and like a point charge far away. For $|x| \ll a$, our result gives $|E_x| = 2kQ/a^2$. With $k = 1/4\pi\epsilon_0$ and $Q/\pi a^2 = \sigma$, this is indeed the field $E = \sigma/2\epsilon_0$ of a charged sheet. You can show in Problem 68 that the case $|x| \gg a$ reduces to the field of a point charge Q, as expected. ■

✓TIP Field and Potential

Note that the *values* of field and potential at a single point aren't related; rather, as Equation 22.9 shows, field measures the *rate of change* of potential. Field and potential are related in the same way as acceleration and velocity; their *values* are independent, but one is the rate of change of the other—although with a negative sign in the case of field and potential. In particular, the field can be zero where the potential isn't, and vice versa.

22.4 Charged Conductors

There's no electric field inside a conducting material in electrostatic equilibrium, and at the conductor surface there's no field component parallel to the surface. Therefore, it takes no work to move a test charge on or inside a conductor—and that means **a conductor in electrostatic equilibrium is an equipotential.**

Consider an isolated, spherical conductor of radius R carrying charge Q. Charge is distributed uniformly over its surface, so the field outside the sphere is that of a point charge. Then the potential at its surface is $V(R) = kQ/R$, as we found in Example 22.3. Now consider two widely separated spheres of different sizes. If we connect them by a thin conducting wire (Fig. 22.17), charge will move through the wire until both spheres are at the same potential. But since the spheres are widely separated, each still has an essentially spherical charge distribution, so $V(R) = kQ/R$ gives each sphere's potential. Because the spheres have the same potential, $kQ_1/R_1 = kQ_2/R_2$. We can write each charge as the surface area multiplied by the surface charge density: $Q = 4\pi R^2\sigma$. Substituting for the Q's in the above equation and solving for the ratio of surface charge densities then give

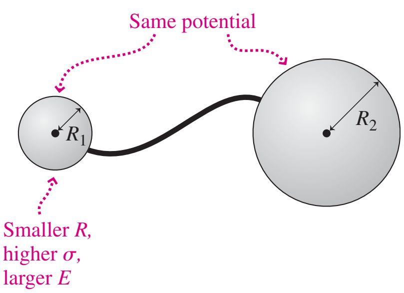


FIGURE 22.17 Two conducting spheres connected by a long conducting wire.

$$\frac{\sigma_1}{\sigma_2} = \frac{R_2}{R_1}$$

Thus the *smaller* sphere has the *higher* surface charge density. Since the electric field at a conductor surface has magnitude $E = \sigma/\epsilon_0$, the field must also be stronger at the smaller sphere. Conceptual Example 22.1 explores this situation further.

CONCEPTUAL EXAMPLE 22.1 An Irregular Conductor

Sketch some equipotentials and electric field lines for an isolated egg-shaped conductor.

EVALUATE Where the conductor surface curves sharply, it's like the small sphere of Fig. 22.17. It therefore has higher surface charge density and a stronger electric field, which means more field lines emerge where the surface curves sharply. Since the field is perpendicular at the conductor surface, equipotentials just above the surface have essentially the same shape as the surface. Far from the charged conductor, on the other hand, its field resembles that of a point charge, with radial field lines and circular equipotentials. Figure 22.18 gives an approximate picture of the field and equipotentials based on these considerations.

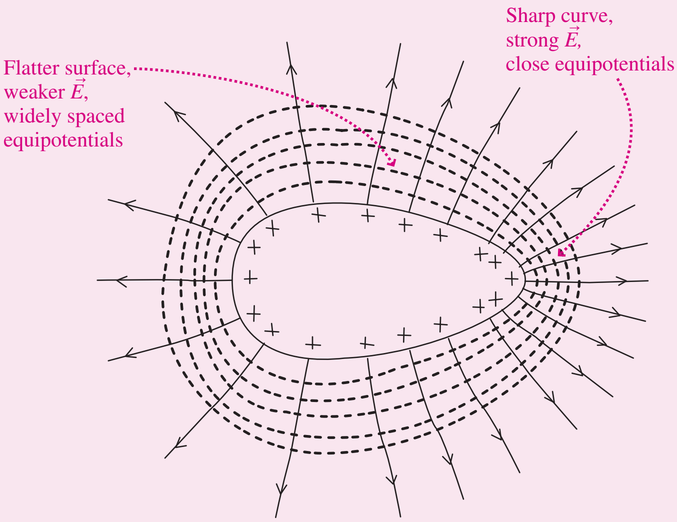


FIGURE 22.18 Equipotentials and field of a charged conductor.

ASSESS Our analysis here applies only to an *isolated* conductor. Figure 22.19 shows how the presence of nearby charges alters the charge distribution on a conductor.

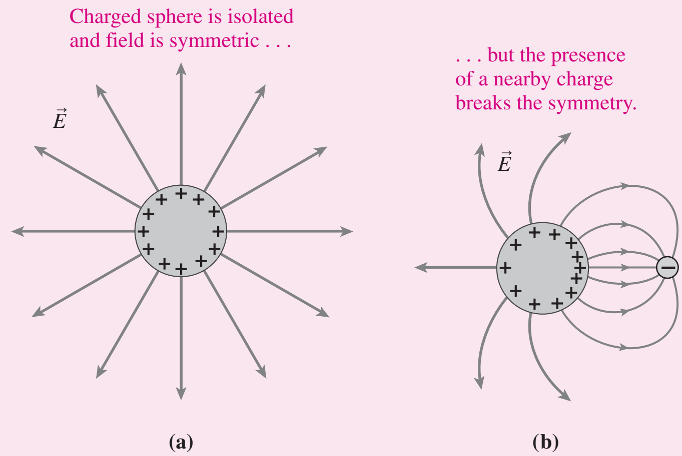
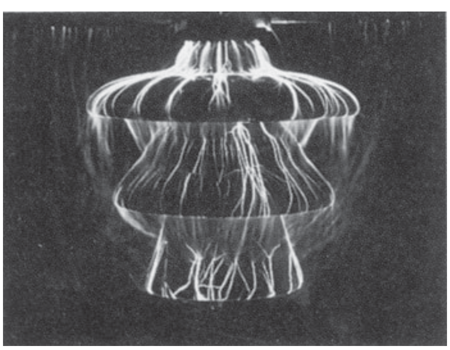


FIGURE 22.19 Distribution of charge on a conductor changes in the presence of another charge.

MAKING THE CONNECTION The potential difference between the conductor in Fig. 22.18 and the outermost equipotential shown is 70 V. Determine approximate values for the strongest and weakest electric fields in the region shown in the figure, assuming it's drawn at actual size.

EVALUATE Electric field is the rate of change of potential. At the tip, where the field is strongest, the outermost equipotential is about 7 mm from the conductor. So the field here is approximately 70 V/7 mm; that's 10 V/mm or 10 kV/m. At its most distant, the outer equipotential is about 12 mm from the conductor, giving a field of 70 V/12 mm or nearly 6 kV/m.

APPLICATION Corona Discharge, Pollution Control, and Xerography



The large electric fields that develop at sharply curved conductors can cause serious problems in electric equipment. Fields stronger than 3 MN/C strip electrons from air molecules, making air a conductor. The result is a blue glow, called **corona discharge**, resulting from electrons recombining with atoms. Corona discharge causes energy loss from high-voltage transmission lines, so engineers try to avoid sharp edges on conducting structures. The photo shows corona discharge leaking current across a power-line insulator.

Corona discharge can also be useful. Pollution-control devices called **electrostatic precipitators** use a thin wire at a high negative potential to produce a strong field that ionizes gas molecules. Ions adhere to pollutant particles, which are then attracted to positively charged collecting plates. Such devices remove up to 99% of particulate pollutants from power plants and factories.

You're using corona discharge when you make a photocopy or use a laser printer. The "ink" in these devices consists of tiny plastic "toner" particles that adhere to charged regions on a special light-sensitive drum. The drum is initially charged uniformly by a corona discharge from a thin wire at high voltage; light then neutralizes all but the dark regions. Toner particles are electrically attracted to those regions and then transferred to paper, where they're melted in to make a permanent image.

CHAPTER 22 SUMMARY

Big Picture

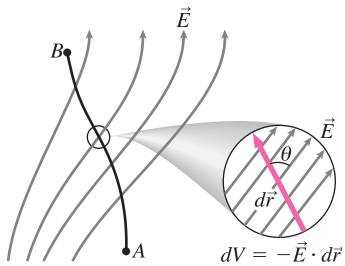
The big idea here is **electric potential difference**—a measure of the energy per unit charge involved in moving charge between two points in an electric field. Because the electric field is conservative, potential difference is path independent and thus depends only on the two points in question.

Key Concepts and Equations

Electric potential difference between points A and B is the negative of the line integral of the electric field over any path from A to B:

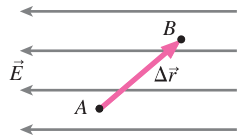
$$\Delta V_{AB} = \frac{\Delta U_{AB}}{q} = -\int_A^B \vec{E} \cdot d\vec{r}$$

When a charge "falls" through a potential difference ΔV, it gains energy $q\Delta V$.

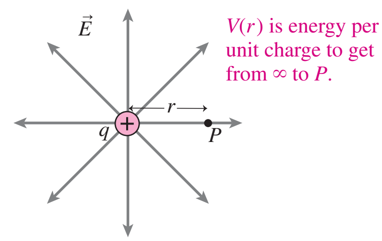


In a uniform field, the potential difference becomes

$$\Delta V_{AB} = -\vec{E} \cdot \Delta\vec{r}$$

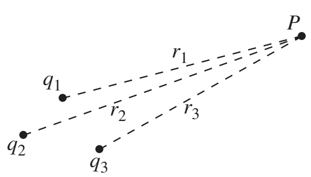


The potential in the field of a point charge is $V(r) = kq/r$, where the zero of potential is taken at infinity and r is the distance from the point charge.

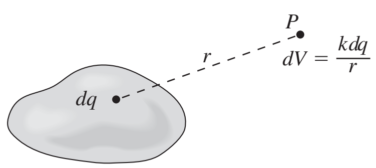


Potentials of charge distributions follow by summing or integrating the fields of pointlike charge elements:

$$V = \sum \frac{kq_i}{r_i} \quad \text{(discrete charges)}$$

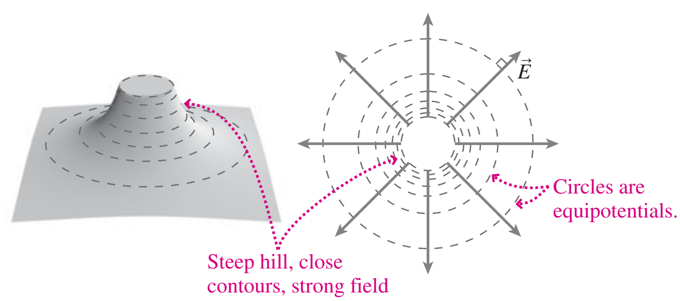


$$V = \int \frac{k\,dq}{r} \quad \text{(continuous charge distribution)}$$



Equipotentials are surfaces of constant potential perpendicular to the electric field. Where equipotentials are close, the field is strong. The field component in a given direction depends on the rate at which potential changes with position; thus,

$$E_x = -\frac{dV}{dx}$$

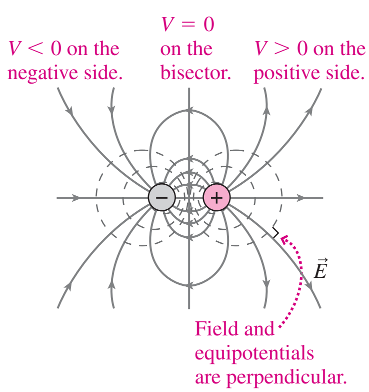


Applications

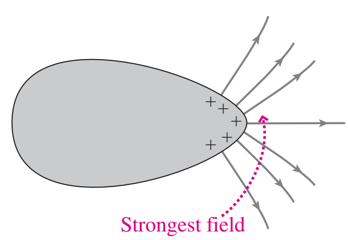
The dipole potential is

$$V = \frac{kp\cos\theta}{r^2}$$

where $p = qd$ is the dipole moment and the angle θ is measured from the dipole axis.



In charged conductors, the charge density is generally highest, and the field strongest, where a conductor curves sharply.



For Thought and Discussion

1. Why can a bird perch on a high-voltage power line without getting electrocuted?
2. One proton is accelerated from rest by a uniform electric field, another proton by a nonuniform electric field. If they move through the same potential difference, how do their final speeds compare?
3. Would a free electron move toward higher or lower potential?
4. The electric field at the center of a uniformly charged ring is obviously zero, yet Example 22.6 shows that the potential at the center isn't zero. How is this possible?
5. Must the potential be zero at any point where the electric field is zero? Explain.
6. Must the electric field be zero at any point where the potential is zero? Explain.
7. The potential is constant throughout an entire volume. What must be true of the electric field within that volume?
8. In considering the potential of an infinite flat sheet, why isn't it useful to take the zero of potential at infinity?
9. "Cherry picker" trucks for working on power lines often carry electrocution hazard signs. Explain how this hazard arises and why it might be more of a danger to someone on the ground than to a worker on the truck.
10. Can equipotential surfaces intersect? Explain.
11. Is the potential at the center of a hollow, uniformly charged spherical shell higher than, lower than, or the same as at the surface?
12. A solid sphere contains positive charge uniformly distributed throughout its volume. Is the potential at its center higher than, lower than, or the same as at the surface?
13. Two equal but opposite charges form a dipole. Describe the equipotential surface on which $V = 0$.
14. The electric potential in a region increases linearly with distance. What can you conclude about the electric field in this region?

Exercises and Problems

Exercises

Section 22.1 Electric Potential Difference

15. How much work does it take to move a 50-μC charge against a 12-V potential difference?
16. The potential difference between the two sides of an ordinary electric outlet is 120 V. How much energy does an electron gain when it moves from one side to the other?
17. It takes 45 J to move a 15-mC charge from point A to point B. What's the potential difference ΔV_{AB}?
18. Show that 1 V/m is the same as 1 N/C.
19. Find the magnitude of the potential difference between two points located 1.4 m apart in a uniform 650-N/C electric field, if a line between the points is parallel to the field.
20. A charge of 3.1 C moves from the positive to the negative terminal of a 9.0-V battery. How much energy does the battery impart to the charge?
21. A proton, an alpha particle (a bare helium nucleus), and a singly ionized helium atom are accelerated through a 100-V potential difference. How much energy does each gain?
22. BIO The potential difference across a typical cell membrane is about 80 mV. How much work is done on a singly ionized potassium ion moving through the membrane?

Section 22.2 Calculating Potential Difference

23. An electric field is given by $\vec{E} = E_0\hat{j}$, where E_0 is a constant. Find the potential as a function of position, taking $V = 0$ at $y = 0$.
24. The classical picture of the hydrogen atom has the electron orbiting 0.0529 nm from the proton. What's the electric potential associated with the proton's electric field at this distance?
25. The potential at the surface of a 10-cm-radius sphere is 4.8 kV. What's the sphere's total charge, assuming charge is distributed in a spherically symmetric way?
26. You're developing a switch for high-voltage power lines. The smallest part in your design is a 5.0-cm-diameter metal sphere. What do you specify for the maximum potential on your switch if the electric field at the sphere's surface isn't to exceed the 3-MV/m breakdown field of air?
27. A 3.5-cm-diameter isolated metal sphere carries 0.86 μC. (a) Find the potential at the sphere's surface. (b) If a proton were released from rest at the surface, what would be its speed far from the sphere?

Section 22.3 Potential Difference and the Electric Field

28. In a uniform electric field, equipotential planes that differ by 1.0 V are 2.5 cm apart. What's the field strength?
29. Figure 22.20 shows a plot of potential versus position along the x-axis. Make a plot of the x-component of the electric field for this situation.

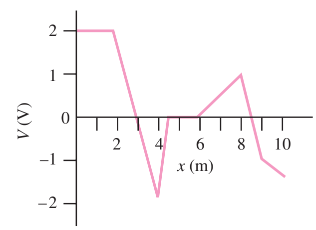


FIGURE 22.20 Exercise 29

30. Figure 22.21 shows some equipotentials in the x-y plane. (a) In what region is the electric field strongest? What are (b) the direction and (c) the magnitude of the field in this region?

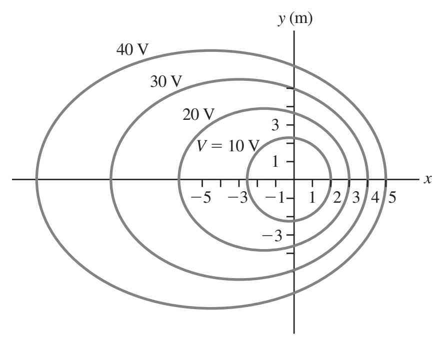


FIGURE 22.21 Exercise 30

31. The electric potential in a region is $V = 2xy - 3zx + 5y^2$, with V in volts and the coordinates in meters. Find (a) the potential and (b) the components of the electric field at the point $x = 1$ m, $y = 1$ m, $z = 1$ m.

Section 22.4 Charged Conductors

32. Dielectric breakdown of air occurs at fields of 3 MV/m. Find (a) the maximum potential (measured from infinity) for the sphere of Example 22.3 before dielectric breakdown occurs at the sphere's surface and (b) the charge on the sphere at this potential.
33. You're an automotive engineer working on the ignition system for a new engine. Its spark plugs have center electrodes made from 2.0-mm-diameter wire. The electrode ends gradually wear to a hemispherical shape, so they behave approximately like charged spheres. Your job is to specify the minimum potential that ensures these plugs will spark in air, neglecting the presence of the second electrode.
34. A large metal sphere has three times the diameter of a smaller sphere and carries three times the charge. Both spheres are isolated, so their surface charge densities are uniform. Compare (a) the potentials and (b) the electric field strengths at their surfaces.

Problems

35. Two points A and B lie 15 cm apart in a uniform electric field, with the path AB parallel to the field. If the potential difference ΔV_{AB} is 840 V, what's the field strength?
36. BIO The electric field within a cell membrane is approximately 8.0 MV/m and is essentially uniform. If the membrane is 10 nm thick, what's the potential difference across the membrane?
37. What's the potential difference between the terminals of a battery that can impart 7.2×10^{-19} J to each electron that moves between the terminals?
38. What's the charge on an ion that gains 1.6×10^{-15} J when it moves through a potential difference of 2500 V?
39. Two flat metal plates are a distance d apart, where d is small compared with the plate size. If the plates carry surface charge densities $\pm\sigma$, show that the magnitude of the potential difference between them is $V = \sigma d/\epsilon_0$.
40. An electron passes point A moving at 6.5 Mm/s. At point B it's come to a stop. Find the potential difference ΔV_{AB}.
41. A 5.0-g object carries 3.8 μC. It acquires speed v when accelerated from rest through a potential difference V. If a 2.0-g object acquires twice the speed under the same circumstances, what's its charge?
42. Points A and B lie 20 cm apart on a line extending radially from a point charge Q, and the potentials at these points are $V_A =$ 280 V and $V_B = 130$ V. Find Q and the distance r between A and the charge.
43. A sphere of radius R carries negative charge of magnitude Q, distributed in a spherically symmetric way. Find the escape speed for a proton at the sphere's surface—that is, the speed that would enable the proton to escape to arbitrarily large distances starting at the sphere's surface.
44. BIO Proton-beam therapy is preferable to X rays for cancer treatment because protons deliver most of their energy to the tumor, with less damage to healthy tissue. A cyclotron used to accelerate protons for cancer treatment repeatedly passes the protons through a 15-kV potential difference. (a) How many passes are needed to bring the protons' kinetic energy to 1.2×10^{-11} J? (b) What's the resulting proton energy in electronvolts?
45. A thin spherical shell has radius R and total charge Q distributed uniformly over its surface. Find the potential at its center.
46. A solid sphere of radius R carries charge Q distributed uniformly throughout its volume. Find the potential difference from the sphere's surface to its center. (*Hint:* Consult Example 21.1.)
47. Find the potential as a function of position in the electric field $\vec{E} = ax\hat{\imath}$, where a is a constant and $V = 0$ at $x = 0$.
48. Your radio station needs a new coaxial cable to connect the transmitter and antenna. One possible cable consists of a 2.0-mm-diameter inner conductor and an outer conductor with diameter 1.6 cm and negligible thickness (Fig. 22.22); the maximum safe potential difference between the conductors is 2 kV. In your application, the conductors carry charge densities ±62 nC/m. Will this cable work for you?

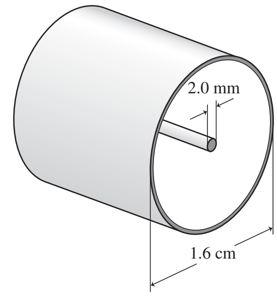


FIGURE 22.22 Problem 48

49. The potential difference between the surface of a 3.0-cm-diameter power line and a point 1.0 m distant is 3.9 kV. Find the line charge density on the power line.
50. Three equal charges q form an equilateral triangle of side a. Find the potential at the center of the triangle.
51. A charge $+Q$ lies at the origin and $-3Q$ at $x = a$. Find two points on the x-axis where $V = 0$.
52. Two identical charges q lie on the x-axis at $\pm a$. (a) Find an expression for the potential at all points in the x-y plane. (b) Show that your result reduces to the potential of a point charge for distances large compared with a.
53. A dipole of moment $p = 2.9$ nC $\cdot$ m consists of two charges separated by far less than 10 cm. Find the potential 10 cm from the dipole (a) on its axis, (b) at 45° to its axis, and (c) on its perpendicular bisector.
54. A thin plastic rod 20 cm long carries 3.2 nC distributed uniformly over its length. (a) If the rod is bent into a ring, find the potential at its center. (b) If it's bent into a semicircle, find the potential at the center (i.e., at the center of the circle of which the semicircle is part).
55. A thin ring of radius R carries charge $3Q$ distributed uniformly over three-fourths of its circumference, and $-Q$ over the rest. Find the potential at the ring's center.
56. The potential at the center of a uniformly charged ring is 45 kV, and 15 cm along the ring axis the potential is 33 kV. Find the ring's radius and total charge.
57. The annulus shown in Fig. 22.23 carries a uniform surface charge density σ. Find an expression for the potential at an arbitrary point P on its axis.

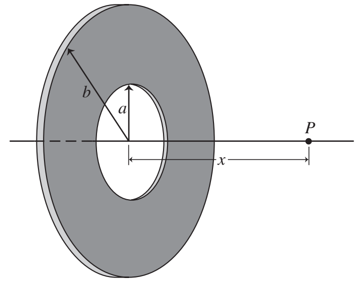


FIGURE 22.23 Problem 57

58. The potential in a region is $V = axy$, where a is a constant. (a) Determine the electric field in the region. (b) Sketch some equipotentials and field lines.
59. Use Equation 22.6 to calculate the electric field on the perpendicular bisector of a dipole, and show that your result is equivalent to Equation 20.6a.
60. Use the result of Example 22.6 to determine the on-axis field of a charged ring, and verify that your answer agrees with the result of Example 20.6.
61. The electric potential in a region is $V = -V_0(r/R)$, where V_0 and R are constants and r is the radial distance from the origin. Find expressions for the magnitude and direction of the electric field in this region.
62. Two metal spheres each 1.0 cm in radius are far apart. One sphere carries 38 nC, the other -10 nC. (a) What's the potential on each? (b) If the spheres are connected by a thin wire, what will be the potential on each once equilibrium is reached? (c) How much charge moves between the spheres in order to achieve equilibrium?
63. Two 5.0-cm-diameter conducting spheres are 8.0 m apart, and each carries 0.12 μC. Determine (a) the potential on each sphere, (b) the field strength at the surface of each sphere, (c) the potential midway between the spheres, and (d) the potential difference between the spheres.
64. A 2.0-cm-radius metal sphere carries 75 nC and is surrounded by a concentric spherical conducting shell of radius 10 cm carrying -75 nC. (a) Find the potential difference between shell and sphere. (b) How would your answer change if the shell's charge were $+150$ nC?
65. A sphere of radius R carries a nonuniform but spherically symmetric volume charge density that results in an electric field in the sphere given by $\vec{E} = E_0(r/R)^2\hat{r}$, where E_0 is a constant. Find the potential difference from the sphere's surface to its center.
66. The potential as a function of position in a region is $V(x) = 3x - 2x^2 - x^3$, with x in meters and V in volts. Find (a) all points on the x-axis where $V = 0$, (b) an expression for the electric field, and (c) all points on the x-axis where $E = 0$.
67. A conducting sphere 5.0 cm in radius carries 60 nC. It's surrounded by a concentric spherical conducting shell of radius 15 cm carrying -60 nC. (a) Find the potential at the sphere's surface, taking $V = 0$ at infinity. (b) Repeat for the case when the shell carries $+60$ nC.
68. Show that the result of Example 22.8 approaches the field of a point charge for $x \gg a$. (*Hint:* You'll need to apply the binomial approximation from Appendix A to the expression $1/\sqrt{x^2 + a^2}$.)
69. The potential on the axis of a uniformly charged disk at 5.0 cm from the disk center is 150 V; the potential 10 cm from the disk center is 110 V. Find the disk radius and its total charge.
70. A uranium nucleus (mass 238 u, charge $92e$) decays, emitting an alpha particle (mass 4 u, charge $2e$) and leaving a thorium nucleus (mass 234 u, charge $90e$). At the instant the alpha particle leaves the nucleus, the centers of the two are 7.4 fm apart and essentially at rest. Treating each particle as a spherical charge distribution, find their speeds when they're a great distance apart.
71. A disk of radius a carries nonuniform surface charge density $\sigma = \sigma_0(r/a)$, where σ_0 is a constant. (a) Find the potential at an arbitrary point x on the disk axis, where $x = 0$ is the disk center. (b) Use the result of (a) to find the electric field on the disk axis, and (c) show that the field reduces to an expected form for $x \gg a$.
72. An open-ended cylinder of radius a and length $2a$ carries charge q spread uniformly over its surface. Find the potential at the center of the cylinder. (*Hint:* Treat the cylinder as a stack of charged rings, and integrate.)
73. A line charge extends along the x-axis from $-L/2$ to $L/2$. Its line charge density is $\lambda = \lambda_0(x/L)^2$, where λ_0 is a constant. Find an expression for the potential on the x-axis for $x > L/2$. Check that your expression reduces to an expected result for $x \gg L$.
74. Repeat Problem 73 for the charge distribution $\lambda = \lambda_0 x/L$. (*Hint:* What does this charge distribution resemble at large distances?)
75. C You're sizing a new electric transmission line, and you can save money with thinner wire. The potential difference between the line and the ground, 60 m below, is 115 kV. The field at the wire surface cannot exceed 25% of the 3-MV/m breakdown field in air. Neglecting charges in the ground itself, what minimum wire diameter do you specify? (*Hint*: You'll have to do a numerical calculation.)

Passage Problems

BIO Standard electrocardiography measures time-dependent potential differences between multiple points on the body, giving cardiologists multiple perspectives on the heart's electrical activity. In contrast, Fig. 22.24 is a "snapshot" showing a more detailed picture at an instant of time. The lines are equipotentials on the surface of a human torso, associated with the heart's electrical activity. Relative to the line marked $V = 0$, the potential is negative to the upper left (black) and positive to the lower right (green).

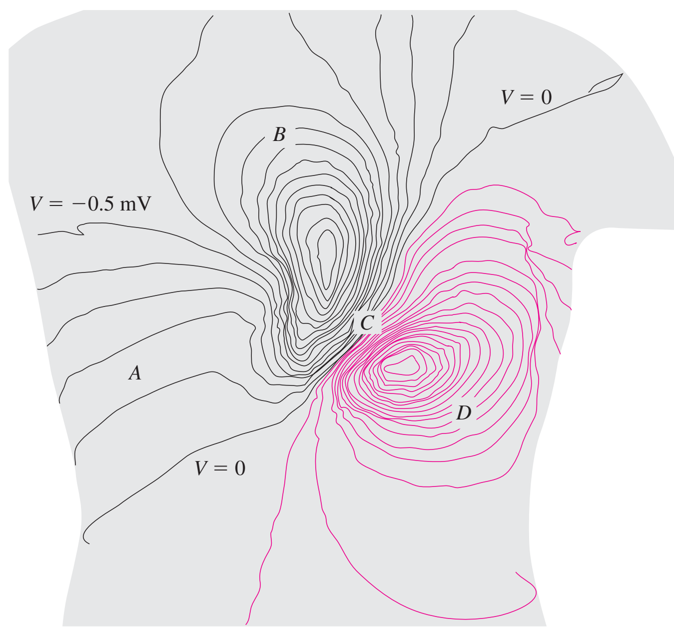


FIGURE 22.24 Equipotentials on a human torso (Passage Problems 76–79)

76. From the equipotentials, you can infer that the heart's electrical structure resembles that of a
 a. uniform charged sheet.
 b. dipole.
 c. point charge.
 d. uniformly charged sphere.
77. The electric field in the vicinity of the heart points approximately
 a. from upper left to lower right.
 b. from lower left to upper right.
 c. from upper right to lower left.
 d. from lower right to upper left.

78. The electric field is strongest in the region marked
   a. A.
   b. B.
   c. C.
   d. D.

79. The electric field in region A is approximately
   a. 20 μN/C.
   b. 2 mN/C.
   c. 20 mN/C.
   d. 2 kN/C.

Answers to Chapter Questions

Answer to Chapter Opening Question

138,000 volts is a measure of electric potential difference—the energy per unit charge involved in moving electric charge between two points. Luckily, the parasailer is in contact with only one wire, so he doesn't experience that lethal potential difference.

Answers to GOT IT? Questions

22.1. (a) doubles; (b) doubles; (c) becomes zero; (d) reverses sign.
22.2. (a) 10 eV; (b) 20 eV; (c) 10 eV.
22.3. (c) has the highest ΔV_{AB}; (b) has the lowest.
22.4. (a), because the field is stronger.
22.5. They're all equal to zero, because the potential anywhere on the perpendicular bisector of a dipole is zero, and they're all the same, because potential difference is path independent.
22.6. (a), because the equipotentials are closer nearer the center, indicating a stronger field. In (b) the field actually gets stronger farther from the center.

23 Electrostatic Energy and Capacitors

The lifesaving jolt of a defibrillator requires a large amount of energy delivered in a short time. Where does that energy come from?

New Concepts, New Skills

By the end of this chapter you should be able to

- Calculate the electrostatic energy of a system of discrete charges (23.1).
- Explain the concept of capacitance, and relate charge, potential difference, and capacitance (23.2).
- Calculate the capacitance of a parallel-plate capacitor (23.2).
- Find the charge and energy stored in a capacitor (23.2).
- Calculate the equivalent capacitance of parallel and series capacitor combinations, and determine maximum working voltages (23.3).
- Describe how all electric fields represent stored energy, and calculate that energy from the electric energy density (23.4).

Connecting Your Knowledge

- This chapter builds on the concept of electric potential (22.1, 22.2).
- You should be familiar with the electric fields of charged conductors (21.6, 22.4).
- You should also be familiar with the concept of dielectrics (20.5).

Figure 23.1 shows three positive charges arranged to form a triangle. Stored in this charge distribution is **electrostatic energy** representing the work done against the repulsive electric forces as the charges were brought into proximity. Although this example may seem trivial, its implications are not. Energy storage in configurations of electric charge is a vital aspect of natural and technological systems. The energy of chemical reactions—including metabolizing food and burning fuels—is ultimately electric energy released in the rearrangement of molecular charge distributions. Energy storage using charged conductors is essential in technologies ranging from computer memories to cameras to high-powered lasers.

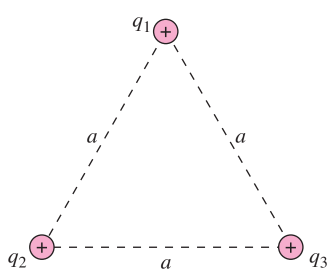


FIGURE 23.1 Electrostatic energy is stored in this configuration of three point charges.

23.1 Electrostatic Energy

Let's find the energy stored in the configuration of Fig. 23.1, assuming we start with widely separated charges and bring them in sequentially. It takes no work to bring in charge q_1, since there's initially no electric field. But with q_1 in place, bringing in q_2 means doing work against q_1's electric field. In Chapter 22 we found that the potential of a point charge q is $V = kq/r$. So the potential V_1 *due* to q_1 at the eventual location of q_2 is kq_1/a, where a is the side of the triangle. That potential is the energy per unit charge; given q_2's charge, the work needed to bring in q_2 is $W_2 = q_2V_1 = kq_1q_2/a$. Then we bring in q_3, which experiences the electric fields of both q_1 and q_2. This requires work done against both fields; following our reasoning for q_2, that work is $W_3 = kq_1q_3/a + kq_2q_3/a$. The denominator is the same in both terms because the charges form an equilateral triangle. So the total work done to assemble this charge distribution is

$$W_2 + W_3 = kq_1q_2/a + kq_1q_3/a + kq_2q_3/a.$$

Because the electric field is conservative, this work becomes the stored electrostatic energy, U.

Although we considered the three charges in Fig. 23.1 to be positive, our expression for work holds no matter what the signs. That means electrostatic energy can be positive or negative, depending on the sign of the work done in assembling a charge distribution. If it's negative, then it takes work to separate the charges. Although we considered assembling our charges in the order 1, 2, 3, the expression for work would have been the same no matter what the order—showing that electrostatic energy is a property of a charge distribution, independent of how it's assembled. Figure 23.1 is a simple metaphor for a molecule. Water, for example, consists of a negatively charged oxygen atom and two positively charged hydrogen atoms. The electrostatic energy is negative and represents the energy it would take to dissociate the molecule; equivalently, it's the energy released when the water forms from individual atoms.

23.2 Capacitors

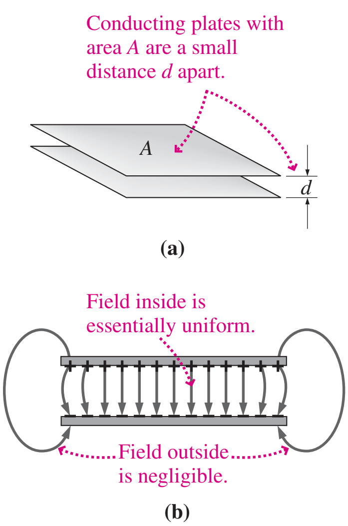


FIGURE 23.2 (a) A parallel-plate capacitor consists of closely spaced conducting plates with area A and spacing d. (b) Edge-on view, showing the electric field.

In technological applications, we often store energy in **capacitors**—pairs of electrical conductors that carry equal but opposite charges. Although capacitors come in many configurations, it's easiest to analyze the **parallel-plate capacitor** consisting of two closely spaced conducting plates (Fig. 23.2*a*). Understanding this device not only is technologically valuable, but will also give us deep insights into the electric field and electrostatic energy.

Initially both capacitor plates are electrically neutral. We charge the capacitor by transferring charge between the plates, building up positive charge on one plate and equal negative charge on the other. In practice, we accomplish this by connecting the capacitor to a battery, but here it's easier to imagine grabbing charge from one plate and physically moving it to the other. Charge on the plates produces an electric field between them, as shown in Fig. 23.2*b*. With closely spaced plates that field is essentially uniform in the region between the plates, except right near the edges. Outside, the field is so small as to be negligible. So we can approximate the parallel-plate capacitor as having a uniform field confined entirely to the region between its plates.

In Chapter 21 we showed that the electric field at the surface of a conductor is $E = \sigma/\epsilon_0$, with σ the charge per unit area. Here we've got charge spread uniformly over the capacitor plates, so if there's charge Q on a plate, then $\sigma = Q/A$, and the uniform field between the plates is $E = Q/\epsilon_0 A$. (If you think this should be doubled because there are two plates, reread the discussion around Figs. 21.26 and 21.27 to see why not.) In this uniform field, the potential difference between the plates is the product of the field and the plate separation: $V = Ed = Qd/\epsilon_0 A$.

Capacitance

We can rewrite our expression for the potential difference between the plates of the capacitor in the form $Q = (\epsilon_0 A/d)V$. We added the parentheses to emphasize two things: First, charge is linearly proportional to potential difference and, second, the proportionality factor depends only on the constant ϵ_0 and on the geometry—here the plate area and spacing—of the two charged conductors. This factor gives the ratio of charge to potential difference, which defines the **capacitance** of a configuration of two conductors:

$$C = \frac{Q}{V} \qquad \text{(capacitance)} \tag{23.1}$$

Capacitance depends on the physical arrangement of the conductors, and it's a constant for a given capacitor. Our expression $Q = (\epsilon_0 A/d)V$ shows that the capacitance of a parallel-plate capacitor is

$$C = \frac{\epsilon_0 A}{d} \qquad \text{(parallel-plate capacitor)} \tag{23.2}$$

Problems 40 and 41 explore capacitance for other configurations.

Equation 23.1 shows that the units of capacitance are coulombs/volt. This unit has its own name, the **farad** (F), in honor of the 19th-century scientist Michael Faraday. One farad is a large capacitance; practical capacitors are often measured in μF (10^{-6} F) or pF (10^{-12} F). Incidentally, Equation 23.2 shows that the units of ϵ_0 may be expressed as F/m.

Energy Storage in Capacitors

Imagine moving a small charge dQ from the negative to the positive plate of a capacitor when there's a potential difference V between the plates. Since potential difference is work per unit charge, this takes work $dW = V\,dQ$. The additional charge increases the electric field in the capacitor, resulting in an increase dV in the potential difference. Equation 23.1 shows that the increases dQ and dV are related by $dQ = C\,dV$. So the work involved in moving the charge dQ between the plates becomes $dW = V\,dQ = CV\,dV$.

If we start with the capacitor uncharged and then begin transferring charge between the plates, we'll need to do increasing amounts of work because the electric field and potential difference increase continuously with the charge we've already transferred. The total work involved will be the sum of all the dW values. Here the potential difference increases continuously, so that sum becomes an integral:

$$W = \int dW = \int_0^V CV\,dV = \tfrac{1}{2}CV^2$$

where the last step follows because the integral has the familiar form $\int x\,dx = \frac{1}{2}x^2$. The work we do in charging the capacitor is stored as potential energy U, so

$$U = \tfrac{1}{2}CV^2 \qquad \text{(energy in a capacitor)} \tag{23.3}$$

We can measure potential difference V directly, with a voltmeter, so it's more useful to express the energy in terms of voltage rather than charge.

✓TIP Charged but Neutral

A "charged capacitor" means a capacitor with one plate positive and the other negative; overall, the capacitor remains neutral (Fig. 23.3). The charge Q refers to the *magnitude* of the charge on either plate—*not* to the capacitor's net charge, which is zero.

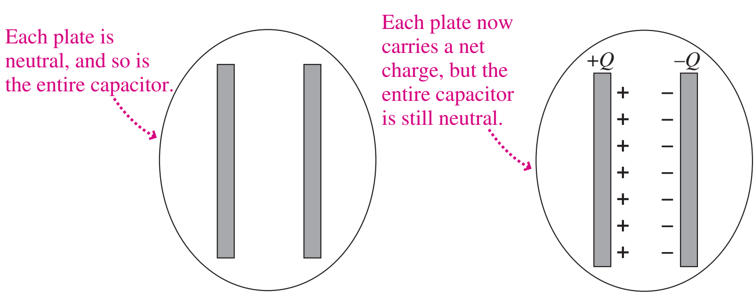


FIGURE 23.3 The net charge on the entire capacitor is zero, whether it's uncharged (left) or charged (right).

EXAMPLE 23.1 Capacitance, Charge, and Energy: A Parallel-Plate Capacitor

A capacitor consists of two circular metal plates of radius $R = 12$ cm, separated by $d = 5.0$ mm. (a) Find its capacitance. Find (b) the charge on the plates and (c) the stored energy when the capacitor is connected to a 12-V battery.

INTERPRET Because the plates' area is much larger than their separation, we can treat the field between them as uniform. So we identify the configuration as a parallel-plate capacitor.

DEVELOP We've sketched the capacitor in Fig. 23.4. Equation 23.2, $C = \epsilon_0 A/d$, determines the capacitance for part (a) from the separation distance and plate area $(A = \pi R^2)$. For parts (b) and (c) the 12-V battery maintains a 12-V potential difference across the capacitor. Knowing that voltage and the capacitance, we can find the capacitor's charge from Equation 23.1, $C = Q/V$, and the stored energy from Equation 23.3, $U = \frac{1}{2}CV^2$.

EVALUATE We first solve part (a) for the capacitance:

$$C = \frac{\epsilon_0 A}{d} = \frac{\epsilon_0 \pi R^2}{d} = 80 \text{ pF}$$

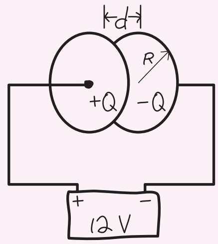


FIGURE 23.4 Sketch for Example 23.1.

For part (b) the definition of capacitance then gives

$$Q = CV = (80 \text{ pF})(12 \text{ V}) = 960 \text{ pC}$$

or just under 1 nC. Then (c) the stored energy is

$$U = \tfrac{1}{2}CV^2 = \tfrac{1}{2}(80 \text{ pF})(12 \text{ V})^2 = 5760 \text{ pJ}$$

or about 5.8 nJ.

ASSESS Make sense? At 80 pF, this is a pretty small capacitor, so no wonder the charge and energy are measured in nano-units (nC and nJ).

FIGURE 23.5 Typical capacitors. The large unit is an 18-mF electrolytic capacitor. At top right is an air-insulated variable capacitor in which one set of plates rotates to change the capacitance. The smaller capacitors range from 43 pF to 10 μF.

23.3 Using Capacitors

Capacitors are essential in modern technology. They range from the billions of 25-fF $(10^{-15}$ F) capacitors that store individual bits of information in your computer's memory, to millifarad-range capacitors that smooth 60-Hz AC power to provide steady current to your stereo, to ultracapacitors measuring hundreds of farads that store electric energy for short bursts of power in some gas–electric hybrid and fuel-cell vehicles (Fig. 23.5).

Practical Capacitors

Equation 23.2 shows that the way to achieve a large capacitance is with large plate area and small spacing. That's true in general, whether or not a capacitor has parallel-plate geometry. Inexpensive capacitors are often made from two long strips of aluminum foil separated by thin plastic insulation. This foil "sandwich" is rolled into a compact cylinder, wires are attached, and the whole thing is dipped in a protective coating. Very large capacitances are achieved with electrolytic capacitors, in which a thin insulating layer develops chemically under the influence of the applied voltage. Capacitors are among the hardest components to fabricate on integrated-circuit chips, but small-capacitance units can be made by alternating conductive material with an insulating layer.

Our analysis of the parallel-plate capacitor assumed air between the plates. But most capacitors have solid insulating materials, or **dielectrics**, that contain molecular dipoles but no free charge. In Section 20.5 we showed how the alignment of molecular dipoles in a dielectric reduces the field in the material. In a capacitor, the effect is to reduce also the potential difference V between the plates (Fig. 23.6). The factor by which the field and potential difference decrease is the **dielectric constant κ** (Greek kappa). For a given charge Q, decreased potential difference means a larger capacitance $C = Q/V$. Thus a parallel-plate capacitor with a dielectric between its plates has capacitance

$$C = \kappa \frac{\epsilon_0 A}{d} \qquad \text{(parallel-plate capacitor with dielectric)} \qquad (23.4)$$

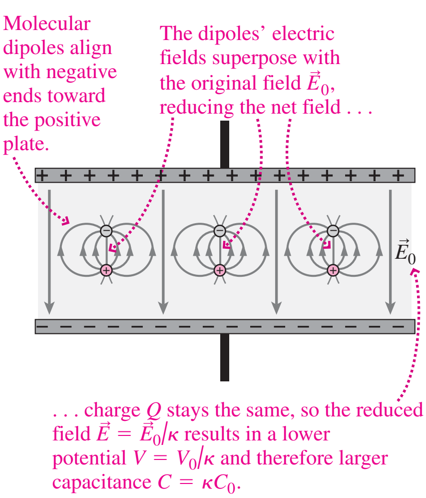


FIGURE 23.6 A capacitor with a dielectric.

Most materials have dielectric constants between about 2 and 10; see Table 23.1. Some tantalum compounds have much higher values of κ, making this rare element a crucial material in today's electronic age.

Table 23.1 Properties of Some Common Dielectrics

Dielectric Material	Dielectric Constant	Breakdown Field (MV/m)
Air	1.0006	3
Aluminum oxide	8.4	670
Glass (Pyrex)	5.6	14
Paper	3.5	14
Plexiglas	3.4	40
Polyethylene	2.3	50
Polystyrene	2.6	25
Quartz	3.8	8
Tantalum oxide	26	500
Teflon	2.1	60
Water	80	depends on time and purity

Another practical consideration is a capacitor's **working voltage**, the maximum safe potential difference, beyond which there's a risk of dielectric breakdown. For a given material, breakdown occurs at a fixed electric field; for air it's 3 MV/m, while polyethylene breaks down at 50 MV/m. In a parallel-plate capacitor the field is $E = V/d$, so the smaller the spacing, the lower the allowed voltage before breakdown. Thus there's a trade-off between large capacitance (small d) and high working voltage (large d). Large-capacitance, high-voltage capacitors are expensive!

EXAMPLE 23.2 Finding Charge and Energy: Which Capacitor?

A 100-μF capacitor has a working voltage of 20 V, while a 1.0-μF capacitor is rated at 300 V. Which can store more charge? More energy?

INTERPRET This problem involves the charge and energy stored in capacitors, now constrained by the working voltage.

DEVELOP Equation 23.1, in the form $Q = CV$, determines the charge, and Equation 23.3, $U = \frac{1}{2}CV^2$, determines the stored energy. Setting V equal to the working voltage will give the maximum charge and energy.

EVALUATE For the charges on the two capacitors, we get from Equation 23.1,

$$Q_{100\mu\text{F}} = CV = (100\ \mu\text{F})(20\ \text{V}) = 2.0\ \text{mC}$$

and, similarly, $Q_{1\mu\text{F}} = 0.30\ \text{mC}$. The energies follow from Equation 23.3:

$$U_{100\mu\text{F}} = \tfrac{1}{2}CV^2 = \tfrac{1}{2}(100\ \mu\text{F})(20\ \text{V})^2 = 20\ \text{mJ}$$

and, similarly, $U_{1\mu\text{F}} = \frac{1}{2}(1.0\ \mu\text{F})(300\ \text{V})^2 = 45\ \text{mJ}$. So the 100-$\mu$F capacitor stores more charge, but the 1-μF capacitor stores more energy.

ASSESS Make sense? The larger capacitor holds more *charge*, despite its lower working voltage. But the *energy* depends on V^2, so the smaller capacitor wins out because of its much higher working voltage. ■

GOT IT? 23.1 You need to replace a capacitor with one that can store more energy. Which will give you greater energy increase: (a) a capacitor with twice the capacitance and the same working voltage as the old one, or (b) a capacitor with the same capacitance but twice the working voltage?

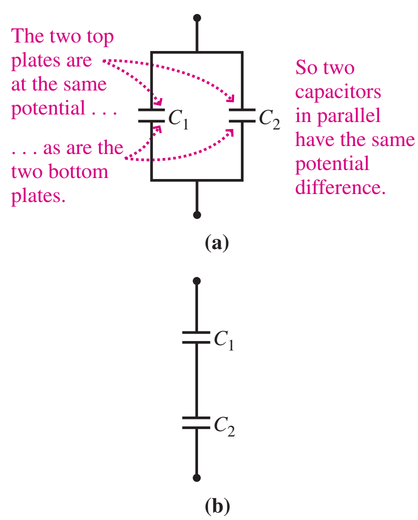


FIGURE 23.7 Connecting capacitors: (a) parallel and (b) series. is the standard circuit symbol for a capacitor.

Connecting Capacitors: Parallel

Connecting capacitors together lets us achieve capacitance or working voltage that might not be available in a single capacitor. There are two simple ways to connect capacitors and other electronic components: **parallel** and **series** (Fig. 23.7).

With capacitors in parallel, a conducting wire connects the top plates of each capacitor and another connects the bottom plates. Therefore, both top plates are at the same potential, and so are both bottom plates. That means **two capacitors in parallel have the same potential difference between their plates**. We'll find that's always true for electric components in parallel. We want the equivalent capacitance of the parallel combination, meaning the ratio of the total charge on both capacitors to their common voltage V. Given the definition $C = Q/V$, we can write $Q_1 = C_1V$ and $Q_2 = C_2V$. So the total charge is $Q = Q_1 + Q_2 = C_1V + C_2V$. The equivalent capacitance is then

$C = Q/V$ or $C = C_1 + C_2$. So capacitors in parallel add, a result that generalizes to any number of capacitors:

$$C = C_1 + C_2 + C_3 + \cdots \qquad \text{(parallel capacitors)} \qquad (23.5)$$

Connecting Capacitors: Series

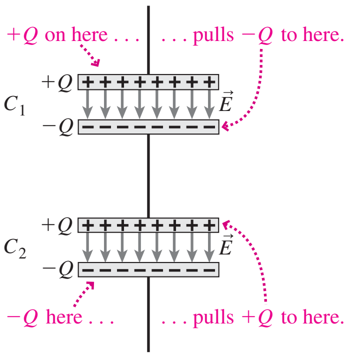


FIGURE 23.8 Capacitors in series carry the same charge.

Figure 23.8 is a closer look at the series combination of Fig. 23.7*b*, showing what happens if we put charge $+Q$ on the top plate of C_1 and charge $-Q$ on the lower plate of C_2. Each of these charged plates pulls the opposite charge to the other plate of the individual capacitors—and that means **two capacitors in series carry the same charge**. But now the voltages can be different; they're given by Equation 23.1 as $V_1 = Q/C_1$ and $V_2 = Q/C_2$, where Q is the common charge. Since the electric fields in the two capacitors point the same way, the voltage across the series combination is $V = V_1 + V_2 = Q/C_1 + Q/C_2$. Dividing through by Q gives V/Q, which is the inverse of the equivalent capacitance Q/V. Thus

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2}$$

More generally, capacitors in series add reciprocally:

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots \qquad \text{(series capacitors)} \qquad (23.6a)$$

With two capacitors it's straightforward to invert Equation 23.6a to get

$$C = \frac{C_1 C_2}{C_1 + C_2} \qquad (23.6b)$$

Either way, the combined capacitance is less than the individual capacitances.

CONCEPTUAL EXAMPLE 23.1 Parallel and Series Capacitors

Using parallel-plate capacitors, explain why capacitance should increase with capacitors in parallel and decrease with capacitors in series. What happens to the working voltage in each case?

EVALUATE Equation 23.2 shows that capacitance increases with increasing plate area and decreases with increasing plate separation. Figure 23.7*a* shows that two capacitors in parallel have greater area, with no change in spacing, so the combined capacitance increases. In contrast, the series combination in Figure 23.7*b* effectively increases the plate separation because it's the sum of the individual separations, so the capacitance goes down.

What about working voltage? In Fig. 23.7*a*, the parallel capacitors have the same voltage, so the working voltage of the combination is that of whichever capacitor has the lower working voltage. But in Fig. 23.7*b*, each series capacitor gets less than the total voltage, so the working voltage increases. How much depends on the ratio of the capacitances.

ASSESS Series and parallel combinations let us build arbitrary capacitances and working voltages from standard capacitors available commercially. You might wonder about the wire connecting the series capacitors in Fig. 23.7*b*: Does it also affect the separation? No, because it doesn't separate charge, which is free to move along the conducting wire.

MAKING THE CONNECTION You've got two 10-μF capacitors rated at 15 V. What are the capacitances and working voltages of their parallel and series combinations?

EVALUATE Applying Equation 23.5 to equal capacitors shows that the capacitance doubles with two capacitors in parallel. So the parallel combination has $C = 20\ \mu$F, and its working voltage is still 15 V because each capacitor gets the full voltage. Apply Equation 23.6b to equal capacitances and you'll see that the series capacitance is half that of either capacitor, in this case 5 μF. Since the individual capacitances are equal, each must get half the applied voltage, giving the combination a working voltage of 30 V.

GOT IT? 23.2 You have two identical capacitors with capacitance C. How would you connect them to get equivalent capacitances (a) $2C$ and (b) $\frac{1}{2}C$? Which combination would have the higher working voltage?

EXAMPLE 23.3 Equivalent Capacitance: Connecting Capacitors

Find the equivalent capacitance of the combination shown in Fig. 23.9*a*. If the maximum voltage to be applied between points *A* and *B* is 100 V, what should be the working voltage of C_1?

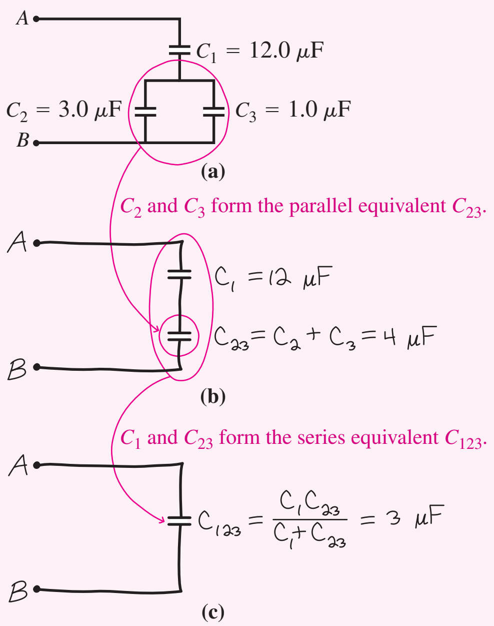


FIGURE 23.9 Finding the equivalent capacitance.

INTERPRET This problem is about an electric circuit—in this case, an assemblage of three capacitors.

DEVELOP To handle such circuit problems, we find combinations of series and parallel components, and then simplify the circuit by treating each combination as a single component. Here all components are capacitors, and each time we compute an equivalent capacitance for two capacitors, we'll redraw the circuit with the new equivalent capacitance. We begin by noting that C_2 and C_3 are in parallel, so the equivalent capacitance is given by Equation 23.5: $C_{23} = C_2 + C_3 = 4.0\ \mu\text{F}$. In Fig. 23.9*b* we've redrawn the original circuit showing this combination of the two individual capacitors. Next we see that C_1 is in series with C_{23}, so their equivalent capacitance follows from Equation 23.6b:

$$C_{123} = \frac{C_1 C_{23}}{C_1 + C_{23}} = \frac{(12\ \mu\text{F})(4.0\ \mu\text{F})}{12\ \mu\text{F} + 4.0\ \mu\text{F}} = 3.0\ \mu\text{F}$$

We've redrawn the circuit again with this equivalent capacitance (Fig. 23.9*c*).

Now we want the working voltage of C_1. Our plan is to go backward from the simple circuit of Fig. 23.9*c* until we have enough information to find the voltage on C_1. With the voltage V_{AB} across *A* and *B* known, we can calculate the charge on C_{123} using Equation 23.1: $Q_{123} = C_{123}V_{AB}$. But C_{123} is the series combination of C_1 and C_{23}, and we know that series capacitors carry the same charge—and that's the charge of their equivalent capacitance. So $Q_1 = Q_{123}$, and we can apply Equation 23.1 again, this time to C_1 alone, to get $V_1 = Q_1/C_1$.

EVALUATE With $V_{AB} = 100$ V across the combination C_{123}, the corresponding charge is $Q_{123} = C_{123}V_{AB} = (3.0\ \mu\text{F})(100\ \text{V}) = 300\ \mu\text{C}$. Because $Q_1 = Q_{123}$, the charge on C_1 is also 300 μC. We then substitute this into $V_1 = Q_1/C_1$ to get $V_1 = (300\ \mu\text{C})/(12.0\ \mu\text{F}) = 25$ V, the minimum working voltage for C_1.

ASSESS Make sense? Since C_1 is in series with C_{23}, it doesn't "feel" the full 100 V applied across *AB*, so its working voltage can be lower. And because its capacitance is *larger*, its share of the voltage is *smaller*, thanks to the relation $V = Q/C$ and the fact that series capacitors carry the *same charge*.

✓TIP Series and Parallel

Parallel components have their ends connected directly together; series components are connected in such a way that if you move through one component, the only place you can go is into the next. In Fig. 23.9*a*, C_2 and C_3 are definitely in parallel. But C_1 isn't in series with either of the other single capacitors because after C_1, the circuit splits and you could go into either C_2 or C_3. Equations 23.5 and 23.6 apply *only* to true parallel and series combinations. C_1 *is* in series with the combination C_{23}, so we could apply Equation 23.6b in analyzing Fig. 23.9*b*.

APPLICATION Bursts of Power

As San Francisco's BART trains decelerate, their kinetic energy is stored as electric energy in an ultracapacitor. The stored energy is then used to accelerate the train. This system saves BART some 320 megawatt-hours of energy each year.

Capacitors are excellent devices for short-term storage of electric energy because they can deliver their stored energy very quickly—much faster than a battery that might contain a lot more total energy.

When you use a flash camera, you have to wait a few seconds before the flash is ready to fire again. That's because the flash requires power—energy per time—far greater than the camera's battery could supply. So the battery gradually charges a capacitor, whose energy is then dumped abruptly to power the brief flash. It takes a while to recharge the capacitor before it's ready again. Much the same thing happens in a defibrillator, which delivers several hundred joules to restore a heart's normal beating. Again, the energy is stored in capacitors, which discharge in milliseconds. On a much larger scale, whole rooms full of capacitors store the energy that drives nanosecond laser pulses pouring millions of joules into tiny targets in experiments aimed at making nuclear fusion a viable energy source. And increasingly, so-called ultracapacitors supply extra energy for bursts of power in machinery from amusement park rides to mass-transit trains to hybrid cars.

23.4 Energy in the Electric Field

What's the difference between a charged and an uncharged capacitor? Not the total charge, which is zero, but the arrangement of charge. And with the charge arrangement comes stored energy. Where, exactly, is this energy? We can ask the same question for the triangular charge distribution we assembled in Fig. 23.1. The individual charges didn't change, but their arrangement did. With the new arrangement came energy, but where is that energy?

What's changed in both cases is the electric field. There's no electric field in the uncharged capacitor, but once charged, there's a field between the plates. The triangular distribution started with three isolated point-charge fields and ended with a more complex field. So where's the stored energy? It's in the electric field. In fact, *every* electric field represents stored energy. Rearrange the charges to their original state—by discharging the capacitor or letting the three point charges fly apart—and you get back that energy. Because electric forces govern much of the behavior of everyday matter, many seemingly different forms of energy are actually electric. Burn gasoline or metabolize food, and you're rearranging the charge distributions we call molecules into new configurations whose electric fields contain less energy.

Since electric fields can vary with position, we specify the **energy density**, or energy stored per unit volume. For a capacitor, we can use Equation 23.1 in the form $V = Q/C$ to write the stored energy $U = \frac{1}{2}CV^2$ as $U = Q^2/2C$. For a parallel-plate capacitor, Equation 23.2 gives $C = \epsilon_0 A/d$, so the stored energy becomes $U = Q^2 d/2\epsilon_0 A$. This energy is associated with the uniform electric field inside the capacitor, where it occupies a volume Ad. So the energy density is $U/Ad = Q^2/2\epsilon_0 A^2$. We can rewrite this in terms of the field, which we found to be $E = Q/\epsilon_0 A$. Then $Q = \epsilon_0 AE$, and the energy density becomes

$$u_E = \tfrac{1}{2}\epsilon_0 E^2 \qquad \text{(electric energy density)} \tag{23.7}$$

Although we derived Equation 23.7 for the uniform field of a parallel-plate capacitor, it is in fact universal. Anywhere there's an electric field, there's also stored energy with density, in J/m^3, given by $\frac{1}{2}\epsilon_0 E^2$. That's the deep significance of Equation 23.7: *Every* electric field represents stored energy. The energy that drives much of the physical universe, from everyday events here on Earth to happenings in distant galaxies, results from the release of energy stored in electric fields.

EXAMPLE 23.4 Electric Energy: A Thunderstorm

Typical electric fields in thunderstorms average around 10^5 V/m. Consider a cylindrical thundercloud with height 10 km and diameter 20 km, and assume a uniform electric field of 1×10^5 V/m. Find the electric energy contained in this cloud.

INTERPRET This problem is about stored electric energy.

DEVELOP Since the field and hence the energy density are uniform, our plan is to find the energy density and then multiply it by the cloud's cylindrical volume to calculate the total electric energy. We'll use Equation 23.7, $u_E = \frac{1}{2}\epsilon_0 E^2$, for the energy density.

EVALUATE The energy density is

$$u_E = \tfrac{1}{2}\epsilon_0 E^2 = \tfrac{1}{2}\epsilon_0 (1\times10^5\ \text{V/m})^2 = 4.4\times10^{-2}\ \text{J/m}^3$$

The cylindrical cloud has volume

$$V = \pi r^2 h = \pi(10\ \text{km})(10\ \text{km})^2 = 3.1\times10^{12}\ \text{m}^3$$

Multiplying energy density by volume gives the total stored energy:

$$U = u_E V = (4.4\times10^{-2}\ \text{J/m}^3)(3.1\times10^{12}\ \text{m}^3) = 140\ \text{GJ}$$

ASSESS Make sense? A gallon of gasoline contains about 0.1 GJ (see Appendix C), so the thundercloud stores the energy equivalent of about 1400 gallons of gasoline. That's not a whole lot for such a vast volume, showing that the energy density of macroscopic electric fields can't compare with the electric energy density locked into the molecular structure of a fuel. You'll never see cars running on the energy stored in atmospheric electric fields! ■

When the electric field is uniform, as in our thundercloud, the total energy is the product of energy density and volume. But when the field changes with position, we need calculus. Consider a small volume dV, so small that the electric field is essentially uniform over this volume. The stored energy is then $dU = u_E\, dV = \frac{1}{2}\epsilon_0 E^2\, dV$. The total energy in the field is the sum—here the integral—of all the dU values:

$$U = \tfrac{1}{2}\epsilon_0 \int E^2\, dV \qquad (23.8)$$

Because Equation 23.8 gives the energy stored in an electric field, it also represents the work done in assembling the charge distribution resulting in that field. The next example illustrates this point.

EXAMPLE 23.5 Work and Energy: A Shrinking Sphere

A sphere of radius R_1 carries charge Q distributed uniformly over its surface. How much work does it take to compress the sphere to a smaller radius R_2?

INTERPRET This problem asks for the work done in rearranging a charge distribution, which we know is equal to the change in stored electric energy. Here we start with a charged sphere already assembled, and rearrange the charge by shrinking the sphere to a smaller radius.

DEVELOP We have spherical symmetry, so the field and thus the stored energy outside the original radius R_1 don't change. Therefore, we need to find the energy stored in the new field created when the sphere shrinks. Figure 23.10 is our sketch of the situation before and after the sphere shrinks. Here the field varies with position, so Equation 23.8,

$$U = \tfrac{1}{2}\epsilon_0 \int E^2\, dV$$

gives the stored energy. Our plan is to evaluate the field in the region $R_2 < r < R_1$ and use the result in Equation 23.8. Given the spherical symmetry, the new field is a point-charge field: $E = kQ/r^2$. To use Equation 23.8 we need an appropriate volume element dV. With spherical symmetry, Fig. 23.11 shows that we can use a thin spherical shell of volume $dV = 4\pi r^2\, dr$. Then Equation 23.8 becomes

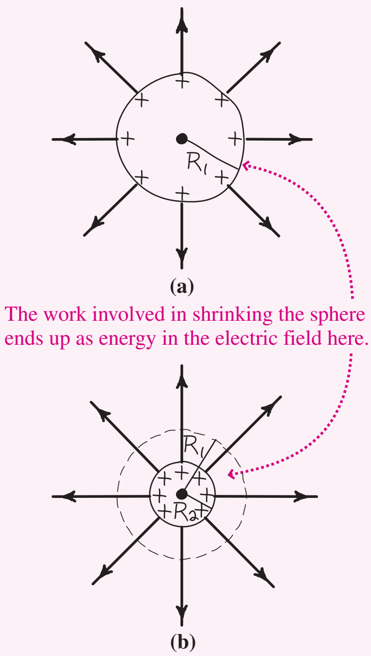


FIGURE 23.10 (a) A charged sphere and its electric field. (b) Shrinking the sphere creates field and energy in the region $R_2 < r < R_1$.

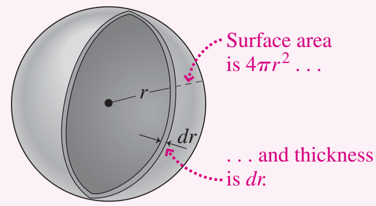


FIGURE 23.11 A thin spherical shell has volume $dV = 4\pi r^2\, dr$.

$$U = \tfrac{1}{2}\epsilon_0 \int E^2\, dV = \tfrac{1}{2}\epsilon_0 \int_{R_2}^{R_1} \left(\frac{kQ}{r^2}\right)^2 4\pi r^2\, dr = \frac{kQ^2}{2}\int_{R_2}^{R_1} r^{-2}\, dr$$

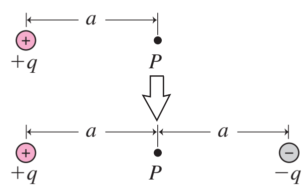
where we substituted $1/4\pi k$ for ϵ_0.

EVALUATE The integral is $\int r^{-2}\, dr = \dfrac{r^{-1}}{-1} = -\dfrac{1}{r}$, so

$$U = \frac{kQ^2}{2}\left(-\frac{1}{r}\right)\Bigg|_{R_2}^{R_1} = \frac{kQ^2}{2}\left(\frac{1}{R_2} - \frac{1}{R_1}\right)$$

ASSESS Make sense? Here $R_2 < R_1$, so the stored energy is positive and indicates that this much work had to be done to shrink the sphere. Of course: The sphere carries charge of the same sign, and shrinking it moves that charge closer together, against the repulsive electric force. Letting R_1 go to infinity gives the work needed to assemble a spherical surface charge distribution. Putting $R_2 = 0$ makes the work and therefore the stored energy infinite—suggesting that the notion of a point charge is an impossible idealization. Problem 68 explores some implications of this result in the theory of elementary particles. ■

GOT IT? 23.3 You're at a point P a distance a from a point charge $+q$. You then place a point charge $-q$ a distance a on the opposite side of P as shown. What happens to (a) the electric field strength and (b) the electric energy density at P? (c) Does the total electric energy $U = \int u_E\, dV$ of the entire field increase, decrease, or remain the same?



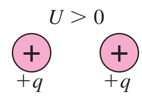
CHAPTER 23 SUMMARY

Big Picture

The big idea here is that *all* electric fields represent stored energy. This energy is associated with the work needed to assemble a distribution of electric charge, and may be negative or positive.

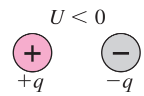
You do positive work to assemble this charge distribution . . .

. . . and therefore the stored electric energy U is positive.



You do negative work to assemble this charge distribution . . .

. . . and therefore the stored electric energy U is negative.

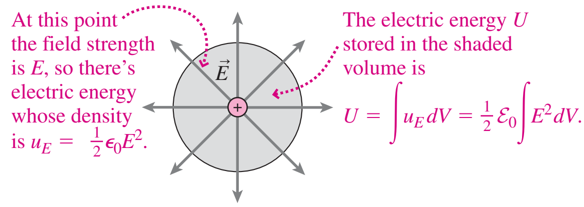


Key Concepts and Equations

The **energy density** in an electric field E is $u_E = \frac{1}{2}\epsilon_0 E^2$.

Integrating over volume gives the total electric energy U stored in the field:

$$U = \int u_E \, dV.$$



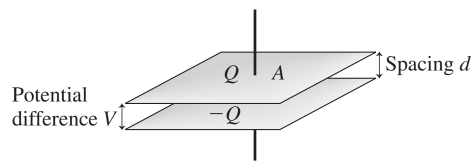
Applications

A **capacitor** is a pair of insulated conductors used to store electric energy. **Capacitance** is the ratio of charge to potential difference:

$$C = Q/V$$

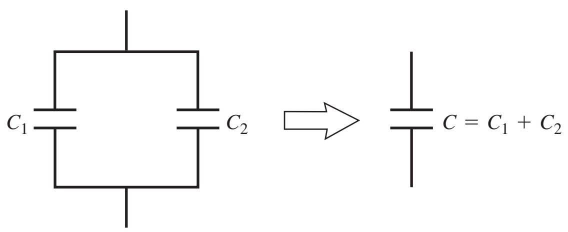
For a parallel-plate capacitor:

$$C = \epsilon_0 A/d$$



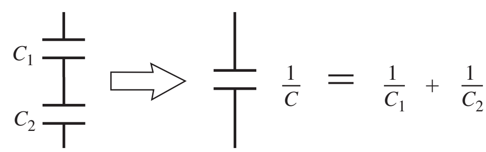
The energy stored in a capacitor is $U = \frac{1}{2}CV^2$.

Capacitors in parallel add: $C = C_1 + C_2$.



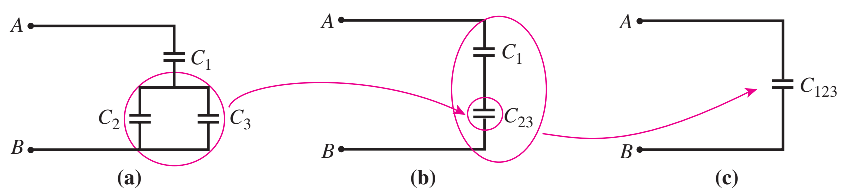
Capacitors in parallel have the same voltage.

Capacitors in series add reciprocally: $\dfrac{1}{C} = \dfrac{1}{C_1} + \dfrac{1}{C_2}$.



Capacitors in series have the same charge.

Complicated circuits are analyzed by breaking them into parallel and series combinations:



A **dielectric** between capacitor plates increases the capacitance, as determined by the **dielectric constant** κ of the material: $C \to \kappa C_0$.

For Thought and Discussion

1. Two positive point charges are infinitely far apart. Is it possible, using a finite amount of work, to move them until they're a small distance d apart?
2. How does the energy density at a certain distance from a negative point charge compare with the energy density at the same distance from a positive point charge of equal magnitude?
3. A dipole consists of two equal but opposite charges. Is the total energy stored in the dipole's electric field zero? Why or why not?
4. Charge is spread over the surface of a balloon, which is then allowed to expand. What happens to the energy of the electric field?
5. Does the superposition principle hold for electric-field energy densities? That is, if you double the field strength at some point, do you double the energy density as well?
6. A student argues that the total energy associated with the electric field of a charged sphere must be infinite because its field extends throughout an infinite volume. Critique this argument.
7. A capacitor is said to carry a charge Q. What's the net charge on the entire capacitor?
8. Does the capacitance describe the maximum amount of charge a capacitor can hold, in the same way that a bucket's capacity describes the maximum amount of water it can hold? Explain.
9. Is a force needed to hold the plates of a charged capacitor in place? Explain.
10. A solid conducting slab is inserted between the plates of a capacitor, not touching either plate. Does the capacitance increase, decrease, or remain the same?
11. Two capacitors contain equal amounts of energy, yet one has twice the capacitance. How do their voltages compare?
12. A parallel-plate capacitor is connected to a battery that imposes a potential difference V between its plates. If a dielectric slab is inserted between the plates, what happens to (a) the potential difference, (b) the capacitor charge, and (c) the capacitance?

Exercises and Problems

Exercises

Section 23.1 Electrostatic Energy

13. Four 50-μC charges, initially far apart, are brought onto a line where they're spaced at 2.0-cm intervals. How much work does it take to assemble this charge distribution?
14. Three point charges $+q$ and a fourth, $-\frac{1}{2}q$, are assembled to form a square of side a. Find an expression for the electrostatic energy of this charge distribution.
15. Repeat Exercise 14 for the case when the fourth charge is $-q$.
16. If the three particles in Fig. 23.1 have identical charge q and mass m, and if they're released from their positions on the triangle, what speed v will they have when they're far away?
17. A crude model of the water molecule has a negatively charged oxygen atom and two protons, as shown in Fig. 23.12. Calculate the electrostatic energy of this configuration, which is therefore the magnitude of the energy released in forming this molecule. (*Note*: Your answer is an overestimate because electrons are actually "shared" among the three atoms, spending more time near the oxygen.)

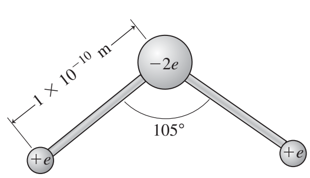


FIGURE 23.12 Exercise 17

Section 23.2 Capacitors

18. A capacitor consists of square conducting plates 25 cm on a side and 5.0 mm apart, carrying charges ± 1.1 μC. Find (a) the electric field, (b) the potential difference between the plates, and (c) the stored energy.
19. An uncharged capacitor has parallel plates 5.0 cm on a side, spaced 1.2 mm apart. (a) How much work is required to transfer 7.2 μC from one plate to the other? (b) How much work is required to transfer an additional 7.2 μC?
20. (a) How much charge must be transferred between the initially uncharged plates of Exercise 19 in order to store 15 mJ of energy? (b) What will be the resulting potential difference between the plates?
21. A capacitor's plates hold 1.3 μC when charged to 60 V. What's its capacitance?
22. The "memory" capacitor in a video recorder stores program recording information during power outages. It has capacitance 4.0 F and is charged to 3.5 V. What's the charge on its plates?
23. What voltage is needed to put 1.6 mC on a 100-μF capacitor?
24. Show that the units of ϵ_0 may be written as F/m.
25. Find the capacitance of a parallel-plate capacitor with circular plates 20 cm in radius separated by 1.5 mm.
26. A parallel-plate capacitor with 1.1-mm plate spacing has ± 2.3 μC on its plates when charged to 150 V. What's the plate area?
27. The power supply in a stereo receiver contains a 2500-μF capacitor charged to 35 V. How much energy does it store?
28. Find the capacitance of a capacitor that stores 350 μJ when the potential difference across its plates is 100 V.

Section 23.3 Using Capacitors

29. You have a 1.0-μF and a 2.0-μF capacitor. What capacitances can you get by connecting them in series or in parallel?
30. Two capacitors are connected in series and the combination is charged to 100 V. If the voltage across each capacitor is 50 V, how do their capacitances compare?
31. (a) Find the equivalent capacitance of the combination shown in Fig. 23.13. Find (b) the charge and (c) the voltage on each capacitor when a 100-V battery is connected across the combination.

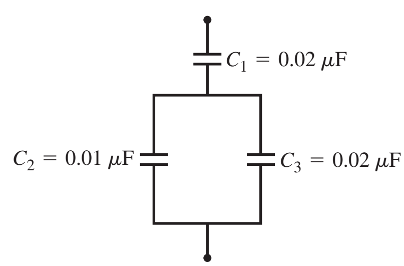


FIGURE 23.13 Exercise 31

32. You're given three capacitors: 1.0 μF, 2.0 μF, and 3.0 μF. Find (a) the maximum, (b) the minimum, and (c) two intermediate capacitances you could achieve using combinations of all three capacitors.

Section 23.4 Energy in the Electric Field

33. The energy density in a uniform electric field is 3.0 J/m^3. What's the field strength?
34. A car battery stores about 4 MJ of energy. If this energy were used to create a uniform 30-kV/m electric field, what volume would it occupy?
35. Air undergoes dielectric breakdown at a field strength of 3 MV/m. Could you store energy in an electric field in air with the same energy density as gasoline? (*Hint:* See Appendix C.)
36. Consider a proton to be a uniformly charged sphere 1 fm in radius. Find the electric energy density at the proton's surface.

Problems

37. A charge Q_0 is at the origin. A second charge, $Q_x = 2Q_0$, is brought from infinity to the point $x = a, y = 0$. Then a third charge Q_y is brought from infinity to $x = 0, y = a$. If it takes twice as much work to bring in Q_y as it did Q_x, what's Q_y in terms of Q_0?
38. A conducting sphere of radius a is surrounded by a concentric spherical shell of radius b. Both are initially uncharged. How much work does it take to transfer charge from one to the other until they carry charges $\pm Q$?
39. Two closely spaced square conducting plates measure 10 cm on a side. The electric-field energy density between them is 4.5 kJ/m^3. What's the charge on the plates?
40. A capacitor consists of two long concentric metal cylinders (Fig. 23.14). Find an expression for its capacitance in terms of the dimensions shown.

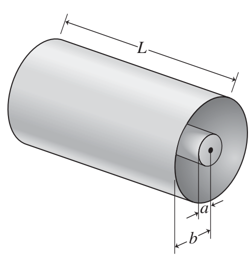


FIGURE 23.14 Problem 40

41. A capacitor consists of a conducting sphere of radius a surrounded by a concentric conducting shell of radius b. Show that its capacitance is $C = ab/k(b - a)$.
42. **BIO** The potential difference across a cell membrane is 65 mV. On the outside are 1.5×10^6 singly ionized potassium atoms. Assuming an equal negative charge on the inside, find the membrane's capacitance.
43. A capacitor stores 40 mJ of energy when charged to 100 V. (a) How much would it store at 25 V? (b) What's its capacitance?
44. Which can store more energy: a 1.0-μF capacitor rated at 250 V or a 470-pF capacitor rated at 3 kV?
45. As an electrical engineer, you're asked to specify a capacitor that can store 12 mJ of energy. The largest capacitor that will physically fit on your circuit board is 10 μF. The manufacturer produces capacitors with voltage ratings in multiples of 25 V. What voltage do you specify?
46. A 0.01-μF, 300-V capacitor costs 25¢; a 0.1-μF, 100-V capacitor costs 35¢; and a 30-μF, 5-V capacitor costs 88¢. (a) Which can store the most charge? (b) Which can store the most energy? (c) Which is the most cost-effective energy-storage device, measured in J/¢?
47. **BIO** A medical defibrillator stores 950 J in a 100-μF capacitor. (a) What is the voltage across the capacitor? (b) If the capacitor discharges 300 J of its stored energy in 2.5 ms, what's the power delivered during this time?
48. A camera requires 5.0 J of energy for a flash lasting 1.0 ms. (a) What power does the flashtube use *while it's flashing*? (b) If the flashtube operates at 200 V, what size capacitor is needed to supply the flash energy? (c) If the flashtube is fired once every 10 s, what's its *average* power consumption?
49. What is the equivalent capacitance of the four identical capacitors in Fig. 23.15, measured between A and B?

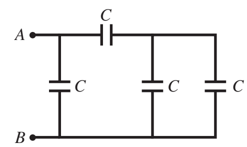


FIGURE 23.15 Problem 49

50. Your company's purchasing department bought lots of cheap 2.0-μF, 50-V capacitors. Your budget is maxed out and they won't let you buy additional capacitors for a circuit you're designing. You need 2.0-μF, 100-V capacitors and 0.5-μF, 50-V capacitors. How will you combine the available capacitors to make these?
51. What's the equivalent capacitance measured between A and B in Fig. 23.16?

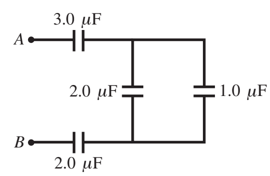


FIGURE 23.16 Problems 51 and 52

52. In Fig. 23.16, find the energy stored in the 1-μF capacitor when a 50-V battery is connected between A and B.
53. Capacitors C_1 and C_2 are in series, with voltage V across the combination. Show that the voltages across the individual capacitors are $V_1 = C_2V/(C_1 + C_2)$ and $V_2 = C_1V/(C_1 + C_2)$.
54. You're evaluating a new hire in your company's engineering department. Together you're working on a circuit where a 0.1-μF, 50-V capacitor is in series with a 0.2-μF, 200-V capacitor. The new engineer claims you can safely put 250 V across the combination. What do you say?
55. A parallel-plate capacitor has plates with area 50 cm^2 separated by 25 μm of polyethylene. Find its (a) capacitance and (b) working voltage.
56. A 470-pF capacitor consists of two 15-cm-radius circular plates, insulated with polystyrene. Find (a) the thickness of the polystyrene and (b) the capacitor's working voltage.
57. **BIO** The first accurate estimate of cell membrane thickness used a capacitive technique, which determined the capacitance per unit area of cell membrane in a macroscopic suspension of cells; the result was about 1 μF/cm^2. Assuming a dielectric constant of about 3 for the membrane, find the membrane's thickness. (*Note:* Your answer is the thickness of the bipolar lipid layer alone, and is lower by a factor of about 3 than values based on X-ray techniques.)
58. Your company is still stuck with those 2-μF capacitors from Problem 50. They turn out to be so cheap that their capacitances are all too low, ranging from 1.7 μF to 1.9 μF. A colleague suggests you put variable "trimmer" capacitors in parallel with

the cheap capacitors and adjust the combination to precisely 2.00 μF. The available trimmers have variable capacitance from 25 nF to 350 nF. Will they work?

59. A cubical region 1.0 m on a side is located between $x = 0$ and $x = 1$ m. The region contains an electric field whose magnitude varies with x but is independent of y and z: $E = E_0(x/x_0)$, where $E_0 = 24$ kV/m and $x_0 = 6.0$ m. Find the total energy in the region.
60. A sphere of radius R contains charge Q spread uniformly throughout its volume. Find an expression for the electrostatic energy contained within the sphere itself. (*Hint:* Consult Example 21.1.)
61. A sphere of radius R carries total charge Q distributed uniformly over its surface. Show that the energy stored in its electric field is $U = kQ^2/2R$.
62. A uranium-235 nucleus has diameter 6.6 fm and contains 92 protons and 143 neutrons. Assuming that charge is distributed uniformly throughout the nucleus, use the results of Problems 60 and 61 to calculate the total electrostatic energy of this configuration.
63. Two widely separated 4.0-mm-diameter water drops each carry 15 nC. Assuming all charge resides on the drops' surfaces, find the change in electrostatic potential energy if they're brought together to form a single spherical drop.
64. A 2.1-mm-diameter wire carries a uniform line charge density $\lambda = 28\ \mu$C/m. Find the energy in a region 1.0 m long within one wire diameter of the wire surface.
65. A typical lightning flash transfers 30 C across a potential difference of 30 MV. Assuming such flashes occur every 5 s in the thunderstorm of Example 23.4, roughly how long would the storm last if its electric energy were not replenished?
66. Show that the result of Problem 41 reduces to that of a parallel-plate capacitor when the separation $b - a$ is much less than the radius a.
67. A solid sphere contains a uniform volume charge density. What fraction of the total electrostatic energy of this configuration is contained *within* the sphere?
68. A classical view of the electron pictures it as a purely electric entity, whose Einstein rest mass energy, $E = mc^2$, is the energy stored in its electric field. If the electron were a sphere with charge distributed uniformly over its surface, what radius would it have in order to satisfy this condition? (*Note:* Your answer, and the picture of the electron as a sphere, aren't consistent with quantum theory.)
69. An air-insulated parallel-plate capacitor of capacitance C_0 is charged to voltage V_0 and then disconnected from the charging battery. A slab with dielectric constant κ and thickness equal to the capacitor spacing is then inserted halfway into the capacitor

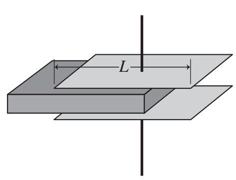


FIGURE 23.17 Problems 69 and 70

(Fig. 23.17). Determine (a) the new capacitance, (b) the stored energy, and (c) the force on the slab in terms of C_0, V_0, κ, and the plate length L.

70. Repeat parts (b) and (c) of Problem 69, now assuming the battery remains connected while the slab is inserted.
71. A transmission line consists of two parallel wires, of radius a and separation b, carrying uniform linear charge densities $\pm\lambda$, respectively. With $a \ll b$, their electric field is the superposition of the fields from two long straight lines of charge. Find the capacitance per unit length for this transmission line.
72. An infinitely long rod of radius R carries uniform volume charge density ρ. Find an expression for the electrostatic energy per unit length contained *within* the rod. (*Hint:* See Problem 21.54.)
73. (a) Write the electrostatic potential energy of a pair of oppositely charged, closely spaced parallel plates as a function of their separation x, their area A, and the charge magnitude Q. (b) Differentiate with respect to x to find the magnitude of the attractive force between the plates. Why isn't the force equal to the charge on one plate times the electric field between the plates?
74. An unknown capacitor C is connected in series with a 3.0-μF capacitor; this pair is placed in parallel with a 1.0-μF capacitor, and the entire combination is put in series with a 2.0-μF capacitor. (a) Make a circuit diagram of this network. (b) When a potential difference of 100 V is applied across the open ends of the network, the total energy stored in all the capacitors is 5.8 mJ. Find C.

Passage Problems

Nuclear fusion could provide humankind with limitless energy, making a gallon of seawater the energy equivalent of 300 gallons of gasoline. The National Ignition Facility (NIF) at Lawrence Livermore National Laboratory was designed for the "ignition" of nuclear fusion by bombarding a tiny deuterium-tritium pellet with energy from 192 converging laser beams. The NIF lasers deliver 2 MJ of energy in about 1 ns; Fig. 23.18 shows the target chamber where the laser beams converge. The energy is stored in capacitors that, because of conversion inefficiencies, have to store some 400 MJ. (*Note:* NIF is more complicated than described here, and the numbers and technical descriptions are only approximate.)

FIGURE 23.18 The NIF target chamber, shown during installation (Passage Problems 75–78)

75. What total capacitance is required if the capacitor system is charged to 20 kV?
    a. 100 μF
    b. 200 μF
    c. 1 F
    d. 2 F
76. If it were technically and economically feasible to double the voltage, how would the required capacitance change?
    a. drop to 1/4 its original value
    b. drop to 1/2 its original value
    c. would not change
    d. would double

77. While they're firing, the power delivered by the laser beams is
   a. 2 MW $(2\times10^6\ \mathrm{W})$.
   b. 2 GW $(2\times10^9\ \mathrm{W})$.
   c. 2 TW $(2\times10^{12}\ \mathrm{W})$.
   d. 2 PW $(2\times10^{15}\ \mathrm{W})$.
78. Among the capacitors that store energy at NIF are 1200 300-μF units charged to about 20 kV. The energy stored in each capacitor is about
   a. 3 J.
   b. 20 kJ.
   c. 60 kJ.
   d. 400 MJ.

Answers to Chapter Questions

Answer to Chapter Opening Question

The energy is stored in the electric field of a pair of charged conductors—a capacitor—and dumped quickly to the defibrillator when it's needed.

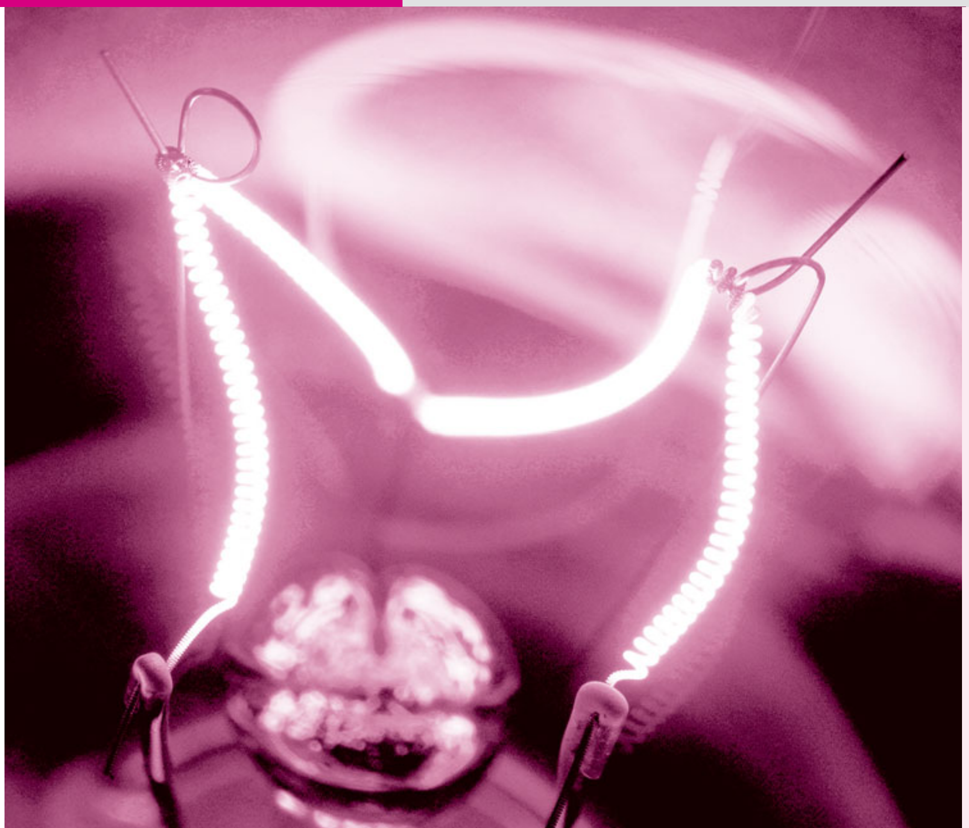
Answers to GOT IT? Questions

23.1. (b), because U depends on V^2.

23.2. (a) parallel; (b) series. The working voltage of the series combination is twice that of the parallel combination, which is the same as that of the individual capacitors.

23.3. (a) $E(P)$ doubles; (b) $u_E(P)$ quadruples; (c) U decreases, since the charges are attracted and therefore you do negative work to bring in the negative charge.

24 Electric Current



How does electric current heat this lightbulb filament? Where does the energy come from?

New Concepts, New Skills

By the end of this chapter you should be able to

- Describe quantitatively electric current and current density in terms of the microscopic properties characterizing the flow of electric charge (24.1).
- Explain the mechanisms of electrical conduction in different materials, especially metals (24.2).
- Describe electrical resistance, and relate current, voltage, and resistance using Ohm's law (24.3).
- Calculate electric power (24.4).
- Use electricity safely (24.5).

Connecting Your Knowledge

- This chapter draws on the concepts of electric charge and conductors (20.1, 20.5).
- We'll look again at potential difference (22.1, 22.3).
- We'll be venturing away from electrostatic equilibrium, so we'll revisit the behavior of charge in conductors (21.6).

We now move beyond electrostatic equilibrium and consider situations in which charges are moving. The flow of charge constitutes **electric current**, and it occurs in materials containing free charges—that is, in conductors.

Electric current is essential in many technological and natural processes. Currents in lightbulbs, toasters, and stoves produce heat and light. Currents in electric motors run refrigerators, hybrid cars, and subway trains. In computers, currents move and process data. In your body, they regulate heartbeat and control muscles. Currents in Earth's liquid outer core generate the planet's magnetism, protecting us from cosmic radiation. And currents in the Sun are responsible for giant eruptions that can spew high-energy particles toward Earth.

24.1 Electric Current

Quantitatively, current is the net rate of charge crossing an area. Its units are coulombs per second, which is given the name **ampere** (A) after the French physicist André Marie Ampère (1775–1836). In electronics and biomedical applications, currents are small enough that milliamperes (mA) and microamperes (μA) are widely used. When current I is steady or a time average will do, we write

$$I = \frac{\Delta Q}{\Delta t} \qquad \text{(steady current)} \tag{24.1a}$$

where ΔQ is the charge crossing the given area in time Δt. For time-varying currents we take the limit of small time intervals:

$$I = \frac{dQ}{dt} \quad \text{(instantaneous current)} \qquad (24.1b)$$

Current is in the direction in which *positive* charge flows. If the moving charge is negative, as with electrons in a metal, then the current is opposite the charge motion.

A current may consist of one kind of moving charge, or both. If it's both, then the net current is the sum of the currents carried by positive and negative charges (Fig. 24.1*a*). That's why the bulk motion of a neutral object—even though it contains lots of positive and negative charge—doesn't constitute a current (Fig. 24.1*b*).

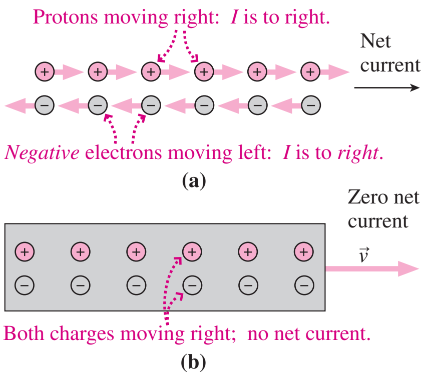


FIGURE 24.1 The net current is the sum of the currents carried by both positive and negative charges.

GOT IT? 24.1 Which of the following represents a nonzero current? What's its direction? (a) a beam of electrons moves from left to right; (b) a beam of protons moves upward; (c) in a solution, positive ions move left and negative ions move right; (d) blood, carrying positive and negative ions at the same speed, moves up through a vein; (e) a metal car with no net charge speeds westward.

Current: A Microscopic Look

Current depends on the speed of the charge carriers, their density, and their charge. In some cases, like a beam of electrons in vacuum, "speed" here means the actual speed of the charges. But in typical conductors, charges are moving about at high speed with random thermal velocities that don't result in a net flow of charge. When a current is present, there's an additional and usually very small **drift velocity** superposed on the charges' random motion, and it's this drift velocity that determines the current. We'll see this in more detail when we consider metallic conductors.

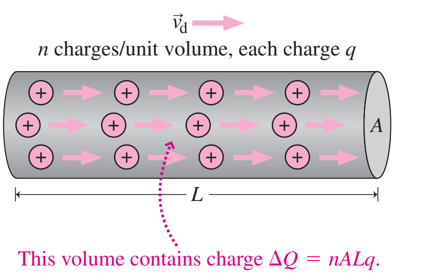


FIGURE 24.2 A conductor of cross-sectional area A containing n charges per unit volume.

Figure 24.2 shows a conductor that contains n charges per unit volume, each with charge q and drift speed v_d. We want to express the current in terms of these microscopic properties and macroscopic properties like length and area. With A the conductor's cross-sectional area, a length L of conductor has volume AL and contains nAL individual charges for a total charge $\Delta Q = nALq$. Moving at v_d, this charge takes time $\Delta t = L/v_d$ to pass a given point. Then the current is

$$I = \frac{\Delta Q}{\Delta t} = \frac{nALq}{L/v_d} = nAqv_d \qquad (24.2)$$

EXAMPLE 24.1 Finding Current: A Copper Wire

A 5.0-A current flows in a copper wire with cross-sectional area 1.0 mm^2, carried by electrons with number density $n = 1.1\times10^{29}\ \text{m}^{-3}$. Find the electrons' drift speed.

INTERPRET We're given microscopic parameters, so this is a problem about the relation between current and the parameters n, q, and v_d.

DEVELOP Figure 24.3 is our sketch. Equation 24.2, $I = nAqv_d$, relates current to the macroscopic parameters, so our plan is to solve for v_d.

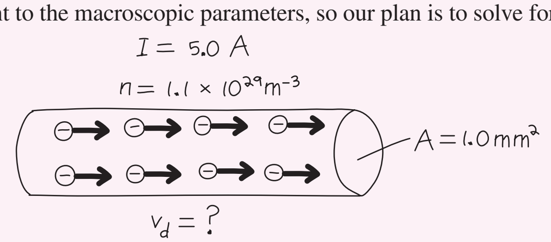


FIGURE 24.3 Sketch for Example 24.1.

EVALUATE Solving, we get

$$v_d = \frac{I}{nAq} = \frac{5.0\ \text{A}}{(1.1\times10^{29}\ \text{m}^{-3})(10^{-6}\ \text{m}^2)(1.6\times10^{-19}\ \text{C})} = 0.28\ \text{mm/s}$$

ASSESS Make sense? Our answer seems awfully small. After all, when you flip a light switch, the light comes on immediately, not several thousand seconds later as our answer might imply. But the answer is right. Electrons in the wire all get their "marching orders" from the electric field, and that's established almost instantaneously. As a result electrons throughout the wire start moving almost simultaneously. That's why the light comes on immediately. ■

✓**TIP** Drift Speed and Signal Speed

Example 24.1 points to an important distinction between the drift speeds of electrons and speed of electric signals. The former is typically about 1 mm/s, but the latter is close to the speed of light. When you connect a wire across a battery, for example, an electric field develops, starting at both battery terminals and moving along the wire at nearly the speed of light. As soon as electrons experience the field, they start to move. So there's almost no time delay before the start of the current.

Current Density

Currents aren't always confined to wires. Currents in the Earth, in chemical solutions, in your body, and in ionized gases flow in ill-defined paths, and their magnitude and direction may vary with position. We characterize such diffuse currents in terms of **current density**, $\vec{J}$, a vector whose direction at each point is that of the local current and whose magnitude is the current per unit area. Dividing Equation 24.2 by area and using the drift velocity vector $\vec{v}_d$ instead of speed v_d gives the current density:

$$\vec{J} = nq\vec{v}_d \quad \text{(current density)} \tag{24.3}$$

When the current density is uniform, as in a wire, the total current is just the product of the current density and the wire's cross-sectional area. But when the current density varies, it's necessary to integrate to find the total current; see Problem 57.

EXAMPLE 24.2 Current and Current Density: Through the Cell Membrane

Ion channels are narrow pores that allow ions to pass through cell membranes (Fig. 24.4). A particular channel has a circular cross section 0.15 nm in radius; it opens for 1 ms and passes 1.1×10^4 singly ionized potassium ions. Find both the current and the current density in the channel.

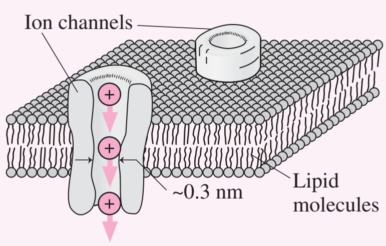


FIGURE 24.4 Diagram of a cell membrane, showing ions passing through an ion channel.

INTERPRET This problem describes a flow of individual ions and asks for two distinct but related quantities: current and current density.

DEVELOP Current is the *rate* of charge passing through a given area, here the opening of an ion channel. Equation 24.1a, $I = \Delta Q/\Delta t$, determines the current. Current density, however, is *current per unit area,* which we can compute from $J = I/A$.

EVALUATE With each ion carrying charge e, a total charge $\Delta Q = 1.1\times10^4 e = 1.8\times10^{-15}$ C flows through the channel in $\Delta t = 10^{-3}$ s, giving a current $I = \Delta Q/\Delta t = 1.8$ pA. For current density we then find

$$J = \frac{I}{A} = \frac{1.8\times10^{-12}\ \text{A}}{\pi(0.15\times10^{-9}\ \text{m})^2} = 2.5\times10^7\ \text{A/m}^2$$

ASSESS Make sense? How can something so tiny as a cell have a current density of 25 million amperes per square meter? No problem: Current density measures current *per unit area.* The ion channel is so small that the total current—1.8 picoamperes—is tiny. But that channel is impressive in its own right; its 25 MA/m^2 is about four times the maximum safe current density in typical household wiring. ■

24.2 Conduction Mechanisms

Electric fields exert forces on charges, so it's the presence of electric fields in conductors that results in electric current. Fields in conductors? Yes. With moving charge we no longer have electrostatic equilibrium, so the field inside a conductor need not be zero. Newton's law suggests that an electric field should *accelerate* free charges in a conductor, resulting in an ever-increasing current. But in most conductors charges collide, usually with ions, and lose energy they've gained from the field. These collisions provide an effective force that counters the electric force, and the end result is that it takes an electric field to sustain a steady current. In most materials the field and current are in the same direction, and we can therefore express the relation between the two as

$$\vec{J} = \sigma\vec{E} \qquad \text{(Ohm's law, microscopic version)} \tag{24.4a}$$

where the quantity σ is the material's **conductivity**.

Ohm's Law: A Microscopic View

For many common conductors, including metals, conductivity σ is independent of electric field. Such materials are called **ohmic**, and for them Equation 24.4a states that current density and electric field are linearly proportional. In **nonohmic** materials, conductivity does depend on field, and thus the relationship between $\vec{J}$ and $\vec{E}$ isn't linear.

You may be familiar with the *macroscopic* version of **Ohm's law**, which relates electric current and voltage in a piece of conducting material. Equation 24.4a is the *microscopic* version of Ohm's law, describing the relation between electric field and current density *at each point* within a conductor. The macroscopic version is helpful in analyzing electric circuits, and we'll derive it in the next section. But the microscopic version is important in biophysics, geophysics, astrophysics, semiconductor engineering, and other areas where electric fields vary with position and we want to know what's going on at each point.

Conductivity σ tells how large a current density will result from a given electric field; it's a measure of how easily charges in a material can move. A perfect conductor would have $\sigma = \infty$; a perfect insulator, $\sigma = 0$. A related quantity is **resistivity**, ρ, defined as the inverse of conductivity: $\rho = 1/\sigma$. Then Equation 24.4a can be written

$$\vec{J} = \frac{\vec{E}}{\rho} \tag{24.4b}$$

Resistivity tells how hard it is for charge to move; the higher a material's resistivity, the stronger the electric field needed to produce a given current density. You may be familiar with electrical *resistance*, and we'll soon see how resistance and resistivity are related.

Equation 24.4b shows that the units of resistivity are V·m/A. One V/A is given the name **ohm**, Ω, after the German physicist Georg Ohm (1789–1854), who explored the relation between voltage and current. Thus the SI units of resistivity are Ω·m; reciprocally, those of conductivity are $(\Omega\cdot\text{m})^{-1}$. Conductivity and resistivity range widely, spanning some 24 orders of magnitude. Table 24.1 lists the resistivities of some typical materials. Measurement of electrical resistivity provides information on the composition of materials in fields from medicine to geophysics.

Table 24.1 Resistivities

Material	Resistivity $(\Omega\cdot\text{m})$
Metallic conductors (20°C)	
Aluminum	2.65×10^{-8}
Copper	1.68×10^{-8}
Gold	2.24×10^{-8}
Iron	9.71×10^{-8}
Mercury	9.84×10^{-7}
Silver	1.59×10^{-8}
Ionic solutions (in water, 18°C)	
1-molar $CuSO_4$	3.9×10^{-4}
1-molar HCl	1.7×10^{-2}
1-molar NaCl	1.4×10^{-4}
H_2O	2.6×10^{5}
Blood, human	0.70
Seawater (typical)	0.22
Insulators	
Ceramics	10^{11}–10^{14}
Glass	10^{10}–10^{14}
Polystyrene	10^{15}–10^{17}
Rubber	10^{13}–10^{16}
Wood (dry)	10^{8}–10^{14}

EXAMPLE 24.3 Finding the Electric Field: Household Wiring

A 1.8-mm-diameter copper wire carries 15 A to a household appliance. Find the magnitude of the electric field in the wire.

INTERPRET This problem asks us to calculate the electric field within a conductor carrying an electric current.

DEVELOP Equation 24.4b, $\vec{J} = \vec{E}/\rho$, relates electric field and current density. Here we're given the total current I and the wire diameter, so we can write the current density as $J = I/A$, with A the wire's cross-sectional area. We also need the resistivity of copper from Table 24.1. Then we can solve Equation 24.4b for the electric field.

EVALUATE Solving for the field magnitude, we have

$$E = J\rho = \frac{I\rho}{A} = \frac{(15\ \text{A})(1.68\times10^{-8}\ \Omega\cdot\text{m})}{\pi(0.90\times10^{-3}\ \text{m})^2} = 99\ \text{mV/m}$$

ASSESS Make sense? This number is a lot smaller than the electric fields we discussed in electrostatic situations. Because copper is such a good conductor, a weak field can drive a substantial current. In well-engineered circuits, the field inside conducting wires is often so small as to be negligible, even when the current is large. ■

GOT IT? 24.2 Two wires carry the same current I. Wire A has a larger diameter, a higher density of current-carrying electrons, and a lower resistivity than wire B. Rank in order (a) the current densities, (b) the electric fields, and (c) the drift speeds in the two wires.

Conduction in Metals

Metals are good conductors because they contain abundant free electrons, which respond readily to electric fields. Each atom in a metal typically contributes one or more electrons to this "sea" of free electrons. The remaining ions form a regular crystal lattice (Fig. 24.5). Electrons move through the lattice at about 10^6 m/s, colliding frequently with ions and bouncing off in random directions. In the absence of an electric field, there's no net flow of electrons in any particular direction, and so no current.

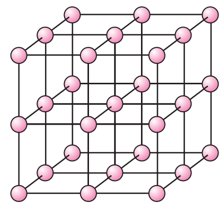


FIGURE 24.5 Atoms of a metal form a regular crystal lattice.

We'll now consider what happens when an electric field is applied to a metal, and we'll show why metals obey Ohm's law. However, our explanation is necessarily incomplete because a full description of metallic conduction involves quantum mechanics.

An electric field accelerates negative electrons in the direction opposite the field. But like a car in stop-and-go traffic, the electron soon gives up the energy and speed it gained from the field. For the car, that happens at the next stoplight; for the electron, it's at the next collision with an ion, where it rebounds in a random direction (Fig. 24.6). Like the car, the electron thus acquires an average velocity that's proportional to the acceleration it experiences between collisions—that is, proportional to the electric field. There's one difference, though, between the electron and the car: The electron has also a high random thermal velocity, so the average velocity is a tiny effect superposed on the electron's random thermal motion. That average velocity is the drift velocity, $\vec{v}_d$. All electrons share this common drift velocity, so their motion constitutes a current proportional to v_d.

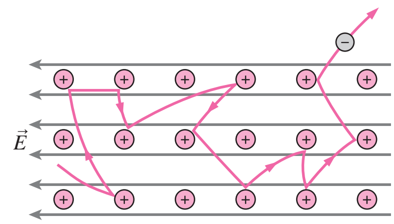


FIGURE 24.6 An electron's path in a metal is almost completely random, but in the presence of an electric field there's a slight drift antiparallel to the field.

The drift velocity depends on two things: the electrons' acceleration and the rate at which they undergo collisions. The electric field provides the acceleration, so v_d is proportional to E. The collision rate depends on how fast the electrons are moving, and here's the important point: Because thermal motions are so fast, the additional drift velocity makes essentially no difference in the collision rate, so the latter is constant. Therefore, the drift velocity and hence the current are proportional to the electric field—and that makes Equation 24.4a a linear relationship between current density and field. That's why metals are ohmic.

Although a metal's conductivity is independent of the applied field, it does depend on temperature T. That's because the thermal speed and hence the collision rate increase with temperature, decreasing conductivity and increasing resistivity. Classical physics gives thermal speed proportional to $\sqrt{T}$, as we saw in Section 17.1, so we might expect resistivity to depend similarly on temperature. Experiment, however, shows that resistivity is nearly linear with temperature (Fig. 24.7)—a result that can be explained using quantum mechanics.

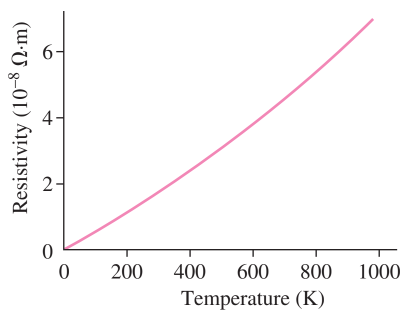


FIGURE 24.7 Resistivity of copper has a nearly linear dependence on temperature, in contrast to the classical prediction of a dependence on $\sqrt{T}$.

Although the current associated with random thermal motions averages to zero, at any given instant short-term fluctuations can result in more electrons moving in a particular direction. The result is a very small current whose direction and magnitude fluctuate randomly. This **thermal noise** can overwhelm currents of interest in sensitive electronic equipment. Circuits like the amplifiers in radio telescopes are often cooled to decrease thermal noise.

Ionic Solutions

Liquid solutions contain positive and negative ions that respond to an electric field by moving in opposite directions, resulting in a net current. Conductivity is limited by collisions between ions and neutral atoms and, as Table 24.1 suggests, ionic solutions are poorer conductors than metals. Ionic conduction is essential to life, as the transport of ions through cell membranes in Example 24.2 suggests. Electric eels use ionic conduction to sense and kill their prey. Batteries and fuel cells use ionic conduction, which also plays a role in the corrosion of metals. And an ionic solution—sweat—increases our vulnerability to electric shock.

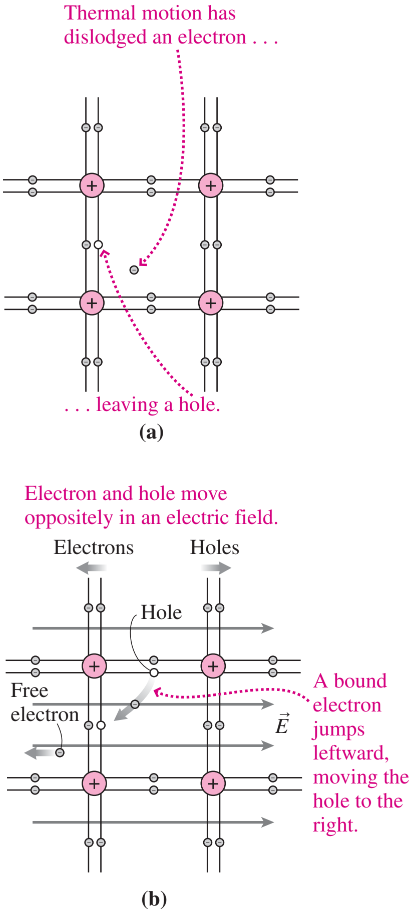


FIGURE 24.8 Structure of a silicon crystal, showing each atom bound to each of its neighbors by two shared electrons. (a) Thermal motion dislodges electrons, creating electron-hole pairs. (b) An electric field drives electrons and holes in opposite directions, creating an electric current.

Plasmas

Plasma is ionized gas that conducts because it contains free electrons and ions. It takes substantial energy to ionize atoms, so plasmas usually exist only at high temperatures. Plasmas are rare on Earth; they're in fluorescent lamps, plasma TVs, neon signs, the ionosphere, flames, and lightning flashes. Yet much of the universe's ordinary matter is in the plasma state; stars, in particular, are mostly plasma.

The electric properties of plasma make it so different from ordinary gas that plasma is often called "the fourth state of matter." Some plasmas—like the Sun's corona—are so diffuse and therefore collisions so rare as to make them far better conductors than metals. These "collisionless" plasmas can sustain large currents with minimal electric fields.

Semiconductors

Even in insulators, random thermal motions dislodge a few electrons, giving these materials very modest conductivity. In a few materials—notably the element silicon—this effect is significant at room temperature. Such materials have conductivities between those of good insulators and metals, so they're called **semiconductors**. Semiconductors make possible the microelectronic technology so pervasive in modern civilization. Here we give a qualitative description of semiconductors based on classical physics; we'll revisit semiconductors from a quantum-mechanical viewpoint in Chapter 37.

A dislodged electron leaves behind a "hole," into which an adjacent electron, nudged by the electric field, can "fall" (Fig. 24.8). The result is a movement of holes in the direction of the field. Thus holes act as positive charges, so a pure semiconductor contains equal numbers of negative charge carriers (electrons) and positive carriers (holes).

The key to semiconductor technology lies in **doping** with impurities that greatly alter a semiconductor's intrinsic conductivity. Figure 24.9 shows how a single phosphorus atom, with five valence electrons, fits into silicon's crystal structure but leaves a free electron. It doesn't take much phosphorus for these extra electrons to constitute the vast majority of charge carriers. Since the charge carriers are negative, the material is called an ***N*-type semiconductor**. Doping with trivalent atoms like boron, in contrast, leaves extra holes and makes a ***P*-type** semiconductor.

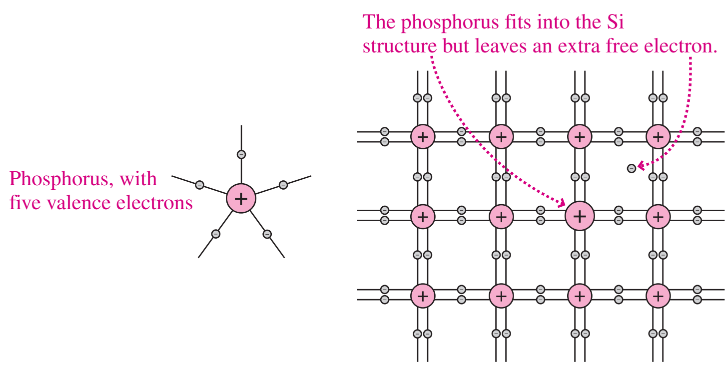


FIGURE 24.9 Phosphorus-doped silicon has extra free electrons, making it an *N*-type semiconductor.

The essential element of nearly every semiconductor device is the ***PN* junction**. Electrons and holes diffuse across such a junction and recombine, depleting the junction region of charge carriers and making it a poor conductor. Applying a voltage from the *P* to the *N* region—but not the other way—lets charge flow through the junction. So the *PN* junction conducts in one direction but not the other (Fig. 24.10). The wide range of semiconductor devices in use today results largely from carefully engineered combinations of *PN* junctions.

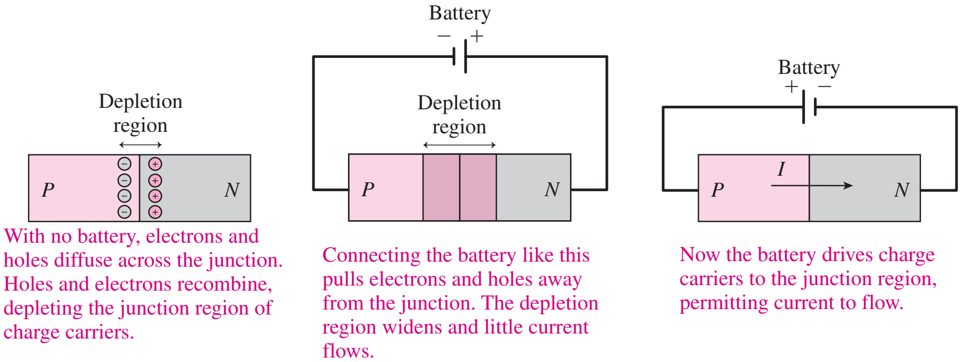


FIGURE 24.10 A *PN* junction conducts in only one direction.

Superconductors

In 1911 the Dutch physicist H. Kamerlingh Onnes found that the resistivity of mercury dropped to zero at a temperature of 4.2 K. Today we know thousands of substances that become **superconductors** at sufficiently low temperatures. Currents in superconductors persist for years without measurable decrease, suggesting the resistivity is truly zero (Fig. 24.11). For decades the known superconductors were metals and metal alloys that required cooling with liquid helium. Then, in 1986, physicists at IBM's Zurich laboratory stunned the scientific community with the discovery of ceramic materials that become superconducting at around 100 K—high enough to cool with inexpensive liquid nitrogen. The search for higher-temperature superconductors continues, with the highest temperature reported now over 160 K. Development of a room-temperature superconductor could revolutionize much of electrical technology.

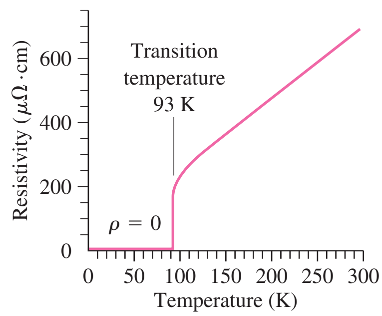
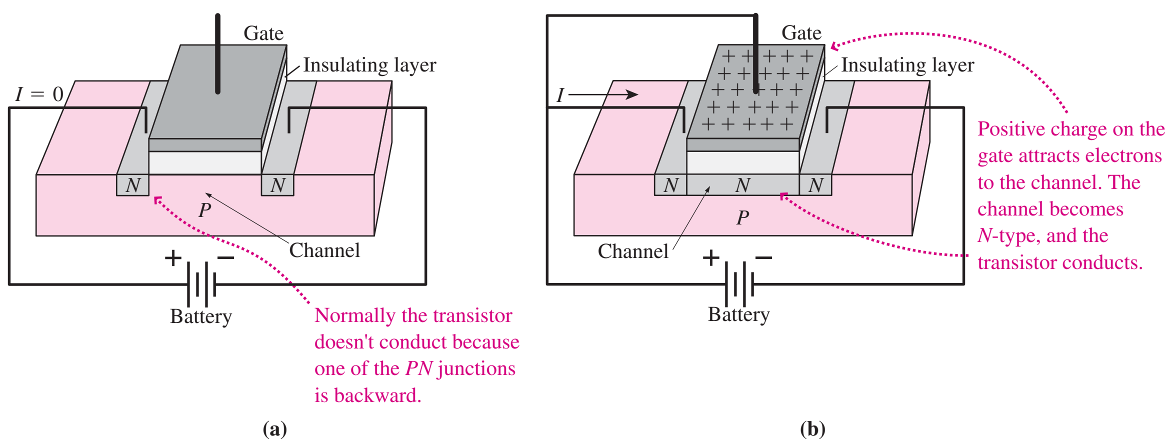


FIGURE 24.11 Resistivity versus temperature for a thin film of yttrium–barium–copper-oxide superconductor.

APPLICATION The Transistor

Few inventions have revolutionized society as much as the transistor, the semiconductor device at the root of all modern electronics. The figure shows one type, the field-effect transistor, or FET. This particular FET is a slab of *P*-type semiconductor with two embedded *N*-type regions. Normally no current can flow through the transistor because one of its two *PN* junctions is backward, as shown in part (a) of the figure. But atop the so-called *channel* between the *N*-type regions is a thin insulating layer, and over it a metal layer called the *gate*. Make the gate positive, and it pulls electrons into the channel, making it temporarily *N*-type, as in part (b). That eliminates the *PN* junctions, and now current flows through the transistor.

Varying the gate voltage continuously makes the transistor an *amplifier*, in which a weak gate signal controls a larger current. Swinging between fully on and off makes the transistor a digital switch, providing the binary 1 and 0 from which all digital information is built. Today, transistors by the billions are fabricated on single chips of silicon, making the powerful microprocessors that are the "brains" of computers.



Superconductors offer loss-free flow of electric power. Today, superconductors are widely used in high-strength electromagnets, including those in MRI scanners; in devices that measure weak magnetic fields in biomedical, geophysical, and other applications; for electric-power transmission in high-density urban applications; and in motors for ship propulsion. Expect more applications in the near future.

24.3 Resistance and Ohm's Law

How much current does it take to run this hair dryer? Do I risk a fatal shock if I touch this wire? How long an extension cord can I use with this electric saw? How long will it take to recharge my cell phone? Is the wiring in my house safe? All these questions are ultimately about the electric current flowing in wires, bodies, and batteries. The answer in each case depends on two things: the voltage V applied across the object and the **resistance** R that the object offers to the flow of electric current.

The macroscopic version of Ohm's law relates voltage, current, and resistance. Ohm's law states that the current through an object is proportional to the voltage across it and inversely proportional to the object's resistance:

$$I = \frac{V}{R} \qquad \text{(Ohm's law, macroscopic version)} \tag{24.5}$$

Ohm's law shows that a given voltage can push more current through a lower resistance. It's worth noting two extreme cases: An **open circuit** is a nonconducting gap with infinite resistance. No matter what the voltage is across an open circuit, Equation 24.5 shows that no current can flow. A switch in its "off" position is an open circuit. A **short circuit**, in contrast, has zero resistance. In a short circuit, current of any magnitude is possible without any voltage or electric field. A switch in its "on" position approximates a short circuit. An unintentional short circuit is dangerous; short circuits in household wiring, for example, are a leading cause of fires because they allow large currents to flow, resulting in excessive heating. All real situations, with the exception of superconductors, lie between the extremes of short and open circuits.

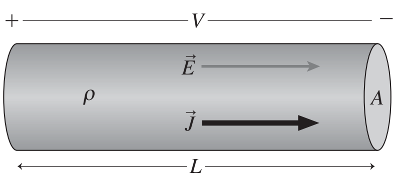


FIGURE 24.12 A cylinder of conducting material with resistivity ρ.

We can understand how the macroscopic version of Ohm's law follows from the microscopic version by considering the conductor shown in Fig. 24.12. Suppose there's a uniform electric field $\vec{E}$ within the conductor. Then there must be a uniform current density given by Equation 24.4b: $\vec{J} = \vec{E}/\rho$, where ρ is the resistivity of the material. Then the total current is $I = JA = EA/\rho$, where A is the conductor's cross-sectional area. If the conductor has length L, then the potential difference between its ends is $V = EL$, since the electric field is uniform. Solving to get $E = V/L$ and using the result in our expression for I give

$$I = \frac{VA}{L\rho} = \frac{V}{\rho L/A}$$

Comparison with the macroscopic Ohm's law, Equation 24.5, lets us identify the resistance with the term $\rho L/A$. Thus resistance depends on the resistivity ρ and the geometry—length and area—of the particular piece of material:

$$R = \frac{\rho L}{A} \tag{24.6}$$

We derived this expression for a conductor of uniform cross section; although Ohm's law still holds for a nonuniform conductor, integration is required to calculate the resistance (see Problem 61). Equations 24.5 and 24.6 both show that the units of resistance are ohms (Ω).

We emphasize that Ohm's law is not fundamental; rather, it's an empirical law that describes electrical conduction in some materials. Table 24.2 summarizes the relation between microscopic and macroscopic quantities in Ohm's law.

Table 24.2 Microscopic and Macroscopic Quantities and Ohm's Law

Microscopic	**Macroscopic**	**Relation**
Electric field, $\vec{E}$	Voltage, V	$\vec{E}$ is defined at each point in a material; V is the integral of $\vec{E}$ over a path. In a uniform field, $V = EL$.
Current density, $\vec{J}$	Current, I	$\vec{J}$ is defined at each point in a material; I is the integral of $\vec{J}$ over an area. With uniform current density, $I = JA$.
Resistivity, ρ	Resistance, R	ρ is a property of a given material; R is a property of a particular piece of that material. In a piece with uniform cross section, $R = \rho L/A$.
Ohm's law $\vec{J} = \dfrac{\vec{E}}{\rho}$	Ohm's law $I = \dfrac{V}{R}$	Microscopic version relates current density to electric field at a point in a material. Macroscopic version relates current through to voltage across a given piece of material.

EXAMPLE 24.4 Resistance and Ohm's Law: Starting Your Car

A copper wire 0.50 cm in diameter and 70 cm long connects your car's battery to the starter motor. What's the wire's resistance? If the starter motor draws a current of 170 A, what's the potential difference across the wire?

INTERPRET This problem involves Ohm's law, and we identify the wire as the object in which we want to relate current, voltage, and resistance.

DEVELOP Figure 24.13 shows the wire. Equation 24.6, $R = \rho L/A$, determines the resistance, so our plan is first to use the resistivity of copper from Table 24.1 in Equation 24.6 and then to use the resulting resistance in Ohm's law (Equation 24.5), $I = V/R$, to find the potential difference.

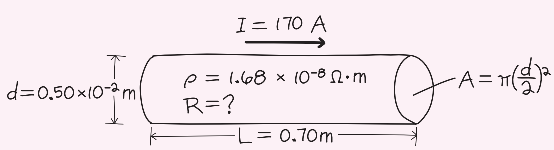


FIGURE 24.13 Sketch for Example 24.4.

EVALUATE Table 24.1 gives $\rho = 1.68\times10^{-8}\ \Omega\cdot\text{m}$ for copper, so for the resistance we get

$$R = \frac{\rho L}{A} = \frac{(1.68\times10^{-8}\ \Omega\cdot\text{m})(0.70\ \text{m})}{\pi(0.25\times10^{-2}\ \text{m})^2} = 0.60\ \text{m}\Omega$$

Then Ohm's law gives the voltage: $V = IR = (170\ \text{A})(0.60\ \text{m}\Omega) = 0.10\ \text{V}$.

ASSESS Make sense? These numbers seem awfully small. They should be! A wire carrying a large amount of current needs to have a very low resistance so the voltage across the wire remains low. We want that 12-V potential difference from the battery to appear across the starter motor, not the connecting wires. A thinner, higher-resistance wire would mean lower voltage across the starter and a significant reduction in current. ■

A resistor is a piece of conductor made to have a specific resistance. Heating elements in electric stoves, hair dryers, irons, space heaters, and the like are all essentially resistors; so are the filaments of incandescent lightbulbs. In all these cases the resistance—ultimately resulting from collisions between conduction electrons and lattice ions—provides a means of turning electric energy into heat. Resistors also set appropriate values of current and voltage in electronic circuits; for this purpose, they're made in a wide range of resistances. Resistors are rated not only by their resistance but also by the maximum power they can dissipate without overheating.

GOT IT? 24.3 The figure shows three pieces of wire. (1) and (2) are made from the same material, while (3) is made from a material with twice the resistivity. (1) and (3) have twice the diameter of (2), while (2) is twice as long as the others. (a) Which has the highest resistance? (b) If the same voltage is applied across each, which will pass the largest current?

24.4 Electric Power

Impose a potential difference V across a resistor, and a current I flows through it. The quantity V is the energy gained per unit charge as charge "falls" through the potential difference. In a resistor, that energy is dissipated through collisions, heating the material. So V is also the energy per unit charge going into heating. Meanwhile, the current I is the rate at which charge flows through the resistor. Then the energy per unit time—that is, the power dissipated in heating the resistor—is the product of the energy per unit charge and the rate at which charge moves through the conductor:

$$P = IV \qquad \text{(electric power)} \tag{24.7}$$

Although we developed Equation 24.7 for power dissipated as heat in a resistor, it holds any time electrical energy is being converted to some other form. If we measure 5 V across an electric motor and 2 A through the motor, we can conclude that the motor is converting electrical to mechanical energy at the rate of 10 W (actually less because some of the power goes into heating).

Solving Ohm's law for V and putting the result in Equation 24.7 give

$$P = I^2R \tag{24.8a}$$

Solving instead for I gives

$$P = \frac{V^2}{R} \tag{24.8b}$$

These are useful forms when we know the resistance and either the voltage or the current.

✓TIP What's Constant?

Equation 24.8a seems to imply that power increases with increasing resistance, while Equation 24.8b seems to suggest the opposite. Both implications are correct—*if* I in Equation 24.8a and V in Equation 24.8b are constants. But there's no contradiction because I and V can't both be constant while the resistance R—the ratio of V to I—changes. In most cases we work with sources of constant voltage, and then the power dissipated is inversely proportional to the resistance.

CONCEPTUAL EXAMPLE 24.1 Power Transmission

Long-distance power transmission lines operate at very high voltages—often hundreds of kilovolts. Why?

EVALUATE Equation 24.7, $P = IV$, shows that we can get the same electric power from low voltage V and high current I, or vice versa. But Equation 24.8a shows that power loss in a transmission line increases as the *square* of the current. So use of high voltage and low current minimizes transmission losses (Fig. 24.14).

ASSESS As a user, you don't encounter these high voltages. That's because transformers "step down" the voltage before it reaches the end user (we'll explore transformers in Chapter 28). The lower voltages

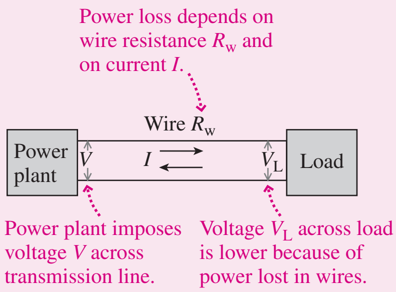


FIGURE 24.14 High voltage and low current minimize losses in power transmission.

are safer and easier to handle, although even standard 120-V household power is far from "safe."

MAKING THE CONNECTION What's the current in a typical 120-V, 100-W lightbulb? What's the bulb's resistance?

EVALUATE Solving Equation 24.7 for I gives $I = P/V = 100\text{ W}/120\text{ V} = 0.833\text{ A}$. Knowing the current, you can get the resistance either from Ohm's law or from Equation 24.8a. Or you can get it directly from Equation 24.8b. All three approaches give $R = 144\ \Omega$. The filament temperature is 3000 K, so this resistance is much higher than what you'd measure with the bulb off.

24.5 Electrical Safety

Whether you're in a lab hooking up electronic equipment, or in a hospital connecting instrumentation to a patient, or on a job designing electric devices, or simply at home plugging in appliances and tools, you should be concerned with electrical safety.

Everyone knows enough to be wary of "high voltage." People with a little more sophistication say, "It isn't the voltage but the current that kills." In fact, both points of view are partially correct. Current through the body is dangerous, but as with any resistor it takes voltage to drive that current.

Table 24.3 shows typical effects of electric currents entering the body through skin contact. A primary danger is disturbance of the electric signals that pace heartbeat; this is reflected in the lethal zone of 100–200 mA at which the heart goes into fibrillation—uncontrolled spasms of the cardiac muscle. With electric signals applied internally, much lower currents can be lethal. Doctors performing cardiac catheterization worry about currents at the microampere level.

Table 24.3 Effects of Externally Applied Current on Humans

Current Range	Effect
0.5–2 mA	Threshold of sensation
10–15 mA	Involuntary muscle contractions; can't let go
15–100 mA	Severe shock; muscle control lost; breathing difficult
100–200 mA	Fibrillation of heart; death within minutes
>200 mA	Cardiac arrest; breathing stops; severe burns

Above 200 mA, complete cardiac arrest may occur, breathing may stop, and burns may occur. Sometimes high currents are useful: Emergency defibrillators briefly apply a high enough current to stop the heart, which often restarts normal beating. The figures in Table 24.3 are rough averages and vary from person to person as well as with duration of the shock and whether alternating or direct current is involved. Very young children and people with heart conditions are at higher risk.

Under dry conditions, the typical human has a resistance of about $10^5\ \Omega$ between two points on unbroken skin. What voltages are dangerous to such a person? At $10^5\ \Omega$ it takes

$$V = IR = (0.1\text{ A})(10^5\ \Omega) = 10{,}000\text{ V}$$

to drive the fatal 100 mA. But a person who's wet or sweaty has a much lower resistance and may be electrocuted by 120-V household electricity or even lower.

To be dangerous, an electric circuit must have high voltage *and* be capable of driving sufficient current. For example, a car battery can deliver 300 A, but it can't electrocute you because its 12 V won't drive much current through you. On the other hand, the 20,000 V that runs your car's spark plugs won't electrocute you either, since the high-voltage circuit can't deliver more than a few mA.

Because potential difference is a property of two points, receiving an electric shock requires that two parts of the body contact conductors at different potentials; this chapter's opening photo provides a dramatic example. In typical 120-V wiring used throughout North America, one of the two wires is connected physically to the ground. This ground connection prevents the wiring from reaching arbitrarily high potentials, as might otherwise happen in a thunderstorm or if a short circuit occurred in a power line. At the same time it means that an individual contacting the "hot" side of the circuit and any grounded conductor such as the ground, a water pipe, or a bathtub will receive a shock.

Many devices use three-wire cords to reduce shock hazard. Exposed metal parts connect directly to a third ground wire that normally carries no current. If something goes wrong and a "hot" wire accidentally short-circuits to the metal case, this wire provides a low-resistance path to ground (Fig. 24.15). A large current flows and blows the fuse or circuit breaker, shutting off the current. Even better are *ground fault circuit interrupters* used in kitchens, bathrooms, and other high-risk locations. These devices sense a slight imbalance in current on the two wires, and shut off the circuit on the assumption that the "missing" current is leaking to ground, perhaps through a person.

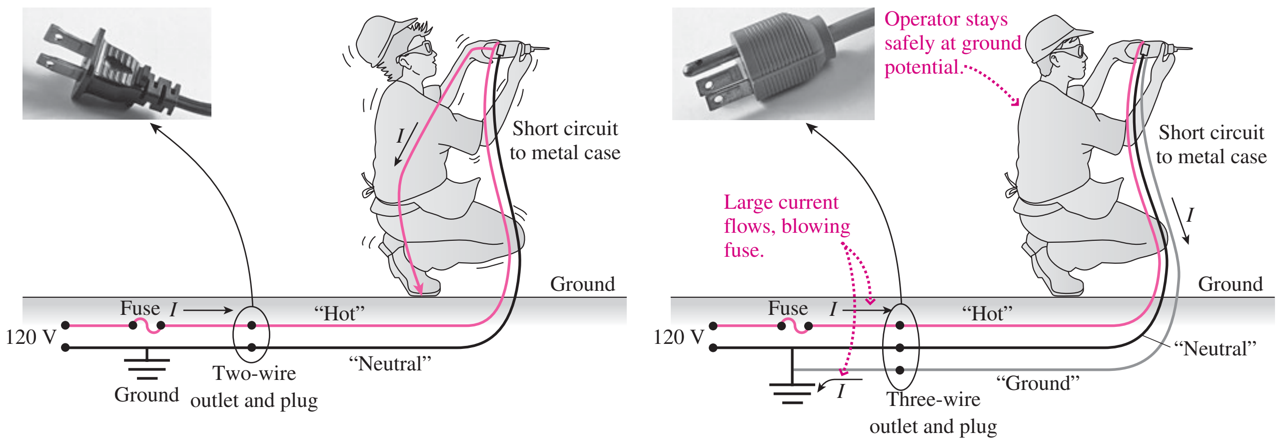


FIGURE 24.15 (a) A short circuit in an ungrounded tool could result in a lethal shock. (b) With a grounded tool, the fuse blows and the operator is safe.

CHAPTER 24 SUMMARY

Big Picture

The big idea here is **electric current**—the flow of electric charge—and its microscopic cousin, **current density**. With current we don't have electrostatic equilibrium, and there's usually an electric field in a current-carrying conductor. **Ohm's law** is an empirical statement—not a fundamental law of physics—that relates current and voltage, or current density and electric field.

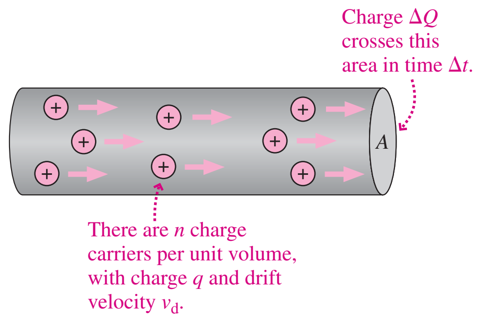
Key Concepts and Equations

Quantitatively, **current** is defined as the rate of charge flow:

$$I = \frac{\Delta Q}{\Delta t}$$

Current density is the current per unit area. Its magnitude is

$$J = \frac{I}{A}$$



Microscopically, current depends on the density of charge carriers, their charge, and the **drift velocity**:

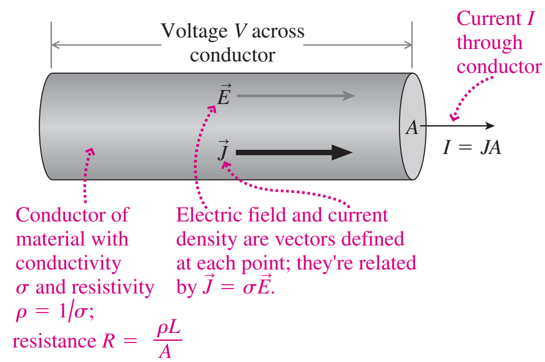
$$I = nqAv_{\mathrm{d}} \quad \text{and}$$
$$\vec{J} = nq\vec{v}_{\mathrm{d}}$$

The microscopic version of **Ohm's law** relates electric field, current density, and **conductivity** σ (or its inverse, **resistivity** ρ):

$$\vec{J} = \sigma\vec{E}$$

The macroscopic version relates voltage, current, and resistance:

$$I = V/R$$



Electric power is the product of voltage and current:

$$P = IV$$

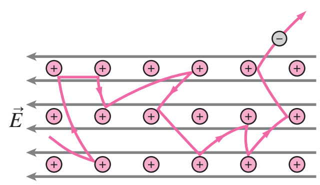
Using Ohm's law, this can also be written

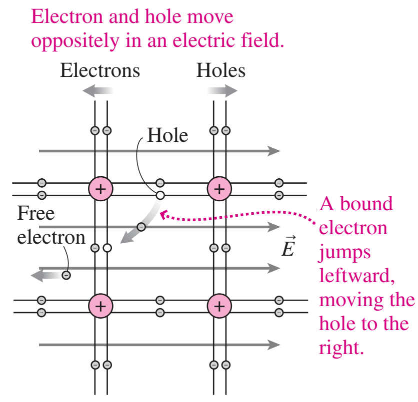
$$P = I^2R$$

$$P = \frac{V^2}{R}$$

Applications

Different types of conductors have different conduction mechanisms. In **metals**, free electrons carry the current; in **ionic solutions**, both positive and negative ions are involved; in **plasmas**, the charge carriers are free electrons and ions; and in **semiconductors**, both electrons and positive holes carry current, with semiconductor conduction properties readily adjustable. **Superconductors** are materials that exhibit zero resistance at sufficiently low temperatures.





Electrical safety is a matter of avoiding currents high enough to cause biological harm, and that means avoiding voltages high enough to drive such currents.

MP For homework assigned on MasteringPhysics, go to www.masteringphysics.com | BIO Biology and/or medical-related problems C Computer problems

For Thought and Discussion

1. Explain the difference between current and current density.
2. A constant electric field generally produces a constant drift velocity. How is this consistent with Newton's assertion that force results in acceleration, not velocity?
3. When caught in the open in a lightning storm, a person should crouch low with feet close together rather than lie flat on the ground. Why?
4. Good conductors of electricity are often good conductors of heat. Why might this be?
5. Why can current persist forever in a superconductor with no applied voltage?
6. Does an electric stove burner draw more current when it's first turned on or when it's fully hot?
7. A person and a cow are standing in a field when lightning strikes the ground nearby. Why is the cow more likely to be electrocuted?
8. You put a 1.5-V battery across a piece of material, and a 100-mA current flows. With a 9-V battery, the current increases to 400 mA. Is the material ohmic or not?
9. The resistance of a metal increases with increasing temperature, while the resistance of a semiconductor decreases. Why the difference?
10. A 50-W and a 100-W lightbulb are both designed to operate at 120 V. Which has the lower resistance?
11. Equation 24.8a suggests that no power can be dissipated in a superconductor because $R = 0$. But Equation 24.8b suggests the power should be infinite. Which is right, and why?
12. What's wrong with this news report: "A power-line worker was injured when 4000 volts passed through his body"?

Exercises and Problems

Exercises

Section 24.1 Electric Current

13. A wire carries 1.5 A. How many electrons pass through the wire in one second?
14. A 12-V car battery is rated at 80 ampere-hours, meaning it can supply 80 A of current for 1 hour before it becomes discharged. If you accidentally leave the headlights on until the battery discharges, how much charge moves through the lights?
15. BIO Biologists measure the total current due to potassium ions moving through the membrane of a rock crab neuron cell as 30 nA. How many ions pass through the membrane each second?
16. The National Electrical Code specifies a maximum current of 10 A in 16-gauge (1.29-mm-diameter) copper wire. What's the corresponding current density?

Section 24.2 Conduction Mechanisms

17. The electric field in an aluminum wire is 85 mV/m. Find the current density in the wire.
18. What electric field is necessary to drive a 7.5-A current through a 0.95-mm-diameter silver wire?
19. A cylindrical tube of seawater carries 350 mA of current. If the electric field in the water is 21 V/m, what's the tube's diameter?
20. A 1.0-cm-diameter rod carries a 50-A current when the electric field in the rod is 1.4 V/m. What's the resistivity of the rod material?
21. Use Table 24.1 to determine the conductivity of (a) copper and (b) seawater.

Section 24.3 Resistance and Ohm's Law

22. Find the resistance of a heating coil that draws 4.8 A when the voltage across it is 120 V.
23. What voltage does it take to drive 300 mA through a 1.2-kΩ resistance?
24. What's the current in a 47-kΩ resistor with 110 V across it?
25. The "third rail" that carries electric power to a subway train is an iron bar whose rectangular cross section measures 10 cm by 15 cm. Find the resistance of a 5.0-km length of this rail.
26. What current flows when a 45-V potential difference is imposed across a 1.8-kΩ resistor?
27. A uniform wire of resistance R is stretched until its length doubles. Assuming its density and resistivity remain constant, what's its new resistance?

Section 24.4 Electric Power

28. A car's starter motor draws 125 A with 11 V across its terminals. What's its power consumption?
29. A 4.5-W flashlight bulb draws 750 mA. (a) At what voltage does it operate? (b) What's its resistance?
30. A watch uses energy at the rate of 240 μW. What current does it draw from its 1.5-V battery?
31. A 35-Ω electric stove burner consumes 1.5 kW of power. At what voltage does it operate?
32. An incandescent lightbulb draws 0.50 A, while a compact fluorescent with the same light output draws 125 mA. Both operate on standard 120-V household power. How do their energy-consumption rates compare?

Section 24.5 Electrical Safety

33. Though rare, electrocution has been reported under wet conditions with voltages as low as 30 V. What resistance would be necessary for this voltage to drive a fatal current of 100 mA?
34. You touch a defective appliance while standing on the ground, and you feel the tingle of a 2.5-mA current. What's your resistance, assuming you're touching the "hot" side of the 120-V household wiring?
35. You have a typical resistance of 100 kΩ. (a) How much current could a 12-V car battery pass through you? (b) Would you feel this?

Problems

36. BIO An ion channel in a cell membrane carries 2.4 pA when it's open, which is only 20% of the time. (a) What's the average current in the channel? (b) If the channel opens for 1.0 ms, how many singly ionized ions pass through in this time?
37. A lightbulb filament has diameter 0.050 mm and carries 0.833 A. Find the current density (a) in the filament and (b) in the 12-gauge wire (diameter 2.1 mm) supplying current to the lightbulb.
38. A gold film in an integrated circuit measures 2.5 μm thick by 0.18 mm wide. It carries a current density of 0.68 MA/m^2. What's the total current?
39. A copper wire joins an aluminum wire whose diameter is twice that of the copper. The same current flows in both wires. The

density of conduction electrons in copper is $1.1\times10^{29}\ \mathrm{m}^{-3}$; in aluminum it's $2.1\times10^{29}\ \mathrm{m}^{-3}$. Compare (a) the drift speeds and (b) the current densities in each.

40. In Fig. 24.16, a 100-mA current flows through a copper wire 0.10 mm in diameter, a salt solution in a 1.0-cm-diameter glass tube, and a vacuum tube where the current is carried by an electron beam 1.0 mm in diameter. The density of conduction electrons in copper is $1.1\times10^{29}\ \mathrm{m}^{-3}$. The current in the solution is carried equally by positive and negative ions with charges $\pm2e$; the density of each ion species is $6.1\times10^{23}\ \mathrm{m}^{-3}$. The electron density in the beam is $2.2\times10^{16}\ \mathrm{m}^{-3}$. Find the drift speed in each region.

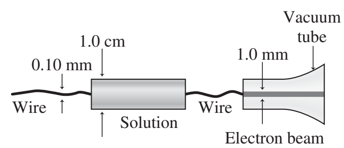


FIGURE 24.16 Problem 40

41. **BIO** In a study of proteins mediating cell membrane transport, biologists measure current versus time through the cell membranes of oocytes (nearly mature egg cells) taken from the African clawed frog, *Xenopus*. The measured current versus time is given approximately by $I = 60t + 200t^2 + 4.0t^3$ with t in seconds and I in nA. Find the total charge that flows through the cell membrane in the interval from $t = 0$ to $t = 5.0$ s.
42. There's a 2.5-V potential difference between opposite ends of a 6.0-m-long iron wire 1.0 mm in diameter. Assuming a uniform electric field in the wire, find (a) the current density and (b) the total current.
43. The maximum safe current in 12-gauge (2.1-mm-diameter) copper wire is 20 A. Find (a) the current density and (b) the electric field under these conditions.
44. Silver and iron wires of the same length and diameter carry the same current. How do the voltages across the two compare?
45. You have a cylindrical piece of material 2.4 cm long and 2.0 mm in diameter. When you attach a 9-V battery to its ends, a 2.6-mA current flows. Which material from Table 24.1 do you have?
46. How must the diameters of copper and aluminum wire be related if they're to have the same resistance per unit length?
47. You're writing the instruction manual for a power saw, and you have to specify the maximum permissible length for an extension cord made from 18-gauge copper wire (diameter 1.0 mm). The saw draws 7.0 A and needs a minimum of 115 V across its motor when the outlet supplies 120 V. What do you specify for the maximum length extension cord, given that they come in 25-foot increments?
48. **BIO** An implanted pacemaker supplies the heart with 72 pulses per minute, each pulse providing 6.0 V for 0.65 ms. The resistance of the heart muscle between the pacemaker's electrodes is 550 Ω. Find (a) the current that flows during a pulse, (b) the energy delivered in one pulse, and (c) the average power supplied by the pacemaker.
49. A solid rectangular iron bar measures 0.50 cm by 1.0 cm by 20 cm. Find the resistance between each of the three pairs of opposing faces, assuming the faces in question are equipotentials.
50. You're heading out for spring break, but your car won't start. Your friend says you might have corrosion at the battery terminals—a frequent cause of hard starting because of increased resistance. Having read Example 24.4, you know that the resistance between battery and starter should be around 1 mΩ. While your friend cranks the starter, you measure 4.2 V between the battery terminal and the wire carrying current to the starter motor. If the motor draws 125 A, is the resistance in its normal range?
51. Two cylindrical resistors are made from the same material and have the same length. When connected across the same battery, one dissipates twice as much power as the other. How do their diameters compare?
52. You're working on a new high-speed rail system. It uses 6000-horsepower electric locomotives, getting power from a single overhead wire with resistance 15 mΩ/km, at 25 kV potential relative to the track. Current returns through the track, whose resistance is negligible. Energy-efficiency standards call for no more than 3% power loss in the wire. How far from the power plant can the train go and still meet this standard?
53. A 100%-efficient electric motor is lifting a 15-N weight at 25 cm/s. How much current does it draw from a 6.0-V battery?
54. A power plant produces 1000 MW to supply a city 40 km away. Current flows from the power plant on a single wire with resistance 50 mΩ/km through the city and returns via the ground, with negligible resistance. At the power plant the voltage between wire and ground is 115 kV. Find (a) the current in the wire and (b) the fraction of the power lost in transmission.
55. You're estimating costs for a new power line with your company's financial group. Engineering specifies a resistance per unit length of 50 mΩ/km. The costs of copper and aluminum wire are \$4.65/kg and \$2.30/kg and their densities are 8.9 g/cm^3 and 2.7 g/cm^3, respectively. Which material is more economical?
56. A 240-V electric motor is 90% efficient, meaning that 90% of the energy supplied to it ends up as mechanical work. If the motor lifts a 200-N weight at 3.1 m/s, how much current does it draw?
57. A metal bar has rectangular cross section 5.0 cm by 10 cm, as shown in Fig. 24.17. The bar has a nonuniform conductivity, and as a result the current density increases linearly from zero at the bottom to 0.10 A/cm^2 at the top. Find the total current in the bar.

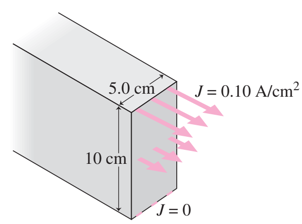


FIGURE 24.17 Problem 57

58. An immersion-type heating coil is connected to a 120-V outlet and immersed in a 250-mL cup of water initially at 10°C. The water comes to a boil in 85 s. Assuming no heat loss, and neglecting the heater's mass, find (a) the power and (b) the heater's resistance.
59. The resistivity of copper as a function of temperature is given approximately by $\rho = \rho_0[1 + \alpha(T - T_0)]$, where ρ_0 is Table 24.1's entry for 20°C, $T_0 = 20$°C, and $\alpha = 4.3\times10^{-3}\ {}^\circ\mathrm{C}^{-1}$. Find the temperature at which copper's resistivity is twice its room-temperature value.
60. Each atom in aluminum contributes about 3.5 conduction electrons. Find the drift speed in a 2.1-mm-diameter aluminum wire carrying 20 A.

61. A circular pan of radius b has a plastic bottom and metallic sidewall of height h. It's filled with a solution of resistivity ρ. A metal disk of radius a and height h is at the center, as shown in Fig. 24.18. The side and disk are essentially perfect conductors. Show that the resistance measured from side to disk is $R = \rho \ln(b/a)/2\pi h$.

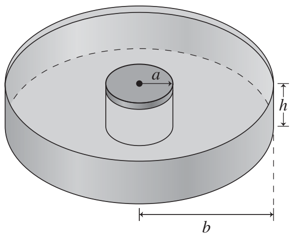


FIGURE 24.18 Problem 61

62. Figure 24.19 shows a truncated cone of material with resistivity ρ. Assume the equipotentials are planes parallel to the two faces, and integrate over slices of thickness dx like the one shown to find an expression for the total resistance between the faces.

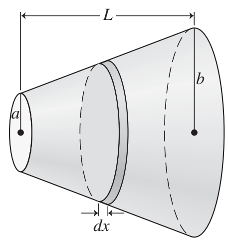


FIGURE 24.19 Problem 62

63. The current density in a particle beam with circular cross section of radius a points along the beam axis with a magnitude that decreases linearly from J_0 at the center $(r = 0)$ to half that value at the edge $(r = a)$. Find an expression for the total current in the beam.
64. A cylindrical resistor is 5.0 mm in diameter and 1.5 cm long. It's made of a composite material whose resistivity varies from one end to the other according to the equation $\rho = \rho_0(1 + x/L)e^{x/L}$, for $0 \leq x \leq L$, where $\rho_0 = 2.41\times10^{-3}\ \Omega\cdot\text{m}$. Find its resistance.
65. You work for an automobile manufacturer developing a new plug-in hybrid car. The car's mass is 1200 kg, and it uses a 360-V battery driving an electric motor that can handle a maximum current of 180 A. You're to specify the greatest slope the car can climb, maintaining 60 km/h, without its gasoline engine coming on to assist.

Passage Problems

A *brownout* occurs when an electric utility can't supply enough power to meet demand. Rather than cut off some customers completely, the utility reduces the voltage across its system. Brownouts are most likely on hot summer days, when heavy air-conditioning loads drive up demand for electricity. In a particular brownout, the utility reduces the voltage by 10%.

66. During the brownout, the current in conductors whose resistance is nearly independent of temperature
    a. decreases by approximately 10%.
    b. decreases by approximately 20%.
    c. decreases by approximately 5%.
    d. You can't tell without knowing the resistance.
67. Which of the following occurs in the conductors of the preceding problem during the brownout?
    a. Both the electric field and electron drift speed decrease.
    b. The electric field decreases but the electron drift speed doesn't.
    c. The current is carried by fewer electrons.
    d. The electrons undergo more frequent collisions.
68. During the brownout, the power dissipated in conductors whose resistance is nearly independent of temperature
    a. decreases by approximately 10%.
    b. decreases by approximately 20%.
    c. decreases by approximately 5%.
    d. You can't tell without knowing the resistance.
69. Metallic conductors like lightbulb filaments and electric stove burners have resistance that increases with increasing temperature. During the brownout, the current in such devices
    a. decreases by 10%.
    b. decreases by more than 10%.
    c. decreases by less than 10%.
    d. You can't tell without knowing more about how the resistance varies.

Answers to Chapter Questions

Answer to Chapter Opening Question

Collisions between electrons and the metal ions in the filament dissipate electric energy as heat. The energy results from the electrons' being accelerated by an electric field.

Answers to GOT IT? Questions

24.1. (a) current, right to left; (b) current, up; (c) current, left; (d), (e) no current.
24.2. (a) $J_A < J_B$; (b) $E_A < E_B$; (c) $v_{dA} < v_{dB}$.
24.3. (a) (2); it's twice as long as (3) but with one-fourth the area and half the resistivity; (b) (1), because it has the lowest resistance.

25 Electric Circuits

Festive lights decorate a city. If one of them burns out, they all go out. Are they connected in series or in parallel?

New Concepts, New Skills

By the end of this chapter you should be able to

- "Read" an electric-circuit diagram, identifying individual components and their interconnections (25.1).
- Analyze simple circuits using series and parallel combinations (25.2).
- Analyze more complex circuits by applying the node and loop laws (25.3).
- Use electrical measuring instruments (25.4).
- Describe the time-dependent behavior of circuits that include capacitors (25.5).

Connecting Your Knowledge

- This chapter draws on the concepts of electric current (24.1) and voltage (22.1).
- We'll use the ideas of Ohm's law and resistance as well as electric power (24.3, 24.4).
- We'll work with series and parallel connections of electrical components (23.3).
- We'll revisit capacitors, now using them in circuits (23.2).

An **electric circuit** is a collection of electrical components connected by conductors. Human-made circuits range from simple flashlights to computers. Electric circuits also exist in nature, including your own nervous system and Earth's atmospheric circuit in which thunderstorms are the batteries and the atmosphere a resistor. Understanding circuits will help you use effectively and safely the myriad electrical devices in your life, and can even help you design new devices or troubleshoot old ones.

25.1 Circuits, Symbols, and Electromotive Force

We diagram circuits using standard symbols for circuit components and lines to represent wires (Fig. 25.1). We usually approximate wires as perfect conductors; then all points connected by a wire are at the same potential and are electrically equivalent. Realizing this will help you understand circuit diagrams.

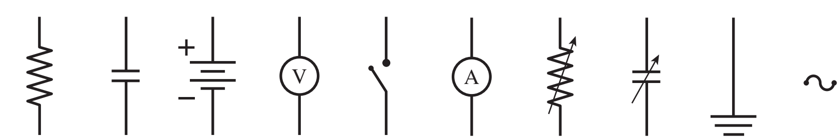


FIGURE 25.1 Common circuit symbols.

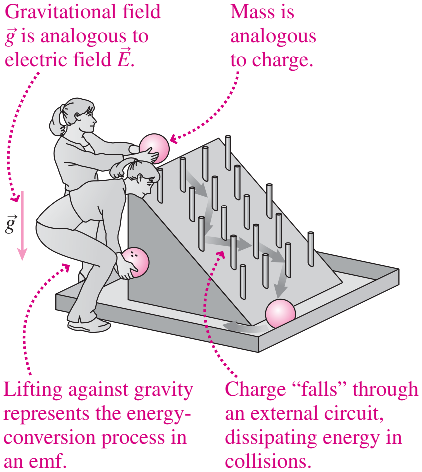


FIGURE 25.2 Gravitational analog for emf.

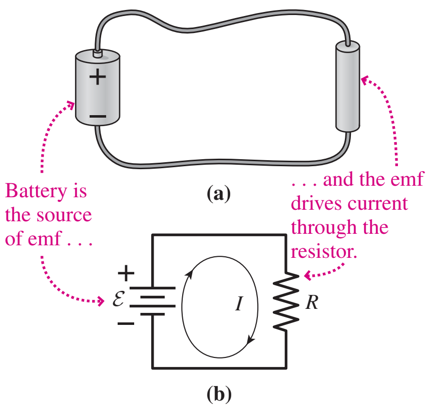


FIGURE 25.3 A circuit consisting of a battery and a resistor: (a) physical circuit; (b) schematic diagram.

It takes an electric field to drive current through a conductor with nonzero resistance. But unless we actively maintain the field, charge will quickly move to establish electrostatic equilibrium, with no field inside the conductor and no current. So we need a device that can maintain a fixed potential difference and therefore an electric field in a current-carrying conductor. Such a device is called a source of **electromotive force**, or **emf**. (The name "force" here is inaccurate and is used only for historical reasons.) Most sources of emf have two **terminals** for connection to other circuit components. An emf converts some other form of energy to electrical energy by separating positive and negative charge to maintain a fixed potential difference between its terminals. The most familiar example is a battery, in which chemical reactions drive charge to the two terminals. Others include electric generators, which convert mechanical to electrical energy; photovoltaic cells, which use sunlight to separate charge; and cell membranes, which control ion flow into and out of the cell.

When a source of emf is connected to an external circuit, current flows through the circuit from the emf's positive terminal to the negative terminal. Energy-conversion processes in the emf then "lift" charge against the emf's internal electric field, maintaining a fixed potential difference across its terminals. The charge then "falls" through the external circuit, dissipating its energy in the circuit resistance. The result is a steady current, driven by the constant voltage across the emf. Figure 25.2 shows a gravitational analogy for an emf connected across an external circuit.

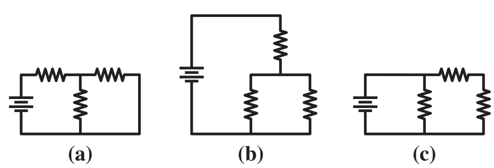
Quantitatively, emf is the work per unit charge involved in "lifting" charge against the electric field. Its units are therefore volts. An **ideal emf** maintains the same voltage across its terminals under all conditions. Real emfs have internal energy losses, and the terminal voltage may not equal the rated emf.

In Fig. 25.3 an ideal battery of emf $\mathcal{E}$ drives current through resistor R. We're assuming the wires connecting the battery and the resistor are perfect conductors, so the voltage across the resistor is equal to the battery's emf. Ohm's law then gives the resistor current: $I = \mathcal{E}/R$. Energetically, this circuit is analogous to Fig. 25.2: Charge gains $\mathcal{E}$ joules per coulomb as it's "lifted" against the electric field inside the battery, then dissipates that energy in heating the resistor.

✓TIP Don't Get Hung Up on Wires

We approximate wires as perfect conductors, so it takes no potential difference to drive current through a wire. Thus all points on the wire are at the same potential and are electrically equivalent. That means there are many ways to draw the same circuit; as long as two points are connected by a wire, that's all that matters. Of course real wires have some resistance, but if it's negligible compared with other resistances in the circuit, then we can approximate the wires as being ideal.

GOT IT? 25.1 The figure shows three circuits. Which are electrically equivalent?



25.2 Series and Parallel Resistors

We considered series and parallel capacitors in Section 23.3. Series and parallel are the two simplest ways to connect *any* electric components. Two components are in series if the current flowing through one component has nowhere to go but through the other component. Two components are in parallel if they're connected together at each end. Here we'll consider series and parallel resistors.

Series Resistors

Figure 25.4 shows a circuit with two resistors in series. We'd like to know the current through and the voltage across each resistor. Neither is connected directly across the battery, so we can't argue that either resistor "sees" the battery emf. But the resistors are in series, and that means the only place for current to go after R_1 is through R_2. In a steady state, with no charge buildup in the circuit, that means the current through both resistors—and through the battery as well—must be the same. This is true whenever circuit components are in series:

> The current through circuit components in series is the same.

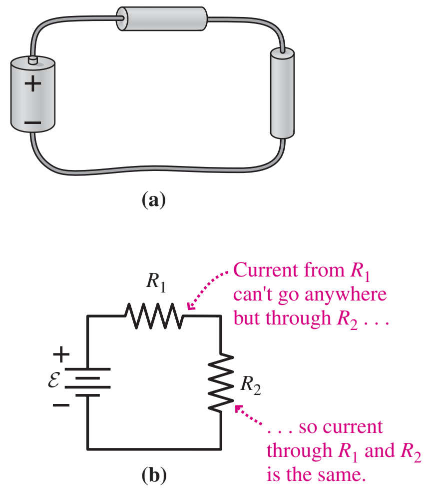


FIGURE 25.4 A battery and two resistors in series: (a) physical circuit; (b) schematic diagram.

If I is the current in Fig. 25.4, then by Ohm's law there must be a voltage $V_1 = IR_1$ across R_1 to drive the current through this resistor. Similarly, the voltage across R_2 is $V_2 = IR_2$. Thus, the voltage across the two resistors together is $V_1 + V_2 = IR_1 + IR_2$. But the battery is connected directly across this series combination, so we have $IR_1 + IR_2 = \mathcal{E}$, or

$$I = \frac{\mathcal{E}}{R_1 + R_2}$$

Comparison with Ohm's law in the form $I = V/R$ shows that the two resistors in series behave like an equivalent resistance equal to the sum of their resistances. In an obvious generalization to more resistors in series, we have

$$R_{\text{series}} = R_1 + R_2 + R_3 + \cdots \qquad (\text{series resistors}) \tag{25.1}$$

In other words, resistors in series add.

Given the current, we can use Ohm's law in the form $V = IR$ to solve for the voltage across each resistor:

$$V_1 = \frac{R_1}{R_1 + R_2}\mathcal{E} \quad \text{and} \quad V_2 = \frac{R_2}{R_1 + R_2}\mathcal{E} \tag{25.2a, b}$$

These expressions show that the battery voltage divides between the two resistors in proportion to their resistance. For this reason a series combination of resistors is called a **voltage divider**.

✓TIP How Does the Battery Know?

How does the battery in Fig. 25.4 "know" how much current to supply? For a brief instant when the circuit is first connected, it doesn't. But in a very short time an electric field is established throughout the wires and resistors, and the circuit settles into a steady state, with the same current everywhere. Later, with circuits including capacitors, we'll analyze the approach to the steady state; for now, assume that the circuit reaches that state essentially instantaneously.

EXAMPLE 25.1 Series Resistors: Designing a Voltage Divider

A lightbulb with resistance 5.0 Ω is designed to operate at a current of 600 mA. To operate this lamp from a 12-V battery, what resistance should you put in series with it?

INTERPRET This problem is about a series circuit like Fig. 25.4, with a lightbulb and unknown resistance for the two resistors.

DEVELOP We've sketched the circuit in Fig. 25.5, taking R_1 as the unknown and R_2 as the 5-Ω lightbulb. The same current flows through series resistors, so our plan is to find an expression for that current and

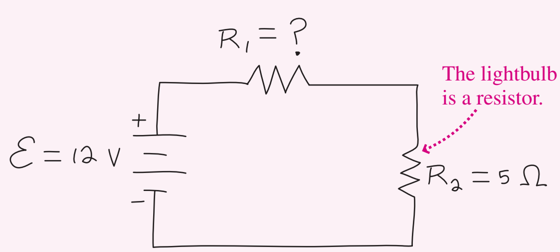


FIGURE 25.5 Sketch for Example 25.1.

(continued)

then solve for the value of R_1 that will make the current 600 mA. Since resistors in series add, the current through both resistors follows from Ohm's law: $I = \mathcal{E}/(R_1 + R_2)$.

EVALUATE We solve for R_1 to get

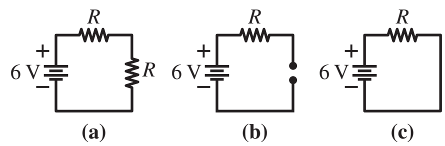
$$R_1 = \frac{\mathcal{E} - IR_2}{I} = \frac{12\ \text{V} - (0.60\ \text{A})(5.0\ \Omega)}{0.60\ \text{A}} = 15\ \Omega$$

ASSESS Make sense? The lightbulb's operating voltage is

$$V = IR_2 = (0.60\ \text{A})(5.0\ \Omega) = 3.0\ \text{V}$$

This is one-fourth of the battery voltage, so Equation 25.2b shows that the bulb's 5-Ω resistance should be one-fourth of the total. That makes the total 20 Ω, leaving 15 Ω for R_1. This isn't a very efficient way to run the bulb, since a lot more energy gets dissipated in R_1 than goes into lighting the bulb. Better to use a 3-V battery and no resistor. ■

GOT IT? 25.2 Rank from highest to lowest the voltages across the identical resistors R at the top of each circuit shown, and give the actual voltage for each. In (a) the second resistor has the same resistance R, and in (b) the gap is an open circuit (infinite resistance).



Real Batteries

FIGURE 25.6 Both batteries are rated at 1.5 V, but they have different internal resistances. Which do you think has the higher R_{int}?

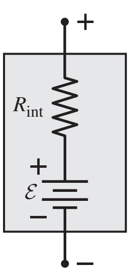


FIGURE 25.7 A real battery modeled as an ideal emf in series with an internal resistance.

What's the difference between the two 1.5-V batteries in Fig. 25.6? If they were ideal, both would maintain 1.5 V across their terminals no matter how much current was flowing. But these are real batteries. Chemical reaction rates limit the current, so it's not surprising that the larger battery can deliver more current.

We model a real battery as an ideal emf in series with an **internal resistance** (Fig. 25.7). Of course there is no ideal emf! The internal resistance is intrinsic to the battery, and there's no way to circumvent it. Some of it is actual resistance, but most represents the limited rate at which chemical reactions can separate charge. For a given battery voltage, lower internal resistance implies a more powerful battery—one that can deliver more current.

Figure 25.8 shows that the internal resistance R_{int} is in series with the external load R_L to which the battery supplies power; the resulting circuit is a voltage divider. If R_{int} is small compared with R_L, Equation 25.2b shows that the voltage across the load will be very nearly the battery voltage. Then the battery is behaving nearly ideally because it has essentially $\mathcal{E}$ volts across its terminals. But if we lower R_L, more current flows and more voltage drops across R_{int}—and that leaves less voltage at the battery terminals and across the load. Even if we short-circuit the battery (not a good idea!), we won't get an infinite current; in fact, we'll get $I = \mathcal{E}/R_{int}$, the most current this battery can deliver.

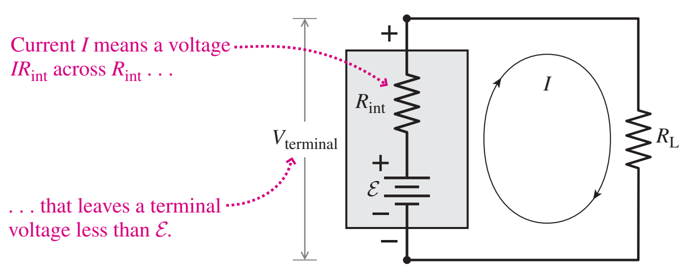


FIGURE 25.8 A real battery connected to an external load. Some voltage drops across the internal resistance, making the terminal voltage less than the battery's rated voltage.

EXAMPLE 25.2 Internal Resistance: Starting a Car

Your car has a 12-V battery with internal resistance 0.020 Ω. When the starter motor is cranking, it draws 125 A. What's the voltage across the battery terminals while starting?

INTERPRET This problem is about a real battery connected to a load, as in Fig. 25.8. We identify one resistor as the internal resistance and the load resistance as the starter motor.

DEVELOP Figure 25.9 is our sketch, showing the internal resistance in series with the load. The current is the same everywhere in a series circuit, so we can use Ohm's law to find the voltage across R_{int}. Subtracting that voltage from the battery's emf will then tell what's left across the load.

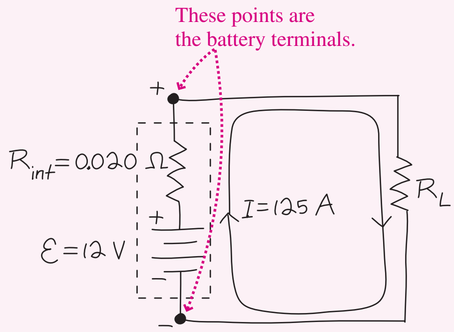


FIGURE 25.9 Sketch for Example 25.2.

EVALUATE For the internal resistance, Ohm's law gives

$$V_{\text{int}} = IR_{\text{int}} = (125\ \text{A})(0.020\ \Omega) = 2.5\ \text{V}$$

That leaves 12 V − 2.5 V or 9.5 V across the battery terminals.

ASSESS Make sense? That 9.5 V is substantially less than the battery's 12-V rating, so we're hardly treating it ideally. But the starter motor runs only briefly; most of the time the load on the battery—headlights, ignition system, electronics, and so on—draws far less current and so the battery behaves essentially like an ideal 12-V emf. A battery voltage of 9–11 V is typical during starting; much less than 9 V indicates a weak battery, a defective starter, or very cold weather. ■

Parallel Resistors

Figure 25.10 shows two resistors in parallel, connected across an ideal battery. Since the two resistors are connected at top and bottom by ideal wires, the voltage across each must be the same. We made this point in Chapter 23 when we discussed parallel capacitors, and it's worth repeating here:

> The voltage across circuit elements in parallel is the same.

The parallel resistors are connected directly across the battery, so their common voltage is the battery emf $\mathcal{E}$. Applying Ohm's law then gives the current through each resistor:

$$I_1 = \frac{\mathcal{E}}{R_1} \qquad \text{and} \qquad I_2 = \frac{\mathcal{E}}{R_2}$$

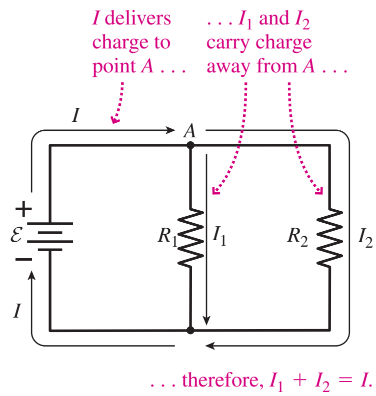


FIGURE 25.10 Parallel resistors connected across a battery.

At point A in Fig. 25.10, a current I brings in charge from the battery, while the currents I_1 and I_2 carry charge away. Charge can't accumulate at this point (see Problem 63), so the incoming and outgoing currents must be equal: $I = I_1 + I_2$. Using our expressions for the two resistor currents gives

$$I = \frac{\mathcal{E}}{R_1} + \frac{\mathcal{E}}{R_2} = \mathcal{E}\left(\frac{1}{R_1} + \frac{1}{R_2}\right)$$

Comparison with Ohm's law in the form $I = V/R$ shows that the equivalent resistance of the parallel combination is given by

$$\frac{1}{R_{\text{parallel}}} = \frac{1}{R_1} + \frac{1}{R_2}$$

This result readily generalizes to more parallel resistors:

$$\frac{1}{R_{\text{parallel}}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots \qquad \text{(parallel resistors)} \qquad (25.3\text{a})$$

In other words, resistors in parallel add reciprocally. Equation 25.3a shows that the resistance of a parallel combination is always lower than that of the lowest resistance in the combination. You should confirm this for yourself.

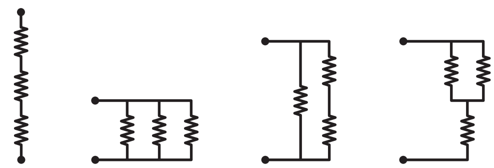
An analogy with highway traffic shows why this makes sense: Adding a lane to a crowded highway eases congestion (i.e., lowers the overall resistance), allowing a greater traffic flow (i.e., greater current). Putting one resistor in parallel with another is like adding an extra traffic lane.

When there are only two parallel resistors, we can rewrite Equation 25.3a using a common denominator to obtain

$$R_{\text{parallel}} = \frac{R_1 R_2}{R_1 + R_2} \tag{25.3b}$$

Note that *parallel* resistors combine in the same way as *series* capacitors, and vice versa.

GOT IT? 25.3 The figure shows all four possible combinations of three identical resistors. Rank them in order of highest to lowest resistance.



(a) **(b)** **(c)** **(d)**

Analyzing Circuits

Many circuits contain series and parallel combinations. We analyze these circuits using the tactics outlined next, following the approach we used with series and parallel capacitors in Example 23.3.

TACTICS 25.1 Analyzing Circuits with Series and Parallel Components

1. Identify series and parallel combinations. Remember that components are in parallel *only* if they're connected directly together at each end. Components are in series *only* if current through one component has no place to go but through the next component. If you can't find at least one series or parallel combination, then you have to use the methods of Section 25.3.
2. Solve for the series and parallel equivalents using Equations 25.1 and 25.3 for resistors:

$$R_{\text{series}} = R_1 + R_2 + R_3 + \cdots \tag{25.1}$$

$$\frac{1}{R_{\text{parallel}}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots \tag{25.3a}$$

$$R_{\text{parallel}} = \frac{R_1 R_2}{R_1 + R_2} \tag{25.3b}$$

   If you're dealing with capacitors, use Equations 23.6 and 23.5, respectively.

3. Redraw the circuit, replacing series and parallel combinations with their one-component equivalents.
4. Repeat Steps 1–3, each time identifying series and parallel combinations and then reducing each to a single equivalent. Continue until either you've found the quantity you're asked for or the circuit consists of just an emf and one other component. You can then solve for the current in this component.
5. Work backward, replacing series and parallel equivalents with combinations of individual components. At each point apply Ohm's law, $I = V/R$, to find the currents through and/or the voltages across the individual components. As you work backward, remember that series components carry the same current as their series equivalent, and parallel components have the same voltage as their parallel equivalent. Continue until you're able to evaluate the quantity you're asked for.

EXAMPLE 25.3 Analyzing a Circuit: Series and Parallel Components

Find the current through the 2-Ω resistor in the circuit of Fig. 25.11*a*.

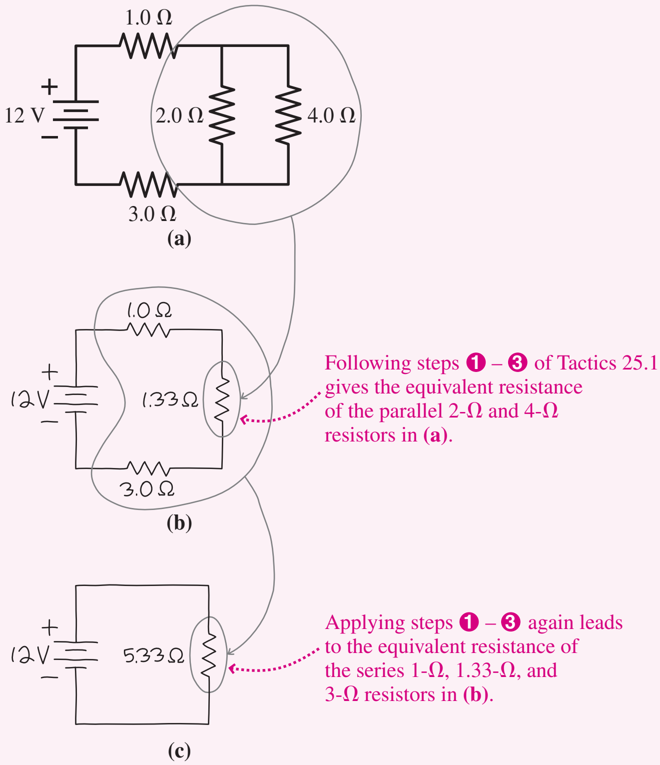


FIGURE 25.11 Analyzing a circuit.

INTERPRET This problem asks for the current in one resistor that's part of a more complex circuit. So it's about analyzing a circuit with series and parallel components.

DEVELOP We follow the steps in Tactics 25.1:

1. We identify the 2-Ω and 4-Ω resistors as being in parallel, and we find no other series or parallel resistor combinations. The 1-Ω resistor, for example, is *not* in series with either the 2-Ω or 4-Ω resistor because current leaving the 1-Ω resistor can take either of two paths.
2. We apply Equation 25.3b, $R_{\text{parallel}} = R_1R_2/(R_1 + R_2)$, to find the parallel combination: $(2\ \Omega)(4\ \Omega)/(2\ \Omega + 4\ \Omega) = 1.33\ \Omega$.
3. We redraw the circuit as Fig. 25.11*b*, replacing the two parallel resistors with their 1.33-Ω equivalent.
4. We repeat Steps 1–3 for the circuit in Fig. 25.11*b*, this time finding a series combination of three resistors. Applying Equation 25.1, $R_{\text{series}} = R_1 + R_2 + R_3$, gives 5.33 Ω for the equivalent resistance, and we redraw the circuit to get the simple circuit of Fig. 25.11*c*. Ohm's law, $I = V/R$, gives the current in the 5.33-Ω equivalent resistance: $I_{5.33\ \Omega} = (12\ \text{V})/(5.33\ \Omega) = 2.25$ A.
5. Now we work backward, "unsimplifying" the circuit. That 5.33-Ω resistor is really the series combination in Fig. 25.11*b*; since the current through series components is the same, 2.25 A flows through each resistor—including the 1.33-Ω resistor that's really the parallel combination shown in Fig. 25.11*a*. We want the current in the 2-Ω member of that combination, and we could get that if we knew the voltage across it. But the voltage across parallel components is the same, and the same as the voltage across their parallel equivalent—in this case the 1.33-Ω resistance. We've found the current through that resistance, so Ohm's law gives the voltage: $V_{1.33\ \Omega} = I_{1.33\ \Omega}R_{1.33\ \Omega} = (2.25\ \text{A})(1.33\ \Omega) = 3.0$ V.

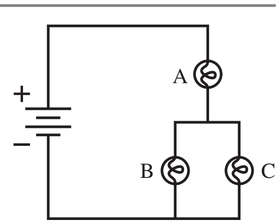
EVALUATE Our last result is the voltage across each of the original parallel resistors, including the 2-Ω resistor whose current we want. So we're finally ready to compute our answer: $I_{2\ \Omega} = V_{2\ \Omega}/R_{2\ \Omega} = (3.0\ \text{V})/(2.0\ \Omega) = 1.5$ A. Done!

ASSESS Make sense? A total of 2.25 A is flowing around the circuit; when it encounters the parallel combination, more should flow through the lower resistance, which is just what we found. Quantitatively, the current divides in inverse proportion to the parallel resistances, with 1.5 A through 2 Ω, and half as much, 0.75 A, through 4 Ω. Note how, in solving this problem, we used Ohm's law to find, alternately, voltage and then current in different resistances. ■

✓TIP Using Ohm's Law

Ohm's law relates the *voltage across a resistor* to the *current through that resistor*. It does *not* relate arbitrary voltages and currents anywhere in a circuit. Just because there's a 12-V battery in Fig. 25.11 doesn't mean there's 12 V across the 2-Ω resistor. And just because we found a total current of 2.25 A in Fig. 25.11*c* doesn't mean that's the current through the 2-Ω resistor.

GOT IT? 25.4 The figure shows a circuit with three identical lightbulbs and a battery. (a) Which, if any, of the bulbs is brightest? (b) What happens to each of the other two bulbs if you remove bulb C?



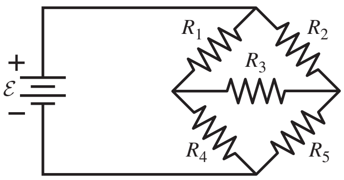


FIGURE 25.12 This circuit can't be analyzed using series and parallel combinations.

25.3 Kirchhoff's Laws and Multiloop Circuits

Some circuits can't be simplified using series and parallel combinations. This happens when there's more than one emf, or when components are connected in complex ways. In Fig. 25.12, are R_1 and R_2 in parallel? No, because R_3 separates their lower ends. Are R_1 and R_4 in series? No, because the current splits after leaving R_1. Analyzing such circuits requires a more general technique.

Kirchhoff's Laws

Charges moving through a circuit gain energy at emfs and lose energy in resistors. If we go completely around a circuit, the *changes* in energy per unit charge—that is, increases and decreases in voltage—sum to zero. This is **Kirchhoff's loop law**, and it holds for any closed loop even if it's part of a more complex circuit: **The sum of voltage changes around a closed loop is zero**. The loop law is essentially a statement of energy conservation for circuits.

In analyzing parallel resistors, we saw that the current flowing into point A in Fig. 25.10 had to equal the current flowing out. That's because charge is conserved, and in a steady state charge can't be accumulating anywhere in a circuit. A junction of three or more wires, like point A in Fig. 25.10, is a **node**. If we count the currents flowing into a node as positive and those flowing out as negative, then we can state **Kirchhoff's node law**: **The sum of currents at any node is zero**.

Multiloop Circuits

Kirchhoff's laws allow us to analyze even the most complex circuits; the following strategy details the approach.

PROBLEM-SOLVING STRATEGY 25.1 Multiloop Circuits

INTERPRET

- Identify circuit loops and nodes. A loop is any complete closed path; a node is any point where three or more wires meet.
- Label the currents at each node, assigning a direction to each. The directions are arbitrary, and the actual direction may not be obvious.

DEVELOP

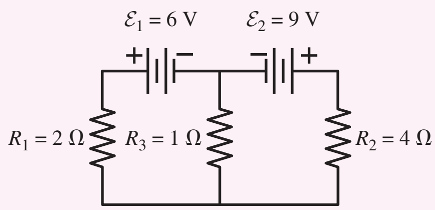
- For all but one node, write equations expressing Kirchhoff's node law: The sum of the currents at each node is zero. Take a current flowing into the node as positive, a current flowing out as negative.
- For as many independent loops as necessary, write equations expressing Kirchhoff's loop law: The sum of the voltage changes around a closed loop is zero. You can go either clockwise or counterclockwise, following these rules:
  - The voltage change going through a battery from the negative to the positive terminal is $+\mathcal{E}$; the voltage change from $+$ to $-$ is $-\mathcal{E}$.
  - For resistors traversed in the direction you've assigned to the current, the voltage change is $-IR$; for the opposite direction, it's $+IR$.
  - For other circuit components, use each component's characteristics to determine the voltage change.
- You don't need equations for all the nodes and loops because some are redundant, as you'll see in the next example.

EVALUATE Solve the equations to determine the unknown currents or other quantities.

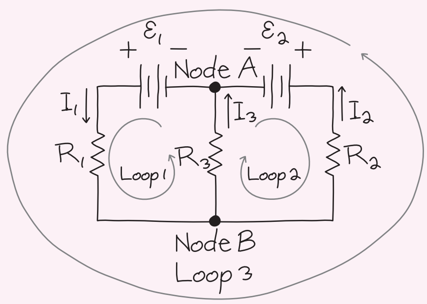
ASSESS Assess your answer to see that it makes sense, paying particular attention to signs. A negative answer for a current means that the current actually flows opposite the direction you arbitrarily assigned.

EXAMPLE 25.4 Using Kirchhoff's Laws: A Multiloop Circuit

Find the current in resistor R_3 of Fig. 25.13*a*, following Problem-Solving Strategy 25.1.



(a)



(b)

FIGURE 25.13 Example 25.4.

INTERPRET In Fig. 25.13*b* we've redrawn the circuit with three loops and two nodes identified. We also labeled three currents at node A. The directions shown are completely arbitrary and may or may not be the actual directions of current flow. Because currents in series elements are the same, we can put the current labels anywhere on the series paths leading to the node. Our drawing shows that the same currents flow at node B; that's why one of the node equations is redundant. Note also that loop 3 comprises parts of loops 1 and 2, so equations for any two of these loops contain all the information we need. So one of the loop equations is redundant.

DEVELOP We need a current equation for one of the two nodes. Given the arbitrary current directions, the equation at node A is

$$-I_1 + I_2 + I_3 = 0 \qquad \text{(node A)}$$

To get the equation for loop 1, let's go counterclockwise around the loop, as shown. Starting at node A, we first encounter a positive voltage change $+\mathcal{E}_1$, then $-I_1R_1$, then $-I_3R_3$. So the loop 1 equation is $\mathcal{E}_1 - I_1R_1 - I_3R_3 = 0$. Here it's simplest if we substitute the numerical values shown in Fig. 25.13*a* and temporarily drop the units to avoid clutter. Then we have

$$6 - 2I_1 - I_3 = 0 \qquad \text{(loop 1)}$$

Loop 2 is similar except here we're going "backward" through R_2, so its term is positive:

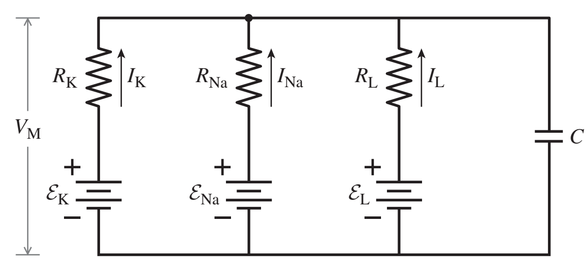
$$9 + 4I_2 - I_3 = 0 \qquad \text{(loop 2)}$$

EVALUATE We want I_3, so we eliminate the other two currents. The node equation gives $I_1 = I_2 + I_3$; substituting in the loop 1 equation gives $6 - 2I_2 - 3I_3 = 0$ or $I_2 = \frac{1}{2}(6 - 3I_3)$. We use this result in the loop 2 equation to get $9 + 2(6 - 3I_3) - I_3 = 0$, or $21 - 7I_3 = 0$. Solving gives our answer: $I_3 = 3$ A.

ASSESS We assigned I_3 an upward direction through R_3, so our positive answer means that this is indeed the direction of the current in R_3. This makes sense because both batteries have their negative terminals at node A. If either battery had been reversed, however, the situation wouldn't have been so clear and we would have had to rely on the algebraic sign to determine the current direction. Even with the circuit as shown in Fig. 25.13, the directions of I_1 and I_2 depend on the relative strengths of the two batteries. With the values we're given, I_2 comes out -1.5 A, showing that the current actually flows downward in R_2. But if we reduce $\mathcal{E}_2$ to 2 V, I_2 becomes zero; lower still, and it flows upward and "backward" through $\mathcal{E}_2$. ■

APPLICATION The Cell Membrane

Many natural systems can be modeled as electric circuits. In 1952, Alan L. Hodgkin and Andrew F. Huxley developed a circuit model for the cell membrane; their work won them a share of the 1963 Nobel Prize for Physiology or Medicine. The figure shows a simplified version of the Hodgkin–Huxley model. The batteries $\mathcal{E}_K$, $\mathcal{E}_{Na}$, and $\mathcal{E}_L$ represent the electrochemical effects of potassium, sodium, and other ions, respectively; their emfs have values in the tens of millivolts. R_K, R_{Na}, and R_L are the resistances the cell membrane offers to each ionic species. The currents I_K, I_{Na}, and I_L represent ion flows across the membrane, and their values and signs follow from solving a multiloop-circuit problem similar to Example 25.4. The voltage V_M is the membrane potential between the inside and outside of the cell. The Hodgkin–Huxley model also contains a capacitor, which causes time-dependent behavior of the sort we'll see in Section 25.5.



25.4 Electrical Measurements

Voltmeters

A **voltmeter** is a device that indicates the potential difference across its two terminals. The indicator is usually a digital readout, although older meters use a moving needle. Potential difference—voltage—is a property of two points, and therefore to measure the voltage between two points, we connect the two terminals of the voltmeter to those points. So to measure the voltage across resistor R_2 in Fig. 25.14*a* we connect the voltmeter *across* R_2, as shown. We do *not* break the circuit and insert the meter as in Fig. 25.14*b*, for then we wouldn't be measuring the voltage *across* the resistor; in fact, as Conceptual Example 25.1 makes clear, we would radically alter the circuit.

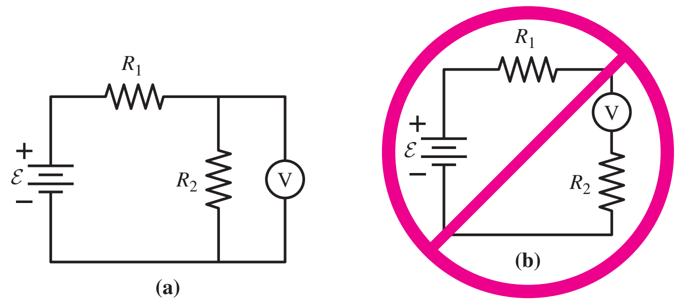


FIGURE 25.14 Correct (a) and incorrect (b) ways to measure the voltage across R_2.

CONCEPTUAL EXAMPLE 25.1 Measuring Voltage

What should be the electrical resistance of an ideal voltmeter?

EVALUATE Before we attach the voltmeter in Fig. 25.14*a*, the battery "sees" the resistors R_1 and R_2 in series. When we connect a meter with resistance R_m, then we've got a parallel combination of R_1 and R_m where before we had just R_1. Since two parallel resistors have a lower resistance than either of the individual resistors, the overall circuit current increases—and so does the voltage across R_1, which carries the total current. That in turn leaves the voltage across R_2 *lower* than before. Even if the meter is perfectly accurate, the voltage it reads will be lower than it was before we connected the meter.

How can we avoid this effect? By giving the meter a high resistance—ideally, infinite resistance, so the meter won't draw any current and therefore won't affect the circuit. The meter will therefore read the voltage that was there before we connected it.

ASSESS Truly infinite resistance is impossible—but as long as the meter's resistance is much larger than resistances in the circuit, the effect of finite meter resistance will be negligible. Modern digital meters come close to the ideal, with resistances of 10 MΩ and higher.

MAKING THE CONNECTION What do you get if you measure the voltage across the 40-Ω resistor in Fig. 25.15 with (a) an ideal voltmeter and (b) a voltmeter whose resistance is 1000 Ω?

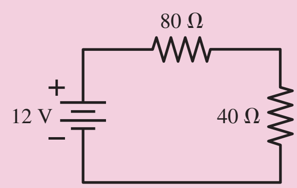


FIGURE 25.15 What's the voltage across the 40-Ω resistor?

EVALUATE (a) The circuit is a simple voltage divider, and Equation 25.2b shows that the voltage across the 40-Ω resistor is one-third of the battery voltage, or 4.00 V. With its infinite resistance, the ideal voltmeter doesn't alter the circuit, so it reads 4.00 V. (b) Connecting the voltmeter gives the circuit of Fig. 25.16; now Equation 25.3b gives 38.5 Ω for the parallel combination of the 1000-Ω meter and 40-Ω resistor. Applying the voltage divider equation (25.2b) gives 3.95 V, 2.5% lower than the ideal voltmeter.

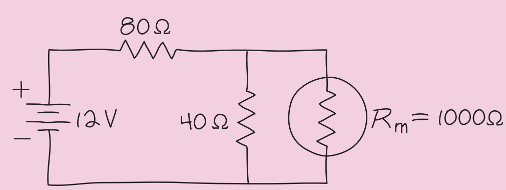
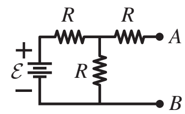


FIGURE 25.16 A nonideal voltmeter (R_m) alters the circuit.

GOT IT? 25.5 If an ideal voltmeter is connected between points A and B in the figure, what will it read? All the resistors have the same resistance R.



Ammeters

An **ammeter** measures the current flowing *through* itself. To measure the current through a circuit component, it's necessary to break the circuit and insert the ammeter in *series* with that component (Fig. 25.17*a*); only then will all the current also go through the meter. Connecting the ammeter across the resistor as in Fig. 25.17*b* is wrong because then the current through the resistor isn't going through the meter.

If the ammeter has any resistance, the total resistance of the circuit will increase with the meter connected in series. This in turn will decrease the current, giving an incorrect reading. So an ideal ammeter should have zero resistance. In practice, ammeter resistance should be much lower than typical resistances in the circuit being measured.

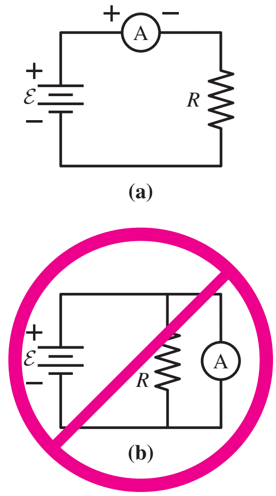


FIGURE 25.17 Correct (a) and incorrect (b) ways to connect an ammeter.

✓TIP Watch Your Language

A voltmeter measures potential difference *between* two points; hence, we connect it *across*—that is, in parallel with—the circuit element whose voltage we wish to measure. An ammeter measures the current *through* itself; hence, we connect it in *series* with the circuit element whose current we wish to measure. If you get used to voltages appearing *across* things and currents flowing *through* them, you'll have no trouble connecting meters. The ways to connect meters, and the words *across* for voltage and *through* for current, go back to the definitions of potential difference as a property of two points and of current as a flow.

Ohmmeters and Multimeters

Often we want the resistance of a particular component. Connecting a known voltage in series with an ammeter and the unknown resistance gives both current and voltage, letting us calculate the unknown resistance. A meter used for this purpose can be calibrated directly in ohms even though it's really measuring current; it's then an **ohmmeter**. The functions of voltmeter, ammeter, and ohmmeter are often combined in a single instrument called a **multimeter**.

25.5 Capacitors in Circuits

So far we've considered only circuits with steady current. A flashlight is a good example: Turn it on and current starts almost instantaneously, and it then continues flowing steadily until you turn the flashlight off.

Capacitors alter this picture, causing circuit quantities to change more slowly. Recall that a capacitor is a pair of insulated conductors with charge and voltage related by $Q = CV$, where Q is the magnitude of the charge on either conductor, V is the potential difference between them, and C is the capacitance. Because charge and voltage are proportional in a capacitor, a change in voltage requires a change in charge. Charge changes when current flows through the wires connecting the capacitor to the rest of a circuit, and the magnitude of the current gives the rate at which capacitor charge increases or decreases. Since the current in any real circuit is finite, the charge on the capacitor cannot change instantaneously. But capacitor voltage is proportional to charge, so:

The voltage across a capacitor cannot change instantaneously.

This statement is the key to understanding circuits with capacitors. It says that the voltage on a capacitor can't jump abruptly from one value to another; mathematically, capacitor voltage V_C must be a continuous function of time, its derivative always finite. Just how rapidly the voltage can change depends on the capacitance and other circuit quantities, as we'll now see.

We consider an ***RC* circuit**, one that includes a resistor and capacitor. *RC* circuits are ubiquitous, appearing everywhere from microbiological structures to stereo amplifiers to giant energy-storage systems. We examine separately the two cases in which the capacitor is (1) charging and (2) discharging.

The *RC* Circuit: Charging

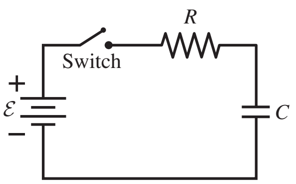


FIGURE 25.18 An *RC* circuit. The switch is closed at time $t = 0$.

In Fig. 25.18 the capacitor is initially uncharged, so the voltage across it is zero. Closing the switch connects the left side of the resistor to the battery, bringing its potential immediately to $\mathcal{E}$ volts (we're taking $V = 0$ at the negative battery terminal). The right end of the resistor remains at the same voltage as the upper capacitor plate—and that's still zero because the voltage across the capacitor can't change instantaneously. So there's $\mathcal{E}$ volts across the resistor, and therefore a current $I = \mathcal{E}/R$ through it. This current delivers positive charge to the upper capacitor plate and negative charge to the lower plate.

As charge accumulates on the plates, the capacitor voltage increases proportionately. But the capacitor and resistor voltages sum to the battery voltage $\mathcal{E}$, so as the capacitor voltage increases, the resistor voltage drops. By Ohm's law, the resistor current $I = V/R$ drops as well. This in turn decreases the *rate* at which charge accumulates in the capacitor. The capacitor voltage continues to increase as charge accumulates, but at an ever-slower rate.

Eventually the capacitor voltage approaches the battery voltage, and the resistor voltage and current tend to zero; so, therefore, does the rate at which charge accumulates on the capacitor. The whole system approaches a final state in which the capacitor is charged to the full battery voltage and the current in the circuit is zero. Figure 25.19 summarizes the interplay among current, charge, and voltage.

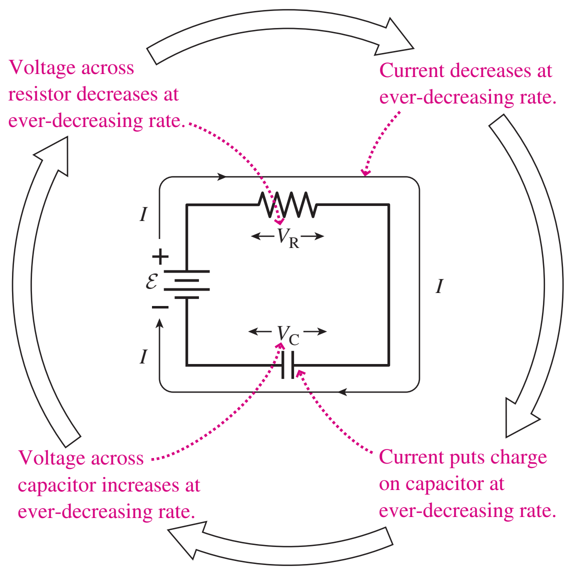


FIGURE 25.19 Interrelationships among quantities in a charging *RC* circuit.

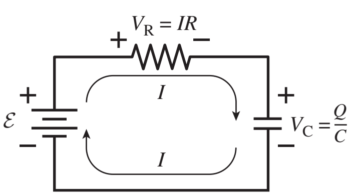


FIGURE 25.20 Voltage changes in a charging *RC* circuit.

We can analyze the circuit of Fig. 25.18 quantitatively using the loop law. Going clockwise around the loop, we first encounter a voltage increase $\mathcal{E}$ across the battery, then a drop IR across the resistor, then a drop V_C from the upper to lower capacitor plate (Fig. 25.20). But $V_C = Q/C$, so the loop equation becomes

$$\mathcal{E} - IR - \frac{Q}{C} = 0$$

This equation contains two unknowns, I and Q, but they're related because the current is the rate at which charge is accumulating on the capacitor: $I = dQ/dt$. To use this relation, we take the time derivative of the loop equation:

$$-R\frac{dI}{dt} - \frac{1}{C}\frac{dQ}{dt} = 0$$

The battery voltage $\mathcal{E}$ doesn't appear here because it's constant. Using $I = dQ/dt$ and rearranging the equation give

$$\frac{dI}{dt} = -\frac{I}{RC} \tag{25.4}$$

This equation shows that the rate of change of current is proportional to the current itself. Equations like this arise whenever a quantity changes at a rate proportional to the quantity itself. Population growth, the increase of money in a bank account, and the decay of a radioactive element are all described by similar equations.

Like the equation for simple harmonic motion in Chapter 13, Equation 25.4 is a *differential equation* because the unknown quantity I occurs in a derivative. The solution to a differential equation isn't a single number but a function expressing the relation between the unknown quantity—in this case current—and the independent variable—in this case time. We can solve this particular differential equation by multiplying both sides by dt/I in order to collect all terms involving I on one side of the equation. This gives

$$\frac{dI}{I} = -\frac{dt}{RC}$$

We can then integrate both sides, noting that RC is constant:

$$\int_{I_0}^{I}\frac{dI}{I} = -\frac{1}{RC}\int_0^t dt$$

where $I_0 = \mathcal{E}/R$ is the initial current at the time $t = 0$ just after the switch is closed and where the integration runs to an arbitrary time t. The integral on the left is the natural logarithm of I, and on the right it's just t. Then we have

$$\ln\left(\frac{I}{I_0}\right) = -\frac{t}{RC}$$

where we used $\ln I - \ln I_0 = \ln(I/I_0)$. To get an equation for I we exponentiate both sides, recalling that $e^{\ln x} = x$. This gives $I/I_0 = e^{-t/RC}$, or, since $I_0 = \mathcal{E}/R$,

$$I = \frac{\mathcal{E}}{R}e^{-t/RC} \tag{25.5}$$

Thus the current in the circuit decreases exponentially with time, in agreement with our qualitative analysis. The capacitor voltage is $V_C = \mathcal{E} - V_R$, or, since $V_R = IR = \mathcal{E}e^{-t/RC}$,

$$V_C = \mathcal{E}(1 - e^{-t/RC}) \qquad (RC \text{ circuit, charging}) \tag{25.6}$$

Equation 25.6 shows the capacitor voltage starting at zero and rising, with its rate of rise ever slowing as it gradually approaches the battery voltage $\mathcal{E}$—just as we reasoned in our qualitative analysis. Figure 25.21 plots capacitor voltage and current using the equations we've just derived.

When is the capacitor fully charged? Never, according to our equations! But the rate at which it approaches full charge is determined by the so-called **time constant**, RC—a characteristic time for changes to occur in a circuit containing a capacitor. Equation 25.6 shows that in one time constant, the voltage rises to $\mathcal{E}(1 - 1/e)$, or to about two-thirds of the battery voltage. A practical rule of thumb says that in five time constants $(t = 5RC)$ a capacitor is 99% charged (see Exercise 33). The RC time constant clarifies our statement that the voltage across a capacitor can't change instantaneously. We can now say that the voltage can't change appreciably in times small compared with the time constant. On the other hand, after many time constants, we'll find essentially no current flowing to the capacitor. We've shown quantitatively the role of the time constant RC by marking the time in units of RC on Fig. 25.21.

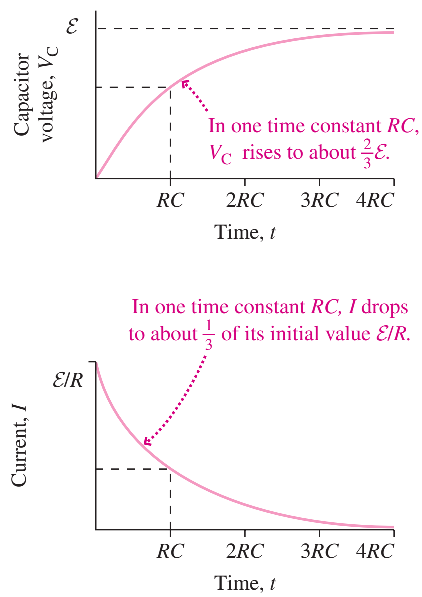


FIGURE 25.21 Time dependence of capacitor voltage and circuit current in a charging RC circuit. The approximate values $\frac{2}{3}$ and $\frac{1}{3}$ are actually $1 - 1/e$ and $1/e$, respectively.

Resistors and capacitors are available in a wide range of values, so practical values for *RC* span many orders of magnitude. *RC* circuits with time constants from microseconds to hours are widely used in electronic devices to control the rates at which electric quantities vary. For example, circuits with *RC* many times the sixtieth-of-a-second period of standard AC power produce steady, direct-current power for audio and video equipment. Equalizers in audio systems are variable resistances in *RC* circuits; changing the resistance changes the time constant and therefore the way the circuit handles rapidly changing audio signals. Sometimes, though, the time constant can be a nuisance. Capacitance in audio systems can limit high-frequency response, decreasing the quality of music reproduction. With computer speeds in the GHz range—meaning basic operations occur billions of times a second—even tiny *RC* time constants associated with the resistance of wires and the capacitance of adjacent conductors can cause trouble.

The *RC* Circuit: Discharging

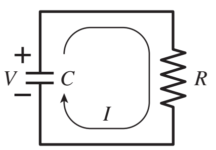


FIGURE 25.22 A discharging *RC* circuit.

Suppose we connect a charged capacitor across a resistor, as shown in Fig. 25.22. If the capacitor voltage is initially V_0, then when the circuit is connected, this voltage will drive a current $I_0 = V_0/R$ through the resistor. This current transfers charge from the positive to the negative capacitor plate, lowering the charge on the capacitor. Since capacitor charge and voltage are proportional, the capacitor voltage drops, too. So, therefore, does the current and therefore the rate at which the capacitor discharges. We expect both the voltage and current in this circuit to decay toward zero. In terms of energy, that happens because the energy stored in the capacitor's electric field is gradually dissipated as heat in the resistor.

The loop equation for Fig. 25.22 is particularly simple; going clockwise, we have $Q/C - IR = 0$, where the two terms are the voltage changes across the capacitor and resistor, respectively. Since we've indicated positive current in Fig. 25.22 in the direction that would *reduce* the capacitor charge Q, the rate of change dQ/dt and the current have opposite signs: $I = -dQ/dt$. Differentiating our loop equation and substituting this expression for I give $dI/dt = -I/RC$. This is Equation 25.4; the solution is therefore Equation 25.5, but with $I_0 = V_0/R$ instead of $\mathcal{E}/R$:

$$I = \frac{V_0}{R}e^{-t/RC} \tag{25.7}$$

In this circuit the capacitor and resistor voltages are the same, since the two are in parallel (in this simplest of circuits, they're also in series). Because the resistor voltage and the current are proportional, the voltage across the capacitor and resistor is

$$V = V_0 e^{-t/RC} \qquad (RC \text{ circuit, discharging}) \tag{25.8}$$

Equations 25.7 and 25.8 show that the capacitor discharges with the same characteristic time constant *RC* that governs its charging.

EXAMPLE 25.5 Charging Capacitors: A Camera Flash

A camera flash gets its energy from a 150-μF capacitor and requires 170 V to fire. If the capacitor is charged by a 200-V source through an 18-kΩ resistor, how long must the photographer wait between flashes? Assume the capacitor is fully discharged with each flash.

INTERPRET This is a problem about a charging capacitor, and we want to find the time to reach a given voltage.

DEVELOP Equation 25.6, $V_C = \mathcal{E}(1 - e^{-t/RC})$, gives the voltage across a charging capacitor, so our plan is to solve this equation for the time t.

EVALUATE First we solve for the exponential term that contains the time:

$$e^{-t/RC} = 1 - \frac{V_C}{\mathcal{E}}$$

Then we take the natural logarithm of both sides, recalling that $\ln e^x = x$, so

$$-\frac{t}{RC} = \ln\left(1 - \frac{V_C}{\mathcal{E}}\right)$$

Solving for t and setting $V_C = 170$ V, $\mathcal{E} = 200$ V, $R = 18$ kΩ, and $C = 150\ \mu$F give

$$t = -RC\ln\left(1 - \frac{V_C}{\mathcal{E}}\right) = 5.1 \text{ s}$$

ASSESS The time constant here is $RC = 2.7$ s, and 170 V is well over two-thirds of the 200-V source. Therefore, we expect a charging time longer than one time constant. Our 5.1-s answer is nearly $2RC$. Problem 66 explores energy and power in this circuit. ■

RC Circuits: Long- and Short-Term Behavior

It's not always necessary to solve exponential equations in analyzing *RC* circuits. If we're interested only in times much shorter than the time constant, then it's enough to remember that the voltage across a capacitor can't change instantaneously. And after many time constants, a capacitor has essentially reached its final voltage, and no current is flowing to it. These conditions are sufficient to analyze circuits on short and long time scales.

TACTICS 25.2 Analyzing Long- and Short-Term Behavior of *RC* Circuits

Short-Term Behavior

For times much shorter than the time constant *RC*, capacitor voltage remains essentially unchanged. Therefore, you can replace the capacitor with a short circuit if it's uncharged or, if it's charged, with a battery whose emf is the capacitor's initial voltage. Then solve the circuit using the techniques of Section 25.2 or 25.3.

Long-Term Behavior

For times much longer than *RC*, no current is flowing to a capacitor. Therefore, you can replace the capacitor with an open circuit, and again solve using earlier techniques.

EXAMPLE 25.6 An *RC* Circuit: Long and Short Times

The capacitor in Fig. 25.23*a* is initially uncharged. Find the current through R_1 (a) the instant the switch is closed and (b) a long time after the switch is closed.

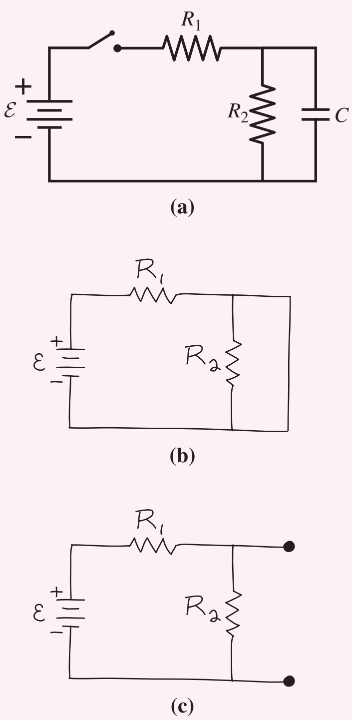


FIGURE 25.23 Original (a) and equivalent short-term (b) and long-term (c) circuits for Example 25.6.

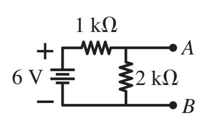
INTERPRET We interpret "the instant the switch is closed" to mean a time much shorter than the time constant *RC*, and "a long time" to mean a time much longer than *RC*. Then this is a problem involving the long- and short-term behavior of a circuit containing a capacitor.

DEVELOP We follow Tactics 25.2 and first redraw the circuit with the capacitor replaced by a short circuit (Fig. 25.23*b*). Solving this circuit will give the current in R_1 right after the switch is closed. For the long-term behavior we redraw the circuit with the capacitor an open circuit (Fig. 25.23*c*).

EVALUATE There can't be any voltage across a short circuit—a perfect conductor—so there's no voltage across R_2 in Fig. 25.23*b*. Thus for part (a) the entire battery voltage appears across R_1, giving a current $I = \mathcal{E}/R_1$. In Fig. 25.23*c* we have two resistors in series and the current in both is $I = \mathcal{E}/(R_1 + R_2)$, our answer to part (b).

ASSESS The current through R_1 starts out at $\mathcal{E}/R_1$ and gradually drops to $\mathcal{E}/(R_1 + R_2)$. That makes sense because the uncharged capacitor initially "shorts out" R_2, making it irrelevant. But as the capacitor charges, current starts flowing through R_2 and its presence is "felt." Without solving more complicated equations, we can't describe the intermediate behavior of the circuit, but getting the short- and long-term behavior is straightforward. ■

GOT IT? 25.6 A capacitor is charged to 12 V and then connected between points *A* and *B* in the figure, with its positive plate at *A*. What's the current through the 2-kΩ resistor (a) immediately after the capacitor is connected and (b) a long time after it's connected?



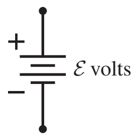
CHAPTER 25 SUMMARY

Big Picture

The big idea here is the **electric circuit**—an interconnection of electric components that usually includes one or more sources of electric energy, such as batteries.

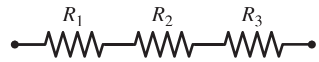
Key Concepts and Equations

A source of **emf** (or an emf, for short) is a battery or other device that imparts energy to electric charge flowing through it. The value of the emf, $\mathcal{E}$, is the energy imparted per unit charge, measured in volts. An ideal emf maintains a fixed potential difference (voltage) across its terminals.



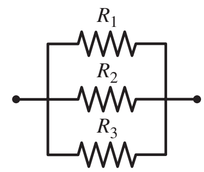
Resistors in series add:

$$R_{\text{series}} = R_1 + R_2 + R_3 + \cdots$$

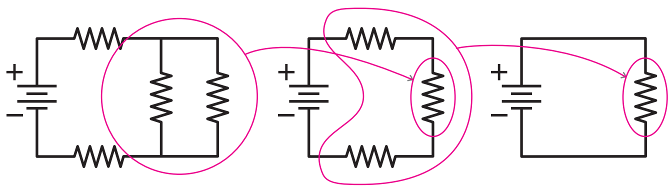


Resistors in parallel add reciprocally:

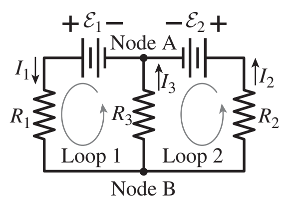
$$\frac{1}{R_{\text{parallel}}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots$$



Simple circuits are analyzed by evaluating series and parallel combinations.



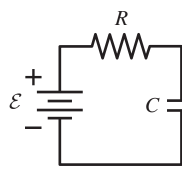
To analyze more complicated circuits, use Kirchhoff's node and loop laws.



Current into node A is 0:
$-I_1 + I_2 + I_3 = 0$
Voltage changes around loops sum to 0:
loop 1: $\mathcal{E}_1 - I_1R_1 - I_3R_3 = 0$
loop 2: $\mathcal{E}_2 + I_2R_2 - I_3R_3 = 0$

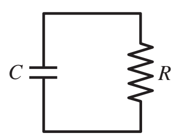
Capacitors result in time-changing behavior of circuit quantities.

The **time constant** RC governs the time scales.



$V_C = \mathcal{E}(1 - e^{-t/RC})$

Charging

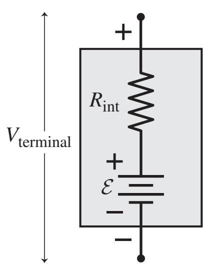


$V_C = V_0e^{-t/RC}$

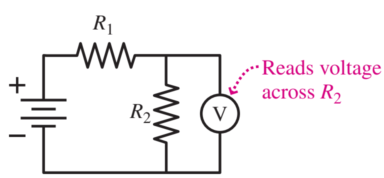
Discharging

Applications

Batteries and other electric-energy sources have **internal resistance**. When they supply current, their terminal voltage is therefore less than their rated voltage $\mathcal{E}$.

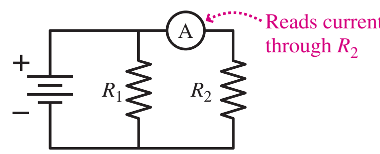


A **voltmeter** measures the voltage *across* its two terminals. It goes in parallel with the component whose voltage you want to measure.



An ideal voltmeter has infinite resistance.

An **ammeter** measures the current *through* itself. It goes in series with the component whose current you want to measure.



An ideal ammeter has zero resistance.

For Thought and Discussion

1. Are household electrical outlets connected in series or parallel? How do you know?
2. All the resistors in Fig. 25.24 have the same resistance. In which circuits does the battery supply the same current?

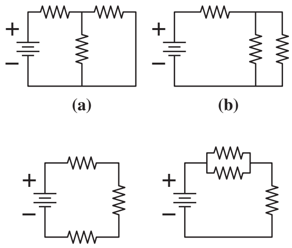


FIGURE 25.24 For Thought and Discussion 2

3. Can the voltage across a battery's terminals differ from the battery's rated voltage? Explain.
4. Can the voltage across a battery's terminals be higher than the battery's rated voltage? Explain.
5. In some cities, streetlights are wired in such a way that when one burns out, they all go out. Are the lights in series or parallel?
6. When the switch in Fig. 25.25 is open, what's the voltage across the resistor? Across the switch?

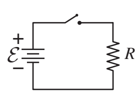


FIGURE 25.25 For Thought and Discussion 6

7. Two identical resistors in series dissipate equal power. How can this be, when electric charge loses energy in flowing through the first resistor?
8. When a large electric load such as a washing machine or oven comes on, lights throughout a house often dim. Why?
9. How would you connect a pair of equal resistors across an ideal battery in order to get the greatest power dissipation?
10. You have a battery whose voltage and internal resistance are unknown. Using an ideal voltmeter and an ideal ammeter, how would you determine each of these characteristics?
11. A student who's confused about voltage and current hooks a nearly ideal ammeter across a car battery. What happens?
12. A student who's confused about voltage and current tries to measure the voltage across a lighted lightbulb by inserting a voltmeter in series with the bulb. What happens to the bulb? Explain.

Exercises and Problems

Exercises

Section 25.1 Circuits, Symbols, and Electromotive Force

13. Sketch a circuit diagram for a circuit that includes a resistor R_1 connected to the positive terminal of a battery, a pair of parallel resistors R_2 and R_3 connected to the lower-voltage end of R_1 and then returned to the battery's negative terminal, and a capacitor across R_2.
14. Sketch a diagram for a circuit consisting of two batteries, a resistor, and a capacitor, all in series. Does the circuit description allow you any flexibility?
15. Resistors R_1 and R_2 are in series, and the series combination is in parallel with R_3. This parallel combination is connected across a battery. Draw a diagram of this circuit.
16. What's the emf of a battery that delivers 27 J of energy as it moves 3.0 C between its terminals?
17. A 1.5-V battery stores 4.5 kJ of energy. How long can it light a flashlight bulb that draws 0.60 A?
18. If you accidentally leave your car headlights (current 5 A) on for an hour, how much energy drains from the car's 12-V battery?

Section 25.2 Series and Parallel Circuits

19. A 47-kΩ resistor and a 39-kΩ resistor are in parallel, and the pair is in series with a 22-kΩ resistor. What's the resistance of the combination?
20. What resistance should you place in parallel with a 56-kΩ resistor to make an equivalent resistance of 45 kΩ?
21. A defective starter motor draws 300 A from a car's 12-V battery, dropping the battery terminal voltage to 6 V. A good starter should draw only 100 A. What will the battery terminal voltage be with a good starter?
22. Find the internal resistance of the battery in Exercise 21.
23. When a 9-V battery is temporarily short-circuited, a 200-mA current flows. What's the battery's internal resistance?
24. You have a 1.0-Ω, a 2.0-Ω, and a 3.0-Ω resistor. What equivalent resistances can you form using all three?

Section 25.3 Kirchhoff's Laws and Multiloop Circuits

25. Find all three currents in the circuit of Fig. 25.13, but now with $\mathcal{E}_2 = 1.0$ V.
26. What's the current through the 3-Ω resistor in Fig. 25.26? (*Hint:* This is trivial. Can you see why?)

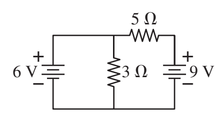


FIGURE 25.26 Exercise 26

27. Find I_2 in Example 25.4 for the case $\mathcal{E}_2 = 2.0$ V.

Section 25.4 Electrical Measurements

28. A voltmeter with 200-kΩ resistance is used to measure the voltage across the 10-kΩ resistor in Fig. 25.27. By what percentage is the measurement in error because of the finite meter resistance?

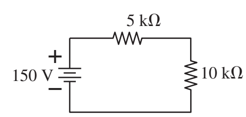


FIGURE 25.27 Exercises 28 and 29

29. An ammeter with 100-Ω resistance is inserted in the circuit of Fig. 25.27. By what percentage is the measured current in error because of the nonzero meter resistance?
30. A new mechanic foolishly connects an ammeter with 0.1-Ω resistance directly across a 12-V car battery with internal resistance 0.01 Ω. What's the power dissipation in the meter? (No wonder it gets destroyed!)

Section 25.5 Capacitors in Circuits

31. Show that the quantity RC has the units of time (seconds).
32. If capacitance is in μF, what will be the units of the time constant RC when resistance is in (a) Ω, (b) kΩ, and (c) MΩ? (Your answers eliminate the need for tedious power-of-10 conversions.)
33. Show that a capacitor is charged to approximately 99% of the applied voltage in five time constants $(5RC)$.
34. An uncharged 10-μF capacitor and a 470-kΩ resistor are in series, and 250 V is applied across the combination. How long does it take the capacitor voltage to reach 200 V?
35. Find an expression for the voltage across the capacitor in Example 25.6 when it's fully charged.

Problems

36. In Fig. 25.28, all resistors have the same value, R. What will be the resistance measured (a) between A and B or (b) between A and C?

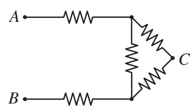


FIGURE 25.28 Problems 36 and 37

37. In Fig. 25.28, take all resistors to be 1 kΩ. Find the current in the vertical resistor when a 6.0-V battery is connected between A and B.
38. Three 1.5-V batteries, with internal resistances 0.01 Ω, 0.1 Ω, and 1 Ω, each have 1-Ω resistors connected across their terminals. What's the voltage between each battery's terminals, to three significant figures?
39. A partially discharged car battery can be modeled as a 9-V emf in series with a 0.08-Ω internal resistance. Jumper cables connect this battery to a fully charged battery, modeled as a 12-V emf in series with a 0.02-Ω internal resistance. The cables connect + to + and − to −. What current flows through the discharged battery?
40. Your company is overstocked on 50-Ω, $\frac{1}{2}$-W resistors. Your project requires 50-Ω resistors that can be safely connected across a 12-V power source. How many of the available resistors will you need, and how will you connect them?
41. A 6.0-V battery has internal resistance 2.5 Ω. If the battery is short-circuited, what's the rate of energy dissipation in its internal resistance?
42. How many 100-W, 120-V lightbulbs can be connected in parallel before they trip a 20-A circuit breaker?
43. **BIO** You company is designing a battery-based backup power source, and your job is to assess its safety. You know that under damp or sweaty conditions, the resistance between two points of unbroken skin on the human body can be as low as 500 Ω. Your product uses a 72-V battery whose internal resistance is 100 Ω. Is it capable of passing a fatal 100 mA (Table 24.3) through a damp human body?
44. Take $\mathcal{E} = 12$ V and $R_1 = 270\ \Omega$ in Fig. 25.4. (a) What's the resistance R_2 if there's 4.5 V across it? (b) What will be the power dissipation in R_2?
45. In Fig. 25.29, R_1 is a variable resistor and the other two resistors have equal resistances R. (a) Find an expression for the voltage across R_1, and (b) sketch a graph of this voltage as R_1 varies from 0 to $10R$.

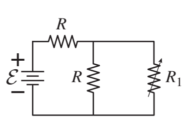


FIGURE 25.29 Problem 45

46. In the circuit of Fig. 25.30, find (a) the current supplied by the battery and (b) the current through the 6-Ω resistor.

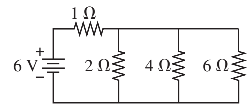


FIGURE 25.30 Problems 46 and 47

47. In Fig. 25.30, how much power is dissipated in the 4-Ω resistor?
48. What's the ammeter reading in Fig. 25.31?

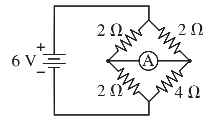


FIGURE 25.31 Problem 48

49. In Fig. 25.32, find the equivalent resistance measured between A and B.

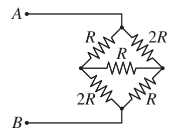


FIGURE 25.32 Problem 49

50. Find all three currents in the circuit of Fig. 25.13 with the values given, but with battery $\mathcal{E}_2$ reversed.
51. The voltage across the 30-kΩ resistor in Fig. 25.33 is measured with (a) a 50-kΩ voltmeter, (b) a 250-kΩ voltmeter, and (c) a 10-MΩ digital meter. What does each read, to two significant figures?

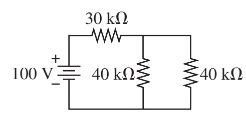


FIGURE 25.33 Problem 51

52. In Fig. 25.34, what are the meter readings when an ideal (a) voltmeter or (b) ammeter is connected between A and B?

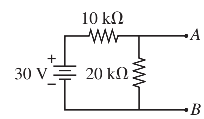


FIGURE 25.34 Problem 52

53. A resistor draws 1.00 A from an ideal 12.0-V battery. (a) If an ammeter with 0.10-Ω resistance is inserted in the circuit, what will it read? (b) If this current is used to calculate the resistance, by what percent will the result be in error?
54. The voltage across a charging capacitor in an RC circuit rises to $1 - 1/e$ of the battery voltage in 5.0 ms. (a) How long will it take to reach $1 - 1/e^3$ of the battery voltage? (b) If the capacitor is charging through a 22-kΩ resistor, what's the capacitance?
55. **BIO** You're designing an external defibrillator that discharges a capacitor through the patient's body, providing a pulse that stops ventricular fibrillation. Specifications call for a capacitor storing 250 J of energy; when discharged through a body with 40-Ω transthoracic resistance, the capacitor voltage is to drop to half its initial value in 10 ms. Determine the capacitance (to the nearest

10 μF) and initial capacitor voltage (to the nearest 100 V) that meet these specs.

56. A capacitor used to provide steady voltages in the power supply of a stereo amplifier charges rapidly to 35 V every 1/60 second. It must then hold that voltage to within 1.0 V for the next 1/60 s while it discharges through the amplifier. If the amplifier draws 1.2 A from the 35-V supply, (a) what's its effective resistance, and (b) what capacitance is needed?

57. A capacitor is charged until it holds 5.0 J of energy, then connected across a 10-kΩ resistor. In 8.6 ms, the resistor dissipates 2.0 J. Find the capacitance.

58. In Fig. 25.35 the 2.0-μF capacitor is charged to 150 V, while the 1.0-μF capacitor is initially uncharged. Switch S is then closed. Find the total energy dissipated in the resistor as the circuit comes to equilibrium. (*Hint:* Think about charge conservation.)

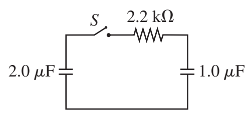


FIGURE 25.35 Problem 58

59. For the circuit of Example 25.6, take $\mathcal{E} = 100$ V, $R_1 = 4.0\,\text{k}\Omega$, and $R_2 = 6.0\,\text{k}\Omega$, and assume the capacitor is initially uncharged. Find the capacitor voltage and the currents in both resistors (a) just after the switch is closed, and (b) a long time after the switch is closed. Long after the switch is closed it's reopened. What are V_C, I_1, and I_2 (c) just after this switch opening, and (d) a long time later?

60. In Fig. 25.36, the switch is initially open and both capacitors are initially uncharged. All resistors have the same value R. Find expressions for the current in R_2 (a) just after the switch is closed, and (b) a long time after the switch is closed.

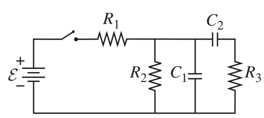


FIGURE 25.36 Problem 60

61. A battery's voltage is measured as 4.36 V with a voltmeter whose resistance is 1000 Ω. When measured with a 1500-Ω meter, it's 4.41 V. Find (a) the battery's voltage and (b) its internal resistance.

62. Find the resistance needed in an RC circuit to bring a 20-μf capacitor from zero charge to 45% charge in 140 ms.

63. Suppose the currents into and out of a circuit node differ by 1 μA. If the node consists of a small metal sphere with diameter 1 mm, how long would it take for the electric field around the node to reach the 3-MV/m breakdown field in air?

64. Show that a battery delivers the most power when the load resistance across its terminals is equal to its internal resistance. (This is not the way to treat a battery, but it's the basis for load matching in amplifiers; see Problem 65.)

65. You're writing the instruction manual for a stereo amplifier with a maximum output of 100 W. The amplifier can be modeled as an emf in series with an 8-Ω resistance. What should you specify for the loudspeaker resistance to be used with the amplifier? How much power can the amplifier deliver to a speaker with half the optimum resistance?

66. Show that only half the total energy drawn from a battery in charging an RC circuit ends up stored in the capacitor. (*Hint:* What happens to the rest? You'll need to integrate.)

67. Find the equivalent resistance between A and B for the circuits in Fig. 25.37.

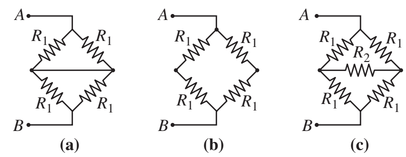


FIGURE 25.37 Problem 67

68. A 50-Ω resistor is connected across a battery, and a 26-mA current flows. When the resistor is replaced with a 22-Ω resistor, 43 mA flows. Find the battery's voltage and internal resistance.

69. Obtain an expression for the rate of increase (dV/dt) of the voltage across a charging capacitor in an RC circuit. Evaluate your result at time $t = 0$, and show that if the capacitor continued charging steadily at this rate, it would reach full charge in exactly one time constant.

70. The circuit in Fig. 25.38 extends forever to the right, and all the resistors have the same value R. Show that the equivalent resistance measured across the two terminals at left is $R(1 + \sqrt{5})/2$. (*Hint:* You don't need to sum an infinite series.)

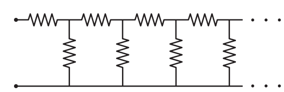


FIGURE 25.38 Problem 70

71. Figure 25.39 shows the voltage across a capacitor that's charging through a 4700-Ω resistor in the circuit of Fig. 25.18. Use the graph to determine (a) the battery voltage, (b) the time constant, and (c) the capacitance.

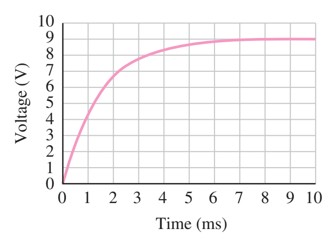


FIGURE 25.39 Problem 71

72. **BIO** Figure 25.40 shows a portion of a circuit used to model muscle cells and neurons. All resistors have the same value $R = 1.5\,\text{M}\Omega$, and the emfs are $\mathcal{E}_1 = 75\,\text{mV}$, $\mathcal{E}_2 = 45\,\text{mV}$, and $\mathcal{E}_3 = 20\,\text{mV}$. Find the current through $\mathcal{E}_3$, including its direction.

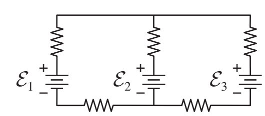


FIGURE 25.40 Problems 72 and 73

73. **BIO** An electrochemical impulse traveling along the cell modeled in Fig. 25.40 changes the value of $\mathcal{E}_3$ so now it supplies a 40-nA upward current. Assuming the rest of the circuit remains as described in Problem 72, what's the new value of $\mathcal{E}_3$?

74. A parallel-plate capacitor has plates of area 10 cm^2 separated by a 0.10-mm layer of glass insulation with resistivity $\rho = 1.2\times10^{13}\ \Omega\cdot\text{m}$ and dielectric constant $\kappa = 5.6$. Because of the finite resistivity, charge leaks through the insulation. (a) How

can such a leaky capacitor be represented in a circuit diagram? (b) Find the time constant for this capacitor to discharge through its insulation, and show that it depends only on the properties of the insulating material and not on its dimensions.

75. Write the node and loop equations for the circuit in Fig. 25.23*a* (Example 25.6), and find the time constant.
76. In Problem 60, take $C_1 = C_2 = C$, and find the current through R_2 as a function of time. (*Hint:* Use the node and loop laws to get a differential equation for the current, and use the initial conditions on current and its derivative to evaluate the constants of integration.)
77. You're about to purchase a battery. Normally, batteries are rated in ampere-hours—the total charge they can deliver. Your application calls for a 5-A·h battery. But the 6-V battery you see while shopping online is rated at 50 watt-hours. Will it work?

Passage Problems

BIO *Stray voltage* is a serious problem on dairy farms, often resulting from corroded wiring or poor wiring practices. These conditions can produce several volts between the ground and metal watering bowls, feed troughs, or milking equipment. Cows feel shocks that make them nervous, reducing milk output and sometimes leading to mammary gland infections. As a result, farmers can face serious financial losses. Figure 25.41 shows a typical stray-voltage situation, with the source of stray voltage modeled as a 6-V emf in series with a 1-kΩ resistance.

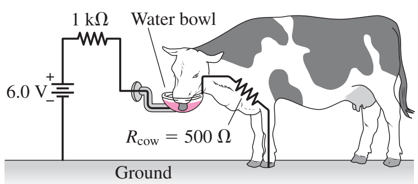


FIGURE 25.41 Stray voltage can bankrupt a dairy farm (Passage Problems 78–81)

78. The current through the 500-Ω cow will be
    a. 3 mA.
    b. 4 mA.
    c. 6 mA.
    d. 12 mA.
79. The voltage across the cow shown is
    a. 2 V.
    b. 4 V.
    c. 6 V.
    d. nearly 0 V.
80. In an effort to diagnose the problem, a farmer connects an ideal voltmeter between the water bowl and ground, with the cow absent. The voltmeter reading is
    a. 2 V.
    b. 4 V.
    c. 6 V.
    d. none of the above.
81. To explore the problem further, a farmer connects an ideal ammeter between the water bowl and ground, with the cow absent. The ammeter reading is
    a. 4 mA.
    b. 6 mA.
    c. 12 mA.
    d. infinite.

Answers to Chapter Questions

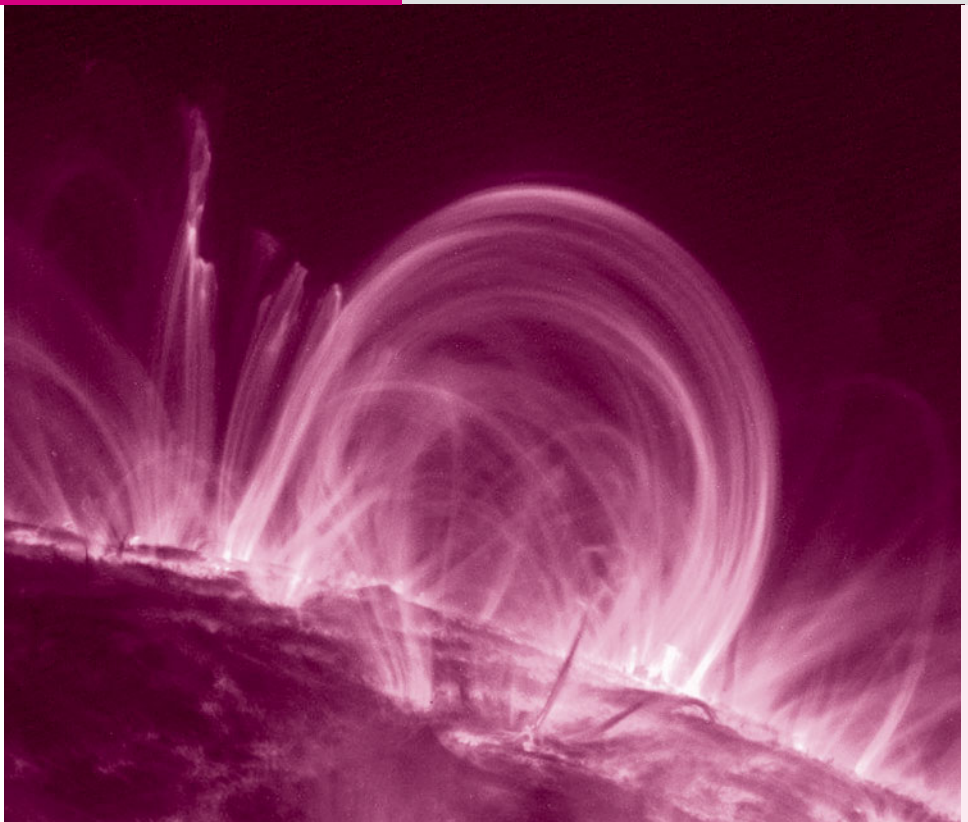
Answer to Chapter Opening Question

Series.

Answers to GOT IT? Questions

25.1. (a) and (b).
25.2. (c) 6 V > (a) 3 V > (b) 0 V.
25.3. $R_a > R_d > R_c > R_b$.
25.4. (a) A is brightest because it carries more current; after A the current splits between B and C. (b) A and B become equally bright, with A dimming and B brightening relative to when C was in the circuit.
25.5. There's no current through the top right resistor, so the voltage is divided evenly across the other two resistors and the meter reads $\frac{1}{2}\mathcal{E}$.
25.6. (a) 6 mA; (b) 2 mA.

26 Magnetism: Force and Field



This ultraviolet image shows delicate loops of million-kelvin ionized gas—plasma—in the Sun's atmosphere. What force shapes the gas into such intricate structures, and why don't we see similar things in Earth's atmosphere?

New Concepts, New Skills

By the end of this chapter you should be able to

- Describe the fundamental nature of magnetism and its relation to electric charge (26.1).
- Explain the relation between magnetic force and magnetic field (26.2).
- Calculate the motion of charged particles in magnetic fields (26.3).
- Calculate magnetic forces on electric currents (26.4).
- Explain the origin of magnetic fields (26.5).
- Recognize the essential role of magnetic dipoles and describe the magnetic field produced by a dipole as well as the interaction of a dipole with an external field (26.6).
- Describe the effects of magnetism in matter (26.7).
- Understand Ampère's law and use it to find the magnetic fields of symmetric currents (26.8).

Connecting Your Knowledge

- This chapter introduces a lot of new material but builds on the fundamental ideas of electric charge and force (20.1, 4.2).
- You should also review vector cross products and line integrals (11.2, 6.2) and uniform circular motion (3.6, 5.3).

People are fascinated with magnets and the mysterious, invisible force they produce. Magnetism plays essential roles in technology and the natural universe. We use magnetism for everything from holding notes on refrigerators to storing computer data to propelling high-speed trains. Earth's magnetism protects us from dangerous solar radiation, which itself originates in violent magnetic storms on the Sun. Without magnetism we wouldn't even see, for light itself results from an interaction between magnetism and electricity. In fact, magnetism and electricity are intimately related, and you'll soon see them as inseparable aspects of the same underlying phenomenon.

26.1 What Is Magnetism?

You know from experience that magnets exert forces on each other and on certain materials, like iron. As we did with the gravitational and electric forces, it's convenient to describe this interaction in terms of a **magnetic field** (symbol $\vec{B}$). One magnet produces a magnetic field, and another responds to the field in its vicinity. We use field lines to picture the field, and we

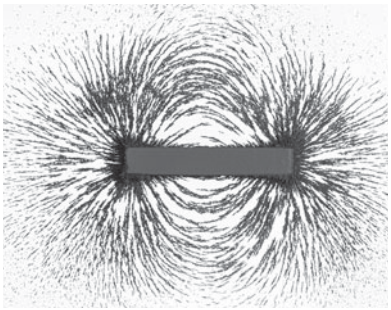


FIGURE 26.1 Iron filings align with the magnetic field, tracing out the field of a bar magnet.

can trace those lines using small iron filings that align with the field (Fig. 26.1). In our illustrations, we'll use color for magnetic field lines to distinguish them from electric fields.

But the magnetism you're most familiar with is only one manifestation of a much more fundamental and universal phenomenon that's intimately linked with electricity. Here we'll go straight to the essence of magnetism; later, we'll see how familiar magnets fit into the big picture.

In Chapter 20 we introduced electric charge, a fundamental property of matter, and described its interactions using the concept of electric field. Magnetism, too, is based in *electric* charge. One crucial point both distinguishes and relates electricity and magnetism:

The phenomena of magnetism involve *moving* electric charge.

In particular, *moving* electric charge is the source of magnetic fields, and *moving* electric charge is what responds to magnetic fields.

26.2 Magnetic Force and Field

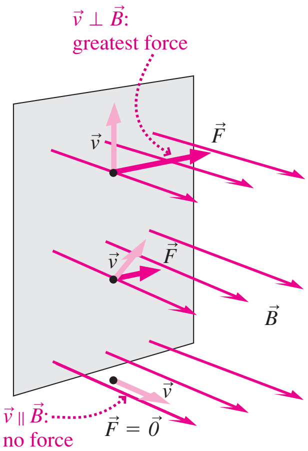


FIGURE 26.2 The magnetic force on a charged particle is perpendicular to both the particle's velocity $\vec{v}$ and the magnetic field $\vec{B}$.

In Chapter 20 we defined the electric field $\vec{E}$ with the equation $\vec{F}_E = q\vec{E}$, where $\vec{F}_E$ is the electric force on a charge q. Now consider a region where there's no electric field, but where there is a magnetic field. You could confirm the presence of the field and determine its direction with a compass, which is just a small magnet free to pivot into alignment with the field. Or, more fundamentally, you could explore the behavior of an electric point charge q in this field. If the charge is at rest, nothing happens. But if it's moving, it experiences a **magnetic force** as shown in Fig. 26.2. Experiment shows that:

1. The magnetic force is always at right angles to both the velocity $\vec{v}$ of the charge and the magnetic field $\vec{B}$.
2. The magnitude of the force is proportional to the product of the charge q, its speed v, and the magnetic field strength B.
3. The force is greatest when the charge moves at right angles to the field and is zero for motion parallel to the field. In general, the force is proportional to $\sin\theta$, where θ is the angle between the velocity $\vec{v}$ and the field $\vec{B}$.

Putting these facts together lets us write the magnetic force compactly using the vector cross product introduced in Chapter 11:

$$\vec{F}_B = q\vec{v} \times \vec{B} \qquad \text{(magnetic force)} \tag{26.1}$$

Recall that the cross product $\vec{v} \times \vec{B}$ is a vector of magnitude $vB\sin\theta$, so the magnitude of the magnetic force is

$$|\vec{F}_B| = |q|vB\sin\theta$$

The direction of $\vec{v} \times \vec{B}$ is given by the right-hand rule (Fig. 26.3), and Equation 26.1 shows that the magnetic force has that same direction for positive q and the opposite direction for negative q.

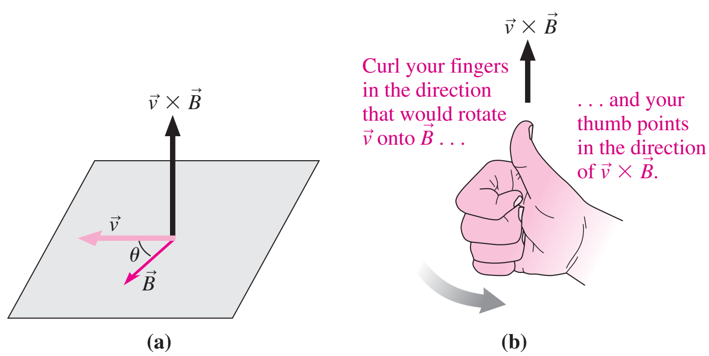
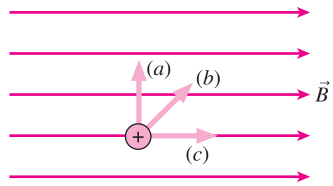


FIGURE 26.3 Finding the direction of the cross product $\vec{v} \times \vec{B}$ with the right-hand rule.

Equation 26.1 shows that the units of magnetic field are N·s/(C·m), a unit given the name **tesla** (T) after the Serbian-American inventor Nikola Tesla (1856–1943). One tesla is a strong field, and a smaller unit called the gauss (G), equal to 10^{-4} T, is often used. Earth's magnetic field is a little less than 1 G, while the field of a refrigerator magnet is about 100 G. The fields used in magnetic resonance imaging (MRI) may be as strong as several tesla, while the incredibly dense, rapidly rotating collapsed stars called magnetars have fields up to 10^{11} T.

GOT IT? 26.1 The figure shows a proton in a magnetic field. For which of the three proton velocities shown will the magnetic force be greatest? What will be the direction of the force in all three cases?



EXAMPLE 26.1 Finding the Magnetic Force: Steering Protons

Figure 26.4 shows three protons entering a 0.10-T magnetic field. All three are moving at 2.0 Mm/s. Find the magnetic force on each.

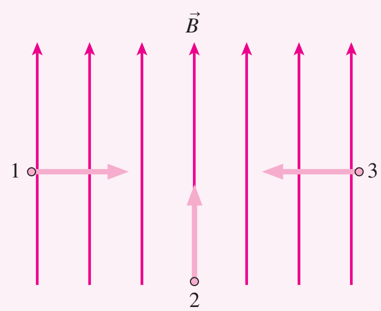


FIGURE 26.4 What's the magnetic force on each proton?

INTERPRET This problem is about the magnetic force on moving charged particles with the same speed but different directions of motion.

DEVELOP Equation 26.1, $\vec{F}_B = q\vec{v} \times \vec{B}$, gives the magnetic force, so we'll apply it to each of the particles.

EVALUATE Proton 2 is moving parallel to the field, so $\vec{v}_2 \times \vec{B} = 0$ and it experiences no magnetic force. Protons 1 and 3 are moving at right angles to the field, so $\sin\theta = 1$, and the magnitude of the force on each is $F_B = qvB\sin\theta = qvB$. Using the proton charge $q = e = 1.6\times10^{-19}$ C and the given values for B and v yields $F = 32$ fN. Since protons are positive, the direction of $\vec{F}_B$ is the same as that of $\vec{v} \times \vec{B}$; applying the right-hand rule shows that the direction is out of the page for proton 1 and into the page for proton 3.

ASSESS Our answer 32 fN $(32\times10^{-15}\text{ N})$ is a tiny force, but that's not surprising given the proton's tiny charge. Note that the magnetic field alone doesn't determine the force; in this example identical particles experience different forces because they're moving in different directions relative to the field. ■

Although electricity and magnetism are related, the electric and magnetic forces are distinct. Both may be present simultaneously, in which case a charged particle experiences both an electric force $\vec{F}_E = q\vec{E}$ and a magnetic force $\vec{F}_B$ given by Equation 26.1. The result is an **electromagnetic force**:

$$\vec{F} = q\vec{E} + q\vec{v} \times \vec{B} \quad \text{(electromagnetic force)} \qquad (26.2)$$

Because the magnetic force depends on velocity but the electric force doesn't, it's possible to use perpendicular electric and magnetic fields to select particles of a particular velocity (Fig. 26.5). Such *velocity selectors* serve to prepare particle beams with uniform velocity as well as to analyze charged-particle populations in interplanetary space.

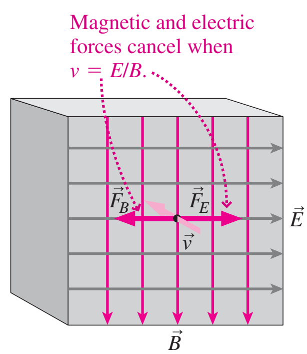


FIGURE 26.5 A velocity selector. The electric and magnetic forces cancel when $qE = qvB$, so only particles with speed $v = E/B$ pass through undeflected. The velocity $\vec{v}$ points into the page.

26.3 Charged Particles in Magnetic Fields

Following Newton's law, the magnetic force deflects charged particles from their otherwise straight-line paths. Magnetic forces "steer" charged particles in a host of practical devices ranging from microwave ovens to giant particle accelerators, and they shape particle trajectories throughout the astrophysical universe.

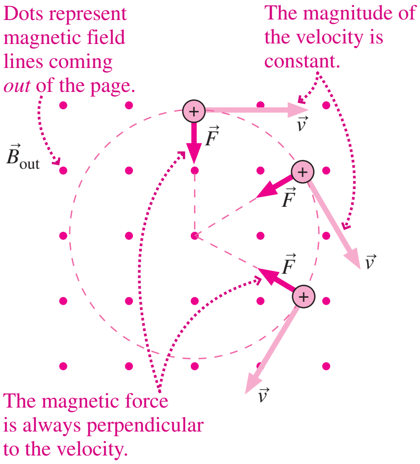


FIGURE 26.6 A charged particle moving at right angles to a uniform magnetic field describes circular motion.

The magnetic force always acts at right angles to a particle's velocity. **Therefore, it changes the direction of motion but not the speed, and it does no work.** In the special case of a particle moving at right angles to a uniform field, the magnetic force has a constant magnitude and the result, as Fig. 26.6 shows, is uniform circular motion. With $\vec{v}$ perpendicular to $\vec{B}$, the magnetic force of Equation 26.1 has magnitude qvB. This force provides the acceleration v^2/r that characterizes circular motion with radius r. Then Newton's law, $F = ma$, reads $qvB = mv^2/r$. We can solve for the radius of the particle's circular path to get

$$r = \frac{mv}{qB} \tag{26.3}$$

This result makes sense: The greater the particle's momentum mv, the harder it is for the magnetic force to bend its path and the larger the radius. On the other hand, a larger charge or field increases the force and makes a tighter orbit.

GOT IT? 26.2 A uniform magnetic field points out of this page. Will an electron that's moving in the plane of the page circle (a) clockwise or (b) counterclockwise as viewed from above the page?

EXAMPLE 26.2 Magnetic Deflection: A Mass Spectrometer

A mass spectrometer separates ions according to their ratio of charge to mass. Such devices are widely used in science and engineering to analyze unknown mixtures and to separate isotopes of chemical elements. Figure 26.7 shows ions of charge q and mass m first being accelerated from rest through a potential difference V and then entering a region of uniform magnetic field B pointing out of the page. Only the magnetic force acts on the ions in this region, so they undergo circular motion and, after half an orbit, land on a detector. Find an expression for the horizontal distance x from the entrance slit to the point where an ion lands on the detector.

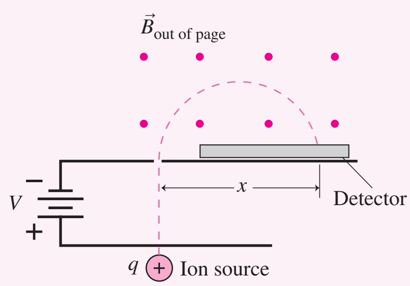


FIGURE 26.7 A mass spectrometer.

INTERPRET This problem is about charged particles undergoing circular motion in a uniform magnetic field. The distance we're asked for is the diameter of the particles' circular path.

DEVELOP Equation 26.3, $r = mv/qB$, shows that the path radius depends on the field and on the particle's mass, charge, and speed. We know everything but the speed, so this becomes a two-step problem in which we'll first find the speed. We're given the potential difference—energy per unit charge—so we can use energy conservation to find the kinetic energy and hence the ions' speed in the magnetic-field region. Then we'll use Equation 26.3 to find the radius of the ions' circular path.

EVALUATE A charge q gains kinetic energy qV in "falling" through a potential difference V, so an ion's kinetic energy once it enters the magnetic field is $\frac{1}{2}mv^2 = qV$. Solving for v gives $v = \sqrt{2qV/m}$. Our answer, the path diameter x, is then twice the radius given in Equation 26.3:

$$x = 2r = \frac{2mv}{qB} = \frac{2m\sqrt{2qV/m}}{qB} = \frac{2}{B}\sqrt{\frac{2mV}{q}}$$

ASSESS Make sense? The greater the mass or speed—which increases with the accelerating voltage V—the harder it is to deflect the ion and the larger the diameter of its semicircular path. The larger the field or charge, the larger the force and the smaller the semicircle. Note that for a fixed voltage and magnetic field, this device sorts ions by their charge-to-mass ratio q/m. ■

The Cyclotron Frequency

What's the period of a particle's circular orbit in a uniform magnetic field? The orbit circumference is $2\pi r$, so the period is $T = 2\pi r/v$. Using Equation 26.3 for the radius gives

$$T = \frac{2\pi r}{v} = \frac{2\pi}{v}\frac{mv}{qB} = \frac{2\pi m}{qB}$$

Remarkably, the period is independent of the particle's speed and orbital radius. Equation 26.3 shows why: The higher the speed v, the larger the radius r and hence the circumference. So a faster particle describes a larger circle and ends up taking the same amount of time to go around.

Equivalently, we can describe the particle's circular motion in terms of its frequency f, in revolutions per second, which is just the inverse of the period:

$$f = \frac{qB}{2\pi m} \quad \text{(cyclotron frequency)} \tag{26.4}$$

This quantity is the **cyclotron frequency**. Because it depends only on the field and the charge-to-mass ratio, cyclotron motion provides astrophysicists with a direct measure of magnetic fields in distant objects. Conversely, a fixed magnetic field guarantees a specific cyclotron frequency regardless of the particles' speeds. Microwave ovens exploit this fact, with their microwaves generated by electrons circling 2.4 billion times per second in a special tube called a magnetron.

Particle Trajectories in Three Dimensions

When a charged particle moves in an arbitrary direction, we consider velocity components both perpendicular and parallel to the magnetic field. Our previous analysis applies to the perpendicular component, giving circular motion in a plane perpendicular to the field. And because there's no magnetic force with velocity parallel to the field, the parallel component is unaffected. The result, in a uniform field, is a spiral path the field direction (Fig. 26.8).

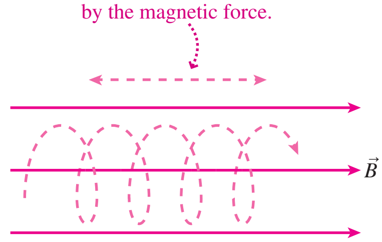
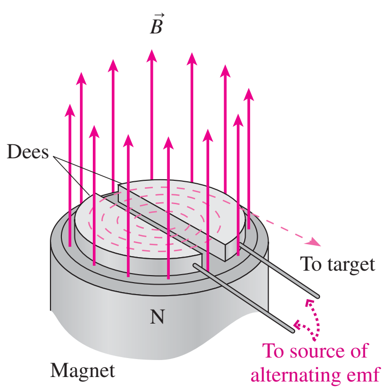


FIGURE 26.8 A particle in a uniform magnetic field describes a spiral path.

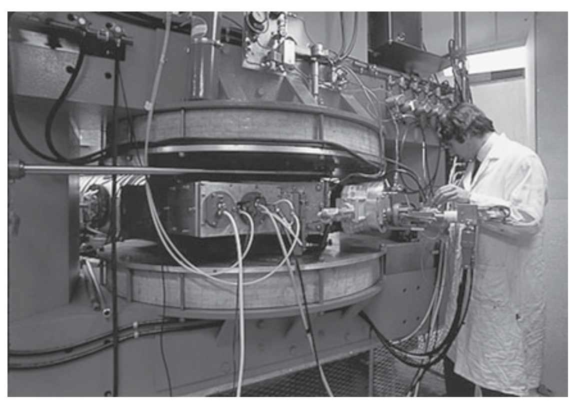
APPLICATION **The Cyclotron**

Physicists use high-energy particles to probe the structure of matter; engineers and physicians need high-energy particle beams in manufacturing, diagnostic, and therapeutic procedures. The easiest way to produce such beams is to accelerate ions through a potential difference, but the difficulties of handling high voltages make that impractical for all but the lowest energies. One of the earliest and most successful devices to circumvent this problem is the **cyclotron**, whose essential parts are shown in the figure. The device consists of an evacuated chamber between the poles of a magnet. Ions are produced at the center and undergo circular motion in the magnetic field.



Also in the chamber are two hollow conducting structures shaped like the letter D. A modest potential difference is applied across these "dees," and it alternates polarity at the cyclotron frequency. As ions circle around inside the cyclotron, they gain energy from the strong electric field associated with the potential difference at the gap. Inside the hollow conducting dee there's no electric field, so here the particles follow circular paths in the magnetic field. Halfway around they again encounter the dee gap. Because the potential is changing polarity in step with the particles' cyclotron motion, they again gain energy as they cross the gap. They move faster and in ever-larger circles, but always with the same orbital period. When they approach the edge of the machine, an electric field deflects the ions and they emerge as a high-energy beam.

Cyclotrons produce ions with energies of millions of electronvolts. This is high enough to cause nuclear reactions, and many medically useful radioactive isotopes are made using cyclotrons. In particular, the diagnostic procedure called PET (positron emission tomography) relies on cyclotron-produced radioisotopes; the photo shows a hospital-based cyclotron used for this purpose. At higher energies the theory of relativity alters our conclusion that the cyclotron frequency is independent of energy, and the cyclotron becomes useless. An alternative design is the **synchrotron**, in which both the magnetic field and the frequency vary to account for increasing particle energy.



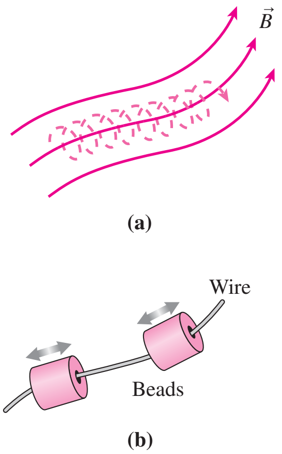


FIGURE 26.9 (a) Charged particles undergoing spiral motion about the magnetic field are "frozen" to the field like (b) beads sliding along a wire.

The absence of magnetic force in the field direction means it's easy to move charged particles along the field. But try to push a charged particle at right angles to the field, and it goes into circular motion; push harder and the circle only gets bigger. As a result, charged particles are effectively "frozen" to the field lines and move along the field like beads strung on a wire (Fig. 26.9). Nonuniform fields and particle collisions make this "freezing" less than perfect, but in many cases particle density is low enough that the "frozen" assumption is an excellent approximation. The coronal loops in this chapter's opening photo are a beautiful example of charged particles "frozen" to the solar magnetic field. Similarly, high-energy particles from the Sun get trapped on Earth's magnetic field lines; where the field intersects the atmosphere, particles collide with atmospheric nitrogen and oxygen to produce the spectacular displays we call the aurora. Here on Earth, trapping of charged particles on magnetic field lines enables researchers to confine plasmas at temperatures of 100 MK in attempts to harness the energy of nuclear fusion.

26.4 The Magnetic Force on a Current

An electric current consists of charges in motion, so a current in a magnetic field should experience a magnetic force. Figure 26.10 shows a straight wire in a magnetic field $\vec{B}$. Charges in the wire are moving about with thermal motions, but because these are random, the magnetic force on all the charges averages to zero. But if there's a current I in the wire, then the charges share a common drift velocity $\vec{v}_d$, and thus each experiences a magnetic force given by Equation 26.1: $\vec{F}_q = q\vec{v}_d \times \vec{B}$. If the wire has cross-sectional area A and contains n charges per unit volume, then the force on all the charge carriers in a length l of wire is $\vec{F} = nAlq\vec{v}_d \times \vec{B}$. But $nAqv_d$ is the current, I, as we found in Chapter 24. If we define a vector $\vec{l}$ whose magnitude is the wire length l and whose direction is along the current, then we can write

$$\vec{F} = I\vec{l} \times \vec{B} \quad \text{(magnetic force on a current)} \qquad (26.5)$$

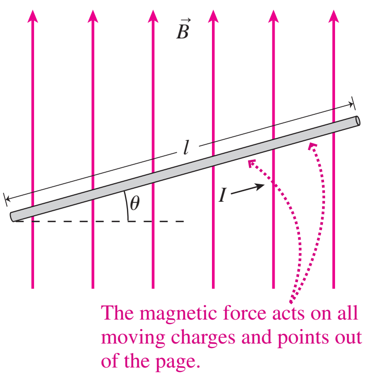


FIGURE 26.10 A straight wire carrying current I through a uniform magnetic field.

This force is perpendicular to both the current and the magnetic field, or out of the page in Fig. 26.10. The direction of the magnetic force doesn't depend on the sign of the charge carriers; they could be negative electrons moving leftward, opposite the current direction in Fig. 26.10, or positive charges moving rightward. For a given current, changing the sign of the charge carriers reverses both the sign of q and the direction of $\vec{v}_d$, leaving the force unchanged.

Equation 26.5 gives the net force on the charge carriers in the wire. In a physical wire, the magnetic force deflects charge carriers to one side of the wire, producing a charge separation and an electric field that results in a force on the rest of the wire (Fig. 26.11). Although its origin is not entirely magnetic, we loosely call the force in Equation 26.5 "the magnetic force on a wire." The magnetic force on a current-carrying wire is the basis for many practical devices, including loudspeakers and the electric motors that start cars and run refrigerators, disk drives, subway trains, pumps, food processors, power tools, and myriad other instruments of modern society.

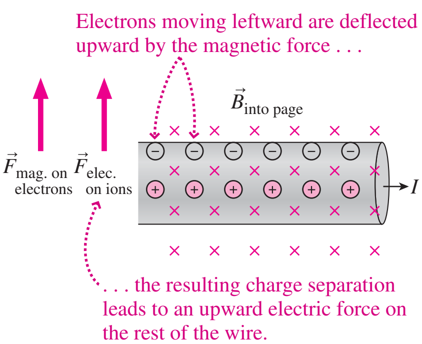
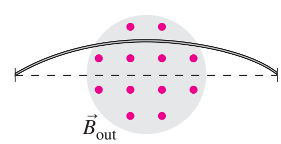


FIGURE 26.11 Origin of the magnetic force on a current-carrying wire.

Equation 26.5 holds for straight wires in uniform magnetic fields. In other cases we apply Equation 26.5 to very short segments of a wire that's either curved or in a nonuniform field, and we integrate to find the net force. Problem 57 explores this situation.

GOT IT? 26.3 The figure shows a flexible wire passing through a magnetic field that points out of the page. The wire is deflected upward, as shown. Is the current flowing (a) to the left or (b) to the right?



CONCEPTUAL EXAMPLE 26.1 **Magnetic Force: A Power Line**

A power line runs along Earth's equator, where the magnetic field points horizontally from south to north; the line carries current flowing from west to east. What's the direction of the magnetic force on the power line?

EVALUATE We've sketched the situation in Fig. 26.12. Using the right-hand rule with the current eastward and magnetic field northward shows that the force is vertically upward.

ASSESS As always, the force is at right angles to both the current and the magnetic field. As you'll see below, this force is pretty feeble compared with the power line's weight.

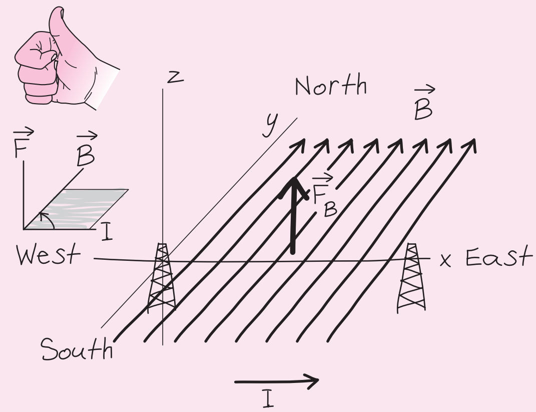


FIGURE 26.12 Sketch for Conceptual Example 26.1.

MAKING THE CONNECTION Earth's equatorial field strength is 30 μT, and the power line carries 500 A. What's the magnetic force on a kilometer of the line?

EVALUATE Equation 26.5 gives

$$F = |I\vec{l} \times \vec{B}| = IlB\sin 90° = (500\text{ A})(1.0\text{ km})(30\ \mu\text{T})(1) = 15\text{ N}$$

That's far less than the line's weight, on the order of 10 kN.

The Hall Effect

We noted earlier that the direction of the magnetic force depends on the direction of the current, not on the sign of the charge carriers. However, there's a subtle difference between two conductors with the same current yet different charge carriers. In Fig. 26.13, moving charge carriers of either sign are deflected to the upper surface of the conductor. Again, that's because the signs of both charge and velocity are opposite in the two cases. In both cases the result is a small electric field and its associated potential difference *across* the wire. The *direction* of the electric field and the *sign* of the potential difference depend on the sign of the charge carriers.

The separation of charges across a current-carrying wire is the **Hall effect**, and the potential difference is the **Hall potential**. In a steady state, the magnetic force on the charge carriers is just balanced by the electric force associated with charge separation, giving $qE = qv_dB$, or simply $E = v_dB$. In the rectangular conductor of Fig. 26.13, the electric field is uniform and the Hall potential is then $V_H = Eh = v_dBh$. Using $I = nqAv_d$ and solving for v_d, we can then write $V_H = IBh/nAq$. Since $A = ht$, with t the conductor thickness in the field direction (see Fig. 26.13), this becomes

$$V_H = \frac{IB}{nqt} \quad \text{(Hall potential)} \qquad (26.6)$$

The quantity $1/nq$ is the **Hall coefficient**. Measuring the Hall coefficient gives information on the nature and density of the charge carriers. Alternatively, measuring V_H in a material of known coefficient carrying a known current gives a direct measure of the magnetic field strength.

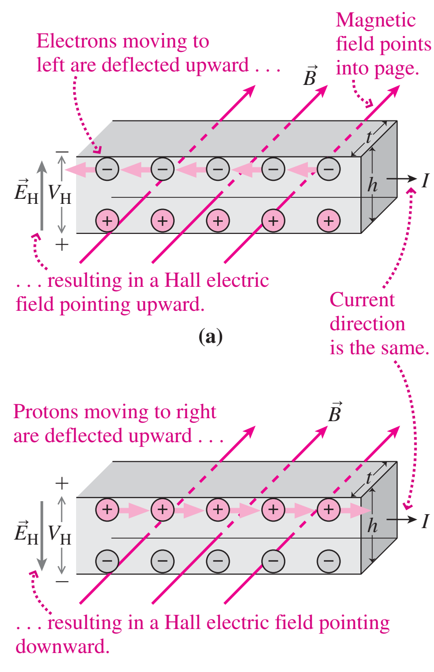


FIGURE 26.13 The Hall electric field $\vec{E}_H$ and Hall potential V_H arise from the magnetic deflection of charge carriers. In both (a) and (b) the current is to the right, carried in (a) by negative charge moving to the left and in (b) by positive charge moving to the right.

26.5 Origin of the Magnetic Field

Electric charges respond to electric fields, and electric charges produce electric fields. So it is with magnetism. We've just explored how *moving* electric charges respond to magnetic fields; we'll now see how *moving* electric charges produce magnetic fields. The first inkling of a relation between electricity and magnetism came in 1820 when the Danish scientist Hans Christian Oersted discovered that a compass needle is deflected by an electric current. A month after Oersted's discovery became known in Paris, the French scientists Jean Baptiste Biot (rhymes with "Leo") and Félix Savart (rhymes with "bazaar") had experimentally determined the form of the force arising from a steady current.

The Biot–Savart Law

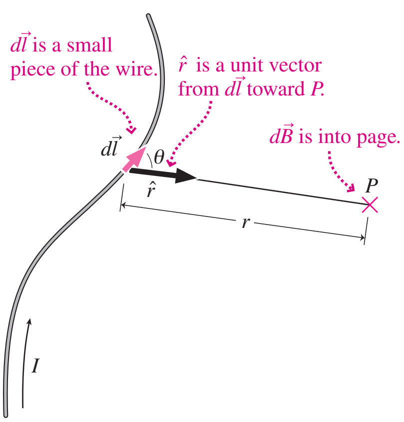


FIGURE 26.14 The Biot–Savart law gives the magnetic field $d\vec{B}$ at the point P arising from the current I flowing along the infinitesimal vector $d\vec{l}$.

The Biot–Savart law gives the contribution $d\vec{B}$ to the magnetic field at a point P due to a small element of current, in much the way that Coulomb's law gives the electric field $d\vec{E}$ due to a charge element dq. Figure 26.14 shows a wire carrying a steady current I and the contribution $d\vec{B}$ to the field at P from a small length dl of the wire. The current element $I\,dl$ is the source of the field; it plays the same role as the charge dq in Coulomb's law. The magnetic field decreases with the inverse square of the distance, just as in Coulomb's law for the electric field.

There are important differences between the Coulomb and Biot–Savart laws. Charge—the source of electric field—is a scalar quantity. But *moving* charge—the source of magnetic field—has direction. The Biot–Savart law accounts for that direction by defining a vector $d\vec{l}$ along the current; then the field contribution $d\vec{B}$ from the source element $I\,d\vec{l}$ depends on the sine of the angle between $d\vec{l}$ and the unit vector $\hat{r}$ from the source toward the point where we're evaluating the field. Mathematically, all this is summarized in a compact vector equation:

$$d\vec{B} = \frac{\mu_0}{4\pi}\frac{I d\vec{l} \times \hat{r}}{r^2} \quad \text{(Biot–Savart law)} \tag{26.7}$$

where μ_0 is the **permeability constant**, whose exact value is $4\pi \times 10^{-7}$ N/A^2 (equivalent and often-used units are T·m/A).

Besides the more complicated directionality evidenced by the cross product in the Biot–Savart law, there's another distinction between the Coulomb and Biot–Savart laws. Both describe fields of localized structures—namely, point charges and current elements. It makes sense to talk about an isolated point charge. But an isolated current element is impossible in the steady state; any steady current must flow in a complete circuit. So a Biot–Savart calculation necessarily involves the fields produced by current elements around an entire circuit. The magnetic field obeys the superposition principle, so the net field at any point is the vector sum, or integral, of the field contributions of all the individual current elements:

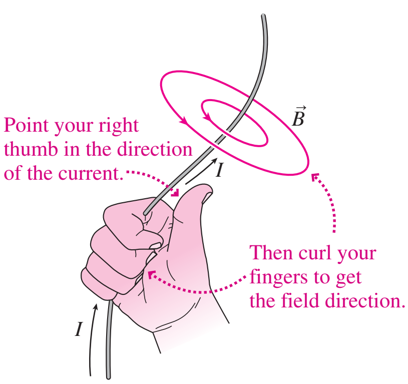


FIGURE 26.15 Magnetic field lines generally encircle a current, with direction given by the right-hand rule.

$$\vec{B} = \int d\vec{B} = \frac{\mu_0}{4\pi}\int\frac{I d\vec{l} \times \hat{r}}{r^2} \quad \text{(Biot–Savart law, integrated)} \tag{26.8}$$

The field given in Equation 26.8 depends on the details of the current distribution, but the directionality associated with the cross product means that, quite generally, magnetic field lines encircle the current that is their source (Fig. 26.15). The next two examples use the Biot–Savart law; later we'll find a simpler way to calculate magnetic fields for some current distributions.

EXAMPLE 26.3 Using the Biot–Savart Law: The Field of a Current Loop

Find the magnetic field at an arbitrary point P on the axis of a circular loop of radius a carrying current I.

INTERPRET This is a problem involving the magnetic field produced by a specified current distribution.

DEVELOP Figure 26.16a shows the current loop with the point P a distance x along the axis. The Biot–Savart law determines the field at

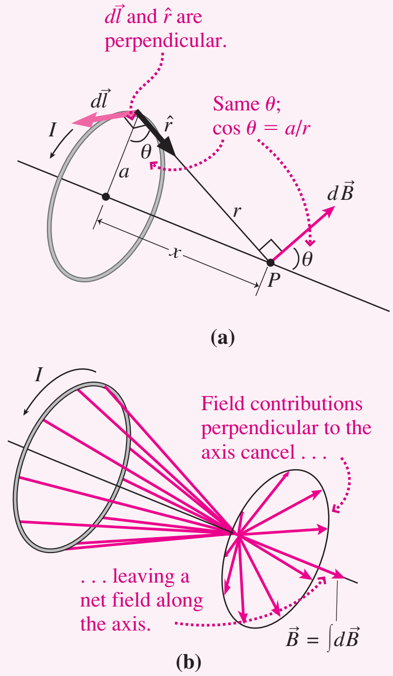


FIGURE 26.16 Finding the magnetic field on the axis of a current loop.

P, and we've identified the vectors $d\vec{l}$ and $\hat{r}$ that appear in the law. As Fig. 26.16b shows, the individual field components perpendicular to the axis cancel, giving a net field that's along the axis. So our plan is to find an expression for the x-components of the field contributions $d\vec{B}$, and then integrate to get the net field.

EVALUATE Figure 26.16a shows that the x-component of any $d\vec{B}$ is $dB_x = dB\cos\theta$, where $\cos\theta = a/r = a/\sqrt{x^2 + a^2}$. The figure also shows that $d\vec{l}$ and $\hat{r}$ are perpendicular; since $\hat{r}$ is a unit vector, the product $d\vec{l} \times \hat{r}$ has magnitude dl. Then the term $d\vec{l} \times \hat{r}/r^2$ in the Biot–Savart law has magnitude $dl/(x^2 + a^2)$, and we have

$$B = \int dB_x = \frac{\mu_0 I}{4\pi}\int_{\text{loop}} \frac{dl}{x^2 + a^2}\frac{a}{\sqrt{x^2 + a^2}}$$

$$= \frac{\mu_0 I a}{4\pi(x^2 + a^2)^{3/2}}\int_{\text{loop}} dl$$

where the integral reduces to a simple form because the distance x is the same for all points on the loop. The remaining integral is the sum of infinitesimal lengths around the loop, or the loop circumference $2\pi a$. So we have

$$B = \frac{\mu_0 I a^2}{2(x^2 + a^2)^{3/2}} \qquad (26.9)$$

The direction of the field, as suggested in Fig. 26.16b, is along the axis.

ASSESS The field is strongest right at the loop center $(x = 0)$ because here we're closest to the loop and so the contributions from all segments of the loop are greatest. The field decreases as we move away from the loop. In general the field is a complicated function of distance, but for large distances $(x \gg a)$ it falls off as $1/x^3$. That should remind you of the field we found for an electric dipole in Chapter 20. We'll have more to say about this dipole-like behavior in Section 26.6. ■

EXAMPLE 26.4 Using the Biot–Savart Law: The Field of a Straight Wire

Find the magnetic field produced by an infinitely long straight wire carrying steady current I.

INTERPRET This example, too, is about the field produced by a specified current distribution.

DEVELOP Figure 26.17 is our drawing of the wire on a coordinate system with the x-axis along the wire. Since the wire is infinite, the field magnitude must be the same at all points equidistant from the wire. We show one such point P, a distance y from the wire. We also show

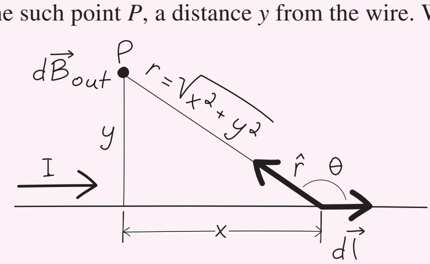


FIGURE 26.17 Calculating the magnetic field at P due to an infinite straight wire carrying current I along the x-axis.

an infinitesimal segment $d\vec{l}$ of the wire and the unit vector $\hat{r}$ toward the field point. Our plan is to calculate the field contributions $d\vec{B}$ from all such current elements, and then integrate to find the field $\vec{B}$.

EVALUATE Both $d\vec{l}$ and $\hat{r}$ lie in the plane of the page, so at P the vector $d\vec{l} \times \hat{r}$ in the Biot–Savart law is out of the page. This is true for any segment of the wire. Therefore, we can sum the magnitudes of the contributions $d\vec{B}$ to find the magnitude of the net field, and we know its direction at P will be out of the page. With $\hat{r}$ a unit vector, $|d\vec{l} \times \hat{r}| = dl\sin\theta$, where Fig. 26.17 shows that $\sin\theta = y/r = y/\sqrt{x^2 + y^2}$. Then the Biot–Savart law gives a field contribution of magnitude

$$dB = \frac{\mu_0 I}{4\pi}\frac{|d\vec{l} \times \hat{r}|}{r^2} = \frac{\mu_0 I}{4\pi}\frac{dl\sin\theta}{r^2} = \frac{\mu_0 I}{4\pi}\frac{y\,dl}{(x^2 + y^2)^{3/2}}$$

Since the segment $d\vec{l}$ lies along the x-axis, $dl = dx$. Also, y is a constant here, so the net field becomes

$$B = \int dB = \frac{\mu_0 I y}{4\pi}\int_{-\infty}^{\infty}\frac{dx}{(x^2 + y^2)^{3/2}}$$

(continued)

where we chose the limits to include the entire infinite wire. The integral is a standard one, given in the integral tables of Appendix A; the result is

$$B = \frac{\mu_0 I}{2\pi y} \tag{26.10}$$

ASSESS This result for the *magnetic* field of a long current-carrying wire should remind you of our earlier finding for the *electric* field of a line charge; both decrease as the inverse of the distance from the line. But where the electric field of a line charge points radially outward, the magnetic field of a line current encircles the current, as shown in Fig. 26.18. Of course, an infinite line current is impossible, but our result is a good approximation to the fields of finite wires if we're close compared with the wire's length. ■

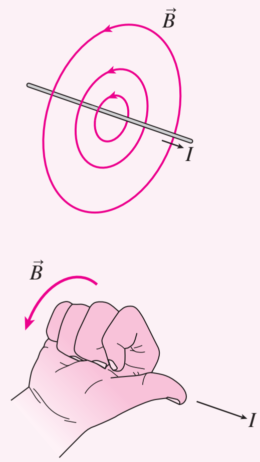


FIGURE 26.18 Magnetic field lines encircle a straight wire, with their direction given by the right-hand rule.

The Magnetic Force Between Conductors

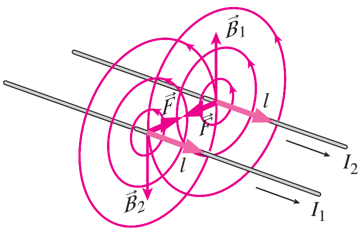


FIGURE 26.19 The magnetic force between parallel currents in the same direction is attractive.

In Section 26.4 we found the force on a current-carrying wire in a magnetic field: $\vec{F} = I\vec{l} \times \vec{B}$. Now we find that a straight wire produces a magnetic field. That means current-carrying wires exert magnetic forces on each other. Figure 26.19 shows the situation for two parallel wires carrying currents in the same direction. The wires are a distance d apart, so the field of wire 1 at the location of wire 2 follows from Equation 26.10: $B_1 = \mu_0 I_1/2\pi d$. The field is perpendicular to wire 2, so the force on a length l of wire 2 is

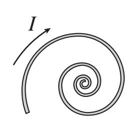
$$F_2 = I_2 l \times B_1 = \frac{\mu_0 I_1 I_2 l}{2\pi d} \quad \text{(magnetic force between two wires)} \tag{26.11}$$

Figure 26.19 shows that the direction of this force is toward wire 1, so the parallel currents *attract*. Analyzing the force on wire 1 from wire 2 amounts to interchanging the subscripts 1 and 2, giving an attractive force of the same magnitude. Reversing one of the currents would change the signs of both forces, showing that antiparallel currents *repel*.

The force between nearby conductors can be quite large, so engineers who design high-strength electromagnets must provide enough physical support to withstand the magnetic force. The hum you often hear around electrical equipment comes from the mechanical vibration of nearby conductors in transformers and other devices, as they respond to the changing force associated with 60-Hz alternating current.

The magnetic force between conductors is the basis for the SI definition of the ampere and, consequently, the coulomb. One ampere is defined as the current in two long, parallel conductors 1 m apart when they experience a magnetic force of 2×10^{-7} N per meter of length. Then 1 C is the amount of charge passing in 1 s through a wire carrying 1 A.

GOT IT? 26.4 A flexible wire is wound into a flat spiral as shown in the figure. If a current flows in the direction shown, will the coil (a) tighten or (b) become looser? Does your answer depend on the current direction?



26.6 Magnetic Dipoles

The current loop of Example 26.4 shows the essential characteristic of all steady-state currents—namely, a closed loop with current everywhere the same. Equation 26.9 gives the field on the loop axis: $B = \mu_0 I a^2/2(x^2 + a^2)^{3/2}$. For $x \gg a$ we can ignore a^2 compared with x^2 in the denominator, giving $B \simeq \mu_0 I a^2/2x^3$. Multiply both sides by 2π to get

$B \simeq 2\mu_0 IA/4\pi x^3$, where A is the loop area. Compare this result with the field on the axis of an *electric* dipole, Equation 20.6b: Both show the inverse-cube dependence of the dipole field, and both involve fundamental constants from the Coulomb and Biot–Savart laws that relate fields and their sources. Where the electric-field expression contains the electric dipole moment p, the product of charge and separation, the magnetic-field expression contains IA, the product of the loop current and loop area. We identify IA as the magnitude, μ, of the current loop's **magnetic dipole moment**. Then the on-axis magnetic dipole field becomes

$$B = \frac{\mu_0}{2\pi}\frac{\mu}{x^3} \quad \text{(on-axis field, magnetic dipole)} \tag{26.12}$$

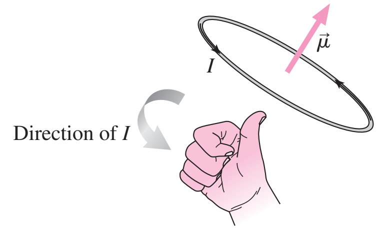


FIGURE 26.20 Finding the direction of a current loop's magnetic dipole moment.

The magnetic dipole moment is a vector whose direction follows from the right-hand rule shown in Fig. 26.20. If we describe the loop by a vector of magnitude A whose direction is perpendicular to the loop as shown in Fig. 26.20, then we can write the magnetic dipole moment as $\vec{\mu} = I\vec{A}$. Practical current loops often have multiple turns; since each carries the same current, an N-turn loop has effective current NI, so its dipole moment becomes

$$\vec{\mu} = NI\vec{A} \quad \left(\begin{array}{c}\text{magnetic dipole moment,}\\ N\text{-turn current loop}\end{array}\right) \tag{26.13}$$

Although we've found the magnetic field for a current loop only on the loop axis, a more elaborate calculation shows that the magnetic field anywhere far from the loop has exactly the same configuration as the electric field far from an electric dipole. And although we developed the magnetic dipole moment for a circular loop, Equation 26.13 in fact gives the dipole moment of *any* closed loop of current. We conclude that **any current loop constitutes a magnetic dipole**, and that far from the loop, its field will be that of a dipole. Electric and magnetic dipoles are analogous: Both have the same field configuration and mathematical form far from their sources (Fig. 26.21), and both are characterized by their respective dipole moments. But they aren't the same. One is an electric field, its origin in static electric charge; the other is a magnetic field, its origin in *moving* charge—specifically, charge moving in a closed loop. And the similarity in field configurations holds only at large distances; as Fig. 26.21 shows, the fields near electric and magnetic dipoles are very different, reflecting the different structures that give rise to each.

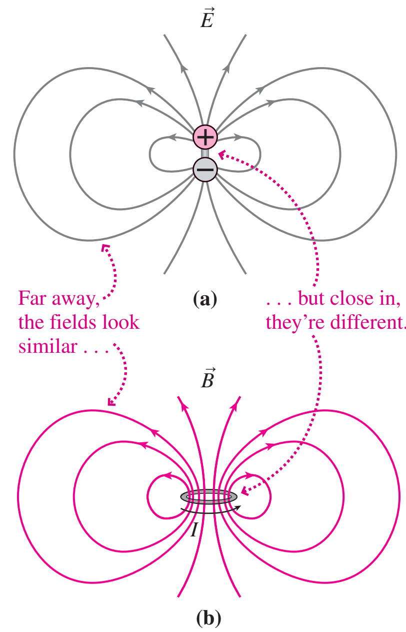


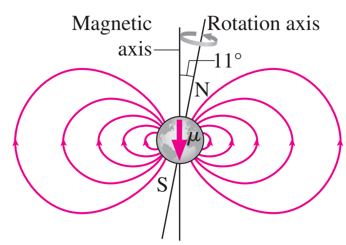
FIGURE 26.21 (a) The electric field of an electric dipole and (b) the magnetic field of a current loop. Far from their sources, both have the shape and the $1/r^3$ dependence of the dipole field.

Current loops are ubiquitous, and so are dipole magnetic fields. Multiple turns of current-carrying wire produce the strong magnetic fields of electromagnets, and superconducting loops provide the fields in MRI scanners. At the atomic level, orbiting and spinning electrons constitute miniature magnetic dipoles. Even planets and stars have magnetic dipole fields.

APPLICATION Magnetic Fields of Earth and Sun

Many astrophysical objects have magnetic fields resulting from the interaction of conducting fluids with the objects' rotation. Earth's field arises in its liquid-iron outer core, where convective flows work with Earth's rotation to produce electric currents. The figure shows that Earth's field approximates that of a dipole; the magnitude of the dipole moment is approximately $\mu = 8.0\times10^{22}\ \text{A}\cdot\text{m}^2$. The direction of the dipole moment vector differs from that of Earth's rotation axis, which accounts for the difference between magnetic and true north. Earth's field reverses roughly every million years, and geologists track seafloor spreading from the resulting magnetization in rocks. Farther out, Earth's magnetic field traps high-energy particles and thus protects us from dangerous radiation.

The Sun's gaseous nature makes its magnetic field much more dynamic, and magnetism is the dominant force in its hot, electrically conducting atmosphere. The Sun's field reverses approximately every 11 years, coinciding with the rise and fall of sunspots—regions of intense magnetic field that are often sources of violent outbursts.



Dipoles and Monopoles

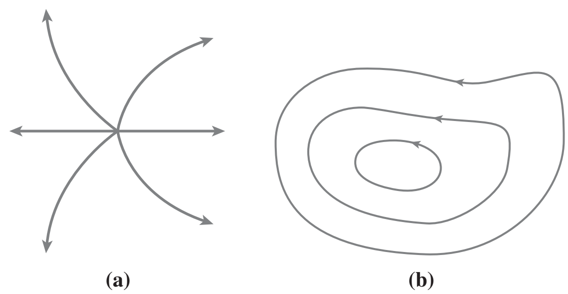
Atoms, molecules, and radio antennas are among the many structures that behave as *electric* dipoles. In all these, separation of positive and negative electric charge gives rise to the dipole. Magnetism is different. No one has ever found an isolated magnetic north or south pole analogous to an electric charge. Electromagnetic theory doesn't rule out such **magnetic monopoles**, and indeed some theories suggest that monopoles might have formed in the Big Bang. But they've never been found. All magnetic fields we've ever seen come from moving *electric* charge. As we'll see in Section 26.7, that includes the fields of permanent magnets. Because steady currents form closed loops, the simplest magnetic entity is the dipole.

Electric field lines generally begin or end on electric charges. But there aren't any "magnetic charges"—magnetic monopoles. Magnetic field lines don't begin or end, but form closed loops encircling the moving electric charges that are their sources. In Chapter 21 we developed Gauss's law to quantify the statement that the number of electric field lines emerging from any closed surface depends only on the charge enclosed. Because there's no magnetic charge, the net number of magnetic field lines and therefore the **magnetic flux** $\oint \vec{B}\cdot d\vec{A}$ emerging from any closed surface is always zero. Thus **Gauss's law for magnetism** is

$$\oint \vec{B}\cdot d\vec{A} = 0 \qquad \text{(Gauss's law for magnetism)} \tag{26.14}$$

Like Gauss's law for electricity, Equation 26.14 is one of the four fundamental laws that govern all electromagnetic phenomena in the universe. We'll meet the remaining two laws shortly. Although Gauss's law for magnetism has zero on its right side, it's not devoid of content; rather, it says that all magnetic fields are configured so that their field lines have no beginnings or endings.

GOT IT? 26.5 The figure shows two fields. Which could be a magnetic field?



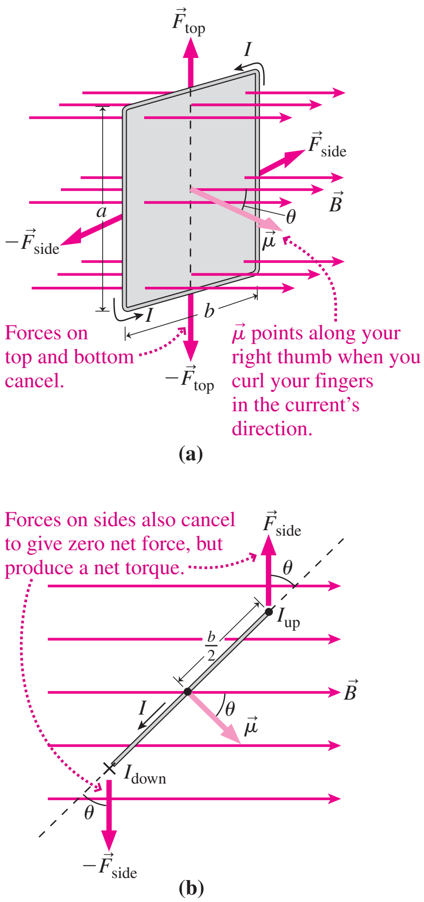


FIGURE 26.22 (a) A rectangular current loop in a uniform magnetic field. (b) Top view of the loop, showing that magnetic forces on the vertical sides result in a net torque.

The Torque on a Magnetic Dipole

In Section 20.5 we found that an electric dipole $\vec{p}$ in a uniform electric field $\vec{E}$ experiences a torque $\vec{\tau} = \vec{p} \times \vec{E}$; in a nonuniform field there's a net force as well. The same is true for a magnetic dipole in a magnetic field, as we can see by considering the rectangular current loop in a uniform field shown in Fig. 26.22*a*. Current flowing along the top and bottom of the loop results in upward and downward forces of equal magnitude, and neither a net force nor a net torque is associated with these forces. Currents flowing along the vertical sides also result in equal but opposite forces. However, as Fig. 26.22*b* shows, these forces result in a net torque about a vertical axis through the center of the loop. The vertical sides have length a and the currents are perpendicular to the horizontal magnetic field, so the force on each has magnitude $F_{side} = IaB$. The vertical sides are half the loop width b from the axis, so the torque due to each is $\tau_{side} = \frac{1}{2}bF_{side}\sin\theta = \frac{1}{2}bIaB\sin\theta$. Torques on the two sides are in the same direction (out of the page in Fig. 26.22*b*), so the net torque is $\tau = IabB\sin\theta = IAB\sin\theta$, with A the loop area. We've already identified IA as the magnitude of the loop's magnetic dipole moment $\vec{\mu}$ and, given the direction of $\vec{\mu}$ as shown in Figs. 26.20 and 26.22*b*, we can incorporate the directionality and the factor $\sin\theta$ into a cross product:

$$\vec{\tau} = \vec{\mu} \times \vec{B} \qquad \text{(torque on a magnetic dipole)} \tag{26.15}$$

analogous to the torque on an electric dipole.

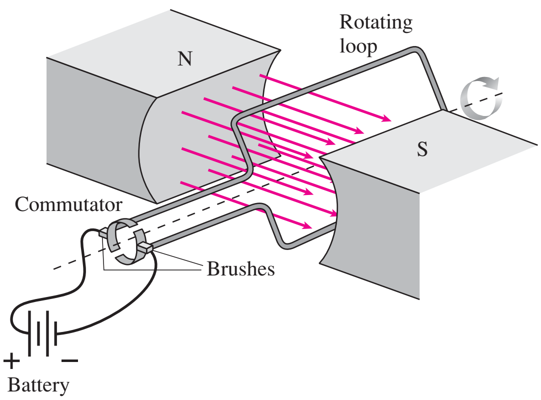
The magnetic torque of Equation 26.15 causes magnetic dipoles—current loops—to align with their dipole moment vectors along the magnetic field. It takes work to rotate a dipole out of alignment with the field, and in analogy with Equation 20.11 the associated potential energy is

$$U = -\vec{\mu} \cdot \vec{B} \quad (26.16)$$

In a nonuniform field, a dipole also experiences a net force. That's why the nonuniform field near the poles of a bar magnet attracts magnetic materials that, as we'll see in the next section, contain magnetic dipoles.

The torque on a magnetic dipole is important in many technologies, including electric motors and MRI imaging. Some satellites use the torque produced by Earth's magnetic field to orient themselves in space; with electricity generated from solar panels powering current loops, there's no fuel to run out.

APPLICATION Electric Motors



Electric motors are so much a part of our lives that we hardly think of them. Yet refrigerators, disk drives, subway trains, vacuum cleaners, power tools, food processors, fans, washing machines, water pumps, hybrid cars, and most industrial machinery would be impossible without electric motors.

At the heart of every electric motor is a current loop in a magnetic field. But instead of a steady current, the loop carries a current that reverses to keep the loop always spinning. In direct-current (DC) motors, this is achieved through the electrical contacts that provide current to the loop. The figure shows how current flows to the loop through a pair of stationary brushes that contact rotating conductors called the commutator. The current loop rotates to align with the field, but just as it does so, the brushes cross the gaps in the commutator and reverse the loop's current and therefore its dipole moment vector. Now the loop swings another 180° to its new "desired" position, but again the commutator reverses the current and so the loop rotates continuously. A rigid shaft spinning with the coils delivers mechanical energy. Thus the motor is a device that converts electrical energy to mechanical energy; the magnetic field is an intermediary in this energy conversion.

EXAMPLE 26.5 Torque on a Current Loop: Designing a Hybrid-Car Motor

Toyota's Prius gas–electric hybrid car uses a 60-kW electric motor that develops a maximum torque of 207 N·m. Suppose you want to produce this torque in a motor like the one in the preceding Application, consisting of a 700-turn rectangular coil measuring 30 cm by 20 cm in a uniform field of 50 mT. How much current does the motor need?

INTERPRET This problem is about an electric motor, which according to the Application is essentially a current loop in a magnetic field. We're given the torque and asked for the current.

DEVELOP Equation 26.15, $\vec{\tau} = \vec{\mu} \times \vec{B}$, determines the torque on a current loop. Figure 26.23 is a sketch of the loop at the point of maximum torque, $\tau_{max} = \mu B$, which occurs when $\sin\theta = 1$. To solve for the current, we need the magnetic dipole moment from Equation 26.13, $\mu = NIA$. Then $\tau_{max} = NIAB$.

EVALUATE Solving for I using the maximum torque and the loop dimensions gives

$$I = \frac{\tau_{max}}{NAB} = \frac{207\ \text{N}\cdot\text{m}}{(700)(0.30\ \text{m})(0.20\ \text{m})(0.050\ \text{T})} = 99\ \text{A}$$

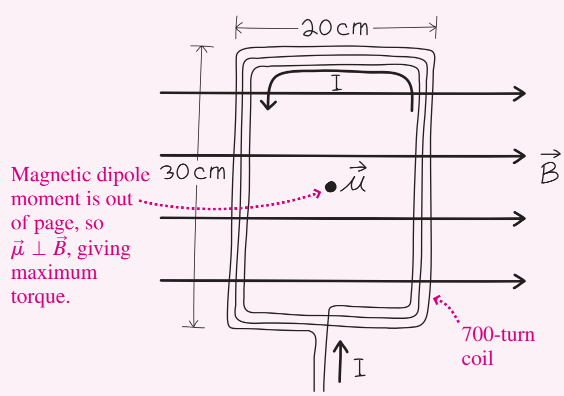


FIGURE 26.23 Loop for the motor of Example 26.5, shown in the position of maximum torque.

ASSESS That's a large current, but propelling a car is a big job. The actual Prius motor operates at 650 V, so its 60-kW power requires current $I = P/V = 92$ A, close to our answer. ■

26.7 Magnetic Matter

So far, we've said remarkably little about magnets. That's because magnetism is fundamentally about moving electric charge. Magnets and magnetic matter are just a minor manifestation of this universal phenomenon.

The magnetism of everyday magnets and of magnetic materials like iron results from atomic-scale current loops. An electron orbiting a nucleus constitutes a simple current loop and therefore has a magnetic dipole moment (Fig. 26.24). More importantly, an electron possesses an intrinsic magnetic dipole moment associated with a quantum-mechanical angular momentum called "spin." Interactions among these magnetic moments determine the magnetic properties of atoms and of bulk matter. The details necessarily involve quantum mechanics; here we give a qualitative overview of magnetism in matter, which manifests itself in three distinct forms.

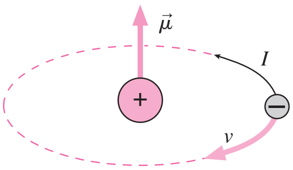


FIGURE 26.24 In the classical model of the atom, the circling electron constitutes a miniature current loop. The current is opposite the motion because the electron is negative. Drawing is only suggestive; the electron's intrinsic magnetic dipole moment is usually more important than that resulting from its orbital motion.

Ferromagnetism

The magnetism you're familiar with is **ferromagnetism**, which is limited to a few substances, including iron, nickel, cobalt, and some alloys and compounds. Strong interactions among atomic magnetic moments result in **magnetic domains**, regions that contain 10^{17}–10^{21} atoms whose magnetic moments are all aligned in the same direction. Normally the magnetic moments of different domains point in random directions, so there's no net magnetic moment. But when an external magnetic field is applied, the domains all align and the material acquires a net magnetic moment. If the field is nonuniform, the material then experiences a net force, which is why ferromagnetic materials are attracted to magnets.

So-called **hard** ferromagnetic materials retain their magnetism even after the applied field is removed; the result is a permanent magnet. A bar magnet, for example, has its internal magnetic moments aligned along its long dimension. You can think of its field as arising from currents circulating around the surface of the magnet (Fig. 26.25)—currents that ultimately result from the superposition of individual atomic current loops. Computer disks, credit card strips, and audio/video tapes use hard ferromagnetic materials that retain information as patterns of permanent magnetization. **Soft** ferromagnetic materials, in contrast, don't hold magnetization. They're used where magnetization must be turned on and

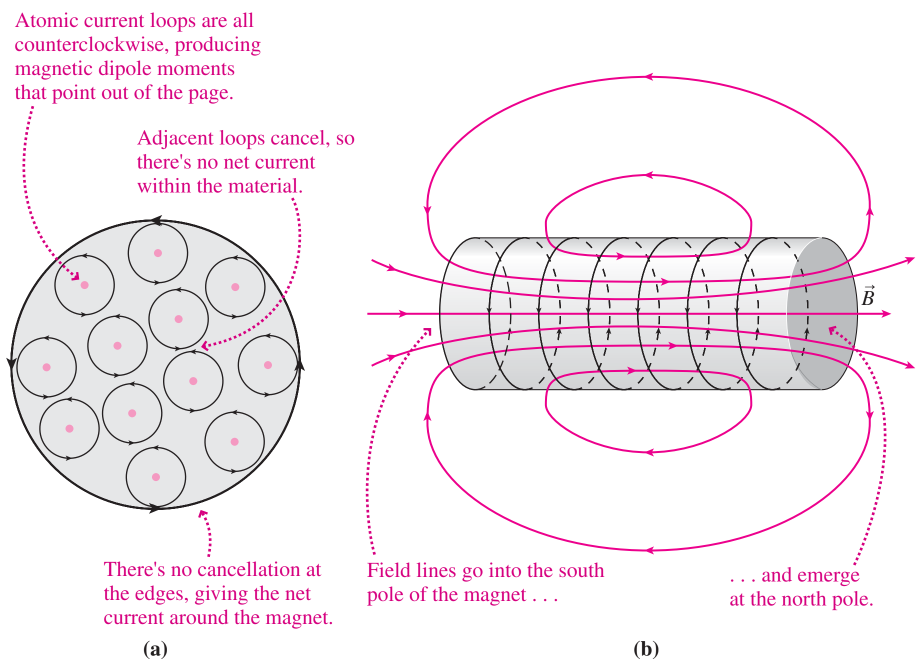


FIGURE 26.25 (a) Cross section of a bar magnet, showing atomic current loops all aligned the same way and making a net current around the magnet. (b) Side view showing the field that results from this magnetization current.

off rapidly, as in the "heads" that write information to computer disks and tapes. Ferromagnetism disappears at the so-called **Curie temperature**, as random thermal motions disrupt the organized alignment of magnetic dipoles; for iron this phase transition occurs at 1043 K.

Paramagnetism

Many substances that aren't ferromagnetic nevertheless consist of atoms or molecules that have permanent magnetic dipole moments. There's no strong interaction among the individual dipoles, so these **paramagnetic** materials respond only weakly to external magnetic fields. Paramagnetic effects are generally significant only at very low temperatures.

Diamagnetism

Materials without intrinsic magnetic moments can have moments induced by changes in an applied magnetic field. Whereas ferromagnetic and paramagnetic materials are attracted to magnets, these **diamagnetic** materials are repelled. We'll explore the origins of diamagnetism in Chapter 27.

Magnetic Permeability and Susceptibility

We found in Chapter 20 that the alignment of molecular electric dipoles reduces the electric field in a material. In paramagnetic and ferromagnetic materials, alignment of magnetic dipoles causes an *increase* in the field. Figure 26.26 shows that this difference occurs because the magnetic field within a current loop points in the *same* direction as the loop's magnetic dipole moment, whereas the internal field of an electric dipole is *opposite* the dipole moment. Ferromagnetic behavior is further complicated because it depends on the material's past history, which is what makes permanent magnets possible. Coils for electromagnets and computer disk "heads" are wound on ferromagnetic cores to provide a much stronger magnetic field than the coil current alone could produce.

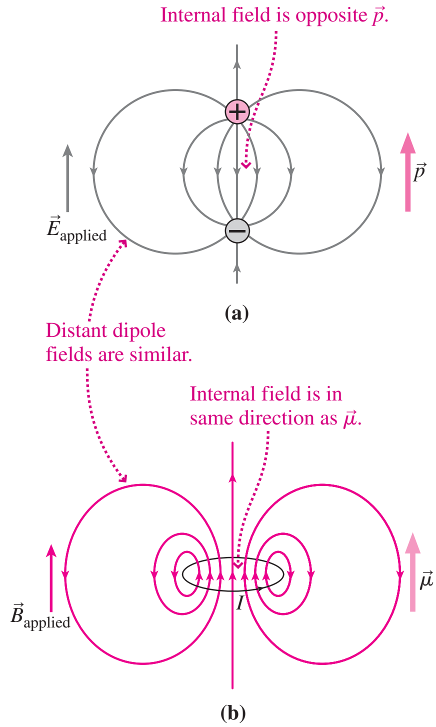


FIGURE 26.26 Internal fields of electric and magnetic dipoles have opposite directions. (a) Electric dipoles reduce an applied electric field; (b) magnetic dipoles increase an applied magnetic field.

26.8 Ampère's Law

Computing electric fields with Coulomb's law in Chapter 20 was cumbersome for all but the simplest charge distributions. In Chapter 21 we saw how Gauss's law greatly simplified electric-field calculations for symmetric charge distributions. Is there an analogous approach for magnetic fields? Gauss's law for magnetism, Equation 26.14, won't do because it doesn't relate a magnetic field to its source—namely, moving charge.

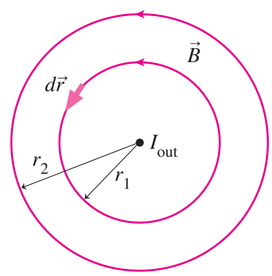


FIGURE 26.27 Two magnetic field lines surrounding a wire carrying current out of the page.

Figure 26.27 shows two of the circular magnetic field lines surrounding a long wire carrying a current I out of the page. Imagine moving around the inner circle, and as you go a little way, take the product of the displacement dr with the magnetic field in the direction you're going. Here you're moving in the direction of the field, so that product is $B\,dr$; more generally, it's the dot product $\vec{B}\cdot d\vec{r}$. Now add up all these products around the circle. Formally, the result is the line integral $\oint \vec{B}\cdot d\vec{r}$, where the circle indicates that we're integrating around a closed path. In this case the integral becomes just $\oint B\,dr$ because $\vec{B}$ and $d\vec{r}$ are in the same direction. But here the field magnitude is given by Equation 26.10: $B = \mu_0 I/2\pi r$, where we've replaced y with the radius r. Since r has the constant value r_1 on the inner circle in Fig. 26.27, the integral becomes $(\mu_0 I/2\pi r_1)\oint dr$. Now $\oint dr$ is the total length of the circular path, or its circumference $2\pi r_1$. So the value of $\oint \vec{B}\cdot d\vec{r}$ is $\mu_0 I$. If you try the same thing for the outer circle in Fig. 26.27, r_2 replaces r_1, but the result is the same: $\oint \vec{B}\cdot d\vec{r} = \mu_0 I$, independent of the radius.

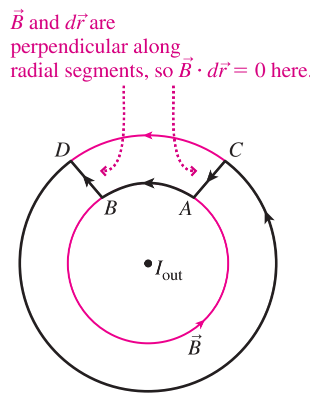


FIGURE 26.28 A closed loop that does not coincide with a field line. The line integral $\oint \vec{B}\cdot d\vec{r}$ around this loop has the same value $\mu_0 I$ that it has around a circular loop.

We get the same result even if the path doesn't coincide with a field line, as Fig. 26.28 suggests. On the radial segments of the path shown, $\vec{B}\cdot d\vec{r} = 0$ and there's no contribution to the integral. On segment AB, the field is stronger than if we had stayed on CD, but the segment is proportionately shorter and the integral remains unchanged. We could approximate any arbitrary path as a sequence of radial segments and circular arcs, showing that the value of $\oint \vec{B}\cdot d\vec{r}$ is independent of path as long as the path surrounds the current I. The value of that integral is simply $\mu_0 I$. Magnetic fields obey the superposition principle, so this result must be true for any current distribution, not just a single line current. That is, the line integral $\oint \vec{B}\cdot d\vec{r}$ around *any* closed path is directly proportional to the current encircled by that path. This result is **Ampère's law**, a universal statement about current and magnetic field:

$$\oint \vec{B}\cdot d\vec{r} = \mu_0 I_{\text{encircled}} \quad \text{(Ampère's law, steady currents)} \tag{26.17}$$

Ampère's law is another of the four fundamental laws of electromagnetism, although in the form of Equation 26.17 it's limited to steady currents; it also provides a decent approximation for slowly varying currents. In Chapter 29 we'll generalize Ampère's law to remove the restriction to steady currents.

Ampère's law relates the magnetic field to its source—namely, moving charge in the form of electric current—as does the Biot–Savart law. In fact, the laws of Ampère and Biot–Savart are related in the same way as Gauss's and Coulomb's laws. Coulomb and Biot–Savart show how fields arise from pointlike sources—charge elements dq and current elements $I\,d\vec{l}$. Gauss and Ampère are global descriptions, telling how the field must behave over a geometric structure (a closed surface for Gauss, a closed loop for Ampère) in relation to the source (charge or current) enclosed or encircled by that structure. In both cases the field $\vec{E}$ or $\vec{B}$ that appears in the integral is the *net* field arising from *all* sources, not just the enclosed charge or encircled currents.

Ampère's law, like Gauss's, is a truly universal statement. It holds in the electromagnetic devices we build, in atomic and molecular systems, and in distant astrophysical objects. Find a path around which $\oint \vec{B}\cdot d\vec{r}$ isn't zero, and you've found a region where electric current must be flowing (Fig. 26.29). Although it's difficult to show mathematically, the Biot–Savart law follows from Ampère's law just as Coulomb's law follows from Gauss's.

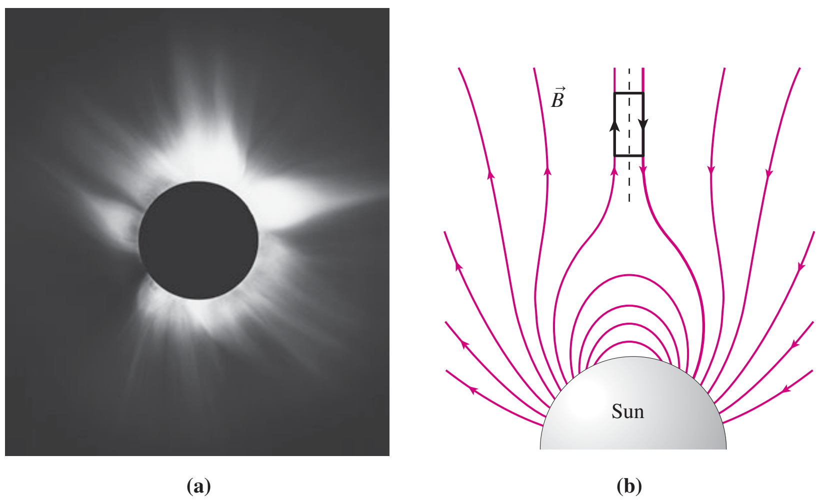


FIGURE 26.29 (a) Coronal streamers in the Sun's atmosphere contain oppositely directed magnetic fields. (b) A model calculation of the magnetic field in a single streamer. Since $\oint \vec{B}\cdot d\vec{r}$ is clearly nonzero around the loop shown, there must be an encircled current.

EXAMPLE 26.6 Ampère's Law: Solar Currents

The long dimension of the rectangular loop in Fig. 26.29*b* is 400 Mm, and the magnetic field strength near the loop has a constant magnitude of 2 mT. Estimate the total current encircled by the rectangle.

INTERPRET This is a problem about currents encircled by a loop, so it must involve Ampère's law.

DEVELOP Figure 26.29*b* provides our drawing. Ampère's law (Equation 26.17) equates $\oint \vec{B}\cdot d\vec{r}$ to $\mu_0 I_{\text{encircled}}$, so we want to evaluate the integral around the loop shown and then solve for $I_{\text{encircled}}$.

EVALUATE On the short segments of the path, $\vec{B}\perp d\vec{r}$, so there's no contribution to the integral. On both long segments, $\vec{B}$ is constant and lies in the direction we're traversing the path, so here $\int \vec{B} \cdot d\vec{r}$ becomes simply Bl, where l is the length of the path segments. Each side contributes this much, so $\oint \vec{B} \cdot d\vec{r} = 2Bl$. Ampère's law equates this quantity to $\mu_0 I_{\text{encircled}}$, so we can solve to get

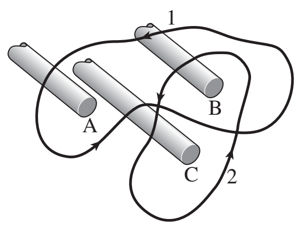
$$I_{\text{encircled}} = \frac{2Bl}{\mu_0} = \frac{(2)(2\text{ mT})(400\text{ Mm})}{4\pi \times 10^{-7}\text{ N/A}^2} = 10^{12}\text{ A}$$

ASSESS This is a large current, but we're dealing with a region much larger than the Earth, so that shouldn't be too surprising. We can get the direction of the current from the right-hand rule: Curl your fingers around the loop in the direction that gives positive $\oint \vec{B} \cdot d\vec{r}$, and your thumb points in the direction of the current. Here that's into the page. In three dimensions this current actually flows around the Sun in approximately the equatorial plane. Note that our result depends crucially on the field reversing across the equatorial plane. In a truly uniform field, one side of the loop would have contributed Bl to the line integral, the other $-Bl$. That would make the integral zero and imply no encircled current. This shows that we can have a uniform field in a current-free region, but not a field that reverses sign—at least not abruptly as in this solar example.

✓TIP Ampèrian Loops

The loop used with Ampère's law is truly arbitrary. It needn't coincide with a field line. In this example, the rectangular loop coincided with the field over its long sides but not along its ends. The loop used with Ampère's law is called an **Ampèrian loop**. Don't confuse Ampèrian loops with field lines; they might coincide, but they don't have to.

GOT IT? 26.6 The figure shows three parallel wires carrying current of the same magnitude I, but in one of them the current direction is opposite that of the other two. If $\oint \vec{B} \cdot d\vec{r} \neq 0$ around loop 2, (a) what's $\oint \vec{B} \cdot d\vec{r}$ around loop 1, and (b) which current is the opposite one?



Using Ampère's Law

For charge distributions with sufficient symmetry, we used Gauss's law to solve for the electric field in a simple and elegant way. We can do the same with Ampère's law for sufficiently symmetric current distributions. Here's the approach:

PROBLEM-SOLVING STRATEGY 26.1 Ampère's Law

INTEPRET Interpret the problem to be sure it's about magnetic field and current. Identify the symmetry.

DEVELOP Based on the symmetry, sketch some field lines. Then find an Ampèrian loop over which you'll be able to evaluate $\oint \vec{B} \cdot d\vec{r}$. That means the field should be constant and parallel to the loop over all or part of the loop; where they're not parallel they should be perpendicular. Like a gaussian surface, the Ampèrian loop is a purely mathematical construct; it need not correspond to anything physical. Draw your loop.

EVALUATE

- Evaluate $\oint \vec{B} \cdot d\vec{r}$ for your loop. This should be easy because you'll be able to take B outside the integral over those segments where it's constant, and segments where $\vec{B}$ is perpendicular to the loop won't contribute to the integral.
- Evaluate the *encircled* current.
- Equate your result for $\oint \vec{B} \cdot d\vec{r}$ to $\mu_0 I_{\text{encircled}}$, and solve for B. Symmetry should give you the direction of $\vec{B}$.

ASSESS Does your answer make sense in terms of what you know about the fields of simple charge and current distributions?

EXAMPLE 26.7 Using Ampère's Law: Outside and Inside a Wire

A long, straight wire of radius R carries a current I distributed uniformly over its cross section. Find the magnetic field (a) outside and (b) inside the wire.

INTERPRET We follow our strategy and identify this as a situation with line symmetry. Therefore, the field depends only on the radial distance from the wire's central axis.

DEVELOP Magnetic field lines encircle their source, so the only field lines consistent with the symmetry are concentric circles. We sketched some in Fig. 26.30. The field is tangent to the field lines and, by symmetry, has the same magnitude B everywhere on a field line. So the field lines themselves make good Ampèrian loops.

EVALUATE

- The field is everywhere parallel to a circular Ampèrian loop, and its magnitude is constant on the loop, so for a loop of radius r, $\oint \vec{B} \cdot d\vec{r}$ becomes just $2\pi rB$. This is true both outside and inside the wire.

We'll first answer (a):

- For any loop *outside* the wire, the encircled current is the total current I.
- Equating our expression for $\oint \vec{B} \cdot d\vec{r}$ to μ_0 times the encircled current gives $2\pi rB = \mu_0 I$, so

$$B = \frac{\mu_0 I}{2\pi r} \quad \left(\begin{array}{c}\text{field outside any current}\\ \text{distribution with line symmetry}\end{array}\right) \qquad (26.18)$$

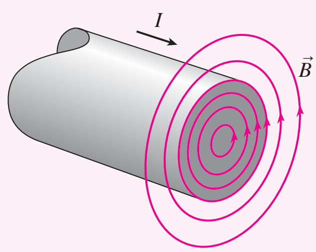


FIGURE 26.30 Cross section of a long cylindrical wire. Any field line can serve as an Ampèrian loop. Inside the wire, the loop's radius r is less than the wire's radius R; outside, $r > R$.

Now on to (b):

- *Inside* the wire, a circular Ampèrian loop encloses only some of the current. How much? With the current uniformly distributed over the wire's cross section, there's a uniform current density $J = I/A = I/\pi R^2$. The encircled current is the current density times the area πr^2 within our loop, so $I_{\text{encircled}} = I(r^2/R^2)$.
- Then we have $2\pi rB = \mu_0 I(r^2/R^2)$, which gives

$$B = \frac{\mu_0 I r}{2\pi R^2} \quad \left(\begin{array}{c}\text{field inside a uniform current}\\ \text{distribution with line symmetry}\end{array}\right) \qquad (26.19)$$

In both cases application of the right-hand rule shows that the field circles counterclockwise, as shown in Fig. 26.30.

ASSESS Equation 26.18 is identical to our result for the line current of Example 26.4, and shows that the field outside *any* current distribution with line symmetry is the same as that of a line current at the symmetry axis. We found the same thing for the electric fields outside cylindrical charge distributions, including the $1/r$ decrease with distance from the axis. Inside the wire, meanwhile, the field increases linearly with distance as we encircle more current in proportion to r^2, while the field decreases as $1/r$. You found a similar result for a uniformly charged cylinder if you worked Problem 54 in Chapter 21. Of course, the electric and magnetic fields of cylindrical distributions look very different—$\vec{E}$ is radial, while $\vec{B}$ forms circles—but the dependence on distance is the same in both cases.

✓TIP Symmetry Is Crucial

Our use of Ampère's law to derive the field of a long wire depends crucially on symmetry. We can't arbitrarily pull B outside the integral unless we know—as we do here from symmetry—that it's constant in magnitude and in direction relative to our Ampèrian loop.

■

EXAMPLE 26.8 Ampère's Law: A Current Sheet

An infinite flat sheet carries current out of this page. The current is distributed uniformly along the sheet, with current per unit width given by J_s. Find the magnetic field of this sheet.

INTERPRET We follow our strategy, identifying the current distribution as having plane symmetry. Then the only thing the field might depend on is the distance from the current-carrying sheet.

DEVELOP The only field lines consistent with the symmetry are straight lines parallel to the plane; we've drawn the current and some field lines in Fig. 26.31. The situation is similar to Example 26.6, and a suitable Ampèrian loop is a rectangle with sides along the field lines and perpendicular edges; we've sketched one such rectangle of width l.

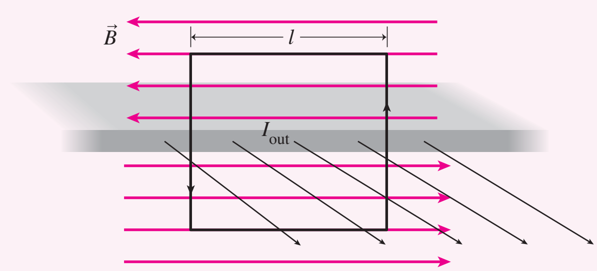


FIGURE 26.31 A current sheet extends infinitely to the left and right, as well as in and out of the page. Field lines and a rectangular Ampèrian loop are shown.

We drew the field in opposite directions on either side of the sheet; it had better be that way if, as we discussed in Example 26.7, $\oint \vec{B} \cdot d\vec{r}$ is to be nonzero. And it has to be nonzero because we know that our Ampèrian loop encircles current.

EVALUATE

- We evaluate $\oint \vec{B} \cdot d\vec{r}$ just as in Example 26.6, getting $2Bl$.
- The sheet carries current J_s per unit width, so our rectangle of width l encircles a current $I_{\text{encircled}} = J_s l$.
- Equating our expression for $\oint \vec{B} \cdot d\vec{r}$ to $\mu_0 I_{\text{encircled}}$ gives $2Bl = \mu_0 J_s l$, or

$$B = \tfrac{1}{2}\mu_0 J_s \quad \text{(field of an infinite current sheet)} \qquad (26.20)$$

ASSESS Make sense? Like the electric field of an infinite plane charge, the magnetic field of an infinite current sheet doesn't depend on distance from the sheet. Of course, there's no truly infinite sheet, so our result is an approximation valid near a finite sheet but not close to its edges. As Fig. 26.32 shows, the lines of a finite sheet wrap around the ends to form closed loops, and far from the loop the field begins to resemble that of a wire. But close in, Equation 26.20 holds. ■

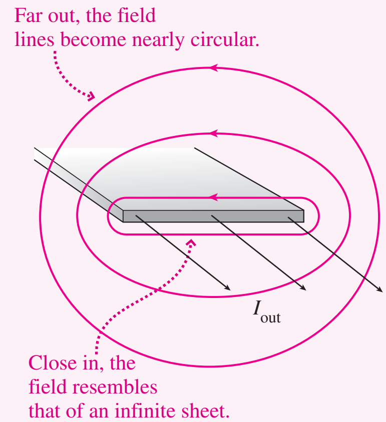
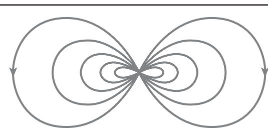
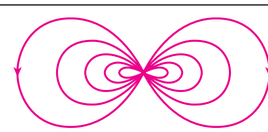
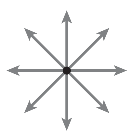
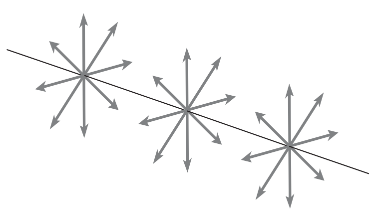
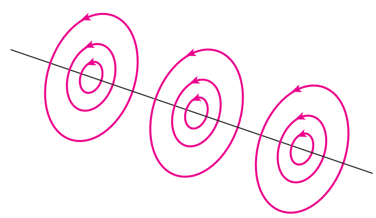
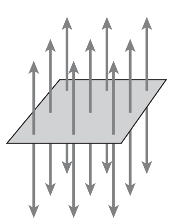
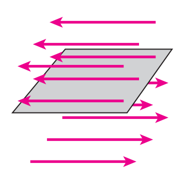


FIGURE 26.32 Field of a finite-width current sheet.

Fields of Simple Current Distributions

We've just used Ampère's law to calculate the fields of two symmetric current distributions, and we compared them with analogous results for electric fields. Table 26.1 summarizes these and other analogies. Although the magnetic and electric fields may look

Table 26.1 Fields of Some Simple Charge and Current Distributions

Field Dependence on Distance[a]	Charge Distribution	Electric Field	Current Distribution	Magnetic Field
$\frac{1}{r^3}$	Electric dipole		Magnetic dipole	
$\frac{1}{r^2}$	Point charge or spherically symmetric		Impossible for steady current	
$\frac{1}{r}$	Charge distribution with line symmetry		Current distribution with line symmetry	
Uniform field; no variation	Infinite flat sheet of charge		Current sheet	

[a]For field *outside* distribution

different, they exhibit the same general relationships between geometry and the way the fields decrease with distance. Real distributions are more complicated, but may often be approximated by these simple cases. Far from *any* current loop, for example, its field approximates that of a dipole. Very near *any* wire, its field is essentially that of a long, straight wire. Very near *any* flat sheet of current, the field is essentially that of Example 26.8.

Solenoids

We found in Chapter 23 that there's an essentially uniform electric field inside a parallel-plate capacitor. Here we explore a current configuration that produces an analogously uniform magnetic field.

Figure 26.33*a* shows a single current loop and its magnetic field. Add a few turns to form an extended coil, and the field isn't much different (Fig. 26.33*b*); more turns (Fig. 26.33*c*), and the region of strongest field is increasingly confined within the coil. With a very long coil (Fig. 26.33*d*), the field is strong and uniform deep within the coil and very weak outside. The limit of an infinitely long, tightly wound coil would produce a uniform field within and no field outside.

A tightly wound coil is a **solenoid**. For a long solenoid—much longer than its diameter—we can use Ampère's law to find the magnetic field inside the solenoid. Figure 26.34 shows a cross section through a solenoid, with a rectangular Ampèrian loop of width l. Since the field is zero outside, the only contribution to $\oint \vec{B} \cdot d\vec{r}$ is from the interior segment parallel to the field, and with a uniform field that gives Bl. If the solenoid carries current I and consists of n turns of wire per unit length, then Fig. 26.34 shows that our Ampèrian loop encircles a total current nlI. So Ampère's law reads $Bl = \mu_0 nlI$, or

$$B = \mu_0 nI \quad \text{(solenoid field)} \tag{26.21}$$

Since the rectangle's vertical dimension never entered the calculation, the field has this same magnitude everywhere inside the solenoid. Although Fig. 26.33 depicts circular coils, Equation 26.21 holds for a solenoid of any cross section.

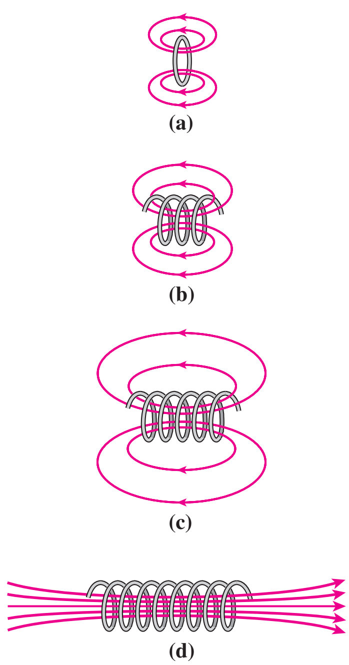


FIGURE 26.33 As the coil gets longer, the interior field stays nearly constant but the exterior field weakens as the field lines spread ever farther apart.

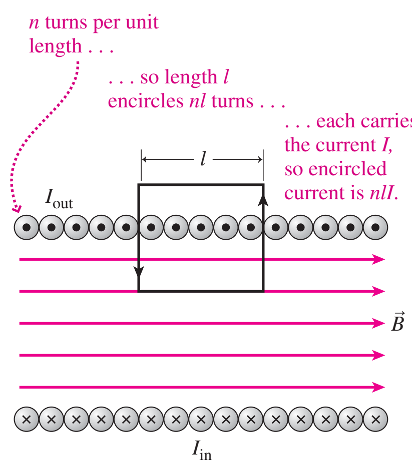


FIGURE 26.34 Cross section of a long solenoid, showing a rectangular Ampèrian loop straddling the region where solenoid coils emerge from the plane of the page.

Solenoids produce uniform magnetic fields in a variety of applications, including the long cylindrical "tunnel" of an MRI scanner. Because the field becomes nonuniform at the ends of a solenoid, ferromagnetic materials are attracted into the coil. Small solenoids can thus produce straight-line motion of an iron plunger. One application is the solenoid on a car starter, which engages the starter motor's gear with the gasoline engine. Solenoid-operated valves are widely used in controlling fluid flows; the valves that admit water to your washing machine and dishwasher are solenoid valves.

EXAMPLE 26.9 A Solenoid: The Current in an MRI Scanner

The solenoid used in an MRI scanner is 2.4 m long and 95 cm in diameter. It's wound from superconducting wire 2.0 mm in diameter, with adjacent turns separated by an insulating layer of negligible thickness. Find the current that will produce a 1.5-T magnetic field inside the solenoid.

INTERPRET This is a problem about a solenoid, which involves relating current and field.

DEVELOP Equation 26.21, $B = \mu_0 nI$, provides the relation we need. To use it we need n, the number of turns per unit length. Figure 26.35 shows how we find n from the wire diameter. Knowing n, we can use Equation 26.21 to find the current.

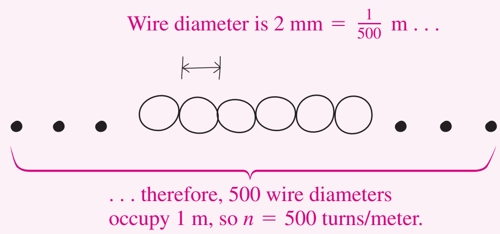


FIGURE 26.35 Finding n.

EVALUATE Figure 26.35 shows that $n = 500$ turns per meter. So now we can solve Equation 26.21 to get

$$I = \frac{B}{\mu_0 n} = \frac{1.5\ \text{T}}{(4\pi \times 10^{-7}\ \text{N/A}^2)(500\ \text{m}^{-1})} = 2.4\ \text{kA}$$

ASSESS That's a large current, but it's readily handled by the niobium–titanium superconductor in the MRI scanner. Notice that more turns per unit length would reduce the current demand; that's because each turn carries the same current I, so more turns increase the encircled current and thus the field for a given total current. ■

With its current flowing around an essentially cylindrical surface, the solenoid might remind you of the bar magnet in Fig. 26.25. There, atomic current loops produce a magnetization current flowing around the cylindrical magnet. Indeed, a solenoid and a bar magnet are very similar, and they produce similar magnetic fields (Fig. 26.36). Wrap a solenoid around on itself and you've got a **toroid**—a donut-shaped coil whose circular field lines close back on themselves. Passage Problems 86–89 explore toroids.

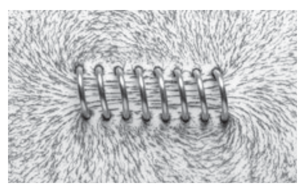


FIGURE 26.36 Iron filings trace the magnetic field of a loosely wound solenoid. Compare with the field of a bar magnet shown in Fig. 26.1.

CHAPTER 26 SUMMARY

Big Picture

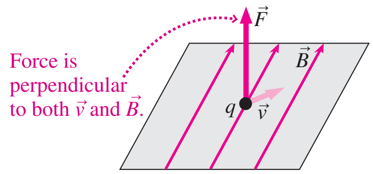
The big new idea here is magnetism—an interaction that fundamentally involves *moving electric charge*. Moving charge produces magnetic fields, and moving charges respond to magnetic fields by experiencing a magnetic force.

Key Concepts and Equations

The **magnetic force** on a charge q moving with velocity $\vec{v}$ in a magnetic field $\vec{B}$ is

$$\vec{F} = q\vec{v} \times \vec{B} \qquad \text{(magnetic force)}$$

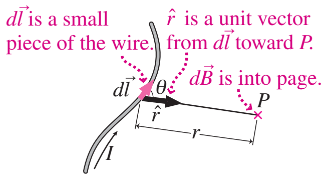
The force acts at right angles to both $\vec{v}$ and $\vec{B}$, and therefore it does no work.



The **Biot–Savart law** describes the magnetic field $d\vec{B}$ arising from a small element of steady current:

$$d\vec{B} = \frac{\mu_0}{4\pi}\frac{I\,d\vec{l} \times \hat{r}}{r^2}$$

Here μ_0 is the **permeability constant**, with value $4\pi \times 10^{-7}$ N/A^2.



Ampère's law provides a more global description of how magnetic fields arise from currents, relating the line integral around any closed loop to the encircled current:

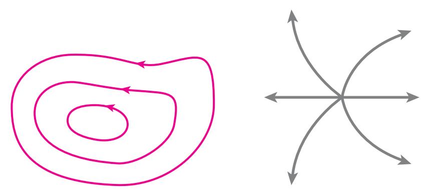
$$\oint \vec{B} \cdot d\vec{r} = \mu_0 I_{\text{encircled}}$$

Ampère's law in this form applies only to steady currents.

Gauss's law for magnetism expresses the fact that there are no magnetic monopoles—magnetic analogs of electric charge—and that magnetic field lines therefore do not begin or end:

$$\oint \vec{B} \cdot d\vec{A} = 0$$

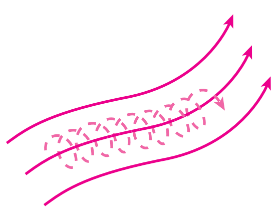
Static electric fields, in contrast, always begin or end on electric charges.



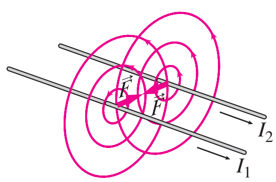
Magnetic field | Electric field

Applications

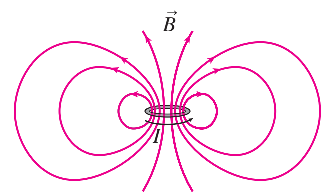
A charged particle moving perpendicular to a uniform magnetic field undergoes circular motion with the **cyclotron frequency** $f = qB/2\pi m$. More generally, charged particles in magnetic fields follow spiral paths, "trapped" on the field lines.



The magnetic force on a straight wire of length l carrying current I in a uniform magnetic field is $\vec{F} = I\vec{l} \times \vec{B}$. Parallel wires a distance d apart experience forces from each other's magnetic field: $F = \dfrac{\mu_0 I_1 I_2 l}{2\pi d}$. The force is attractive for currents in the same direction, repulsive for currents in opposite directions.

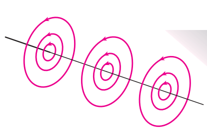


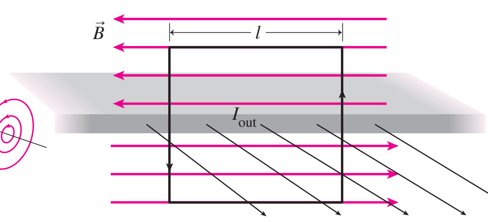
A current loop gives rise to a magnetic field that, at distances large compared with the loop's size, is a dipole field. The loop's magnetic dipole moment has magnitude $\mu = IA$, with A the loop area, and the loop responds to an external magnetic field by experiencing the torque typical of a dipole: $\vec{\tau} = \vec{\mu} \times \vec{B}$.

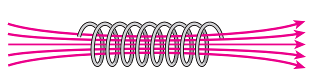


Fields of simple current distributions:

Line current: $B = \dfrac{\mu_0 I}{2\pi r}$ Current sheet: $B = \frac{1}{2}\mu_0 J_s$ Solenoid: $B = \mu_0 n I$







Magnetism in matter arises from the interactions of atomic-scale current loops. **Ferromagnetic** materials have strong interactions and exhibit the bulk magnetism associated with permanent magnets and with magnetic materials like iron. **Paramagnetism** and **diamagnetism** are weaker manifestations of magnetism in matter.

MP For homework assigned on MasteringPhysics, go to www.masteringphysics.com | BIO Biology and/or medical-related problems C Computer problems

For Thought and Discussion

1. A charged particle moves through a region containing only a magnetic field. Under what condition will it experience no force?
2. An electron moving with velocity $\vec{v}$ through a magnetic field $\vec{B}$ experiences a magnetic force $\vec{F}$. Which of the vectors $\vec{F}$, $\vec{v}$, and $\vec{B}$ must be at right angles?
3. A magnetic field points out of this page. Will a positively charged particle moving in the plane of the page circle clockwise or counterclockwise as viewed from above?
4. Do particles in a cyclotron gain energy from the electric field, the magnetic field, or both? Explain.
5. An electron and a proton moving at the same speed enter a region containing a uniform magnetic field. Which is deflected more from its original path?
6. Two identical particles carrying equal charge are moving in opposite directions along a magnetic field, when they collide elastically head-on. Describe their subsequent motion.
7. In what two senses does a current loop behave like a magnetic dipole?
8. The Biot–Savart law shows that the magnetic field of a current element decreases as $1/r^2$. Could you put together a complete circuit whose field exhibits this decrease? Why or why not?
9. Do currents in the same direction attract or repel? Explain.
10. If a current is passed through an unstretched spring, will the spring contract or expand? Explain.
11. Figure 26.37 shows some magnetic field lines associated with two parallel wires carrying equal currents perpendicular to the page. Are the currents in the same or opposite directions? How can you tell?

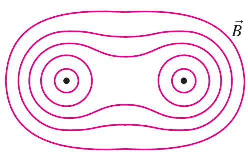


FIGURE 26.37 For Thought and Discussion 11

12. Why is a piece of iron attracted into a solenoid?
13. Would there be a magnetic force on a piece of iron deep inside a long solenoid? Explain.
14. An unmagnetized piece of iron has no net magnetic dipole moment, yet it's attracted to either pole of a bar magnet. Why?

Exercises and Problems

Exercises

Section 26.2 Magnetic Force and Field

15. Find (a) the minimum magnetic field needed to exert a 5.4-fN force on an electron moving at 21 Mm/s and (b) the field strength required if the field were at 45° to the electron's velocity.
16. An electron moving at right angles to a 0.10-T magnetic field experiences an acceleration of 6.0×10^{15} m/s^2. (a) What's its speed? (b) By how much does its speed change in 1 ns?
17. Find the magnitude of the magnetic force on a proton moving at 2.5×10^5 m/s (a) perpendicular; (b) at 30°; (c) parallel to a 0.50-T magnetic field.
18. The magnitude of Earth's magnetic field is about 0.5 gauss near Earth's surface. What's the maximum possible magnetic force on an electron with kinetic energy of 1 keV? Compare with the gravitational force on the electron.
19. A velocity selector uses a 60-mT magnetic field perpendicular to a 24-kN/C electric field. At what speed will charged particles pass through the selector undeflected?

Section 26.3 Charged Particles in Magnetic Fields

20. Find the radius of the path described by a proton moving at 15 km/s in a plane perpendicular to a 400-G magnetic field.
21. How long does it take an electron to complete a circular orbit perpendicular to a 1.0-G magnetic field?
22. Radio astronomers detect electromagnetic radiation at a frequency of 42 MHz from an interstellar gas cloud. If the radiation results from electrons spiraling in a magnetic field, what's the field strength?
23. Microwaves in a microwave oven are produced by electrons circling in a magnetic field at a frequency of 2.4 GHz. (a) What's the magnetic field strength? (b) The electrons' motion takes place inside a special tube called a *magnetron*. If the magnetron can accommodate electron orbits with maximum diameter 2.5 mm, what's the maximum electron energy?
24. Two protons, moving in a plane perpendicular to a uniform 500-G magnetic field, undergo an elastic head-on collision. How much time elapses before they collide again?

Section 26.4 The Magnetic Force on a Current

25. Find the magnitude of the force on a 50-cm-long wire carrying 15 A at right angles to a 500-G magnetic field.
26. A wire carrying 15 A makes a 25° angle with a uniform magnetic field. The magnetic force per unit length of wire is 0.31 N/m. Find (a) the magnetic field strength and (b) the maximum force per unit length that could be achieved by reorienting the wire.
27. You're on a team performing a high-magnetic-field experiment. A conducting bar carrying 4.1 kA will pass through a 1.3-m-long region containing a 12-T magnetic field, making a 60° angle with the field. A colleague proposes resting the bar on wooden blocks. You argue that it will have to be clamped in place, and to back up your argument you claim that the magnetic force will exceed 10,000 pounds. Are you right?
28. A wire with mass per unit length 75 g/m runs horizontally at right angles to a horizontal magnetic field. A 6.2-A current in the wire results in its being suspended against gravity. What's the magnetic field strength?

Section 26.5 Origin of the Magnetic Field

29. A wire carries 15 A. You form it into a single-turn circular loop with magnetic field 80 μT at the loop center. What's the loop radius?
30. A single-turn wire loop is 2.0 cm in diameter and carries a 650-mA current. Find the magnetic field strength (a) at the loop center and (b) on the loop axis, 20 cm from the center.
31. A 2.2-m-long wire carrying 3.5 A is wound into a tight coil 5.0 cm in diameter. Find the magnetic field at its center.
32. What's the current in a long wire if the magnetic field strength 1.2 cm from the wire's axis is 67 μT?
33. In standard household wiring, parallel wires about 1 cm apart carry currents of about 15 A. What's the force per unit length between these wires?

Section 26.6 Magnetic Dipoles

34. Earth's magnetic dipole moment is 8.0×10^{22} A·m^2. Find the magnetic field strength at Earth's magnetic poles.
35. A single-turn square wire loop 5.0 cm on a side carries a 450-mA current. (a) What's the loop's magnetic dipole moment? (b) If the loop is in a uniform 1.4-T magnetic field with its dipole moment vector at 40° to the field, what's the magnitude of the torque it experiences?
36. An electric motor contains a 250-turn circular coil 6.2 cm in diameter. If it develops a maximum torque of 1.2 N·m at a current of 3.3 A, what's the magnetic field strength?

Section 26.8 Ampère's Law

37. The line integral of the magnetic field on a closed path surrounding a wire has the value 8.8 μT·m. Find the current in the wire.
38. The magnetic field shown in Fig. 26.38 has uniform magnitude 75 μT, but its direction reverses abruptly. Find the current encircled by the rectangular loop shown.

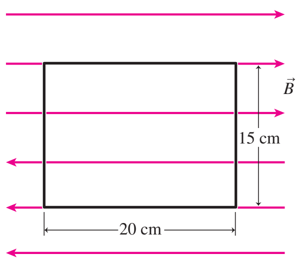


FIGURE 26.38 Exercise 38

39. A wire 1.0 mm in diameter carries 5.0 A distributed uniformly over its cross section. Find the field strength (a) 0.10 mm from its axis and (b) at the wire's surface.
40. Show that Equations 26.18 and 26.19 give the same results when evaluated at the wire's surface.
41. A superconducting solenoid has 3300 turns per meter and carries 4.1 kA. Find the magnetic field strength in the solenoid.

Problems

42. A particle carrying a 50-μC charge moves with velocity $\vec{v} = 5.0\hat{\imath} + 3.2\hat{k}$ m/s through a magnetic field $\vec{B} = 9.4\hat{\imath} + 6.7\hat{\jmath}$ T. (a) Find the magnetic force on the particle. (b) Form the dot products $\vec{F}\cdot\vec{v}$ and $\vec{F}\cdot\vec{B}$ to show explicitly that the force is perpendicular to both $\vec{v}$ and $\vec{B}$.
43. Jupiter has the strongest magnetic field in our solar system, about 14 G at its poles. Approximating the field as that of a dipole, find Jupiter's magnetic dipole moment. (*Hint:* Consult Appendix E.)
44. A proton moving with velocity $\vec{v}_1 = 3.6\times10^4\hat{\jmath}$ m/s experiences a magnetic force of $7.4\times10^{-16}\hat{\imath}$ N. A second proton moving on the x-axis experiences a magnetic force of $2.8\times10^{-16}\hat{\jmath}$ N. Find the magnitude and direction of the magnetic field (assumed uniform), and the velocity of the second proton.
45. A simplified model of Earth's magnetic field has it originating in a single current loop at the outer edge of the planet's liquid core (radius 3000 km). What current would give the 62-μT field measured at the north magnetic pole?
46. A beam of electrons moving in the x-direction at 8.7 Mm/s enters a region where a uniform 180-G magnetic field points in the y-direction. The boundary of the field region is perpendicular to the beam. How far into the field region does the beam penetrate?
47. Show that the orbital radius of a charged particle moving at right angles to a magnetic field B can be written $r = \sqrt{2Km}/qB$, where K is the kinetic energy in joules, m the particle's mass, and q its charge.
48. A 90-cm-diameter cyclotron with a 2.0-T magnetic field is used to accelerate deuterium nuclei (one proton plus one neutron). (a) At what frequency should the dee voltage be alternated? (b) What's the maximum kinetic energy of the deuterons? (c) If the magnitude of the potential difference between the dees is 1500 V, how many orbits do the deuterons complete before reaching maximum energy?
49. An electron is moving in a uniform 0.25-T magnetic field; its velocity components parallel and perpendicular to the field are both 3.1 Mm/s. (a) What's the radius of the electron's spiral path? (b) How far does it move along the field direction in the time it takes to complete a full orbit about the field?
50. A wire of negligible resistance is bent into a rectangle as in Fig. 26.39, and a battery and resistor are connected as shown. The right-hand side of the circuit extends into a region containing a uniform 38-mT magnetic field pointing into the page. Find the magnitude and direction of the net force on the circuit.

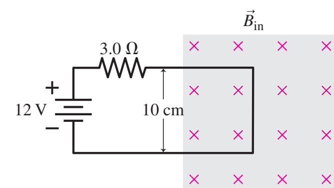


FIGURE 26.39 Problem 50

51. **BIO** You're designing a prosthetic ankle that includes a miniature electric motor containing a 150-turn circular coil 15 mm in diameter. The motor needs to develop a maximum torque of 3.1 mN·m. The strongest magnets available that will fit in the prosthesis produce a 220-mT field. What current do you need in your motor's coil?
52. A 20-cm-long conducting rod with mass 18 g is suspended by wires of negligible mass (Fig. 26.40). A uniform magnetic field of 0.15 T points horizontally into the page, as shown. An external circuit supplies current between the supports A and B. (a) What's the minimum current necessary to move the bar to the upper position, so it's supported against gravity? (b) What direction should the current flow?

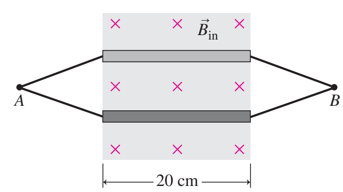


FIGURE 26.40 Problem 52

53. A rectangular copper strip measures 1.0 mm in the direction of a uniform 2.4-T magnetic field. When the strip carries a 6.8-A current perpendicular to the field, a 1.2-μV Hall potential develops across the strip. Find the number density of free electrons in the copper.

54. A single-turn wire loop 10 cm in diameter carries a 12-A current. It experiences a 0.015 N·m torque when the normal to the loop plane makes a 25° angle with a uniform magnetic field. Find the magnetic field strength.
55. A simple electric motor consists of a 100-turn coil 3.0 cm in diameter, mounted between the poles of a magnet that produces a 0.12-T field. When a 5.0-A current flows in the coil, what are (a) its magnetic dipole moment and (b) the motor's maximum torque?
56. **BIO** Nuclear magnetic resonance (NMR) is a technique for analyzing chemical structures and also the basis of magnetic resonance imaging used for medical diagnosis. NMR relies on sensitive measurements of the energy needed to flip atomic nuclei by 180° in a given magnetic field. In an apparatus with a 9.4-T magnetic field, what energy is needed to flip a proton ($\mu = 1.41 \times 10^{-26}$ A·m^2) from parallel to antiparallel to the field?
57. A wire carrying 1.5 A passes through a 48-mT magnetic field. The wire is perpendicular to the field and makes a quarter-circle turn of radius 21 cm in the field region, as shown in Fig. 26.41. Find the magnitude and direction of the magnetic force on the curved section of wire.

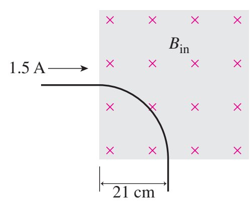


FIGURE 26.41 Problem 57

58. Your company is developing a device incorporating a 20-cm-diameter coil carrying 0.50 A that, when properly oriented, will just cancel Earth's 50-μT magnetic field at the coil's center. How much wire must you requisition for each coil?
59. A single piece of wire carrying current I is bent so it includes a circular loop of radius a, as shown in Fig. 26.42. Find an expression for the magnetic field at the loop center.

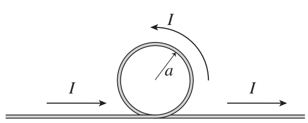


FIGURE 26.42 Problem 59

60. You and a friend get lost while hiking, so your friend pulls out a magnetic compass to get re-oriented. However, you're standing right under a power line carrying 1.5 kA toward magnetic north; it's 10 m above the compass. The horizontal component of Earth's magnetic field at your latitude points northward and has magnitude 0.24 G. Will the compass help you find your way?
61. Part of a long wire carrying current I is bent into a semicircle of radius a, as in Fig. 26.43. Use the Biot–Savart law to find the magnetic field at P, the center of the semicircle.

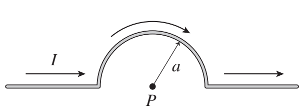


FIGURE 26.43 Problem 61

62. Three parallel wires of length l each carry current I in the same direction. They're positioned at the vertices of an equilateral triangle of side a, and oriented perpendicular to the triangle. Find an expression for the magnitude of the force on each wire.
63. A long, straight wire carries 20 A. A 5.0-cm by 10-cm rectangular wire loop carrying 500 mA is 2.0 cm from the wire, as shown in Fig. 26.44. Find the net magnetic force on the loop.

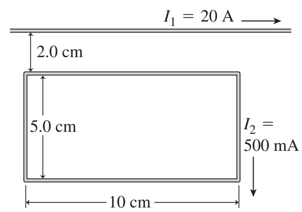


FIGURE 26.44 Problem 63

64. A long conducting rod of radius R carries a nonuniform current density $J = J_0 r/R$, where J_0 is a constant and r is the radial distance from the rod's axis. Find expressions for the magnetic field strength (a) inside and (b) outside the rod.
65. A long, hollow conducting pipe of radius R carries a uniform current I along the pipe, as shown in Fig. 26.45. Use Ampère's law to find the magnetic field strength (a) inside and (b) outside the pipe.

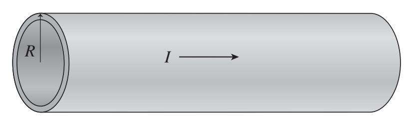


FIGURE 26.45 Problem 65

66. The coaxial cable shown in Fig. 26.46 consists of a solid inner conductor of radius a and a hollow outer conductor of inner radius b and thickness c. The two carry equal but opposite currents I, uniformly distributed. Find expressions for the magnetic field as a function of radial position r (a) within the inner conductor, (b) between the inner and outer conductors, and (c) beyond the outer conductor.

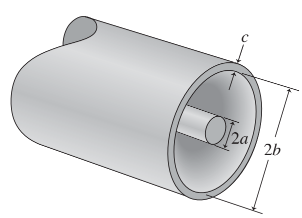


FIGURE 26.46 Problems 66 and 71

67. A solenoid used in a plasma physics experiment is 10 cm in diameter, is 1.0 m long, and carries a 35-A current to produce a 100-mT magnetic field. (a) How many turns are in the solenoid? (b) If the solenoid resistance is 2.7 Ω, how much power does it dissipate?
68. You have 10 m of 0.50-mm-diameter copper wire and a battery capable of passing 15 A through the wire. What magnetic field strengths could you obtain (a) inside a 2.0-cm-diameter solenoid wound with the wire as closely spaced as possible and (b) at the center of a single circular loop made from the wire?

69. Derive Equation 26.21 for the solenoid field by considering the solenoid to be made of infinitesimal current loops. Use Equation 26.9 for the loop fields, and integrate over all loops.
70. The largest lightning strikes have peak currents of around 250 kA, flowing in essentially cylindrical channels of ionized air. How far from such a flash would the resulting magnetic field be equal to Earth's magnetic field strength, about 50 μT?
71. A coaxial cable (see Fig. 26.46) consists of a 1.0-mm-diameter inner conductor and a 0.20-mm-thick outer conductor with interior diameter 1.0 cm. A 100-mA current flows down the inner conductor and back along the outer conductor. Find the magnetic field strength (a) 0.10 mm, (b) 5.0 mm, and (c) 2.0 cm from the cable axis.
72. A circular wire loop of radius 15 cm and negligible thickness carries a 2.0-A current. Use suitable approximations to find the magnetic field of this loop (a) in the loop plane, 1.0 mm outside the loop, and (b) on the loop axis, 3.0 m from the loop center.
73. A long, flat conducting bar of width w carries a total current I distributed uniformly, as shown in Fig. 26.47. Use approximations to write expressions for the magnetic field strength (a) near the conductor surface $(r \ll w)$ but not near its edges and (b) far from the conductor $(r \gg w)$.

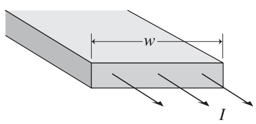


FIGURE 26.47 Problem 73

74. A long, hollow conducting pipe of radius R and length l carries a uniform current I flowing around the pipe (Fig. 26.48). Find expressions for the magnetic field (a) inside and (b) outside the pipe. (*Hint:* What configuration does this resemble?)

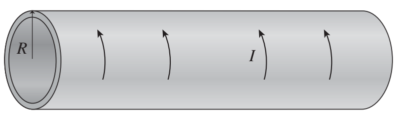


FIGURE 26.48 Problem 74

75. A solid conducting wire of radius R runs parallel to the z-axis and carries a current density given by $\vec{J} = J_0(1 - r/R)\hat{k}$, where J_0 is a constant and r is the distance from the wire axis. Find expressions for (a) the total current in the wire and (b) the magnetic field for $r > R$ and (c) $r < R$.
76. A disk of radius a carries uniform surface charge density σ and rotates with angular speed ω about the disk axis. Show that the magnetic field at the disk's center is $\frac{1}{2}\mu_0\sigma\omega a$.
77. You're developing a system to orient an orbiting telescope. The system uses three perpendicular coils, with torques developed in Earth's magnetic field when current passes through them. Weight limitations restrict you to a length l of wire for each coil. A colleague argues you'll get the greatest dipole moment and therefore the most torque with a multi-turn coil. You say a 1-turn coil is best. Who's right?
78. The structure shown in Fig. 26.49 is made from conducting rods. The upper horizontal rod (mass 22 g, length 95 cm) is free to slide vertically on the uprights while maintaining electrical contact. A battery connected across the insulating gap at the bottom of the left-hand upright drives 66 A through the structure. At what height h will the upper wire be in equilibrium?

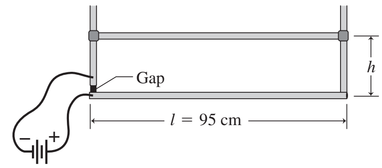


FIGURE 26.49 Problem 78

79. A long, flat conducting ribbon of width w is parallel to a long, straight wire; its near edge is a distance a from the wire (Fig. 26.50). Wire and ribbon carry the same current I; it's distributed uniformly over the ribbon. Use integration to show that the force per unit length between the two has magnitude $\dfrac{\mu_0 I^2}{2\pi w}\ln\left(\dfrac{a+w}{a}\right)$.

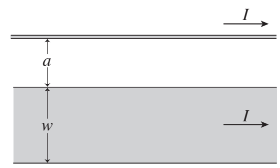


FIGURE 26.50 Problem 79

80. Find an expression for the magnetic field at the center of a square loop of side a carrying current I.
81. Repeat the calculation in Problem 69 for a solenoid of finite length l and cross-sectional radius a to find the magnetic field strength at the center of the solenoid's axis.
82. A magnetic dipole $\vec{\mu} = \mu\hat{\imath}$ is on the axis of a circular current loop of radius a oriented as shown in Fig. 26.16a, a distance x from the center. Differentiate Equation 26.16 to find the force on the dipole, and evaluate its magnitude for $x = a$. Is the force attractive or repulsive?
83. You have a summer job working for an audio equipment manufacturer. The loudspeaker engineer asks you to make a calculation for a prototype speaker. The speaker coil consists of 100 turns of wire, 3.5 cm in diameter, suspended in a uniform magnetic field. When the coil current is 2.1 A, the force on the coil should be 14.8 N. What magnetic field will give this force?
84. Derive Equation 26.20 by considering the current sheet to be made of infinitely many infinitesimal line currents.
85. **BIO** Your roommate is sold on "magnet therapy," a sham treatment using small bar magnets attached to the body. You skeptically ask your roommate how this is supposed to work. He mumbles something about the Hall effect speeding blood flow. In reply, you estimate the Hall potential associated with typical blood parameters in the 100-G field of a bar magnet: red blood cells carrying 2-pC charge in a 12-cm/s flow through a 3.0-mm-diameter blood vessel containing 5 billion red blood cells per mL. To show that the Hall potential is negligible, you compare your estimate with the tens of mV typical of bioelectric activity. How do the two values compare?

Passage Problems

A *toroid* is a solenoid-like coil bent into a circle (Fig. 26.51*a*). Toroids are the configuration of choice in magnetic-confinement nuclear fusion experiments, which, if successful, could provide us with an almost unlimited energy source using deuterium fuel extracted from seawater. The ITER consortium, an international collaboration, is building a large toroidal fusion experiment in France; it's expected to be the first fusion device to produce energy on a large scale. Figure 26.51*b* shows a cross section of a toroid, with current emerging from the page at the inner edge and descending at the outer edge. The black circle is an Ampèrian loop.

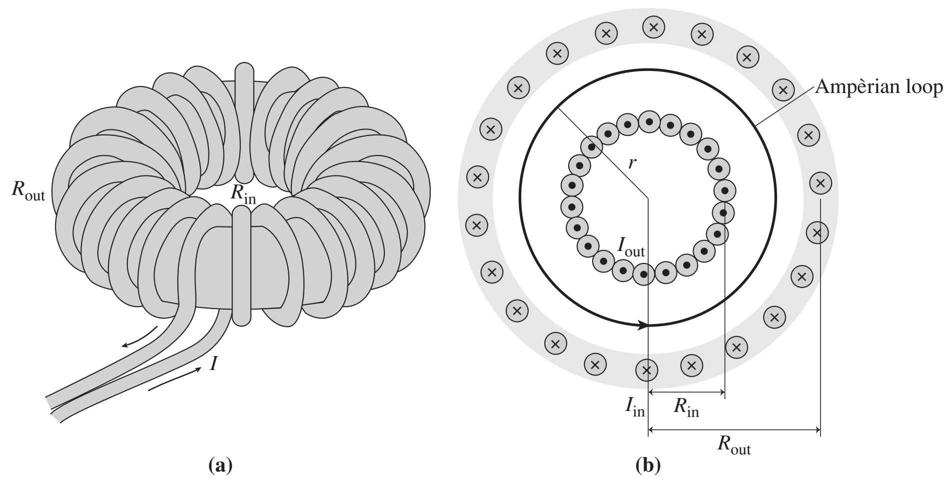


FIGURE 26.51 Diagram of (a) a toroidal coil and (b) a cross section of the coil (Passage Problems 86–89)

86. The magnetic field associated with the toroid is nonzero
    a. only within the "hole" in the donut-shaped coil.
    b. only within the region bounded by the coils.
    c. only outside the coils.
    d. everywhere.
87. In Fig. 26.51*b*, the magnetic field lines must be
    a. straight, and pointing into the page.
    b. straight, and pointing out of the page.
    c. straight, and pointing radially.
    d. circular.
88. Doubling the total number of turns N in the toroid, without changing its size or the current, will
    a. double the magnetic field.
    b. quadruple the magnetic field.
    c. halve the magnetic field.
    d. not change the magnetic field.
89. The toroid has inner radius R_{in} and outer radius R_{out}, while r is the radial coordinate measured from the center. The toroid is made from wire wound into a total of N turns, and carries current I. Which of the following is the correct formula for the magnetic field within the coils?
    a. $B = \mu_0 NI$
    b. $B = \mu_0 NI/2\pi R_{in}$
    c. $B = \mu_0 NI/2\pi R_{out}$
    d. $B = \mu_0 NI/2\pi r$

Answers to Chapter Questions

Answer to Chapter Opening Question

Magnetic force shapes the structure of the solar atmosphere. Magnetism is fundamentally an interaction involving moving electric charge, and the hot, ionized gas of the solar atmosphere contains free charge that responds to magnetism. Earth's cooler atmosphere consists of neutral molecules that don't experience a magnetic force.

Answers to GOT IT? Questions

26.1. Greatest for (*a*), 0 for (*c*); direction for (*a*) and (*b*) is into the page.
26.2. (b); it's a *negative* charge.
26.3. (a).
26.4. (a), because adjacent currents are in the same direction. Changing the current direction doesn't matter because the currents are still parallel.
26.5. (b), because the field lines form closed loops.
26.6. (a) 0; (b) current A.

27 Electromagnetic Induction

It takes fourteen 110-car trainloads of coal each week to fuel this power plant. What feature of the equation $\mathcal{E} = -d\Phi_B/dt$ demands this prodigious fuel consumption?

New Concepts, New Skills

By the end of this chapter you should be able to

- Explain electromagnetic induction (27.1).
- Use Faraday's law to calculate induced emfs and currents (27.2).
- Describe how induction is consistent with energy conservation, and find the direction of induced emfs and currents (27.3).
- Explain inductance, and analyze simple circuits involving inductors (27.4).
- Recognize that magnetic fields store energy, and calculate that energy for simple magnetic-field configurations (27.5).
- Recognize the broader meaning of Faraday's law as it relates electric and magnetic fields (27.6).

Connecting Your Knowledge

- This chapter introduces a new phenomenon, but involves fundamental ideas introduced previously: the electric field (20.3), electric current (24.1), electromotive force (emf) (25.1), the magnetic field (26.1), and magnetic force (26.2).
- We'll also work extensively with solenoids (26.8).
- You'll find the mathematics of line and flux integrals useful (6.2, 26.8, 21.2).

The electric and magnetic fields we've encountered so far originated in electric charge, either stationary or moving. We recognized a link between electricity and magnetism that lies in their common involvement with electric charge. In the remainder of our study, we'll explore a more intimate relation between electricity and magnetism, in which the fields themselves interact directly. This relation is the basis for new electromagnetic technologies, reveals the nature of light, and points toward the theory of relativity.

27.1 Induced Currents

In 1831, the English scientist Michael Faraday and the American Joseph Henry independently found that electric currents arose in circuits subjected to changing magnetic fields. Here are four experiments that illustrate this phenomenon:

1. (Fig. 27.1) Move a bar magnet in the presence of a circuit consisting of a wire coil and an ammeter. There's no battery or other obvious source of emf. As long as you hold the magnet stationary, there's no current. But move the magnet, and the ammeter registers a current—which we call an **induced current**. Move the magnet faster, and the induced current increases. Reverse the direction of motion, and the induced current reverses.

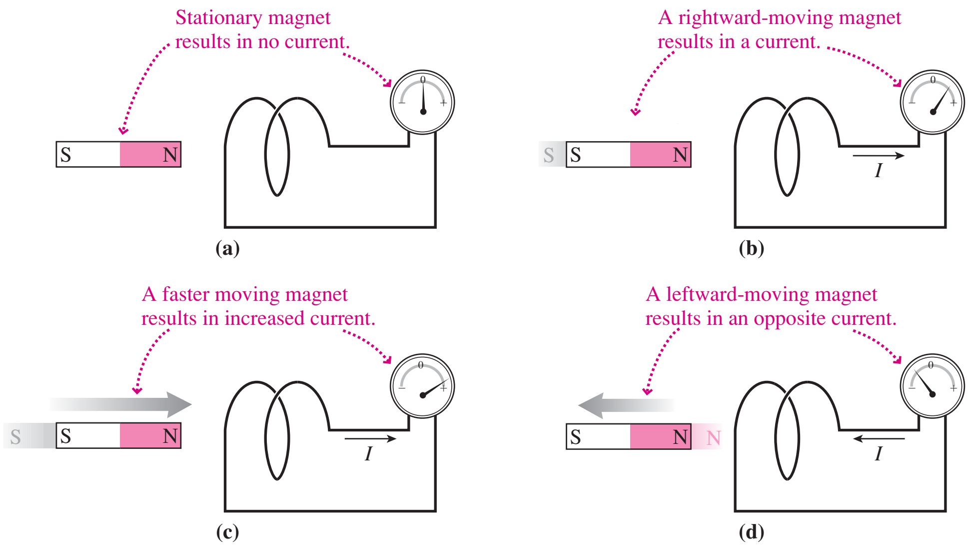


FIGURE 27.1 When a magnet moves near a closed circuit, current flows in the circuit.

2. (Fig. 27.2) Move a coil near a stationary magnet, and a similar induced current results. So the effect is the same whether it's the magnet that moves, or the coil. All that matters is the relative motion between magnet and coil.
3. (Fig. 27.3) Replace the bar magnet with a second coil, this one carrying a steady current from a battery. The new coil creates a magnetic field like that of a bar magnet and, not surprisingly given the results of experiments 1 and 2, an induced current arises in the original coil when the two coils move relative to one another.

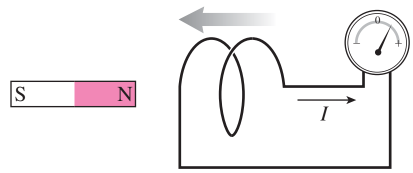


FIGURE 27.2 Moving the coil instead of the magnet gives the same result, as in Fig. 27.1*b*.

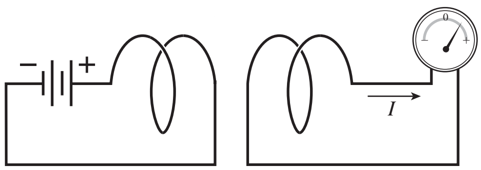


FIGURE 27.3 An induced current also results when a current-carrying circuit replaces the magnet.

4. (Fig. 27.4) Hold both coils stationary, and there's no induced current. But now open the switch connecting the battery to the left-hand coil. The current drops quickly to zero, and during that brief interval the ammeter registers a current in the right-hand coil. Then the induced current ceases as the current in the left-hand coil remains at zero. Now close the switch again; as current briefly rises in the left-hand coil, the ammeter registers an induced current in the right-hand coil—and its direction is opposite what it was when you opened the switch. Once the current in the left coil reaches a steady value, the induced current in the right coil again ceases.

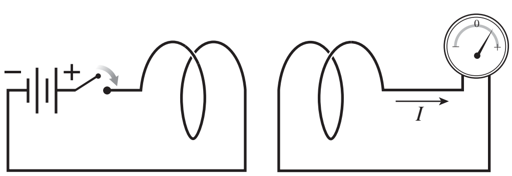


FIGURE 27.4 A current is also induced when the current in an adjacent circuit changes.

The common feature in these experiments is a *changing magnetic field*. It doesn't matter whether the field changes because a magnet moves, or a circuit moves, or because the current giving rise to the field changes. In each case, an induced current appears in a circuit subjected to a changing magnetic field. Here's a new phenomenon—**electromagnetic induction**—whereby electrical effects arise from *changing* magnetic fields.

27.2 Faraday's Law

It takes a source of emf, such as a battery, to drive current in a circuit. With induced currents there's no battery, but there still must be an emf. This **induced emf** isn't usually localized, as with a battery, but may be spread through the conductors making up the circuit.

Magnetic Flux

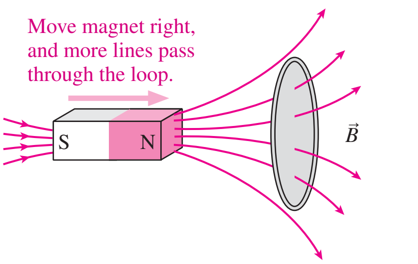


FIGURE 27.5 A circular wire loop in the magnetic field of a bar magnet. As the magnet moves closer, the flux through the loop increases.

We found in Chapter 26 that the magnetic flux through any *closed* surface is zero. Here we're interested in the flux through *open* surfaces, which need not be zero (Fig. 27.5). Like the electric flux defined in Chapter 21, magnetic flux is the integral of the magnetic field over a surface:

$$\Phi_B = \int \vec{B} \cdot d\vec{A} \quad \text{(magnetic flux)} \tag{27.1a}$$

With electromagnetic induction, we're interested in the flux through a surface bounded by a circuit. For a loop like the one in Fig. 27.5, that surface can be the circular disk whose circumference is the loop. More generally, it can be *any* surface bounded by the loop. For a flat surface in a uniform magnetic field, Equation 27.1a reduces to

$$\Phi_B = \vec{B} \cdot \vec{A} = BA\cos\theta \quad \text{(magnetic flux, uniform field and flat area)} \tag{27.1b}$$

where θ is the angle between the field and the normal to the area. When the field and area are perpendicular, as in the next example, Equation 27.1b reduces further to $\Phi_B = BA$.

EXAMPLE 27.1 Magnetic Flux: A Solenoid

A solenoid of circular cross section has radius R, consists of n turns per unit length, and carries current I. Find the magnetic flux through each turn of the solenoid.

INTERPRET The solenoid creates a uniform magnetic field, and we're asked for the flux of this field through an area bounded by one turn of the solenoid.

DEVELOP The solenoid field is perpendicular to the turns of wire that make up the solenoid, as we've drawn in Fig. 27.6. So we have a uniform field at right angles to a flat area, and the flux becomes $\Phi_B = BA$. Equation 26.20 gives the solenoid field, $B = \mu_0 nI$, and the area is that of a circle, πR^2.

EVALUATE $\Phi_B = BA = \mu_0 nI\pi R^2$

ASSESS The flux increases with any factor that increases either the field or the area, so this result makes sense. We're being a little loose in thinking of a single turn of the solenoid as a closed loop, but if the solenoid is tightly wound, then this is an excellent approximation. What we've found is the flux through one turn; if the solenoid has N turns, then the flux through the entire solenoid is N times our result. ■

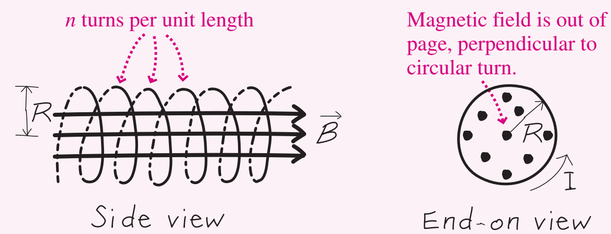


FIGURE 27.6 Sketch for Example 27.1.

EXAMPLE 27.2 Magnetic Flux: A Nonuniform Field

A long, straight wire carries current I. A rectangular wire loop of dimensions l by w lies in a plane containing the wire, with its closest edge a distance a from the wire and its dimension l parallel to the wire. Find the magnetic flux through the loop.

INTERPRET The long, straight wire gives rise to a magnetic field, and we're asked for the flux of this field through an adjacent rectangular area.

DEVELOP Figure 27.7 shows the situation. Field lines encircle the long wire, and at the rectangular loop they're pointing into the page, perpendicular to the loop area. Thus $\vec{B} \cdot d\vec{A}$ in Equation 27.1a becomes just $B\,dA$. Equation 26.17 gives the field strength: $B = \mu_0 I/2\pi r$. Since this varies with distance from the wire, we have

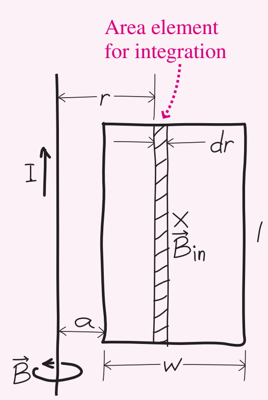


FIGURE 27.7 A rectangular loop in the magnetic field of a long wire.

to integrate. We divide the rectangle into thin strips of width dr and area $dA = l\,dr$. Knowing B and dA, we can integrate over all such strips.

EVALUATE We have

$$\Phi_B = \int B\,dA = \int_a^{a+w} \frac{\mu_0 I}{2\pi r} l\,dr = \frac{\mu_0 I l}{2\pi}\int_a^{a+w} \frac{dr}{r}$$

The integral is the natural logarithm, so

$$\Phi_B = \frac{\mu_0 I l}{2\pi} \ln r \Big|_a^{a+w} = \frac{\mu_0 I l}{2\pi} \ln\left(\frac{a+w}{a}\right)$$

ASSESS This result is directly proportional to the current, which determines the field strength, and to the loop length l. But it isn't proportional to the width w because the field strength falls off, and increasing w would expand the loop into regions of weaker field, contributing less to the overall flux. ■

Flux and Induced emf

Having developed the notion of magnetic flux, we can now state **Faraday's law of induction**, another of the four basic laws of electromagnetism:

> The induced emf in a circuit is proportional to the rate of change of magnetic flux through any surface bounded by that circuit.

This statement is a special case of Faraday's law that describes electromagnetic induction specifically in circuits; later we'll present a more general form that applies even when no circuit is present. The induced emf tends to oppose the change in flux—a crucial point to which we'll devote all of Section 27.3—and so in SI the proportionality between emf and rate of change of flux is -1. Thus Faraday's law is

$$\mathcal{E} = -\frac{d\Phi_B}{dt} \quad \text{(Faraday's law)} \tag{27.2}$$

where $\mathcal{E}$ is the induced emf in a circuit and Φ_B is the magnetic flux through any surface bounded by that circuit.

Faraday's law relates the induced emf to the *change* in flux. It isn't magnetic field or flux that causes an induced emf—it's the *change* in flux. The flux in a uniform field is given by Equation 27.1b, $\Phi_B = \vec{B}\cdot\vec{A} = BA\cos\theta$, which shows that we can change flux by changing the field strength B, the area A, or the angle θ describing the orientation between area and field.

PROBLEM-SOLVING STRATEGY 27.1 Faraday's Law and Induced emf

INTERPRET Make sure the problem involves a circuit in which current flows because of a changing magnetic flux. Identify the circuit and the cause of the changing flux. Possibilities include:

- A changing magnetic field, caused either by relative motion between the circuit and a magnet or by a changing current in an adjacent circuit. Alternatively, the problem may simply state that a magnetic field is changing at some specified rate, without giving the cause.
- A changing area, caused by the circuit expanding or contracting in the presence of a magnetic field.
- A changing orientation of the circuit relative to the field, causing a change in $\cos\theta$.

DEVELOP Find an appropriate expression for the magnetic flux through your circuit. If the field varies with position, you'll have to set up the integral in Equation 27.1a: $\Phi_B = \int \vec{B}\cdot d\vec{A}$; if not, you can use the simpler expression of Equation 27.1b: $\Phi_B = \vec{B}\cdot\vec{A} = BA\cos\theta$. Since the flux is changing, your expression for flux should either have an explicit time dependence or contain a quantity whose rate of change you're given.

EVALUATE Differentiate the flux with respect to time. Faraday's law, $\mathcal{E} = -d\Phi_B/dt$, then gives the induced emf. If you're asked about the circuit current, you can find it using Ohm's law: $I = \mathcal{E}/R$, with R the circuit resistance.

ASSESS Does your answer make sense? Does the induced emf or current increase with an increased rate of whatever quantity is changing? Do the induced effects vanish if you set the rate of change to zero?

EXAMPLE 27.3 Induced Current: A Changing Magnetic Field

A wire loop of radius 10 cm has resistance 2.0 Ω. The plane of the loop is perpendicular to a uniform magnetic field $\vec{B}$ that's increasing at 0.10 T/s. Find the magnitude of the induced current in the loop.

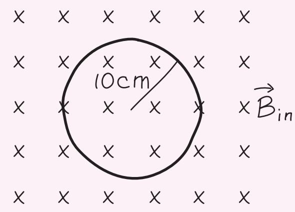


FIGURE 27.8 A circular conducting loop in a plane perpendicular to a uniform magnetic field.

INTERPRET We apply our problem-solving strategy, noting that this is a problem about induction in a circular loop, with the flux change caused by a changing magnetic field.

DEVELOP Figure 27.8 shows the loop with a field pointing into the page. With the field uniform and perpendicular to the loop area, we have $\Phi_B = BA = B\pi r^2$. We're given the rate of change dB/dt, so we can evaluate the derivative $d\Phi_B/dt$.

EVALUATE The rate of change of flux is

$$\frac{d\Phi_B}{dt} = \frac{d}{dt}(B\pi r^2)$$

Since the radius isn't changing, this becomes

$$\frac{d\Phi_B}{dt} = \pi r^2 \frac{dB}{dt}$$

We're given $dB/dt = 0.10$ T/s and $r = 10$ cm. So with $\mathcal{E} = -d\Phi_B/dt$, the magnitude of the induced emf evaluates to 3.14 mV. Then Ohm's law gives the current: $I = \mathcal{E}/R = 3.14\text{ mV}/2.0\ \Omega = 1.6$ mA.

ASSESS Make sense? The induced emf and hence the current scale directly with the value dB/dt, confirming that the changing magnetic field is indeed the cause of the induced effects. Does it bother you that we took $\vec{B}$ as uniform even though it's changing, thus avoiding the integral of Equation 27.1a? The field is indeed changing, but that change is in *time*, not *space*, and the integral for the flux is over space. So at each instant the field is uniform, and we can dispense with the integral.

EXAMPLE 27.4 Induced Current: A Changing Area

Two parallel conducting rails a distance l apart are connected at one end by a resistance R. A conducting bar completes the circuit, joining the two rails electrically but free to slide along them. The whole circuit is perpendicular to a uniform magnetic field $\vec{B}$, as shown in Fig. 27.9. Find the current when the bar is pulled to the right with constant speed v.

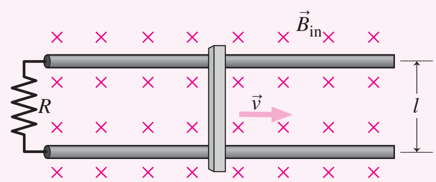


FIGURE 27.9 Pulling the bar to the right increases the circuit area, increasing the magnetic flux and inducing an emf that drives a current.

INTERPRET Here the circuit is formed by the rails, the resistance, and the conducting bar. The circuit area increases as the bar slides along the rails, so we've got a case of induction caused by a changing magnetic flux resulting from a changing area.

DEVELOP In this case of a uniform field perpendicular to the circuit, the flux is the product $\Phi_B = BA$. We can express this flux in terms of the changing position x of the sliding bar; since we're given the bar's speed, we'll be able to evaluate the rate of change of flux. If we take $x = 0$ at the left end of the rails, then the circuit area is $A = lx$, so the flux is $\Phi_B = BA = Blx$.

EVALUATE Differentiating the flux with respect to time gives

$$\frac{d\Phi_B}{dt} = Bl\frac{dx}{dt} = Blv$$

since dx/dt is the bar's velocity v. Faraday's law says that Blv is the magnitude of the induced emf $\mathcal{E}$, so the current in the circuit becomes

$$I = \frac{\mathcal{E}}{R} = \frac{Blv}{R}$$

ASSESS Make sense? Sure: The faster the bar moves, the greater the rate of change of flux, and so the greater the induced emf and current.

✓TIP It's the Change That Matters

You may wonder why, in problems like the preceding two examples, you're not given values for the magnetic field itself or for the location of the sliding bar—quantities that determine the magnetic flux. But the flux itself doesn't matter, only its *rate of change*. And in both cases the rate followed from the given information: in one case the rate of change of the field and in the other the speed of the bar.

Examples 27.3 and 27.4 take care of two ways to change magnetic flux. The third—changing orientation—is at the heart of an important electromagnetic technology, and we'll do an example in the next section.

27.3 Induction and Energy

Move a bar magnet toward a wire loop, as in Fig. 27.10. An induced current flows, dissipating energy as it heats the loop. Where did that energy come from? It came from work you did in moving the magnet.

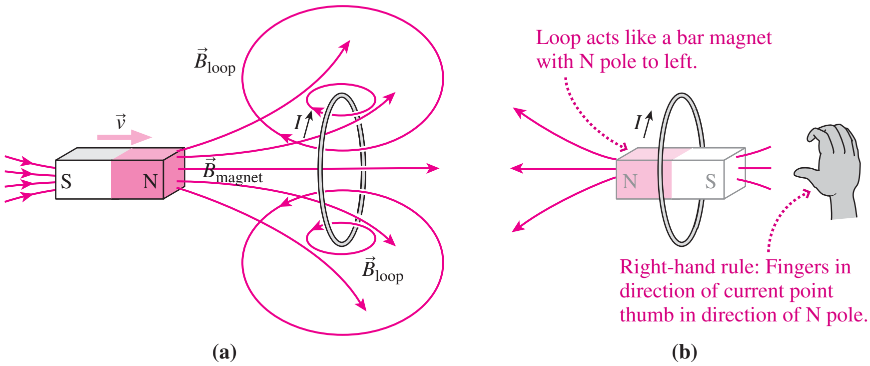


FIGURE 27.10 Conservation of energy determines the direction of the induced current. (a) Fields of bar magnet and loop. (b) The loop acts like a magnet with north pole to the left, making it hard to move the bar magnet.

Normally it doesn't take work to move with constant speed. But the induced current makes the loop a magnetic dipole whose field, as Fig. 27.10 shows, *opposes* the field of the approaching magnet. You have to do positive work to overcome the resulting repulsive force. It had better be this way! Otherwise, you'd get something for nothing, heating the loop without any source of energy.

You can always find the direction of induced emfs and currents by asking: What direction of induced current will make it hard to move the magnet? The answer for Fig. 27.10 is a current that makes the loop a magnet with its north pole on the left, to repel the approaching bar magnet. By the right-hand rule, that gives the current direction shown: into the page at the top of the loop and out at the bottom. If, on the other hand, you move the magnet away from the loop, then the current flows in the opposite direction, putting the loop's south pole on the left and attracting the magnet, making it hard to pull the magnet away (Fig. 27.11).

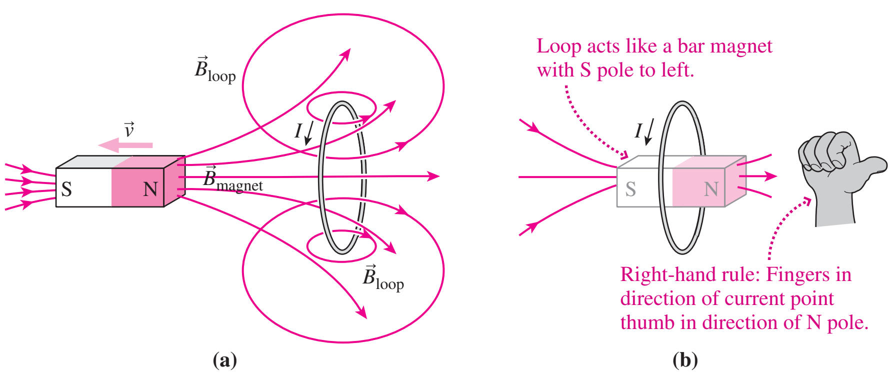


FIGURE 27.11 Now the direction of the induced current puts the loop's south pole to the left, making it hard to pull the magnet away.

This discussion is ultimately about energy conservation in the context of electromagnetic induction. **Lenz's law** summarizes what we've found:

> The direction of an induced emf or current is such that the magnetic field created by the induced current opposes the change in magnetic flux that created the current.

Mathematically, Lenz's law is contained in the minus sign that appears in Faraday's law, but it's usually easier to use Faraday's law to find the magnitude of the induced emf and then reason out the direction using energy conservation.

GOT IT? 27.1 You push a bar magnet toward a loop, with the north pole toward the loop as in Fig. 27.10. If you keep pushing the magnet straight through the loop, what will be the direction of the current as you pull it out the other side? Will you need to do work, or will work be done on you?

Motional emf and Lenz's Law

When a conductor moves through a magnetic field, we can understand the origin of the induced emf in terms of the magnetic force on charge carriers. In this case of **motional emf**, we'll show explicitly that Lenz's law requires energy conservation.

Consider a square conducting loop of side l and resistance R pulled with constant speed v out of a uniform magnetic field $\vec{B}$ (Fig. 27.12). The magnetic flux through the loop is changing, so there's an induced emf that drives a current. Energy is dissipated as heat and so, as we've just argued, the agent pulling the loop must do work. We'll now demonstrate quantitatively that energy is conserved by showing that the rate of heating in the loop is exactly equal to the rate at which the agent pulling the loop does work.

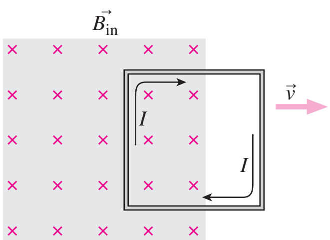


FIGURE 27.12 A conducting loop being pulled out of a magnetic field.

Pulling the loop to the right moves its free electrons through the magnetic field; the magnetic force $q\vec{v} \times \vec{B}$ on these electrons is downward in Fig. 27.12 (opposite $\vec{v} \times \vec{B}$ since electrons are negative). The resulting motion of the negative electrons in the left-hand side of the loop constitutes an upward current, which continues clockwise around the loop.

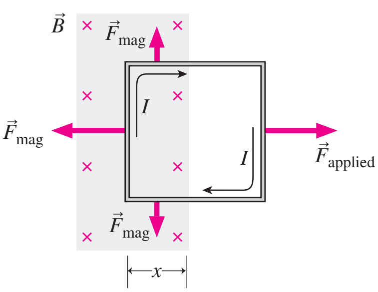


FIGURE 27.13 Forces on the loop.

We found in Chapter 26 that the magnetic force on a current-carrying conductor of length l is $\vec{F} = I\vec{l} \times \vec{B}$. Applying this expression to our conducting loop shows that there's no magnetic force on the right-hand side, where $\vec{B} = \vec{0}$, and that the forces on the top and bottom cancel (Fig. 27.13). So the magnetic force on the loop is that on the left side alone. The magnitude of this force is IlB, and the right-hand rule shows that it points to the left. This leftward magnetic force cancels the rightward applied force, giving the zero net force that Newton's law requires for the loop to move with constant velocity.

We could equally well determine the current direction from magnetic-flux considerations. As the loop leaves the field, the flux decreases. The direction of the induced current is such as to oppose this decrease. Therefore, the magnetic field of the induced current points *into* the page, as the induced current tries to maintain the flux. By the right-hand rule, a field within the loop and into the page requires that the induced current flow clockwise.

To calculate the current, we first find the induced emf. With the field perpendicular to the loop, and uniform in the region where it's nonzero, the magnetic flux is the product of the magnetic-field strength and the loop area that lies within the field: $\Phi_B = Blx$. Here x is the distance between the left edge of the loop and the right edge of the magnetic-field region. The magnetic field remains constant, but as the loop moves, the distance x decreases at the rate $dx/dt = -v$, where the minus indicates a decrease. Then the rate of change of flux is

$$\frac{d\Phi_B}{dt} = \frac{d(Blx)}{dt} = Bl\frac{dx}{dt} = -Blv$$

so Faraday's law gives

$$\mathcal{E} = -\frac{d\Phi_B}{dt} = Blv$$

This induced emf drives a current I around the loop, where $I = \mathcal{E}/R = Blv/R$. The rate of energy dissipation in the loop is the product of the emf and the current (Equation 24.7):

$$P = I\mathcal{E} = \frac{Blv}{R}Blv = \frac{B^2l^2v^2}{R} \qquad \text{(electric power dissipated in loop)}$$

We've found that the magnetic force on the loop has magnitude $F = IlB$; since the loop is moving with constant velocity, this is also the magnitude of the applied force. Equation 6.19 gives $P = \vec{F}\cdot\vec{v}$ for the power supplied. Here, with $\vec{F}$ and $\vec{v}$ in the same direction, we have

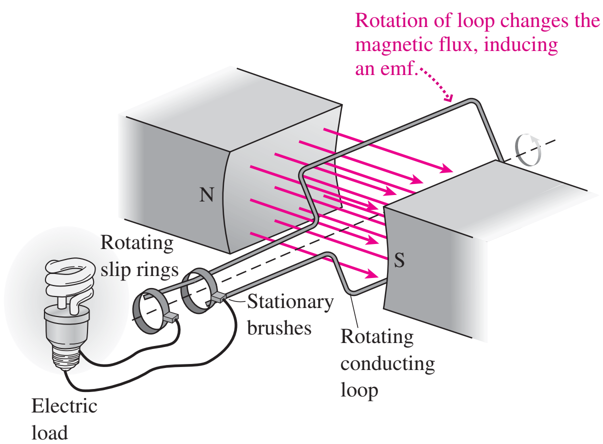
$$P = Fv = IlBv = \frac{Blv}{R}lBv = \frac{B^2l^2v^2}{R} \qquad \text{(mechanical power supplied to pull loop)}$$

the same as our expression for the power dissipated in the loop. Thus, all the work done by the agent pulling the loop ends up heating the resistor, showing explicitly that energy is indeed conserved.

GOT IT? 27.2 What will be the direction of the current when the loop in Fig. 27.12 first enters the field, coming in from the left side?

Electromagnetic induction is the principle behind many important technologies, from credit cards to electric-power generation. Induction also gives us the flexibility to transform voltage levels in electric-power systems, and to provide wireless charging systems for devices ranging from electric cars to toothbrushes.

APPLICATION Electric Generators



Probably the most important technological application of induction is the electric generator. Humanity uses electrical energy at the phenomenal rate of about 2 TW $-$ 2×10^{12} W and roughly equal to the power output of 20 billion human bodies—and virtually all this power comes from generators. A **generator** is just a system of conducting loops in a magnetic field, as shown in the figure. Mechanical energy rotates the conductors, resulting in a changing magnetic flux and therefore an induced emf. Current flows through the generator and on to whatever electrical loads are connected to it. Because the changing flux results from a change in the orientation of the loop relative to the field—that is, a change in θ in the expression $\Phi_B = BA\cos\theta$—a generator such as the one shown here produces an alternating emf that varies sinusoidally with time.

Any source of mechanical energy can power the generator, but the most common is steam from burning fossil fuels or from nuclear fission. Electrical energy is also generated from kinetic energy of water or wind. A small electric generator, driven by the car's engine, is used to recharge a car's battery.

Lenz's law, the conservation of energy in electromagnetic induction, is very much applicable to electric generators. Were it not for Lenz's law, generators would turn on their own and happily supply electricity without coal, oil, or uranium! The voluminous quantities of fuel consumed by power plants are dramatic testimony to the minus sign on the right-hand side of Equation 27.2—as suggested in this chapter's opening photo.

Turn a hand-cranked or pedal-driven generator, and you can literally feel Lenz's law. Without any electrical load, turning the generator is easy. Switch on increasingly heavy loads, and the generator gets harder to turn. Most people find they can just sustain a 100-W lightbulb with a hand generator. Think about this next time you leave a light on!

If the diagram here reminds you of the motor in the Application in Chapter 26, that's no coincidence. Motors and generators are similar devices, just run in opposite ways. A motor converts electrical energy to mechanical energy; a generator converts mechanical energy to electrical energy. Often the same physical device serves both purposes. In a hybrid car, for example, an electric motor takes energy from a battery to provide propulsion. When the car brakes, the wheels turn the motor, which then acts as a generator and puts the car's energy back into the battery instead of dissipating it as heat. Such **regenerative braking** is one of the hybrid's several means of achieving greater energy efficiency.

GOT IT? 27.3 If you lower the electrical resistance connected across a generator while turning the generator at a constant speed, will the generator get easier or harder to turn?

EXAMPLE 27.5 Induction: Designing a Generator

An electric generator consists of a 100-turn circular coil 50 cm in diameter. It's rotated at $f = 60$ rev/s to produce standard 60-Hz alternating current like that used throughout North America. Find the magnetic-field strength needed for a peak output voltage of 170 V (which is the actual peak in standard 120-V household wiring).

INTERPRET Here we have a conducting coil rotating in a fixed magnetic field, so this is an induction problem where changing flux results from a changing orientation.

DEVELOP We sketched the coil and magnetic field in Fig. 27.14. With a uniform field and flat, circular area, the flux through one turn of the coil is given by Equation 27.1b, $\Phi_{1\,\text{turn}} = \vec{B}\cdot\vec{A} = BA\cos\theta = B\pi r^2\cos\theta$. The angle θ changes as the loop rotates, and with it the flux. We need to express the total flux as a function of time so we can evaluate its derivative and thus the emf. Because the loop rotates with constant angular speed $\omega = 2\pi f$, its angular position is $\theta = 2\pi f t$. Then the flux through each turn is $B\pi r^2\cos(2\pi f t)$, and the total flux through all $N = 100$ turns is $NB\pi r^2\cos(2\pi f t)$.

EVALUATE Faraday's law equates the induced emf with the rate of change of this flux:

$$\mathcal{E} = -\frac{d\Phi_B}{dt} = -NB\pi r^2\frac{d}{dt}[\cos(2\pi f t)] = -NB\pi r^2[-2\pi f\sin(2\pi f t)]$$

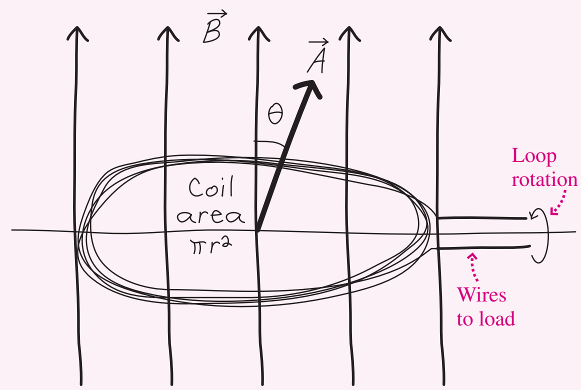


FIGURE 27.14 Coil in the generator of Example 27.5; at this instant the normal to the coil makes an angle θ with the magnetic field.

The emf has its peak value when the sine is 1, so $\mathcal{E}_{\text{peak}} = 2\pi^2 r^2 NBf$. We want this value to be 170 V; using $r = 25$ cm, $N = 100$ turns, and $f = 60$ rev/s then gives $B = 23$ mT.

ASSESS This value is about 200 G, typical of the field strength near the poles of a permanent magnet. Note that you don't need a value for time t to find the peak emf; when a quantity varies sinusoidally, its peak occurs when the sine or cosine function is 1, so the peak value is the magnitude of whatever quantity multiplies the sine or cosine. ■

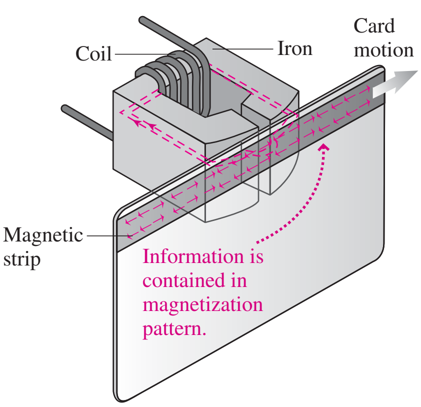


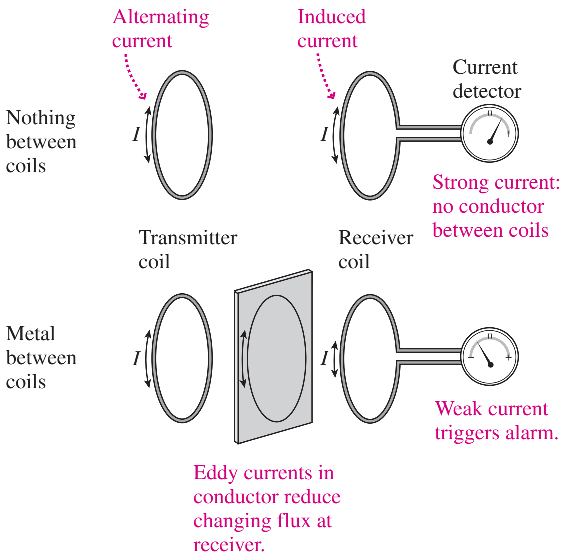
FIGURE 27.15 Swiping a credit card. Patterns of magnetization on the strip induce currents in the coil.

Electromagnetic induction is also the basis of magnetic recording, once the dominant means of storing audio, video, and computer information but now more common in credit cards and similar applications. The magnetic strip on your credit card is a ferromagnetic material that stores information in regions of differing magnetization. Swiping your card induces current in a wire coil, which extracts the stored information as an electrical signal (Fig. 27.15). Early computer disks worked on the same principle, although in today's disks the magnetic field of the spinning disk causes changes in electrical resistance in the "head" that reads the disk information.

Eddy Currents

Induced currents aren't limited to conducting loops and circuits. They also occur in solid conductors subject to changing magnetic flux. The resistance of a solid conductor is low, which can result in large induced currents and significant power dissipation. That can make it hard to move a conducting material into or out of a magnetic field, as it's subject to a changing flux. The result is a kind of magnetic friction that saps energy. On the other hand, the effect can be useful in providing an alternative to friction brakes. Rapidly rotating saw blades or train wheels, for example, can be stopped quickly by turning on a nearby electromagnet; the resulting eddy currents dissipate the rotational kinetic energy. The mechanical resistance you feel in exercise machines like elliptical trainers or stationary cycles results from a magnet positioned near the machine's rotating parts. And eddy currents are guardians of our security, as the next Application shows.

APPLICATION **Metal Detectors**



Metal detectors used in airports and other security checkpoints rely on eddy currents. In one type of detector, shown in the figure, an alternating current in one coil—the transmitter—produces a changing magnetic field that induces a current in a second coil, the receiver. A detector, basically an electronic ammeter, monitors the receiver current. Eddy currents are induced in any conducting material that comes between the two coils, and the direction of the induced currents is, as always, such as to reduce the changing flux. The superposition of the transmitter's changing flux with the changing flux from the eddy currents therefore reduces the changing flux at the receiver, dropping the receiver current and triggering an alarm. Other detectors have a single coil, using a short pulse of current to induce eddy currents and then "listening" for the currents induced back in the coil. Either way, you can thank Faraday's law if you've ever been stopped while going through a metal detector!

GOT IT? 27.4 A copper penny falls on a path that takes it between the poles of a magnet. Does it hit the ground going (a) faster than, (b) slower than, or (c) at the same speed as if the magnet weren't present?

Closed and Open Circuits

Figure 27.16 shows a closed, conducting loop in a magnetic field that points into the page. Suppose the field is increasing in strength; then in order to oppose this change, the induced current must be in such a direction as to oppose the increase. Here that means the field in the interior of the loop needs to come out of the page, and by the right-hand rule that means the induced current is counterclockwise. It's not that the induced field always opposes the inducing field—rather, it opposes the *change*. If the field in Fig. 27.16 had been decreasing, then the induced current would have "tried" to reinforce it by flowing clockwise to make additional field into the page.

What if we have an open circuit, like the conducting loop with a small gap shown in Fig. 27.17? Then there's no induced current whose effects can oppose a change in the field. But we can imagine what would happen if the circuit were completed; as in Fig. 27.16, current would flow counterclockwise. Open the gap, and that means positive charge accumulates at the upper end of the gap and negative charge at the bottom. Charge buildup continues until the potential difference at the gap opposes the induced emf's tendency to move charge. The result is a steady state in which the gap voltage equals the induced emf.

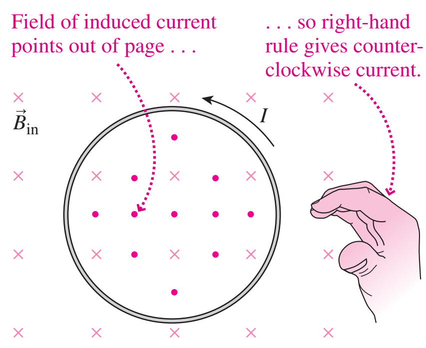


FIGURE 27.16 The field $\vec{B}$ is into the page and increasing; the induced current is counterclockwise, so its field opposes the increase.

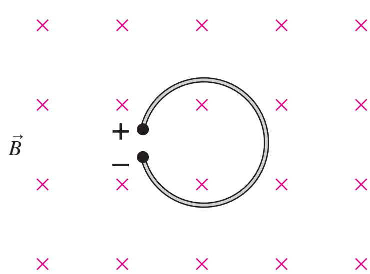
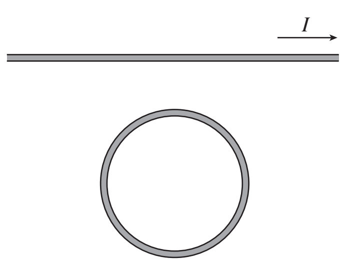


FIGURE 27.17 In a changing magnetic field, the induced emf results in charge buildup at the gap of an open circuit. The polarity shown results when the field is increasing.

GOT IT? 27.5 A long wire carries a current I as shown. What's the direction of the current in the circular conducting loop when I is (a) increasing and (b) decreasing?



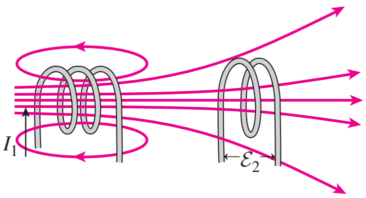


FIGURE 27.18 Mutual inductance. A changing current in either coil induces an emf in the other coil.

27.4 Inductance

There are many ways to change magnetic flux and thus induce emfs and currents. We can move a magnet, or move or rotate the circuit. Or, as in Fig. 27.4 or Fig. 27.18, we can change the magnetic flux by changing the current in a circuit and therefore the magnetic field it produces. In that case we speak of the **inductance** of a circuit or circuits.

Mutual Inductance

Figure 27.18 shows two coils in proximity. If we send a changing current through the left-hand coil, the resulting magnetic field gives rise to a changing magnetic flux in the right-hand coil. There's then an induced emf in the right-hand coil and, if it's connected in a complete circuit, an induced current as well.

The two coils in Fig. 27.18 have **mutual inductance**, meaning a changing current in one coil produces a changing flux at the other coil, thereby inducing an emf. Just how strong this effect is depends on the construction and orientation of the coils; for maximum inductance they should be arranged so that most of the flux from each coil goes through the other. Often coils are wound on iron cores to concentrate the flux and increase the mutual inductance.

Mutual inductance is the basis of transformers, which change voltage levels in alternating-current circuits; more on that in Chapter 28. Your car's ignition coil uses mutual inductance to produce the tens of kilovolts needed to fire the spark plugs and ignite the gasoline–air mixture in the engine. Current to the coil is interrupted at just the right instant, producing a rapid change in magnetic flux and inducing the emf that drives the spark. More mundane is an electric toothbrush, whose batteries charge without electrical connection to a power source. Instead, a small coil in the device is placed in proximity with a coil in the charging base, and alternating current in the base coil transfers energy via mutual inductance to provide the charging current.

Self-Inductance

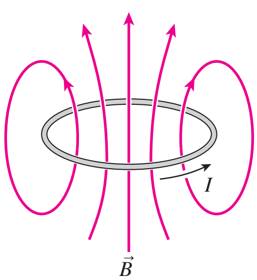


FIGURE 27.19 Magnetic flux from a current loop passes through the loop itself; a change in the current induces an emf that opposes the change.

Inductance isn't limited to two-coil systems. Magnetic flux from current in a single coil or circuit passes through that circuit itself (Fig. 27.19). If the current changes, so does the flux—and that induces an emf. As always, the induced emf opposes the *change* that produces it. Suppose, for example, that the current in Fig. 27.19 is increasing. Then the induced emf will be in the direction that opposes the current increase—clockwise, or opposite the current in Fig. 27.19. The induced emf therefore makes it harder to increase the current. On the other hand, if the current in Fig. 27.19 is decreasing, then the induced emf will try to drive additional current to counter the decrease; the induced emf is therefore in the same direction as the current. Either way, the induced emf makes it hard to change the current in a circuit.

This property whereby a circuit's own magnetic field opposes changes in the circuit current is called **self-inductance**. All circuits have self-inductance, but it's most important in circuits that encircle a great deal of their own magnetic flux, or when currents change rapidly. A simple piece of wire has little impact on the 60-Hz alternating current used for electric power. But in TVs and computers, where currents change billions of times per second, even the slightest self-inductance can have deleterious effects.

An **inductor** is designed specifically to exhibit self-inductance. Inductors have many uses in electric circuits, including establishing the frequencies of radio transmitters and helping "steer" high- and low-frequency signals to the tweeters and woofers of loudspeaker systems. We'll explore some of these uses in the next chapter. A typical inductor consists of a wire coil, sometimes wound on an iron core to promote flux concentration. Ideally, the only electrical property of an inductor is its inductance, but real inductors have resistance as well.

As long as the current in an inductor is steady, the magnetic flux is constant, so there's no induced emf and the inductor acts like a wire. But when the current changes, the changing magnetic flux induces an emf that opposes the change in current. The more rapidly the current changes, the greater the rate of change of flux and so the greater the emf. The induced emf depends also on how much of its own magnetic flux the inductor encircles; consequently, we define self-inductance, L, as the ratio of magnetic flux through the inductor to current in the inductor:

$$L = \frac{\Phi_B}{I} \quad \text{(self-inductance)} \tag{27.3}$$

Equation 27.3 shows that the units of self-inductance are T·m^2/A. This unit is given the name **henry** (H) in honor of the American scientist Joseph Henry (1797–1878). Inductances in common electronic circuits usually range from microhenrys (μH) up to several henrys.

Inductance is a constant determined by the physical design of an inductor. In principle we can calculate the inductance of any inductor, but in practice that's difficult unless the geometry is particularly simple.

EXAMPLE 27.6 Calculating Inductance: A Solenoid

A long solenoid of cross-sectional area A and length l has n turns per unit length. Find its self-inductance.

INTERPRET We're asked for self-inductance, which Equation 27.3 shows is the ratio of magnetic flux through the solenoid to current in the solenoid.

DEVELOP We'll assume a current in the solenoid and find the resulting magnetic flux. Then we can take their ratio to get the self-inductance. We need the magnetic field of a solenoid, which follows from Equation 26.20: $B = \mu_0 nI$. The field is uniform and perpendicular to the solenoid coils, as we showed in our drawing for Example 27.1, so the flux through each turn follows from Equation 27.1b: $\Phi_{1\text{ turn}} = BA$.

EVALUATE With n turns per unit length, the solenoid contains a total of nl turns, so the flux through all the turns is

$$\Phi_B = nlBA = (nl)(\mu_0 nI)A = \mu_0 n^2 AlI$$

Equation 27.3 gives the self-inductance as the ratio of flux to current, so

$$L = \frac{\Phi_B}{I} = \mu_0 n^2 Al \quad \text{(inductance of solenoid)} \tag{27.4}$$

ASSESS Make sense? As the area increases, so does the flux and therefore the inductance. As the length increases, so does the number of turns, so again the total flux increases. And as the number of turns per unit length increases, two things happen. First, the magnetic field of Equation 26.20 increases, increasing the flux BA through each turn. Second, the total number of turns increases, again increasing the total flux. That's why the inductance, L, depends on *n squared*. ■

The induced emf in an inductor is determined by Faraday's law, which relates emf to the rate of change of magnetic flux: $\mathcal{E} = -d\Phi_B/dt$. Differentiating Equation 27.3, the definition of inductance, gives

$$\frac{d\Phi_B}{dt} = L\frac{dI}{dt}$$

Then Faraday's law becomes

$$\mathcal{E}_L = -L\frac{dI}{dt} \quad \text{(inductor emf)} \tag{27.5}$$

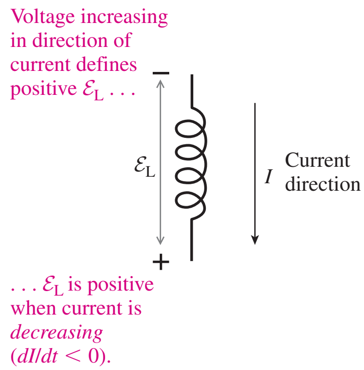
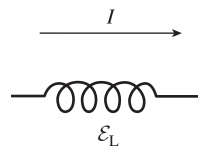


FIGURE 27.20 The direction of induced emf in Equation 27.5 depends on whether the current is increasing or decreasing. The coil is the circuit symbol for an inductor.

This equation gives the emf $\mathcal{E}$ induced in an inductor L when the inductor current is changing at the rate dI/dt. The minus sign again tells us that the emf *opposes* the change in current. For this reason the inductor emf is often called a **back emf**; it works *against* changes brought about by an externally applied emf. Figure 27.20 shows how to interpret the sign in Equation 27.5.

When the current in an inductor is steady, $dI/dt = 0$ and there's no induced emf. In this case, the inductor acts like a piece of wire. But when the current changes, the inductor responds by producing a back emf that opposes the change. Now the inductor acts much like a battery, with the magnitude of its emf dependent on how fast the current changes. If we try to start or stop current suddenly, dI/dt is very large and a very large back emf appears. This isn't just mathematics! Rapid switching of inductive devices such as solenoids, solenoid valves, or motors can destroy delicate electronic devices. And people have been killed opening switches in circuits containing large inductors. In the next section we'll take a closer look at circuits that include inductors.

GOT IT? 27.6 Current flows from left to right through the inductor shown. A voltmeter connected across the inductor gives a constant reading, and shows that the left end of the inductor is positive. Is the current in the inductor (a) increasing, (b) decreasing, or (c) steady? Why?



EXAMPLE 27.7 Back emf: A Dangerous Inductor

A 5.0-A current is flowing in a 2.0-H inductor. The current is then reduced steadily to zero over 1.0 ms. Find the magnitude and direction of the inductor emf during this time.

INTERPRET There's an emf in the inductor because the current is changing; we want the magnitude and direction of that emf.

DEVELOP Figure 27.21 shows the situation, complete with an external circuit that's the source of the decreasing current. Equation 27.5, $\mathcal{E}_L = -L(dI/dt)$, gives the inductor emf in terms of the rate of change of current. Since the current changes steadily, the latter is just the change in current (-5.0 A) divided by the time involved.

EVALUATE

$$\mathcal{E}_L = -L\frac{dI}{dt} = -(2.0\text{ H})\left(\frac{-5.0\text{ A}}{1.0\text{ ms}}\right) = 10{,}000\text{ V}$$

That this answer is positive tells us that the emf increases in the same direction as the current, as we indicated in Fig. 27.21.

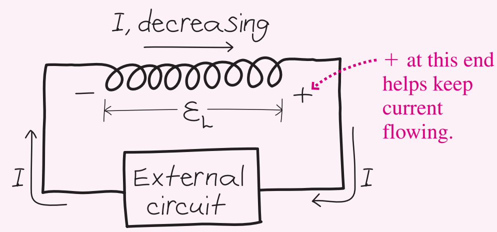


FIGURE 27.21 Sketch for Example 27.7.

ASSESS That's a potentially lethal voltage! Our answer is unrelated to the battery or whatever is supplying the inductor current. One could have a 6-V battery and still be electrocuted opening a circuit with a large inductance. Note that the direction we deduced is consistent with Lenz's law; here the inductor emf opposes the decrease in current, and that means it provides an emf in the direction that would keep current flowing in the external circuit—just the opposite of the situation in Got It? 27.6. ■

Inductors in Circuits

In Chapter 25 we found that the voltage across a capacitor can't change instantaneously. We can make an analogous statement for inductors. Because the inductor emf depends on the rate of change of current and because an infinite emf is impossible, **the current through an inductor can't change instantaneously**. Much of our understanding of capacitors applies to inductors if we interchange the words "voltage" and "current."

Figure 27.22 shows a circuit with a battery, switch, resistor, and inductor. With the switch open there's no current (Fig. 27.22*a*). Close the switch, and the current at that instant is still zero because the inductor current can't change instantaneously. With no current, there's no voltage across the resistor, so the inductor must be producing a back emf equal in magnitude to the battery emf (Fig. 27.22*b*). Although at this instant there's zero current in the inductor, the nonzero emf $\mathcal{E}_L = -L(dI/dt)$ shows that the *rate of change* of current, dI/dt, isn't zero.

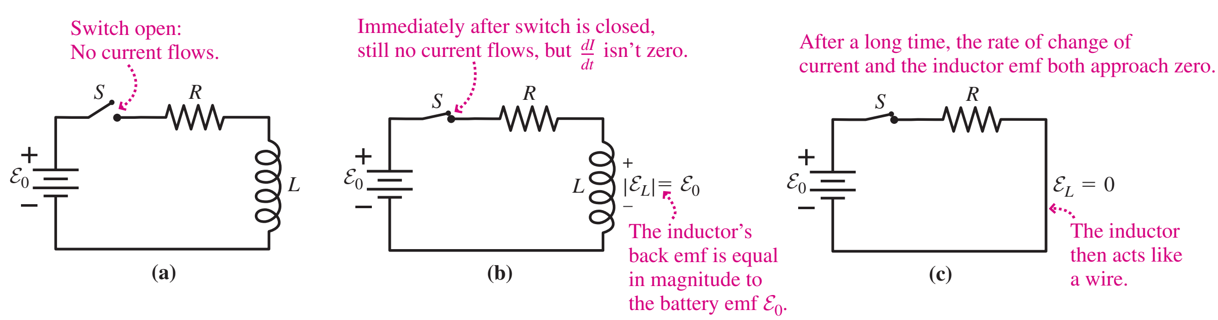


FIGURE 27.22 An *RL* circuit at three times.

So the inductor current rises from zero, and with it the resistor current and therefore the resistor voltage IR. The battery emf $\mathcal{E}_0$ is constant, so as IR increases, the magnitude of the inductor emf drops. Equation 27.5 shows that the rate of change of current drops as well. Eventually the whole circuit reaches a steady state in which dI/dt and therefore the inductor emf are both zero (Fig. 27.22*c*). At this point the inductor acts like a wire, and the resistor determines the current: $I = \mathcal{E}_0/R$. Figure 27.23 summarizes this analysis of the RL circuit.

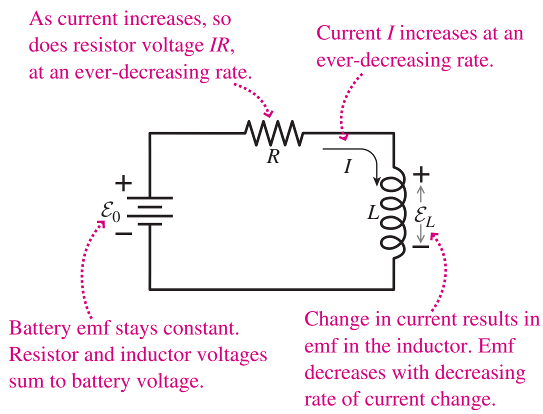


FIGURE 27.23 Interrelationships among circuit quantities as current builds up in an *RL* circuit. Compare with Fig. 25.19 for a charging capacitor.

We can analyze the circuit quantitatively using the loop law. Going clockwise, we encounter a voltage increase $\mathcal{E}_0$ at the battery, a decrease $-IR$ at the resistor, and a change $\mathcal{E}_L$ at the inductor. This change is actually a decrease, but we'll let Equation 27.5 take care of the signs. Then the loop law reads $\mathcal{E}_0 - IR + \mathcal{E}_L = 0$. The battery emf is constant, so if we differentiate this equation, we get

$$\frac{d\mathcal{E}_L}{dt} = R\frac{dI}{dt}$$

But Equation 27.5 gives $dI/dt = -\mathcal{E}_L/L$, so

$$\frac{d\mathcal{E}_L}{dt} = -R\frac{\mathcal{E}_L}{L}$$

This looks like Equation 25.4 for the RC circuit, but with $\mathcal{E}_L$ in place of current I, L in place of C, and $1/R$ in place of R. So the solution is that of Equation 25.4 with the appropriate substitutions:

$$\mathcal{E}_L = -\mathcal{E}_0 e^{-Rt/L} \qquad (27.6)$$

This shows that the inductor emf decays exponentially from its initial value $-\mathcal{E}_0$ (negative because the inductor emf opposes the battery) to zero. Using the undifferentiated loop equation, we can now solve for the current:

$$I = \frac{\mathcal{E}_0 + \mathcal{E}_L}{R} = \frac{\mathcal{E}_0}{R}\left(1 - e^{-Rt/L}\right) \qquad (27.7)$$

With a capacitor, we characterized time-changing quantities with the capacitive time constant RC. Here the **inductive time constant** is L/R. In contrast to the capacitor case, the inductive time constant depends *inversely* on resistance. That's because a lower resistance means a higher steady-state current, which therefore requires a longer time to build up. Significant changes in current can't occur on time scales much shorter than L/R. Wait many time constants, and the circuit approaches a steady state with $\mathcal{E}_L = 0$. Figure 27.24 summarizes the time-dependent behavior of circuit quantities in an RL circuit.

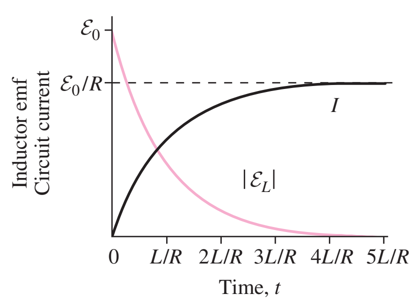


FIGURE 27.24 Inductor current and emf as functions of time.

EXAMPLE 27.8 The Inductive Time Constant: Firing Up an Electromagnet

A large electromagnet used for lifting scrap iron has self-inductance $L = 56$ H. It's connected to a constant 440-V power source; the total resistance of the circuit is 2.8 Ω. Find the time it takes for the current to reach 75% of its final value.

INTERPRET This is a problem about the buildup of current in an *RL* circuit.

DEVELOP Equation 27.7, $I = (\mathcal{E}_0/R)(1 - e^{-Rt/L})$, determines the current; here $\mathcal{E}_0/R$ is the final current, and we want to solve for the time t when I is 75% of this final value. That is, we want $0.75 = 1 - e^{-Rt/L}$.

EVALUATE Rearranging, we have $e^{-Rt/L} = 0.25$; then taking natural logs of both sides and using $\ln e^x = x$ gives $-Rt/L = \ln(0.25)$, or

$$t = -\frac{L}{R}\ln(0.25) = -\frac{56\ \text{H}}{2.8\ \Omega}\ln(0.25) = 28\ \text{s}$$

ASSESS This is a little more than 1 time constant $(L/R = 20\ \text{s})$—not surprising because we found with capacitors that we reach approximately two-thirds of the full charge in 1 time constant. Analogously, with inductors, we reach about two-thirds of the final current in 1 time constant.

Figure 27.25 shows a circuit with a two-way switch. Throw the switch to *A*, and current builds up as we just described. Throw it to *B*, and current continues to flow through the inductor and resistor because the inductor current can't change instantaneously. We won't go through the math, but it's straightforward to show that the current decays exponentially with the same time constant L/R:

$$I = I_0 e^{-Rt/L} \qquad (27.8)$$

This is analogous to our result for the discharging capacitor.

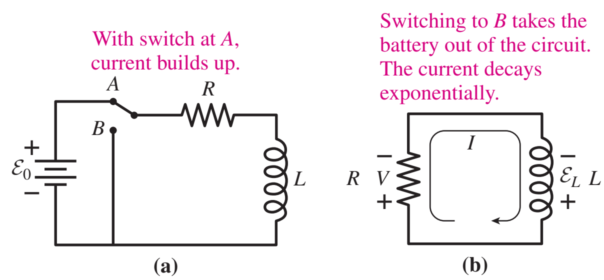


FIGURE 27.25 Buildup and decay of current in an *RL* circuit.

As with capacitors, it's not necessary to use exponential equations to analyze the short- and long-term behavior of circuits with inductors. All you need to remember is that for short times inductor current can't change instantaneously, and for long times inductors have no emfs and therefore act like wires. The next example explores this situation.

CONCEPTUAL EXAMPLE 27.1 Inductors: Short Times, Long Times

The switch in Fig. 27.26*a* is initially open. It's then closed and, a long time later, reopened. What's the direction of the current in R_2 after the switch is reopened?

EVALUATE To see what's happening here, we sketch the circuit in three situations, beginning with the switch closing and ending with it reopening (Fig. 27.26*b*–*d*). There's no inductor current with the switch initially open, so there's no current right after it closes. Then the inductor might as well be an open circuit, so we drew Fig. 27.26*b* with only the two resistors. After a long time the current stops changing, and the inductor behaves like a wire (Fig. 27.26*c*). Finally, whatever current was flowing in the inductor continues to flow after the switch is reopened. That current was flowing downward in the inductor, so, as Fig. 27.26*d* shows, it's flowing *upward* through R_2.

ASSESS Does this surprising result make sense? Yes: Current in an inductor can't change instantaneously, and once the switch opens the current has nowhere to go but upward through R_2. The resistor has no say in the matter; as Making the Connection shows, its current is determined entirely by the battery voltage and R_1. If the switch stays open, the current in Fig. 27.26*d* decays exponentially as the resistor dissipates the energy that was stored in the inductor.

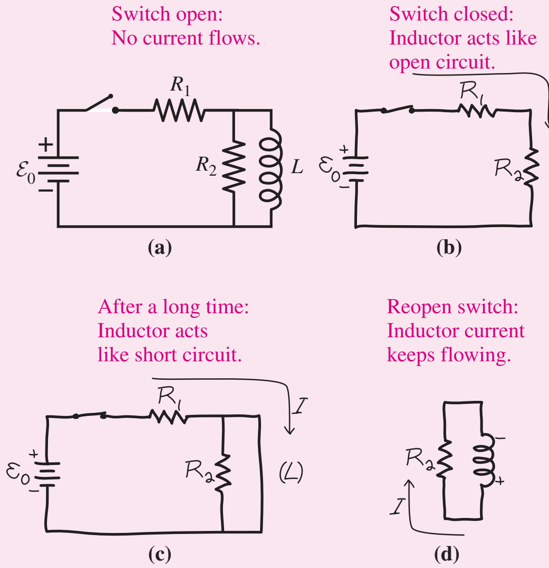


FIGURE 27.26 Conceptual Example 27.1

Figure 27.27 shows currents in the inductor and R_2 as functions of time. If R_2 weren't in the circuit, the voltage would rise dangerously high as the inductor tries to keep the current flowing. Resistors are often wired in parallel with large inductors to alleviate this danger.

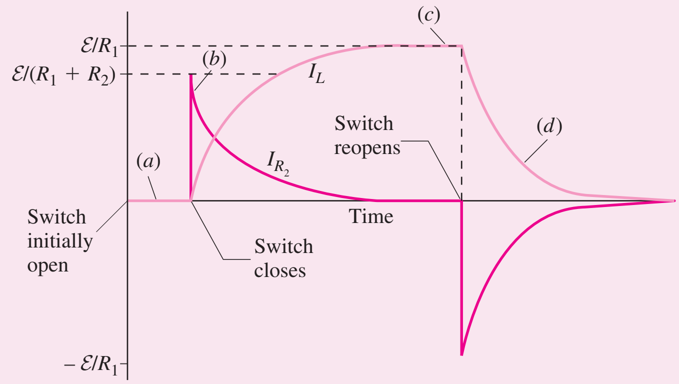


FIGURE 27.27 Currents in R_2 and L for Conceptual Example 27.1.

MAKING THE CONNECTION Verify that the current in R_2 just after the switch is reopened has the value indicated in Fig. 27.27.

EVALUATE Just *before* the switch is reopened, Fig. 27.26*c* shows that the current through the inductor is $I_L = \mathcal{E}_0/R_1$; R_2 is irrelevant here because it's short-circuited by the inductor. Just *after* the switch opens, the current continues flowing, now going upward through R_2 as we reasoned above. So the current in R_2 is $-\mathcal{E}_0/R_1$, with the minus sign designating the upward direction according to the sign conventions in Fig. 27.27.

27.5 Magnetic Energy

In Figs. 27.25*b* and 27.26*d*, current flows in circuits containing only a resistor and an inductor. Energy is dissipated, heating the resistor. Where does this energy come from?

Because there's a current in the inductor, there's also a magnetic field. The change in that magnetic field is what produces the emf that drives the current. As the current decreases, so does the magnetic field. Eventually the circuit reaches a state where there's no current, no magnetic field—and a hot resistor. So where did the resistor's thermal energy come from? It came from the magnetic field.

Like the electric field, the magnetic field contains stored energy. Our decaying *RL* circuit is analogous to a discharging *RC* circuit, in which the electric field between the capacitor plates disappears as thermal energy appears in the resistor. As in the electric case, magnetic energy isn't limited to circuits: *Any* magnetic field contains energy. Release of magnetic energy drives a number of practical devices and also powers violent events throughout the universe (Fig. 27.28).

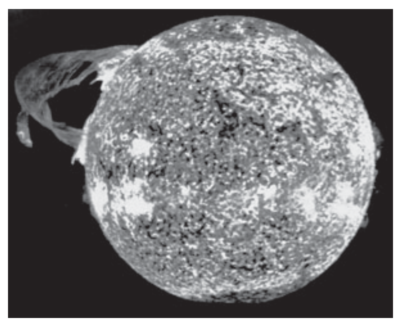


FIGURE 27.28 This eruption of a huge prominence from the Sun's surface releases energy stored in magnetic fields.

Magnetic Energy in an Inductor

We can find the stored energy by reconsidering the buildup of current in the inductor. Earlier we wrote the loop law for the circuit of Fig. 27.22; if we multiply that equation by the current I, we get $I\mathcal{E}_0 - I^2R + I\mathcal{E}_L = 0$ or, using Equation 27.5 for $\mathcal{E}_L$,

$$I\mathcal{E}_0 - I^2R - LI\frac{dI}{dt} = 0$$

The three terms here have the units of voltage times current, or power. The first shows that the battery supplies energy *to* the circuit at the rate $I\mathcal{E}_0$. The second, $-I^2R$, is the rate of energy dissipation in the resistor; the negative sign means the resistor takes energy *from* the circuit. The current is increasing $(dI/dt > 0)$, so the third term is also negative; it describes energy the inductor takes from the circuit. But the inductor doesn't dissipate this energy; rather, it stores the energy in its magnetic field. The rate at which the inductor stores energy is thus

$$P = LI\frac{dI}{dt}$$

Suppose we increase the current in an inductor by some small amount dI over a small time interval dt. Since the power is the rate of energy storage, the energy dU stored during this time is $dU = P\,dt = LI\,dI$. We find the total energy stored in bringing the inductor current from zero to some final value I by summing—that is, integrating—all the dU values:

$$U = \int dU = \int P\,dt = \int_0^I LI\,dI = \tfrac{1}{2}LI^2\Big|_0^I$$

Evaluating at the two limits then gives the stored energy:

$$U = \tfrac{1}{2}LI^2 \qquad \text{(energy stored in inductor)} \qquad (27.9)$$

This much energy is therefore released when the magnetic field decays.

EXAMPLE 27.9 Magnetic Energy: An MRI Disaster

Superconducting electromagnets like the solenoids in MRI scanners store a lot of magnetic energy. Loss of coolant can be dangerous because the current is suddenly left without its zero-resistance path and quickly decays. The result is an explosive release of magnetic energy. A particular MRI solenoid carries 2.4 kA and has a 0.53-H inductance. When it loses superconductivity, its resistance goes abruptly to 31 mΩ. Find (a) the stored magnetic energy and (b) the rate of energy release at the instant superconductivity is lost.

INTERPRET We're asked first for the total stored energy and then for the power dissipated in the resistance of the coils at the instant they cease to be superconducting.

DEVELOP Equation 27.9, $U = \frac{1}{2}LI^2$, determines the stored energy, while $P = I^2R$ determines the resistor power. Just before the coolant loss, the MRI solenoid carries 2.4 kA; since current can't change instantaneously in an inductor, this current remains momentarily unchanged. Therefore, we have everything we need to find the power dissipation.

EVALUATE (a) Equation 27.8 gives

$$U = \tfrac{1}{2}LI^2 = \left(\tfrac{1}{2}\right)(0.53\text{ H})(2.4\text{ kA})^2 = 1.5\text{ MJ}$$

while for (b) we have

$$P = I^2R = (2.4\text{ kA})^2(31\text{ m}\Omega) = 0.18\text{ MW}$$

ASSESS This is substantial power, equivalent to 1800 100-W lightbulbs burning in the space of this roughly human-size device! You can show in Problem 55 that it takes some 20 s before 90% of the energy has dissipated. To prevent explosive energy release, superconducting wires generally incorporate copper or silver to carry the current in the event of coolant loss that quenches the superconductivity. ■

Magnetic-Energy Density

In Example 27.6 we found the inductance of a solenoid with length l and cross-sectional area A: $L = \mu_0 n^2 Al$. Equation 27.9 then gives the magnetic energy stored in the solenoid:

$$U = \tfrac{1}{2}LI^2 = \tfrac{1}{2}\mu_0 n^2 AlI^2 = \frac{1}{2\mu_0}(\mu_0 nI)^2 Al = \frac{B^2}{2\mu_0}Al$$

where we recognized the quantity $\mu_0 nI$ as B, the magnetic field in the solenoid (Equation 26.20). The quantity Al is the volume containing this field, so the energy per unit volume—the **magnetic-energy density**—is

$$u_B = \frac{B^2}{2\mu_0} \qquad \text{(magnetic-energy density)} \qquad (27.10)$$

Although we derived this expression for the field of a solenoid, it is, in fact, a universal expression for the local magnetic-energy density. Wherever there's a magnetic field, there's stored energy.

Equation 27.10 is similar to Equation 23.7 for the energy density in an electric field: $u_E = \frac{1}{2}\epsilon_0 E^2$. Each energy density is proportional to the *square* of the field strength, and each contains the appropriate constant, μ_0 or ϵ_0. That the constant appears in the numerator in one case and in the denominator in the other is merely a consequence of the way SI units are defined.

27.6 Induced Electric Fields

So far we've been talking about induction in terms of emfs and circuits. But what, really, is emf? In the case of a battery, it results from chemical reactions that separate charge. With motional emf (Section 27.3), magnetic forces on a moving conductor act to separate charge. But what causes the emf in a conducting loop subject to a changing magnetic field? There's no motion, yet there must be a force on the free charges in the conductor. The only force we know that acts on stationary charges is the electric force, which results from electric fields. Therefore, there must be an electric field—an **induced electric field**—in the conducting loop. This field has the same effect on charges, exerting a force $q\vec{E}$, as did the electric fields we considered earlier. But the induced field originates not in electric charge but in **changing magnetic field**.

An induced electric field results whenever a magnetic field changes with time—whether or not an electric circuit is present. If there is a circuit, then the field drives induced currents. But the induced field, not the current, is fundamental. A single, stationary electron in a changing magnetic field experiences an *electric* force—clear evidence for the existence of the induced electric field.

We wrote Faraday's law as Equation 27.2, giving the relation between induced emf and changing magnetic flux. But the induced electric field is more fundamental, and emf simply means the work per unit charge gained as charge goes around a circuit—or for that matter any closed loop. Thus we can write $\mathcal{E} = \oint \vec{E} \cdot d\vec{r}$, and Faraday's law becomes

$$\oint \vec{E} \cdot d\vec{r} = -\frac{d\Phi_B}{dt} \qquad \text{(Faraday's law)} \tag{27.11}$$

In this form, Faraday's law is a universal statement about electric fields and changing magnetic flux. The line integral on the left-hand side is over *any* closed loop, which need not coincide with a circuit or conductor. The flux on the right-hand side is the surface integral of the magnetic field over any open surface bounded by the loop on the left-hand side.

Faraday's law tells us that there's another source of electric fields besides electric charge—namely, changing magnetic field:

> A changing magnetic field creates an electric field.

This direct interaction between fields is the basis for many practical devices and, as we'll see in Chapter 29, is essential to the existence of light.

Faraday's law is similar to Ampère's law (Equation 26.16). On the left side, both involve the line integral of a field, $\vec{E}$ for Faraday and $\vec{B}$ for Ampère. On the right is a source of that field, changing magnetic field for $\vec{E}$ and moving electric charge—current—for $\vec{B}$. Both fields *encircle* their sources. That means the configuration of an induced electric field is very different from that of an electric field originating in charge. Field lines of an induced electric field have no beginnings or ends; they generally form closed loops encircling regions of changing magnetic field (Fig. 27.29).

When a changing magnetic field has sufficient symmetry, we can evaluate the induced electric field in the same way we did the magnetic field of a symmetric current distribution.

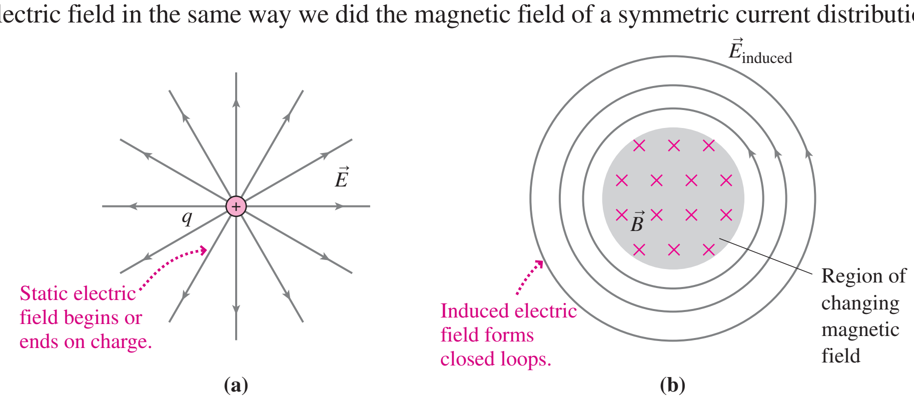


FIGURE 27.29 (a) Static electric fields originate in charges and look very different from (b) induced fields that result from changing magnetic fields.

EXAMPLE 27.10 Finding the Induced Electric Field: A Solenoid

A long solenoid has circular cross section of radius R. The solenoid current is increasing, and as a result so is the magnetic field in the solenoid. The field strength is given by $B = bt$, where b is a constant. Find the induced electric field outside the solenoid, a distance r from the axis.

INTERPRET Here's a problem about a changing magnetic field producing an electric field—that is, about Faraday's law. We'll follow the steps in Chapter 26's strategy for Ampère's law, modifying as appropriate to Faraday's law. We begin by identifying the symmetry, which here is line symmetry.

DEVELOP Symmetry requires that encircling electric field lines be circular. We've drawn some field lines in Fig. 27.30 and marked one of them as a loop for the integration in Faraday's law. We chose a loop coinciding with a field line because symmetry requires that the field strength E be constant over a circle concentric with the symmetry axis.

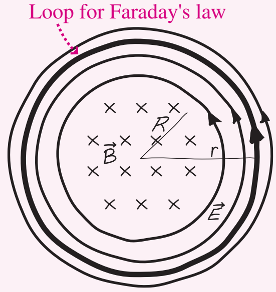


FIGURE 27.30 Cross section of a solenoid whose magnetic field points into the page and is increasing. Field lines of the induced electric field are circles concentric with the solenoid axis.

EVALUATE The situation on the left-hand side of Faraday's law is just as in Example 26.8, except with $\vec{E}$ instead of $\vec{B}$. So $\oint \vec{E} \cdot d\vec{r}$ evaluates immediately to $2\pi rE$. Instead of current on the right-hand side, we have changing magnetic flux $-d\Phi_B/dt$. Here the loop is outside the solenoid, so it encircles the entire flux, which is $\Phi_B = BA = bt\pi R^2$. The right-hand side of Faraday's law has the rate of change of flux, which is $d\Phi_B/dt = \pi R^2 b$. Equating the left and right sides of Faraday's law then gives $2\pi rE = -\pi R^2 b$, or

$$E = -\frac{R^2 b}{2r}$$

ASSESS The $1/r$ dependence here shouldn't surprise you; we found the same thing for other fields resulting from distributions with line symmetry. The minus sign tells about the direction, but as usual it's easiest to invoke Lenz's law to reason out the direction. If a current were flowing as a result of the induced electric field, its direction would be such as to *oppose* the increase in the solenoid's magnetic field. So it would have to produce a field out of the page in Fig. 27.30, and that means a counterclockwise current. So the induced electric field goes counterclockwise. Calculating the field *inside* the solenoid would be similar, but a given field line would encircle only part of the magnetic flux; Exercise 31 covers this situation. ■

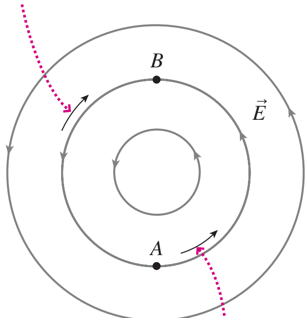


FIGURE 27.31 The work done to move charge in an induced electric field isn't path independent, so the induced field isn't conservative.

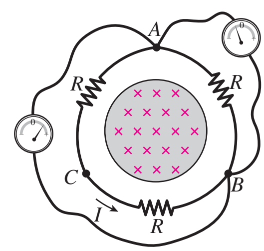
Conservative and Nonconservative Electric Fields

Static electric fields—those beginning and ending on stationary charge distributions—are *conservative*, meaning that the work required to move a charge between two points is path independent. A consequence is that it takes no work to move around a closed path in an electrostatic field; mathematically, we express this by writing

$$\oint \vec{E} \cdot d\vec{r} = 0 \qquad \text{(electrostatic field)}$$

In contrast, induced electric fields generally form closed loops, and here Faraday's law shows that the line integral of the electric field around a closed path is decidedly not zero. That means the induced electric field does work on a charge moved around a *closed* path and that the work done in moving between two points cannot be independent of the path taken (Fig. 27.31). The induced electric field is therefore not conservative.

GOT IT? 27.7 The figure shows three resistors in series surrounding an infinitely long solenoid with a changing magnetic field; the resulting induced electric field drives a current counterclockwise, as shown. Two identical voltmeters are shown connected to the *same* points A and B. What does each read? Explain any apparent contradiction. *Hint:* This is a challenging question!



Diamagnetism

We introduced diamagnetism in Chapter 26 but couldn't explain it there because it involves induced electric fields. Figure 27.32 shows a highly simplified model representing two atomic electrons with equal but opposite magnetic moments. Although a proper treatment of diamagnetism requires quantum mechanics, this model shows qualitatively how diamagnetism arises.

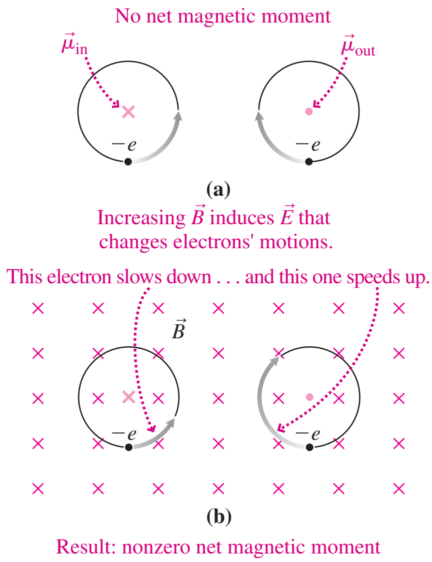


FIGURE 27.32 A simple model for diamagnetism.

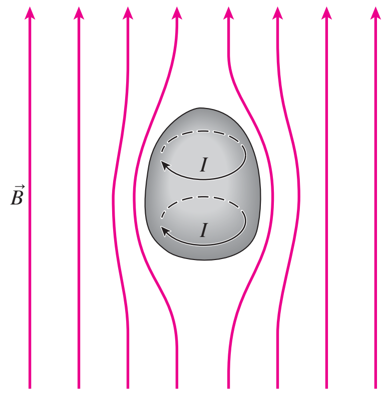


FIGURE 27.33 Induced currents in a superconductor completely cancel an applied magnetic field.

The dipole moments in Fig. 27.32*a* cancel, so the associated atom has no magnetic dipole moment. But what happens when a magnetic field is applied, pointing into the page (Fig. 27.32*b*), perhaps by moving the north pole of a bar magnet toward the page? The changing magnetic field results in an electric field that alters the electrons' speeds. In order to oppose the imposition of the magnetic field, the electron on the right speeds up. Its dipole moment, which points out of the page, increases and opposes the bar magnet's field. Meanwhile the left-hand electron's dipole moment decreases. Now the atom has a net dipole moment pointing out of the page, opposing the incoming magnet and resulting in the repulsive force that characterizes diamagnetism.

A superconductor is perfectly diamagnetic, meaning that the magnetic field resulting from induced currents completely cancels any applied field. Since these induced currents persist in the zero-resistance superconductor, the material completely excludes magnetic fields from its interior, a phenomenon known as the Meissner effect (Fig. 27.33). The repulsive force associated with the magnetic moments of a permanent magnet and a nearby superconductor results in the widely publicized phenomenon of magnetic levitation (Fig. 27.34).

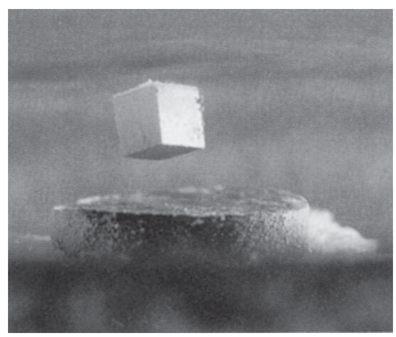
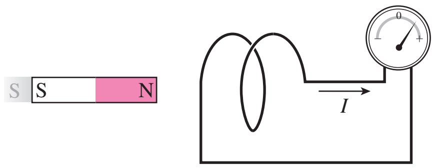


FIGURE 27.34 A small magnet levitates above a wafer of high-temperature superconductor in a bath of liquid nitrogen.

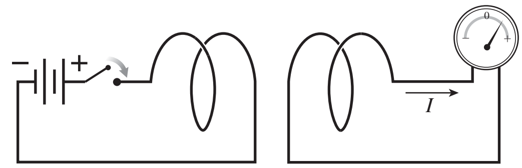
CHAPTER 27 SUMMARY

Big Picture

The big idea here is **electromagnetic induction**, a phenomenon in which **a changing magnetic field produces an electric field**. Applied to circuits, induction results in induced emfs that drive induced currents.



Here a moving magnet produces the changing magnetic field.

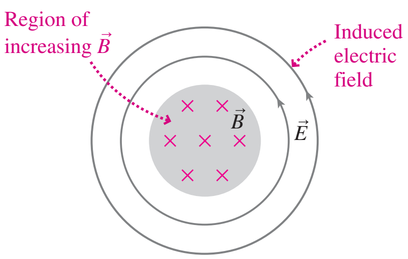


Here a change in current produces the changing magnetic field.

Key Concepts and Equations

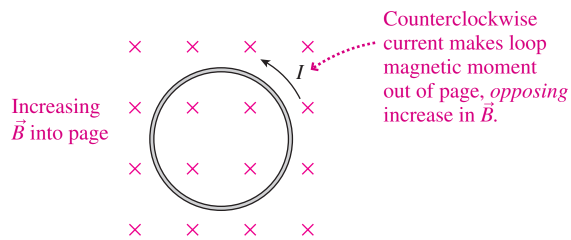
Faraday's law describes induction quantitatively, relating the line integral of the induced electric field to changing magnetic flux:

$$\oint \vec{E} \cdot d\vec{r} = -d\Phi_B/dt$$



In conductors, Faraday's law gives the induced emf: $\mathcal{E} = -d\Phi_B/dt$.

Lenz's law shows that electromagnetic induction is consistent with conservation of energy, which requires that induced effects act to oppose the changes that give rise to them.

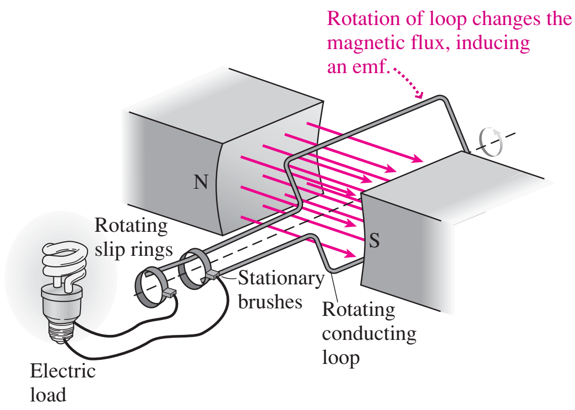


Magnetic fields contain stored energy, as do electric fields. The **magnetic-energy density** is

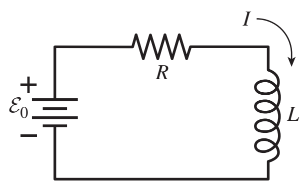
$$u_B = \frac{B^2}{2\mu_0}$$

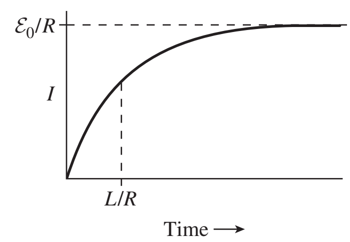
Applications

Electric generators convert mechanical to electrical energy by moving conductors in magnetic fields to induce emfs that drive currents.



Inductors are wire coils that encircle their own magnetic flux, giving **self-inductance** $L = \Phi_B/I$. An inductor opposes changes in current, producing an emf given by $\mathcal{E} = -L(dI/dt)$. Circuit quantities in a simple RL circuit change with **inductive time constant** L/R.





$$I = \frac{\mathcal{E}_0}{R}\left(1 - e^{-Rt/L}\right)$$

Diamagnetism occurs when electromagnetic induction results in atoms acquiring net magnetic moments; the result is a repulsive interaction.

For homework assigned on MasteringPhysics, go to www.masteringphysics.com | BIO Biology and/or medical-related problems C Computer problems

For Thought and Discussion

1. In Fig. 27.35, a bar magnet moves toward a conducting ring. What's the direction of the induced current in the ring?

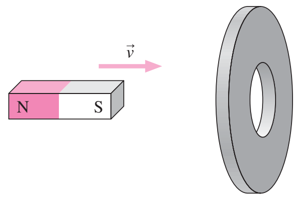


FIGURE 27.35 For Thought and Discussion 1

2. Figure 27.36 shows two concentric conducting loops, the outer connected to a battery and a switch. The switch is initially open. It's then closed, left closed for a while, and then reopened. Describe the currents in the inner loop during the entire procedure.

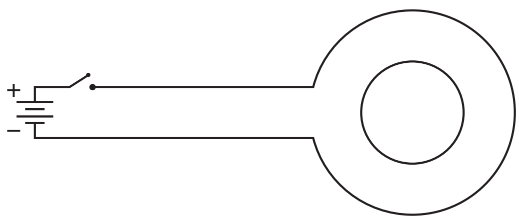


FIGURE 27.36 For Thought and Discussion 2

3. Fluctuations in Earth's magnetic field due to changing solar activity can wreak havoc with communications, even those using underground cables. How is this possible?
4. Chapter 26 claimed that a static magnetic field cannot change the energy of a charged particle. Is this true of a changing magnetic field? Discuss.
5. Can an induced electric field exist in the absence of a conductor?
6. A car battery has a 12-V emf, yet energy from the battery provides the 30,000-V spark that ignites the gasoline. How is this possible?
7. You have a fixed length of wire to wind into an inductor. Will you get more inductance if you wind a short coil with large diameter, or a long coil with small diameter?
8. In a popular demonstration of induced emf, a lightbulb is connected across a large inductor in an *RL* circuit, as shown in Fig. 27.37. When the switch is opened, the bulb flashes brightly and burns out. Why?

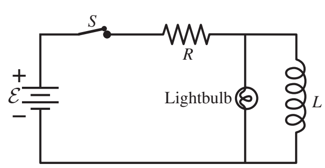


FIGURE 27.37 For Thought and Discussion 8

9. List some similarities and differences between inductors and capacitors.
10. A 1-H inductor carries 10 A, and a 10-H inductor carries 1 A. Which contains more stored energy?
11. It takes work to push two bar magnets together with like poles facing. Where does this energy go?

Exercises and Problems

Exercises

Sections 27.2 Faraday's Law and 27.3 Induction and Energy

12. Show that the volt is the SI unit for the rate of change of magnetic flux, making Faraday's law dimensionally correct.
13. Find the magnetic flux through a 5.0-cm-diameter circular loop oriented with the loop normal at 30° to a uniform 80-mT magnetic field.
14. A circular wire loop 40 cm in diameter has resistance 100 Ω and lies in a horizontal plane. A uniform magnetic field points vertically downward, and in 25 ms it increases linearly from 5.0 mT to 55 mT. Find the magnetic flux through the loop at (a) the beginning and (b) the end of the 25-ms period. (c) What's the loop current during this time? (d) Which way does this current flow?
15. A conducting loop of area 240 cm^2 and resistance 12 Ω is perpendicular to a spatially uniform magnetic field and carries a 320-mA induced current. At what rate is the magnetic field changing?
16. The magnetic field inside a 20-cm-diameter solenoid is increasing at 2.4 T/s. How many turns should a coil wrapped around the outside of the solenoid have so that the emf induced in the coil is 15 V?

Section 27.4 Inductance

17. Find the self-inductance of a 1000-turn solenoid 50 cm long and 4.0 cm in diameter.
18. The current in an inductor is changing at 100 A/s and the inductor emf is 40 V. What's the self-inductance?
19. A 2.0-A current is flowing in a 20-H inductor. A switch opens, interrupting the current in 1.0 ms. Find the induced emf in the inductor.
20. Your little sister is building a radio from scratch. Plans call for a 450-μH inductor wound on a cardboard tube. She brings you the tube from a toilet-paper roll (12 cm long, 4.0 cm diameter), and asks how many turns she should wind on the full length of the tube. Your answer?
21. What inductance should you put in series with a 100-Ω resistor to give a time constant of 2.2 ms?
22. The current in a series *RL* circuit increases to 20% of its final value in 3.1 μs. If $L = 1.8$ mH, what's the resistance?

Section 27.5 Magnetic Energy

23. How much energy is stored in a 5.0-H inductor carrying 35 A?
24. What's the current in a 10-mH inductor storing 50 μJ of energy?
25. A 220-mH inductor carries 350 mA. How much energy must be supplied to the inductor in raising the current to 800 mA?
26. A 500-turn solenoid 23 cm long and 1.5 cm in diameter carries 65 mA. How much magnetic energy does it contain?
27. Show that the quantity $B^2/2\mu_0$ has the units of energy density.
28. The world's strongest magnet that can produce a sustained field is a 45-T device at the National High Magnetic Field Laboratory in Florida. What's the corresponding magnetic-energy density?
29. Find the magnetic-field strength in a region where the magnetic-energy density is 7.8 J/cm^3.

Section 27.6 Induced Electric Fields

30. The induced electric field 12 cm from the axis of a 10-cm-radius solenoid is 45 V/m. Find the rate of change of the solenoid's magnetic field.
31. Find an expression for the electric-field strength *inside* the solenoid of Example 27.10, a distance r from the axis.

Problems

32. A conducting loop of area A and resistance R lies at right angles to a spatially uniform magnetic field. At time $t = 0$, the magnetic field and loop current are both zero. Subsequently, the current increases according to $I = bt^2$, where b is a constant with units A/s^2. Find an expression for the magnetic-field strength as a function of time.
33. A conducting loop with area $0.15\ m^2$ and resistance $6.0\ \Omega$ lies in the x-y plane. A spatially uniform magnetic field points in the z-direction. The field varies with time according to $B_z = at^2 - b$, where $a = 2.0\ T/s^2$ and $b = 8.0\ T$. Find the loop current (a) at $t = 3.0$ s and (b) when $B_z = 0$.
34. A square wire loop of side l and resistance R is pulled with constant speed v from a region of no magnetic field until it's fully inside a region of constant, uniform magnetic field $\vec{B}$ perpendicular to the loop plane. The boundary of the field region is parallel to one side of the loop. Find an expression for the total work done by whatever is pulling the loop.
35. A 5-turn coil 1.0 cm in diameter is rotated at 10 rev/s about an axis perpendicular to a uniform magnetic field. A voltmeter connected to the coil through rotating contacts reads a peak value 360 μV. What's the magnetic-field strength?
36. A magnetic field is given by $\vec{B} = B_0(x/x_0)^2\hat{k}$, where B_0 and x_0 are constants. Find an expression for the magnetic flux through a square of side $2x_0$ that lies in the x-y plane with one corner at the origin and sides coinciding with the positive x- and y-axes.
37. A square wire loop 3.0 m on a side is perpendicular to a uniform 2.0-T magnetic field. A 6-V lightbulb is in series with the loop, as shown in Fig. 27.38. The magnetic field is reduced steadily to zero over time Δt. (a) Find Δt such that the bulb will shine at full brightness. (b) Which way will the loop current flow?

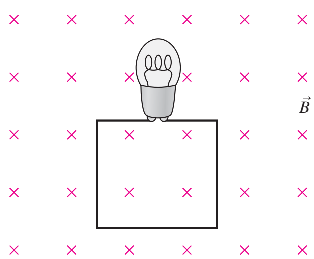


FIGURE 27.38 Problem 37

38. In Example 27.2 take $a = 1.0$ cm, $w = 3.5$ cm, and $l = 6.0$ cm. Suppose the rectangular loop is a conductor with resistance 50 mΩ, and the current I in the long wire is increasing at 25 A/s. Find the induced current in the loop. What's its direction?
39. A 2000-turn solenoid is 2.0 m long and 15 cm in diameter. The solenoid current is increasing at 1.0 kA/s. (a) Find the current in a 10-cm-diameter wire loop with resistance 5.0 Ω lying inside the solenoid and perpendicular to the solenoid axis. (b) Repeat for a similarly oriented 25-cm-diameter loop with the same resistance, lying entirely outside the solenoid.
40. **BIO** A *stent* is a cylindrical tube, often made of metal mesh, that's inserted into a blood vessel to overcome a constriction. It's sometimes necessary to heat the stent after insertion to prevent cell growth that could cause the constriction to recur. One method is to place the patient in a changing magnetic field, so that induced currents heat the stent. Consider a stainless-steel stent 12 mm long by 4.5 mm diameter, with total resistance 41 mΩ. Treating the stent as a wire loop in the optimum orientation, find the rate of change of magnetic field needed for a heating power of 250 mW.
41. A uniform magnetic field is given by $\vec{B} = bt\hat{k}$, where $b = 0.35$ T/s. Find the induced current in a conducting loop with area 240 cm^2 and resistance 0.20 Ω that lies in the x-y plane. In what direction is the current, as viewed from the positive z-axis?
42. You're an electrical engineer designing an alternator (the generator that charges a car's battery). Mechanical engineers specify a 10-cm-diameter rotating coil, and you determine that you can fit 250 turns in this coil. To charge a 12-V battery, you need a peak output of 14 V when the alternator is rotating at 1200 rpm. What do you specify for the alternator's magnetic field?
43. A generator consists of a rectangular coil 75 cm by 1.3 m, spinning in a 0.14-T magnetic field. If it's to produce a 60-Hz alternating emf with peak value 6.7 kV, how many turns must it have?
44. Figure 27.39 shows a pair of parallel conducting rails a distance l apart in a uniform magnetic field $\vec{B}$. A resistor R is connected across the rails, and a conducting bar of negligible resistance is being pulled along the rails with velocity $\vec{v}$ to the right. (a) What direction is the current in the resistor? (b) At what rate does the agent pulling the bar do work?

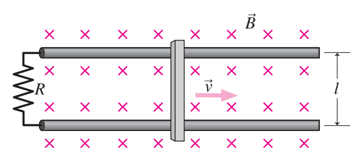


FIGURE 27.39 Problems 44–47 and 71

45. The resistor in Problem 44 is replaced by an ideal voltmeter. (a) To which rail should the positive meter terminal be connected if it's to indicate a positive voltage? (b) At what rate does the agent pulling the bar do work?
46. A battery of emf $\mathcal{E}$ is inserted in series with the resistor in Fig. 27.39, with its positive terminal toward the top rail. The bar is initially at rest, and now nothing's pulling it. (a) Describe the bar's subsequent motion. (b) The bar eventually reaches a constant speed. Why? (c) What is that constant speed, in terms of the magnetic field, the battery emf, and the rail spacing l? Does the resistance R affect the final speed? If not, what role does it play?
47. In Fig. 27.39, $l = 10$ cm, $B = 0.50$ T, $R = 4.0\ \Omega$, and $v = 2.0$ m/s. Find (a) the current in the resistor, (b) the magnetic force on the bar, (c) the power dissipation in the resistor, and (d) the mechanical power supplied by the agent pulling the bar. Compare your answers to (c) and (d).
48. The magnetic field inside a solenoid of circular cross section is given by $\vec{B} = bt\hat{k}$, where $b = 2.1$ T/ms. At time $t = 0.40\ \mu$s, a proton is inside the solenoid at $x = 5.0$ cm, $y = z = 0$, and is moving with velocity $\vec{v} = 4.8\hat{\jmath}$ Mm/s. Find the electromagnetic force on the proton.
49. An electron is inside a solenoid, 28 cm from the axis. It experiences a 1.3-fN electric force. At what rate is the solenoid's magnetic field changing?

50. During lab, you're given a circular wire loop of resistance R and radius a with its plane perpendicular to a uniform magnetic field. You're supposed to increase the field strength from B_1 to B_2 and measure the total charge that moves around the loop. Your lab partner claims that the details of how you vary the field will make a difference in the total charge; your hunch is that it won't. By integrating the loop current over time, determine who's right.
51. A *flip coil* is used to measure magnetic fields. It's a small coil placed with its plane perpendicular to a magnetic field, and then flipped through 180°. The coil is connected to an instrument that measures the total charge Q that flows during this process. If the coil has N turns, area A, and resistance R, show that the field strength is $B = QR/2NA$.
52. The current in a series RL circuit rises to half its final value in 7.6 s. What's the time constant?
53. In a series RL circuit like Fig. 27.22a, $\mathcal{E}_0 = 45$ V, $R = 3.3\ \Omega$, and $L = 2.1$ H. If the current is 9.5 A, how long has the switch been closed?
54. In Fig. 27.22a, take $R = 2.5\ \text{k}\Omega$ and $\mathcal{E}_0 = 50$ V. When the switch is closed, the current through the inductor rises to 10 mA in 30 μs. Find (a) the inductance and (b) the current in the circuit after many time constants.
55. How long does it take to dissipate 90% of the magnetic energy in Example 27.9?
56. A series RL circuit like Fig. 27.22a has $\mathcal{E}_0 = 60$ V, $R = 22\ \Omega$, and $L = 1.5$ H. Find the rate of change of the current (a) immediately after the switch is closed and (b) 100 ms later.
57. You're a safety engineer reviewing plans for a university's new high-rise dorm. The elevator motors draw 20 A and behave electrically like 2.5-H inductors. You're concerned about dangerous voltages developing across the switch when a motor is turned off, and you recommend that a resistor be wired in parallel with each motor. (a) What should be the resistance in order to limit the emf to 100 V? (b) How much energy will the resistor dissipate?
58. In Fig. 27.25, take $\mathcal{E}_0 = 12$ V, $R = 2.7\ \Omega$, and $L = 20$ H. Initially the switch is in position B and there's no current anywhere. At $t = 0$ the switch is thrown to position A, and at $t = 10$ s it's returned to B. Find the inductor current at (a) $t = 5.0$ s and (b) $t = 15$ s.
59. In Fig. 27.40, $\mathcal{E}_0 = 12$ V, $R_1 = 4.0\ \Omega$, $R_2 = 8.0\ \Omega$, and $R_3 = 2.0\ \Omega$. Find current I_2 (a) immediately after the switch is first closed and (b) a long time later. (c) After a long time, the switch is reopened. Now what's I_2?

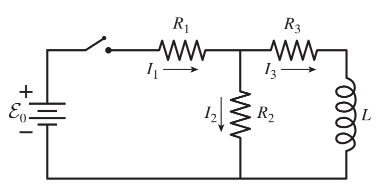


FIGURE 27.40 Problem 59

60. A battery, switch, resistor, and inductor are connected in series. When the switch is closed, the current rises to half its steady-state value in 1.0 ms. How long does it take for the magnetic energy in the inductor to rise to half its steady-state value?
61. When a nonideal 1.0-H inductor is short-circuited, its magnetic energy drops to one-fourth of its original value in 3.6 s. What is its resistance?
62. **BIO** Your hospital is installing a new MRI scanner using a 3.5-H superconducting solenoid carrying 1.8 kA. Copper is embedded in the coils to carry the current in the event of a quench (see Example 27.9). As safety officer, you're to specify (a) the maximum resistance that will limit power dissipation to 100 kW immediately after a loss of superconductivity and (b) the time it will take the power to drop to 50 kW. What specs do you give?
63. A neutron star's magnetic field is about 10^8 T. Consult Appendix C to compare the energy density in this field with that of (a) gasoline and (b) pure uranium-235 (mass density $19 \times 10^3\ \text{kg/m}^3$).
64. A single-turn loop of radius R carries current I. How does the magnetic-energy density at the loop center compare with that of a long solenoid of the same radius, carrying the same current, and consisting of n turns per unit length?
65. A wire of radius R carries current I distributed uniformly over its cross section. Find an expression for the total magnetic energy per unit length *within* the wire.
66. (a) Use Equation 27.8 to write an expression for the resistor's power dissipation as a function of time, and (b) integrate from $t = 0$ to $t = \infty$ to show that the total energy dissipated is equal to the energy initially stored in the inductor.
67. An electric field and a magnetic field have the same energy density. Find an expression for the ratio E/B and evaluate this ratio numerically. What are its units? Is your answer close to any of the fundamental constants listed inside the front cover?
68. A rectangular conducting loop of resistance R, mass m, and width w falls into a uniform magnetic field as shown in Fig. 27.41. (a) Explain why the loop eventually reaches a terminal speed. (b) Find an expression for the terminal speed.

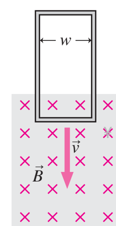


FIGURE 27.41 Problem 68

69. A conducting disk with radius a, thickness h, and resistivity ρ is inside a solenoid of circular cross section, its axis coinciding with the solenoid axis. The magnetic field in the solenoid is given by $B = bt$, where b is a constant. Find expressions for (a) the current density in the disk as a function of the distance r from the disk center and (b) the power dissipation in the entire disk. (*Hint:* Consider the disk as consisting of infinitesimal conducting loops.)
70. A small bar magnet is moved steadily from left to right along the axis of a wire loop, as in Fig. 27.5. Sketch qualitatively the current and power dissipation in the loop as functions of time. Take the current to be positive when it's counterclockwise as viewed from the right.
71. The bar in Problem 44 has mass m and is initially at rest. A constant force $\vec{F}$ to the right is applied to the bar. Formulate Newton's second law for the bar, and find its velocity as a function of time.
72. Use the node and loop laws to determine the current in R_2 as a function of time after the switch is closed in Conceptual Example 27.1.
73. A long, straight coaxial cable consists of two thin tubular conductors, the inner of radius a and the outer of radius b. Current I flows out along one conductor and back along the other. Find the cable's self-inductance per unit length.

74. You and your roommate are headed to Cancún for spring break. Your roommate, who has had only high school physics, has read that an emf can be induced in the wings of an airplane and wonders whether this would give enough voltage to power a portable music player. What's your answer? (Assume that the wingspan of your 747 is 60 m, the plane is flying at 600 mph, and Earth's magnetic field is 0.3 G.)

75. BIO One way to measure blood flow when blood vessels are exposed during surgery is to use an *electromagnetic flowmeter*. This device surrounds the blood vessel with an electromagnet, creating a magnetic field perpendicular to the blood flow. Since blood is a modest conductor, a motional emf develops across the blood vessel. Given vessel diameter d, magnetic field B, and voltage V measured across the vessel, show that the volume blood flow is given by $\pi d^2 V/4Bd$.

Passage Problems

Clever farmers with power lines crossing their land have been known to steal power by stringing wire near the power line and making use of the induced current. At least one such crime went to court and resulted in a conviction—despite the defense's claim that the defendant didn't touch the lines. Figure 27.42 shows a possible crime scene, with a rectangular wire loop mounted in a vertical plane beneath a power line. The power line carries a current of 10^4 A, alternating sinusoidally at 60 Hz.

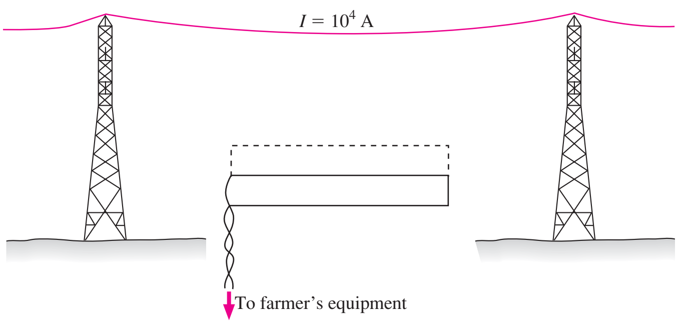


FIGURE 27.42 Crime scene for Passage Problems 76–79

76. If the loop were mounted in a horizontal rather than vertical plane at the same distance from the power line, the induced emf would
    a. increase slightly.
    b. decrease slightly.
    c. remain the same.
    d. become essentially zero.

77. If the loop's vertical dimension were doubled by extending it toward the power line (dashed line in Fig. 27.42), the induced emf would
    a. double.
    b. quadruple.
    c. more than double but not quadruple.
    d. increase but not quite double.

78. Suppose the same crime were committed in Europe, where the standard frequency is 50 Hz. Assuming everything else about the situation were the same, the induced emf would
    a. be greater.
    b. be less.
    c. be unchanged.
    d. depend on the nature of the energy source.

79. When this crime occurs,
    a. more fuel must be consumed at the power plant supplying the line.
    b. the power company does not suffer any economic damage.
    c. the power company can't determine that it's being robbed without an on-site inspection.
    d. there's no power left for customers further down the line.

Answers to Chapter Questions

Answer to Chapter Opening Question

The minus sign, which expresses conservation of energy in electromagnetic induction.

Answers to GOT IT? Questions

27.1. Opposite the direction shown in Fig. 27.10, but you'll still have to do work.
27.2. Counterclockwise.
27.3. It gets harder to turn. Constant rate implies a fixed peak emf, so lowering the resistance increases the current and therefore the power.
27.4. (b), because eddy currents dissipate some of its kinetic energy.
27.5. Changing current in the long wire produces an increasing magnetic field that points into the page at the loop. (a) The loop current opposes the increase in this field by producing a field in its interior that's out of the page. Therefore, the loop current is counterclockwise. (b) The loop current opposes the decrease in the into-the-page field, so it's clockwise.
27.6. (a), because the inductor emf is in such a direction as to oppose the current supplied by the external circuit.
27.7. Left-hand meter reads $2IR$, right-hand meter reads IR—even though they're electrically connected to the same points. There's no contradiction because the field isn't conservative, and electric potential therefore can't be defined unambiguously.

28 Alternating-Current Circuits

Why does alternating current facilitate the transmission and distribution of electric power?

New Concepts, New Skills

By the end of this chapter you should be able to

- Characterize AC circuit quantities in terms of amplitude, frequency, and phase (28.1).
- Explain the relations between current and voltage in resistors, capacitors, and inductors, and describe these relations using equations and phasor diagrams (28.2).
- Describe the oscillatory behavior of *LC* circuits (28.3).
- Explain damped oscillations and resonance (28.3, 28.4).
- Calculate power in AC circuits (28.5).
- Describe the operation of transformers and power supplies (28.6).

Connecting Your Knowledge

- This chapter builds on your knowledge of electric circuits (Chapter 25) and individual circuit elements: resistors, capacitors, and inductors (24.3, 23.2, 23.3, 27.4).
- We'll use ideas from Chapter 13, especially frequency and angular frequency, simple harmonic motion, and damped harmonic motion (13.1, 13.2, 13.6).

So far we've considered electric circuits energized by steady sources like batteries. But many circuits—from household power to audio and video signals to the "clock" that orchestrates events inside your computer—involve time-varying electrical quantities. Here we consider such **alternating-current** (AC) circuits.

28.1 Alternating Current

We saw in Chapter 27 how rotational motion in electric generators naturally leads to voltage and current that vary sinusoidally with time. Audio, video, and computer signals have more complicated time dependence, but, as we showed in Fig. 14.17, those signals can be analyzed as sums of sinusoidal terms. Studying circuits with sinusoidally varying electrical quantities therefore provides insights into all AC circuits.

A sinusoidal AC voltage or current is characterized by its amplitude, frequency, and phase constant—the same quantities we developed in Chapter 13 to describe simple harmonic motion. Amplitude is specified by the peak value $(V_\mathrm{p}, I_\mathrm{p})$ or the **root-mean-square** value $(V_\mathrm{rms}, I_\mathrm{rms})$. The rms is an average obtained by squaring the signal, taking the time average, and then taking the square root. For a sine wave, rms and peak values are related by

$$V_\mathrm{rms} = \frac{V_\mathrm{p}}{\sqrt{2}} \quad \text{and} \quad I_\mathrm{rms} = \frac{I_\mathrm{p}}{\sqrt{2}} \qquad (28.1)$$

The 120 V of household wiring, for example, is the rms value (see Fig. 28.1, next page).

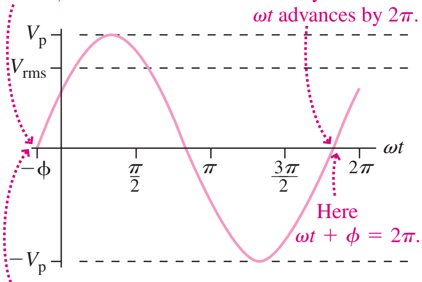


FIGURE 28.1 A sinusoidally varying AC voltage, showing peak and rms amplitudes and phase ϕ.

In practical situations we usually describe frequency f in cycles per second, or hertz (Hz). In mathematical analysis it's more convenient to use the angular frequency ω in radians per second or, equivalently, inverse seconds (s^{-1}). The relation between the two,

$$\omega = 2\pi f \tag{28.2}$$

is the same as for rotational and simple harmonic motion, and for the same reason: A full cycle contains 2π radians.

The phase constant ϕ of an AC signal tells when the sine curve crosses zero with positive slope (Fig. 28.1). A full mathematical description of an AC voltage or current then includes its amplitude (V_p, I_p), frequency (ω), and phase constant (ϕ):

$$V = V_p \sin(\omega t + \phi_V) \quad \text{and} \quad I = I_p \sin(\omega t + \phi_I) \tag{28.3}$$

Here we've labeled the phase constants with subscripts V and I to indicate that voltage and current—even in the same circuit element—need not have the same phase.

EXAMPLE 28.1 AC: Characterizing Household Voltage

Standard household wiring in North America supplies 120 V rms at 60 Hz. Express this mathematically in the form of Equation 28.3, assuming the voltage is rising through zero at time $t = 0$.

INTERPRET We're given an AC voltage in "practical" units, and we're asked to express it in the more mathematical form of Equation 28.3. We identify 120 V as the amplitude V_{rms}, 60 Hz as the frequency f, and the information about timing as describing the phase.

DEVELOP Equation 28.3, $V = V_p \sin(\omega t + \phi_V)$, contains the peak amplitude V_p and angular frequency ω. Equations 28.1, $V_{rms} = V_p/\sqrt{2}$, and 28.2, $\omega = 2\pi f$, determine these quantities from the values we're given.

EVALUATE Equation 28.1 gives $V_p = \sqrt{2}V_{rms} = (\sqrt{2})(120\text{ V}) = 170\text{ V}$, and Equation 28.2 gives $\omega = 2\pi f = (2\pi)(60\text{ Hz}) = 377\text{ s}^{-1}$. We don't have an equation for phase, but the fact that the sine curve rises through zero at $t = 0$ tells us that $\phi = 0$. So Equation 28.3's description of this AC voltage becomes $V = 170\sin(377t)$ V.

ASSESS Make sense? Both the peak voltage and the angular frequency are numerically greater than their more familiar counterparts. That's because the rms voltage is a kind of average, lower than the peak, and because the angular frequency measures radians per second rather than full cycles. Incidentally, wires entering your house actually carry 240 V rms, which is split into separate 120-V circuits except for major appliances like stoves, dryers, and water heaters; these operate at the full 240 V rms. In Europe, standard household voltage is 230 V rms at 50 Hz, and much of the rest of the world uses 220-V, 50-Hz power. ■

28.2 Circuit Elements in AC Circuits

Here we examine separately the AC behavior of resistors, capacitors, and inductors so we can subsequently understand what happens when we combine these elements in AC circuits.

Resistors

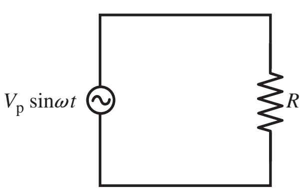


FIGURE 28.2 A resistor connected across an AC generator (symbol ⊗).

An ideal resistor is a device whose current and voltage are proportional: $I = V/R$. Figure 28.2 shows a resistor R connected across an AC generator, making the voltage across the resistor equal to the generator voltage. The generator voltage is described by Equation 28.3, where we take $\phi_V = 0$. Then the current is

$$I = \frac{V}{R} = \frac{V_p \sin\omega t}{R} = \frac{V_p}{R}\sin\omega t$$

The current has the same frequency as the voltage, and, since its phase constant is also zero, voltage and current are *in phase*—they peak at the same time. The peak current is the peak voltage divided by the resistance: $I_p = V_p/R$. Both voltage and current are sinusoidal, so their rms values are in the same ratio as their peak values; thus $I_{rms} = V_{rms}/R$.

Capacitors

Figure 28.3 shows a capacitor connected across an AC generator. In Chapter 23, we saw that voltage and charge are directly proportional in a capacitor: $q = CV$. Differentiating this relation gives

$$\frac{dq}{dt} = C\frac{dV}{dt}$$

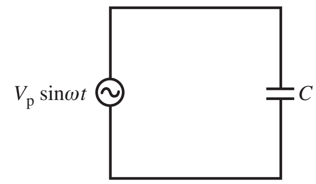


FIGURE 28.3 A capacitor connected across an AC generator.

But dq/dt is the current flowing to the capacitor plates (which we'll call the "capacitor current" even though charge doesn't actually flow through the space between the plates). So we have $I = C(dV/dt)$. The generator voltage $V_p \sin\omega t$ appears directly across the capacitor, so

$$\begin{aligned} I &= C\frac{d}{dt}(V_p \sin\omega t) \\ &= \omega C V_p \cos\omega t = \omega C V_p \sin\left(\omega t + \frac{\pi}{2}\right) \end{aligned} \tag{28.4}$$

Because the cosine curve is just a sine curve shifted to the left by $\pi/2$ or 90°, Equation 28.4 tells us that **in a capacitor, current leads voltage by 90°** (Fig. 28.4).

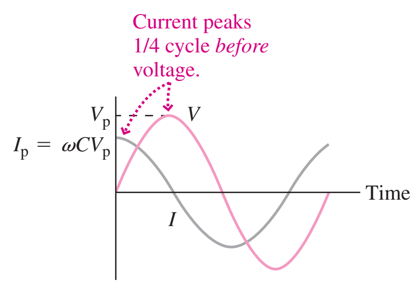


FIGURE 28.4 The current in a capacitor leads the voltage by one-fourth of a cycle, $\pi/2$ radians or 90°.

The term $\omega C V_p$ multiplying the cosine in Equation 28.4 is the peak current, so $I_p = \omega C V_p$ or, in a form resembling Ohm's law,

$$I_p = \frac{V_p}{1/\omega C} = \frac{V_p}{X_C} \tag{28.5}$$

where we've defined $X_C = 1/\omega C$.

Equation 28.5 shows that the capacitor acts somewhat like a resistance $X_C = 1/\omega C$. But not quite! This "resistance" gives the relation between peak voltage and current, but it doesn't tell the whole story. The capacitor also introduces a phase difference between voltage and current. This phase difference reflects a fundamental physical difference between resistors and capacitors. A resistor dissipates electric energy as heat. A capacitor stores and releases electric energy. Over a complete cycle, the agent turning the generator in Fig. 28.3 does no net work, while the agent turning the generator with the resistive load of Fig. 28.2 continuously does work that gets dissipated as heat in the resistor. We give the quantity X_C in Equation 28.5 the name **capacitive reactance**. Like resistance, reactance is measured in ohms (Ω).

Does it make sense that X_C depends on frequency? Yes. As frequency goes to zero, X_C goes to infinity. At zero frequency nothing is changing; there's no charge moving on or off the plates, and the capacitor might as well be an open circuit. As frequency increases, larger currents flow to move charge on and off the plates in ever-shorter times, so the capacitor looks increasingly like a short circuit. To summarize, a capacitor at low frequencies acts like an open circuit, while at high frequencies it acts like a short circuit.

Why does the capacitor current *lead* the voltage? Because the capacitor voltage is proportional to its charge, and it takes current to move charge onto the capacitor plates. Therefore, current flows *before* the voltage changes significantly.

Inductors

Figure 28.5 shows an inductor connected across an AC generator. The loop law for this circuit is $V_p \sin\omega t + \mathcal{E}_L = 0$. From Chapter 27 we know that the inductor emf is $\mathcal{E}_L = -L(dI/dt)$, so the loop law becomes

$$V_p \sin\omega t = L\frac{dI}{dt}$$

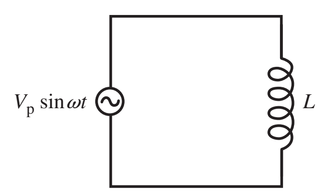


FIGURE 28.5 An inductor connected across an AC generator.

To obtain a relation involving the current I rather than its derivative, we integrate:

$$\int V_p \sin \omega t \, dt = \int L \frac{dI}{dt} dt = \int L \, dI$$

The integral of sine is the negative cosine, so

$$-\frac{V_p}{\omega} \cos \omega t = LI$$

Here we've set the integration constants to zero because nonzero values would represent a DC emf and current that aren't in this circuit. Solving for I gives

$$I = -\frac{V_p}{\omega L} \cos \omega t = \frac{V_p}{\omega L} \sin\left(\omega t - \frac{\pi}{2}\right) \tag{28.6}$$

where the last step follows because $\sin(\alpha - \pi/2) = -\cos \alpha$ for any α.

Equation 28.6 shows that current in the inductor lags the voltage by $\pi/2$ or 90°. Equivalently, **the voltage across an inductor leads the inductor current by 90°** (Fig. 28.6). Equation 28.6 also shows that the peak current is

$$I_p = \frac{V_p}{\omega L} = \frac{V_p}{X_L} \tag{28.7}$$

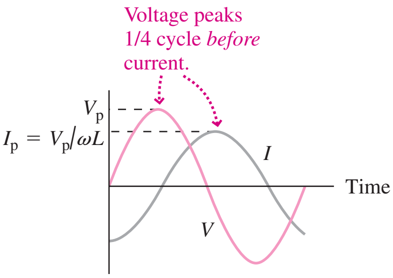


FIGURE 28.6 The voltage across an inductor leads the current by $\pi/2$ or 90°.

Again, this equation resembles Ohm's law, with **inductive reactance** $X_L = \omega L$. As with the capacitor, no power is dissipated; instead, energy is alternately stored and released as the inductor's magnetic field builds and decays.

Does it make sense that inductive reactance increases with ω and L? Through its induced back emf, an inductor opposes changes in current. The greater the inductance, the greater the opposition. And the more rapidly the current is changing, the more vigorously the inductor opposes the change, so inductive reactance increases at high frequencies. At very high frequencies, an inductor looks like an open circuit. But at very low frequencies, it looks more and more like a short circuit, until with direct current (zero frequency), an inductor exhibits zero reactance because current isn't changing.

Why does the inductor voltage *lead* the current? Because a changing current in an inductor induces an emf. *Before* the current can build up significantly, there must first, therefore, be voltage across the inductor.

Table 28.1 summarizes amplitude and phase relations in resistors, capacitors, and inductors.

Table 28.1 Amplitude and Phase Relations in Circuit Elements

Circuit Element	Peak Current versus Voltage	Phase Relation
Resistor	$I_p = \frac{V_p}{R}$	V and I in phase
Capacitor	$I_p = \frac{V_p}{X_C} = \frac{V_p}{1/\omega C}$	I leads V by 90°
Inductor	$I_p = \frac{V_p}{X_L} = \frac{V_p}{\omega L}$	V leads I by 90°

GOT IT? 28.1 A capacitor and inductor are connected across separate but identical electric generators, and the same current flows in each. If the frequency of the generators is doubled, which component will carry more current?

EXAMPLE 28.2 **Inductors and Capacitors: Equal Currents?**

A capacitor is connected across a 60-Hz, 120-V rms power line, and an rms current of 200 mA flows. (a) Find the capacitance. (b) What inductance, connected across the same power line, would result in the same current? (c) How would the phases of the inductor and capacitor currents compare?

INTERPRET We're being asked about the relation between AC voltage and current in capacitors and inductors. The idea here is that the voltage–current relation depends not only on the component values but also on frequency, and it involves phase as well as amplitude.

DEVELOP Equations 28.5, $I_{Cp} = V_{Cp}\omega C$, and 28.7, $I_{Lp} = V_{Lp}/\omega L$, relate the peak current and voltage in the two devices. Since rms and peak values are proportional, similar equations also relate rms current and voltage. The equations and associated phase relations also appear in Table 28.1.

EVALUATE (a) For the capacitor, we know the voltage and current. Equation 28.5 then gives $C = I_{\mathrm{Crms}}/\omega V_{\mathrm{Crms}} = 4.42\ \mu\mathrm{F}$, where we used $I_{\mathrm{Crms}} = 0.20$ A and $\omega = 2\pi f = 377\ \mathrm{s}^{-1}$ as we found in Example 28.1 for 60-Hz AC power. (b) For an inductor to pass the same current, it must have the same reactance; comparing Equations 28.5 and 28.7 shows that $\omega L = 1/\omega C$, or

$$L = \frac{1}{\omega^2 C} = \frac{1}{(377\ \mathrm{s}^{-1})^2(4.42\ \mu\mathrm{F})} = 1.6\ \mathrm{H}$$

(c) Table 28.1 shows that the capacitor current *leads* the voltage by 90°, while the inductor current *lags* by 90°; therefore, the capacitor and inductor currents must be out of phase by 180° or π radians.

ASSESS Our expression for L shows that a larger capacitor would require a smaller inductor for the same current. That's because a larger capacitor has *lower* reactance and so passes more current at a given frequency. But an inductor is the opposite: A larger inductor has *higher* reactance. So at a fixed frequency, the inductance required for a given current scales inversely with the capacitance required for the same current. ■

Phasor Diagrams

Phasor diagrams summarize phase and amplitude relations in AC circuits. A **phasor** is an arrow whose length represents the amplitude of an AC voltage or current, rotating counterclockwise with the angular frequency ω of the AC quantity. The phasor's component on either axis represents the sinusoidally varying AC quantity. We'll use the vertical axis; others, especially electrical engineers, may use the horizontal.

Figure 28.7*a* shows phasors for current and voltage in a resistor. Since they're in phase, the two phasors point in the same direction. For capacitors and inductors, current and voltage phasors are at right angles, indicating 90° phase differences (Fig. 28.7*b, c*). The phasor magnitudes are related by $V_p = I_pX$, with X being the appropriate reactance. As they rotate, the phasors' vertical components trace out current and voltage graphs like those of Figs. 28.4 and 28.6. You should convince yourself that the relations of Table 28.1 are correctly described by the phasor diagrams of Fig. 28.7.

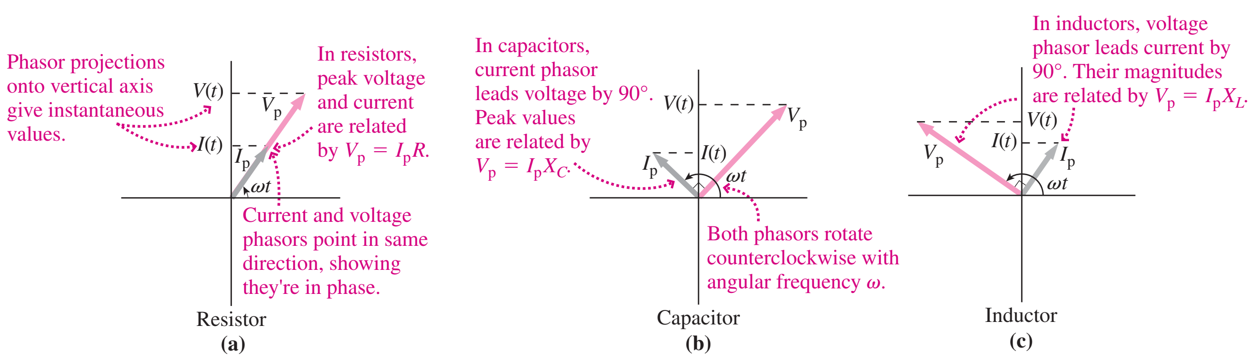


FIGURE 28.7 Phasor diagrams showing voltage and current in (a) a resistor, (b) a capacitor, and (c) an inductor.

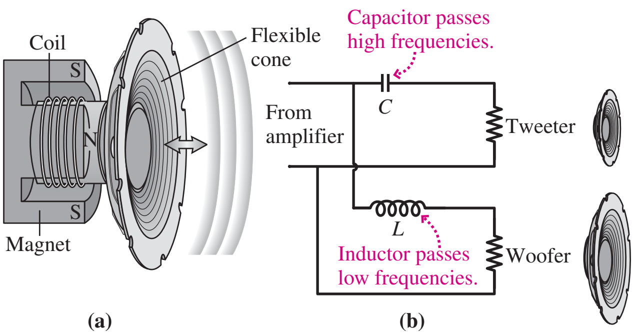
Capacitors and Inductors: A Comparison

Capacitors and inductors are complementary. A capacitor opposes instantaneous changes in voltage; an inductor opposes instantaneous changes in current. In an *RC* circuit, voltage builds up across the capacitor. In an *RL* circuit, current builds up in the inductor. Similar curves describe capacitor voltage and inductor current over time. A capacitor stores electric energy $\frac{1}{2}CV^2$. An inductor stores magnetic energy $\frac{1}{2}LI^2$. A capacitor acts like an open circuit at low frequencies; an inductor like a short circuit at low frequencies. Each exhibits the opposite behavior at high frequencies. These comparisons reflect a deeper complementarity between electric and magnetic fields. Any verbal description of a capacitor applies to an inductor if we replace the words "capacitor" with "inductor," "electric" with "magnetic," and "voltage" with "current." Table 28.2 summarizes the complementary aspects of capacitors and inductors.

Table 28.2 Capacitors and Inductors

	Capacitor	**Inductor**
Defining relation	$C = \dfrac{q}{V}$	$L = \dfrac{\Phi_B}{I}$
Defining relation; differential form	$I = C\dfrac{dV}{dt}$	$\mathcal{E} = -L\dfrac{dI}{dt}$
Opposes changes in	Voltage	Current
Energy storage	In electric field $U = \frac{1}{2}CV^2$	In magnetic field $U = \frac{1}{2}LI^2$
Behavior in low-frequency limit	Open circuit	Short circuit
Behavior in high-frequency limit	Short circuit	Open circuit
Reactance	$X_C = 1/\omega C$	$X_L = \omega L$
Phase	Current leads voltage by 90°	Voltage leads current by 90°

APPLICATION Loudspeaker Systems



Loudspeakers convert electrical energy to sound, using the magnetic force on a coil that fits loosely around a permanent magnet. Part (a) of the figure shows that the coil is attached to a flexible cone. Cone and coil move back and forth as AC current corresponding to the audio signal flows in the coil. The moving cone disturbs the air, producing sound waves.

Good loudspeaker systems include at least two separate units. A small *tweeter* produces high-frequency sound, while a larger, more massive *woofer* handles the low frequencies. A *crossover network* uses inductors and capacitors to "steer" the high- and low-frequency signals to the appropriate speakers. As the circuit diagram in part (b) shows, an inductor in series with the woofer blocks high frequencies but lets low frequencies pass unimpeded; a capacitor in series with the tweeter does the opposite. This circuit is an example of a *filter*, used in electronic systems to pass preferentially a range of frequencies.

28.3 *LC* Circuits

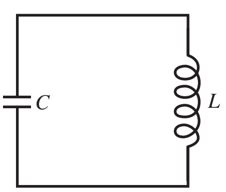


FIGURE 28.8 An *LC* circuit.

Suppose we charge a capacitor to some voltage V_p and corresponding charge q_p, and then connect it across an inductor, as shown in Fig. 28.8. The capacitor contains stored electric energy, but initially there's no current in the inductor and so no stored magnetic energy (Fig. 28.9*a*). The capacitor begins to discharge through the inductor, but slowly at first because the inductor opposes changes in current. Gradually the current rises, and with it the magnetic energy in the inductor. The capacitor voltage, charge, and stored energy decrease. At some time the initial energy is divided equally between capacitor and inductor (Fig. 28.9*b*). But the capacitor keeps discharging, eventually reaching zero charge (Fig. 28.9*c*). Now all the energy that was originally in the electric field of the capacitor is in the magnetic field of the inductor.

Does everything stop at this point? No, because there's current in the inductor, and inductor current can't change instantaneously. So the current keeps flowing and begins piling positive charge on the bottom plate of the capacitor (Fig. 28.9*d*). Stored electric energy increases, and current and magnetic energy both decrease. Eventually

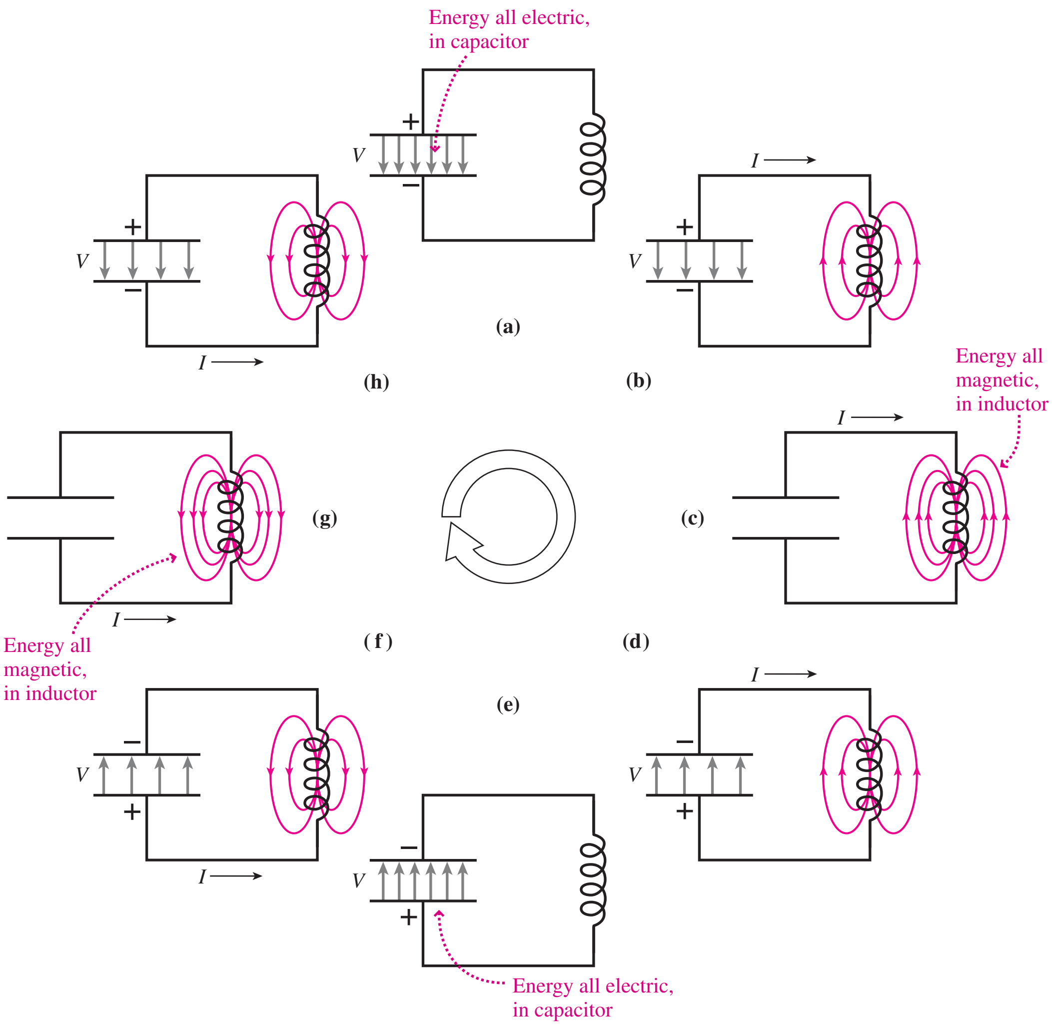


FIGURE 28.9 *LC* oscillations transfer energy between electric and magnetic fields.

the capacitor is fully charged but opposite its initial state (Fig. 28.9*e*). Again all the energy is in the capacitor. Now the capacitor begins to discharge, and the process repeats, with a counterclockwise current (Fig. 28.9*f*). All the energy is transferred to the inductor (Fig. 28.9*g*), and then back to the capacitor (Fig. 28.9*a* again). The circuit is now back to its initial state. Provided there's no energy loss, the oscillation repeats indefinitely.

This *LC* oscillation should remind you of the mass–spring system of Chapter 13. There, energy went back and forth between kinetic energy of the mass and potential energy of the spring. Here, energy goes between magnetic energy of the inductor and electric energy of the capacitor. The mass–spring system oscillates with frequency determined by the mass m and spring constant k. The *LC* circuit oscillates with frequency determined by the inductance L and capacitance C, as we'll show next. Figure 28.10 illustrates this analogy between the mass–spring system and the *LC* circuit.

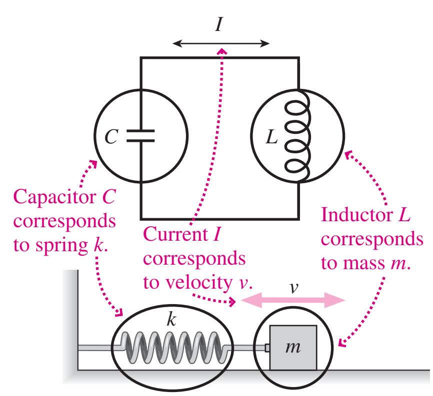


FIGURE 28.10 An *LC* circuit is the electrical analog of a mass–spring system.

Analyzing the *LC* Circuit

The total energy in the *LC* circuit is the sum of magnetic and electric energy:

$$U = U_B + U_E = \tfrac{1}{2}LI^2 + \tfrac{1}{2}CV^2$$

In an ideal LC circuit this quantity doesn't change, so its derivative is zero:

$$\frac{dU}{dt} = \frac{d}{dt}\left(\tfrac{1}{2}LI^2 + \tfrac{1}{2}CV^2\right) = 0$$

Carrying out the differentiation, we have

$$LI\frac{dI}{dt} + CV\frac{dV}{dt} = 0$$

We substitute $V = q/C$, $dV/dt = (1/C)dq/dt$, $I = dq/dt$, and $dI/dt = d^2q/dt^2$ and then divide by I to get

$$L\frac{d^2q}{dt^2} + \frac{1}{C}q = 0 \qquad (28.8)$$

This is a differential equation describing capacitor charge q as a function of time. We encountered a similar equation in Chapter 13 for the mass–spring system: $m(d^2x/dt^2) + kx = 0$. The solution there was a sinusoidal oscillation with angular frequency $\omega = \sqrt{k/m}$. In Equation 28.8, L replaces m and $1/C$ replaces k; therefore, the solution to Equation 28.8 is a sinusoidal oscillation:

$$q = q_{\mathrm{p}}\cos\omega t \qquad (28.9)$$

with angular frequency

$$\omega = \frac{1}{\sqrt{LC}} \qquad (28.10)$$

Equation 28.9 readily provides other electrical quantities in the LC circuit. Using $q = CV$ gives the voltage, and differentiating gives $I = dq/dt$. From there you can get the electric and magnetic energies, $U_E = \frac{1}{2}CV^2$ and $U_B = \frac{1}{2}LI^2$. Sum them to verify that the total energy remains constant (Fig. 28.11; see Problem 64 for details).

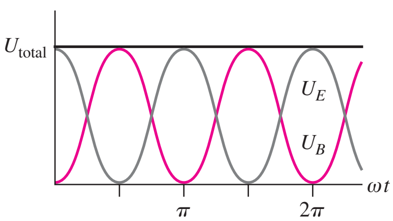


FIGURE 28.11 Electric and magnetic energies in an LC circuit sum to a constant total energy.

GOT IT? 28.2 You have an LC circuit that oscillates at a typical AM radio frequency of 1 MHz. You want to change the capacitor so it oscillates at a typical FM frequency, 100 MHz. Should you make the capacitor (a) larger or (b) smaller? By what factor?

EXAMPLE 28.3 An *LC* Circuit: Tuning a Piano

You wish to make an LC circuit oscillate at 440 Hz (A above middle C) to use in tuning pianos. You have a 20-mH inductor. (a) What value of capacitance should you use? (b) If you charge the capacitor to 5.0 V, what will be the peak current in the circuit?

INTERPRET This problem is about designing an LC circuit for a given frequency; in (b) we want the peak current—that is, the current when all the energy is in the inductor.

DEVELOP Equation 28.10, $\omega = 1/\sqrt{LC}$, relates frequency, capacitance, and inductance; we're given L and f. With $\omega = 2\pi f$, we can solve for C. We've recognized that the peak current comes when all the energy is the magnetic energy $\frac{1}{2}LI^2$ of the inductor. Given the initial capacitor voltage, we can equate this with the initial electric energy $\frac{1}{2}CV^2$ and solve for I.

EVALUATE (a) Equation 28.10 gives $C = 1/\omega^2L = 1/4\pi^2f^2L = 6.5\ \mu\text{F}$, where $f = 440$ Hz and $L = 20$ mH. (b) Now that we know C, we equate the peak magnetic energy with the peak electric energy to get $\frac{1}{2}LI^2 = \frac{1}{2}CV^2$. Solving gives

$$I = \sqrt{\frac{C}{L}}V = \sqrt{\frac{6.54\ \mu\text{F}}{20\ \text{mH}}}(5.0\ \text{V}) = 90\ \text{mA}$$

ASSESS Our expression shows that the higher the initial voltage, the greater the current. Obviously, the higher the voltage, the greater the energy, so the greater the current needed when this energy becomes all magnetic. A larger capacitance also raises the electric energy $\frac{1}{2}CV^2$, while a larger inductance lowers the current needed to achieve the same magnetic energy $\frac{1}{2}LI^2$.

Resistance in *LC* Circuits—Damping

Real inductors, capacitors, and wires have resistance (Fig. 28.12). If the resistance is low enough that only a small fraction of the energy is lost in each cycle, then the analysis in the preceding discussion applies. The circuit oscillates at a frequency given very nearly by Equation 28.10, but the oscillation amplitude slowly declines as energy is dissipated in the resistance.

We can analyze an *RLC* circuit by evaluating dU/dt as before, this time setting the result not to zero but to the rate of energy dissipation, $-I^2R$, where the minus sign indicates energy loss from the circuit:

$$\frac{dU}{dt} = \frac{d}{dt}\left(\tfrac{1}{2}LI^2 + \tfrac{1}{2}CV^2\right) = -I^2R$$

Making the same substitutions as before leads to

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{q}{C} = 0$$

This is mathematically identical to Equation 13.16 for damped harmonic motion, with L again replacing m, $1/C$ replacing k, and now R replacing the damping constant b. The solution follows by analogy with Equation 13.17, which is the solution to Equation 13.16:

$$q(t) = q_p e^{-Rt/2L}\cos\omega t \qquad (28.11)$$

Voltage and current behave similarly, with oscillation amplitude decaying exponentially with time constant $2L/R$ (Fig. 28.13).

As the resistance increases, oscillations decay more rapidly and the frequency of oscillation decreases. Eventually, when the time constant $2L/R$ equals the inverse of the frequency given in Equation 28.10, we have **critical damping**. Then all circuit quantities decay to zero without oscillation, just as we found for mechanical systems. In circuits designed to oscillate, like radio transmitters or TV tuners, engineers obviously want to minimize damping. But in situations where oscillations would be a nuisance, it's important that circuits have enough resistance to suppress oscillation.

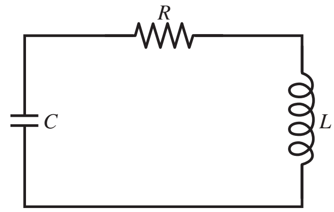


FIGURE 28.12 An *RLC* circuit.

FIGURE 28.13 An oscilloscope displays the capacitor voltage in an *RLC* circuit.

28.4 Driven *RLC* Circuits and Resonance

Figure 28.14 shows an *RLC* circuit connected across an AC generator. Adding the generator is like adding the external driving force on the mechanical oscillator that we considered in Section 13.7. We'll call the generator frequency ω_d, the driving frequency, just as we did in Chapter 13. Pursuing the mechanical analogy, we expect the driven *RLC* circuit to exhibit resonant behavior as we discussed in Section 13.7. Such electrical resonance is crucial to the operation of radio, TV, and other frequency-specific devices.

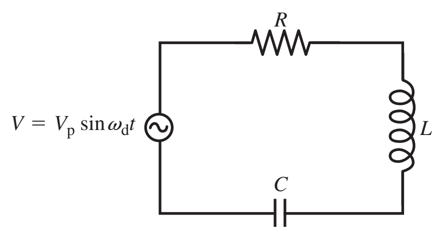


FIGURE 28.14 A series *RLC* circuit driven by an AC generator.

Resonance in the *RLC* Circuit

Suppose we vary the generator frequency ω_d in Fig. 28.14 while keeping the generator's peak voltage constant. At low frequencies the capacitor acts almost like an open circuit (its reactance $X_C = 1/\omega C$ is large), so little current flows. At high frequencies the inductor acts almost like an open circuit (its reactance $X_L = \omega L$ is large), so little current flows. At some intermediate frequency the current must be a maximum. We now show that this **resonant frequency** is the undamped natural frequency $\omega_0 = 1/\sqrt{LC}$.

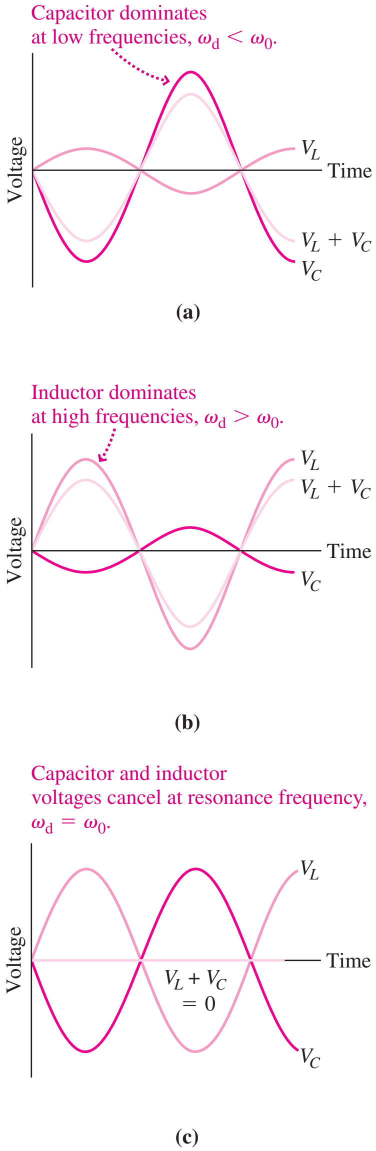


FIGURE 28.15 Capacitor and inductor voltages are 180° out of phase, but their relative magnitudes vary with frequency.

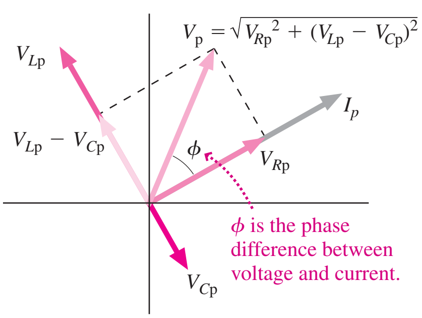


FIGURE 28.16 Phasor diagram for the driven *RLC* circuit, for the case $\omega > \omega_0$.

Figure 28.14 is a series circuit, so the *same* current flows through all components. The voltage in a capacitor lags the current by 90°, while the voltage in an inductor leads by 90°. Since the same current flows in both components, the inductor and capacitor voltages are therefore 180° out of phase and thus they tend to cancel (Fig. 28.15). But that cancellation is complete only when the two voltages have the same peak value. Since the current is the same in both components, comparison of Equations 28.5 for the capacitor, $I_p = V_p/X_C$, and 28.7 for the inductor, $I_p = V_p/X_L$, shows that the peak voltages are the same when the capacitive reactance $X_C = 1/\omega C$ equals the inductive reactance $X_L = \omega L$. Equating the reactances gives

$$\omega_0 = \frac{1}{\sqrt{LC}} \qquad \text{(resonant frequency)}$$

This is precisely the undamped natural frequency of Equation 28.10.

At resonance the capacitor and inductor voltages completely cancel. The voltage across the pair together is zero and—at the resonant frequency only—the pair might just as well be a wire. At resonance the resistance alone determines the circuit current. At any other frequency the capacitor and inductor voltages don't cancel, and the current is lower.

GOT IT? 28.3 You measure the capacitor and inductor voltages in a driven *RLC* circuit, and find 10 V for the rms capacitor voltage and 15 V for the rms inductor voltage. Is the driving frequency (a) above or (b) below resonance?

Frequency Response of the *RLC* Circuit

Here we'll use a phasor diagram (recall Section 28.2) to find the current in an *RLC* circuit as a function of the driving frequency (Fig. 28.16). Since the same current flows through all elements of this series circuit, a single phasor of length I_p represents the current. The resistor voltage is in phase with the current, so its phasor, V_{Rp}, is in the same direction as I_p. But the inductor voltage leads the current and the capacitor voltage lags, each by 90°, so their phasors, V_{Lp} and V_{Cp}, are perpendicular to the current phasor. At each instant the three voltages sum to give the generator voltage; Fig. 28.16 shows that this sum has magnitude $V_p = \sqrt{V_{Rp}^2 + (V_{Lp} - V_{Cp})^2}$. Expressing this in terms of the common current I_p and the resistance and reactances gives $V_p = \sqrt{I_p^2R^2 + (I_pX_L - I_pX_C)^2}$. Solving for I_p gives

$$I_p = \frac{V_p}{\sqrt{R^2 + (X_L - X_C)^2}} = \frac{V_p}{Z} \qquad (28.12)$$

where we've defined the **impedance**, Z, as $Z = \sqrt{R^2 + (X_L - X_C)^2}$. Impedance is a generalization of resistance to include frequency-dependent effects of capacitance and inductance. Equation 28.12 is the corresponding generalization of Ohm's law. Impedance is lowest when $X_L = X_C$ or $\omega = 1/\sqrt{LC}$; then it's equal to the resistance alone. But Z becomes large at high frequencies, where $X_L = \omega L$ becomes large, and at low frequencies, where $X_C = 1/\omega C$ is large.

Figure 28.17 plots resonance curves from Equation 28.12, showing peak current versus frequency for three resistance values. At low resistance, the curve peaks sharply. Such a **high-*Q*** (for high-quality) circuit does a good job distinguishing its resonance frequency from nearby frequencies. High-*Q* circuits are important in applications such as radio, TV, and cell phones, where many signals occupy nearby frequencies. With higher resistance, the resonance curve broadens and the circuit responds to a range of frequencies; such a circuit has low Q. Problem 73 gives a rigorous definition of Q.

Equation 28.12 relates peak current and voltage in the *RLC* circuit, but it doesn't tell the whole story. As Fig. 28.16 shows, current and voltage are out of phase by the angle ϕ. Trigonometry gives $\tan\phi = (V_{Lp} - V_{Cp})/V_{Rp}$ or, since voltages are proportional to reactances and resistance,

$$\tan\phi = \frac{X_L - X_C}{R} = \frac{\omega L - 1/\omega C}{R} \tag{28.13}$$

where $\phi = \phi_V - \phi_I$ is the phase difference between voltage and current. Positive ϕ means voltage leads current; negative ϕ means current leads voltage.

At resonance, $X_L = X_C$ and $\phi = 0$. Here capacitor and inductor voltages cancel, and the circuit behaves like a pure resistance. At low frequencies, capacitive reactance dominates; here ϕ is negative and the current leads the voltage. This is just what we expect in a capacitive circuit. The opposite is true at high frequencies, where the inductive reactance dominates. Figure 28.18 shows the phase difference as a function of frequency for three resistance values.

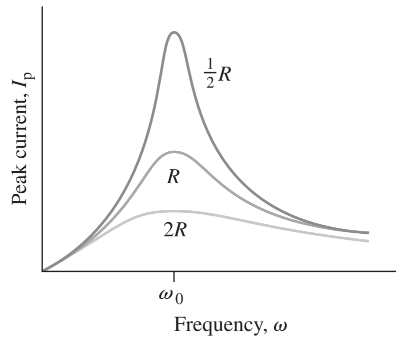


FIGURE 28.17 Resonance curves for an *RLC* circuit with three different resistances.

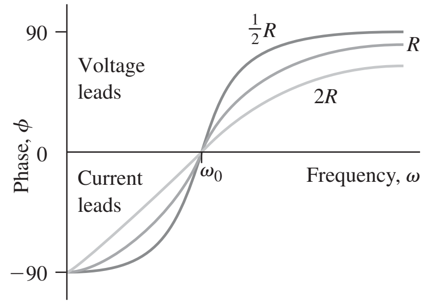


FIGURE 28.18 Phase relations for the *RLC* circuits whose resonance curves are shown in Fig. 28.17.

✓TIP Phase Matters

You can't analyze AC circuits by treating resistors, capacitors, and inductors all as "resistors" with resistances R, X_C, and X_L. That's because each component has a different phase relation between current and voltage. Phasor diagrams correctly account for these relations, which show up in the minus sign joining capacitive and inductive reactance, and in the Pythagorean addition of resistance and reactance in Fig. 28.16 and Equation 28.12.

EXAMPLE 28.4 An *RLC* Circuit: Designing a Loudspeaker System

Current flows to the midrange speaker in a loudspeaker system through a 2.2-mH inductor in series with a capacitor. (a) What should the capacitance be so that a given voltage produces the greatest current at 1 kHz? (b) If the same voltage produces half this current at 618 Hz, what is the speaker's resistance? (c) If the peak output voltage of the amplifier is 20 V, what will the peak capacitor voltage be at 1 kHz?

INTERPRET This is a problem about the peak current and voltage in a series *RLC* circuit, where we identify the speaker as R and the amplifier as the generator in Fig. 28.14.

DEVELOP The peak current is at the resonant frequency of Equation 28.10, $\omega = 1/\sqrt{LC}$, so in (a) we can solve this equation for C. Equation 28.12 relates peak voltage and current to the component values and the frequency, so in (b) we can solve for R. In (c) we'll need to find the current at 1 kHz, and then use Equation 28.5, $I_p = V_p/X_C$, which relates peak voltage and current in a capacitor, to find V_{Cp}.

EVALUATE (a) We solve Equation 28.10 for C, using $\omega = 2\pi f$: $C = 1/[(2\pi f)^2 L]$; with $f = 1000\text{ Hz}$ and $L = 2.2\text{ mH}$, this gives $C = 11.5\ \mu\text{F}$. (b) Equation 28.12 shows that we'll have half the peak current when Z is twice its value $Z = R$ at resonance. So we want $Z = \sqrt{R^2 + (X_L - X_C)^2} = 2R$. Squaring and solving for R give

$$R = \frac{1}{\sqrt{3}}\left|\omega L - \frac{1}{\omega C}\right| = 8.0\ \Omega$$

where we used $X_L = \omega L$ and $X_C = 1/\omega C$ for the reactances. For (c), note that the impedance at the 1-kHz resonant frequency is just R, so the peak current is $I_p = V_p/R$. Then Equation 28.5 gives the peak capacitor voltage:

$$V_{Cp} = I_p X_C = \left(\frac{V_p}{R}\right)\left(\frac{1}{\omega C}\right) = 35\text{ V}$$

where V_p is the 20-V peak voltage applied to the circuit and $\omega = 2\pi f$ with $f = 1$ kHz. (This answer is for the 1-kHz resonant frequency; Problem 76 shows that V_{Cp} is actually somewhat higher than 35 V at frequencies just below resonance.)

ASSESS Our 35-V answer for (c) is greater than the 20-V peak output of the amplifier, so how can it be right? Remember that there's another source of emf in the circuit—the inductor, whose emf depends on the rate of change of current. Although the capacitor and inductor voltages cancel at resonance, individually both can be higher than the applied voltage. In this low-Q circuit, the peak capacitor voltage isn't too much higher than the applied voltage, but in high-Q circuits like radio transmitters, capacitors may have to withstand voltages hundreds of times the applied voltage. Incidentally, that 8-Ω answer in (b) is typical of loudspeakers. ■

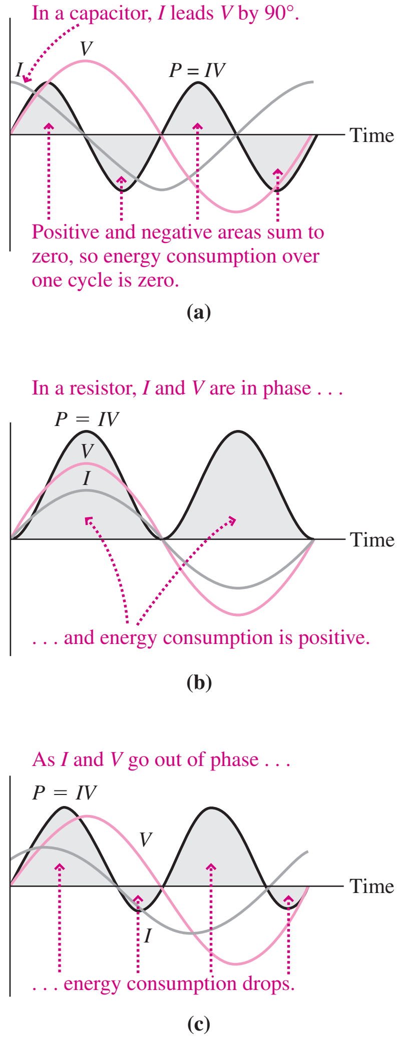


FIGURE 28.19 Energy consumption over one cycle is the area under the curve $P = IV$, with areas below the axis counted as negative.

28.5 Power in AC Circuits

Capacitors and inductors don't dissipate energy; rather, in an AC circuit they alternately store and release it. Therefore, the average power consumption over one cycle is zero in a purely reactive circuit—one containing only capacitance and/or inductance. You can see this mathematically in Fig. 28.19*a*, which shows current, voltage, and power—the product IV—over one cycle for a capacitor. Although the capacitor absorbs energy during part of the cycle—when the power is positive—it returns the same amount later (negative power), giving zero net power over the cycle. That's because current and voltage are out of phase, so their product can be negative or positive at different times. In contrast, a resistor's V and I are in phase (Fig. 28.19*b*), so power is always positive and the resistor takes energy from the circuit. In a circuit containing resistance, capacitance, and inductance, the phase relation between current and voltage depends on the circuit details. Figure 28.19*c* shows a case where I and V are only slightly out of phase; the result is a net power consumption, but less than with a pure resistance.

We can develop a general expression for power in AC circuits by considering the time-average product of voltage and current with arbitrary phase difference ϕ:

$$\langle P \rangle = \langle [I_p \sin(\omega t - \phi)][V_p \sin \omega t] \rangle$$

where $\langle\rangle$ indicates a time average over one cycle. Expanding the current term using a trig identity (see Appendix A) gives

$$\langle P \rangle = I_p V_p \langle (\sin^2 \omega t)(\cos \phi) - (\sin \omega t)(\cos \omega t)(\sin \phi) \rangle$$

The average of $(\sin \omega t)(\cos \omega t)$ is zero, as we've just shown for two signals 90° out of phase. The quantity $\sin^2 \omega t$ swings symmetrically from 0 to 1, so its average is $\frac{1}{2}$. Then we have $\langle P \rangle = \frac{1}{2} I_p V_p \cos \phi$. Writing the peak values as $\sqrt{2}$ times the rms values gives

$$\langle P \rangle = \tfrac{1}{2}\sqrt{2}\, I_{\text{rms}} \sqrt{2}\, V_{\text{rms}} \cos \phi = I_{\text{rms}} V_{\text{rms}} \cos \phi \qquad (28.14)$$

This confirms our earlier graphical arguments. When the voltage and current are in phase, the average power is the product $I_{\text{rms}} V_{\text{rms}}$. But with current and voltage out of phase, the average power is lower; at 90° phase difference it's zero.

The factor $\cos \phi$ is the **power factor**. A purely resistive circuit has power factor 1, while a circuit with only inductance and capacitance has power factor 0. In general, the power factor depends on frequency; in the series *RLC* circuit, for example, it's 1 at resonance but lower at other frequencies.

CONCEPTUAL EXAMPLE 28.1 Managing the Power Factor

You're chief engineer for a power company. Should you strive for a high or a low power factor on your lines?

EVALUATE For a given power, Equation 28.14 shows that the product $I_{\text{rms}} V_{\text{rms}}$ will need to be higher as the power factor drops below 1. Since your equipment operates at fixed voltage, that means more current when $\cos \phi < 1$. Power lost in the lines is $I^2 R$, so you'll have greater transmission loss with a low power factor. Furthermore, you'll risk overloading your lines. So you're best served by keeping the power factor close to 1.

ASSESS Our answer helps explain some real-life power failures: The August 2003 blackout that affected 50 million people in the U.S. and Canada resulted in part from too low a power factor, resulting in an overloaded line that drooped from excessive heating, short-circuited to a tree, and triggered chain-reaction failures throughout the power grid.

MAKING THE CONNECTION Transmission losses on a well-managed electric grid average about 8% of the total power delivered. How does this figure change if the power factor drops from 1 to 0.71?

EVALUATE To get the same power down the line, Equation 28.14 shows that the current must increase by $1/\cos \phi = 1/0.71 = 1.4$. The transmission loss is $I^2 R$, so the loss increases by a factor $1.4^2 = 2$. That will more than double the original 8% loss rate, because the line will need to carry still more power to overcome the loss, and the line will heat more, increasing its resistance.

28.6 Transformers and Power Supplies

A **transformer** is a pair of wire coils, often wound on an iron core to concentrate magnetic flux (Fig. 28.20). A changing current in the **primary** coil results in a changing magnetic flux through the **secondary**, and this induces an emf in the secondary. The induced emf, in turn, drives current in any circuit connected across the secondary. Thus the device transfers electric power between two circuits without direct electrical contact.

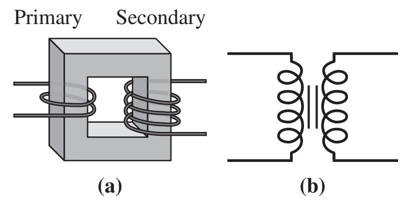


FIGURE 28.20 (a) A transformer consisting of two coils wound on an iron core. (b) Transformer circuit symbol.

The transformer in Fig. 28.20 is a **step-up transformer** because it has more turns in its secondary. Since each turn encircles the same changing magnetic flux, each gets the same induced emf and therefore the emf across the secondary is greater than across the primary. Interchanging primary and secondary in Fig. 28.20 would give a **step-down transformer**. In general, the ratio of the peak (or rms) secondary voltage V_2 to the peak (or rms) primary voltage V_1 is the same as the ratio of turns in the two coils:

$$V_2 = \frac{N_2}{N_1} V_1 \tag{28.15}$$

Aren't we getting something for nothing with a step-up transformer? No. A step-up transformer increases voltage, but not power. An ideal transformer passes all the power supplied to its primary on to the secondary, so $I_1 V_1 = I_2 V_2$. If voltage goes up, current goes down, and vice versa. Real transformers have losses, but good engineering holds these to a few percent of the total power.

Transformers work only with AC because they use electromagnetic induction and therefore require *changing* current. One reason for the near-universal use of AC power is the ease of changing voltage levels (Fig. 28.21). Relatively low voltages are safer for the end user. But since power $P = IV$, using a higher voltage in long-distance transmission means lower current. Power dissipated in the conductors themselves is I^2R, so that in turn means less power lost in transmission. Transformers readily handle the voltage conversions in AC power systems. Changing the voltage from a DC source, in contrast, requires first interrupting the DC to produce a changing current; a car's ignition system is one example.

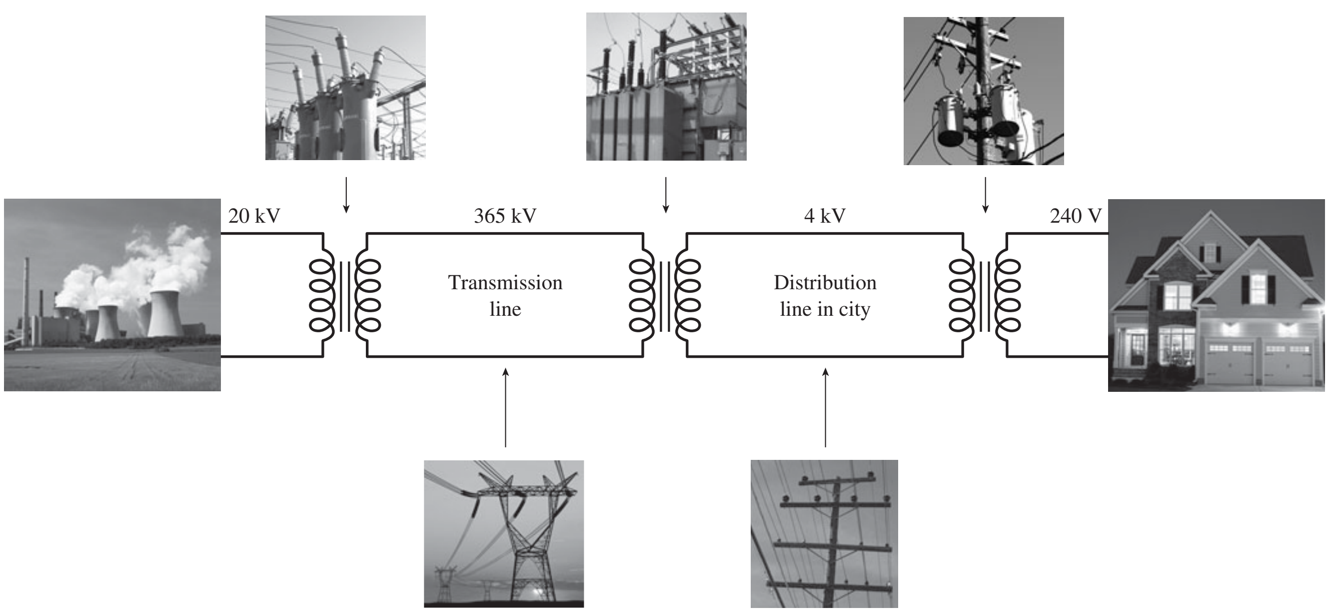


FIGURE 28.21 Transformers change voltage levels throughout the power distribution network.

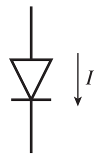


FIGURE 28.22 Circuit symbol for a diode, with preferred current direction indicated.

Direct-Current Power Supplies

Lightbulbs and heaters work equally well on AC or DC, but electronic equipment requires DC. In Chapter 24 we found that a junction between *P*- and *N*-type semiconductors passes current in one direction but not the other. A **diode** is a *PN* junction that serves as a "one-way valve" for electric current. An ideal diode acts like a short circuit in the preferred direction, and like an open circuit in the opposite direction (Fig. 28.22).

Figure 28.23*a* shows a DC power supply using a transformer, diode, and capacitor, and delivering power to a load symbolized by the resistor *R*. The transformer steps the voltage to the desired level, while the diode passes current only in its preferred direction, "chopping off" the negative half of the AC cycle. The capacitor smoothes, or **filters**, the remaining half to produce nearly steady DC. Figure 28.23*b* shows how this works: As the AC voltage rises, the capacitor charges rapidly through the low resistance of the diode in its "on" state. But the diode "turns off" when the AC voltage drops, leaving only the resistor as a discharge path for the capacitor. If the *RC* time constant is long enough—much longer than the typical 1/60-s AC cycle—then the capacitor voltage hardly drops before the next cycle again sends in a surge of charge. Large capacitors are expensive, so practical power supplies often use additional filtering and voltage regulation involving semiconductor devices.

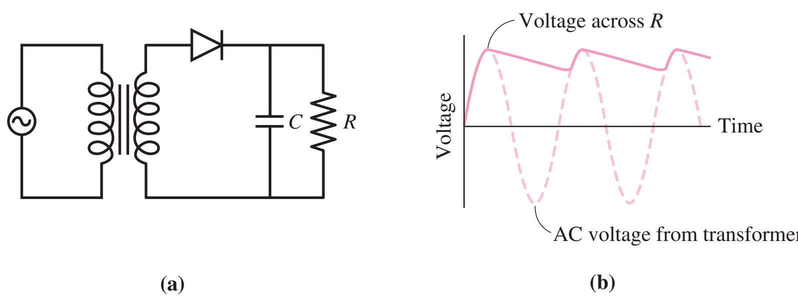


FIGURE 28.23 (a) A simple DC power supply using a diode and capacitive filter. (b) Voltage across *R* exhibits a variation called *ripple* as the capacitor discharges slightly between cycles. A practical power supply would use a larger capacitor, resulting in less ripple.

CHAPTER 28 SUMMARY

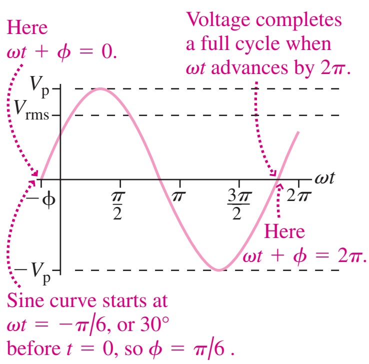
Big Picture

The big idea here is **alternating current** (AC), which in its simplest form exhibits sinusoidal variation in current and voltage. Resistors respond to AC as to DC, with current directly proportional to voltage. In capacitors and inductors, the current–voltage relation depends on frequency, and the current and voltage are out of phase.

Key Concepts and Equations

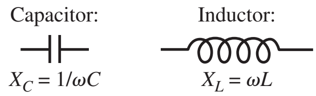
An AC voltage or current is characterized by its amplitude (rms or peak value), its frequency, and phase:

$$V = V_p \sin(\omega t + \phi)$$



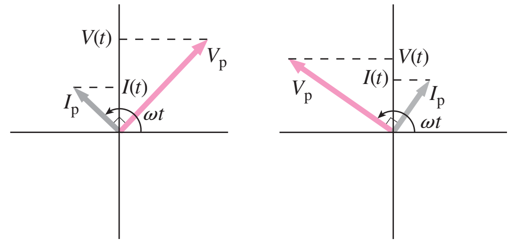
Reactance, X, characterizes the relation between peak (or rms) current and voltage in capacitors and inductors:

$$I_p = \frac{V_p}{X}$$



where $\omega = 2\pi f$ is the angular frequency of the AC voltage and current.

Phasors are arrows used to describe time-varying AC quantities. They rotate with angular velocity equal to the angular frequency ω.



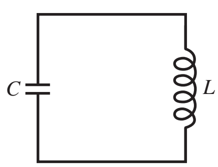
Capacitor: Current leads voltage by 90°.

Inductor: Current lags voltage by 90°.

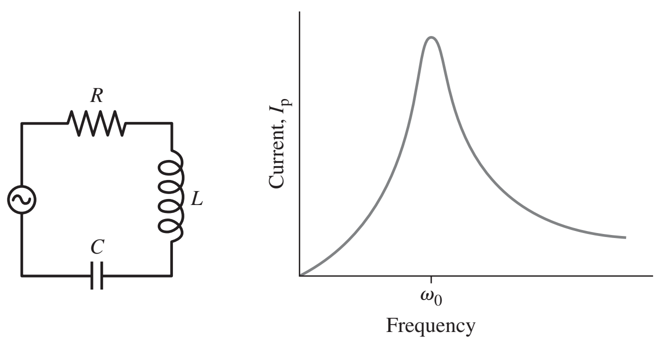
Applications

In an ***LC* circuit**, energy oscillates between electric and magnetic forms with frequency

$$\omega_0 = \frac{1}{\sqrt{LC}}$$



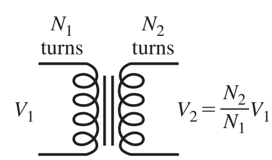
In a series ***RLC* circuit**, capacitor and inductor voltages cancel at the **resonant frequency**, ω_0. Here the circuit exhibits the minimum **impedance**, $Z = \sqrt{R^2 + (X_L - X_C)^2}$, and passes the maximum current. The phase difference between voltage and current is $\tan\phi = (X_L - X_C)/R$.



The average power in an AC circuit depends on the cosine of the phase difference, also called the **power factor**:

$$\langle P \rangle = I_{rms} V_{rms} \cos\phi$$

Transformers use electromagnetic induction to change voltage levels, transferring electric power between two circuits. **Diodes** and capacitive filters change AC to DC.



For Thought and Discussion

1. Two AC signals have the same amplitude but different frequencies. Are their rms amplitudes the same?
2. What's meant by the statement, "A capacitor acts like a DC open circuit"?
3. There's an insulating gap between capacitor plates, so how can current flow in an AC circuit containing a capacitor?
4. Why does it make sense that inductive reactance increases with frequency?
5. The same AC voltage appears across a capacitor and a resistor, and the same rms current flows in each. Is the power dissipation the same in each?
6. When a particular inductor and capacitor are connected across the same AC voltage, the current in the inductor is higher than in the capacitor. Is this true at all frequencies?
7. An inductor and capacitor are connected in series across an AC generator, and the voltage across the inductor is higher than across the capacitor. Is the generator frequency above or below resonance?
8. When the capacitor voltage in an undriven LC circuit reaches zero, why don't the oscillations stop?
9. Why is Equation 28.5 not a full description of the relation between voltage and current in a capacitor?
10. The applied voltage in a series RLC circuit lags the current. Is the frequency above or below resonance?
11. The voltage across two components in series is zero. Is it possible that the voltages across the individual components *aren't* zero? Give an example.
12. If you measure the rms voltages across the resistor, capacitor, and inductor in a series RLC circuit, will they add to the rms generator voltage?
13. A step-up transformer increases voltage, or energy per unit charge. Why doesn't this violate energy conservation?

Exercises and Problems

Exercises

Section 28.1 Alternating Current

14. Much of Europe uses AC power at 230 V rms and 50 Hz. Express this AC voltage in the form of Equation 28.3, taking $\phi_V = 0$.
15. An industrial electric motor runs at 208 V rms and 400 Hz. What are (a) the peak voltage and (b) the angular frequency?
16. An AC current is given by $I = 495\sin(9.43t)$, with I in mA and t in ms. Find (a) the rms current and (b) the frequency in Hz.
17. What are the phase constants for the signals in Fig. 28.24?

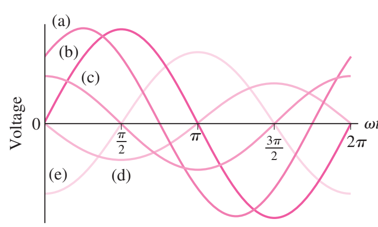


FIGURE 28.24 Exercise 17

Section 28.2 Circuit Elements in AC Circuits

18. Find the rms current in a 1.0-μF capacitor connected across 120-V rms, 60-Hz AC power.
19. A 470-Ω resistor, 10-μF capacitor, and 750-mH inductor are each connected across 6.3-V rms, 60-Hz AC power. Find the rms current in each.
20. Find the reactance of a 3.3-μF capacitor at (a) 60 Hz, (b) 1.0 kHz, and (c) 20 kHz.
21. A 15-μF capacitor carries 1.4 A rms. What's its minimum safe voltage rating if the frequency is (a) 60 Hz and (b) 1.0 kHz?
22. A capacitor and a 1.8-kΩ resistor pass the same current when connected across 60-Hz power. Find the capacitance.
23. A 50-mH inductor is connected across a 10-V rms AC generator, and a 2.0-mA rms current flows. What's the generator frequency?

Section 28.3 *LC* Circuits

24. Find the resonant frequency of an LC circuit consisting of a 0.22-μF capacitor and a 1.7-mH inductor.
25. An LC circuit with $C = 18$ mF undergoes oscillations with period 2.4 s. Find the inductance.
26. Your sister who's building the radio (Chapter 27 Problem 20) wants to use a variable capacitor with her toilet-paper-tube inductor to span the AM radio band (550–1600 kHz). What capacitance range do you suggest?
27. An LC circuit with a 20-μF capacitor oscillates with period 5.0 ms. The peak current is 25 mA. Find (a) the inductance and (b) the peak voltage.
28. Your university's FM station broadcasts at 89.5 MHz. The LC circuit that establishes this frequency has a 47-pF capacitor. What's the corresponding inductance?

Section 28.4 Driven *RLC* Circuits and Resonance

29. A series RLC circuit has $R = 18$ kΩ, $L = 20$ mH, and resonates at 4.0 kHz. (a) What's the capacitance? (b) Find the circuit's impedance at resonance and (c) at 3.0 kHz.
30. Find the impedance at 10 kHz of a circuit consisting of a 1.5-kΩ resistor, 5.0-μF capacitor, and 50-mH inductor in series.
31. A series RLC circuit has $R = 18$ kΩ, $C = 14$ μF, and $L = 0.20$ H. (a) At what frequency is its impedance lowest? (b) What's the impedance at this frequency?
32. If the peak voltage applied to produce the curves in Fig. 28.17 is 100 V, and if $R = 10$ kΩ, what are the peak currents at resonance for the three curves shown?

Section 28.5 Power in AC Circuits
Section 28.6 Transformers and Power Supplies

33. An electric drill draws 4.6 A rms at 120 V rms. If the current lags the voltage by 25°, what's the drill's power consumption?
34. A 40-W fluorescent lamp has power factor 0.85 and operates from the 120-V rms AC power line. How much current does it draw?
35. An electric water heater draws 20 A rms at 240 V rms and is purely resistive. An AC motor has the same current and voltage, but inductance causes the voltage to lead the current by 20°. Find the power consumption in each device.
36. BIO For safety, medical equipment connected to patients is often powered by an *isolation transformer*, whose primary is connected to 120-V AC power and whose secondary delivers 120-V power. What's the turns ratio of such a transformer?
37. You're planning to study in Europe, and you want a transformer designed to step 230-V European power down to 120 V needed to operate your stereo. (a) If the transformer's primary has 460 turns, how many should the secondary have? (b) You can save money with a transformer whose maximum primary current is 1.5 A. If your stereo draws a maximum of 3.3 A, will this transformer work?

Problems

38. (a) A 2.2-H inductor is connected across 120-V rms, 60-Hz power. Find the rms inductor current. (b) Repeat if the same inductor is connected across the 230-V rms, 50-Hz power commonly used in Europe.
39. A 2.0-μF capacitor has 1.0-kΩ reactance. (a) What's the frequency of the applied voltage? (b) What inductance would give the same reactance at this frequency? (c) How would the reactances compare if the frequency were doubled?
40. Show that the unit of both capacitive and inductive reactance is the ohm.
41. **BIO** Electrical activity in the human brain results in AC signals at various frequencies. *Electroencephalography* (EEG) analyzes these signals to provide information on brain functions and to detect abnormalities such as epilepsy. One pattern is the *alpha wave*, with frequency in the range of 8–10 Hz. A particular alpha wave has frequency 10 Hz and amplitude 32 μV rms measured at the scalp. Express this signal in the form of Equation 28.3, assuming zero phase constant.
42. A 1.2-μF capacitor is connected across a generator whose output is $V = V_p \sin 2\pi f t$, with $V_p = 22$ V, $f = 60$ Hz, and t in seconds. Find (a) the peak current and the magnitudes of (b) the voltage and (c) the current at $t = 6.5$ ms.
43. At 10 kHz an inductor has 10 times the reactance of a capacitor. At what frequency will their reactances be equal?
44. A 0.75-H inductor is in series with a fluorescent lamp, and the combination is across 120-V rms, 60-Hz power. If the rms inductor voltage is 90 V, what's the rms lamp current?
45. A 2.2-nF capacitor and one of unknown capacitance are in parallel across a 10-V rms sine-wave generator. At 1.0 kHz, the generator supplies a total current of 3.4 mA rms. The generator frequency is then decreased until the rms current drops to 1.2 mA. Find (a) the unknown capacitance and (b) the lower frequency.
46. **BIO** Connections to the body for electrocardiography (ECG) and electroencephalography (EEG) are normally made with metal electrodes and conductive gels to ensure good electrical contact. An alternative is the *capacitively coupled noncontact electrode*, which uses a conductor near but not contacting the skin, to form a capacitor. Clothing can serve as the capacitor's insulation, eliminating skin contact. A particular EEG instrument calls for capacitive electrodes with maximum reactance 10 MΩ at a typical EEG beta wave frequency of 25 Hz. What's the minimum electrode capacitance?
47. The FM radio band covers the frequency range 88–108 MHz. If the variable capacitor in an FM receiver ranges from 10.9 pF to 16.4 pF, what inductor should be used to make an *LC* circuit whose resonant frequency spans the FM band?
48. An *LC* circuit includes a 0.025-μF capacitor and a 340-μH inductor. (a) If the peak capacitor voltage is 190 V, what's the peak inductor current? (b) How long after the voltage peak does the current peak occur?
49. One-eighth of a cycle after the capacitor in an *LC* circuit is fully charged, what are the following as fractions of their peak values: (a) capacitor charge, (b) energy in the capacitor, (c) inductor current, (d) energy in the inductor?
50. The 2000-μF capacitor in Fig. 28.25 is initially charged to 200 V. (a) Describe how you would manipulate switches *A* and *B* to transfer all the energy from the 2000-μF capacitor to the 500-μF capacitor. Include the times you would throw the switches. (b) What will be the voltage across the 500-μF capacitor once you've finished?

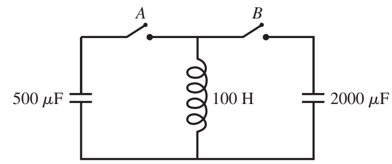


FIGURE 28.25 Problem 50

51. A damped *LC* circuit consists of a 0.15-μF capacitor and a 20-mH inductor with resistance 1.6 Ω. How many oscillation cycles will occur before the peak capacitor voltage drops to half its initial value?
52. A damped *RLC* circuit includes a 5.0-Ω resistor and a 100-mH inductor. If half the initial energy is lost after 15 cycles, what's the capacitance?
53. An *RLC* circuit includes a 1.5-H inductor and a 250-μF capacitor rated at 400 V. The circuit is connected across a sine-wave generator with $V_p = 32$ V. What minimum resistance will ensure that the capacitor voltage does not exceed its rated value when the circuit is at resonance?
54. You're asked to experiment with a series *RLC* circuit consisting of a 10-Ω resistor, 50-mH inductor, and 1.5-μF capacitor rated at 1200 V. You're to apply a sinusoidal AC voltage peaking at 100 V. But you're worried there might be a chance you'll exceed the capacitor's rated voltage. Your lab partner claims this can't happen, since the capacitor rating is 12 times the peak voltage of the AC source. Who's right? To find out, plot the peak capacitor voltage as a function of frequency. Is there a frequency range you should avoid?
55. Figure 28.26 shows the phasor diagram for an *RLC* circuit. (a) Is the driving frequency above or below resonance? (b) Complete the diagram by adding the applied voltage phasor, and from your diagram determine the phase difference between applied voltage and current.

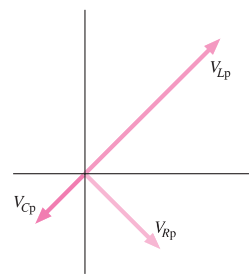


FIGURE 28.26 Problem 55

56. An AC voltage of fixed amplitude is applied across a series *RLC* circuit. The components are such that the current at half the resonant frequency is half the current at resonance. Show that the current at twice the resonant frequency is also half the current at resonance.
57. A series *RLC* circuit has resistance 100 Ω and impedance 300 Ω. (a) What's the power factor? (b) If the rms current is 200 mA, what's the power dissipation?
58. A series *RLC* circuit has power factor 0.80 and impedance 100 Ω at 60 Hz. (a) What's the resistance? (b) If the inductance is 0.10 H, what's the resonant frequency?

59. You're Chief Financial Officer for a power company, and you consult your engineering department in an effort to minimize power-line losses. Your power plant produces 60-Hz power at 365 kV rms and 200 A rms, and delivers it via transmission lines with total resistance 100 Ω. You ask the engineers for the percentage of power that's lost. They reply that it depends on the power factor. What's the percentage loss for power factors of (a) 1.0 and (b) 0.60?
60. A car-battery charger runs off the 120-V rms AC power line and supplies 10-A DC at 14 V. (a) If the charger is 80% efficient in converting the line power to the DC power it supplies to the battery, how much current does it draw from the AC line? (b) If electricity costs 9.5¢/kWh, how much does it cost to run the charger for 10 hours if the power factor is 1?
61. A power supply like that of Fig. 28.23 is supposed to deliver 22-V DC at a maximum current of 150 mA. The transformer's peak output voltage can charge the capacitor to a full 22 V, and the primary is supplied with 60-Hz AC. What capacitance will ensure that the output voltage stays within 3% of the rated 22 V?
62. An *RLC* circuit includes a 3.3-μF capacitor and a 27-mH inductor. The capacitor is charged to 35 V, and the circuit begins oscillating. Ten full cycles later the capacitor voltage peaks at 28 V. Find the resistance.
63. A series *RLC* circuit with $R = 1.3\ \Omega$, $L = 27$ mH, and $C = 0.33\ \mu$F is connected across a sine-wave generator. If the capacitor's peak voltage rating is 600 V, what's the maximum safe value for the generator's peak output voltage when it's tuned to resonance?
64. Differentiate Equation 28.9 to find the current in the *LC* circuit, and use $q = CV$ to find the voltage. From these, obtain the electric energy in the capacitor and the magnetic energy in the inductor, and sum to show that the total energy remains constant. (*Hint:* You'll need Equation 28.10 and a familiar trig identity.)
65. Find a second frequency where the speaker current in Example 28.4 has half its maximum value.
66. A sine-wave generator delivers a signal whose peak voltage is independent of frequency. Two identical capacitors are connected in parallel across the generator, and the generator supplies a peak current I_p at frequency f_1. The capacitors are then connected in series across the generator. What generator frequency will bring the current back to I_p?
67. Two capacitors are connected in parallel across a 10-V rms, 10-kHz sine-wave generator, and the generator supplies a total rms current of 30 mA. With capacitors rewired in series, the rms generator current drops to 5.5 mA. Find the two capacitances.
68. An *LC* circuit starts at $t = 0$ with its 2.0-mF capacitor at its peak voltage of 14 V. At $t = 35$ ms the voltage drops to 8.5 V. (a) What's the peak current? (b) When will the peak current occur?
69. A "black box" has two input connections and two output connections. With a 12-V rms, 60-Hz sine wave across the inputs, the output is a 6.0-V, 60-Hz sine wave leading the input voltage by 45°. Design a circuit that could be in the "black box."
70. A series *RLC* circuit with $R = 47\ \Omega$, $L = 250$ mH, and $C = 4.0\ \mu$F is connected across a sine-wave generator whose peak output voltage is independent of frequency. Find the frequency range over which the peak current will exceed half its value at resonance.
71. A sine-wave generator with 20-V peak output is applied across a series *RLC* circuit. At the resonant frequency of 2.0 kHz, the peak current is 50 mA; at 1.0 kHz, it's 15 mA. Find *R*, *L*, and *C*.
72. Use phasor analysis to show that the parallel *RLC* circuit of Fig. 28.27 has impedance

$$Z = \left[\frac{1}{R^2} + \left(\frac{1}{X_L} - \frac{1}{X_C}\right)^2\right]^{-1/2}$$

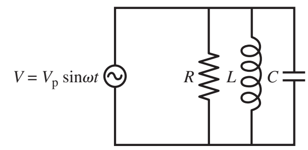


FIGURE 28.27 Problem 72

73. For *RLC* circuits in which the resistance isn't too high, the *Q* factor may be defined as the ratio of the resonant frequency to the difference between the two frequencies where the power dissipated in the circuit is half the power dissipated at resonance. Using suitable approximations, show that this definition leads to $Q = \omega_0 L/R$, with ω_0 the resonant frequency.
74. A *triangle wave* swings linearly between voltages $-V_p$ and $+V_p$. Show that the rms voltage of a triangle wave is $V_p/\sqrt{3}$.
75. Substitute the expression for $q(t)$ in Equation 28.11 into the differential equation for an *LC* circuit with resistance, and find an expression for the angular frequency of the damped oscillations in terms of *R*, *L*, and *C*.
76. Although the maximum current flows in the speaker circuit of Example 28.4 at the 1-kHz resonant frequency, the peak voltage across the capacitor is a maximum at a somewhat lower frequency. Find that frequency and the corresponding peak voltage.
77. You're concerned about a circuit that will be used in a remote communications installation. The series *RLC* circuit with $R = 5.5\ \Omega$, $L = 180$ mH, and $C = 0.12\ \mu$F is connected across a sine-wave generator. The inductor can safely handle 1.5 A of current. The peak generator output when it's tuned to resonance will be 8.0 V. Will the inductor current stay within a safe limit?
78. Your professor tells you about the days before digital computers when engineers used electric circuits to model mechanical systems. Suppose a 5.0-kg mass is connected to a spring with $k = 1.44$ kN/m. This is then modeled by an *LC* circuit with $L = 2.5$ H. What should *C* be in order for the *LC* circuit to have the same resonant frequency as the mass–spring system?

Passage Problems

BIO A *filter* is a circuit designed to pass AC signals in some frequency range and to attenuate others. Common filters include *low-pass filters*, which allow low-frequency signals to pass but attenuate high frequencies; *high-pass filters*, which do the opposite; and *band-pass* filters, which pass a range of frequencies while attenuating signals with frequencies outside the band. Filters are widely used in electronics. Applications include tone and equalizer controls in audio equipment; filters to separate nearby frequencies at cell phone towers; and filters to eliminate unwanted electrical noise in biomedical instruments such as electrocardiographs. A simple design for an *RC* filter is shown in Fig. 28.28.

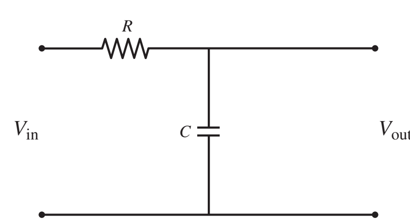


FIGURE 28.28 An *RC* filter (Passage Problems 79–82)

79. The circuit shown in Fig. 28.28 is
  a. a low-pass filter.
  b. a high-pass filter.
  c. a band-pass filter.
  d. impossible to tell without knowing the component values.

80. When the angular frequency ω of the input voltage V_{in} is such that the capacitor's reactance is equal to the resistance, the output voltage is
  a. $V_{in}/4$.
  b. $V_{in}/2$.
  c. $V_{in}/\sqrt{2}$.
  d. $2V_{in}$.

81. The circuit of Fig. 28.28
  a. exhibits resonance at frequency $\omega = 1/RC$.
  b. exhibits resonance at frequency $\omega = 1/\sqrt{RC}$.
  c. produces an output voltage whose frequency differs from that of the input.
  d. produces an output voltage whose phase differs from that of the input.

82. If you replace the capacitor in Fig. 28.28 with an inductor, the circuit
  a. continues to function as before.
  b. becomes the opposite kind of filter.
  c. produces zero output voltage because the inductor is a short circuit.
  d. produces an output voltage that exceeds the input voltage.

Answers to Chapter Questions

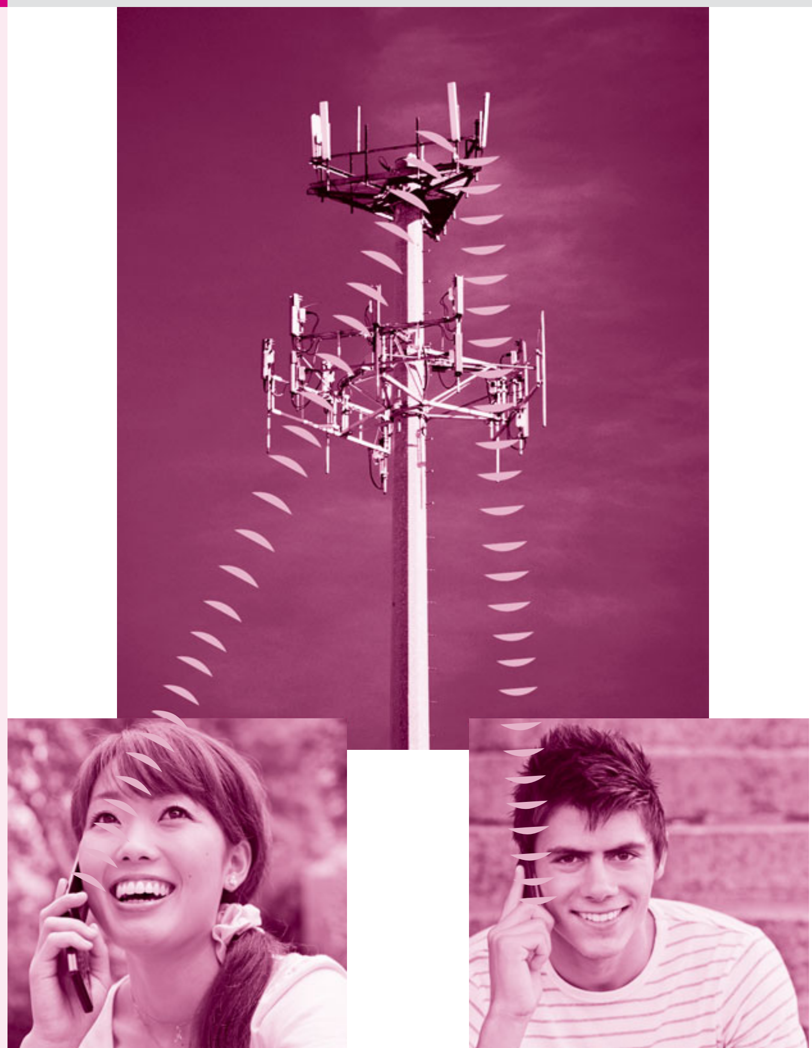
Answer to Chapter Opening Question

Electromagnetic induction lets us change the voltage levels in circuits, making for efficient power transmission but safe end-use voltages.

Answers to GOT IT? Questions

28.1. The capacitor will carry twice as much current because its reactance goes down; the inductor will carry only half as much because its reactance doubles.
28.2. (b), by a factor of 10^{-4}.
28.3. (a), because the inductor's reactance must be greater.

29 Maxwell's Equations and Electromagnetic Waves



How does a conversation travel between cell phones?

New Concepts, New Skills

By the end of this chapter you should be able to

- State both mathematically and in words the four fundamental equations that govern all electromagnetic phenomena (29.1, 29.3).
- Describe electromagnetic waves in terms of frequency, wavelength, amplitude, speed, and polarization (29.4, 29.5).
- Recognize the wide range of frequencies and wavelengths that make up the electromagnetic spectrum (29.6).
- Explain how electromagnetic waves are produced (29.7).
- Calculate energy and momentum in electromagnetic waves (29.8).

Connecting Your Knowledge

- This chapter builds on the four fundamental laws of electromagnetism as we've introduced them throughout Part 4:
  - Gauss's law for electricity (21.3)
  - Gauss's law for magnetism (26.5)
  - Ampère's law (26.8, so far for steady currents only)
  - Faraday's law for the induced electric field (27.6)
- We'll use the mathematical description of waves (14.2); you may also want to brush up on differentiating sines and cosines.

At this point we've introduced the four fundamental laws of electromagnetism—Gauss's law for electricity, Gauss's law for magnetism, Ampère's law, and Faraday's law—that govern the behavior of electric and magnetic fields throughout the universe. We've seen how these laws describe the electric and magnetic interactions that make matter act as it does, and we've explored practical electromagnetic devices. Here we extend the fundamental laws to their most general form and show how they predict the existence of electromagnetic waves. These include the visible light, radio, microwaves, X rays, ultraviolet, and infrared with which we see, communicate, cook our food, diagnose diseases, learn about the universe, and perform myriad other tasks from mundane to profound.

29.1 The Four Laws of Electromagnetism

Table 29.1 summarizes the four laws as we introduced them in earlier chapters. Look at these laws and you'll notice some similarities. On the left-hand sides of the equations, the two laws of Gauss are identical except for the interchanging of $\vec{E}$ and $\vec{B}$; the same is true for Ampère's and Faraday's laws.

The right-hand sides are more different. Gauss's law for electricity involves charge, while Gauss's law for magnetism has zero on the right-hand side. Actually, though, these laws are similar. Since we have no experimental evidence for the existence of isolated magnetic charge, the magnetic charge on the right-hand side of Gauss's law for magnetism is zero. If and when magnetic monopoles are discovered, then the right-hand side of Gauss's law for magnetism would be nonzero for any surface enclosing net magnetic charge.

The right-hand sides of Ampère's and Faraday's laws are distinctly different. Ampère's law has current—the flow of electric charge—as a source of magnetic field. We can understand the absence of a similar term in Faraday's law because we've never observed a flow of magnetic monopoles. If we had such a flow, then we would expect this magnetic current to produce an electric field.

Two of the differences among the laws of electromagnetism would be resolved if we knew for sure that magnetic monopoles exist. That current theories of elementary particles suggest the existence of monopoles is a tantalizing hint that there may be a fuller symmetry between electric and magnetic phenomena.

Table 29.1 Four Laws of Electromagnetism (still incomplete)

Law	Mathematical Statement	What It Says
Gauss for $\vec{E}$	$\oint \vec{E}\cdot d\vec{A} = \dfrac{q}{\epsilon_0}$	How charges produce electric field; field lines begin and end on charges.
Gauss for $\vec{B}$	$\oint \vec{B}\cdot d\vec{A} = 0$	No magnetic charge; magnetic field lines don't begin or end.
Faraday	$\oint \vec{E}\cdot d\vec{r} = -\dfrac{d\Phi_B}{dt}$	Changing magnetic flux produces electric field.
Ampère (steady currents only)	$\oint \vec{B}\cdot d\vec{r} = \mu_0 I$	Electric current produces magnetic field.

29.2 Ambiguity in Ampère's Law

There's one difference that magnetic monopoles won't resolve. On the right-hand side of Faraday's law is the term $d\Phi_B/dt$ that describes changing magnetic flux as a source of electric field. There's no comparable term in Ampère's law. Are we missing something? Is it possible that a changing electric flux produces a magnetic field? So far, we've given no experimental evidence for such a conjecture. It's suggested only by our sense that the near-symmetry between electricity and magnetism is not a coincidence. If a changing electric flux did produce a magnetic field, just as a changing magnetic flux produces an electric field, then we would expect a term $d\Phi_E/dt$ on the right-hand side of Ampère's law.

When we first stated Ampère's law in Chapter 26, we emphasized that it applied only to *steady* currents. Why that restriction? Figure 29.1 shows a situation in which current is *not* steady—namely, an *RC* circuit. Current in this circuit carries charge onto the capacitor plates. The current gradually decreases to zero as the capacitor charges. While it's flowing, the current should produce a magnetic field. Let's try to use Ampère's law to calculate that field.

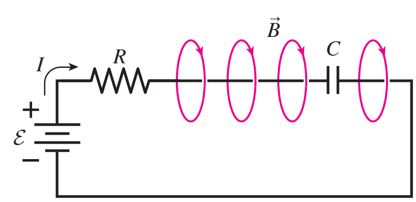


FIGURE 29.1 A charging *RC* circuit, showing some magnetic field lines surrounding current-carrying wire.

Ampère's law says that the line integral of the magnetic field around any closed loop is proportional to the encircled current:

$$\oint \vec{B} \cdot d\vec{r} = \mu_0 I$$

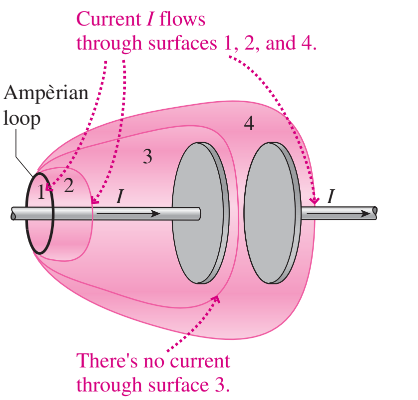


FIGURE 29.2 Four surfaces bounded by the same circular Ampèrian loop. Surface 1 is a flat, circular disk. The others are like soap bubbles in the process of being blown; they're open at the left end, so if current does pass through a surface, it does so at the right end only.

The encircled current is the current through *any open surface* bounded by the loop. Figure 29.2 shows four such surfaces. The same current flows through surfaces 1, 2, and 4 because a current-carrying wire pierces each surface. But no current pierces surface 3 because it's in the gap between the capacitor plates. Charge flows onto the plates of the capacitor, but it doesn't flow through that gap. So for surfaces 1, 2, and 4 the right-hand side of Ampère's law is $\mu_0 I$, but for surface 3 it's zero. Thus Ampère's law is ambiguous in this case of a changing current.

This ambiguity doesn't arise with steady currents. In an *RC* circuit the steady-state current is zero, and thus the right-hand side of Ampère's law is zero for *any* surface. It's only when currents are changing with time that Ampère's law becomes ambiguous. That's why the form of Ampère's law we've used until now is strictly valid only for steady currents.

Can we extend Ampère's law to cover unsteady currents without affecting its validity in the steady case? Symmetry between Ampère's and Faraday's laws suggests that a changing electric flux might produce a magnetic field. Between the plates of a charging capacitor is an electric field whose magnitude is increasing. That means there's a changing electric flux through surface 3 of Fig. 29.2.

It was the Scottish physicist James Clerk Maxwell who, in about 1860, suggested that a changing electric flux should give rise to a magnetic field. Since that time many experiments, including direct measurement of the magnetic field inside a charging capacitor, have confirmed Maxwell's remarkable insight. Maxwell quantified his idea by introducing a new term into Ampère's law:

$$\oint \vec{B} \cdot d\vec{r} = \mu_0 I + \mu_0 \epsilon_0 \frac{d\Phi_E}{dt} \quad \text{(Ampère's law with Maxwell's modification)} \qquad (29.1)$$

Now there's no ambiguity. The integral is taken around any loop, I is the current through *any* surface bounded by the loop, and Φ_E is the electric flux through that surface. With our charging capacitor, Equation 29.1 gives the same magnetic field no matter which surface we choose. For surfaces 1, 2, and 4 of Fig. 29.2, the current I makes all the contribution to the right-hand side of the equation. For surface 3, the right-hand side of Equation 29.1 comes entirely from the changing electric flux.

Although changing electric flux isn't the same thing as electric current, it has the same effect in producing a magnetic field. For this reason Maxwell called the term $\epsilon_0(d\Phi_E/dt)$ the **displacement current**. The word "displacement" has historical roots that don't provide much physical insight. But "current" is meaningful because displacement current is indistinguishable from real current in producing magnetic fields. Although we developed the idea of displacement current using the specific example of a charging capacitor, we emphasize that Ampère's law in its now complete form (Equation 29.1) is truly universal: *Any* changing electric flux results in a magnetic field. That fact will prove crucial in establishing the existence of electromagnetic waves.

EXAMPLE 29.1 Displacement Current: A Capacitor

A parallel-plate capacitor with plate area A and spacing d is charging at the rate dV/dt. Show that the displacement current is equal to the current in the wires feeding the capacitor.

INTERPRET This is about a comparison between a familiar quantity—current—and a new quantity, namely, displacement current.

DEVELOP We're given the rate at which the capacitor voltage increases. Given that $q = CV$ for a capacitor, we can find the rate of charge buildup—and that's equal to the current I delivering charge to the capacitor. Equation 29.1 shows that the displacement current is $\epsilon_0(d\Phi_E/dt)$, so we'll need the rate of change of electric flux. A parallel-plate capacitor produces an essentially uniform field $E = V/d$. Since the field is uniform, the electric flux through a surface within the capacitor is simply the field strength times the plate area.

EVALUATE For the current, we differentiate the capacitor relation $q = CV$ to get $dq/dt = I = C\,dV/dt$. For the flux, we multiply the electric field by the plate area: $\Phi_E = EA = VA/d$. The rate of change of flux is then $d\Phi_E/dt = (A/d)(dV/dt)$, so the displacement current becomes

$$I_d = \epsilon_0 \frac{d\Phi_E}{dt} = \frac{\epsilon_0 A}{d}\frac{dV}{dt}$$

But $\epsilon_0 A/d$ is the capacitance of a parallel-plate capacitor (Equation 23.3), so the displacement current is $I_d = C\,dV/dt$, the same as the actual current I.

ASSESS Make sense? It had better be this way, or Ampère's law would still be ambiguous. For any surface pierced by the wire in Fig. 29.2, the only contribution to the right-hand side of Ampère's law is from the current I. For any surface between the capacitor plates, the only contribution is from the displacement current $I_d = \epsilon_0(d\Phi_E/dt)$. For Ampère's law to give the same magnetic field whichever surface we choose, I and I_d had better be the same. ■

29.3 Maxwell's Equations

It was Maxwell's genius to recognize that Ampère's law needed modifying to reflect the symmetry suggested by Faraday's law. To honor Maxwell, the four complete laws of electromagnetism are called **Maxwell's equations**. The complete set of equations, first published in 1864, governs the behavior of electric and magnetic fields everywhere. Table 29.2 summarizes Maxwell's equations.

Table 29.2 Maxwell's Equations

Law	Mathematical Statement	What It Says	Equation Number
Gauss for $\vec{E}$	$\oint \vec{E}\cdot d\vec{A} = \frac{q}{\epsilon_0}$	How charges produce electric field; field lines begin and end on charges.	(29.2)
Gauss for $\vec{B}$	$\oint \vec{B}\cdot d\vec{A} = 0$	No magnetic charge; magnetic field lines don't begin or end.	(29.3)
Faraday	$\oint \vec{E}\cdot d\vec{r} = -\frac{d\Phi_B}{dt}$	Changing magnetic flux produces electric field.	(29.4)
Ampère	$\oint \vec{B}\cdot d\vec{r} = \mu_0 I + \mu_0\epsilon_0\frac{d\Phi_E}{dt}$	Electric current and changing electric flux produce magnetic field.	(29.5)

These four compact statements are all it takes to describe classical electromagnetic phenomena. Everything electric or magnetic that we've considered and will consider—from polar molecules to electric current; resistors, capacitors, inductors, and transistors; solar flares and cell membranes; electric generators and thunderstorms; computers, iPods, and the northern lights—can be described using Maxwell's equations. And despite this wealth of phenomena, we have yet to discuss a most important manifestation of electromagnetism—namely, electromagnetic waves. We've put off waves until now because they depend crucially on Maxwell's extension of Ampère's law. It's easiest to understand electromagnetic waves when they propagate through empty space, so we'll first simplify Maxwell's equations for the case of a vacuum.

Maxwell's Equations in Vacuum

To express Maxwell's equations in vacuum, we simply remove all reference to matter—that is, to electric charge and current:

$$\oint \vec{E}\cdot d\vec{A} = 0 \quad (\text{Gauss}, \vec{E}) \qquad (29.6)$$

$$\oint \vec{B}\cdot d\vec{A} = 0 \quad (\text{Gauss}, \vec{B}) \qquad (29.7)$$

$$\oint \vec{E}\cdot d\vec{r} = -\frac{d\Phi_B}{dt} \quad (\text{Faraday}) \qquad (29.8)$$

$$\oint \vec{B}\cdot d\vec{r} = \mu_0\epsilon_0\frac{d\Phi_E}{dt} \quad (\text{Ampère}) \qquad (29.9)$$

In vacuum the symmetry is complete, with electric and magnetic fields appearing on an equal footing. With charge and current absent, the only source of either field is a change in the other field—as shown by the time derivatives on the right-hand sides of Faraday's and Ampère's laws.

29.4 Electromagnetic Waves

Faraday's law shows that a changing magnetic field induces an electric field. Ampère's law shows that a changing electric field induces a magnetic field. Together, the two suggest the possibility of **electromagnetic waves**, in which each type of field continuously induces the other, resulting in an electromagnetic disturbance that propagates through space as a wave. We'll now confirm this suggestion with a rigorous demonstration, directly from Maxwell's equations, that electromagnetic waves are indeed possible. In the process we'll discover the properties of electromagnetic waves and come to a deep understanding of the nature of light.

A Plane Electromagnetic Wave

Here we describe the simplest type of electromagnetic wave—a plane wave in vacuum. A plane wave's properties don't vary in directions perpendicular to the wave propagation, so its wavefronts are infinite planes. A plane wave is an approximation to the more realistic case of a spherical wave expanding from a localized source, and it's a good approximation at distances from the wave source that are large compared with the wavelength. Light waves from the Sun, for example, or radio waves miles from the transmitter are essentially plane waves.

In vacuum, it turns out that the electric and magnetic fields of an electromagnetic wave are perpendicular. They're both also perpendicular to the direction of wave propagation—making the electromagnetic wave a transverse wave, as defined in Chapter 14. To be concrete, we'll take the x-direction to be the direction of propagation, the y-direction that of the electric field, and the z-direction that of the magnetic field (Fig. 29.3). We won't prove that a configuration like this is the only one possible for an electromagnetic wave (although in vacuum it is; see Problem 44). What we will do is prove that this configuration satisfies Maxwell's equations—thus showing that such electromagnetic waves are indeed possible. But first we need a mathematical description of our plane electromagnetic wave.

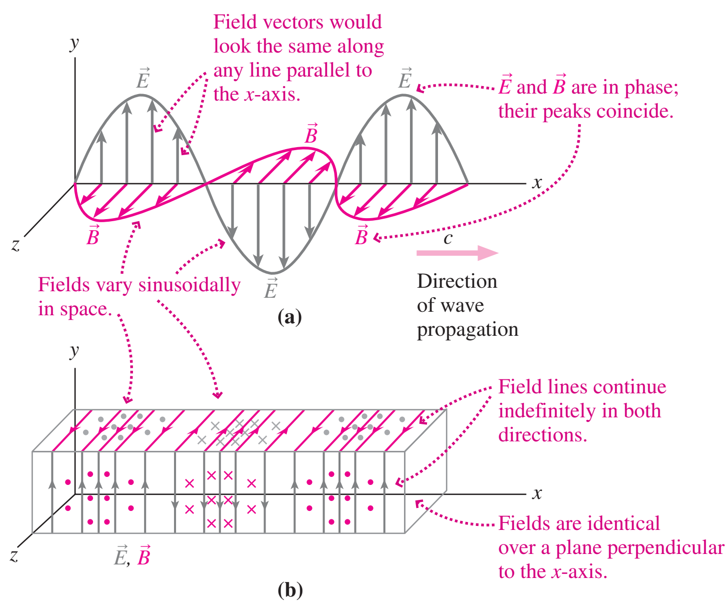


FIGURE 29.3 Fields of a plane electromagnetic wave, shown at a fixed instant of time. (a) Field vectors for points on the x-axis show sinusoidal variation in the fields. (b) A partial representation of the field lines in a rectangular slab. Lines on the facing surfaces of the slab are shown as arrows; lines going through the slab appear as dots or crosses. Spacing of the field lines reflects the sinusoidal variation shown in part (a).

In Chapter 14 we described a sinusoidal wave propagating in the x-direction by a function of the form $A\sin(kx - \omega t)$, where A is the wave amplitude, k the wave number, and ω the angular frequency. For mechanical waves, $A\sin(kx - \omega t)$ described some physical quantity such as the height of a water wave or the pressure variation in a sound wave. In an electromagnetic wave, the corresponding physical quantities are the electric and magnetic fields. It turns out that these two wave fields, though perpendicular, are in phase—meaning that their peaks and troughs coincide, as shown in Fig. 29.3*a*. Having chosen the y-direction for the electric field and z for the magnetic field, we can write the fields of our plane electromagnetic wave as

$$\vec{E}(x,t) = E_p \sin(kx - \omega t)\hat{\jmath} \tag{29.10}$$

and

$$\vec{B}(x,t) = B_p \sin(kx - \omega t)\hat{k} \tag{29.11}$$

where the amplitudes E_p and B_p are constants and where $\hat{\jmath}$ and $\hat{k}$ are unit vectors in the y- and z-directions. Figure 29.3*a* is a "snapshot" of some field vectors of this wave at points on the x-axis, shown at a fixed instant of time. That $\vec{E}$ and $\vec{B}$ are perpendicular is obvious from the figure, as is the fact that they're perpendicular to the propagation direction x. You can also see the sinusoidal variation, as the field vectors get alternately longer, then shorter, then reverse direction, and so on. And you can see that $\vec{E}$ and $\vec{B}$ are in phase because their peaks coincide. We emphasize that Fig. 29.3*a* shows field *vectors* for points on the x-axis only; the fields extend forever throughout space, and because this is a plane wave, a picture of field vectors along any line parallel to the x-axis would look the same.

We can also draw field *lines* for our wave, in contrast to the field vectors of Fig. 29.3*a*. We can't draw complete field lines because they extend forever in both directions. So in Fig. 29.3*b* we've shown the field lines only in a rectangular slab; that's enough to give a picture of what the fields look like everywhere. You should convince yourself that Figs. 29.3*a* and *b* show exactly the same thing—namely, a plane electromagnetic wave described by Equations 29.10 and 29.11. In one case we use field *vectors*, whose lengths are proportional to the field magnitudes, and in the other we use field *lines*, which extend forever and whose spacing indicates the field magnitudes.

We'll now show that the electric and magnetic fields pictured in Fig. 29.3 and described by Equations 29.10 and 29.11 satisfy Maxwell's equations. We've chosen a sinusoidal waveform for our wave fields because of its mathematical simplicity. But the superposition principle holds for electric and magnetic fields, and we know from Section 14.5 that we can represent *any* waveform by superposing sinusoids. So our proof that electromagnetic waves can exist holds for any wave shape. That means we can use electromagnetic waves to communicate the complex waveforms of music, TV images, and computer data.

Gauss's Laws

In vacuum, Gauss's laws for electric and magnetic fields both have zero on the right-hand side, reflecting the absence of charge. That means the electric and magnetic flux through *any* closed surface must be zero, and therefore the field lines can't begin or end. With our plane wave, the field lines shown partially in Fig. 29.3*b* extend straight forever in both directions. So they don't begin or end, and therefore the fields satisfy Gauss's laws.

Faraday's Law

To see that Faraday's law is satisfied, look directly toward the x-y plane in Fig. 29.3*b*. You see electric field lines going up and down and magnetic field lines coming straight in and out, as shown in Fig. 29.4. Consider the small rectangular loop of height h and infinitesimal width dx shown in the figure. Evaluating the line integral of the electric field $\vec{E}$ around this loop, we get no contribution from the short ends at right angles to the field. Going around counterclockwise, we get a contribution $-Eh$ as we go down the left side against the field direction. Then we get a positive contribution going up the right side. Because the field varies with position, the field on the right side of the loop is different from that

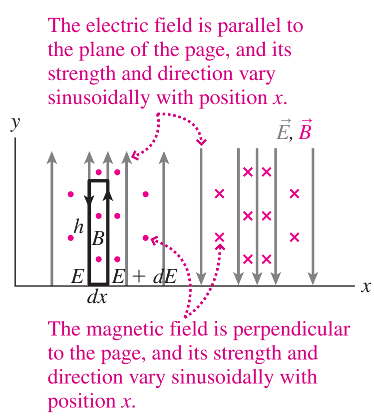


FIGURE 29.4 View of Fig. 29.3*b* in the x-y plane, with a rectangular loop for evaluating the line integral in Faraday's law.

on the left. Let the change be dE, so the field on the right side is $E + dE$, giving a contribution $(E + dE)h$ to the line integral. Then the line integral of $\vec{E}$ around the loop is

$$\oint \vec{E} \cdot d\vec{r} = -Eh + (E + dE)h = h\,dE$$

This nonzero line integral implies an induced electric field. Induced by what? By a changing magnetic flux through the loop. The electric field of the wave arises because of the changing magnetic field of the wave. The area of the loop is $h\,dx$, and the magnetic field $\vec{B}$ is at right angles to this area, so the magnetic flux through the loop is $\Phi_B = Bh\,dx$. The rate of change of flux through the loop is then

$$\frac{d\Phi_B}{dt} = h\,dx\frac{dB}{dt}$$

Faraday's law relates the line integral of the electric field to the rate of change of flux:

$$\oint \vec{E} \cdot d\vec{r} = -\frac{d\Phi_B}{dt}$$

or, using our expressions for the line integral and the rate of change of flux, $h\,dE = -h\,dx(dB/dt)$. Dividing through by $h\,dx$ gives $dE/dx = -dB/dt$. In deriving this equation, we considered changes in E with position at a fixed instant of time. Similarly, the change in B with respect to time is taken at a fixed position. That is, the derivatives are *partial derivatives*—rates of change with respect to one variable while another is held fixed. If you've studied partial derivatives in calculus, you know that the symbol ∂ designates a partial derivative. So our equation $dE/dx = -dB/dt$ should be written with partial derivatives:

$$\frac{\partial E}{\partial x} = -\frac{\partial B}{\partial t} \tag{29.12}$$

This equation—which is Faraday's law applied to our electromagnetic wave—says that the rate at which the electric field changes with *position* depends on the rate at which the magnetic field changes with *time*.

Ampère's Law

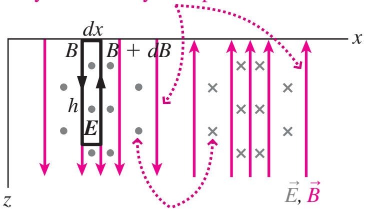


FIGURE 29.5 View of Fig. 29.3*b* in the *x-z* plane, with a rectangular loop for evaluating the line integral in Ampère's law.

Now look at Fig. 29.3*b* from above. You see the magnetic field lines in the *x-z* plane and electric field lines emerging perpendicular to the *x-z* plane (Fig. 29.5). Apply Ampère's law (Equation 29.9) to the rectangle shown. In the line integral there's no contribution from the short sides because they're perpendicular to the field. Going down the left side gives Bh. Going up the right, against the field, gives $-(B + dB)h$, where dB is the change in B across the rectangle. So the line integral in Ampère's law is

$$\oint \vec{B} \cdot d\vec{r} = Bh - (B + dB)h = -h\,dB$$

The electric flux through the rectangle is $Eh\,dx$, so the rate of change of electric flux is

$$\frac{d\Phi_E}{dt} = h\,dx\left(\frac{dE}{dt}\right)$$

Ampère's law relates the line integral of the magnetic field to this time derivative of the electric flux, giving $-h\,dB = \epsilon_0\mu_0 h\,dx(dE/dt)$. Dividing by $h\,dx$ and again using partial derivatives, we have

$$\frac{\partial B}{\partial x} = -\epsilon_0\mu_0\frac{\partial E}{\partial t} \tag{29.13}$$

Equations 29.12 and 29.13—derived from Faraday's and Ampère's laws—express fully the requirements that Maxwell's universal laws of electromagnetism pose on the field structure of Fig. 29.3. Each describes an induced field that arises from the changing of the other field. That other field, in turn, arises from the changing of the first field. Thus we have a self-perpetuating electromagnetic structure, whose fields exist and change without the need for charged matter. If Equations 29.10 and 29.11, which describe the fields in Fig. 29.3, can be made consistent with Equations 29.12 and 29.13, then we'll have shown that our electromagnetic wave satisfies Maxwell's equations and is thus a possible configuration of electric and magnetic fields. An alternative approach, which doesn't require the sinusoidal fields of Equations 29.10 and 29.11, is to show that Equations 29.12 and 29.13 lead to the wave equation that we introduced in Chapter 14. You can explore this approach in Problem 63.

Conditions on the Wave Fields

To see that Equation 29.12 is satisfied, we differentiate the electric field of Equation 29.10 with respect to x and the magnetic field of Equation 29.11 with respect to t:

$$\frac{\partial E}{\partial x} = \frac{\partial}{\partial x}[E_p \sin(kx - \omega t)] = kE_p \cos(kx - \omega t)$$

and

$$\frac{\partial B}{\partial t} = \frac{\partial}{\partial t}[B_p \sin(kx - \omega t)] = -\omega B_p \cos(kx - \omega t)$$

Putting these expressions in for the derivatives in Equation 29.12 gives

$$kE_p \cos(kx - \omega t) = -[-\omega B_p \cos(kx - \omega t)]$$

The cosine cancels, showing that the equation holds if

$$kE_p = \omega B_p \tag{29.14}$$

To see that Equation 29.13 is also satisfied, we differentiate the magnetic field of Equation 29.11 with respect to x and the electric field of Equation 29.10 with respect to t:

$$\frac{\partial B}{\partial x} = kB_p \cos(kx - \omega t) \qquad \text{and} \qquad \frac{\partial E}{\partial t} = -\omega E_p \cos(kx - \omega t)$$

Using these expressions in Equation 29.13 then gives

$$kB_p \cos(kx - \omega t) = -\epsilon_0 \mu_0[-\omega E_p \cos(kx - \omega t)]$$

Again, the cosine cancels, so this equation is satisfied if

$$kB_p = \epsilon_0 \mu_0 \omega E_p \tag{29.15}$$

Our analysis has shown that electromagnetic waves whose form is given by Fig. 29.3 and Equations 29.10 and 29.11 can exist, provided that the amplitudes E_p and B_p, and the frequency ω and wave number k, are related by Equations 29.14 and 29.15. Physically, the existence of these waves is possible because a change in either field—electric or magnetic—induces the other field, giving rise to a self-perpetuating electromagnetic-field structure. Maxwell's theory thus leads to the prediction of an entirely new phenomenon—electromagnetic waves. We'll now explore some properties of these waves.

29.5 Properties of Electromagnetic Waves

Wave Speed

In Chapter 14 we found that the speed of a sinusoidal wave is the ratio of the angular frequency to the wave number: speed $= \omega/k$. To determine the speed of our electromagnetic wave, we use Equation 29.14 to get $E_p = \omega B_p/k$, and then use this expression in Equation 29.15:

$$kB_p = \epsilon_0 \mu_0 \omega E_p = \frac{\epsilon_0 \mu_0 \omega^2 B_p}{k}$$

The amplitude B_p cancels, and we solve for the wave speed ω/k to get

$$\text{wave speed} = \frac{\omega}{k} = \frac{1}{\sqrt{\epsilon_0\mu_0}} \qquad \text{(EM wave speed in vacuum)} \tag{29.16a}$$

This result shows that the speed of an electromagnetic wave in vacuum depends only on the electric and magnetic constants ϵ_0 and μ_0. All electromagnetic waves in vacuum, regardless of frequency or amplitude, share this speed. Although we derived this result for sinusoidal waves, the superposition principle ensures that it holds for any wave shape.

We can easily evaluate the speed given in Equation 29.16a:

$$\frac{1}{\sqrt{\epsilon_0\mu_0}} = \frac{1}{\sqrt{(8.85\times10^{-12}\ \text{C}^2/\text{N}\cdot\text{m}^2)(4\pi\times10^{-7}\ \text{N/A}^2)}} = 3.00\times10^8\ \text{m/s}$$

But this is the speed of light! During the two centuries before Maxwell, scientists had measured light's speed with increasing accuracy. They had also recognized, thanks to Thomas Young's 1801 interference experiment, that light consists of waves. Then, in the 1860s, came Maxwell. Using a theory developed from laboratory experiments on electricity and magnetism, with no reference to optics or light, Maxwell showed how the interplay of electric and magnetic fields results in electromagnetic waves. The wave speed—calculated from the constants ϵ_0 and μ_0—was the known speed of light. Maxwell's conclusion was inescapable: Light is an electromagnetic wave.

Maxwell's identification of light as an electromagnetic phenomenon is a classic example of the unification of knowledge in science. With one simple calculation, Maxwell brought the entire science of optics under the umbrella of electromagnetism. Maxwell's work stands as a crowning intellectual triumph, one whose implications are still expanding our view of the universe.

Maxwell's discovery lets us recast Equation 29.16a in the form

$$\frac{\omega}{k} = c \qquad \text{(EM wave speed in vacuum: the speed of light, } c!\text{)} \tag{29.16b}$$

where $c = 1/\sqrt{\epsilon_0\mu_0}$ is the speed of light. Because $\omega = 2\pi f$ and $k = 2\pi/\lambda$, we can rewrite Equation 29.16b in terms of the more familiar frequency f and wavelength λ as

$$f\lambda = c \qquad \text{(frequency, wavelength, and the speed of light)} \tag{29.16c}$$

As we saw in Chapter 1, the 1983 definition of the meter gives c the exact value 299,792,458 m/s.

Wave Amplitude

The amplitudes E_p and B_p dropped out of our analysis, showing that an electromagnetic wave's speed is independent of amplitude. But the field strengths E and B aren't independent. Using $\omega/k = c$, we can recast Equation 29.14 to show that

$$E = \frac{\omega}{k}B = cB \qquad (E, B \text{ relation in vacuum EM wave}) \tag{29.17}$$

Here we dropped the "peak" subscript because E_p and B_p multiply identical cosine terms in our wave description, so Equation 29.14 applies whether or not we're at the peak field.

Phase, Orientation, and Waves in Matter

The wave of Fig. 29.3 and Equations 29.10 and 29.11 has $\vec{E}$ and $\vec{B}$ in phase in *time*—peaking at the same time—while they're perpendicular in *space* and also perpendicular to the propagation direction. Our derivation of the wave speed used these properties, so we've confirmed that electromagnetic waves in vacuum are transverse waves with $\vec{E}$ and $\vec{B}$ perpendicular and in phase. Specifically, the direction of propagation is that of the cross product $\vec{E} \times \vec{B}$. These geometrical properties also apply to electromagnetic waves in common materials like air and glass. The wave speed in these materials is lower than in vacuum, although for air the difference is minuscule. Electromagnetic waves in more complex materials can have very different properties and propagation speeds.

GOT IT? 29.1 At a particular point the electric field of an electromagnetic wave points in the $+y$-direction, while the magnetic field points in the $-z$-direction. Is the propagation direction (a) x; (b) $-x$; (c) either $+x$ or $-x$ but you can't tell which; (d) $-y$; (e) $+z$; or (f) not along any of the coordinate axes?

EXAMPLE 29.2 Electromagnetic-Wave Properties: Laser Light

A laser beam with wavelength 633 nm is propagating through air in the $+z$-direction. Its electric field is parallel to the x-axis and has amplitude 6.0 kV/m. (a) Find the wave frequency, (b) the amplitude of the magnetic field, and (c) the direction of the magnetic field.

INTERPRET Light is an electromagnetic wave, so the laser beam shares the wave properties we've just discussed. As we noted, its speed in air is nearly the same as in vacuum. Here we're given the wavelength and peak electric field, E_p.

DEVELOP Equation 29.16c, $f\lambda = c$, relates wavelength and frequency, so we'll use that for (a). Equation 29.17, $E = cB$, relates E and B, so we can get (b) from the given value of E_p. For (c), we'll draw a reoriented version of Fig. 29.3*a* to help infer the direction of $\vec{B}$.

EVALUATE (a) Solving for the frequency gives

$$f = c/\lambda = (3.0\times10^8\ \text{m/s})/(633\ \text{nm}) = 4.7\times10^{14}\ \text{Hz}$$

(b) Solving for the magnetic-field amplitude gives $B_p = E_p/c = 20\ \mu\text{T}$. (c) Figure 29.6 shows that with propagation in the z-direction and $\vec{E}$ along the x-axis, $\vec{B}$ must be parallel to the y-axis.

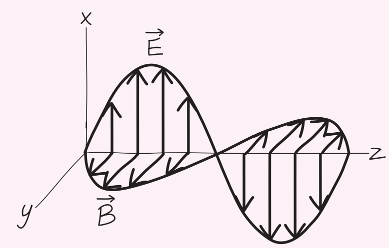


FIGURE 29.6 Reoriented version of Fig. 29.3*a* for Example 29.2. It's important that we still have a right-handed coordinate system, with the x-, y-, and z-axes in the same relation as in Fig. 29.3. Equivalently, the direction of $\vec{E}\times\vec{B}$ is the propagation direction.

ASSESS That 10^{14}-Hz frequency sounds huge, but light has such a short wavelength that its frequency is indeed high; more on this shortly. Notice both in Fig. 29.3 and in our reoriented wave of Fig. 29.6 that the vectors $\vec{E}$ and $\vec{B}$ and the propagation direction form a right-handed coordinate system, so any two of those vectors determine the direction of the third. ■

Polarization

Although $\vec{E}$ and $\vec{B}$ are necessarily perpendicular, their orientation is still arbitrary within a plane perpendicular to the propagation direction. **Polarization** specifies the direction of the electric field and thus determines the perpendicular magnetic-field direction as well (Fig. 29.7).

Electromagnetic waves used in radio and TV originate from antennas that give the waves a definite polarization. Most laser light is also polarized. In contrast, light from hot sources like the Sun or a lightbulb is unpolarized, consisting of a mix of waves with random field orientations. Unpolarized light becomes polarized when it reflects off surfaces

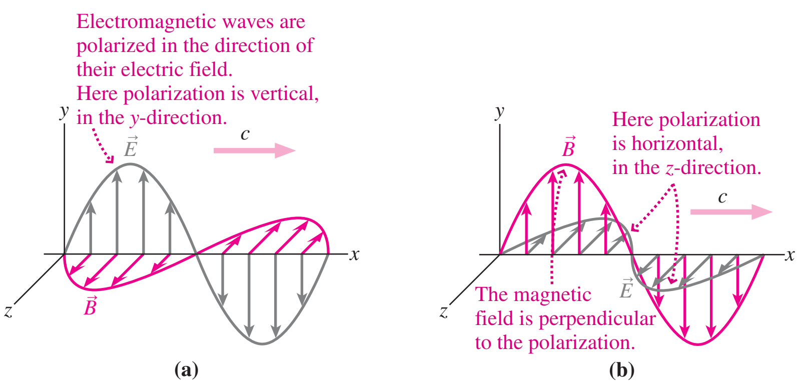


FIGURE 29.7 The polarization direction is the direction of the wave's electric field.

or passes through substances whose structure has a preferred direction. Many crystals and synthetic materials such as Polaroid exhibit such a *transmission axis*. Light reflecting off a car's hood becomes partially polarized in the horizontal direction, and Polaroid sunglasses, with their transmission axis vertical, block the resulting glare.

A polarizing material passes unattenuated only the component of the wave field $\vec{E}$ along the transmission axis—namely, $E\cos\theta$, where θ is the angle between the field and the transmission axis. We'll show shortly that the intensity of an electromagnetic wave is proportional to the square of the field strength. As a result, a wave of intensity S_0 emerges from a polarizer with intensity given by the **Law of Malus**:

$$S = S_0\cos^2\theta \tag{29.18}$$

Thus electromagnetic waves are blocked completely by a polarizer with its transmission axis oriented perpendicular to the waves' polarization (Fig. 29.8).

Measuring polarization tells us about sources of electromagnetic waves and about materials through which they propagate. Many astrophysical processes produce polarized waves; their polarization gives clues to mechanisms operating in the cosmos. Geologists pass polarized light through thin sections of rock to reveal the rocks' composition, and engineers use polarization to locate stresses in mechanical structures. Polarization is essential in many technologies, including the ubiquitous liquid crystal displays (LCDs) in our cell phones, cameras, computers, and TVs (Fig. 29.9).

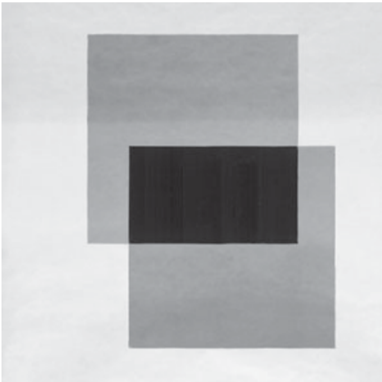


FIGURE 29.8 Two pieces of polarizing material with their transmission axes at right angles. Where they overlap, no light gets through.

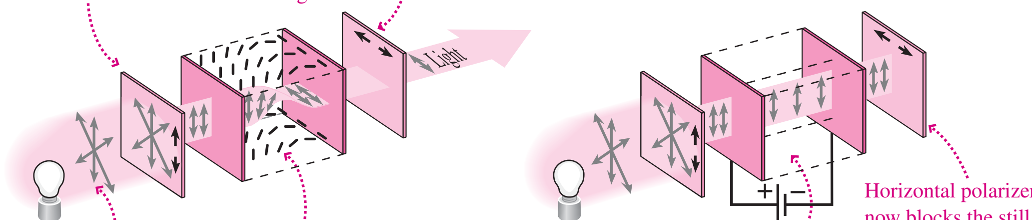


FIGURE 29.9 Polarization plays a central role in the operation of a liquid crystal display. Multiple units like the one shown—millions in a TV or computer screen—produce the individual pixels on an LCD.

CONCEPTUAL EXAMPLE 29.1 Crossed Polarizers

Unpolarized light shines on a pair of polarizers with their transmission axes perpendicular, so no light gets through the combination. What happens when a third polarizer is sandwiched in between, with its transmission axis at 45° to the others?

EVALUATE The middle polarizer's transmission axis isn't perpendicular to the first one's, so some of the light coming through the first polarizer gets through the middle one. That light's polarization isn't perpendicular to the last polarizer's transmission axis, so some light gets all the way through the combination.

ASSESS This result may seem surprising: If the two outer polarizers are perpendicular, how can a third polarizer change the situation? But it does. No pair of adjacent polarizers is perpendicular, so each pair transmits some light. Inserting the third polarizer lets light through where none came through before.

MAKING THE CONNECTION How does the intensity of light emerging from this polarizer "sandwich" compare with the intensity of the incident unpolarized light?

EVALUATE Unpolarized light is a random mix of polarization directions, so $\cos^2\theta$ in Equation 29.18 ranges from 0 to 1 for the first polarizer. Its average is $\frac{1}{2}$, so the intensity emerging from the first polarizer is half the incident intensity. This light is now polarized in the direction of the first polarizer; it then passes through the middle polarizer, oriented at 45°. Since $\cos 45° = 1/\sqrt{2}$, Equation 29.18 shows that its intensity is cut in half again. Light emerging from the middle polarizer then passes through the last one, oriented at 45° to the light's new polarization, so its intensity is halved yet again. The effect of three reductions by one-half each is that light emerges from the "sandwich" with one-eighth its incident intensity.

29.6 The Electromagnetic Spectrum

Although Equations 29.16 relate an electromagnetic wave's frequency and wavelength, one of these quantities remains arbitrary. That means electromagnetic waves can have any frequency or, equivalently, any wavelength. Visible light occupies a wavelength range from about 400 nm to 700 nm, corresponding to frequencies of 7.5×10^{14} Hz to 4.3×10^{14} Hz. The different wavelengths or frequencies correspond to different colors, with red at the long-wavelength, low-frequency end of the visible region and violet at the short-wavelength, high-frequency end (see the enlargement in Fig. 29.10).

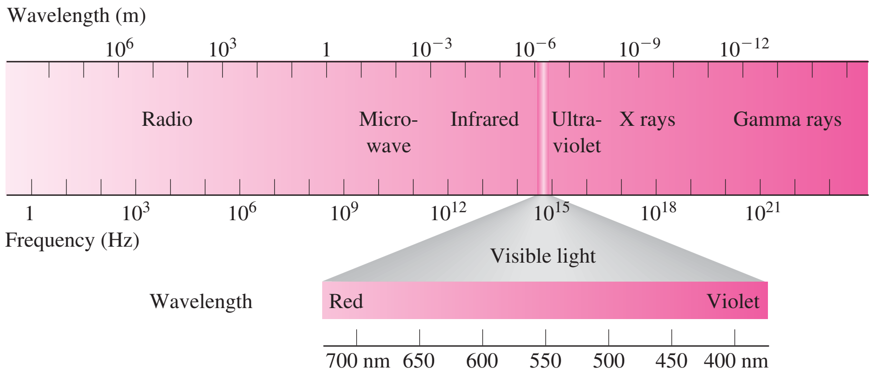


FIGURE 29.10 The electromagnetic spectrum ranges from radio waves to gamma rays, with visible light occupying only a narrow range of wavelengths and frequencies on a logarithmic scale.

Figure 29.10 shows the **electromagnetic spectrum**, including frequencies and wavelengths that differ by many orders of magnitude from those of visible light. The invisible electromagnetic waves beyond the narrow visible range were unknown in Maxwell's time. A brilliant confirmation of Maxwell's theory came in 1888, when the German physicist Heinrich Hertz succeeded in generating and detecting electromagnetic waves of much lower frequency than visible light. Hertz intended his work only to verify Maxwell's modification of Ampère's law, but the practical consequences have proven enormous. In 1901, the Italian scientist Guglielmo Marconi transmitted electromagnetic waves across the Atlantic Ocean, creating a public sensation. From the pioneering work of Hertz and Marconi, spurred by the theoretical efforts of Maxwell, came the entire technology of radio, television, and microwaves that so dominates modern society. We now consider all electromagnetic waves in the frequency range from a few Hz to about 3×10^{11} Hz as radio waves, with AM radio at about 1 MHz, FM at 100 MHz, television in patches of the spectrum from about 50 MHz to 1 GHz, and microwaves for WiFi, radar, cooking, cell phones, and satellite communications at 1 GHz and above.

Between radio waves and visible light lies the infrared frequency range. Electromagnetic waves in this region are emitted by warm objects, even when they're not hot enough to glow visibly. For this reason, infrared cameras are used to determine subtle body-temperature differences in medical diagnosis, to examine buildings for heat loss, and to study the birth of stars in clouds of interstellar gas and dust.

Beyond visible light are the ultraviolet rays responsible for sunburn, then the highly penetrating X rays, and finally the gamma rays whose primary terrestrial source is radioactive decay. All these phenomena, from radio to gamma rays, are fundamentally the same: All are electromagnetic waves, differing only in frequency and wavelength. All travel with speed c in vacuum, and all consist of electric and magnetic fields produced from each other through the induction processes described by Faraday's and Ampère's laws. Naming the different types of electromagnetic waves is just a convenience; there are no gaps in the continuous range of frequencies and wavelengths. Practical differences arise because waves of different wavelengths interact differently with matter; in particular, shorter wavelengths tend to be generated and absorbed most efficiently by smaller systems.

Earth's atmosphere is transparent to visible light and to most radio frequencies. But it's opaque to most infrared, the higher-frequency ultraviolet, X rays, and all but the highest-frequency gamma rays. Earth's surface would be hazardous to life if ultraviolet weren't absorbed by ozone gas high in the atmosphere, and our planet would be a lot cooler if water vapor, carbon dioxide, and other gases didn't absorb outgoing infrared. But that same infrared absorption is at the heart of our worries about global climate change, because we humans are increasing the levels of infrared-trapping gases. Until the space age, our view of the universe beyond Earth was limited to visible light and some weak radio signals. Today, spacecraft observe the cosmos in all wavelengths from radio through gamma rays. Our picture of the universe is far richer than it would be if the only electromagnetic waves we could receive were those that made it through Earth's atmosphere.

29.7 Producing Electromagnetic Waves

All it takes to produce an electromagnetic wave is a changing electric or magnetic field. Once there's a change in one field, induction provides the other field, and together the changing fields continuously regenerate one another. The wave is on its way! Ultimately, changing fields of both types result when we alter the motion of electric charge. Therefore, **accelerated charge is the source of electromagnetic waves.**

In a radio transmitter, the accelerated charges are electrons moving back and forth in an antenna, driven by alternating voltage from an *LC* circuit (Fig. 29.11). In an X-ray tube, high-energy electrons decelerate rapidly as they slam into a target; their deceleration is the source of the electromagnetic waves, now in the X-ray region of the spectrum. In the magnetron tube of a microwave oven, electrons circle in a magnetic field; their centripetal acceleration is the source of the microwaves that cook your food. And the altered movement of electrons in atoms—although described accurately only by quantum mechanics—is the source of most visible light. If the motion of the accelerated charges is periodic, then the wave frequency is that of the motion; more generally, systems are most efficient at producing (and receiving) electromagnetic waves whose wavelength is comparable to the size of the system. That's why TV antennas are on the order of 1 m in size, while nuclei—some 10^{-15} m in diameter—produce gamma rays.

Calculation of electromagnetic waves from accelerated charges presents challenging but important problems for physicists and engineers. Figure 29.12 shows the field of an oscillating dipole—a configuration approximated by many systems from antennas to atoms and molecules. Note that the waves are strongest in the direction at right angles to the acceleration of the charge distribution and that there's no radiation in the direction of the acceleration. This accounts for, among other phenomena, the directionality of radio and TV antennas, which transmit and receive most effectively perpendicular to the long direction of the antenna.

The field shown in Fig. 29.12 seems to bear little resemblance to the plane-wave fields of Fig. 29.3 that we used to demonstrate the possibility of electromagnetic waves. We could produce true plane waves only with an infinite sheet of accelerated charge—an obvious

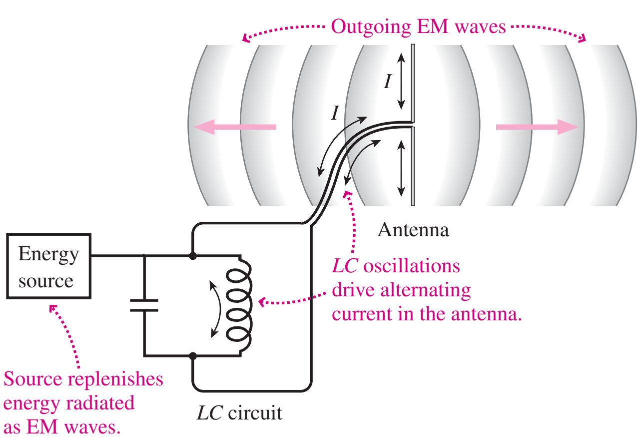


FIGURE 29.11 Simplified diagram of a radio transmitter.

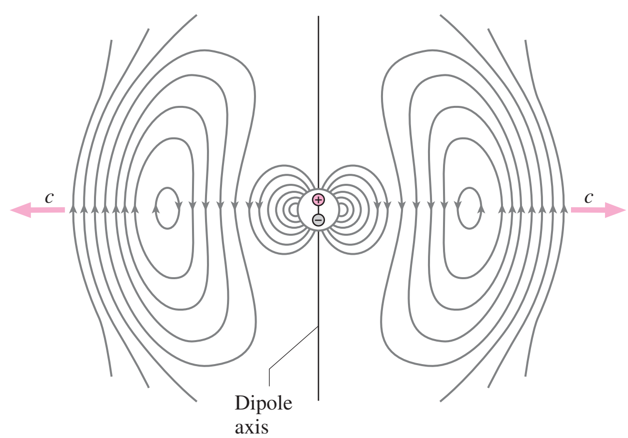


FIGURE 29.12 "Snapshot" showing the electric field of an oscillating electric dipole at an instant in time.

impossibility. But far from the source, the curved field lines in Fig. 29.12 would appear straight, and the wave would approximate a plane wave. So our plane-wave analysis is a valid approximation at great distances—typically many wavelengths—from a localized wave source. Closer to the source more complicated expressions for the wave fields apply, but these, too, satisfy Maxwell's equations.

29.8 Energy and Momentum in Electromagnetic Waves

We showed in earlier chapters that electric and magnetic fields contain energy. Electromagnetic waves are combinations of electric and magnetic fields; as they propagate, they transport the energy contained in those fields.

Wave Intensity

In Chapter 14 we defined wave intensity as the rate at which a wave transports energy across a unit area; its units are W/m^2. We can calculate the intensity S of a plane electromagnetic wave by considering a rectangular box of thickness dx and cross-sectional area A with its face perpendicular to the wave propagation (Fig. 29.13). Within this box are wave fields $\vec{E}$ and $\vec{B}$ whose energy densities are given by Equations 23.7 and 26.9: $u_E = \frac{1}{2}\epsilon_0 E^2$ and $u_B = B^2/2\mu_0$. If dx is small enough that E and B don't vary significantly, the total energy in the box is the sum of the electric and magnetic energy densities multiplied by the box volume $A\,dx$:

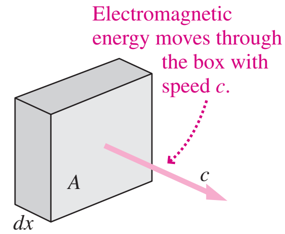


FIGURE 29.13 A box of length dx and cross-sectional area A at right angles to the propagation of an electromagnetic wave.

$$dU = (u_E + u_B)A\,dx = \frac{1}{2}\left(\epsilon_0 E^2 + \frac{B^2}{\mu_0}\right)A\,dx$$

This energy moves with speed c, so all the energy moves out of the box in a time $dt = dx/c$. The rate at which energy moves through the cross-sectional area A is then

$$\frac{dU}{dt} = \frac{1}{2}\left(\epsilon_0 E^2 + \frac{B^2}{\mu_0}\right)\frac{A\,dx}{dx/c} = \frac{c}{2}\left(\epsilon_0 E^2 + \frac{B^2}{\mu_0}\right)A$$

So the intensity S, or rate of energy flow per unit area, is

$$S = \frac{c}{2}\left(\epsilon_0 E^2 + \frac{B^2}{\mu_0}\right)$$

We can recast this equation in simpler form using $E = cB$ and $B = E/c$ for an electromagnetic wave. Replacing one of the E's in E^2 with cB and one of the B's in B^2 with E/c, we have

$$S = \frac{c}{2}\left(\epsilon_0 cEB + \frac{EB}{\mu_0 c}\right) = \frac{1}{2\mu_0}(\epsilon_0\mu_0 c^2 + 1)EB$$

But $c = 1/\sqrt{\epsilon_0\mu_0}$, so $\epsilon_0\mu_0 c^2 = 1$, giving

$$S = \frac{EB}{\mu_0} \tag{29.19a}$$

Although we derived Equation 29.19a for an electromagnetic wave, it is in fact a special case of the more general result that nonparallel electric and magnetic fields entail a flow of electromagnetic energy. In general, the rate of energy flow per unit area is given by

$$\vec{S} = \frac{\vec{E} \times \vec{B}}{\mu_0} \quad \text{(Poynting vector)} \tag{29.19b}$$

Here the vector $\vec{S}$ gives the direction of the energy flow as well as its magnitude. For an electromagnetic wave in vacuum, with $\vec{E}$ and $\vec{B}$ at right angles, Equation 29.19b reduces to Equation 29.19a, with the direction of energy flow the same as the direction of wave travel. The vector intensity $\vec{S}$ is called the **Poynting vector** after the English physicist J. H. Poynting, who suggested it in 1884. Problem 60 explores an important application of the Poynting vector to fields that don't constitute an electromagnetic wave.

In an electromagnetic wave the fields oscillate, and so does the intensity. We're usually not interested in this rapid oscillation. For example, an engineer designing a solar collector doesn't care that sunlight intensity oscillates at about 10^{14} Hz. What she really wants is the *average* intensity, $\bar{S}$. Because the instantaneous intensity of Equation 29.19a contains a product of sinusoidally varying terms, which are in phase, the average intensity is just half the peak intensity:

$$\bar{S} = \frac{\overline{EB}}{\mu_0} = \frac{E_p B_p}{2\mu_0} \qquad \text{(average intensity)} \tag{29.20a}$$

Typical values for $\bar{S}$ in visible light range from a few W/m^2 in the faint light of a candle to many MW/m^2 in the most intense laser beams.

We wrote Equation 29.20a in terms of both the electric and magnetic fields, but we can use the wave condition $E = cB$ to eliminate either field in terms of the other:

$$\bar{S} = \frac{E_p^{\,2}}{2\mu_0 c} \quad \text{and} \quad \bar{S} = \frac{cB_p^{\,2}}{2\mu_0} \tag{29.20b, c}$$

GOT IT? 29.2 Lasers 1 and 2 emit light of the same color, and the electric field in the beam from laser 1 is twice as strong as the field in laser 2's beam. How do their (a) magnetic fields, (b) intensities, and (c) wavelengths compare?

EXAMPLE 29.3 Fields and Power: Solar Energy

The average intensity of noontime sunlight on a clear day is about 1 kW/m^2. (a) What are the peak electric and magnetic fields in sunlight? (b) At this intensity, what area of 40% efficient solar collectors would you need to replace a 4.8-kW water heater?

INTERPRET In (a) we're asked for the peak electric and magnetic fields, and we identify 1 kW/m^2 as the average intensity, $\bar{S}$. In (b) we want solar collectors to replace an electric heater whose power we're given.

DEVELOP For (a), Equations 29.20b, $\bar{S} = E_p^{\,2}/2\mu_0 c$, and 29.20c, $\bar{S} = cB_p^{\,2}/2\mu_0$, relate peak fields and average intensity. We could solve both of these or, more easily, solve one of them and then use Equation 29.17, $E = cB$, to get the other field. In (b) we'll need to use the 40% efficiency to get the effective power per square meter of solar collector, and then find the area needed to replace the 4.8-kW power of the electric heater.

EVALUATE (a) We solve Equation 29.20c for B_p: $B_p = \sqrt{2\mu_0 \bar{S}/c} = 2.9\,\mu$T. Then Equation 29.17 gives $E = cB = 0.87$ kV/m. (b) At 40% efficiency, each square meter of solar collector in 1-kW/m^2 sunlight produces 0.40 kW. So to get 4.8 kW, we need $(4.8\text{ kW})/(0.40\text{ kW/m}^2) = 12\text{ m}^2$ of collector area.

ASSESS The fields we've calculated are relatively modest, showing that even bright sunlight doesn't entail large electric and magnetic fields. That 12-m^2 collector area is also modest—much smaller than the 100 m^2 of a typical house roof—showing that water heating is a practical use of solar technology. Indeed, your author's house, in cloudy Vermont, gets 95% of its summertime hot water from just 9 m^2 of solar collectors. ■

Waves from Localized Sources

When a wave originates in a localized source such as an atom, a radio transmitting antenna, a lightbulb, or a star, its wavefronts aren't planes but expanding spheres (recall Fig. 14.13). As it expands, the wave's energy is spread over the area of an ever-larger sphere—whose area increases as the square of the distance from the source. Therefore the power per unit area—the intensity—decreases as the inverse square of the distance:

$$S = \frac{P}{4\pi r^2} \tag{29.21}$$

Here S and P can be either peak or average intensity and power, and r is the distance from the source. The intensity decreases not because electromagnetic waves "weaken" and lose energy but because their energy gets spread ever more thinly.

Because the intensity of an electromagnetic wave is proportional to the *square* of the field strengths (Equations 29.20), Equation 29.21 shows that the *fields* of a spherical wave decrease as $1/r$. Contrast that with the $1/r^2$ decrease in the electric field of a stationary point charge, and you can see why the wave fields associated with an accelerated charge dominate in all but the immediate vicinity of the charge.

EXAMPLE 29.4 Electromagnetic-Wave Intensity: Cell-Phone Reception

A cell phone's typical average radiated power is about 0.6 W. If the receiver at a cell tower can handle signals with peak electric fields as weak as 1.2 mV/m, what's the maximum distance from phone to tower?

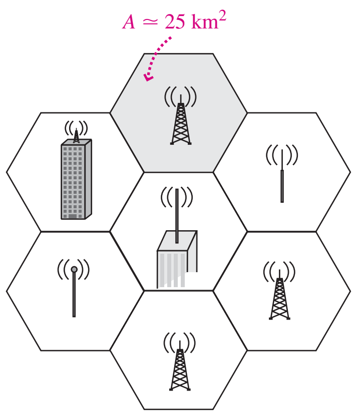
INTERPRET We're asked to find the distance from the 0.6-W cell phone to a cell tower on the condition that the electric field of the cell phone's electromagnetic wave is no weaker than 1.2 mV/m when measured at the tower.

DEVELOP Assuming the 0.6-W signal spreads in all directions, Equation 29.21, $\overline{S} = \overline{P}/4\pi r^2$, gives the average intensity at a distance r from the phone. (Here the bars indicate we're dealing with average quantities.) Using this value for $\overline{S}$ in Equation 29.20b gives $\overline{P}/4\pi r^2 = E_p^2/2\mu_0 c$. Our plan is to solve for the distance r that gives $E_p = 1.2$ mV/m.

EVALUATE Solving for r gives $r = \sqrt{2\mu_0 c P/4\pi E_p^2} = 5$ km, using $P = 0.6$ W and $E_p = 1.2$ mV/m.

ASSESS This answer is about 3 miles, a bit more than the cell radius discussed in the Application below. That's enough to provide a margin of safety, ensuring reliable communications for all phones within the cell. ■

APPLICATION Cell Phones



Your cell phone contains a tiny, low-power radio transmitter whose signal intensity decreases as the inverse square of the distance from the phone. The cell-phone network consists of antennas and associated circuits that receive and transmit signals from and to individual phones. Because of the phones' low power, antennas need to be closely spaced so a phone is rarely out of range. The figure shows a typical urban cell-phone network consisting of multiple cells—hence the term "cell" phone—each with an antenna mounted on a tower or building. Cells are typically hexagonal regions about 25 km^2 in area; approximating them as circles gives a radius of about 2.8 km—roughly the maximum distance between a phone and an antenna. As you move through an urban area, the network automatically "hands off" your phone to the nearest cell tower. Cell phones transmit on one frequency and receive on another, allowing two-way communications with both parties able to talk at once. The system uses hundreds to thousands of frequency channels, and thus a single cell tower can handle many simultaneous calls. Cell towers are more widely spaced in rural regions, and phones automatically boost their power to compensate.

Momentum and Radiation Pressure

Moving objects carry not only energy but also momentum. So do electromagnetic waves. Maxwell showed that wave energy U and momentum p are related by $p = U/c$. The wave intensity $\overline{S}$ is the average rate at which the wave carries energy per unit area, and therefore the wave carries momentum per unit area at the rate $\overline{S}/c$. An object that absorbs the wave energy (like a black object exposed to sunlight) absorbs this momentum as well. Newton's law in its general form $F = dp/dt$ shows that the object then experiences a force. Since $\overline{S}/c$ is the rate of momentum absorption per unit area, the result is a **radiation pressure**:

$$P_{\text{rad}} = \frac{\overline{S}}{c} \qquad \text{(radiation pressure)} \qquad (29.22)$$

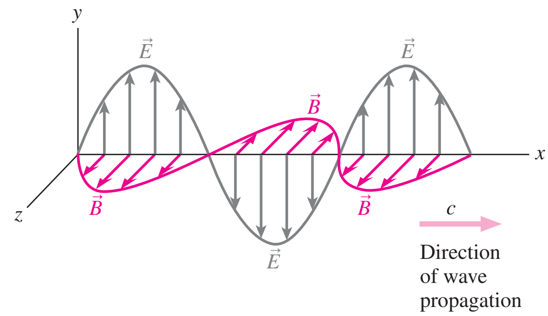
Radiation pressure doubles if an object reflects electromagnetic waves, just as bouncing a basketball off the backboard changes the ball's momentum by $2mv$ and therefore delivers momentum $2mv$ to the backboard.

The pressure of ordinary light is tiny and difficult to measure, but high-power lasers can actually levitate small particles. Light pressure has even been suggested for spacecraft propulsion (see Passage Problems 69–72). Finally, the idea that electromagnetic waves carry momentum played a crucial role in Einstein's development of his famous equation $E = mc^2$.

CHAPTER 29 SUMMARY

Big Picture

The big idea here—and one of the biggest ideas in physics—is that electric and magnetic fields together form self-regenerating structures that propagate through space as **electromagnetic waves**. What makes these waves possible is that changing magnetic fields induce electric fields (Faraday's law), and changing electric fields induce magnetic fields (Ampère's law, with Maxwell's modification). Electromagnetic (EM) waves in vacuum consist of electric and magnetic fields perpendicular to each other and to the direction of wave propagation, and in phase.



Key Concepts and Equations

Maxwell's equations describe completely the behavior of electric and magnetic fields in classical physics:

Law	Mathematical Statement	What It Says
Gauss for $\vec{E}$	$\oint \vec{E}\cdot d\vec{A} = \dfrac{q}{\epsilon_0}$	How charges produce electric fields; field lines begin and end on charges.
Gauss for $\vec{B}$	$\oint \vec{B}\cdot d\vec{A} = 0$	No magnetic charge; magnetic field lines don't begin or end.
Faraday	$\oint \vec{E}\cdot d\vec{r} = -\dfrac{d\Phi_B}{dt}$	Changing magnetic flux produces electric fields.
Ampère	$\oint \vec{B}\cdot d\vec{r} = \mu_0 I + \mu_0\epsilon_0\dfrac{d\Phi_E}{dt}$	Electric current and changing electric flux produce magnetic fields.

Maxwell's equations show that electromagnetic waves are possible and that their speed in vacuum, the speed of light c, is related to the electric and magnetic constants ϵ_0 and μ_0:

$$c = \frac{1}{\sqrt{\epsilon_0\mu_0}}$$

The value of c is very nearly 3.00×10^8 m/s. Its exact value, used in defining the meter, is 299,792,458 m/s.

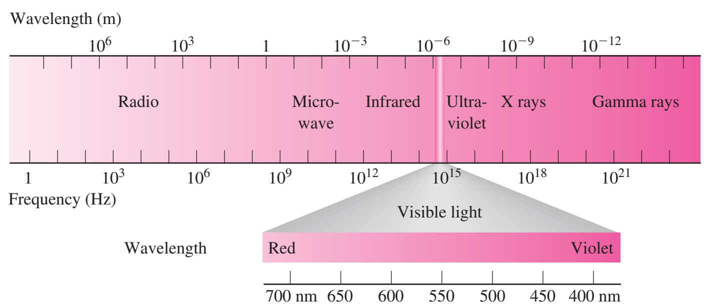
In vacuum, the electric and magnetic fields of a wave are related by

$$E = cB$$

The wave's frequency and wavelength are related by

$$f\lambda = c$$

EM waves can have any wavelength; the whole range constitutes the **electromagnetic spectrum**.



Applications

Polarization describes the direction of an EM wave's electric field and is a property widely used in scientific research and in technological devices including the ubiquitous liquid crystal displays. When polarized light of intensity S_0 is incident on a polarizer with its transmission axis at angle θ to the polarization, the light emerges with intensity

$$S = S_0\cos^2\theta$$

EM waves carry both energy and momentum. The **Poynting vector**

$$\vec{S} = \frac{\vec{E}\times\vec{B}}{\mu_0}$$

describes the rate of energy flow per unit area, while the momentum flow results in a **radiation pressure:**

$$P_{\text{rad}} = \frac{\bar{S}}{c}$$

For Thought and Discussion

1. Why is Maxwell's modification of Ampère's law essential to the existence of electromagnetic waves?
2. The presence of magnetic monopoles would require a modification of Gauss's law for magnetism. Which other Maxwell equation would need modification?
3. Is there displacement current in an electromagnetic wave? Is there ordinary conduction current?
4. List some similarities and differences between electromagnetic waves and sound waves.
5. The speed of an electromagnetic wave is given by $c = \lambda f$. How does the speed depend on frequency? On wavelength?
6. When astronomers observe a supernova explosion in a distant galaxy, they see a sudden, simultaneous rise in visible light and other forms of electromagnetic radiation. How is this evidence that the speed of light is independent of frequency?
7. Turning a TV antenna so its rods point vertically may change the quality of your TV reception. Why?
8. The Sun emits about half of its electromagnetic-wave energy in the visible region of the spectrum. Where do you think it emits most of the remainder?
9. An *LC* circuit is made entirely from superconducting materials, yet its oscillations eventually damp out. Why?
10. If you double the field strength in an electromagnetic wave, what happens to the intensity?
11. The intensity of light drops as the inverse square of the distance from the source. Does this mean that electromagnetic energy is lost? Explain.
12. Electromagnetic waves don't readily penetrate metals. Why not?

Exercises and Problems

Exercises

Section 29.2 Ambiguity in Ampère's Law

13. A uniform electric field is increasing at 1.5 $(\text{V/m})/\mu\text{s}$. Find the displacement current through a 1-cm^2 area perpendicular to the field.
14. A parallel-plate capacitor has square plates 10 cm on a side and 0.50 cm apart. The voltage across the plates is increasing at 220 V/ms. What's the displacement current in the capacitor?

Section 29.4 Electromagnetic Waves

15. The fields of an electromagnetic wave are $\vec{E} = E_\text{p}\sin(kz + \omega t)\hat{\jmath}$ and $\vec{B} = B_\text{p}\sin(kz + \omega t)\hat{\imath}$. Give a unit vector in the wave's propagation direction.
16. A radio wave's electric field is given by $\vec{E} = E\sin(kz - \omega t) \times (\hat{\imath} + \hat{\jmath})$. (a) Find the peak electric field. (b) Give a unit vector in the direction of the magnetic field at a place and time where $\sin(kz - \omega t)$ is positive.

Section 29.5 Properties of Electromagnetic Waves

17. A *light-minute* is the distance light travels in 1 minute. Show that the Sun is about 8 light-minutes from Earth.
18. Your intercontinental telephone call is carried by electromagnetic waves routed via a satellite in geosynchronous orbit at 36,000 km altitude. Approximately how long does it take before your voice is heard at the other end?
19. An airplane's radar altimeter works by bouncing radio waves off the ground and measuring the round-trip travel time. If that time is 50 μs, what's the altitude?
20. Roughly how long does it take light to travel 1 foot?
21. If you speak via radio from Earth to an astronaut on the Moon, how long is it before you can get a reply?
22. What are the wavelengths of (a) a 100-MHz FM radio wave, (b) a 5.0-GHz WiFi signal, (c) a 600-THz light wave, and (d) a 1.0-EHz X ray?
23. A 60-Hz power line emits electromagnetic radiation. What's the wavelength?
24. A microwave oven operates at 2.4 GHz. What's the distance between wave crests in the oven?
25. An electromagnetic wave is propagating in the *z*-direction. What's its polarization direction if its magnetic field is in the *y*-direction?
26. Polarized light is incident on a sheet of polarizing material, and only 20% of the light gets through. Find the angle between the electric field and the material's transmission axis.
27. Vertically polarized light passes through a polarizer with its axis at 70° to the vertical. What fraction of the incident intensity emerges from the polarizer?

Section 29.8 Energy and Momentum in Electromagnetic Waves

28. A typical laboratory electric field is 1000 V/m. Find the average intensity of an electromagnetic wave with this value for its peak field.
29. What would be the average intensity of a laser beam so strong that its electric field produced dielectric breakdown of air (which requires $E_\text{p} = 3$ MV/m)?
30. Estimate the peak electric field inside a 1.1-kW microwave oven under the simplifying approximation that the microwaves propagate as a plane wave through the oven's 750-cm^2 cross-sectional area.
31. Your new radio says it can pick up signals with peak electric fields as weak as 450 μV/m. Will it work if you take it to your remote cabin, where the intensity of your favorite radio station is 0.35 nW/m^2?
32. A laser pointer delivers 0.10-mW average power in a beam 0.90 mm in diameter. Find (a) the average intensity, (b) the peak electric field, and (c) the peak magnetic field.
33. Your university radio station has a 5.0-kW radio transmitter that broadcasts uniformly in all directions; listeners within 15 km have reliable reception. You want to increase the power to double that range. What should be the new power?

Problems

34. A parallel-plate capacitor has circular plates with radius 50 cm and spacing 1.0 mm. A uniform electric field between the plates is changing at the rate of 1.0 MV/m·s. Find the magnetic field between the plates (a) on the symmetry axis, (b) 15 cm from the axis, and (c) 150 cm from the axis.
35. An electric field points into the page and occupies a circular region of radius 1.0 m, as shown in Fig. 29.14. There are no electric

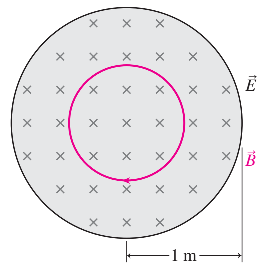


FIGURE 29.14 Problem 35

charges in the region, but there is a magnetic field forming closed loops pointing clockwise, as shown. The magnetic-field strength 50 cm from the center of the region is 2.0 μT. (a) What's the rate of change of the electric field? (b) Is the electric field increasing or decreasing?

36. You're engineering a new cell phone, and you'd like to incorporate the antenna entirely within the phone, which is 9 cm long when closed. The antenna is to be a quarter-wavelength long—a common design for vertically oriented antennas. If the cell-phone frequency is 2.4 GHz, will the antenna fit?
37. BIO The medical profession divides the ultraviolet region of the electromagnetic spectrum into three bands: UVA (320–420 nm), UVB (290–320 nm), and UVC (100–290 nm). UVA and UVB promote skin cancer and premature skin aging; UVB also causes sunburn, but helpfully fosters production of vitamin D. Ozone in Earth's atmosphere blocks most of the more dangerous UVC. Find the frequency range associated with UVB radiation.
38. Dielectric breakdown in air occurs when the electric field is approximately 3 MV/m. What would be the peak magnetic field in an electromagnetic wave with this peak electric field?
39. A radio receiver can detect signals with electric fields as weak as 320 μV/m. Find the corresponding magnetic field.
40. A polarizer blocks 75% of a polarized light beam. What's the angle between the beam's polarization and the polarizer's axis?
41. An electro-optic modulator is a device that switches a laser beam rapidly from off to on by switching the polarization direction through 90° when a voltage is applied. But a brownout results in only enough voltage for a 72° rotation. What fraction of the light is transmitted during the brownout when the beam is supposed to be fully on?
42. Unpolarized light of intensity S_0 passes first through a polarizer with its axis vertical and then through one with its axis at 35° to the vertical. Find the intensity after the second polarizer.
43. Vertically polarized light passes through two polarizers, the first at 60° to the vertical and the second at 90° to the vertical. What fraction of the light gets through?
44. Show that it's impossible for an electromagnetic wave in vacuum to have a time-varying component of its electric field in the direction of its magnetic field. (*Hint:* Assume $\vec{E}$ does have such a component, and show that you can't satisfy both Gauss's and Faraday's laws.)
45. BIO High microwave intensities can cause biological damage through heating of tissue; a particular concern is cataract formation. The U.S. Food and Drug Administration limits microwave radiation near the door of a microwave oven to 5.0 mW/m^2. The window in a particular oven door measures 40 cm by 17 cm and is covered with a metal screen to block microwaves. Assuming power leaks uniformly through the window area, what percent of the oven's 900-W microwave power can leak without exceeding the FDA standards?
46. Use the fact that sunlight intensity at Earth's orbit is 1368 W/m^2 to calculate the Sun's total power output.
47. A quasar 10 billion light-years from Earth appears the same brightness as a star 50,000 light-years away. How do the power outputs of quasar and star compare?
48. BIO Lasers are classified according to the eye-damage danger they pose. Class 2 lasers, including many laser pointers, produce visible light with no greater than 1 mW total power. They're relatively safe because the eye's blink reflex limits exposure time to 250 ms. Find (a) the intensity of a 1-mW class 2 laser with beam diameter 1.0 mm, (b) the total energy delivered before the blink reflex shuts the eye, and (c) the peak electric field in the laser beam.
49. At 1.5 km from a radio transmitter, the peak electric field is 350 mV/m. Assuming the transmitter broadcasts equally in all directions, find (a) the transmitted power and (b) the peak electric field 10 km from the transmitter.
50. Find the peak electric and magnetic fields 1.5 m from a 60-W lightbulb that radiates equally in all directions.
51. A typical fluorescent lamp is a little more than 1 m long and a few cm in diameter. How do you expect the light intensity to vary with distance (a) near the lamp but not near either end and (b) far from the lamp?
52. A camera flash delivers 2.5 kW of light power for 1.0 ms. Find (a) the total energy and (b) the total momentum carried by the flash.
53. A laser produces an average power of 7.0 W in a 1.0-mm-diameter beam. Find (a) the average intensity and (b) the peak electric field of the laser light.
54. A 180-W/cm^2 laser beam shines on a light-absorbing surface. What's the radiation pressure on the surface?
55. A 65-kg astronaut is floating in empty space. If she shines a 1.0-W flashlight in a fixed direction, how long will it take her to accelerate to 10 m/s?
56. A *photon rocket* emits a beam of light instead of hot gas. How powerful a beam would be needed for a thrust equal to that of a space shuttle (35 MN)? Compare your answer with humanity's total electric power-generating capability, about 1 TW.
57. A white dwarf star is approximately the size of Earth but radiates about as much power as the Sun. Estimate the radiation pressure on a light-absorbing object at the white dwarf's surface.
58. Use appropriate data from Appendix E to calculate the radiation pressure on a light-absorbing object at the Sun's surface.
59. A radar system produces pulses consisting of 100 full cycles of a sinusoidal 70-GHz electromagnetic wave. The average power while the transmitter is on is 45 MW, and the waves are confined to a beam 20 cm in diameter. Find (a) the peak electric field, (b) the wavelength, (c) the total energy in a pulse, and (d) the total momentum in a pulse. (e) If the transmitter produces 1000 pulses per second, what's its average power output?
60. A cylindrical resistor of length L, radius a, and resistance R carries current I. Calculate the electric and magnetic fields at the surface of the resistor, assuming the electric field is uniform over the surface. Calculate the Poynting vector and show that it points into the resistor. Calculate the flux of the Poynting vector (that is, $\int \vec{S} \cdot d\vec{A}$) over the resistor's surface to get the rate of electromagnetic energy flow into the resistor, and show that the result is I^2R. Your result shows that the energy heating the resistor comes from the fields surrounding it. These fields are sustained by the source of electric energy that drives the current.
61. In a stack of polarizing sheets, each sheet has its transmission axis rotated 14° with respect to the preceding sheet. If the stack passes 37% of the incident unpolarized light, how many sheets does it contain?
62. You're an astronomer studying the origin of the solar system, and you're evaluating a hypothesis that sufficiently small particles were blown out of the solar system by the force of sunlight. To see how small such particles must be, compare the force of sunlight with the force of solar gravity, and solve for the particle radius at which the two are equal. Assume spherical particles with density 2 g/cm^3. (*Note:* Distance from the Sun doesn't matter. Why not?)
63. Differentiate Equation 29.12 with respect to x and Equation 29.13 with respect to t. Then, using the fact that mixed derivatives are equal (e.g., $\frac{\partial}{\partial t}\left(\frac{\partial B}{dx}\right) = \frac{\partial}{\partial x}\left(\frac{\partial B}{dt}\right)$), combine the resulting

equations and show that the result is the wave equation (Equation 14.5) for waves with speed $c = 1/\sqrt{\epsilon_0 \mu_0}$.

64. Maxwell's equations in a dielectric resemble those in vacuum (Equations 29.6–29.9) but with ϵ_0 replaced by $\kappa\epsilon_0$, where κ is the dielectric constant introduced in Chapter 23. Show that the speed of electromagnetic waves in a dielectric is $c/\sqrt{\kappa}$.
65. A friend buys a used pickup truck that comes with a CB radio. However, the antenna is broken off, and your friend asks you to help make one out of a steel rod that he will affix to the rear bumper. You know that the CB channel frequency is 27.3 MHz and that the antenna must be a quarter-wavelength long. How long should you make the rod?
66. Your roommate's father is CEO of a coal company, so your roommate is understandably skeptical of alternative energy proposals. He claims that there's no future for solar energy, because the power in sunlight is insufficient to meet humankind's energy demand. Is he right? To find out, compare the solar power incident on Earth with our human energy consumption rate of about 15 TW.
67. The Voyager 1 spacecraft is now beyond the outer reaches of our solar system, but earthbound scientists still receive data from the spacecraft's 20-W radio transmitter. Voyager is expected to continue transmitting until about 2025, when it will be some 25 billion km from Earth. What's the diameter of a dish antenna that will receive 10^{-20} W of power from Voyager at this time?
68. Your friend who works for the college radio station must make electric-field measurements for a report to be filed with the station's application for license renewal. The measurement is made 4.6 km from the antenna, where your friend measures the electric field at 380 V/m. The station is allowed to broadcast at no more than 55-kW power. Assuming power spreads equally in all directions, is the station in compliance with its license?

Passage Problems

Proposals have been made to "sail" spacecraft to the outer solar system using the pressure of sunlight, or even to propel interstellar spacecraft with high-powered, Earth-based lasers. Sailing spacecraft would need no fuel—a great advantage because fuel constitutes much of the initial weight of any space mission. The first successful test of sunlight-powered sailing is the Japanese spacecraft IKAROS, launched in 2010 (Fig. 29.15).

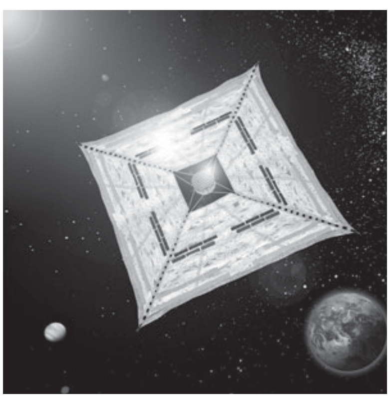


FIGURE 29.15 Launched in 2010, the Japanese IKAROS mission is the first successful test of a sunlight-powered sailing spacecraft. IKAROS's sail is 200 m^2 in area but only 0.0075 mm thick (Passage Problems 69–72).

69. If a sunlight-powered sailing spacecraft accelerated at 1 m/s^2 in the vicinity of Earth's orbit, what would be its acceleration at Mars, about 1.5 times as far from the Sun as Earth?
   a. about 0.25 m/s^2
   b. a little less than 0.5 m/s^2
   c. a little more than 0.5 m/s^2
   d. about 0.66 m/s^2
70. One spacecraft has a sail that absorbs all light incident on it; the other has a perfectly reflective sail. How do their accelerations compare in light with the same intensity?
   a. The absorptive sail gives twice the acceleration.
   b. The reflective sail gives twice the acceleration.
   c. The absorptive sail gives greater acceleration, but not twice as much.
   d. The reflective sail gives greater acceleration, but not twice as much.
71. A sail capable of propelling a spacecraft to the outer solar system must be able to overcome the Sun's gravity. Suppose a spacecraft is designed so the force of sunlight on its sail is 20 times that of solar gravity in the vicinity of Earth's orbit. If the spacecraft reaches Jupiter, some 5 times as far from the Sun as Earth,
   a. the sail force will still exceed solar gravity, now by a factor of 4.
   b. the sail force will be slightly less than solar gravity.
   c. the sail force will now be 25 times solar gravity.
   d. the sail force will still be 20 times solar gravity.
72. The intensity of sunlight at Earth's orbit is about 1.4 kW/m^2. A 100-kg sailing spacecraft with 1-km^2 sail area would experience an acceleration of about
   a. 5 mm/s^2.
   b. 5 cm/s^2.
   c. 5 m/s^2.
   d. 5 km/s^2.

Answers to Chapter Questions

Answer to Chapter Opening Question

Electromagnetic waves, comprising changing electric and magnetic fields, carry not only cell-phone conversations but also TV shows, the energy of sunlight, and signals from physical processes in the farthest reaches of the universe.

Answers to GOT IT? Questions

29.1. (b).
29.2. (a) $B_1 = 2B_2$; (b) $S_1 = 4S_2$; (c) $\lambda_1 = \lambda_2$.

Electromagnetism

Electromagnetism is a fundamental force of nature. The strong attraction between positive and negative charge makes most bulk matter electrically neutral, and hides from us the essential role electricity and magnetism play in the structure of matter.

Electromagnetic interactions are best described in terms of **electric fields** and **magnetic fields**. Electric charges create electric fields, and electric charges respond to the fields of other charges. **Moving electric charges** create magnetic fields, and moving electric charges respond to magnetic fields. Both electric and magnetic fields store energy.

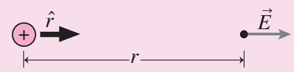
A changing magnetic field creates an electric field, and vice versa. Together, changing fields combine to make **electromagnetic waves**—self-replicating structures that propagate through empty space at the speed of light, c. Light itself is an electromagnetic wave.

Maxwell's equations are the four fundamental laws of electromagnetism:

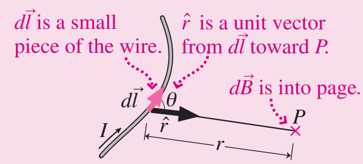
Law	Mathematical Statement	What It Says
Gauss for $\vec{E}$	$\oint \vec{E}\cdot d\vec{A} = \dfrac{q}{\epsilon_0}$	How charges produce electric field; field lines begin and end on charges.
Gauss for $\vec{B}$	$\oint \vec{B}\cdot d\vec{A} = 0$	No magnetic charge; magnetic field lines don't begin or end.
Faraday	$\oint \vec{E}\cdot d\vec{r} = -\dfrac{d\Phi_B}{dt}$	Changing magnetic flux produces electric field.
Ampère	$\oint \vec{B}\cdot d\vec{r} = \mu_0 I + \mu_0\epsilon_0\dfrac{d\Phi_E}{dt}$	Electric current and changing electric flux produce magnetic field.

Coulomb's law and the **Biot–Savart law** provide alternatives to Gauss's and Ampère's laws for determining electric and magnetic fields of pointlike elements of charge and moving charge, respectively:

$$\vec{E} = \frac{kq}{r^2}\hat{r}$$



$$d\vec{B} = \frac{\mu_0}{4\pi}\frac{I\,d\vec{l} \times \hat{r}}{r^2}$$



The **electromagnetic force** on a charged particle consists of the **electric force** and the **magnetic force**. Both are proportional to the charge and to the appropriate field; the magnetic force depends also on the particle's velocity $\vec{v}$:

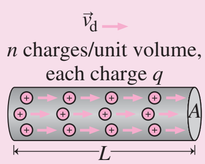
$$\vec{F}_{\mathrm{EM}} = \vec{F}_E + \vec{F}_B = q\vec{E} + q\vec{v} \times \vec{B}$$

The **electric potential difference** describes the work per unit charge needed to move charge between two points in an electric field; its units are N/C or **volts** (V):

$$V_{AB} = -\int_A^B \vec{E}\cdot d\vec{r}$$

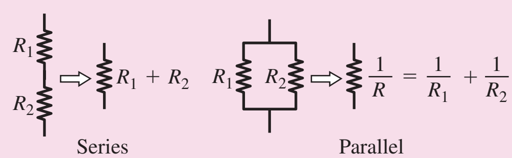
Electric current is a flow of electric charge:

$$I = \frac{\Delta Q}{\Delta t} = nAqv_{\mathrm{d}}$$

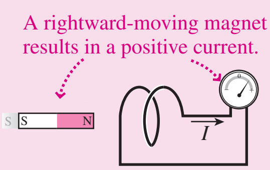


In ohmic materials, **Ohm's law** relates voltage, current, and resistance: $I = V/R$.

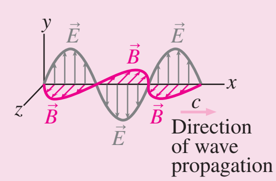
Electric circuits are interconnections of electric components, including batteries, resistors, and others. They can often be analyzed by considering **series** and **parallel** combinations.



Electromagnetic induction, described by Faraday's law, is the basis of electric generators and a host of other electromagnetic technologies and natural phenomena.

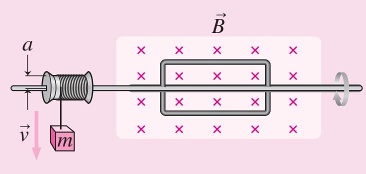


Electromagnetic waves result from changing electric and magnetic fields. EM waves include light, and all EM waves propagate in vacuum at the speed of light, $c = 1/\sqrt{\mu_0\epsilon_0}$.



Part Four Challenge Problem

A wire of length L and resistance R forms a rectangular loop twice as long as it is wide. It's mounted on a nonconducting horizontal axle parallel to its longer dimension, as shown in the figure. A uniform magnetic field $\vec{B}$ points into the page. A long string of negligible mass is wrapped many times around a drum of radius a attached to the axle, and a mass m is attached to the string. When the mass is released, it falls and eventually reaches a speed that, averaged over one cycle of the loop's rotation, is constant. Find an expression for that average speed.



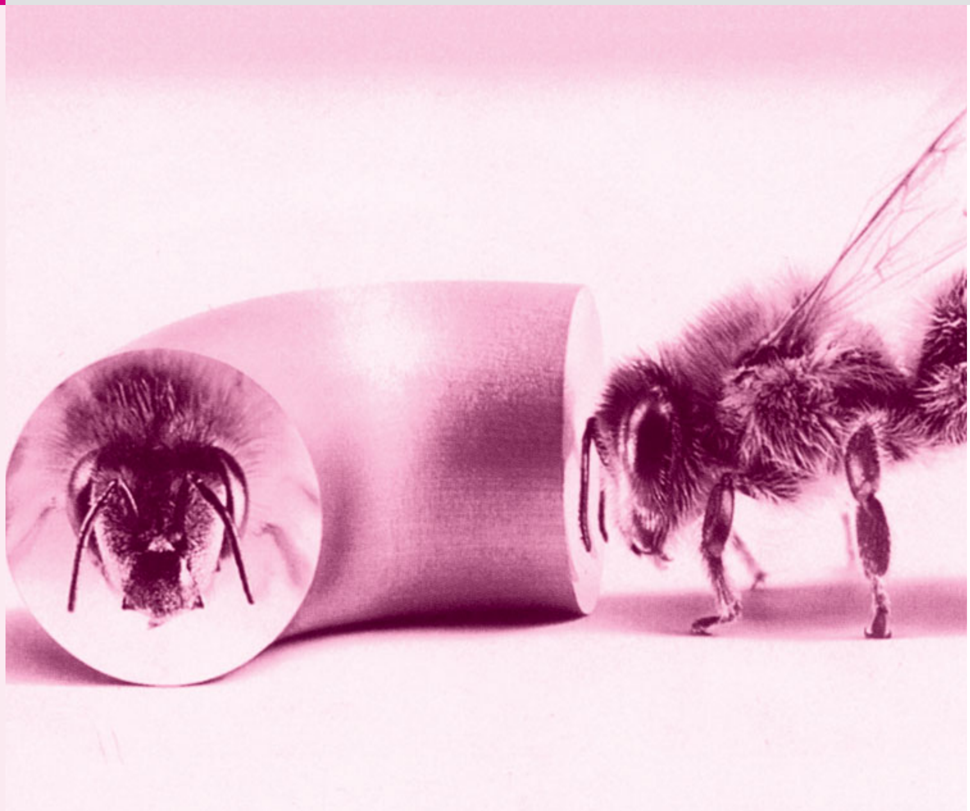
PART FIVE OVERVIEW

Optics

Drops of dew act as miniature optical systems, with light refracting through the drops to form myriad images of the background flowers.

Imagine a world without light. We see because light reflects off objects, and our eyes form images because light refracts in our corneas and lenses. When our built-in optical systems aren't perfect, we correct them with additional lenses or we use lasers to reshape the cornea. Microscopes and telescopes extend the range of our vision. The phenomenon of interference makes possible some of the most precise measurements and is behind the operation of everyday technologies like CDs and DVDs. Light signals carry e-mail, web pages, telephone conversations, and computer data through the optical fibers that form the world's communications networks. Although the behavior of light is ultimately grounded in Maxwell's equations of electromagnetism, we can learn much about light from the simpler perspective of optics. The next three chapters explore the behavior of light, images and optical instruments, and phenomena associated with the wave nature of light.

30 Reflection and Refraction



Why does the bee's image appear at left, and what does this have to do with e-mail and the Internet?

New Concepts, New Skills

By the end of this chapter you should be able to

- Describe the reflection of light from plane surfaces (30.1).
- Describe quantitatively the refraction of light at the interface between two transparent materials (30.2).
- Understand total internal reflection, and determine quantitatively the circumstances when it occurs (30.3).
- Explain how the wavelength dependence of refraction leads to dispersion (30.4).

Connecting Your Knowledge

- This chapter is, in principle, based on the electromagnetic nature of light as described by Maxwell's equations, but here we'll develop a simpler approach that applies when light interacts with objects much larger than its wavelength.

Maxwell's brilliant work shows that the phenomena of **optics**—the behavior of light—are manifestations of electromagnetism. Except in the atomic realm, where quantum physics reigns, all optical phenomena are understandable in terms of electromagnetic-wave fields described by Maxwell's equations. But when objects with which light or any other wave interacts are much larger than the wavelength, light generally travels in straight lines called **rays. Geometrical optics** describes the behavior of light in this approximation. Here we'll use geometrical optics to explore the behavior of light at interfaces between different materials. In Chapter 31 we'll see how geometrical optics explains lenses, the human eye, and many optical instruments.

30.1 Reflection

Some materials, notably metals, **reflect** nearly all the light incident on them. It's no coincidence that these materials are also good electrical conductors. The oscillating electric field of a light wave drives a metal's free electrons into oscillatory motion, which, in turn, produces electromagnetic waves. The net effect is to reradiate the wave back into the original medium. Other materials reflect only part of the incident light. Either way, reflection satisfies the same geometrical conditions: The incident ray, the reflected ray, and the normal to the interface between two materials all lie in the same plane. The **angle of reflection** θ_1' that the reflected ray makes with the normal is the same as the **angle of incidence** θ_1 made by the incident ray (Fig. 30.1*a*):

$$\theta_1' = \theta_1 \qquad (30.1)$$

where the subscript 1 denotes the first medium.

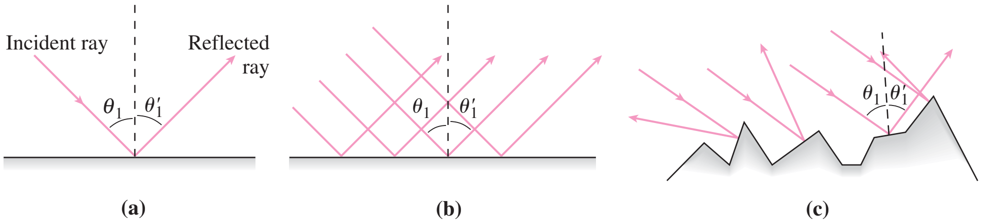


FIGURE 30.1 (a) Angles of reflection and incidence are equal. (b) In specular reflection, a smooth surface reflects a light beam undistorted. (c) A rough surface results in diffuse reflection.

In **specular reflection**, parallel rays reflect off a smooth surface and the entire beam is reflected without distortion (Fig. 30.1*b*). In contrast, a rough surface reflects individual rays in different directions (Fig. 30.1*c*)—even though each ray still obeys Equation 30.1. This is **diffuse reflection**. White paper is a diffuse reflector, while the aluminum coating of a mirror is an excellent specular reflector.

EXAMPLE 30.1 Reflection: The Corner Reflector

Two mirrors join at right angles. Show that any light ray incident in the plane of the page will return antiparallel to its incident direction.

INTERPRET We've sketched the situation in Fig. 30.2. We're asked to show that the lines representing incident and outgoing rays in the figure are parallel.

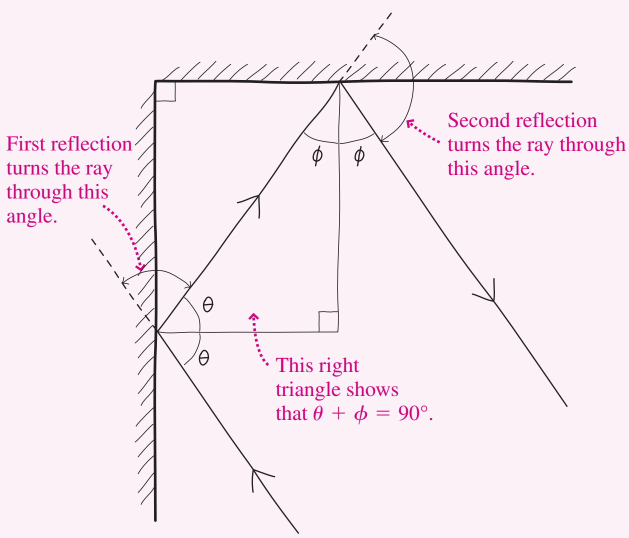


FIGURE 30.2 A two-dimensional corner reflector.

DEVELOP The outgoing ray will be antiparallel to the incident ray if the light turns through a total of 180°. Applying Equation 30.1 to each of the two reflections shows that the angles θ are the same, and so are the angles ϕ. Figure 30.2 shows that the first reflection turns the incident ray through an angle $180° - 2\theta$. Similarly, the second turns it through $180° - 2\phi$.

EVALUATE The pair of mirrors thus turns the ray through a total angle of

$$(180° - 2\theta) + (180° - 2\phi) = 360° - 2(\theta + \phi)$$

But Fig. 30.2 shows that $\theta + \phi = 90°$, so the total angle is $360° - 180° = 180°$—which is what we set out to prove.

ASSESS Here we've explored a remarkable device, a pair of perpendicular mirrors that returns a light ray in exactly the direction from which it came—provided the light is in the plane perpendicular to both mirrors. Add a third mirror at right angles to the other two, and you have a **corner reflector**—a device that returns a light ray in the direction from which it came, period. Corner reflectors, often made with prisms rather than mirrors, are widely used in optics. A corner reflector left on the Moon allows laser-based measurements of the Moon's distance with an accuracy of a few centimeters. (See Problem 58.) ■

Partial Reflection

Some light is reflected even at the interface with a transparent material. The detailed description of such partial reflection follows from Maxwell's equations, and is akin to the partial reflection of waves on strings described in Chapter 14. The least reflection occurs with normal incidence; for glass, about 4% of normally incident light is reflected. Reflection increases with larger incidence angles. Camera lenses, solar photovoltaic cells, and other devices often have special antireflection coatings to reduce light loss.

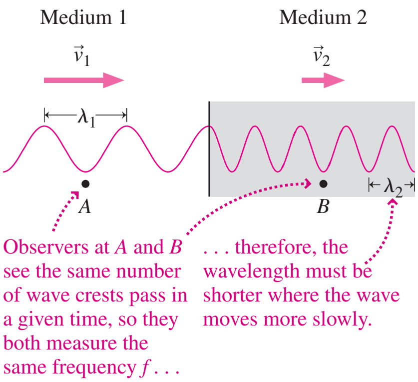


FIGURE 30.3 Wave frequency doesn't change as a wave goes from one medium to another, but wavelength does change.

Table 30.1 Indices of Refraction*

Substance	Index of Refraction, n
Gases	
Air	1.000293
Carbon dioxide	1.00045
Liquids	
Water	1.333
Ethyl alcohol	1.361
Glycerine	1.473
Benzene	1.501
Diiodomethane	1.738
Solids	
Ice (H_2O)	1.309
Polystyrene	1.49
Glass	1.5–1.9
Sodium chloride (NaCl)	1.544
Diamond (C)	2.419
Rutile (TiO_2)	2.62

*At 1 atm pressure and temperatures ranging from 0°C to 20°C, measured at a wavelength of 589 nm (the yellow line of sodium).

30.2 Refraction

We saw in Chapter 14 that wave speeds differ in different media. With light, the speed in transparent media is lower than in vacuum. We characterize a transparent medium by its **index of refraction**, defined as the ratio of the speed of light c in vacuum to the speed of light v in the medium:

$$n = \frac{c}{v} \qquad \text{(index of refraction)} \tag{30.2}$$

Although the wave speed changes when light enters a new medium, Fig. 30.3 shows that its frequency f can't change and therefore, since the wave speed is $v = f\lambda$, the wavelength must change. Equation 30.2 shows that the wavelength in a medium with refractive index n is $\lambda = v/f = c/nf$; since c and f don't change, the wavelength is inversely proportional to n. Table 30.1 lists some refractive indices.

When light is incident at an angle on a transparent material, the light transmitted into the material undergoes **refraction**—a change in its propagation direction (Fig. 30.4). Figure 30.5 shows how refraction results from the change in wave speed and therefore wavelength. Here we assume the refractive index is higher in medium 2; our result $\lambda = c/nf$ then shows that the wavelength is shorter in medium 2. We've shaded two right triangles with a common hypotenuse and one side equal to the appropriate wavelength. The angles opposite these sides are the angles of incidence and refraction. In each case the hypotenuse is given by $\lambda/\sin\theta$. Equating expressions for this common hypotenuse gives $\lambda_1/\sin\theta_1 = \lambda_2/\sin\theta_2$. Since $\lambda = c/nf$ with f the same in both media, we get **Snell's law**:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \qquad \text{(Snell's law)} \tag{30.3}$$

First developed geometrically in 1621 by van Roijen Snell of the Netherlands, and described analytically in the 1630s by René Descartes in France, Snell's law lets us predict what will happen at an interface given the refractive indices of the two media.

Snell's law applies whether light goes from a medium of lower to higher refractive index or vice versa. When going from lower to higher index, the beam bends *toward* the normal; when going from higher to lower index, it bends *away* from the normal.

In some situations, including the human eye and Earth's atmosphere, the refractive index varies continuously with position, so light refracts continuously, following a curved path. You can explore two examples in Passage Problems 62–65.

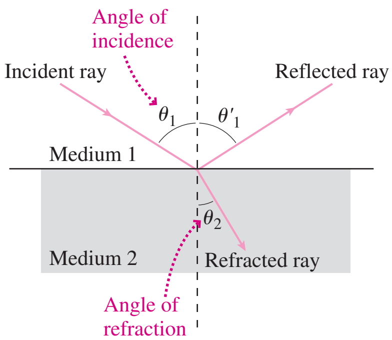


FIGURE 30.4 Refraction and reflection at an interface, here when medium 2 has the higher refractive index.

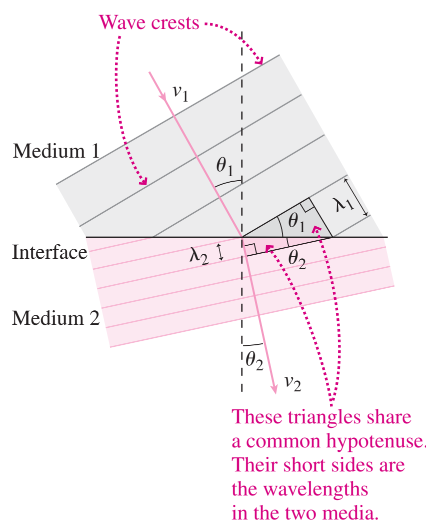


FIGURE 30.5 Refraction occurs because wave speed and wavelength differ in the two media.

EXAMPLE 30.2 Refraction: A Plane Slab

A light ray propagating in air strikes a glass slab of thickness d and refractive index n at incidence angle θ_1. Show that it emerges from the slab propagating parallel to its original direction.

INTERPRET This is a problem about refraction, which in this case occurs twice. We identify the two interfaces as first the air–glass interface where the light enters the glass and then the glass–air interface where the light exits the glass.

DEVELOP Figure 30.6 is a sketch showing the path of the light through the glass. There are two pairs of incidence and refractive angles, which we labeled θ_1, θ_2 and θ_3, θ_4. Our plan is to apply Snell's law at each interface and thus prove that $\theta_4 = \theta_1$. At the air–glass interface, we have $n_1 \sin\theta_1 = n_2 \sin\theta_2$, and at the glass–air interface, $n_3 \sin\theta_3 = n_4 \sin\theta_4$. Note that θ_1 and θ_4 are in the same medium (air), so we set $n_4 = n_1$. Similarly, $n_3 = n_2$ for θ_2 and θ_3 in glass. Then our equations become $n_1 \sin\theta_1 = n_2 \sin\theta_2$ and $n_2 \sin\theta_3 = n_1 \sin\theta_4$.

EVALUATE Taking $n_1 = 1$ for air and $n_2 = n$ for glass at the air–glass interface, we have $\sin\theta_2 = \sin\theta_1/n$. But at the glass–air interface, $n_1 = n$ and $n_2 = 1$, so here $\sin\theta_4 = n\sin\theta_3$. But the slab faces are parallel, so $\theta_3 = \theta_2$. So $\sin\theta_4 = n\sin\theta_2$. Using our expression for θ_2 from the first interface then gives

$$\sin\theta_4 = n\left(\frac{\sin\theta_1}{n}\right) = \sin\theta_1$$

showing that the incident and outgoing rays are indeed parallel.

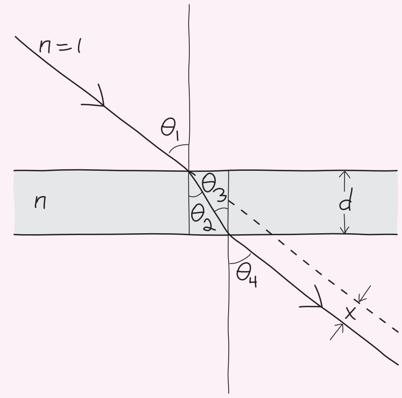


FIGURE 30.6 Light passing through a transparent slab.

ASSESS This result shows that light isn't deflected when it passes through a parallel-faced slab of transparent material. It is, however, displaced by the distance x shown in Fig. 30.6. You can find that displacement in Problem 54.

EXAMPLE 30.3 Refraction: CD Music

The laser beam that "reads" information from a compact disc is 0.737 mm wide when it strikes the disc, and it forms a cone with half-angle $\theta_1 = 27.0°$ as shown in Fig. 30.7. It then passes through a 1.20-mm-thick layer of plastic with refractive index 1.55 before reaching the reflective information layer near the disc's top surface. What's the beam diameter d at the information layer?

INTERPRET Rays defining the beam refract toward the normal, making a smaller convergence angle within the disc. We're asked how much this converging beam narrows when it reaches the information layer.

DEVELOP Snell's law will give us the angles θ_2 in the figure: $n_1 \sin\theta_1 = n_2 \sin\theta_2$. Given θ_2, we can use trigonometry to find the distance x marked in the drawing: $x = t\tan\theta_2$, with t the 1.2-mm thickness. The beam diameter d then follows from $d = D - 2x$, where D is the 0.737-mm beam diameter at the disc surface.

EVALUATE With $n_1 = 1$ and $n_2 = 1.55$, Snell's law gives $\theta_2 = \sin^{-1}(\sin\theta_1/n_2) = 17.03°$, so $d = D - 2x = D - 2t\tan\theta_2 = 1.80\ \mu\text{m}$.

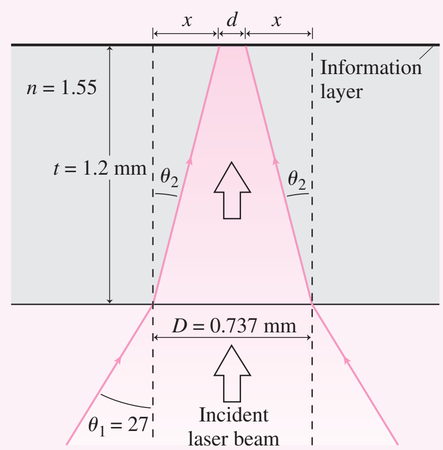
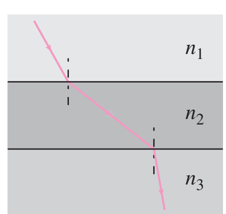


FIGURE 30.7 Section through a compact disc, showing convergence of the laser beam to a narrow spot at the information layer.

ASSESS This answer makes sense because d is just a bit larger than the "pits" cut into the CD to store information. Narrowing of the laser beam plays a crucial role in keeping CDs noise free. The tiniest dust speck would blot out information at the μm-scale information layer, but at the point where the beam enters the disc, it would take mm-size dust to cause problems. We'll explore CD and DVD technology further in Chapter 32.

GOT IT? 30.1 The figure shows the path of a light ray through three different media. Rank the media according to their refractive indices in decreasing order.



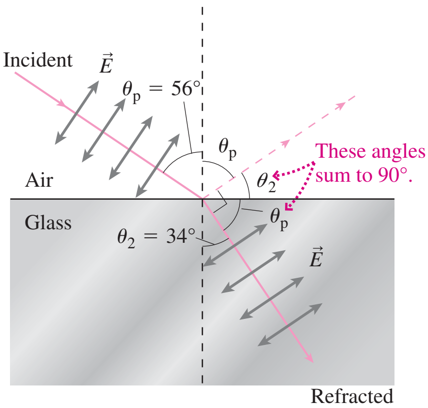


FIGURE 30.8 The polarizing incidence angle θ_p occurs when the angles of incidence and refraction sum to 90°.

Refraction, Reflection, and Polarization

Both reflection and refraction ultimately involve interactions between the incident light wave's electric field and charges in a material. The oscillation of molecular dipoles in response to the field gives rise to both refracted and reflected light. It's not surprising, therefore, that details depend on the direction of the electric field—that is, on polarization. When the field lies in the plane defined by the incident and reflected rays, there's a special angle of incidence at which no reflection occurs. This is the **Brewster angle**, or **polarizing angle**, and it occurs when the reflected ray would be perpendicular to the transmitted ray (Fig. 30.8). Then the molecular dipoles are oscillating along the direction a reflected ray would take, and, as we saw in Section 29.7, there's no electromagnetic radiation from an oscillating dipole along the oscillation direction.

Figure 30.8 shows that the polarizing incidence angle θ_p occurs when θ_p and the angle θ_2 of the reflected ray sum to 90°; equivalently, $\theta_2 = 90° - \theta_p$. Since $\sin\theta = \cos(90° - \theta)$, that means $\sin\theta_2 = \cos\theta_p$. Now, Snell's law gives $\sin\theta_2 = (n_1/n_2)\sin\theta_p$; substituting $\cos\theta_p$ for $\sin\theta_2$, this becomes $\cos\theta_p = (n_1/n_2)\sin\theta_p$. Multiplying both sides by (n_2/n_1) and dividing by $\cos\theta_p$ then gives

$$\tan\theta_p = \frac{n_2}{n_1} \qquad \text{(polarizing angle)} \qquad (30.4)$$

For the air–glass interface shown in Fig. 30.8, θ_p is about 56°.

When unpolarized light is incident at the polarizing angle, only the component of the light's electric field that's perpendicular to the plane of Fig. 30.8 gets reflected, and the result is polarized light. Polarization using this effect is important in a number of technologies, including lasers. The window through which light emerges from a laser is usually cut at the polarizing angle, and as a result most laser light is intrinsically polarized. A similar polarizing phenomenon occurs for reflection from metals and other opaque surfaces, and it results in a glare that polarizing sunglasses can eliminate.

30.3 Total Internal Reflection

Light propagating from a medium with a higher refractive index into one with a lower index is bent *away* from the normal, as shown for a glass–air interface in Fig. 30.9. In other words, the angle of refraction in this case is larger than the angle of incidence. So at some incidence angle, the angle of refraction becomes 90°. Then what?

As Fig. 30.9 shows, light incident at this **critical angle** or larger cannot escape from the glass. Instead, **total internal reflection** occurs, returning all the light to the medium with the larger refractive index. We can find the critical angle by setting $\theta_2 = 90°$ in Snell's law (Equation 30.3). The critical angle is then θ_1, and we have

$$\sin\theta_c = \frac{n_2}{n_1} \qquad \text{(critical angle)} \qquad (30.5)$$

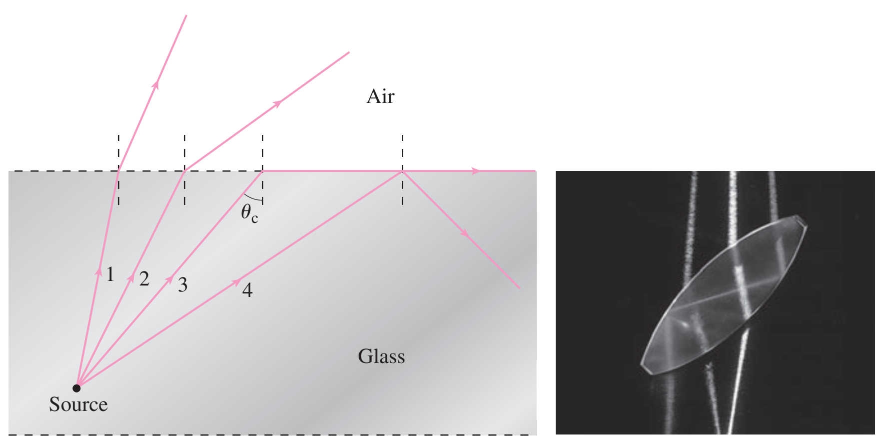


FIGURE 30.9 Light propagating in glass is refracted away from the normal at the glass–air interface. Ray 3, incident at the critical angle θ_c, just skims along the interface. At larger incidence angles (ray 4), the light undergoes total internal reflection. The rightmost beam in the photo (incident from above) undergoes two total internal reflections.

Total internal reflection makes uncoated glass an excellent reflector when it's oriented appropriately (Fig. 30.10). Binoculars owe their compact size to glass prisms that reflect light internally to provide a longer light path. For an underwater observer, the existence of the critical angle affects the view of the outside world, as the next example shows. Finally, total internal reflection is the basis of the optical fibers that carry signals over the global Internet, as the Application below describes.

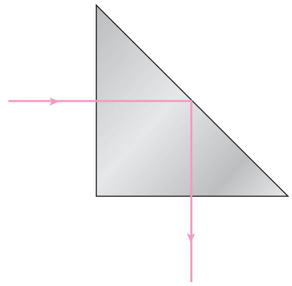


FIGURE 30.10 Light undergoes total internal reflection in a glass prism.

GOT IT? 30.2 The glass prism in Fig. 30.10 has $n = 1.5$ and is surrounded by air $(n = 1)$. What would happen to the incident light ray shown if the prism were immersed in water $(n = 1.333)$?

CONCEPTUAL EXAMPLE 30.1 Total Internal Reflection: A Watching Whale

Planeloads of whale watchers fly over the ocean. What does a whale see when it looks up at them?

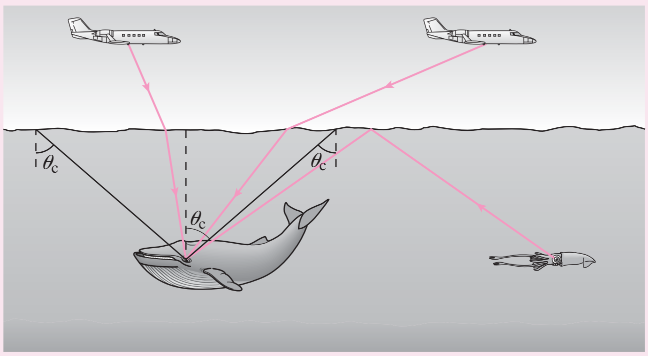


FIGURE 30.11 The whale sees the entire world above the surface in a cone of half-angle θ_c; beyond that, it sees reflections of objects below the surface.

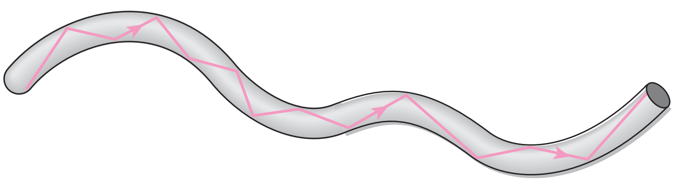
EVALUATE The whale is underwater, so it's in a medium with a higher refractive index than air. Some of the light reaching the whale is from objects above the water surface, like the planes in Fig. 30.11. But the whale also sees objects below the surface, like the squid in Fig. 30.11, light from which is totally reflected at the water–air interface.

ASSESS The whale sees the entire above-surface world within a cone, as Fig. 30.11 shows. If you've ever looked upward from underwater, you've experienced the same phenomenon.

MAKING THE CONNECTION What's the half-angle of the cone in which the whale sees objects above the water surface?

ANSWER The cone's half-angle is the critical angle θ_c, as shown in Fig. 30.11. For water, Table 30.1 gives $n = 1.333$, so, by Equation 30.5, $\theta_c = \sin^{-1}(1/1.333) = 48.6°$.

APPLICATION Optical Fibers



Refraction and total internal reflection are the basis for **optical fibers,** which carry much of the world's communications. Optical fibers provide the physical connectivity of the global Internet and handle information ranging from telephone and television to light signals within medical, astronomical, and industrial instruments.

A typical fiber consists of a glass core only 8 μm in diameter, surrounded by a so-called cladding consisting of glass with a lower refractive index. Total internal reflection at the core–cladding interface guides light along the fiber, as shown in the figure. The glass used in optical fibers is so clear that a 1-km-thick slab would be as transparent as an ordinary window pane. Today's fibers carry light produced by semiconductor lasers at infrared wavelengths of 850, 1350, or 1550 nm.

An optical fiber's main advantage over copper wire is its huge rate of information flow, called **bandwidth**. Communicating information—audio, video, or digital data—requires a range of frequencies, and the greater the rate of information transfer, the wider that range. With its frequency of around 10^{14} Hz, light can accommodate a much wider frequency range within a channel than can radio-frequency communication systems at 10^{10} Hz. A single optical fiber, for example, can carry tens of thousands of simultaneous telephone conversations. Fibers are also lighter and more rugged than copper cables, and they're less vulnerable than copper or open-air transmission to illicit tapping. And because they're made from insulators, optical fibers are less susceptible to electrical noise. The photo shows two cables that can carry information at the same rate. One consists of a few optical fibers while the other is a thick bundle of copper wires.

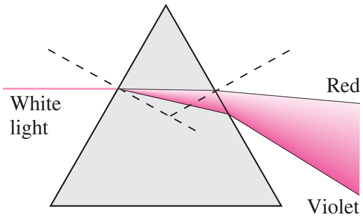


FIGURE 30.12 Dispersion separates the colors in white light, with shorter-wavelength violet experiencing the greatest refraction.

30.4 Dispersion

Refraction ultimately involves the interaction of electromagnetic-wave fields with atomic electrons. It's not surprising that the electrons' behavior and therefore also the refractive index depend on frequency. That means different frequencies—different colors of visible light—refract through different angles. The classic example of this **dispersion** is Newton's demonstration that white light is a mixture of all colors in the visible spectrum (Fig. 30.12). The rainbow is a beautiful natural manifestation of dispersion combined with internal reflection, as the Application below describes.

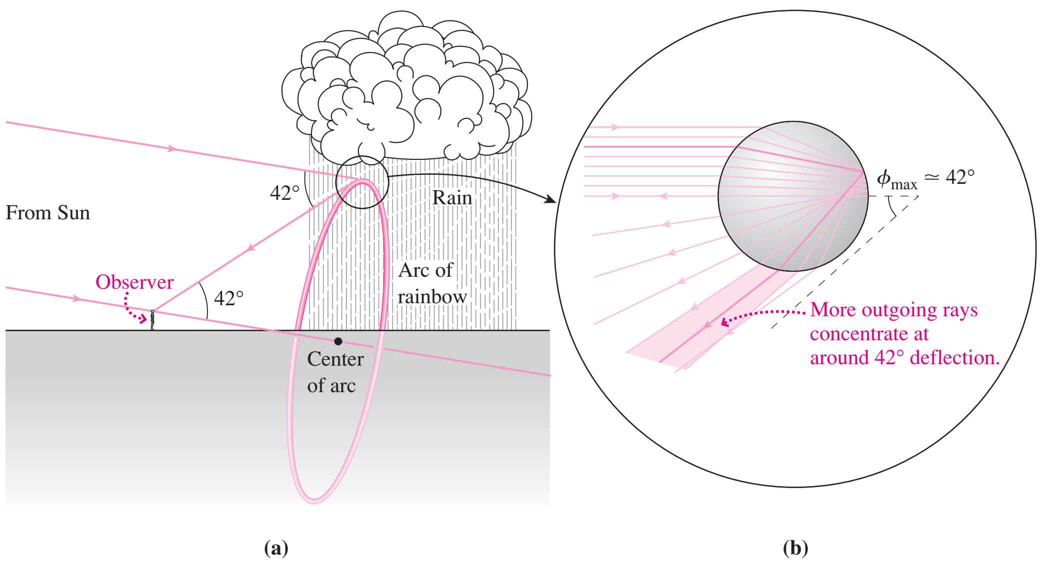
APPLICATION The Rainbow

Rainbows occur when sunlight strikes rain or other airborne water droplets. An observer standing between the Sun and the rain then sees a circular arc of colored bands. Part (a) of the figure shows that the center of that arc lies on the line joining the Sun to the observer's head. That means each observer sees a different rainbow! Furthermore, the rainbow's arc always subtends an angle of about 42°.

Isaac Newton provided the first full explanation of the rainbow, invoking both internal reflection and dispersion. Part (b) of the figure shows light passing through a spherical raindrop. Parallel rays striking the curved drop experience a range of incidence angles, giving a range of angles ϕ between incident and outgoing rays. As the figure shows, however, there's a maximum angle ϕ_{max} of about 42°, and more light returns at angles close to ϕ_{max} than at other angles. That's why the rainbow appears as a broad arc at an angle of about 42° to the direction of the Sun's rays. Problems 55 and 56 detail how to find ϕ_{max}.

The "bunching" of light rays near ϕ_{max} shows why a bright band appears, but why the different colors? The refractive index varies with wavelength, and so, therefore, does ϕ_{max}. Thus each color appears at a slightly different angle. For water, the refractive index ranges from $n_{red} = 1.330$ to $n_{violet} = 1.342$. Using these values with the results of Problems 55 and 56 yields $\phi_{red} = 42.53°$ and $\phi_{violet} = 40.78°$. Thus the rainbow appears as a band of colors subtending an angle of about 1.75°, with red at the top.

You'll occasionally see a fainter *secondary rainbow* above the primary arc. This results from two internal reflections, which causes the order of colors to be reversed. Problem 57 explores the secondary rainbow.



(a) The rainbow is a circular arc at 42° from the line connecting the Sun, the observer, and the center of the arc.
(b) The rainbow results from total internal reflection in raindrops, concentrating light at about 42° deflection. Dispersion separates wavelengths slightly, resulting in the rainbow's colors.

Dispersion is the basis of **spectroscopy**, the analysis of light and other electromagnetic radiation in terms of its constituent wavelengths. Hot, dense objects emit a continuous range of wavelengths, while diffuse gases emit and absorb radiation at only a few specific wavelengths (Fig. 30.13). Such discrete spectra provide some of the strongest evidence for the nature of atoms, and today spectroscopy is a powerful tool throughout the sciences. Spectroscopy helps astronomers to determine the composition and motions of distant astrophysical objects, geologists to identify minerals, and chemists to study molecules. Although early spectroscopy used prisms, most modern instruments use instead diffraction gratings, which we'll describe in Chapter 32.

FIGURE 30.13 The emission spectrum of a hot, diffuse gas—here hydrogen—consists of light at discrete wavelengths.

Dispersion can be a nuisance in optical systems. Glass lenses, for example, focus different colors at different points, resulting in distortion known as *chromatic aberration*. Dispersion in optical fibers—based not on wavelength but on different paths taken by rays reflecting at different angles—can degrade digital information. So-called single-mode fibers reduce this effect by passing only those rays that have a single specific reflection angle. On the other hand, dispersion of radio waves provides a crucial correction to the global positioning system (GPS). Ionization in the upper atmosphere introduces an uncertain but frequency-dependent variation in the travel time for radio waves from GPS satellites. It's this travel time that provides GPS location information. Sending waves at two different frequencies and comparing their travel times reveals the atmospheric conditions, and makes dual-frequency GPS receivers accurate to within a few centimeters.

CHAPTER 30 SUMMARY

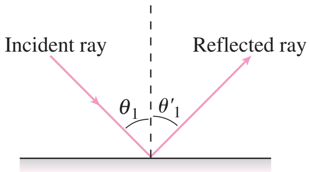
Big Picture

The big idea here is that light can be considered to travel in straight **rays** when the objects with which it interacts are much larger than the wavelength. Under these conditions, light rays **reflect** and **refract** at interfaces between different materials.

Key Concepts and Equations

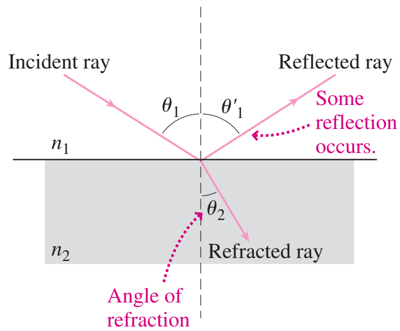
The **angle of incidence** and **angle of reflection** are equal:

$$\theta_1' = \theta_1$$



Snell's law relates the angle of incidence and angle of refraction:

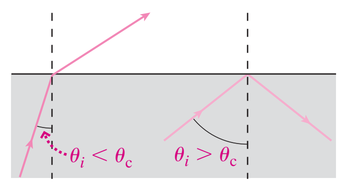
$$n_1 \sin\theta_1 = n_2 \sin\theta_2$$



Applications

Total internal reflection results when light is incident at greater than the **critical angle**, θ_c, on an interface with a medium with lower refractive index n_2:

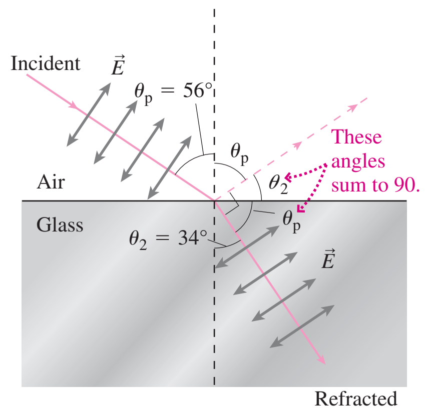
$$\sin\theta_c = \frac{n_2}{n_1}$$



Light polarized in the plane of the incident and refracted rays undergoes no reflection at an interface; this special **polarizing angle**, θ_p, is given by

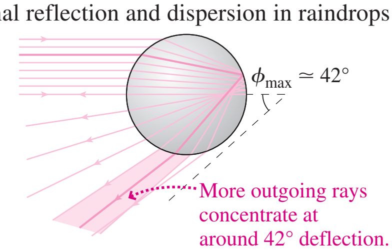
$$\tan\theta_p = \frac{n_2}{n_1}$$

For an air–glass interface, $\theta_p \simeq 56°$.

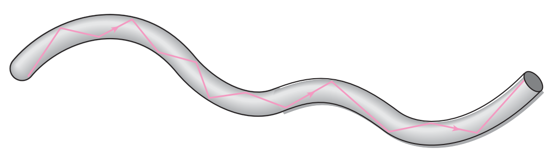


Dispersion results from the wavelength dependence of the refractive index and causes different colors to refract at different angles.

A combination of total internal reflection and dispersion in raindrops accounts for the rainbow.



Total internal reflection guides signals in optical fibers.



For Thought and Discussion

1. Why is it usually inappropriate to consider low-frequency sound waves as traveling in rays? Why is the ray approximation more appropriate for high-frequency sound and for light?
2. Why does a spoon appear bent when it's in a glass of water?
3. Why do a diamond and an identically shaped piece of glass sparkle differently?
4. White light goes from air through a glass slab with parallel surfaces. Will its colors be dispersed when it emerges from the glass?
5. You send white light through two identical glass prisms, oriented as shown in Fig. 30.14. Describe the beam that emerges from the right-hand prism.

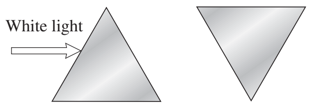


FIGURE 30.14 For Thought and Discussion 5

6. In glass, which end of the visible spectrum has the smallest critical angle for total internal reflection?
7. Why can't you walk to the end of the rainbow?
8. What's wrong with Fig. 30.15, which shows rainbows over Niagara Falls? (*Hint:* The rainbow subtends a half-angle of 42°.)

FIGURE 30.15 For Thought and Discussion 8. The painting is Harry Fenn's *Niagara*.

9. Why are polarizing sunglasses better than glasses that simply reduce the total amount of light?
10. Under what conditions will the polarizing angle be smaller than 45°?

Exercises and Problems

Exercises

Section 30.1 Reflection

11. Through what angle should you rotate a mirror so that a reflected ray rotates through 30°?
12. The mirrors in Fig. 30.16 make a 60° angle. A light ray enters parallel to the symmetry axis, as shown. (a) How many reflections does it make? (b) Where and in what direction does it exit the mirror system?

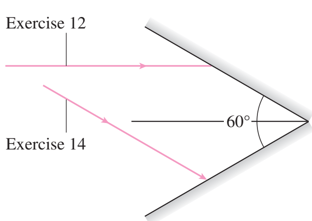


FIGURE 30.16 Exercises 12 and 14 and Problem 28

13. To what angular accuracy must two ostensibly perpendicular mirrors be aligned so that an incident ray returns within 1° of its incident direction?
14. If a light ray enters the mirror system of Fig. 30.16 propagating in the plane of the page and parallel to one mirror, through what angle will it be turned?

Section 30.2 Refraction

15. In which substance in Table 30.1 does the speed of light have the value 2.292×10^8 m/s?
16. Information in a compact disc is stored in "pits" whose depth is essentially one-fourth the wavelength of the laser light used to "read" the information. That wavelength is 780 nm in air, but the wavelength on which the pit depth is based is measured in the $n = 1.55$ plastic that makes up most of the disc. Find the pit depth.
17. Light is incident on an air–glass interface, and the refracted light in the glass makes a 40° angle with the normal to the interface. The glass has refractive index 1.52. Find the incidence angle.
18. A light ray propagates in a transparent material at 15° to the normal to the surface. It emerges into the surrounding air at 24° to the normal. Find the material's refractive index.
19. Light propagating in the glass $(n = 1.52)$ wall of an aquarium tank strikes the wall's interior surface with incidence angle 12.4°. What's the angle of refraction in the water?
20. Find the polarizing angle for diamond when light is incident from air.
21. Find the refractive index of a material for which the polarizing angle in air is 62°.

Section 30.3 Total Internal Reflection

22. Find the critical angle for total internal reflection in (a) ice, (b) polystyrene, and (c) rutile, when the surrounding medium is air.
23. A drop of water is trapped in a block of ice. What's the critical angle for total internal reflection at the water–ice interface?
24. What is the critical angle for light propagating in glass with $n = 1.52$ when the glass is immersed in (a) water, (b) benzene, and (c) diiodomethane?
25. Total internal reflection occurs at an interface between plastic and air at incidence angles greater than 37°. Find the plastic's refractive index.

Section 30.4 Dispersion

26. Blue and red laser beams strike an air–glass interface with incidence angle 50°. If the glass has refractive indices of 1.680 for the blue light and 1.621 for the red, what will be the angle between the two beams in the glass?
27. White light propagating in air is incident at 45° on the equilateral prism of Fig. 30.17. Find the angular dispersion γ of the outgoing beam if the prism has refractive indices $n_{\text{red}} = 1.582$ and $n_{\text{violet}} = 1.633$.

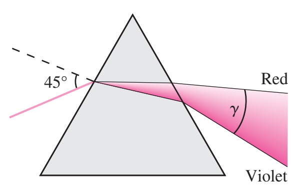


FIGURE 30.17 Exercise 27 (angles of dispersed rays aren't accurate)

Problems

28. Suppose the 60° angle in Fig. 30.16 is changed to 75°. A ray enters the mirror system parallel to the axis. (a) How many reflections does it make? (b) Through what angle is it turned when it exits the system?

29. **BIO** The refractive index of a human cornea is 1.40. If 550-nm light strikes a cornea at incidence angle 25°, find (a) the angle of refraction and (b) the wavelength in the cornea.
30. Two plane mirrors make an angle ϕ. A light ray enters the system and is reflected once off each mirror. Show that the ray is turned through an angle $360° - 2\phi$.
31. An unlabeled bottle of liquid has spilled, and you're trying to find out whether it's relatively harmless ethyl alcohol or toxic benzene. You submerge a glass block with $n = 1.52$ in the liquid, and shine a laser beam so it strikes the submerged glass with incidence angle 31.5°. You measure the angle of refraction in the glass at 27.9°. Which liquid is it? (See Table 30.1.)
32. A meter stick lies on the bottom of the rectangular tank in Fig. 30.18, with its zero mark at the tank's left edge. You look into the long dimension of the tank at a 45° angle, with your line of sight just grazing the top edge, as shown. What mark on the meter stick do you see when the tank is (a) empty, (b) half full of water, and (c) full of water?

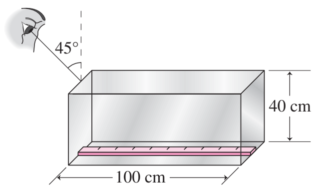


FIGURE 30.18 Problem 32

33. You look at the center of one face of a solid glass cube of glass, on a line of sight making a 55° angle with the normal to the cube face. What minimum refractive index of the glass will let you see through the cube's opposite face?
34. At the aquarium where you work, a fish has gone missing in a 10-m-deep, 11-m-diameter cylindrical tank. You shine a flashlight in from the top edge of the tank, hoping to see if the missing fish is on the bottom. What's the smallest angle your flashlight beam can make with the horizontal if it's to illuminate the bottom?
35. You're standing 2.3 m horizontally from the edge of a 4.5-m-deep lake, with your eyes 1.7 m above the water's surface. A diver holding a flashlight at the lake bottom shines the light so you can see it. If the light in the water makes a 42° angle with the vertical, at what horizontal distance is the diver from the edge of the lake?
36. You've dropped your car keys at night off the end of a dock into water 1.6 m deep. A flashlight held directly above the dock edge and 0.50 m above the water illuminates the keys when it's aimed at 40° to the vertical, as shown in Fig. 30.19. What's the horizontal distance x from the edge of the dock to the keys?

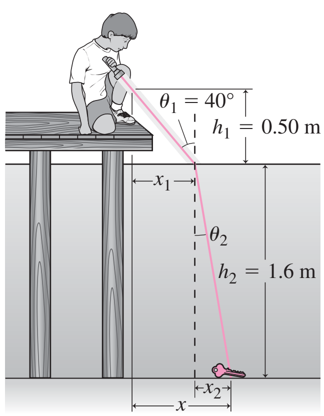


FIGURE 30.19 Problem 36

37. **BIO** Laser eye surgery uses ultraviolet light with wavelength 193 nm. What's the UV light's wavelength within the eye's lens, where $n = 1.39$?
38. The prism in Fig. 30.20 has $n = 1.52$ and $\alpha = 60°$ and is surrounded by air. A light beam is incident at $\theta_1 = 37°$. Find the angle δ through which the beam is deflected.

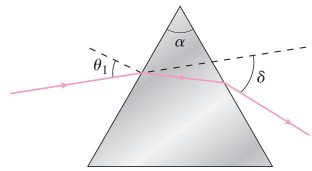


FIGURE 30.20 Problems 38 and 39

39. Repeat Problem 38 for the case $n = 1.75$, $\alpha = 40°$, and $\theta_1 = 25°$.
40. Find the minimum refractive index for the prism in Fig. 30.10 if total internal reflection occurs as shown when the prism is surrounded by air.
41. Where and in what direction would the main beam emerge if the prism in Fig. 30.10 were made of ice, surrounded by air?
42. Find the speed of light in a material for which the critical angle at an interface with air is 61°.
43. The prism of Fig. 30.10 has $n = 1.52$. When it's immersed in a liquid, a beam incident as shown in the figure ceases to undergo total reflection. What's the minimum value for the liquid's refractive index?
44. For the interface between air (refractive index 1) and a material with refractive index n, show that the critical angle and the polarizing angle are related by $\sin\theta_c = \cot\theta_p$.
45. A scuba diver sets off a camera flash at depth h in water with refractive index n. Show that light emerges from the water's surface through a circle of diameter $2h/\sqrt{n^2 - 1}$.
46. Suppose the red and blue beams of Exercise 26 are now propagating in the same direction *inside* the glass. For what range of incidence angles on the glass–air interface will one beam be totally reflected and the other not?
47. A compound lens is made from crown glass $(n = 1.52)$ bonded to flint glass $(n = 1.89)$. What's the critical angle for light incident on the flint–crown interface?
48. Find a simple expression for the speed of light in a material in terms of c and the critical angle at an interface between the material and vacuum.
49. Find the polarizing angle for light incident from below on the surface of a pond.
50. A cylindrical tank 2.4 m deep is full to the brim with water. Sunlight first hits part of the tank bottom when the rising Sun makes a 22° angle with the horizon. Find the tank's diameter.
51. For what diameter tank in Problem 50 will sunlight strike some part of the tank bottom whenever the Sun is above the horizon?
52. Light is incident from air on the flat wall of a polystyrene water tank. If the incidence angle is 40°, what is the angle of refraction in the water?
53. **BIO** You're an optometrist, mounting a projector at the back of your 4.2-m-long exam room, 2.6 m above the floor. It shines an eye-test pattern on the opposite wall. Patients will sit with their eyes 3.3 m from the wall and 1.4 m above the floor to view the pattern. At what height should you center the pattern on the wall?
54. Find an expression for the displacement x in Fig. 30.6, in terms of θ_1 d, and n.
55. Figure 30.21 shows light passing through a spherical raindrop, undergoing two refractions and total internal reflection, resulting

in an angle ϕ between the incident and outgoing rays. Show that $\phi = 4\sin^{-1}(\sin\theta/n) - 2\theta$, where θ is the incidence angle.

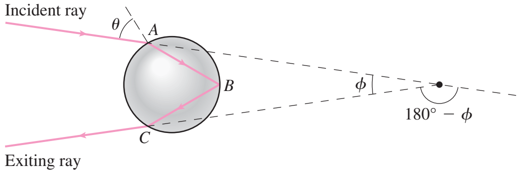


FIGURE 30.21 Problem 55

56. (a) Differentiate the result of Problem 55 to show that the maximum value of ϕ occurs when the incidence angle θ is given by $\cos^2\theta = \frac{1}{3}(n^2 - 1)$. (b) Use this result and that of Problem 55 to find the maximum ϕ in a raindrop with $n = 1.333$. This is the angle at which the rainbow appears, as shown in the Application on page 538.
57. Figure 30.22 shows the approximate path of a light ray that undergoes internal reflection twice in a spherical water drop. Repeat Problems 55 and 56 for this case to find the angle at which the secondary rainbow occurs.

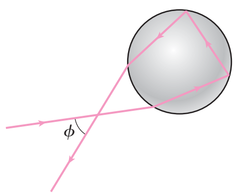


FIGURE 30.22 Problem 57

58. Show that a three-dimensional corner reflector (three mutually perpendicular mirrors, or a solid cube in which total internal reflection occurs) turns an incident light ray through 180°. (*Hint:* Let $\vec{q} = q_x\hat{\imath} + q_y\hat{\jmath} + q_z\hat{k}$ be a vector in the propagation direction. How does this vector get changed on reflection by a mirror in a plane defined by two of the coordinate axes?)
59. *Fermat's principle* states that a light ray's path is such that the time to traverse that path is an extremum (a minimum or a maximum) when compared with times for nearby paths. Show that Fermat's principle implies Snell's law by proving that a light ray going from point *A* in one medium to point *B* in a second medium will take the least time if it obeys Snell's law.
60. You're an automotive engineer charged with evaluating safety glass, which is made by bonding a layer of flexible plastic between two layers of glass, thus eliminating dangerous glass fragments during accidents. A new product uses glass with refractive index $n = 1.55$ and plastic with $n = 1.48$. You're asked to determine whether total internal reflection at the glass–plastic interface could cause problems with visibility. What do you conclude, and why?
61. A slab of transparent material has thickness d and refractive index n that varies across the material: $n(x) = n_1 + (n_2 - n_1)(x/d)^2$, where x is measured from one face of the slab. A light ray is incident normally on the slab. Find an expression for the time it takes to traverse the slab.

Passage Problems

Mirages occur when air's refractive index varies with position as a result of uneven heating. Under such conditions, light undergoes refraction continually and thus follows a curved path. Other examples where a varying refractive index is important include the eye's lens and Earth's *ionosphere*, an electrically conductive layer in the upper atmosphere, where the refractive index for radio waves varies with altitude.

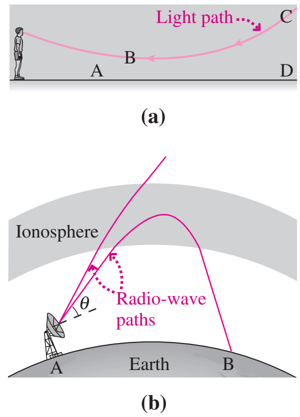


FIGURE 30.23 Passage Problems 62–65. (a) Light path in a mirage. (b) Long-distance radio communication via ionospheric refraction (not to scale).

62. Figure 30.23*a* depicts light's path over a hot road, producing a mirage. From the path shown, you can conclude that the air's refractive index
    a. increases from left to right.
    b. increases from right to left.
    c. increases upward.
    d. increases downward.
63. The observer in Fig. 30.23*a* sees a shimmering mirage that looks like water but actually results from sky light following the curved path. To the observer, the mirage appears to be at
    a. point A.
    b. point B.
    c. point C.
    d. point D.
64. Figure 30.23*b* shows how continuous refraction in the ionosphere enables long-distance radio communication. Waves launched at angles steeper than θ don't refract enough to return to Earth, so they propagate through the ionosphere and on to space. You can therefore conclude that
    a. all points between A and B receive stronger signals from A than point B receives.
    b. points between A and B can't receive signals from A via the ionosphere.
    c. the refractive index must become infinite at the maximum altitude of the radio signal.
65. The refractive index in the ionosphere is strongly dependent on radio-wave frequency, approaching 1 for high frequencies. Therefore,
    a. long-distance communication via the ionosphere is more likely at higher frequencies.
    b. higher frequencies won't penetrate as far into the ionosphere.
    c. higher frequencies are more appropriate for satellite-based communication.

Answers to Chapter Questions

Answer to Chapter Opening Question

Light from the bee propagates along the solid, curved plastic "light pipe" via total internal reflection. This same process carries e-mail and other computer data on the fiber-optic cables that constitute the physical structure of the Internet.

Answers to GOT IT? Questions

30.1. $n_3 > n_1 > n_2$.
30.2. Most would emerge into the water from the diagonal interface; some would still be reflected as shown.

31 Images and Optical Instruments

How does laser surgery provide permanent vision correction?

New Concepts, New Skills

By the end of this chapter you should be able to

- Describe graphically how plane and curved mirrors form images (31.1).
- Explain both graphically and quantitatively how lenses form images (31.2).
- Describe the difference between real and virtual images (31.1, 31.2).
- Give the details of refraction at curved surfaces, and be able to calculate the focal lengths of thin lenses (31.3).
- Explain the operation of common optical instruments and the human eye, including vision correction (31.4).

Connecting Your Knowledge

- This chapter is based on the optical phenomena of reflection and refraction introduced in Chapter 30 (30.1, 30.2).
- We'll be working extensively with geometrical constructions, especially similar triangles.

Reflection and refraction alter the direction of light propagation, according to the laws developed in Chapter 30. Microscopes, telescopes, cameras, contact lenses, scanners, and your own eyes use reflection or refraction to form **images** that provide visual representations of reality. Here we study image formation using geometrical optics—a valid approximation provided the objects we're imaging are much larger than the wavelength of light.

When we view an object through an optical system, light reflects or refracts, so it doesn't propagate in straight paths from the object. As a result we see an image that may differ in size, orientation, or apparent position from the actual object. In some cases light actually comes from the image to our eyes; the image is then a **real image**. In other cases light only *appears* to come from the image location; then the image is a **virtual image**.

31.1 Images with Mirrors

Plane Mirrors

In Fig. 31.1*a* we show three light rays that leave the tip of an arrowhead and reflect off a plane mirror to reach the observer's eye. The light looks to the observer like it's coming from a point behind the mirror, so that's the location of the arrowhead's image. The image is virtual because no light actually comes from behind the mirror—even though that's where the image is.

Since two lines define a point, we need only two rays to locate the arrowhead in Fig. 31.1*a*. We've repeated the image-location procedure in Fig. 31.1*b*, now using as one of the rays the ray that reflects normally. The same procedure locates the bottom of the arrow, and we could easily fill in to get the entire arrow; the resulting image is shown in Fig. 31.1*b*.

Because the angles of incidence and reflection are equal, angles OQP and $O'QP$ in Fig. 31.1*b* are equal. The right triangles OPQ and $O'PQ$ share a common side as well, so they're congruent. Thus the distances OP and $O'P$ are equal, so the image is located behind the mirror as far as the object is in front of it. Furthermore, rays from the top and bottom of the arrow and normal to the mirror are parallel, so the image is the same height as the actual arrow.

Images in plane mirrors preserve an object's length and upright orientation, but they reverse the object. When you look in the mirror, you face the mirror. So does your image—meaning the reversal is front to back, not left to right as you might think. This front-to-back reversal makes the image of your right hand look like a left hand (Fig. 31.2). Mathematically, the mirror reverses the coordinate axis perpendicular to the mirror plane. This alters handedness, rotation, and all other phenomena connected with the right-handed coordinate systems we've been using.

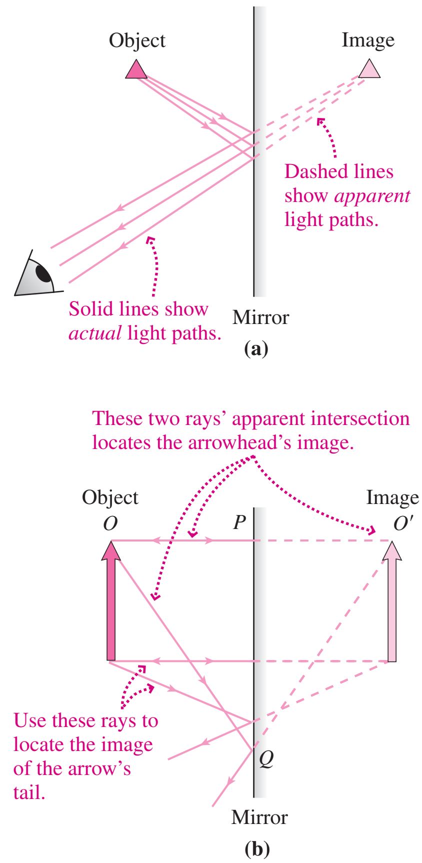


FIGURE 31.1 Image formation in a plane mirror.

FIGURE 31.2 The palm of a right hand faces the mirror. So does the image's palm. That makes the image look like a left hand, but it's still the image of a *right* hand.

GOT IT? 31.1 You stand in front of a plane mirror whose top is at the same height as the top of your head. Approximately how far down must the mirror extend for you to see your full image?

Curved Mirrors

In contrast to plane mirrors, curved mirrors form images that may be upright or inverted, virtual or real, large or small. The best curved mirrors are parabolic. That's because any line parallel to the parabola's axis makes the same angle to the normal of the parabola as does a second line drawn to a special point called the **focus** or **focal point** (Fig. 31.3). Because the angles of incidence and reflection are equal, this means a parabolic mirror reflects rays parallel to the mirror axis so they converge at the focus. This effect is used to concentrate light to high intensities or, conversely, to create a parallel beam from a point source of light at the focus.

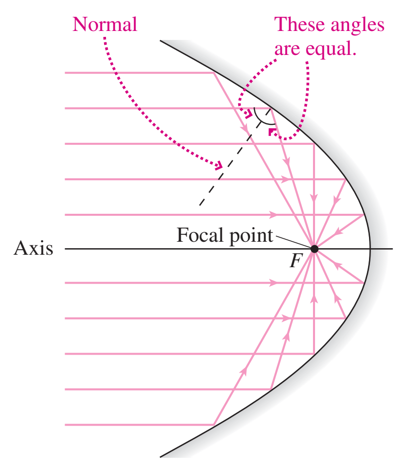


FIGURE 31.3 A parabolic mirror reflects rays parallel to its axis to a common focus.

Near the apex of the parabolic mirror in Fig. 31.3, you can't tell whether the shape is parabolic or spherical; a sphere closely approximates the parabola. Because a spherical surface

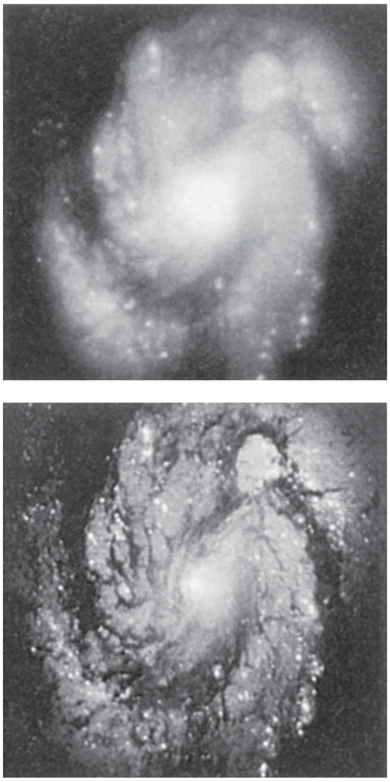


FIGURE 31.4 Incorrect curvature gave the Hubble Space Telescope mirror substantial spherical aberration. Astronauts later installed corrective optics. Images are of the same galaxy before and after the repair.

is easier to form, many focusing mirrors are, in fact, spherical. The slight distortion this causes is called **spherical aberration**; a notable case is the Hubble Space Telescope mirror, which was ground to the wrong curve and had substantial aberration (Fig. 31.4). Normally spherical aberration is minimized by making the mirror only a tiny fraction of the entire sphere. In that case the **focal length**—the distance from the mirror's apex to its focal point—is much larger than the mirror, so most rays striking the mirror are nearly parallel to the mirror axis. It's only for such **paraxial rays** that the approximation of a parabola by a sphere results in accurate focusing. But for clarity our diagrams will often show mirrors with exaggerated curvature, and consequently not all rays will seem paraxial.

We can see how spherical mirrors form images by tracing two rays from each of several points on the object, as we did for plane mirrors. Some special rays simplify this process; their properties all follow from the law of reflection and the properties of a spherical mirror in the paraxial approximation.

TACTICS 31.1 Ray Tracing with Mirrors

Figure 31.5 shows four special rays, any two of which suffice to locate an image:

1. Any ray parallel to the mirror axis reflects through the focal point F.
2. Conversely, any ray that passes through F reflects parallel to the axis.
3. Any ray that strikes the center of the mirror reflects symmetrically about the mirror axis.
4. Any ray through the center of curvature, C, strikes the mirror normal to the mirror surface and thus returns on itself.

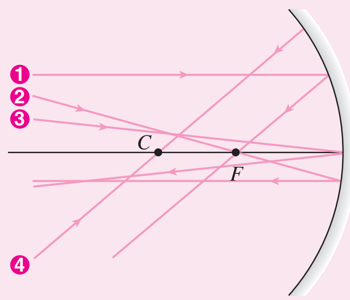


FIGURE 31.5 Four special rays for locating images in curved mirrors. Any two suffice to locate the image.

Figure 31.6 shows ray tracings, using our special rays 1 and 2 that go through the focal point, to find the image location in three cases. In each case symmetry ensures that the bottom of the image arrow is on the axis, so we haven't bothered to trace it. In

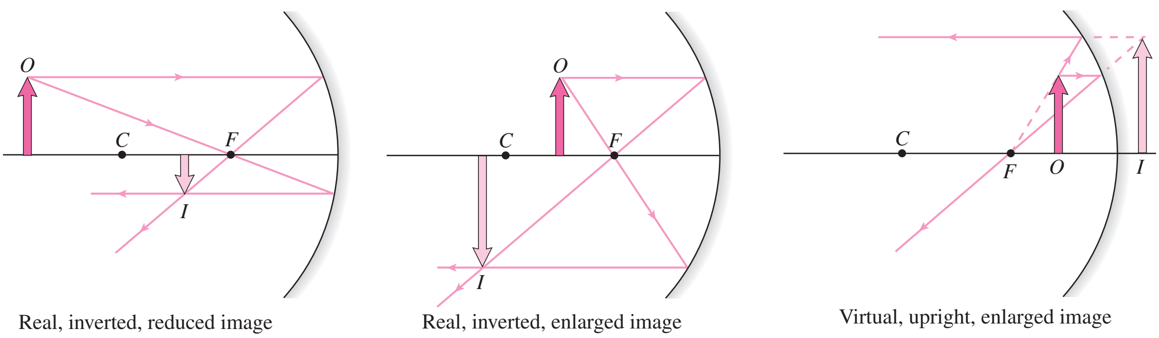


FIGURE 31.6 Image formation with a concave spherical mirror, using rays 1 and 2 described in Tactics 31.1. O denotes the object and I its image.

Fig. 31.6*a* we see that an object beyond the mirror's center of curvature, *C*, forms a smaller, inverted image. Light actually emerges from this image, so it's a *real* image. If you looked from the left toward the mirror in Fig. 31.6*a* you would actually see the image in space in front of the mirror (Fig. 31.7).

As the object moves closer to the mirror, the real image grows; with the object between the center of curvature and the focus, the image is larger than the object and farther from the mirror (Fig. 31.6*b*). As the object moves toward the focus, the image grows larger and moves rapidly away from the mirror. With the object right at the focus, the rays emerge in a parallel beam and there's no image. Finally, rays from an object closer to the mirror than the focus diverge after reflection. To an observer they appear to come from a point behind the mirror. Thus there is a virtual image, in this case upright and enlarged (Fig. 31.6*c*).

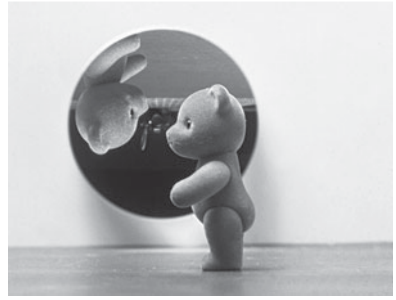


FIGURE 31.7 A bear meets its real image, formed by the concave mirror at the rear. Bear and image are both in *front* of the mirror.

GOT IT? 31.2 Where would you place an object so that its real image is the same size as the object, as in Fig. 31.7?

Convex Mirrors

A convex mirror reflects on the outside of its spherical curvature, causing light to diverge and therefore to form only virtual images (Fig. 31.8). Although the focus has less obvious physical significance in this case, its location still controls the geometry of reflected rays. As Fig. 31.8 shows, we can still draw a ray parallel to the axis and another ray that would go through the focus if the mirror weren't in its way. The reflected rays appear to diverge from a common point behind the mirror, showing a virtual image that's upright and reduced in size. By considering different object positions, you can convince yourself that the image in a convex mirror always has these characteristics. Convex mirrors are widely used where an image of a broad region needs to be captured in a small space (Fig. 31.9).

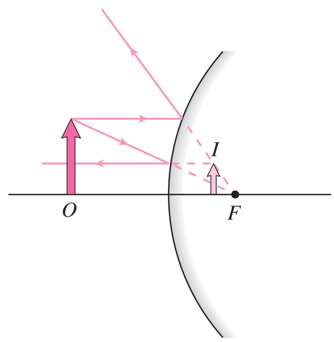


FIGURE 31.8 Image formation with a convex mirror. The image is always virtual, upright, and reduced in size.

FIGURE 31.9 A convex mirror gives a wide-angle view.

The Mirror Equation

Drawing ray diagrams gives an intuitive feel for image formation. More precise image locations and sizes follow from the **mirror equation**, which we now derive. This time we'll find the image using our special rays parallel to the axis and striking the mirror's center, as shown in Fig. 31.10*a*. The ray that strikes the center of the mirror reflects symmetrically about the axis; therefore, the two shaded triangles are similar. Then the **magnification** *M*—the ratio of image height h' to object height h—is the same as the ratio of image and object distances from the mirror. We'll consider the image height negative if the image is inverted; then from Fig. 31.10*a* we have

$$M = \frac{h'}{h} = -\frac{s'}{s} \qquad \text{(magnification)} \tag{31.1}$$

Here object and image are both in front of the mirror, so we take object and image distances s and s' as positive quantities; the negative sign in Equation 31.1 then shows that in

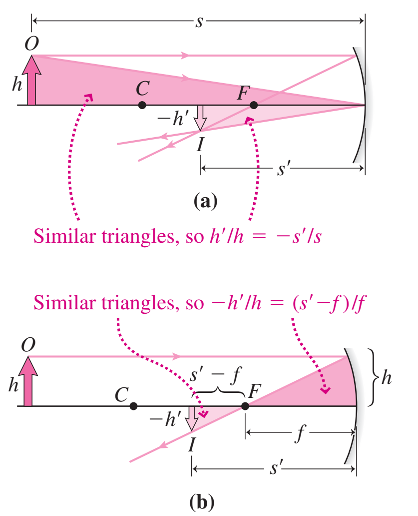


FIGURE 31.10 Finding the image *I* using rays 1 and 3 of Tactics 31.1. For an inverted image the height h' is negative, so we've marked the arrow length—a positive quantity—as $-h'$.

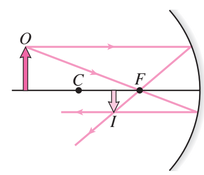
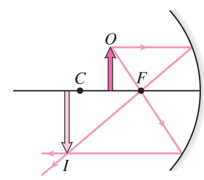
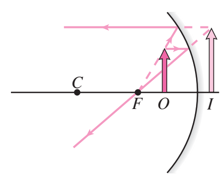
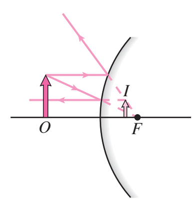
this case the image is inverted. Also, for the object location of Fig. 31.10*a*, it's clear that $|M| < 1$, meaning the image is reduced rather than enlarged.

Figure 31.10*b* is the same as Fig. 31.10*a* except that now we show only reflects through the focus. We've also labeled the focal length f and shaded a different pair of similar triangles. From these you can see that $-h'/h = (s' - f)/f$. Here we put the minus sign on h' because we've defined h' as negative for the inverted image, but comparing similar triangles requires a ratio of positive quantities. Equation 31.1 shows that the ratio h'/h is the magnification $M = -s'/s$. So we have $s'/s = (s' - f)/f$ or, dividing both sides by s' and doing a little algebra,

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} \qquad \text{(mirror equation)} \tag{31.2}$$

Although we derived the mirror equation using a real image, the equation applies to virtual images with the convention that a *negative* image distance s' means the image is *behind* the mirror. And we can handle *convex* mirrors as well by taking the focal length to be a *negative* quantity. Table 31.1 summarizes image formation with mirrors, including these sign conventions.

Table 31.1 Image Formation with Mirrors: Sign Conventions

Focal Length, *f*	**Object Distance, *s***	**Image Distance, *s'***	**Type of Image**	**Ray Diagram**
+ (concave)	+ (in front of mirror) $s > 2f$	+ (in front of mirror) $s' < 2f$	Real, inverted, reduced	
+ (concave)	+ (in front of mirror) $2f > s > f$	+ (in front of mirror) $s' > 2f$	Real, inverted, enlarged	
+ (concave)	+ (in front of mirror) $s < f$	− (behind mirror)	Virtual, upright, enlarged	
− (convex)	+ (in front of mirror)	− (behind mirror)	Virtual, upright, reduced	

A diagram similar to Fig. 31.10 but using the ray through the center of curvature gives another useful fact about curved mirrors: The magnitude of the focal length is half the radius:

$$|f| = \frac{R}{2} \tag{31.3}$$

You can prove this in Problem 77.

EXAMPLE 31.1 A Concave Mirror: The Hubble Space Telescope

During assembly of the Hubble Space Telescope, a technician stood 3.85 m in front of the telescope's concave mirror (Fig. 31.11). Given the telescope's 5.52-m focal length, find (a) the location and (b) the magnification of the technician's image.

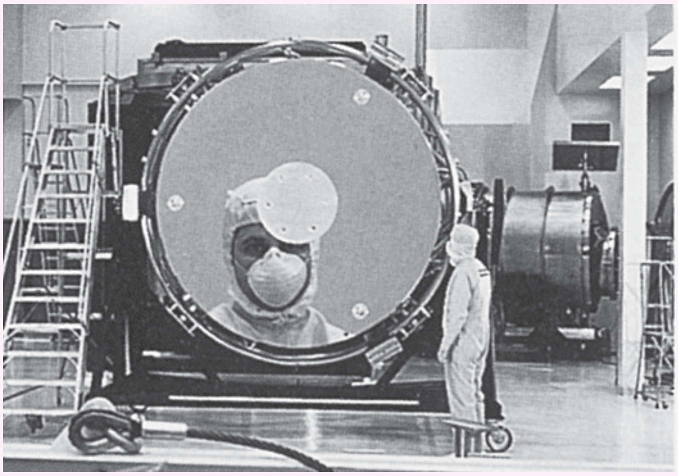


FIGURE 31.11 A technician standing in front of the Hubble Space Telescope mirror.

INTERPRET This problem is about image formation in a concave mirror. We identify the technician as the object, the 5.52-m focal length as f, and the 3.85-m distance as the object distance s.

DEVELOP We've sketched the situation in Fig. 31.12. With the object closer than the focal length, our sketch resembles Fig. 31.6c, so we anticipate an enlarged, virtual image, as shown. For (a), we'll solve the mirror equation 31.2 for the image distance s' to get the image location. Then for (b), we can find the magnification from Equation 31.1, $M = -s'/s$.

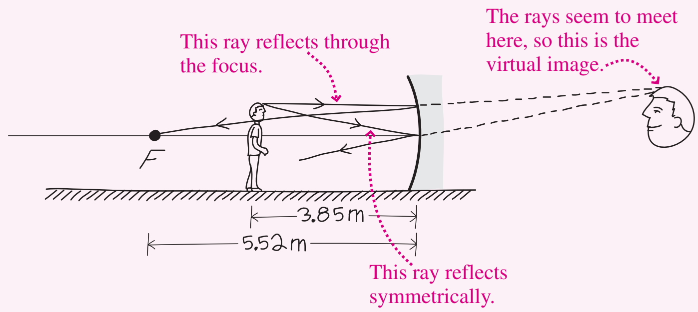


FIGURE 31.12 Sketch for Example 31.1, showing two rays that locate the virtual image of the technician's head.

EVALUATE (a) Solving Equation 31.2 for s' gives

$$s' = \frac{fs}{s-f} = \frac{(5.52\text{ m})(3.85\text{ m})}{3.85\text{ m} - 5.52\text{ m}} = -12.7\text{ m}$$

(b) Using this result in Equation 31.1 gives the magnification:

$$M = -\frac{s'}{s} = -\frac{-12.7\text{ m}}{3.85\text{ m}} = 3.30$$

ASSESS The *negative* image distance confirms what our sketch anticipated: This is a virtual image, located behind the mirror. The negative distance cancels the minus sign in Equation 31.1, giving a positive threefold magnification. Thus the image is upright and enlarged, just as Fig. 31.11 shows. ■

EXAMPLE 31.2 A Convex Mirror: Jurassic Park

In the film *Jurassic Park*, horrified passengers watch in a car's convex side-view mirror as a *Tyrannosaurus rex* pursues them. Printed on the mirror is the warning "OBJECTS IN MIRROR ARE CLOSER THAN THEY SEEM." If the mirror's curvature radius is 12 m and the *T. rex* is actually 9.0 m from the mirror, by what factor does the dinosaur appear reduced in size?

INTERPRET This is about a convex mirror, governed by the same basic equations as the concave mirror in Example 31.1. We identify the 12-m length as the curvature radius R and the 9.0-m distance as the object distance s.

DEVELOP We've sketched the situation in Fig. 31.13. Since this resembles Fig. 31.8, we see that the image will indeed be reduced in size. Equation 31.1, $M = -s'/s$, gives the magnification we want, but to use it we need the image distance s'. In Example 31.1 we solved Equation 31.2 to get $s' = fs/(s-f)$. Although that mirror was concave and this one is convex, Equation 31.2 applies to both mirrors. So we can use $s' = fs/(s-f)$ in Equation 31.1 to get the magnification:

$$M = -\frac{s'}{s} = -\frac{fs/(s-f)}{s} = -\frac{f}{s-f}$$

We aren't given the focal length, but Equation 31.3, $|f| = R/2$, shows that its magnitude is half the 12-m radius. Since this is a *convex* mirror, Table 31.1 shows that the focal length is *negative*, so $f = -6.0$ m.

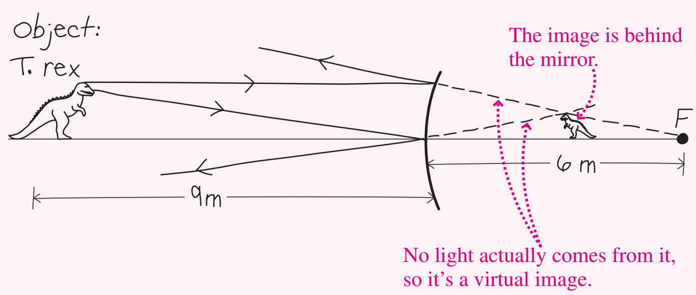


FIGURE 31.13 Sketch for Example 31.2, showing two rays that locate the image of *T. rex*. Mirror is not to scale.

EVALUATE Using our expression for M gives

$$M = -\frac{f}{s-f} = -\frac{(-6.0\text{ m})}{9.0\text{ m} - (-6.0\text{ m})} = 0.40$$

where the answer comes out positive because f is negative.

ASSESS Our result shows that *T. rex* appears in the mirror at only 40% of its actual size; in other words, it looks farther away than it really is. ■

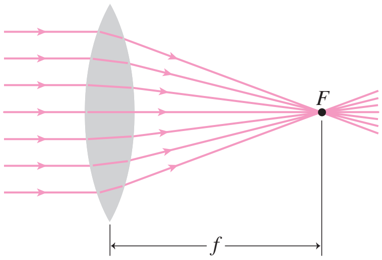


FIGURE 31.14 A convex lens brings parallel light rays to a focus at *F*.

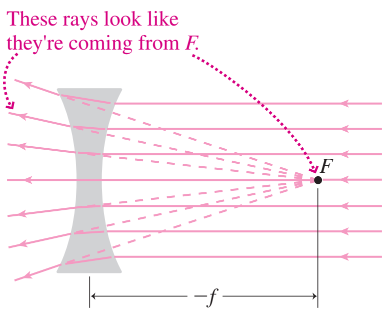


FIGURE 31.15 Parallel light passing through a concave lens diverges so it looks as though it's coming from a common focus.

31.2 Images with Lenses

A **lens** is a piece of transparent material that uses refraction to form images. Like mirrors, lenses can be either concave or convex. But light goes through a lens while it reflects off a mirror, so the roles of the two are reversed. A convex lens focuses parallel rays to a **focal point** and is therefore a **converging lens** (Fig. 31.14). As we'll see, a convex lens can form real and virtual images, depending on the object location. A concave lens, in contrast, is a **diverging lens**; it refracts parallel rays so they appear to diverge from a common focus. Like a *convex* mirror, a *concave* lens forms only virtual images (Fig. 31.15).

We first explore a **thin lens**—one whose thickness is small compared with the curvature radii of its two surfaces. Although light refracts as it enters a lens and again as it exits, in the thin-lens approximation the two surfaces are so close that it suffices to consider that the light bends just once, as it crosses the center plane of the lens. Unlike a mirror, light can go either way through a lens; that means the lens has two focal points, one on either side. For a thin lens, the focal length proves to be the same in either direction, so it doesn't matter which way we orient the lens. We'll consider all lenses to be thin unless otherwise stated, and later we'll justify the thin-lens approximation mathematically.

Lens Images by Ray Tracing

As with mirrors, two rays serve to locate the image formed by a lens. For lenses, two special rays simplify ray tracing.

TACTICS 31.2 Ray Tracing with Lenses

Figure 31.16 shows two special rays:
1. Any ray parallel to the lens axis refracts through the focal point.
2. Any ray through the center of the lens passes undeflected.

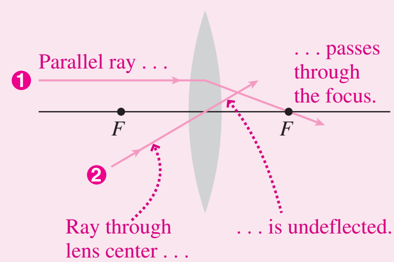


FIGURE 31.16 Two special rays for locating images formed with lenses.

Figure 31.17 shows ray tracings for different object placements in relation to a converging lens. In Fig. 31.17*a* we see that an object farther out than two focal lengths produces a smaller, inverted, real image on the other side of the lens. Since light really emanates from this image, you could see it without looking through the lens. As the object moves toward the lens, the image moves away and grows. When the object is between one and two focal lengths from the lens, the image has moved beyond $2f$ and is enlarged (Fig. 31.17*b*). The image on a movie screen is formed in this way. Moving the object closer than the focal point produces an enlarged, virtual image that can be seen only by an observer looking *through* the lens (Fig. 31.17*c*).

Figure 31.18 shows ray tracings for a diverging lens. Like a convex mirror, this lens produces only virtual images that are upright and reduced in size; they're visible only through the lens. The basic geometry of Fig. 31.18 doesn't change even if the object moves within the focal length.

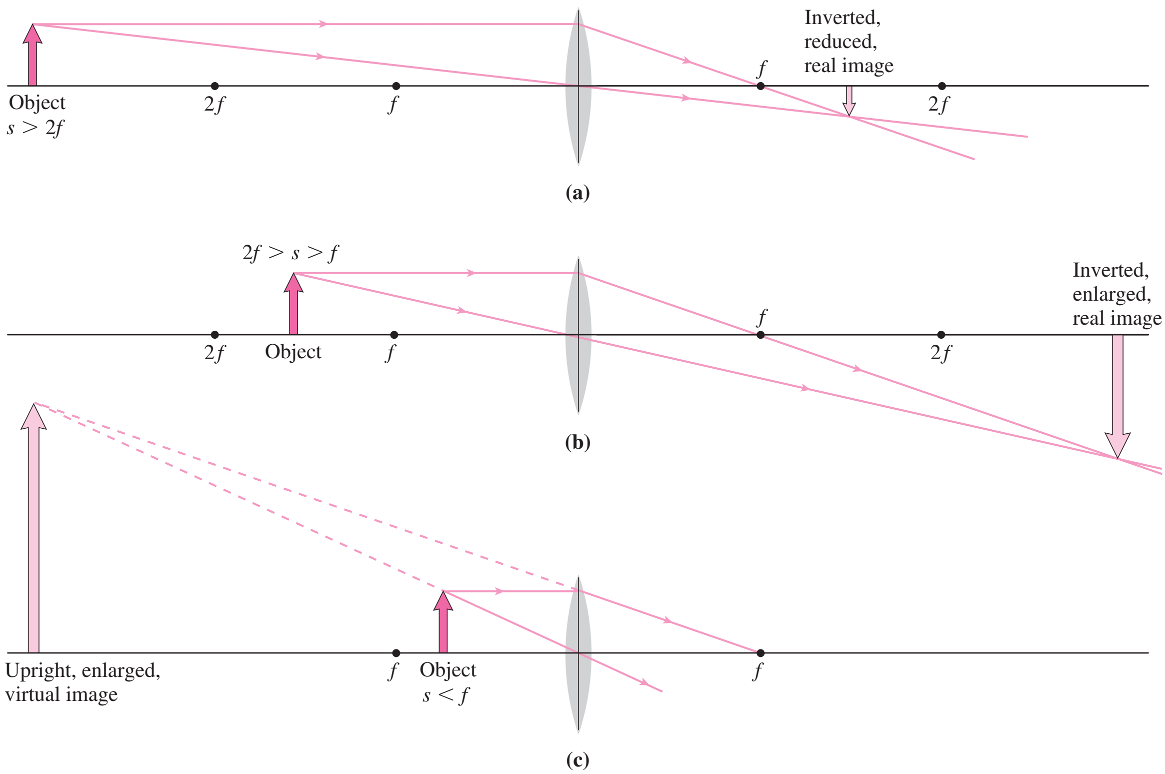


FIGURE 31.17 Image formation with a converging lens, shown for three object locations.

Getting Quantitative: The Lens Equation

Study Fig. 31.19 and you'll see that triangles OAB and IDB are similar. Therefore, as for mirrors, the image magnification is

$$M = \frac{h'}{h} = -\frac{s'}{s} \tag{31.4}$$

where again a negative height means an inverted image. The shaded triangles in Fig. 31.19 are also similar, so $-h'/(s' - f) = h/f$. Combining this result with Equation 31.4 and doing some algebra then give

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} \qquad \text{(lens equation)} \tag{31.5}$$

which is identical to the mirror equation 31.2. Although we derived Equation 31.5 for the case of a real image, it holds for virtual images if we consider the image distance negative; in that case the image is on the same side of the lens as is the object. And it holds for diverging lenses if we consider the focal length negative. Table 31.2 summarizes image formation with lenses, including these sign conventions. Figure 31.20 (next page) describes graphically the sizes and types of images formed at different object distances.

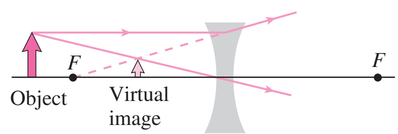


FIGURE 31.18 A diverging lens always forms a reduced, upright, virtual image, visible only through the lens.

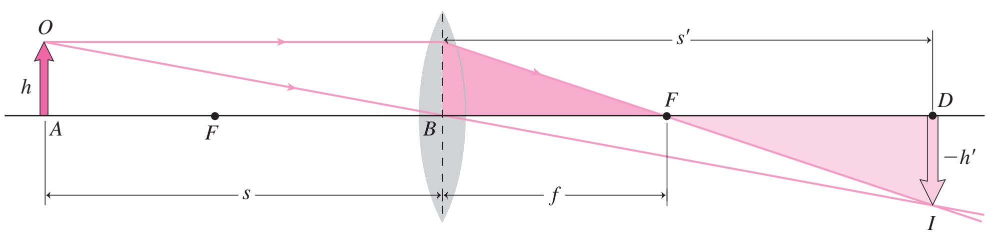


FIGURE 31.19 Ray diagram for deriving the lens equation. Triangles *OAB* and *IDB* are similar, as are the shaded triangles.

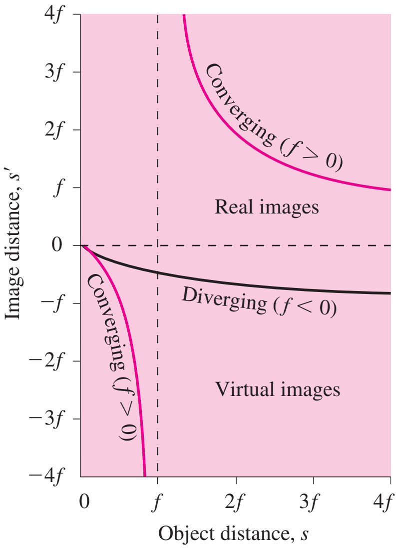
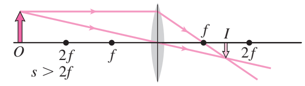
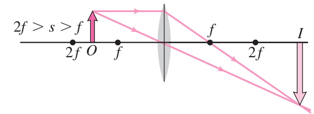
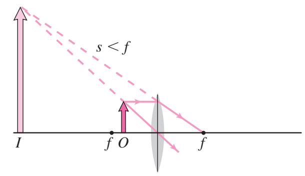
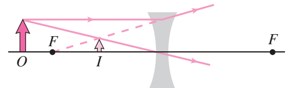


FIGURE 31.20 Image distance versus object distance for lenses.

Table 31.2 Image Formation with Lenses: Sign Conventions

Focal Length, f	Object Distance, s	Image Distance, s'	Type of Image	Ray Diagram
+ (convex)	+ $s > 2f$	+ (opposite side of lens) $2f > s' > f$	Real, inverted, reduced	
+ (convex)	+ $2f > s > f$	+ (opposite side of lens) $s' > 2f$	Real, inverted, enlarged	
+ (convex)	+ $s < f$	− (same side of lens)	Virtual, upright, enlarged	
− (concave)	+	− (same side of lens)	Virtual, upright, reduced	

GOT IT? 31.3 You look through a lens at this page and see the words enlarged and right-side up. Is the image you observe real or virtual? Is the lens concave or convex?

EXAMPLE 31.3 Using the Lens Equation: Fine Print

You're using a magnifying glass (a converging lens) with 30-cm focal length to read a telephone book (Fig. 31.21). How far from the page should you hold the lens in order to see the print enlarged three times?

FIGURE 31.21 Using a converging lens as a magnifying glass (Example 31.3).

INTERPRET This is a problem involving image formation with a converging lens. The object is the phone book, so we identify the book-to-lens distance we're asked for as the object distance s. Since this is a converging (i.e., convex) lens, the focal length is positive: $f = +30$ cm. The factor-of-3 magnification we want is the quantity M.

DEVELOP The situation is like Fig. 31.17*c*, with an enlarged, upright, virtual image. We're given the focal length and magnification, but we don't know either the object distance s, which we're looking for, or the image distance s'. Equation 31.4, $M = -s'/s$, relates the two, so we can first use that equation to eliminate s' in terms of s and then solve the lens equation 31.5, $1/s + 1/s' = 1/f$, for s.

EVALUATE With $M = 3$, Equation 31.4 gives $s' = -3s$. Then Equation 31.5 becomes

$$\frac{1}{s} - \frac{1}{3s} = \frac{2}{3s} = \frac{1}{f} = \frac{1}{30\text{ cm}}$$

so $s = (2)(30\text{ cm})/3 = 20$ cm.

ASSESS Our answer is less than the focal length, as Fig. 31.17*c* shows is required for a virtual image. Figure 31.21 confirms that the image is enlarged, upright, and virtual, and it appears farther away than the object. It's also on the same side of the lens as the object, which explains the negative image distance $s' = -60$ cm. ■

31.3 Refraction in Lenses: The Details

So far we've treated lenses as being arbitrarily thin and neglected details of the refraction process. Here we develop a more general description of refraction in lenses, of which our thin-lens approximation is a special case.

Refraction at a Curved Surface

Figure 31.22 shows a transparent material with refractive index n_2 and a curved surface of radius R. Outside the material is a medium with refractive index n_1. We'll now prove what

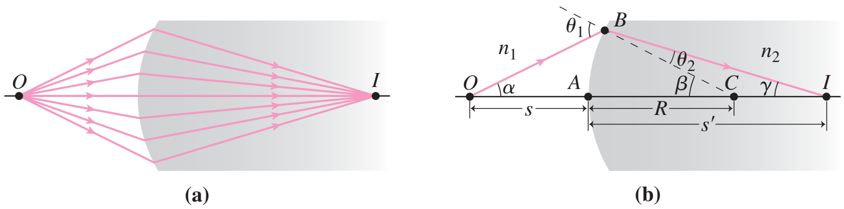


FIGURE 31.22 Refraction at an interface with a curved surface. All labeled angles are considered small, even though the drawing doesn't show them as such.

Fig. 31.22*a* shows: that rays from a point object O are refracted to a common image point I. Our proof is valid only in the paraxial approximation that all rays make small angles with the optic axis; as with mirrors, our drawings won't always show these angles as being small.

Figure 31.22*b* shows a single ray. With all the labeled angles small, we can approximate $\sin x \simeq \tan x \simeq x$, with x in radians. Then Snell's law, $n_1 \sin\theta_1 = n_2 \sin\theta_2$, becomes $n_1\theta_1 = n_2\theta_2$. Triangles BCI and OBC give $\theta_2 = \beta - \gamma$ and $\theta_1 = \alpha + \beta$, so Snell's law becomes $n_1(\alpha + \beta) = n_2(\beta - \gamma)$. Furthermore, in the small-angle approximation the arc BA is so close to a straight line that we can write $\alpha \simeq \tan\alpha \simeq BA/s$, with $s = OA$ the object's distance from the refracting surface. Similarly, $\beta \simeq BA/R$ and $\gamma \simeq BA/s'$. Thus our expression of Snell's law becomes

$$n_1\left(\frac{BA}{s} + \frac{BA}{R}\right) = n_2\left(\frac{BA}{R} - \frac{BA}{s'}\right)$$

or, on canceling BA and rearranging,

$$\frac{n_1}{s} + \frac{n_2}{s'} = \frac{n_2 - n_1}{R} \tag{31.6}$$

The angle α doesn't appear here, showing that this relation between object and image distances holds for *all* rays that satisfy the small-angle approximation. So Fig. 31.22*a* is correct: All such rays do indeed come to a common focus at I.

We derived Equation 31.6 for the case of a real image, but as usual it applies to virtual images if we take the image distance as negative. And it applies to concave surfaces if we take R to be negative. It even works for flat surfaces, with $R = \infty$.

EXAMPLE 31.4 Refraction at a Curved Surface: A Cylindrical Aquarium

An aquarium consists of a thin-walled plastic tube 70 cm in diameter. For a cat looking directly into the aquarium, what's the apparent distance to a fish 15 cm from the aquarium wall?

INTERPRET We interpret the cylindrical aquarium as a two-dimensional version of the spherical surface we analyzed with Fig. 31.22 and Equation 31.6. The plastic tube is thin, so we can neglect refraction within the plastic and consider that we have just a cylinder of water. The object is the fish, and since it's *inside* the water, the cylindrical edge of the aquarium is concave *toward* the object. Then the curvature radius is *negative*; we're given the diameter as 70 cm, so $R = -35$ cm. With the object in the water, $n_1 = 1.333$ from Table 30.1 and $n_2 = 1$ for air. The 15-cm distance from the edge to the fish is the object distance s.

DEVELOP Figure 31.23 shows the physical situation and a ray diagram viewed from above. Our plan is to solve Equation 31.6, $n_1/s + n_2/s' = (n_2 - n_1)/R$, for s' and evaluate using $R = -35$ cm, $n_1 = 1.333$, $n_2 = 1$, and $s = 15$ cm.

EVALUATE Solving, we have

$$s' = n_2\left(\frac{n_2 - n_1}{R} - \frac{n_1}{s}\right)^{-1} = -12.6 \text{ cm}$$

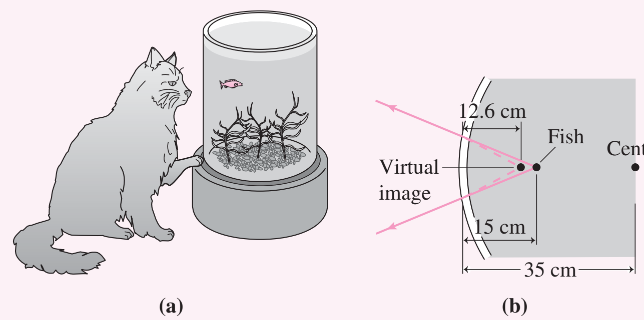


FIGURE 31.23 (a) A cylindrical aquarium. (b) Top view, showing the formation of a virtual image of a fish that's actually 15 cm from the edge.

ASSESS Make sense? The fish is actually 15 cm from the edge, but refraction makes the image distance s' shorter. The same effect occurs when you look down into a swimming pool or lake: Objects on the bottom look closer, and you can find out how much by applying Equation 31.6 with $R = \infty$ (see Exercise 30).

Lenses, Thick and Thin

Figure 31.24 shows a lens of thickness t with refractive index n, surrounded by air with $n = 1$. Object O_1 lies a distance s_1 from the left-hand surface. This surface forms an image I_1, which we'll also call O_2 because it acts as an object for the right-hand surface. Refraction at that surface forms a second image I_2. We want to relate the original image O_1 and the final image I_2.

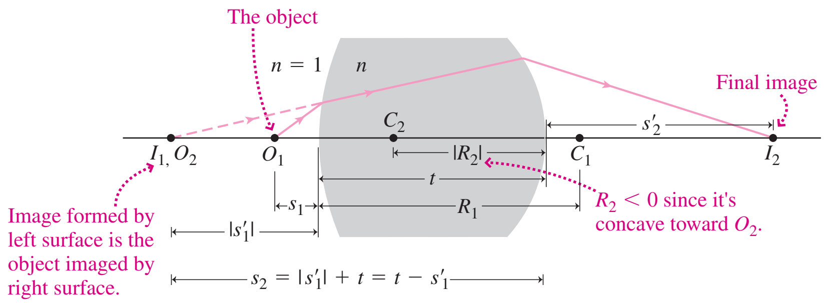


FIGURE 31.24 Analysis of a thick lens with different curvature radii. C_1 and C_2 are the centers of curvature of the left and right sides, respectively; t is the lens thickness.

At the left-hand surface, the quantities in Equation 31.6 become

$$\frac{1}{s_1} + \frac{n}{s_1'} = \frac{n-1}{R_1} \qquad \text{(left-hand surface)}$$

We've placed O_1 so close that I_1 is a *virtual* image, so s_1' is *negative*. Now we set up another instance of Equation 31.6, this time for the right-hand surface. Here I_1 is the object, whose distance s_2 is $t - s_1'$ because s_1' is negative. Also at the right-hand surface, $s' = s_2'$, $n_1 = n$, and $R = R_2$, where, for the case shown, R_2 is negative because the right-hand surface is concave toward the object (I_1, O_2) that's being imaged. So at the right-hand surface, Equation 31.6 reads

$$\frac{n}{t - s_1'} + \frac{1}{s_2'} = \frac{1-n}{R_2} \qquad \text{(right-hand surface)}$$

Now we'll let the lens become arbitrarily thin, so $t \to 0$. Then we add the two equations; the intermediate-image term n/s_1' cancels, leaving only the first object distance and final image distance. So we drop the subscripts 1 and 2, and the result is

$$\frac{1}{s} + \frac{1}{s'} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$$

The left-hand side here is identical to the left-hand side of Equation 31.5, and Equation 31.5's right-hand side is $1/f$. Equating the two right-hand sides results in the **lensmaker's formula**, which gives the focal length:

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \qquad \text{(lensmaker's formula)} \qquad (31.7)$$

Again, the radii here can be positive or negative; in Fig. 31.24 R_1 is positive because the left-hand surface is convex toward the object, while R_2 is negative because the right-hand surface is concave toward its object, the intermediate image I_1. Although we derived Equation 31.7 for the case of a virtual intermediate image, the lensmaker's formula is a general result for the focal length of a thin lens.

Lenses come in a variety of shapes (Fig. 31.25). Those that are thicker in the center are converging lenses, for which Equation 31.7 gives a positive focal length. Those that are thinner in the center are diverging lenses, with negative f. These behaviors reverse if the medium surrounding the lens has a higher refractive index (see Problem 76).

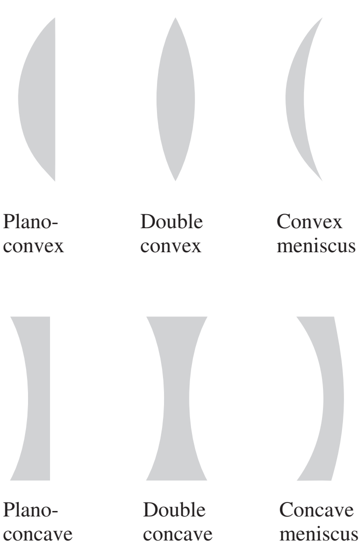


FIGURE 31.25 Common lens types.

EXAMPLE 31.5 The Lensmaker's Formula: A Plano-Convex Lens

Find an expression for the focal length of the plano-convex lens in Fig. 31.25, given refractive index n and radius R for the one curved surface.

INTERPRET This is a thin lens, such as we just analyzed in deriving the lensmaker's formula. With an object at the left of the lens, we identify $R_1 = R$ and $R_2 = \infty$ for the flat surface.

DEVELOP The lensmaker's formula, Equation 31.7, relates the focal length, the lens surface radii, and the refractive index. Our plan is to solve for f.

EVALUATE With $R_1 = R$ and $R_2 = \infty$, Equation 31.7 gives

$$f = \left[(n-1)\left(\frac{1}{R} - \frac{1}{\infty}\right)\right]^{-1} = \frac{R}{n-1}$$

ASSESS Make sense? The smaller R, the more curved the lens and the more it bends light; the result is a shorter focal length. The higher n, the greater the refraction, and with n in the denominator, the result is again a shorter focal length. We asserted earlier that a thin lens works the same either way. You can see that explicitly here by putting the object beyond the flat side; then $R_1 = \infty$ and $R_2 = -R$, but the result for f is unchanged. ■

Lens Aberrations

Lenses exhibit several optical defects. We described **spherical aberration** in mirrors; this same defect occurs with spherical lenses (Fig. 31.26*a*). Our lens analysis required that all rays make small angles with the lens axis; if not, then they don't share a common focus, causing spherical aberration. Small angles occur naturally with distant objects, but not with objects close to the lens. Using only the central portion of the lens can eliminate those rays with larger angles (Fig. 31.26*b*), leading to sharper focus. That's why a camera focuses over a wider range when it's "stopped down," with an opaque iris covering the outer part of the lens. The trade-off is that less light is available.

We mentioned **chromatic aberration** in Chapter 30; it occurs because the refractive index varies with wavelength, causing different colors to focus at different points. High-quality optical systems minimize this effect by using composite lenses of materials with different refractive indices. **Astigmatism** occurs when a lens has different curvature radii in different directions. This is a common defect in the human eye, corrected with glasses or contact lenses that have compensating asymmetric curvature.

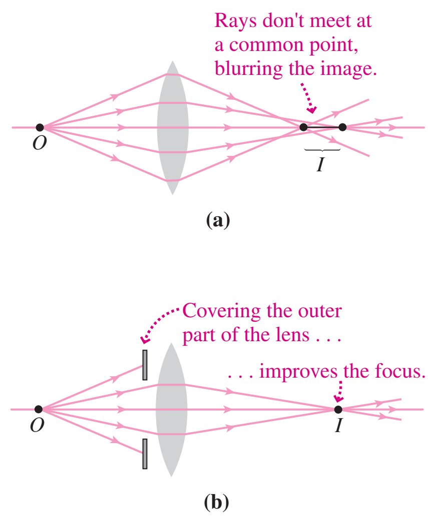


FIGURE 31.26 (a) Spherical aberration. (b) Using only the central portion of the lens minimizes aberration, but at the expense of a dimmer image.

31.4 Optical Instruments

Numerous optical instruments use lenses, mirrors, or both to form images. All but the simplest have more than one optical element, but the principles we've developed here still apply. We analyze such instruments by tracing light through the sequence of optical elements, using the image formed by one element as the object for the next.

The Eye

Our eyes are complex optical systems with several refracting surfaces and mechanisms to vary the focal length and amount of light admitted (Fig. 31.27). Light enters through the cornea and traverses a liquid called the aqueous humor before passing through the lens. It then traverses the vitreous humor, a liquid in the main body of the roughly 2.3-cm-diameter eyeball. Finally it strikes the retina, where special cells called rods and cones produce electrochemical signals that carry visual information to the brain.

A properly functioning eye produces well-focused real images on the retina, with the cornea providing most of the refractive focusing. Muscles adjust the lens, changing its focal length to compensate for different object distances. Other muscles adjust the iris, resizing the pupil opening to adjust for different light levels.

In nearsighted (myopic) people, the image forms in front of the retina, causing distant objects to appear blurred (Fig. 31.28*a*, next page). Diverging corrective lenses produce closer intermediate images that the myopic eye can then focus (Fig. 31.28*b*, next page). In

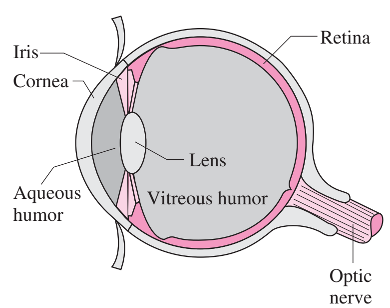


FIGURE 31.27 The human eye.

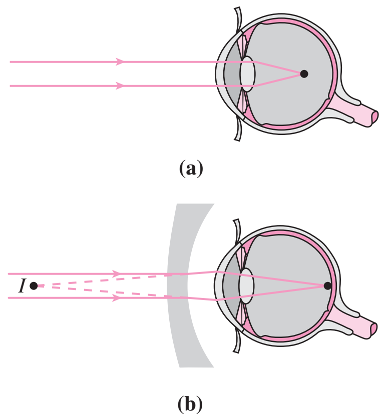


FIGURE 31.28 (a) A nearsighted eye focuses light from distant objects in front of the retina. (b) A diverging lens corrects the problem, creating a closer virtual image that the eye can focus.

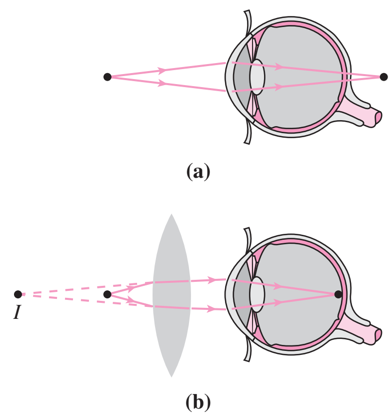


FIGURE 31.29 (a) A farsighted eye can't focus light from nearby objects. (b) A converging lens produces a more distant image that the eye can focus.

farsighted (hyperopic) people, the image of nearby objects would form behind the retina, and converging corrective lenses are used (Fig. 31.29). Even normal eyes can't focus much closer than the so-called **near point** at about 25 cm. This distance greatly increases with age, a condition called "presbyopia."

Prescriptions for corrective lenses specify the corrective power, P, in **diopters**, which is the inverse of the focal length in meters. Thus a 1-diopter lens has $f = 1$ m, while a 2-diopter lens has $f = 0.5$ m and is more powerful in that it refracts light more sharply. Like f itself, the sign of a lens's corrective power is positive or negative depending on whether the lens is converging or diverging.

It doesn't matter whether a corrective lens is several centimeters from the eye, as with glasses, or right on the cornea, as with contact lenses. Contact lenses can be thin because, as Equation 31.7 shows, it's the curvature radii and not the thickness that determine the focal length. A more radical approach to vision correction is laser surgery, described in the Application.

APPLICATION Laser Vision Correction

The cornea provides most of the eye's refractive power, with the adjustable lens compensating for different object distances. The popular LASIK procedure corrects vision by reshaping the cornea. In LASIK, the surgeon begins by mechanically cutting a flap of the outermost corneal layer. Then a precision laser beam breaks molecular bonds in the corneal tissue, vaporizing the material and reshaping the cornea according to a prescription customized for the individual eye. With a nearsighted eye, the laser thins the central cornea, making it less sharply curved and thus reducing its refractive power. This has the same effect as the corrective lens in Fig. 31.28*b*. It's harder, but still possible, to correct farsightedness with LASIK. This involves thinning a ring-shaped region around the central cornea, making the cornea more steeply curved and thus increasing its refractive power. The corrective lens in Fig. 31.29*b* accomplishes the same thing. The corneal reshaping doesn't have to be symmetric, so LASIK can also correct an asymmetric cornea that causes astigmatism. What it can't do is restore the ability to focus both near and far, since that's handled by the lens, which stiffens with age. So older LASIK patients still need reading glasses. It's possible to correct one eye for near vision and another for distance, but then the patient loses some of the depth perception that comes with binocular vision.

The laser used in vision correction is a precisely controllable *excimer laser*, which produces intense bursts of ultraviolet light. Each pulse removes only 0.25 μm of tissue—one four-thousandth of a millimeter. The laser is so precise that it can cut notches in a human hair! The surgeon determines the necessary corneal adjustments and feeds the information to a computer that controls the laser. Thanks to the laser's precision, most patients achieve nearly complete vision correction.

CONCEPTUAL EXAMPLE 31.1 Contact Lens Mix-Up

You and your roommate have gotten your boxes of disposable contact lenses mixed up. One box is marked "−1.75 diopter," the other "+2.5 diopter." You're farsighted and your roommate is nearsighted. Which lenses are yours?

EVALUATE Figure 31.29 shows that you need a converging lens to correct your farsightedness. From Table 31.2's sign conventions, that means a positive focal length and therefore a positive corrective power $P = 1/f$. So yours are the +2.5-diopter lenses.

ASSESS Your lenses don't actually look like the lens in Fig. 31.29*b*. Since your cornea is curved, they're more like the convex meniscus lens of Fig. 31.25. The important point is that they're thicker in the middle, which makes them converging lenses.

MAKING THE CONNECTION What's the focal length of your contact lenses?

EVALUATE The diopter measure is the inverse of the focal length in meters, so, conversely, $f = 1/P = (1/2.5)\text{ m} = 40\text{ cm}$.

EXAMPLE 31.6 The Power of Lenses: Lost Your Glasses!

You're on vacation and have lost your reading glasses; without them, your eyes can't focus closer than 70 cm. Fortunately, you can buy nonprescription reading glasses at the pharmacy, where they come in 0.25-diopter increments. Which glasses should you buy so you can focus at the standard 25-cm near point?

INTERPRET Your eyes can't focus closer than 70 cm, so this problem is asking for the power of a lens that will make an object at 25 cm appear as if it's at 70 cm. In other words, when the object distance s is 25 cm, the image distance s' should be −70 cm. We put the minus sign here because, as Fig. 31.29*b* shows, the image is on the same side of the lens as the object, so this is a *virtual* image.

DEVELOP Equation 31.5, $1/s + 1/s' = 1/f$, relates the inverse of the focal length to the object and image distances. But with the focal length in meters, $1/f$ is just the power, P, in diopters. So we can get the required power directly from Equation 31.5.

EVALUATE Applying Equation 31.5 gives

$$P = \frac{1}{f} = \frac{1}{s} + \frac{1}{s'} = \frac{1}{0.25\text{ m}} + \frac{1}{-0.70\text{ m}} = 2.57\text{ diopters}$$

ASSESS The closest available power is 2.5 diopters, so that's what you should buy. ■

Cameras

A camera is much like the eye, except that an electronic detector or film replaces the light-sensitive retina. Where the eye changes the lens shape to accommodate different object distances, a camera moves its rigid lens to change the image distance. Simple "point and shoot" cameras use infrared beams to determine the object distance, and then automatically adjust the lens position for optimum focus. The camera also adjusts the lens aperture and exposure time for ambient light conditions. Zoom lenses have moveable elements that alter the focal length for wide-angle to telephoto views.

Magnifiers and Microscopes

Examining very small objects requires bringing them closer than the 25-cm near point below which the human eye can't focus. We therefore use lenses to put enlarged images at greater distances, where we can focus. What matters is not the actual image size, but how much bigger an object looks to us—and that depends on how much of our field of view it occupies. **Angular magnification**, m, is the ratio of the angle an object subtends when seen through a lens to the angle subtended when it's at the 25-cm near point and viewed with the naked eye. Figure 31.30*a* shows that the latter angle, measured in radians, is approximately $\alpha = h/25\text{ cm}$, where h is the object height. We get the most comfortable viewing with the eye close to the lens and the object just inside the focal length, forming a large and distant virtual image. With this geometry, Fig. 31.30*b* shows that the image angle is very nearly h/f. Then the angular magnification is

$$m = \frac{\beta}{\alpha} = \frac{h/f}{h/25\text{ cm}} = \frac{25\text{ cm}}{f} \quad \text{(simple magnifier)} \tag{31.8}$$

Single lenses produce angular magnifications up to about 4 before aberrations compromise image quality. Greater magnification requires more than one lens. In a

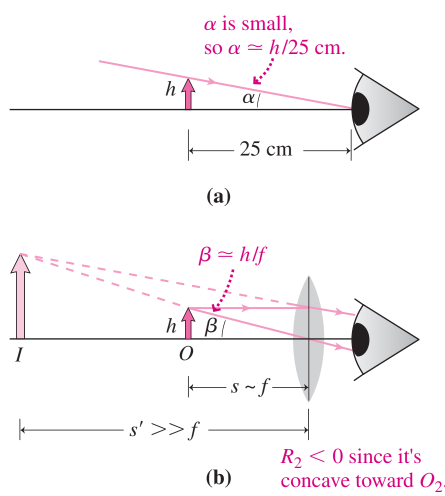


FIGURE 31.30 Calculating the angular magnification $m = \beta/\alpha$.

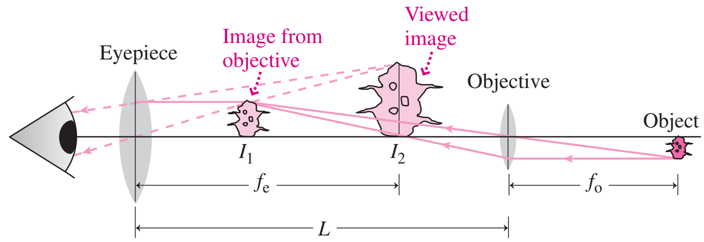


FIGURE 31.31 Image formation in a compound microscope. Figure is not to scale; L should be much greater than either focal length, and image I_1 should be very near the eyepiece's focus, resulting in greater magnification.

compound microscope, an **objective lens** of short focal length forms a magnified real image. This image is viewed through a second lens, the **eyepiece**, used as a simple magnifier (Fig. 31.31). The object being viewed is positioned just beyond the focus of the objective lens, and its image falls just inside the focal length of the eyepiece. If both focal lengths are small compared with the distance between the lenses, then the object distance for the objective lens is approximately the objective focal length f_o, and the resulting image distance is approximately the lens spacing L. The real image formed by the objective lens is larger than the object by the ratio of the image and object distances, or $M_o = -L/f_o$. The eyepiece makes the real image look larger still, by a factor of its angular magnification $m = 25\text{ cm}/f_e$. So the overall magnification of the microscope is

$$M = M_o m_e = -\frac{L}{f_o}\left(\frac{25\text{ cm}}{f_e}\right) \qquad \text{(compound microscope)} \qquad (31.9)$$

where, as usual, the minus sign signifies an inverted image.

Optical microscopes work well as long as the approximation of geometrical optics holds—that is, when the object is much larger than the wavelength of light. Viewing smaller objects requires shorter wavelengths than those of visible light. In the electron microscope, those "waves" are electrons, whose wavelike nature we'll examine in Chapter 34.

Telescopes

A telescope collects light from distant objects, either forming an image or supplying light to instruments for analysis. Modern astronomical instruments are invariably **reflectors**, whose main light-gathering element is a mirror. Small handheld telescopes, binoculars, and telephoto lenses are **refractors**, which use lenses to gather light.

A simple refractor consists of an objective lens that images distant objects at essentially its focal point, followed by an eyepiece to view this image (Fig. 31.32). The focal points of objective and eyepiece are nearly coincident, so the real image at the objective's focus is then seen through the eyepiece as a greatly enlarged, virtual image. The angular

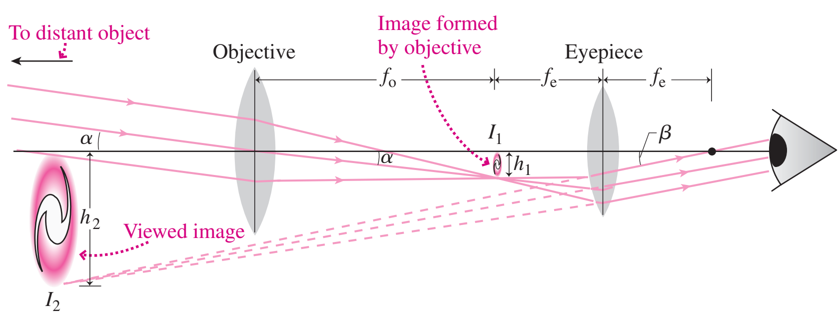


FIGURE 31.32 Image formation in a refracting telescope. A distant object is imaged first at the focus of the objective lens (image I_1). An eyepiece with its focus at nearly the same point then gives an enlarged virtual image (I_2). The angles α and β are given by $\alpha \simeq h_1/f_o$ and $\beta \simeq h_1/f_e$, leading to Equation 31.10.

magnification is the ratio of the angle β subtended by the final image to the angle α subtended by the actual object; Fig. 31.32 shows that this ratio is

$$m = \frac{\beta}{\alpha} = \frac{f_o}{f_e} \qquad (\text{refracting telescope}) \tag{31.10}$$

Since a real image is inverted and a virtual image is upright, a two-lens refracting telescope gives an inverted image. This is fine for astronomical work, but telescopes designed for terrestrial use have an extra lens, a diverging eyepiece, or a set of reflecting prisms to produce an upright image.

Reflecting telescopes offer many advantages over refractors. Mirrors have reflective coatings on their front surfaces, eliminating chromatic aberration because light doesn't pass through glass. Reflectors can be much larger since mirrors are supported across their entire back surfaces—unlike lenses, which must be supported at the edges. Whereas the largest refracting telescope ever built has a 1-m-diameter lens, large reflectors boast diameters of 10 m or greater. These designs incorporate segmented and/or flexible mirrors whose shape can be adjusted under computer control for optimum focusing. With so-called adaptive optics, such systems may adjust rapidly enough to compensate for the atmospheric turbulence that has traditionally limited the resolution of ground-based telescopes.

The simplest reflecting telescope is a curved mirror with a detector at its focus. Superb image quality results, in principle limited only by wave effects we'll discuss in the next chapter. More often the telescope is used as a "light bucket," collecting light from distant sources too small to image even with today's large optical telescopes. Then a secondary mirror sends light to a focus at a point that's convenient for telescope-mounted instrumentation. Optical fibers may also be used to bring light collected by the primary mirror to fixed instruments. Figure 31.33 shows three common designs for reflecting telescopes.

Magnification is not a particularly important quantity in astronomical telescopes, which are used more for spectral and other analysis than for direct imaging. More important is the light-gathering power of the instrument, which is determined simply by the area of its objective lens or primary mirror. Each of the two 10-m Keck Telescopes, for instance, has 100 times the light-gathering power of the 1-m Yerkes refractor and more than 17 times the power of the 2.4-m Hubble Space Telescope. The Giant Magellan Telescope, scheduled for operation in 2016, boasts seven mirrors equivalent to a single 21-m mirror, with more than four times the light-gathering power of the 10-m Keck instruments (Fig. 31.34).

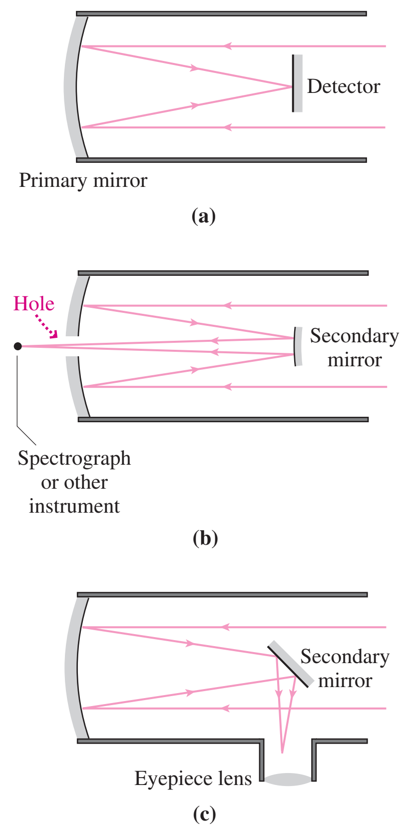


FIGURE 31.33 Reflecting telescopes. (a) A detector at the prime focus gives the best image quality. (b) The Cassegrain design is widely used in large telescopes. (c) The Newtonian design is used primarily in small telescopes.

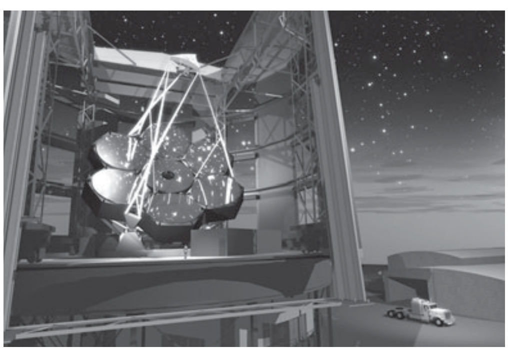


FIGURE 31.34 A painting of the Giant Magellan Telescope as it will look when completed in 2016.

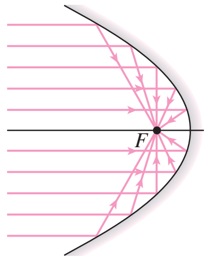
CHAPTER 31 SUMMARY

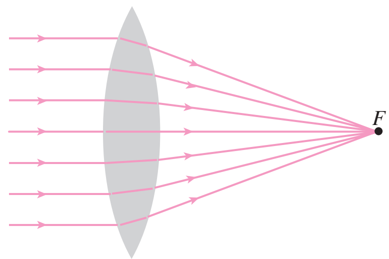
Big Picture

The big idea here is how reflection and refraction form **images**. Images are **real** or **virtual** depending on whether or not light actually comes from the image location.

Key Concepts and Equations

A curved mirror or lens has a **focal point**, F, at which parallel light rays converge:





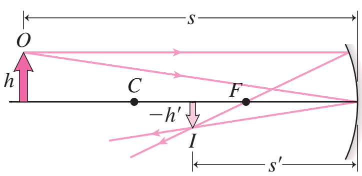
The same equation describes image formation with mirrors and lenses:

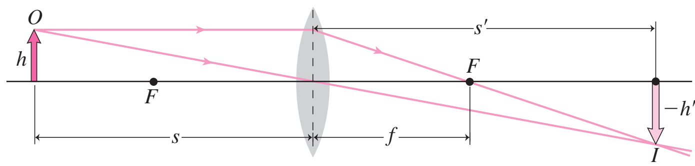
$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f}$$

The table summarizes the sign conventions for each term.

Value	Symbol	Condition	Sign
Object distance	s	Object on same side as *incoming* light rays	+
		Object on opposite side from *incoming* light rays	−
Image distance	s'	Image on same side as *outgoing* light rays	+
		Image on opposite side from *outgoing* light rays	−
Focal length	f	Focus on same side as *outgoing* light rays	+
		Focus on opposite side from *outgoing* light rays	−

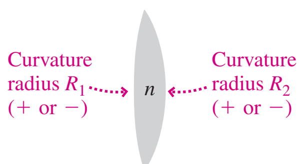
Image formation with mirrors and lenses:





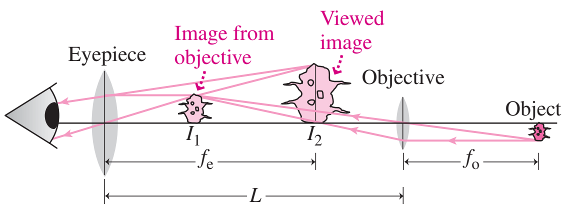
The **lensmaker's formula** is

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$$



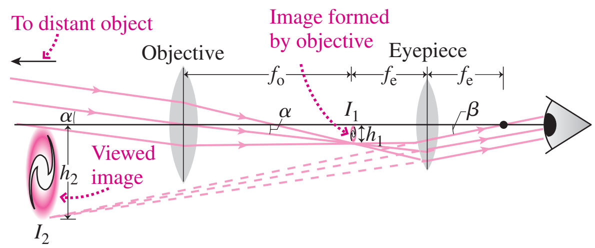
Applications

Compound microscope



Magnification: $M = -\frac{L}{f_o}\left(\frac{25\text{ cm}}{f_e}\right)$

Refracting telescope



Angular magnification: $m = \frac{f_o}{f_e}$

For homework assigned on MasteringPhysics, go to www.masteringphysics.com | BIO Biology and/or medical-related problems C Computer problems

For Thought and Discussion

1. How can you see a virtual image, when it's not "really there"?
2. You lay a magnifying glass (a converging lens) on a printed page. Looking toward the lens, you move it toward you, away from the page, eventually going well beyond its focal length. Explain the changes in what you see.
3. Under what circumstances will the image in a concave mirror be the same size as the object?
4. If you're handed a converging lens, what can you do to estimate its focal length quickly?
5. What's the meaning of a negative object distance? A negative focal length?
6. A diverging lens always makes a reduced image. Could you use such a lens to start a fire by focusing sunlight? Explain.
7. Is there any limit to the temperature you can achieve by focusing sunlight? (*Hint:* Think about the second law of thermodynamics.)
8. Can a concave mirror make a reduced real image? A reduced virtual image? An enlarged real image? An enlarged virtual image? Specify conditions for each possible image.
9. If you placed a screen at the location of a virtual image, would the image appear on the screen? Why or why not?
10. If you look into the bowl of a metal spoon, you see yourself upside down. Flip the spoon and you're right-side up. Explain why.
11. Is the image on a movie screen real or virtual? How do you know?
12. Does a fish in a spherical bowl appear larger or smaller than it actually is?
13. A block of ice contains a hollow, air-filled space in the shape of a double-convex lens. Describe the optical behavior of this space.
14. The refractive index of the human cornea is about 1.4. If you can see clearly in air, why can't you see clearly underwater? Why do goggles help?
15. Do you want a long or short focal length for a telescope's objective lens? What about a microscope's?
16. Give at least three reasons why reflecting telescopes are superior to refractors.

Exercises and Problems

Exercises

Section 31.1 Images with Mirrors

17. A shoe store uses small floor-level mirrors to let customers view prospective purchases. At what angle should such a mirror be inclined so that a person standing 50 cm from the mirror with eyes 140 cm off the floor can see her feet?
18. A candle is on the axis of a 15-cm-focal-length concave mirror, 36 cm from the mirror. (a) Where is its image? (b) How do the image and object sizes compare? (c) Is the image real or virtual?
19. An object is five focal lengths from a concave mirror. (a) How do the object and image heights compare? (b) Is the image upright or inverted?
20. A virtual image is located 40 cm behind a concave mirror with focal length 18 cm. (a) Where is the object? (b) By how much is the image magnified?
21. (a) Where on the axis of a concave mirror would you place an object in order to produce a full-size image? (b) Will the image be real or virtual?

Section 31.2 Images with Lenses

22. A lightbulb is 56 cm from a convex lens. Its image appears on a screen 31 cm from the lens, on the other side. Find (a) the lens's focal length and (b) how much the image is enlarged or reduced.
23. By what factor is the image magnified for an object 1.5 focal lengths from a converging lens? Is the image upright or inverted?
24. A lens with 50-cm focal length produces a real image the same size as the object. How far from the lens are image and object?
25. By holding a magnifying glass 25 cm from your desk lamp, you can focus an image of the lamp's bulb on a wall 1.6 m from the lamp. What's the focal length of your magnifying glass?
26. A real image is four times as far from a lens as is the object. What's the object distance, measured in focal lengths?
27. A magnifying glass enlarges print by 50% when it's 9.0 cm from a page. What's its focal length?

Section 31.3 Refraction in Lenses: The Details

28. You're writing specifications for a new line of magnifying glasses that have double-convex lenses with equal 32-cm curvature radii, made from glass with $n = 1.52$. What do you list for the focal length?
29. You're standing in a wading pool and your feet appear to be 30 cm below the surface. How deep is the pool?
30. The bottom of a swimming pool looks to be 1.5 m below the surface. Find the pool's actual depth.
31. A tiny insect is trapped 1.0 mm from the center of a spherical dewdrop 4.0 mm in diameter. As you look straight into the drop, what's the insect's apparent distance from the drop's surface?
32. You're underwater, looking through a spherical air bubble (Fig. 31.35). What's its actual diameter if it appears, along your line of sight, to be 1.5 cm in diameter?

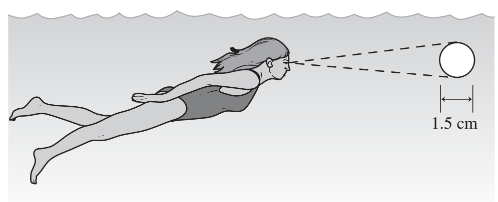


FIGURE 31.35 Exercise 32

Section 31.4 Optical Instruments

33. BIO You have to hold a book 55 cm from your eyes for the print to be in focus. What power lens would correct your farsightedness?
34. What focal length is needed for a simple magnifier with angular magnification 3?
35. BIO You're an optometrist helping a nearsighted patient who claims he can't see clearly beyond 80 cm. Prescribe a lens that will put the images of distant objects at 80 cm, giving your patient clear vision at all distances beyond the normal near point.
36. BIO A particular eye has a focal length of 2.0 cm instead of the 2.2 cm that would put a sharply focused image on the retina. (a) Is this eye nearsighted or farsighted? (b) What corrective lens is needed?
37. A compound microscope has objective and eyepiece focal lengths of 6.1 mm and 1.7 cm, respectively. If the lenses are 8.3 cm apart, what is the instrument's magnification?

Problems

38. (a) Find the focal length of a concave mirror if an object placed 50 cm in front of the mirror has a real image 75 cm from the mirror. (b) Where and what type will the image be if the object is moved to a point 20 cm from the mirror?
39. A 12-mm-high object is 10 cm from a concave mirror with focal length 17 cm. (a) Where is the image, (b) how high is it, and (c) what type is it?
40. Repeat Problem 39 for a convex mirror, assuming all numbers stay the same.
41. An object's image in a 27-cm-focal-length concave mirror is upright and magnified by a factor of 3. Where is the object?
42. You're asked to design a concave mirror that will produce a virtual image, enlarged 1.8 times, of an object 22 cm from the mirror. What do you specify for the mirror's curvature radius?
43. Viewed from Earth, the Moon subtends an angle of 0.5° in the sky. How large an image of the Moon will be formed by the 3.6-m-diameter, 8.5-m-focal-length mirror of the Canada–France–Hawaii Telescope?
44. At what two distances could you place an object from a 45-cm-focal-length concave mirror to get an image 1.5 times the object's size?
45. LCD projectors commonly used for computer and video projection create an image on a small LCD display (see Application on page 345). The display is mounted before a lens and illuminated from behind. In a projector using a 7.50-cm-focal-length convex lens, where should the LCD display be located so the projected image is focused on a screen 6.30 m from the lens?
46. An object 15 cm from a concave mirror has a virtual image magnified 2.5 times. What's the mirror's focal length?
47. How far from a page should you hold a lens with 32-cm focal length in order to see the print magnified 1.6 times?
48. A converging lens has focal length 4.0 cm. A 1.0-cm-high arrow is located 7.0 cm from the lens with its lowest point 5.0 mm above the lens axis. Make a full-scale ray-tracing diagram to locate both ends of the image. Confirm using the lens equation.
49. A lens has focal length $f = 35$ cm. Find the type and height of the image produced when a 2.2-cm-high object is placed at distances (a) $f + 10$ cm and (b) $f - 10$ cm.
50. How far apart are the object and image produced by a converging lens with 35-cm focal length when the object is (a) 40 cm and (b) 30 cm from the lens?
51. A candle and a screen are 70 cm apart. Find two points between candle and screen where you could put a convex lens with 17-cm focal length to give a sharp image of the candle on the screen.
52. An object is placed two focal lengths from a diverging lens. (a) What type of image forms, (b) what's the magnification, and (c) where is the image?
53. How far from a 25-cm-focal-length lens should you place an object to get an upright image magnified 1.8 times?
54. An object and its lens-produced real image are 2.4 m apart. If the lens has 55-cm focal length, what are the possible values for the object distance and magnification?
55. An object is 68 cm from a plano-convex lens whose curved side has curvature radius 26 cm. The refractive index of the lens is 1.62. Where is the image, and what type is it?
56. Use Equation 31.6 to show that an object at the center of a glass sphere will appear to be its actual distance—one radius—from the edge. Draw a ray diagram showing why this makes sense.
57. Rework Example 31.4 for a fish 15 cm from the *far* wall of the tank.
58. Consider the inverse of Example 31.4: You're inside a 70-cm-diameter hollow tube containing air, and the tip of your nose is 15 cm from the tube's wall. The tube is immersed in water, and a fish looks in. To the fish, what's the apparent distance from your nose to the tube wall?
59. Two specks of dirt are trapped in a crystal ball, one at the center and the other halfway to the surface. If you peer into the ball on a line joining the two specks, the outer one appears to be only one-third of the way to the other. Find the refractive index of the ball.
60. BIO A contact lens is in the shape of a convex meniscus (see Fig. 31.25). The inner surface is curved to fit the eye, with curvature radius 7.80 mm. The lens is made from plastic with refractive index $n = 1.56$. If it has a 44.4-cm focal length, what's the curvature radius of its outer surface?
61. For what refractive index would the focal length of a plano-convex lens be equal to the curvature radius of its one curved surface?
62. An object is 28 cm from a double-convex lens with $n = 1.5$ and curvature radii 35 cm and 55 cm. Where is the image, and what type is it?
63. BIO You're an optician who's been asked to design a new replacement lens for cataract patients. The lens must be 5.5 mm in diameter, with focal length 17 mm, and it can't be thicker than 0.8 mm. For the lens material, you have a choice of plastic with refractive index 1.49 or more expensive silicone with $n = 1.58$. Which material do you choose, and why?
64. A double-convex lens with equal 38-cm curvature radii is made from glass with refractive indices $n_{\text{red}} = 1.51$ and $n_{\text{violet}} = 1.54$. If a point source of white light is on the lens axis 95 cm from the lens, over what range will its visible image be smeared?
65. An object placed 15 cm from a plano-convex lens of crown glass focuses to a virtual image twice the object's size. If the lens is replaced with an identically shaped one made from diamond, what type of image will appear and what will be its magnification? (See Table 30.1.)
66. You're taking a photography class, working with a camera whose zoom lens covers the focal-length range 38–110 mm. Your instructor asks you to compare the sizes of the images of a distant object when photographed at the two zoom extremes. Your answer?
67. A camera can normally focus as close as 60 cm, but it has provisions for mounting additional lenses just in front of the main lens to provide close-up capability. What type and power of auxiliary lens will allow the camera to focus as close as 20 cm?
68. A 300-power compound microscope has a 4.5-mm-focal-length objective lens. If the distance from objective to eyepiece is 10 cm, what should be the focal length of the eyepiece?
69. To the unaided eye, Jupiter has an angular diameter of 50 arcseconds. What will its angular size be when viewed through a 1-m-focal-length refracting telescope with a 40-mm-focal-length eyepiece?
70. A Cassegrain telescope like that shown in Fig. 31.33*b* has 1.0-m focal length, and the convex secondary mirror is located 0.85 m from the primary. What should be the focal length of the secondary in order to put the final image 0.12 m behind the front surface of the primary mirror?
71. You stand with your nose 6.0 cm from the surface of a reflecting ball, and your nose's image appears three-quarters full size. What's the ball's diameter?
72. BIO A contact lens prescription calls for +2.25-diopter lenses with inner curvature radius 8.6 mm to fit the patient's cornea. (a) If the lenses are plastic with $n = 1.56$, what should be the outer curvature radius? (b) With these lenses, the patient comfortably reads

a newspaper 30 cm from her eyes. Where's the image as viewed through the lenses?

73. Show that placing a 1-diopter lens in front of a 2-diopter lens gives the equivalent of a single 3-diopter lens (i.e., the powers of closely spaced lenses add).
74. Derive an expression for the thickness t of a plano-convex lens with diameter d, focal length f, and refractive index n.
75. Show that identical objects placed equal distances on either side of the focal point of a concave mirror or converging lens produce images of equal size. Are the images of the same type?
76. Generalize the derivation of the lensmaker's formula (Equation 31.7) to show that a lens of refractive index n_{lens} in an external medium with index n_{ext} has focal length given by

$$\frac{1}{f} = \left(\frac{n_{\text{lens}}}{n_{\text{ext}}} - 1\right)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$$

77. Draw a diagram like Fig. 31.10, but showing a ray from the arrowhead through the center of curvature. Using the fact that this ray reflects back on itself, draw similar triangles with object and image as their vertical sides, and show that the center of curvature is twice as far from the mirror as the focal point—that is, $R = 2f$, with R the curvature radius.
78. Galileo's first telescope used the arrangement shown in Fig. 31.36, with a double-concave eyepiece slightly before the focus of the objective lens. Use ray tracing to show that this design gives an upright image, which makes the Galilean telescope useful in terrestrial observing.

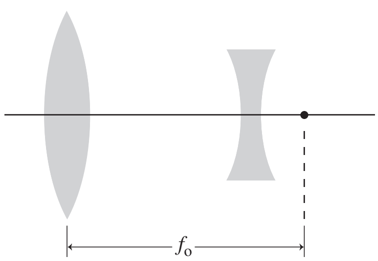


FIGURE 31.36 A Galilean telescope (Problem 78)

79. The maximum magnification of a simple magnifier occurs with the image at the 25-cm near point. Show that the angular magnification is $m = 1 + (25\text{ cm}/f)$, where f is the focal length.
80. *Chromatic aberration* results from variation of the refractive index with wavelength. Starting with the lensmaker's formula, find an expression for the fractional change df/f in the focal length of a thin lens in terms of the change dn in refractive index.
81. For visible wavelengths, the refractive index of a thin glass lens is $n = n_0 - b\lambda$, where $n_0 = 1.546$ and $b = 4.47\times10^{-5}\text{ nm}^{-1}$. If its focal length is 30 cm at 550 nm, how much does the focal length vary over a wavelength spread of 10 nm centered on 550 nm? (*Hint:* See Problem 80.)

Passage Problems

The *speed* of a camera lens measures its ability to photograph in dim light. Speed is characterized by *f-ratio*, also called the *f-number*, defined as the ratio of focal length f to lens diameter d. Thus an $f/2.8$ lens, for example, has diameter $d = f/2.8$. The actual amount of light a lens admits depends on its area A, but the inverse-square law shows that the light intensity at the camera's imaging sensor is proportional to A/f^2. Most cameras have an adjustable iris that obscures part of the lens to change the f-ratio in response to available light. Point-and-shoot cameras adjust the f-ratio automatically, but serious photographers use their camera's manual f-ratio adjustment (Fig. 31.37). "Stopping down" is the photographer's term for reducing the lens area using the adjustable iris.

FIGURE 31.37 A 35-mm camera lens (Passage Problems 82–85). The numbers from 22 to 2.8 at the bottom are values for the f-ratio, *f/d*. Turning the ring with these numbers adjusts the iris that covers the outer part of the lens, thus changing the f-ratio.

82. Zooming your camera's lens for telephoto shots increases the focal length. With no change in the lens area, this will
    a. increase the f-ratio and increase the lens speed.
    b. decrease the f-ratio and decrease the lens speed.
    c. increase the f-ratio and decrease the lens speed.
    d. not change the f-ratio or the lens speed.
83. Increasing the f-ratio from 2.8 to 5.6
    a. decreases the light admitted by a factor of 2.
    b. decreases the light admitted by a factor of 4.
    c. increases the light admitted by a factor of 2.
    d. increases the light admitted by a factor of 4.
84. You're given two lenses with different diameters. Knowing nothing else, you can conclude that
    a. the larger lens is faster.
    b. the smaller lens has the shorter focal length.
    c. the smaller lens suffers less spherical aberration.
    d. none of the above
85. If a lens suffers from spherical aberration, stopping down will
    a. worsen the focus.
    b. improve the focus.
    c. not affect the focus.

Answers to Chapter Questions

Answer to Chapter Opening Question

High-intensity laser light reshapes the cornea, so that refracted light converges to produce sharp images.

Answers to GOT IT? Questions

31.1. About half your height.
31.2. At the mirror's center of curvature, which is at twice the focal length.
31.3. Virtual image; convex lens.

32 Interference and Diffraction

How do optical principles ultimately govern the amount of information that can be stored on a DVD or Blu-ray disc?

New Concepts, New Skills

By the end of this chapter you should be able to

- Explain how interference occurs with light waves and why coherent light is essential in forming stable interference patterns (32.1).
- Describe double-slit interference quantitatively (32.2).
- Extend the description of double-slit interference to multiple slits and diffraction gratings (32.3).
- Explain the principles involved in grating spectroscopy and X-ray diffraction (32.3).
- Describe how interferometry provides a powerful technique for making sensitive distance measurements (32.4).
- Explain diffraction, and describe quantitatively the limitations it puts on our ability to form perfect optical images (32.5, 32.6).

Connecting Your Knowledge

- This chapter builds on your understanding of wave phenomena from Chapter 14, particularly wave interference and phase changes involved in reflection (14.5, 14.6).
- We'll draw on the electromagnetic nature of light waves as described in Chapter 29 (29.4, 29.8).

The preceding chapters described the behavior of light using geometrical optics—an approximation that's valid when we're dealing with length scales much larger than the wavelength of light, so we can ignore light's wave nature. We now turn to **physical optics**, which treats optical phenomena for which the wave nature of light plays an essential role. Two related phenomena, interference and diffraction, are central in physical optics.

32.1 Coherence and Interference

In Chapter 14 we showed how wave displacements add to produce **constructive interference** or **destructive interference**. Electromagnetic waves, including light, are no exception: Since electric and magnetic fields obey the superposition principle, the net fields at any point are the vector sums of individual wave fields. That sum may increase (constructive interference) or reduce (destructive interference) the net field.

Coherence

Steady interference patterns occur only when waves are **coherent**, meaning they maintain a constant phase relation. Ordinary light sources like lightbulbs or the Sun produce short wavetrains with random phases, so their light doesn't stay coherent for long (Fig. 32.1*a*). Lasers, in contrast, produce long wavetrains that maintain coherence over many wavelengths (Fig. 32.1*b*). The typical length of a wavetrain is the **coherence length** for a given light source.

It takes at least two waves to interfere, and even with lasers it's virtually impossible for two different sources to maintain coherence. Therefore, interference usually occurs when light from a single source is split, travels different paths, and is then recombined. Even light from an ordinary lightbulb will produce interference provided light in the recombining beams originated at very nearly the same place and time. Coherent laser light relaxes this restriction considerably. Coherence also requires that interfering light beams have exactly the same frequency and therefore wavelength or color. Again, lasers make things easier: Their light is very nearly **monochromatic**, consisting of a narrow band of wavelengths.

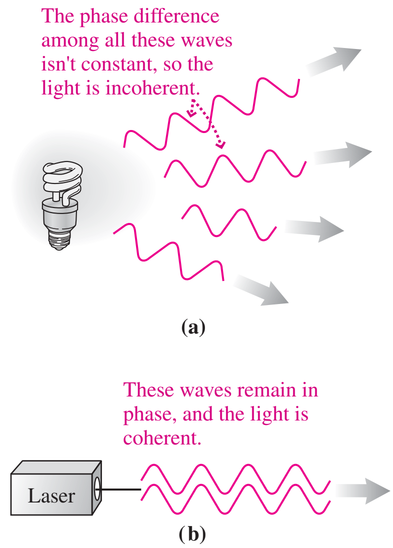


FIGURE 32.1 (a) Lightbulbs emit incoherent light consisting of short wavetrains with random phases. (b) Lasers produce coherent light, which facilitates stable interference.

Destructive and Constructive Interference

Consider light waves that originate together at a single source, travel two different paths, and then rejoin. Suppose one wave's path is exactly half a wavelength longer than the other. Then, when the waves recombine, they'll be out of phase by half a wavelength (Fig. 32.2*a*) and thus their superposition has smaller amplitude (zero, if the two interfering waves have exactly the same amplitude). If, on the other hand, the path lengths don't differ, or they differ by a full wavelength, then the two waves recombine in phase (Fig. 32.2*b*) and their superposition has larger amplitude. These two cases correspond, respectively, to destructive interference and constructive interference. It doesn't matter whether path lengths for the waves in Fig. 32.2*a* differ by half a wavelength, or $1\frac{1}{2}$ wavelengths, or $2\frac{1}{2}$ wavelengths; as long as the difference is an odd multiple of a half-wavelength, the waves recombine out of phase and destructive interference results. Thus **destructive interference results when light paths differ by an odd-integer multiple of a half-wavelength**. Similarly, it doesn't matter whether the path lengths in Fig. 32.2*b* are the same, or differ by 1, 2, 3, or any other integer number of wavelengths. Thus **constructive interference results when light paths differ by an integer multiple of the wavelength**.

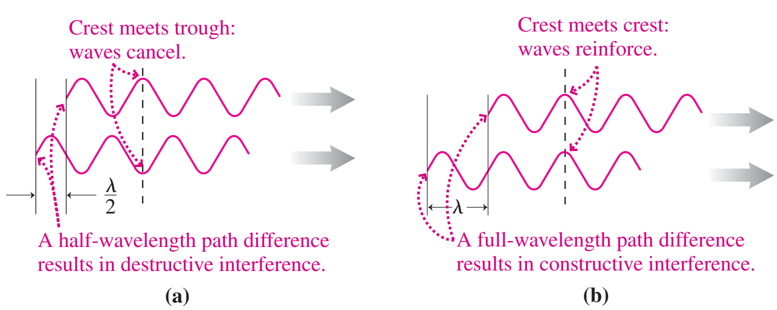
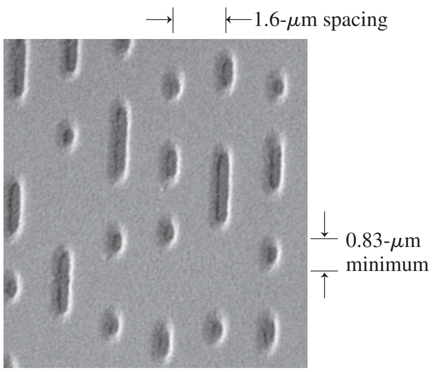


FIGURE 32.2 Two waves that start out in phase but travel different paths before rejoining.

There's one caveat to our statements: The path difference can't be greater than the coherence length; otherwise, the waves won't be coherent when they recombine. Once again, laser light has the advantage here because of its greater coherence length.

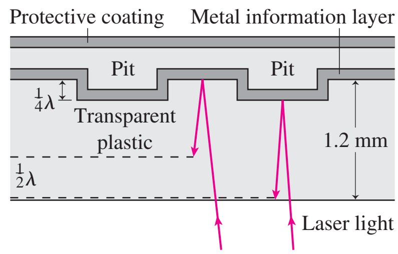
Of course, light paths don't have to differ by half or full multiples of the wavelength. In intermediate cases interfering waves superpose to make a composite wave whose amplitude may be enhanced or diminished, depending on the relative phase.

APPLICATION **More CD Music**



Example 30.3, "Refraction: CD Music," showed how refraction helps focus the laser beam that reads information from a compact disc. Interference, too, plays a crucial role in reading a CD.

Information on a CD, DVD, or Blu-ray disc is stored digitally in a sequence of pits stamped into a reflective metallic information layer, as shown in the photo. The pits' depth is very nearly one-quarter wavelength of the laser light used to read the disc. From the transparent underside of the disc, each pit appears as an elevated bump. Since the bumps stick down one-quarter wavelength, light reflecting off a bump follows a round-trip path that's shorter by half a wavelength than that of light reflecting off the undisturbed information layer (see the figure below). The laser beam is wider than the pit, so the reflected beam includes light both from the undisturbed disc and from the bump. The two interfere destructively, making the reflected beam less intense when a bump is present. As the disc spins, the result is a pattern of fluctuating light intensity conveying the information associated with the pattern of pits. A photodetector then converts that pattern to electrical signals that ultimately drive loudspeakers, headphones, or a video display.



32.2 Double-Slit Interference

In Chapter 14 we looked briefly at interference patterns produced by a pair of coherent sources. Such a pair can be made by passing light through two narrow slits. In 1801 Thomas Young used this approach in a historic experiment that confirmed the wave nature of light. Young first admitted sunlight to his laboratory through a hole small enough to ensure coherence of the incoming light. The light then passed through a pair of narrow, closely spaced slits, after which it illuminated a screen. Each slit acts as a source of cylindrical wavefronts that interfere in the region between slits and screen (Fig. 32.3*a*). Constructive and destructive interference produce **interference fringes**—alternating bright and dark bands (Fig. 32.3*b*).

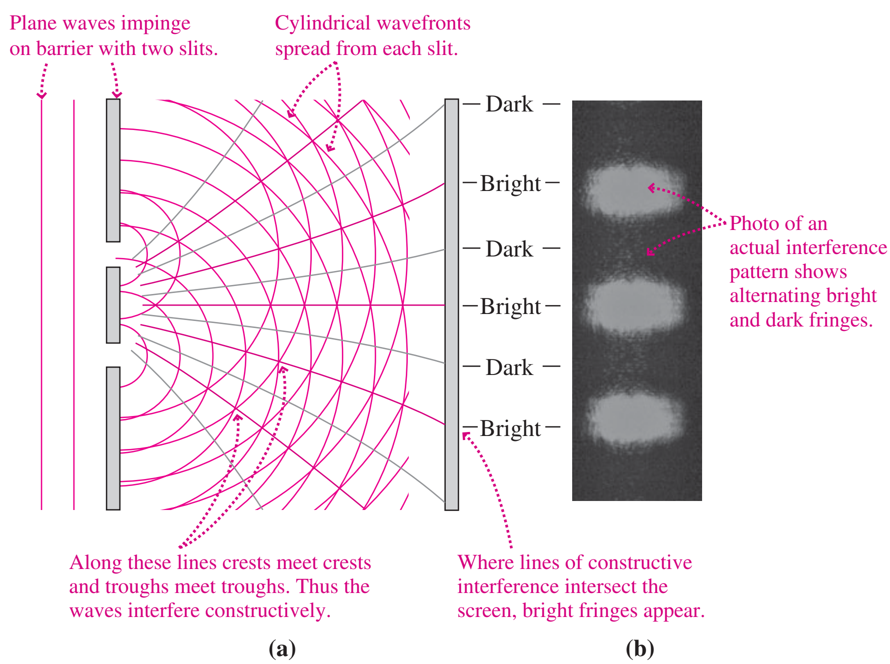


FIGURE 32.3 Double-slit interference results when light from a single source passes through closely spaced slits.

Bright fringes represent constructive interference, and therefore they occur where the difference in the path length for light traveling from the two slits is a multiple of the wavelength. When the distance L from slits to screen is much greater than the slit spacing d, Fig. 32.4 shows that the path difference to a point on the screen is $d\sin\theta$, where θ is the angular position of a point on the screen measured from an axis perpendicular to slits and screen. So our criterion for constructive interference—that this difference be an integer number of wavelengths—becomes

$$d\sin\theta = m\lambda \qquad (\text{bright fringes, } m = 0, 1, 2, \ldots) \tag{32.1a}$$

The integer m is the **order** of the fringe, with the central bright fringe being the zeroth-order fringe and with higher-order fringes on either side.

Waves interfere destructively when their path lengths differ by an odd-integer multiple of a half-wavelength:

$$d\sin\theta = \left(m + \tfrac{1}{2}\right)\lambda \qquad (\text{dark fringes, } m = 0, 1, 2, \ldots) \tag{32.1b}$$

where m is any integer.

In a typical double-slit experiment, L may be on the order of 1 m, d a fraction of 1 mm, and λ the sub-μm wavelength of visible light. Then we have the additional condition that $\lambda \ll d$. This makes the fringes very closely spaced on the screen, so the angle θ in Fig. 32.4 is small even for large orders m. Then $\sin\theta \simeq \tan\theta = y/L$, and a fringe's position y on the screen, measured from the central maximum, becomes

$$y_{\text{bright}} = m\frac{\lambda L}{d} \quad \text{and} \quad y_{\text{dark}} = \left(m + \tfrac{1}{2}\right)\frac{\lambda L}{d} \qquad \begin{pmatrix}\text{fringe position,}\\ \lambda \ll d\end{pmatrix} \tag{32.2a, b}$$

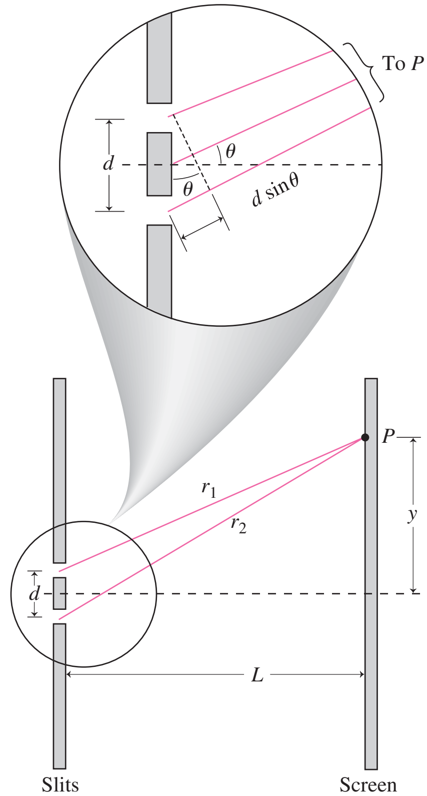


FIGURE 32.4 Geometry for finding locations of the interference fringes. In the blowup you can see that for $L \gg d$, the paths to P are nearly parallel and differ by $d\sin\theta$.

GOT IT? 32.1 If you increase the slit separation in a two-slit system, do the interference fringes become closer together or farther apart?

EXAMPLE 32.1 Measuring Wavelength: Laser Light

Two slits 0.075 mm apart are located 1.5 m from a screen. Laser light shining through the slits produces an interference pattern whose third-order bright fringe is 3.8 cm from the screen center. Find the light's wavelength.

INTERPRET The concept behind this problem is two-slit interference. The phrase "third order" tells us we're dealing with the $m = 3$ bright fringe located at $y_{\text{bright}} = 3.8$ cm.

DEVELOP Our plan is to use Equation 32.2a, $y_{\text{bright}} = m(\lambda L/d)$, and solve for λ. Since that equation requires $\lambda \ll d$, we'll then check to see whether our answer is consistent with this condition.

EVALUATE Solving, we have

$$\lambda = \frac{y_{\text{bright}}d}{mL} = \frac{(0.038\ \text{m})(0.075\times10^{-3}\ \text{m})}{(3)(1.5\ \text{m})} = 633\ \text{nm}$$

ASSESS This is indeed much less than the slit spacing of 0.075 mm or 75,000 nm. Our 633-nm result is in fact the wavelength of the red light from widely used helium–neon lasers.

Intensity in the Interference Pattern

We located the maxima and minima in two-slit interference using geometrical arguments alone. To find the actual intensity we need to superpose the interfering waves, which means adding the wave fields. So consider again a point P in the interference pattern (Fig. 32.5, next page). In the approximation $d \ll L$, the difference in path lengths is so small that we can neglect any difference in the amplitudes of the two waves. However, the difference in phase is crucial; it's what causes the interference. So we consider waves whose electric fields at P vary sinusoidally in time, with equal amplitude E_p but an explicit phase difference ϕ:

$$E_1 = E_p\sin\omega t \qquad \text{and} \qquad E_2 = E_p\sin(\omega t + \phi)$$

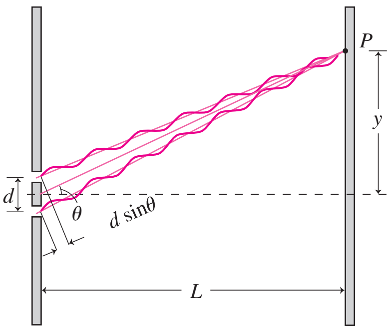


FIGURE 32.5 Waves from the slits arrive at P displaced by the path-length difference $d\sin\theta$. For $L \gg d$, $\sin\theta \simeq \tan\theta = y/L$.

We aren't bothering with vectors because the two waves are polarized in the same direction and therefore their fields add algebraically. Then the net electric field at P is

$$E = E_1 + E_2 = E_p[\sin\omega t + \sin(\omega t + \phi)]$$

Appendix A gives the trig identity $\sin\alpha + \sin\beta = 2\sin[(\alpha+\beta)/2]\cos[(\alpha-\beta)/2]$, which, with $\alpha = \omega t$ and $\beta = \omega t + \phi$, gives

$$E = 2E_p \sin\left(\omega t + \frac{\phi}{2}\right)\cos\left(\frac{\phi}{2}\right)$$

where we also used $\cos(-x) = \cos x$. Thus, the electric field at P oscillates with the wave frequency ω, and its amplitude is $2E_p\cos(\phi/2)$. Since the phase difference ϕ depends on the difference in path lengths from the two slits, this amplitude varies across the screen, giving the interference pattern.

We've seen that the path-length difference is $d\sin\theta$, with d the slit spacing and θ the angle to P; under our approximation $d \ll L$, θ is small and $\sin\theta \simeq \tan\theta = y/L$, where y is the position of P as shown in Fig. 32.5. Then the path difference becomes yd/L. Now, that all-important phase difference ϕ is whatever fraction of a full cycle (2π radians) this path difference is of the wavelength λ; that is, $\phi = 2\pi(yd/\lambda L)$. Then the amplitude $2E_p\cos(\phi/2)$ becomes $2E_p\cos(\pi yd/\lambda L)$. The average intensity follows from Equation 29.20b:

$$\overline{S} = \frac{[2E_p\cos(\pi yd/\lambda L)]^2}{2\mu_0 c} = 4\overline{S}_0\cos^2\left(\frac{\pi d}{\lambda L}y\right) \qquad (32.3)$$

where $\overline{S}_0 = E_p^2/2\mu_0 c$ is the average intensity of either wave alone. Now, $\cos^2$ has its maximum value, 1, when its argument is a multiple of π. Thus Equation 32.3 gives maximum intensity when $yd/\lambda L$ is an integer m, or when $y = m\lambda L/d$. This is just the condition 32.2a, showing that our intensity calculation is fully consistent with the simpler geometrical analysis. But the intensity calculation tells more: It gives not only the fringe positions, but also the intensity variation in between.

32.3 Multiple-Slit Interference and Diffraction Gratings

Systems with multiple slits play a crucial role in optical instrumentation and in the analysis of materials. As we'll see, gratings with several thousand slits per centimeter make possible high-resolution spectroscopic analysis. At a much smaller scale the regularly spaced rows of atoms in a crystal act much like a multiple-slit system for X rays, and the resulting X-ray patterns reveal the crystal structure.

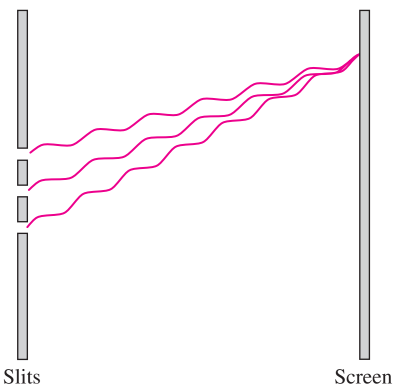


FIGURE 32.6 Waves from three evenly spaced slits interfere constructively when they reach the screen in phase.

Figure 32.6 shows waves from three evenly spaced slits interfering at a screen. Maximum intensity requires that all three waves be in phase or, equivalently, travel paths differing by an integer number of wavelengths. Our criterion for the maximum in a two-slit pattern, $d\sin\theta = m\lambda$, ensures that waves from two adjacent slits will add constructively. Since the slits are evenly spaced with distance d between each pair, waves coming through a third slit will be in phase with the other two if this criterion is met. So the criterion for a maximum in an N-slit system is still Equation 32.1a:

$$d\sin\theta = m\lambda \qquad \text{(maxima in multiple-slit interference, } m = 0, 1, 2, \ldots) \qquad (32.1a)$$

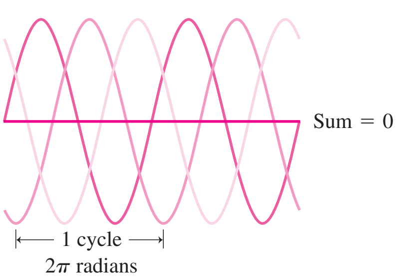


FIGURE 32.7 Waves from three slits must be out of phase by one-third of a cycle in order to interfere destructively.

With more than two waves, however, the criterion for destructive interference is more complicated. Somehow all the waves need to sum to zero. Figure 32.7 shows that this happens for three waves when each is out of phase with the others by one-third of a cycle. Thus, the path-length difference $d\sin\theta$ must be either $\left(m + \frac{1}{3}\right)\lambda$ or $\left(m + \frac{2}{3}\right)\lambda$, where m is an integer. The case $\left(m + \frac{3}{3}\right)\lambda$ is excluded because then the path lengths differ by a full wavelength, giving constructive interference and thus a maximum in the interference pattern. More generally we can write

$$d \sin \theta = \frac{m}{N}\lambda \tag{32.4}$$

for destructive interference in an N-slit system, where m is an integer *but not an integer multiple of N*.

Figure 32.8 shows interference patterns and intensity plots from some multiple-slit systems. Note that the bright, or *primary*, maxima are separated by several minima and fainter, or *secondary*, maxima. Why this complex pattern? Our analysis of the three-slit system shows two minima between every pair of primary maxima; for example, we considered the minima at $d \sin \theta$ equal to $\left(m + \frac{1}{3}\right)\lambda$ or $\left(m + \frac{2}{3}\right)\lambda$, which lie between the maxima at $d \sin \theta$ equal to $m\lambda$ and $(m + 1)\lambda$. More generally, Equation 32.4 shows that there are $N - 1$ minima between each pair of primary maxima given by Equation 32.1a. The secondary maxima that lie between these minima result from interference that is neither fully destructive nor fully constructive. The figure shows that the primary maxima become brighter and narrower as the number of slits increases, while the secondary maxima become relatively less bright. With a large number N of slits, then, we should expect a pattern of bright but narrow primary maxima, with broad, essentially dark regions in between.

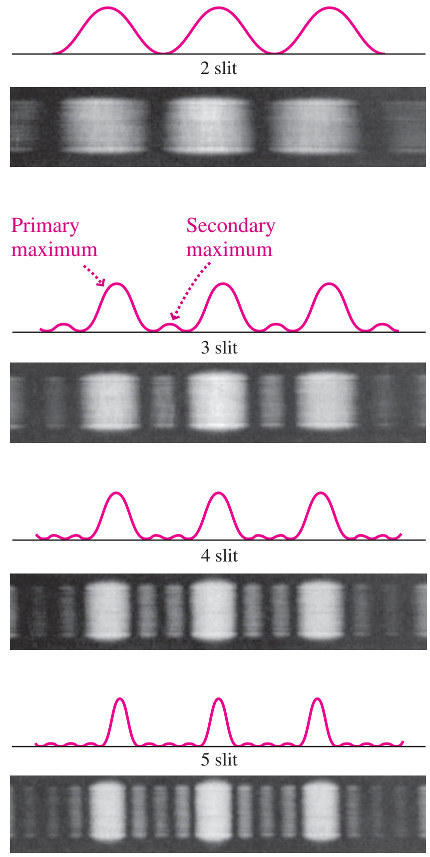


FIGURE 32.8 Interference patterns for multiple-slit systems with the same slit spacing. The bright fringes stay in the same place but become narrower and brighter as the number of slits increases. Intensity plots don't have the same vertical scale; peak intensity scales as the *square* of the number of slits.

Diffraction Gratings

A set of many closely spaced slits is called a **diffraction grating** and proves very useful in the spectroscopic analysis of light. Diffraction gratings are commonly several centimeters across and have several thousand slits—usually called lines—per cm. Gratings are made by photoreducing images of parallel lines or by ruling with a diamond stylus on aluminum-plated glass. Gratings like the slit systems we've been discussing are **transmission gratings**, since light passes through the slits. **Reflection gratings** produce similar interference effects by reflecting incident light.

We've seen that the maxima of the multiple-slit interference pattern are given by the same criterion, $d \sin \theta = m\lambda$, that applies to a two-slit system. For $m = 0$ this equation implies that all wavelengths peak together at the central maximum, but for larger values of m the angular position of the maximum depends on wavelength. Thus, a diffraction grating can be used in place of a prism to disperse light into its component wavelengths, and the integer m is therefore called the **order** of the dispersion. Figure 32.9 shows a spectrometer that works on this principle. Because the maxima in N-slit interference are very sharp for large N (recall Fig. 32.8), a grating with many slits diffracts individual wavelengths to very precise locations.

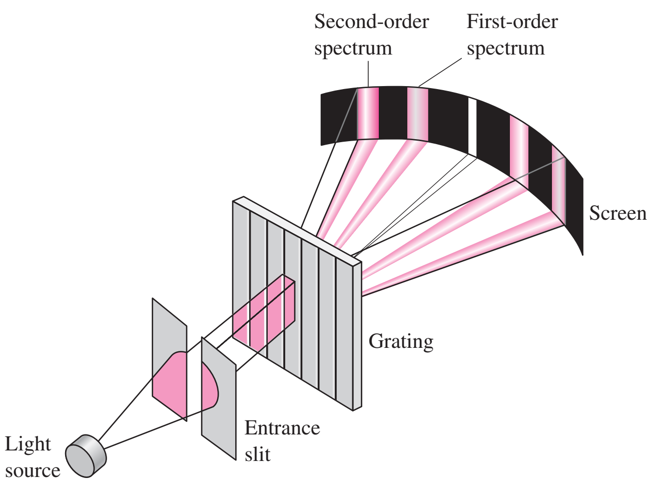


FIGURE 32.9 Essential elements of a grating spectrometer. An electronic detector normally replaces the screen.

EXAMPLE 32.2 Finding the Separation: A Grating Spectrometer

Light from glowing hydrogen contains many discrete spectral lines, of which two are Hα (hydrogen-alpha) and Hβ (hydrogen-beta), with wavelengths of 656.3 nm and 486.1 nm, respectively. Find the first-order angular separation between these wavelengths in a spectrometer that uses a grating with 6000 slits per cm.

INTERPRET The concept behind the grating spectrometer is multiple-slit interference, so our job is to find the angles to which the grating sends the given wavelengths. "First-order" means we have $m = 1$.

DEVELOP Equation 32.1a, $d\sin\theta = m\lambda$, gives the location of the interference maxima as a function of wavelength λ, order m, and slit spacing d. We're given m and two values for λ, but we don't know d. However, we're told that there are 6000 slits per cm, so we can find d. Our plan is first to calculate d, then use Equation 32.1a to find the angular positions for the two wavelengths, and finally take their difference to get the angular separation.

EVALUATE With 6000 slits/cm, the spacing is $d = 1/6000\text{ cm} = 1.667\ \mu\text{m}$. Applying Equation 32.1a with $m = 1$ gives

$$\theta_\alpha = \sin^{-1}\left(\frac{\lambda}{d}\right) = \sin^{-1}\left(\frac{0.6563\ \mu\text{m}}{1.667\ \mu\text{m}}\right) = 23.2^\circ$$

A similar calculation gives $\theta_\beta = 17.0^\circ$. Thus the angular separation is 6.2°.

ASSESS Our 6.2° result is certainly adequate to distinguish clearly these two wavelengths. For greater angular separation, or to separate closer wavelengths, we could look at the higher-order dispersion (see Exercise 20). ■

Resolving Power

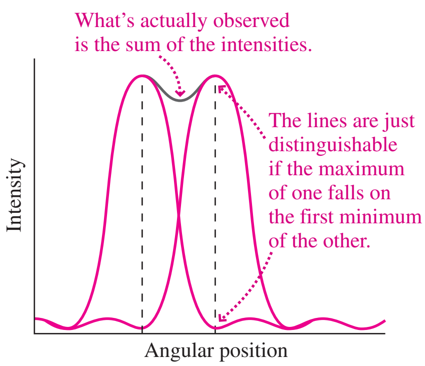


FIGURE 32.10 Intensity versus angular position for spectral lines with slightly different wavelengths, as dispersed with a grating.

The detailed shapes and wavelengths of spectral lines contain a wealth of information about the systems in which light originates. Studying these details requires a high dispersion in order to separate nearby spectral lines or to analyze the intensity-versus-wavelength profile of a single line. Suppose we pass light containing two spectral lines of nearly equal wavelengths λ and λ' through a grating. Figure 32.10 shows that we'll just be able to distinguish them if the peak of one line corresponds to the first minimum of the other; any closer and the lines blur together. Suppose wavelength λ has its mth-order maximum at angular position θ, so $d\sin\theta_{\text{max}} = m\lambda$. We can write this as $d\sin\theta_{\text{max}} = (mN/N)\lambda$, with N the number of slits in the grating. Equation 32.4 then shows that we get an adjacent minimum if we add 1 to the numerator mN. Thus the adjacent minimum satisfies

$$d\sin\theta_{\text{min}} = \frac{mN + 1}{N}\lambda$$

Our criterion that the two wavelengths λ and λ' be distinguishable is that the maximum for λ' fall at the location of this minimum for λ. But the maximum for λ' satisfies $d\sin\theta'_{\text{max}} = m\lambda' = (mN/N)\lambda'$, so for $\theta'_{\text{max}} = \theta_{\text{min}}$ we must have $(mN + 1)\lambda = mN\lambda'$. Expressing this in terms of the wavelength difference $\Delta\lambda = \lambda' - \lambda$ leads to

$$\frac{\lambda}{\Delta\lambda} = mN \qquad \text{(resolving power)} \tag{32.5}$$

The quantity $\lambda/\Delta\lambda$ is the grating's **resolving power**, a measure of its ability to distinguish closely spaced wavelengths. The higher the resolving power, the smaller the wavelength difference $\Delta\lambda$ that we can distinguish. Equation 32.5 shows that the resolving power increases with the number of lines, N, on the grating and also with the order, m, of the spectrum we observe.

EXAMPLE 32.3 Resolving Power: "Seeing" a Double Star

A double-star system consists of a massive star essentially at rest, with a smaller companion in circular orbit. It's far too distant for the pair to appear as anything but a single point even to the largest telescopes. Yet astronomers can "see" the companion star through the Doppler shift in wavelengths of its spectral lines. The Hα spectral line from the stationary massive star is at $\lambda = 656.272$ nm; for the companion when it's moving away from Earth, the Hα line Doppler-shifts to 656.329 nm (corresponding to a speed of about 26 km/s). If a spectrometer has 5000 lines, what order spectrum will resolve the Hα lines from the two stars?

INTERPRET The concept here is resolution of distinct spectral lines using a grating spectrometer.

DEVELOP Equation 32.5, $\lambda/\Delta\lambda = mN$, determines the resolving power. We can solve for m to get the order: $m = \lambda/(N\,\Delta\lambda)$. We're given the two wavelengths, so we can readily find $\Delta\lambda$.

EVALUATE We have $\Delta\lambda = 656.329\text{ nm} - 656.272\text{ nm} = 0.057\text{ nm}$. Then

$$m = \frac{\lambda}{N\,\Delta\lambda} = \frac{656.272\text{ nm}}{(5000)(0.057\text{ nm})} = 2.3$$

ASSESS Since m must be an integer, we'll have to use the third-order spectrum.

X-Ray Diffraction

The wavelengths of X rays, on the order of 0.1 nm, are far too short for diffraction with gratings produced mechanically or photographically. Instead, **X-ray diffraction** occurs when X rays interact with the regularly spaced atoms in a crystal. At the microscopic level, reflection of an electromagnetic wave occurs when the wave's electric field sets electrons oscillating. The electrons re-radiate, producing the reflected beam. With X rays reflecting from a crystal, the regular atomic spacing results in interference that enhances the reflected radiation at certain angles. Figure 32.11*a* shows an X-ray beam interacting with the atoms in a crystal. In Fig. 32.11*b* we see that waves reflecting at one layer of atoms travel a distance $2d\sin\theta$ farther than those reflecting at the layer above, where θ is the angle between the incident beam and the atomic planes. Constructive interference occurs when this difference is an integer number of wavelengths:

$$2d\sin\theta = m\lambda \qquad (\text{Bragg condition, } m = 1, 2, 3 \ldots) \tag{32.6}$$

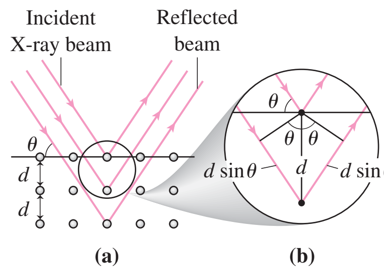


FIGURE 32.11 (a) X rays reflecting off the planes of atoms in a crystal. (b) Constructive interference enhances the outgoing beam when the extra distance $2d\sin\theta$ is an integer multiple of the X-ray wavelength.

This **Bragg condition** lets us use a crystal with known spacing as a diffraction grating for X rays. More important is the converse: Much of what we know about crystal structure comes from probing crystals with X rays and using the resulting patterns to deduce positions of their atoms. X-ray diffraction measurements by British scientist Rosalind Franklin in 1952 were crucial in establishing the structure of DNA.

Other Gratings

Anything with regularly spaced structures can act as a diffraction grating for waves of suitable wavelength. The rainbow of colors you see on the underside of a CD or DVD (see this chapter's opening photo) results because adjacent pits of CD tracks (shown in the Application earlier in this chapter) act as a diffraction grating. Sound waves in a solid set up refractive index variations that act as diffraction gratings; changing the wavelength of the sound changes the "slit" spacing and therefore the diffraction angle. **Acousto-optic modulators** (AOMs) based on this principle are widely used to "steer" light beams in lightwave communication and other opto-electronic technologies. Laser printers and digital copiers, for example, use an AOM to control the laser beam that "paints" the image of the printed page on a light-sensitive surface.

GOT IT? 32.2 If you increase the number of slits in a grating while keeping the spacing the same, what happens to (a) the positions of the intensity maxima in the interference pattern for a given wavelength; (b) the intensity at the maxima; and (c) the width of the maxima?

32.4 Interferometry

Passing light through multiple slits isn't the only way to produce interference. So will any process that separates light into several beams, sends them on different paths, and then rejoins them. Such processes are the basis of **interferometry**, an exquisitely sensitive technique for measuring small displacements, time intervals, and other quantities.

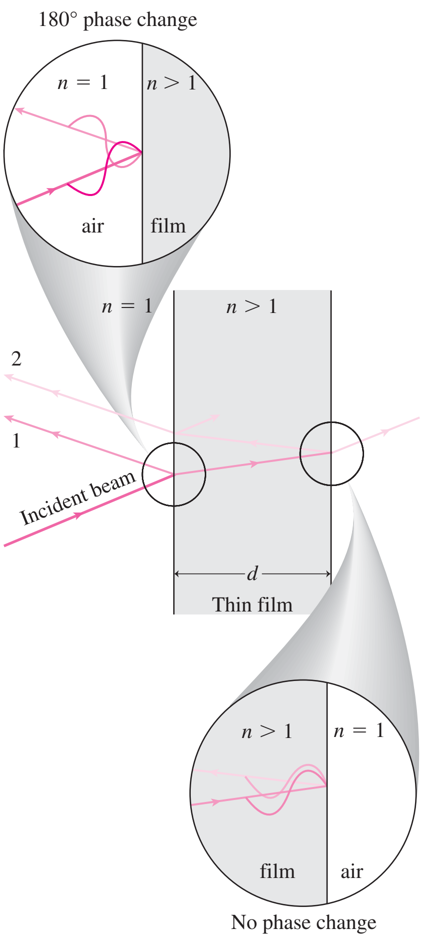


FIGURE 32.12 Reflection and refraction at a thin film of transparent material, showing a 180° phase change at the first interface and none at the second.

Thin Films

Light passing through thin, transparent films is partially reflected at both the front and back surfaces, and the resulting beams recombine to produce interference. In Section 14.6 we saw how waves on a string reflect where one string joins another with different properties; in particular, the reflected wave is inverted if the second string has greater mass per unit length. Light behaves analogously: It's reflected with a 180° phase change at the interface between a material with lower refractive index and one with higher refractive index. But it reflects without a phase change at an interface from higher to lower refractive index. For a thin film with refractive index n higher than its surroundings, Fig. 32.12 shows that there's a 180° phase change at the first interface and no change at the second.

If the film in Fig. 32.12 has thickness d, there's also a phase change due to the additional path length for beam 2; for the case of normal incidence, the extra length is $2d$. Because reflected beam 1 forms with a 180° phase change but beam 2 has no phase change, it takes another 180° phase shift—half a wavelength path difference—to put beams 1 and 2 back in phase to give constructive interference. That occurs if beam 2's extra path length $2d$ is half a wavelength, or $1\frac{1}{2}$ wavelengths, or any odd-integer multiple of one-half wavelength: $2d = \left(m + \frac{1}{2}\right)\lambda_n$, where $m = 0, 1, 2, 3, \ldots,$ and where the subscript n indicates that this is the wavelength as measured in the material. In Chapter 30 we found that the wavelength in a material with refractive index n is reduced by a factor $1/n$ from its value in air or vacuum; thus $\lambda_n = \lambda/n$, and our condition for constructive interference becomes

$$2nd = \left(m + \tfrac{1}{2}\right)\lambda \qquad \text{(constructive interference, thin film)} \qquad (32.7)$$

Interference in thin layers is the basis of some very sensitive optical techniques. The shape of a lens, for example, can be measured to within a fraction of a wavelength of light using interference in a thin "film" consisting of the air between the lens and a flat glass plate (Fig. 32.13).

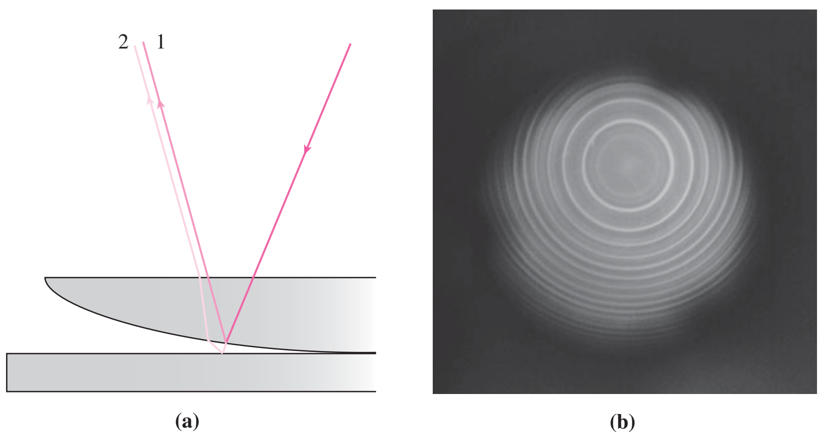


FIGURE 32.13 (a) A portion of a lens sitting on a flat glass plate. (b) Newton's rings arise from the difference in path lengths between rays like 1 and 2, and provide precise information about the lens shape.

CONCEPTUAL EXAMPLE 32.1 Interference: A Soap Film

Figure 32.14 shows a soap film in a circular ring with different colored bands that run horizontally across the film. Why do these bands occur?

EVALUATE The bands must be the result of interference, as described in Fig. 32.12. But why are they different colors? The soap film is vertical and consists mostly of water, so gravity makes it thicker at the bottom. Therefore, the wavelengths that undergo constructive interference vary with vertical position on the film. There are many bands, with colors repeating, because there are multiple orders of interference, with different values of m.

FIGURE 32.14 Interference in a soap film illuminated with white light.

ASSESS The dark region at the top confirms our explanation: Here the film is so thin that no visible wavelength undergoes constructive interference, so it appears dark. The film is probably about to break!

MAKING THE CONNECTION A 20-cm-high soap film is 1 μm thick at the bottom, tapering to near zero thickness at the top. If it's illuminated with 650-nm laser light, how many bright bands appear?

EVALUATE Equation 32.7 gives the condition for constructive interference; the number of bands will be the number of interference orders m possible in this film. Solving for m at the bottom of the film gives $m = 2n\lambda/d - 1/2 = 3.6$, with $n = 1.333$ for water. Since m must be an integer, there's no bright band right at the bottom. The $m = 3$ band is higher up, and above it are the $m = 2$, $m = 1$, and $m = 0$ bands, for a total of four bright bands.

In analyzing thin films, we've considered only the first reflection at each interface. Actually, multiple reflections occur within the film, producing ever-weaker rays. A fuller treatment involving Maxwell's equations shows that when a film of refractive index n_2 is sandwiched between materials with indices n_1 and n_3, complete cancellation of reflected rays in the incident medium occurs if the thickness is right and if $n_2 = \sqrt{n_1 n_3}$. This is the basis of the antireflection coatings, mentioned in Chapter 30, which ensure maximum light transfer in camera lenses, solar photovoltaic cells, and other applications.

The Michelson Interferometer

Several optical instruments use interference for precise measurement. Among the simplest and most important is the **Michelson interferometer**, invented by the American physicist Albert Michelson and used in the 1880s in a famous experiment that paved the way for the theory of relativity. We discuss this experiment in the next chapter; here we describe the interferometer, which is still used for precision measurements.

Figure 32.15 shows the basic Michelson interferometer. The key idea is that light from a monochromatic source is split into two beams by a half-silvered mirror called a **beam splitter**. The beam splitter is set at a 45° angle, so the reflected and transmitted beams travel perpendicular paths. Each then reflects off a flat mirror and returns to the beam splitter. The beam splitter again transmits and reflects half the light incident on it, with the result that some light from the originally separated beams is recombined. The recombined beams interfere, and the interference pattern is observed with a viewing lens; an example of the resulting pattern is shown at the bottom of Fig. 32.15.

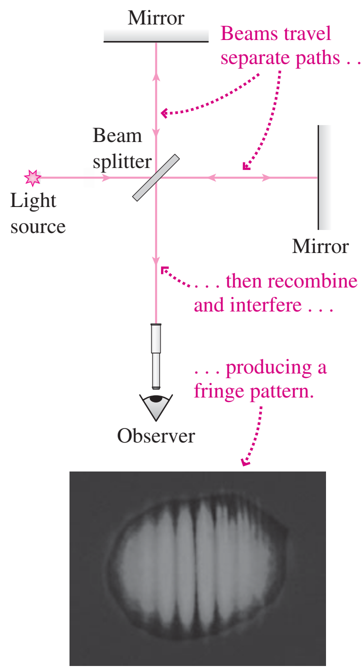


FIGURE 32.15 Schematic diagram of a Michelson interferometer, with a photo of the interference fringes.

If the path lengths for the two beams were exactly the same, they would recombine in phase and interfere constructively. In reality, the path lengths are never exactly the same, the mirrors are never exactly perpendicular, and the beams aren't perfectly parallel. But that's no problem: What happens is that different parts of the beams recombine with different phase differences, and the result is a pattern of light and dark interference fringes, as shown in Fig. 32.15. The distance between successive fringes corresponds to a path-length difference of one full wavelength.

Now suppose one mirror moves slightly. The path-length differences change and therefore the interference pattern shifts. A mere quarter-wavelength mirror movement adds an extra half-wavelength to the round-trip path. That results in a 180° phase shift, moving dark fringes to where light ones were. Shifts a fraction of this amount are readily detected, allowing the measurement of mirror displacements to within a small fraction of a wavelength. A similar shift occurs if a transparent material is placed in one path, retarding the beam because of its refractive index. This provides accurate measures of the refractive indices of gases, which are so close to 1 that less-sensitive techniques don't work.

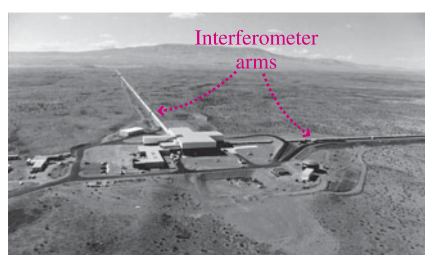


FIGURE 32.16 The LIGO instrument at Hanford, Washington, is an interferometer with 4-km arms. The light undergoes multiple reflections, giving an effective arm length of 300 km.

The largest Michelson interferometers ever built are twin instruments with 4-km arms (Fig. 32.16). These comprise LIGO, the Laser Interferometer Gravitational Wave

Observatory. LIGO is designed to detect mirror displacements on the order of 10^{-18} m resulting from gravitational waves—"ripples" in the structure of space and time caused by distant cosmic events. LIGO will eventually be dwarfed by a space-based interferometer with arms 5 million km long!

EXAMPLE 32.4 An Interferometric Measurement: Sandstorm!

A sandstorm has pitted the aluminum mirrors of a desert solar-energy installation, and engineers want to know the depths of the pits. They construct a Michelson interferometer with a sample from one of the pitted mirrors in place of one flat mirror. With 633-nm laser light, the interference pattern in Fig. 32.17 results. What is the approximate depth of the pit?

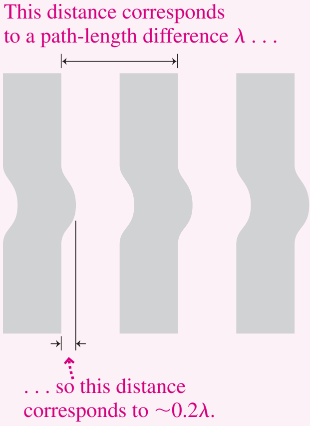


FIGURE 32.17 Fringe pattern resulting from a pitted mirror.

INTERPRET The concept here is interferometry—inferring distance from observations of light interference, in this case with the Michelson configuration of Fig. 32.15. The distortions of the interference fringes shown in Fig. 32.17 result because some of the light travels a little farther—namely, into the bottom of a pit and back out.

DEVELOP The full fringe spacing corresponds to a path difference of one wavelength, and Fig. 32.17 shows that the fringe distortion due to the pit gives a shift about one-fifth of the distance between fringes. We need to use this information to find the extra distance traveled by light reflecting from the bottom of the pit, and from that the pit depth. We're given the wavelength, so we can estimate the round trip as approximately 0.2λ. The pit depth will be half this quantity.

EVALUATE The extra path length for the light reflecting off the pit is 0.2λ, so the pit depth is about 0.1λ. With $\lambda = 633$ nm, the pit depth is about 63 nm.

ASSESS Try measuring that with a meter stick! Interferometry provides an exquisitely sensitive measurement of small distances. ■

32.5 Huygens' Principle and Diffraction

The interference we've been studying in this chapter isn't the only optical phenomenon where the wave nature of light is important. There's also **diffraction**—the bending of light or other waves as they pass by objects. Interference and diffraction are closely related, and the double- and multiple-slit interference we've studied actually involves diffraction as well—hence the term "diffraction grating."

Diffraction, like other optical phenomena, is ultimately governed by Maxwell's equations. But we can understand diffraction more readily using **Huygens' principle**, articulated in 1678 by the Dutch scientist Christian Huygens, who was the first to suggest that light might be a wave. Huygens' principle states:

> All points on a wavefront act as point sources of spherically propagating "wavelets" that travel at the speed of light appropriate to the medium. At a short time Δt later, the new wavefront is the unique surface tangent to all the forward-propagating wavelets.

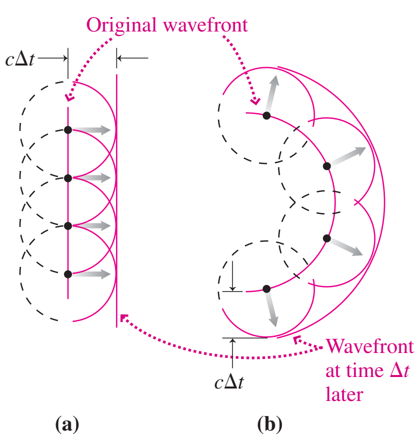


FIGURE 32.18 Application of Huygens' principle to (a) plane and (b) spherical waves. In each case the wavefront acts like a set of point sources emitting circular waves that expand to produce a new wavefront.

Figure 32.18 shows how Huygens' principle accounts for the propagation of plane and spherical waves.

Diffraction

Figure 32.19 shows plane waves incident on an opaque barrier containing a hole. Since the waves are blocked by the barrier, Huygens' wavelets produced near each barrier edge cause the wavefronts to bend at the barrier. When the width of the hole is much greater than the wavelength, as in Fig. 32.19a, this diffraction is of little consequence, and the waves effectively propagate straight through the hole in a beam defined by the hole size. But when the hole size and wavelength are comparable, wavefronts emerging from the hole spread in a broad pattern (Fig. 32.19*b*). Thus diffraction, although it always occurs, is significant only on length scales comparable to or smaller than the wavelength. That's why we could ignore diffraction and assume that light always travels in straight lines when we considered optical systems with dimensions much larger than the wavelength of light.

Diffraction ultimately limits our ability to image small objects and to focus light precisely. Next, we'll see why this is so by examining the behavior of light as it passes through a single slit. The result will help us understand optical challenges ranging from telescopic imaging of distant astrophysical objects to the development of Blu-ray discs.

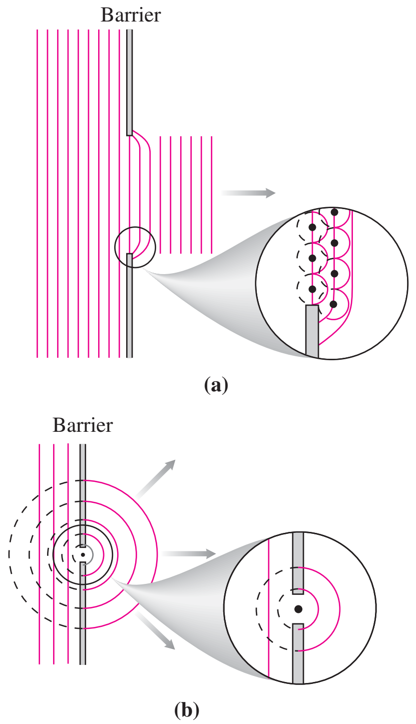


FIGURE 32.19 Plane waves incident on an opaque barrier with a hole. Diffraction is negligible for a hole large compared with the wavelength (a), but pronounced for a small hole (b).

Single-Slit Diffraction

In treating double-slit and multiple-slit interference, we assumed that plane waves passing through a slit emerged with circular wavefronts. According to Fig. 32.19*b*, that's true only if the slit width is small compared with the wavelength, so the slit can be treated as a single, localized source of new waves. When the slit width isn't small, Huygens' principle implies that we have to consider each point in the slit as a separate source—and then we can expect interference from waves originating at different points in the same slit. Thus a single wide slit is really like a multiple-slit system with infinitely many slits!

Figure 32.20*a* shows light incident on a slit of width a. Each point in the slit acts as a source of spherical wavelets propagating in all directions to the right of the slit. We focus on a particular direction described by the angle θ, and we'll look at interference of light from the five points shown. Figure 32.20*b* concentrates on the points from which rays 1, 2, and 3 originate and shows that the path lengths for rays 1 and 3 differ by $\frac{1}{2}a\sin\theta$. These two beams will interfere destructively if this distance is half the wavelength—that is, if $\frac{1}{2}a\sin\theta = \frac{1}{2}\lambda$ or $a\sin\theta = \lambda$. But if rays 1 and 3 interfere destructively, so do rays 3 and 5, which have the same geometry, and so do rays 2 and 4, for the same reason. In fact, a ray leaving *any* point in the lower half of the slit will interfere destructively with the point located a distance $a/2$ above it. Therefore, an observer viewing the slit system at the angle θ satisfying $a\sin\theta = \lambda$ will see no light.

Similarly, the sources for rays 1 and 2 are $a/4$ apart and will therefore interfere destructively if $\frac{1}{4}a\sin\theta = \frac{1}{2}\lambda$, or $a\sin\theta = 2\lambda$. But then so will rays 2 and 3, and rays 3 and 4; in fact, any ray from a point in the lower three-quarters of the slit will interfere destructively with a ray from the point $a/4$ above it, and therefore, an observer looking at an angle θ satisfying $a\sin\theta = 2\lambda$ will see no light.

We could equally well have divided the slit into six sections with seven evenly spaced points; we would then have found destructive interference if $\frac{1}{6}a\sin\theta = \frac{1}{2}\lambda$, or $a\sin\theta = 3\lambda$. We could obviously continue this process for any number of points in the slit, and therefore, we conclude that destructive interference occurs for all angles θ satisfying

$$a\sin\theta = m\lambda \qquad \text{(destructive interference, single-slit diffraction)} \qquad (32.8)$$

with m any nonzero integer and a the slit width. Note that the case $m = 0$ is excluded; it produces not destructive interference but a central maximum in which all waves are in phase.

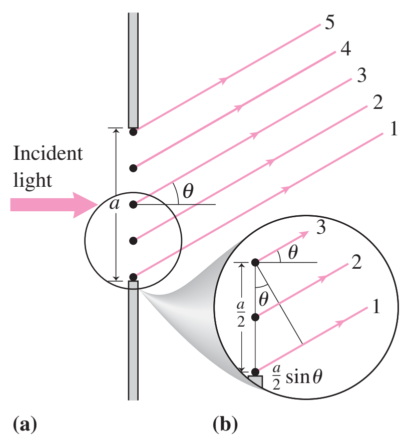


FIGURE 32.20 Each point in a slit acts as a source of Huygens' wavelets, which interfere in the region to the right of the slit.

✓TIP Interference and Diffraction

Equation 32.8 for the *minima* of a single-slit diffraction pattern looks just like Equation 32.1a for the *maxima* of a multiple-slit interference pattern, except that the slit width a replaces the slit spacing d. Why does the same equation give the minima in one case and the maxima in another? Because we're dealing with two distinct but related phenomena. In the multiple-slit case, each slit was so narrow that it could be considered a single source, neglecting the interference of waves originating within the same slit. In the single-slit case, the diffraction pattern occurs precisely because of the interference of waves from different points within the same slit.

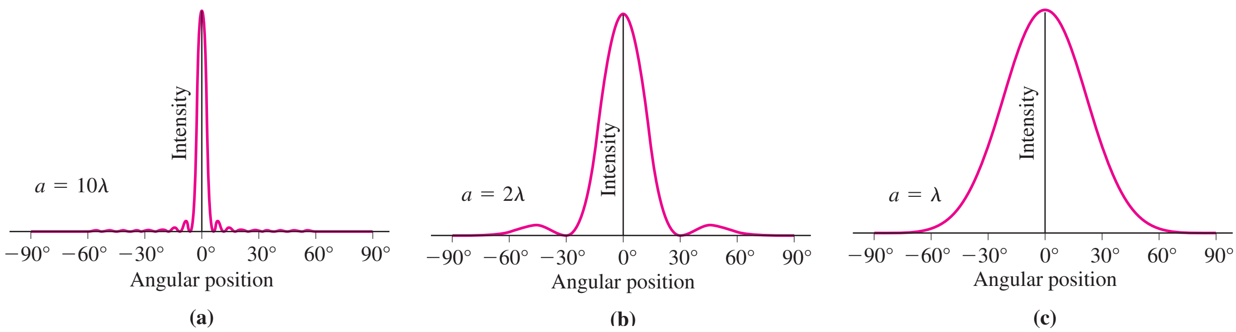


FIGURE 32.21 Intensity in single-slit diffraction, as a function of the angle θ from the centerline, for three values of slit width a.

Intensity in Single-Slit Diffraction

Geometry gave us the *positions* of the maxima in single-slit diffraction, just as it did for multiple-slit interference. But to get the *intensity* across the diffraction pattern, we'd have to superpose electric fields of the interfering waves, as we did in deriving Equation 32.3 for two-slit interference. But now we've got infinitely many fields to sum, corresponding to waves from every part of the slit. It's possible to do this using a calculus-like graphical technique involving the phasor concept introduced in Chapter 28. We won't go through the derivation; however, we'll motivate the result by noting that the diffraction pattern at any point occurs because of phase differences among waves originating at different parts of the slit. It's not surprising, therefore, that a key factor is the phase difference between waves from opposite ends of the slit. Applying the analysis of Fig. 32.20 to the entire slit width a gives a path-length difference $a\sin\theta$ for these waves. As usual, this path difference is to the wavelength as the phase difference ϕ is to a full cycle, 2π radians. Thus

$$\phi = \frac{2\pi}{\lambda} a \sin\theta \tag{32.9}$$

is the phase difference between rays from the ends of the slit to a point at angular position θ. The phasor-summing process relates the amplitude of the net electric field to this phase difference ϕ and shows that the field is proportional to $\sin(\phi/2)/(\phi/2)$. Used in Equation 29.20 to get the intensity from the electric field, this result gives the intensity as a function of angle in single-slit diffraction:

$$\bar{S} = \bar{S}_0 \left[\frac{\sin(\phi/2)}{\phi/2}\right]^2 \tag{32.10}$$

Here $\bar{S}_0$ is the average intensity at the central maximum of the pattern $(\theta = \phi = 0)$, and ϕ is given by Equation 32.9. At $\theta = \phi = 0$, Equation 32.10 appears to be indeterminate, but using the limit $\sin x/x \to 1$ as $x \to 0$ shows that the result is indeed $\bar{S}_0$. Problem 68 explores another approach to Equation 32.10.

Figure 32.21 plots Equation 32.10 for three values of the slit width a in relation to the wavelength λ. For wide slits—a large compared with λ—the central peak is narrow and the secondary peaks are much lower and also half as wide as the central peak. Here diffraction is negligible, and the beam essentially propagates through the slit in the ray approximation of geometrical optics. But as the slit narrows, the diffracted beam spreads until, with $a = \lambda$, it covers an angular width of some 120°.

The intensity given by Equation 32.10 will be zero when the numerator on the right-hand side is zero—that is, when the argument of the sine function is an integer multiple of π. That occurs when $\phi/2 = (\pi a/\lambda)\sin\theta = m\pi$, or when $a\sin\theta = m\lambda$. Thus, we recover our result of Equation 32.8 for the angular positions where destructive interference gives zero intensity.

Multiple Slits and Other Diffracting Systems

In treating multiple-slit systems in Section 32.2, we assumed the slits were so narrow compared with the wavelength that the central diffraction peak spread into the entire space beyond the slit system. When the slit width isn't negligible, each slit produces a

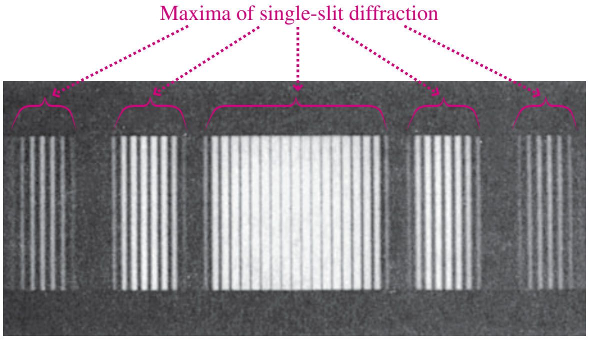


FIGURE 32.22 When the slit width is not negligible, a double-slit system produces the regular variations of double-slit interference within a single-slit diffraction pattern.

single-slit diffraction pattern. The result is a pattern that combines single-slit diffraction with multiple-slit interference (Fig. 32.22).

Diffraction occurs any time light passes a sharp, opaque edge like the edges of the slits we've been considering. Close examination of the shadow produced by a sharp edge shows parallel fringes resulting from interference of the diffracting wavefronts (Fig. 32.23*a*). More complex diffraction patterns result from objects of different shape (Fig. 32.23*b*). Such diffraction limits our ability to form sharp optical images, as we show in the next section.

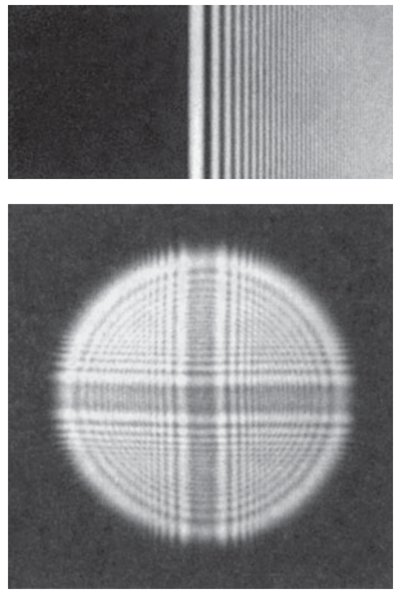


FIGURE 32.23 Diffraction patterns from light passing sharp edges. (a) Straight edge of an opaque barrier. (b) Circular aperture with crosshairs.

32.6 The Diffraction Limit

Diffraction imposes a fundamental limit on the ability of optical systems to distinguish closely spaced objects. Consider two point sources of light illuminating a slit. The sources are so far from the slit that waves reaching the slit are essentially plane waves, but the different source positions mean the waves reach the slit at different angles. We assume the sources are incoherent, so they don't produce a regular interference pattern. Then light diffracting at the slit produces two single-slit diffraction patterns, one for each source. Because the sources are at different angular positions, the central maxima of these patterns don't coincide, as shown in Fig. 32.24.

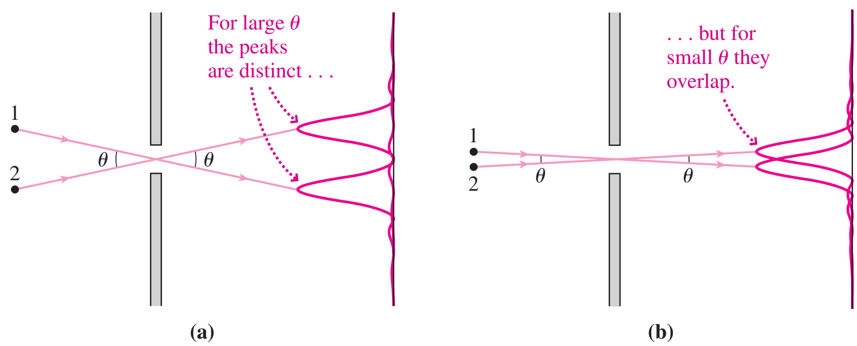


FIGURE 32.24 Two distant light sources at different angular positions produce diffraction patterns whose central peaks have the same angular separation θ as the sources.

If the angular separation between the sources is great enough, then the central maxima of the two diffraction patterns will be entirely distinct. In that case we can clearly distinguish the two sources (Fig. 32.24*a*). But as the sources get closer, the central maxima begin to overlap (Fig. 32.24*b*). They remain distinguishable as long as the total intensity pattern shows two peaks. Since the sources are incoherent, the total intensity is just the sum of the individual intensities. Figure 32.25 shows how that sum loses its two-peak structure as the diffraction patterns merge. In general, two peaks are barely distinguishable if the central maximum of one coincides with the first minimum of the other. This condition is called the **Rayleigh criterion**, and when it's met the two sources are just barely **resolved**.

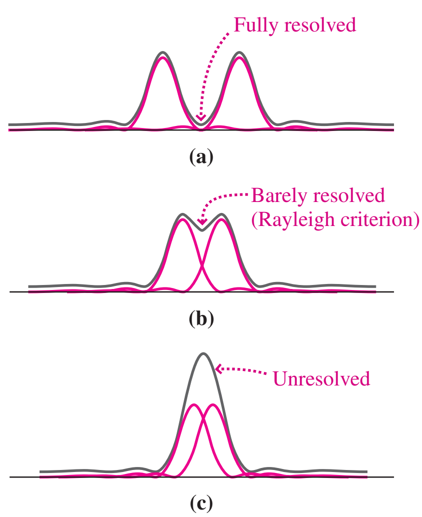


FIGURE 32.25 Since the two sources are incoherent, the total intensity is just the sum (gray curve) of the intensities of the two diffraction patterns.

Optical systems are analogous to the single slit we've just considered. Every system has an aperture of finite size through which light enters. That aperture may be an actual slit or hole, like the diaphragm that "stops down" a camera lens, or it may be the full size of a lens or a telescope mirror. So all optical systems ultimately suffer loss of resolution if two sources—or two parts of the same object—have too small an angular separation. Thus, diffraction fundamentally limits our ability to probe the structure of objects that are either very small or very distant. Figure 32.26 shows the loss of resolution as diffraction patterns overlap.

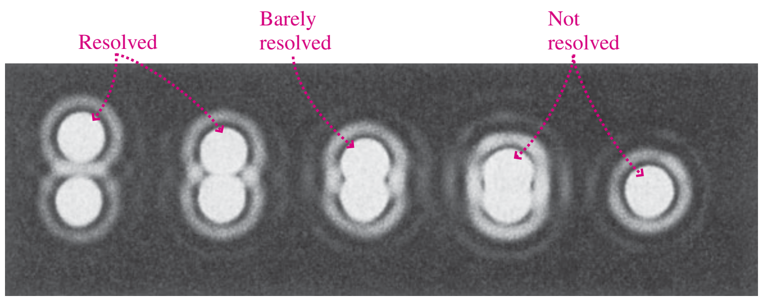


FIGURE 32.26 Diffraction patterns produced by a pair of point sources. The angular separation of the sources decreases until they can't be resolved.

Figure 32.24 shows that the angular separation between the diffraction peaks is equal to the angular separation between the sources themselves. Then the Rayleigh criterion is just met if the angular separation between the two sources is equal to the angular separation between a central peak and the first minimum. We found earlier that the first minimum in single-slit diffraction occurs at the angular position given by $\sin\theta = \lambda/a$, with a the slit width and with θ measured from the central peak. In most optical systems the wavelength is much less than the size of any apertures, so we can use the small-angle approximation $\sin\theta \simeq \theta$. Then the Rayleigh criterion—the condition that two sources be just resolvable—for single-slit diffraction becomes

$$\theta_{\text{min}} = \frac{\lambda}{a} \qquad \text{(Rayleigh criterion, slit)} \tag{32.11a}$$

Most optical systems have circular apertures rather than slits. The diffraction pattern from such an aperture is a series of concentric rings (Fig. 32.27). Mathematical analysis shows that the angular position of the first ring and therefore the minimum resolvable source separation for a circular aperture is

$$\theta_{\text{min}} = \frac{1.22\lambda}{D} \qquad \text{(Rayleigh criterion, circular aperture)} \tag{32.11b}$$

with D the aperture diameter.

FIGURE 32.27 3-D plot of intensity versus position in circular diffraction. The right-most image in Fig. 32.26 shows the corresponding diffraction pattern.

Equations 32.11 show that increasing the aperture size allows smaller angular differences to be resolved. In optical instrument design, that means larger mirrors or lenses. An alternative is to decrease the wavelength, which may or may not be an option depending on the source. In high-quality optical systems, diffraction is often the limiting factor preventing perfectly sharp image formation; such systems are said to be **diffraction limited**. For example, the diffraction limit sets a minimum size for objects resolvable with optical microscopes; that's why electron microscopes—with shorter effective wavelength—are used to image the smallest biological structures. Large ground-based telescopes are an exception to the diffraction limit; their image quality is limited by atmospheric turbulence, although this can be reduced with adaptive optics. From its vantage point above the atmosphere, the Hubble Space Telescope is the first large diffraction-limited astronomical telescope.

Astronomers circumvent the diffraction limit by combining data from several telescopes to produce, in effect, a single instrument with aperture equal to the telescope separation. Radio astronomers achieve exquisite resolution by combining telescopes on different continents; for optical astronomy the technique is limited to smaller separations. You can explore astronomical interferometry further in the Passage Problems.

GOT IT? 32.3 You're a biologist trying to resolve details of structures within a cell, but they look fuzzy even at the highest power of your microscope. Which of the following might help: (a) substituting an eyepiece with shorter focal length, as suggested by Equation 31.10; (b) putting a red filter over the white light source used to illuminate the microscope slide; or (c) putting a blue filter over the white light source?

EXAMPLE 32.5 The Diffraction Limit: Asteroid Alert

An asteroid 20×10^6 km away appears on a collision course with Earth. What's the minimum size for the asteroid that could be resolved with the 2.4-m-diameter diffraction-limited Hubble Space Telescope, using 550-nm reflected sunlight?

INTERPRET This is a problem about the diffraction limit with a circular aperture. We're after the minimum physical size for the asteroid at a given distance. We identify $D = 2.4$ m as the aperture size and $\lambda = 550$ nm as the wavelength of the light.

DEVELOP Equation 32.11b, $\theta_{\text{min}} = 1.22\lambda/D$, determines the diffraction limit, expressed as the minimum angular size that can be resolved. So our plan is to express the unknown physical size l in terms of angular size θ and then apply Equation 32.11b.

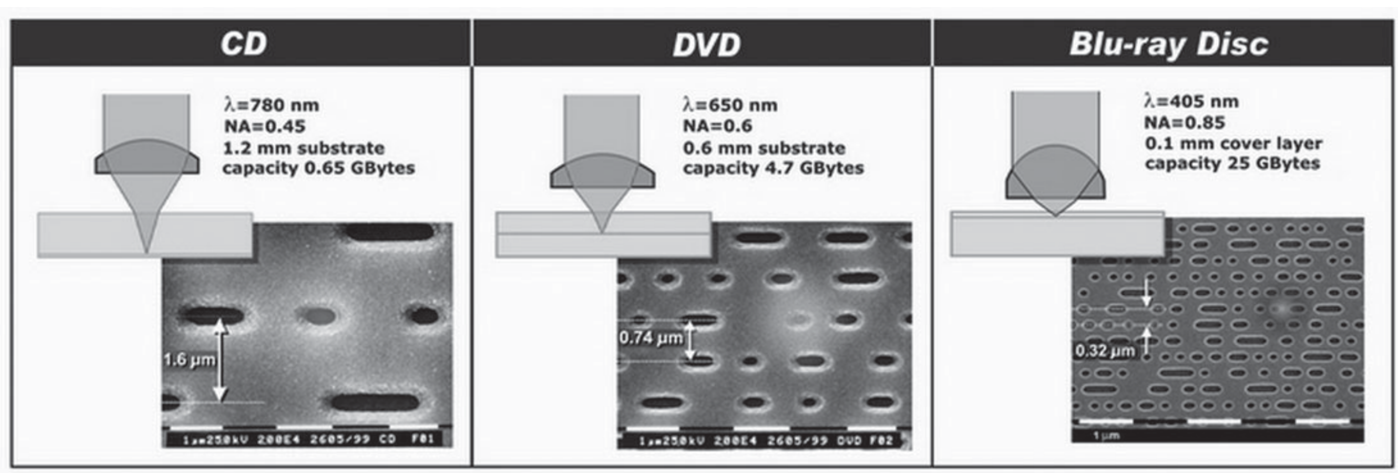
EVALUATE If we consider the opposite ends of the asteroid to be like the two peaks in Fig. 32.24, then the angular size for small θ is just l/L, where L is the distance to the asteroid. Then Equation 32.11b becomes

$$\frac{l}{L} = \frac{1.22\lambda}{D}$$

or $l = 1.22\lambda L/D = 5.6$ km using the numbers given.

ASSESS An object this size poses a grave danger, being comparable to the asteroid whose impact caused the extinction of the dinosaurs. If astronomers see only a fuzzy blur, then they'll have to wait until the asteroid is closer to resolve its physical size and assess the danger. ■

APPLICATION Movies on Disc: CD to DVD to Blu-ray



The Application earlier in this chapter described how a CD encodes information in pits 1.6 μm apart and as short as 0.83 μm. CDs are read with 780-nm infrared laser light. The pit size and spacing are chosen so diffraction effects at that wavelength don't cause the CD player's optical system to confuse adjacent pits. The result is a maximum capacity of about 650 MB (megabytes; 1 byte is 8 binary bits, with a bit the fundamental piece of binary information represented by a digital 1 or 0). This translates into 74 minutes of audio.

CDs were developed in the 1980s, when inexpensive semiconductor lasers were available only in the infrared. By the 1990s inexpensive visible-light lasers became available, and that enabled the development of DVDs (for "digital video disc" or "digital versatile disc"). Read with 635-nm or 650-nm red light, DVDs can use smaller pit size and spacing because of the lower diffraction limit. That, coupled with a two-layer structure and more sophisticated data-compression schemes, gives standard DVDs a capacity of about 4.7 GB—enough for 2 hours or more of video, depending on quality.

Despite their large capacity, DVDs aren't adequate for today's high-definition TV (HDTV). But improvements in laser technology give us a 405-nm violet laser that enables high-definition video discs. Again, the shorter wavelength and hence lower diffraction limit, along with other improvements, allow much more information to fit on a disc. The resulting Blu-ray technology stores 25 GB on a single-layer disc. That corresponds to 4.5 hours of high-definition video, or 12 hours of standard video. The figure compares CDs, DVDs, and Blu-ray discs.

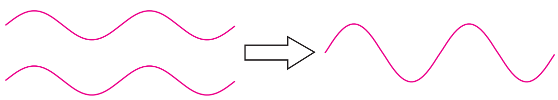
CHAPTER 32 SUMMARY

Big Picture

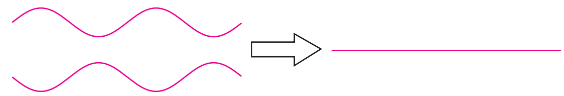
The big idea here is that—despite our use of the geometrical-optics approximation in the preceding chapters—light is indeed a wave and therefore exhibits the two related phenomena of **interference** and **diffraction**. These wave effects are important whenever light or any other wave interacts with objects whose size is comparable to or smaller than the wavelength.

Key Concepts and Equations

Constructive interference occurs when two waves combine in phase:



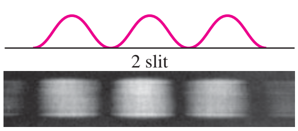
Destructive interference occurs when two waves combine 180° out of phase:

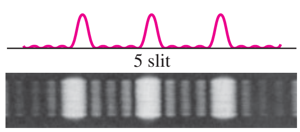


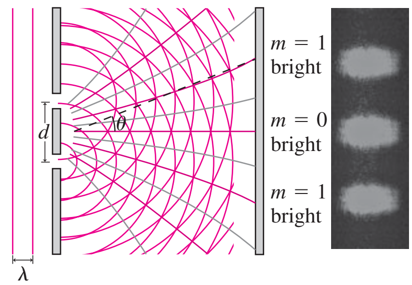
When light of wavelength λ passes through two or more narrow slits, the resulting interference shows maxima when

$$d \sin\theta = m\lambda$$

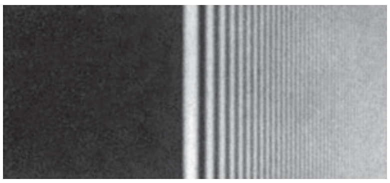
where m is an integer called the order. With multiple slits the maxima become stronger and narrower, but their position doesn't change.





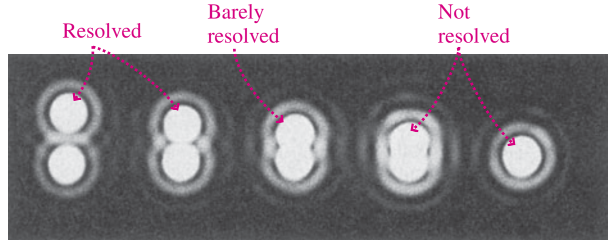


Diffraction occurs because, according to **Huygens' principle**, each point on a wavefront acts as a source of spherical waves, causing light to bend as it encounters sharp edges. Waves from different parts of a wavefront interfere to produce diffraction patterns.



The **diffraction limit** is a fundamental restriction on our ability to image small or distant objects. For a circular aperture of diameter D, the **Rayleigh criterion** gives the minimum angular separation that can be resolved with light of wavelength λ:

$$\theta_{\text{min}} = \frac{1.22\lambda}{D}$$



Applications

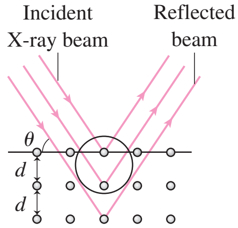
A **diffraction grating** consists of multiple slits or lines that result in constructive interference at different positions for different wavelengths. Diffraction gratings are used in spectrometers to disperse individual wavelengths. A grating's **resolving power**, the ratio of wavelength to the minimum resolvable difference in wavelengths, is given by

$$\frac{\lambda}{\Delta\lambda} = mN$$

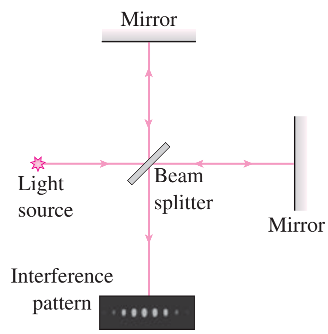
where N is the number of lines in the grating and m is the order of the dispersion.

X-ray diffraction uses regularly positioned atoms as a grating, and is a powerful technique for analyzing crystal and molecular structure. Maximum intensity occurs when

$$2d \sin\theta = m\lambda$$



The **Michelson interferometer** splits light into two beams that travel on perpendicular paths. They recombine, and the resulting interference allows precision measurements.



MP For homework assigned on MasteringPhysics, go to www.masteringphysics.com | BIO Biology and/or medical-related problems C Computer problems

For Thought and Discussion

1. A prism bends blue light more than red. Is the same true of a diffraction grating?
2. Why does an oil slick show colored bands?
3. Why does a soap bubble turn colorless just before it dries up and pops?
4. Why don't you see interference effects between the front and back of your eyeglasses?
5. You can hear around corners, but you can't see around corners. Why?
6. In deriving the intensity in double-slit interference, why can't you simply add the intensities from the two slits?
7. The primary maxima in multiple-slit interference are in the same angular positions as those in double-slit interference. Why, then, do diffraction gratings have thousands of slits instead of just two?
8. When the Moon passes in front of a star, the starlight intensity fluctuates before going to zero instead of dropping abruptly. Explain.
9. Sketch roughly the diffraction pattern you would expect for light passing through a square hole a few wavelengths wide.

Exercises and Problems

Exercises

Section 32.2 Double-Slit Interference

10. A double-slit system is used to measure the wavelength of light. The system has slit spacing $d = 15\ \mu\text{m}$ and slit-to-screen distance $L = 2.2$ m. If the $m = 1$ maximum in the interference pattern occurs 7.1 cm from screen center, what's the wavelength?
11. A double-slit experiment with $d = 0.025$ mm and $L = 75$ cm uses 550-nm light. Find the spacing between adjacent bright fringes.
12. A double-slit experiment has slit spacing 0.12 mm. (a) What should be the slit-to-screen distance L if the bright fringes are to be 5.0 mm apart when the slits are illuminated with 633-nm laser light? (b) What will be the fringe spacing with 480-nm light?
13. The interference pattern from two slits separated by 0.37 mm has bright fringes with angular spacing 0.065°. Find the light's wavelength.
14. The 546-nm green line of gaseous mercury falls on a double-slit apparatus. If the fifth dark fringe is at 0.113° from the centerline, what's the slit separation?

Section 32.3 Multiple-Slit Interference and Diffraction Gratings

15. In a five-slit system, how many minima lie between the zeroth-order and first-order maxima?
16. In a three-slit system, the first minimum occurs at angular position 5°. Where is the next maximum?
17. A five-slit system with 7.5-μm slit spacing is illuminated with 633-nm light. Find the angular positions of (a) the first two maxima and (b) the third and sixth minima.
18. Green light at 520 nm is diffracted by a grating with 3000 lines/cm. Through what angle is the light diffracted in (a) first and (b) fifth order?
19. Light is incident normally on a grating with 10,000 lines/cm. Find the maximum order in which (a) 450-nm and (b) 650-nm light will be visible.
20. Find the second-order angular separation of the two wavelengths in Example 32.2.

Section 32.4 Interferometry

21. Find the minimum thickness of a soap film $(n = 1.333)$ in which 550-nm light will undergo constructive interference.
22. Light of unknown wavelength shines on a precisely machined glass wedge with refractive index 1.52. The closest point to the apex of the wedge where reflection is enhanced occurs where the wedge is 98 nm thick. Find the wavelength.
23. Monochromatic light shines on a glass wedge with refractive index 1.65, and enhanced reflection occurs where the wedge is 450 nm thick. Find all possible values for the wavelength in the visible range.
24. White light shines on a 100-nm-thick sliver of fluorite $(n = 1.43)$. What wavelength is most strongly reflected?
25. For the soap film described in Conceptual Example 32.1's "Making the Connection," what portion of the film will appear dark when it's illuminated with white light?

Section 32.5 Huygens' Principle and Diffraction

26. For what ratio of slit width to wavelength will the first minima of a single-slit diffraction pattern occur at $\pm 90°$?
27. Light with wavelength 633 nm is incident on a 2.5-μm-wide slit. Find the angular width of the central peak in the diffraction pattern, taken as the angular separation between the first minima.
28. A beam of parallel rays from a 29-MHz Citizen's Band radio transmitter passes between two electrically conducting (hence opaque to radio waves) buildings located 45 m apart. What's the beam's angular width when it emerges from between the buildings?
29. Find the intensity as a fraction of the central peak intensity for the second secondary maximum in single-slit diffraction, assuming the peak lies midway between the second and third minima.

Section 32.6 The Diffraction Limit

30. Find the minimum angular separation resolvable with 633-nm laser light passing through a circular aperture of diameter 2.1 cm.
31. Find the minimum telescope aperture that could resolve an object with angular diameter 0.35 arcsecond, observed at 500-nm wavelength. (*Note*: 1 arcsec = 1/3600°.)
32. What's the longest wavelength of light you could use to resolve a structure with angular diameter 0.44 mrad, using a microscope with aperture 1.2 mm in diameter?
33. BIO In bright light, the human eye's pupil diameter is about 2 mm. If diffraction were the limiting factor, what's the eye's minimum angular resolution under these conditions, assuming 550-nm light?

Problems

34. Find the angular position θ of the second-order bright fringe in a double-slit system with 1.5-μm slit spacing if the light's wavelength is (a) 400 nm and (b) 700 nm.
35. A double-slit experiment has slit spacing 0.035 mm, slit-to-screen distance 1.5 m, and wavelength 500 nm. What's the phase difference between two waves arriving at a point 0.56 cm from the center line?
36. For a double-slit experiment with slit spacing 0.25 mm and wavelength 600 nm, at what angular position is the path difference a quarter wavelength?
37. A screen 1.0 m wide is 2.0 m from a pair of slits illuminated by 633-nm laser light, with the screen's center on the centerline of the slits. Find the highest-order bright fringe that will appear on the screen if the slit spacing is (a) 0.10 mm and (b) 10 μm.
38. A tube of glowing gas emits light at 550 nm and 400 nm. In a double-slit apparatus, what's the lowest-order 550-nm bright

fringe that will fall on a 400-nm dark fringe, and what are the fringes' corresponding orders?

39. On the screen of a multiple-slit system, the interference pattern shows bright maxima separated by 0.86° and seven minima between each bright maximum. (a) How many slits are there? (b) What's the slit separation if the incident light has wavelength 656.3 nm?
40. You're designing a spectrometer whose specifications call for a minimum of 5° separation between the red hydrogen-α line at 656 nm and the yellow sodium line at 589 nm when the two are observed in third order with a grating spectrometer. Available gratings have 2500 lines/cm, 3500 lines/cm, or 4500 lines/cm. What's the coarsest grating you can use?
41. For visible light with wavelengths from 400 nm to 700 nm, show that the first-order spectrum is the only one that doesn't overlap with the next higher order.
42. Find the total number of lines in a 2.5-cm-wide diffraction grating whose third-order spectrum puts the 656-nm hydrogen-α spectral line 37° from the central maximum.
43. What order is necessary to resolve 647.98-nm and 648.07-nm spectral lines using a 4500-line grating?
44. A thin film of toluene $(n = 1.49)$ floats on water. Find the minimum film thickness if the most strongly reflected light has wavelength 460 nm.
45. NASA asks you to assess the feasibility of a single-mirror space-based optical telescope that could resolve an Earth-size planet 5 light-years away. What do you conclude?
46. Echelle spectroscopy uses relatively coarse gratings in high order. Compare the resolving power of an 80-lines/mm echelle grating used in 12th order with a 600-lines/mm grating used in first order, assuming the two have the same width.
47. X-ray diffraction in potassium chloride (KCl) results in a first-order maximum when 97-pm-wavelength X rays graze the crystal plane at 8.5°. Find the spacing between crystal planes.
48. As a soap bubble with $n = 1.333$ evaporates and thins, reflected colors gradually disappear. What are (a) the bubble thickness just as the last vestige of color vanishes and (b) the last color seen?
49. An oil film with refractive index 1.25 floats on water. The film thickness varies from 0.80 μm to 2.1 μm. If 630-nm light is incident normally on the film, at how many locations will it undergo enhanced reflection?
50. You know that it's safe to microwave plastic containers, since their molecules don't respond significantly to 2.4-GHz microwaves. But since you've learned about thin-film interference, you worry about enhanced microwave intensity due to multiple reflections in plastic cookware. You calculate the minimum thickness for a plastic tray with refractive index 1.3 that will cause enhanced reflection of microwaves incident normal to the plate. Are your worries assuaged?
51. Two perfectly flat glass plates are separated at one end by a sheet of paper 0.065 mm thick. 550-nm light illuminates the plates from above, as shown in Fig. 32.28. How many bright bands appear to an observer looking down on the plates?

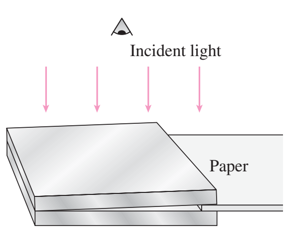


FIGURE 32.28 Problems 51, 52, and 64

52. An air wedge like that of Fig. 32.28 shows N bright bands when illuminated from above. Find an expression for the number of bands if the air is replaced by a liquid of refractive index n different from that of the glass.
53. A Michelson interferometer uses light from glowing hydrogen at 486.1 nm. As you move one mirror, 530 bright fringes pass a fixed point in the viewer. How far did the mirror move?
54. Find the wavelength of light used in a Michelson interferometer if 550 bright fringes go by a fixed point when the mirror moves 0.150 mm.
55. One arm of a Michelson interferometer is 42.5 cm long and is enclosed in a box that can be evacuated. The box initially contains air, which is gradually pumped out. In the process, 388 bright fringes pass a point in the viewer. If the interferometer uses light with wavelength 641.6 nm, what's the air's refractive index?
56. Your stereo is in a dead spot caused by direct reception from an FM radio station at 89.5 MHz interfering with the signal reflecting off a wall behind you. How much farther from the wall should you move so that the interference is fully constructive?
57. A proposed "star wars" antimissile laser is to focus 2.8-μm-wavelength infrared light to a 50-cm-diameter spot on a missile 2500 km distant. Find the minimum diameter for a concave mirror that can achieve this spot size, given the diffraction limit. (Your answer suggests one of many technical difficulties faced by antimissile defense systems.)
58. Suppose one of the 10-m-diameter Keck Telescopes in Hawaii is trained on San Francisco, 3400 km away. Assuming 550-nm light, and ignoring atmospheric distortion, would it be possible to read (a) newspaper headlines or (b) a billboard sign at this distance? (c) Repeat for the case of the Keck optical interferometer, formed from the two 10-m Keck Telescopes and several smaller ones, with a 50-m effective aperture.
59. A camera has an $f/1.4$ lens, meaning the ratio of focal length to lens diameter is 1.4. Find the smallest spot diameter (i.e., the diameter of the first diffraction minimum) to which this lens can focus parallel light with 580-nm wavelength.
60. The CIA wants your help identifying individual terrorists in a photo of a training camp taken from a spy satellite at 100-km altitude. You ask for details of the optical system used, but they're classified. However, they do tell you that the optics are diffraction limited and can resolve facial features as small as 5 cm. Assuming a typical optical wavelength of 550 nm, what do you conclude about the size of the mirror or lens in the satellite camera?
61. **BIO** While driving at night, your eyes' irises dilate to 3.1-mm diameter. If your vision were diffraction limited, what would be the greatest distance at which you could see as distinct the two headlights of an oncoming car, spaced 1.5 m apart? Take $\lambda = 550$ nm.
62. Under the best conditions, atmospheric turbulence limits ground-based telescopes' resolution to about 1 arcsecond (1/3600 of a degree). For what apertures is this limitation more severe than that of diffraction at 550 nm? (Your answer shows why large ground-based telescopes don't generally produce better images than small ones, although they do gather more light.)
63. **BIO** Your molecular biology lab studies proteins, and you're frustrated because your microscopes can't quite resolve crystallized proteins. A sales rep touts the advantages of an expensive microscope using 200-nm ultraviolet light, saying you'll be able to resolve structures less than half the size that's resolvable with your optical microscopes. Is the sales rep correct?
64. An air wedge like that of Fig. 32.28 displays 10,003 bright bands when illuminated from above. If the region between the plates is

then evacuated, the number of bands drops to 10,000. Find the refractive index of the air.

65. A thin-walled glass tube of length L containing a gas of unknown refractive index is placed in one arm of a Michelson interferometer using light of wavelength λ. The tube is then evacuated. During the process, m bright fringes pass a fixed point in the viewer. Find an expression for the refractive index of the gas.
66. Light is incident on a diffraction grating at angle α to the normal. Show that the condition for maximum light intensity becomes $d(\sin\theta \pm \sin\alpha) = m\lambda$.
67. An arrangement known as Lloyd's mirror (Fig. 32.29) allows interference between direct and reflected beams from the same source. Find an expression for the separation of bright fringes on the screen, given the distances d and D and the light's wavelength λ.

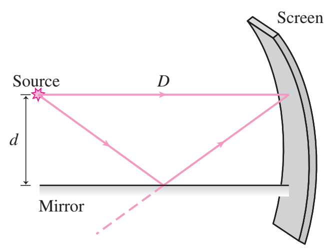


FIGURE 32.29 Lloyd's mirror (Problem 67)

68. The intensity in single-slit diffraction can be calculated by summing infinitely many electric-field amplitudes corresponding to waves from every infinitesimal part of the slit. (a) Referring to Fig. 32.20, show that the field from an element of width dy in the slit, a distance y from the bottom edge, is $dE = (E_p\,dy/\lambda)\sin(\omega t + \phi(y))$, where $\phi(y) = (2\pi y/a)\sin\theta$. (b) Integrate dE over the entire slit (y from 0 to a) and use trig identities from Appendix A to find the total amplitude, and from there show that the intensity is given by Equation 32.10.
69. You're on an international panel charged with allocating "real estate" for communications satellites in geosynchronous orbit. The panel needs to know how many satellites could fit in geosynchronous orbit without receivers on the ground picking up multiple signals. Assume all satellites broadcast at 12 GHz and that receiver dishes are 45 cm in diameter. Begin by calculating the angular size of the beam associated with such a receiver dish, defined as the full width of its central diffraction peak. Use your result to find the number of satellites allowed in geosynchronous orbit if each receiver dish is to "see" just one satellite. (*Hint:* Consult Example 8.3.)
70. You're investigating an oil spill for your state environmental protection agency. There's a thin film of oil on water, and you know its refractive index is $n_{oil} = 1.38$. You shine white light vertically on the oil, and use a spectrometer to determine that the most strongly reflected wavelength is 580 nm. Assuming first-order thin-film interference, what do you report for the thickness of the oil slick?

Passage Problems

Even the nearest stars are so distant that a single diffraction-limited telescope capable of imaging Earth-size planets orbiting them would be hopelessly large (see Problem 45). Astronomers get around this limitation using *interferometry* to combine data from several telescopes, producing an instrument that acts like a single telescope with aperture equal to the distance between the individual telescopes (Fig. 32.30). The technological challenge is to combine the signals with their relative phase intact; for this reason, interferometry has been used successfully for decades in radio astronomy but is just beginning to be used with optical telescopes.

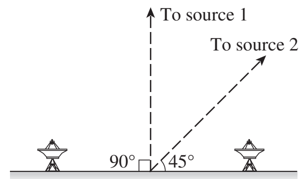


FIGURE 32.30 A two-dish interferometer used for radio astronomy (Passage Problems 71–74). Dashed lines show directions to sources in Problems 73 and 74.

71. If the separation of two telescopes comprising an interferometer is doubled, the angular separation between two sources just barely resolvable by the interferometer will
    a. not change.
    b. decrease by a factor of $1/\sqrt{2}$.
    c. halve.
    d. double.
72. If the separation of two telescopes comprising an interferometer is doubled, the instrument's light-collecting power will
    a. not change.
    b. increase by a factor of $\sqrt{2}$.
    c. double.
    d. quadruple.
73. If a point source is located directly above a two-telescope interferometer, on the perpendicular bisector of the line joining the telescopes (source 1 in Fig. 32.30), electromagnetic waves reaching the two will be
    a. in phase.
    b. out of phase by 45°.
    c. out of phase by 90°.
    d. you can't tell without further information
74. If a point source is located on a line at 45° to the line joining the two telescopes (source 2 in Fig. 32.30), electromagnetic waves reaching the two will be
    a. in phase.
    b. out of phase by 45°.
    c. out of phase by 90°.
    d. you can't tell without further information

Answers to Chapter Questions

Answer to Chapter Opening Question

The individual bits of information need to be stored in physical structures no smaller than the wavelength of light. The progression to greater information storage from CD to DVD to Blu-ray disc is largely a result of shorter-wavelength lasers being used to read the discs.

Answers to GOT IT? Questions

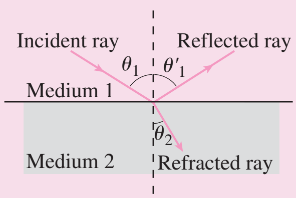
32.1. Closer together.
32.2. (a) No change; (b) intensity increases; (c) peaks become narrower.
32.3. (c) A blue filter means you're using a shorter wavelength, decreasing λ and therefore θ in Equations 32.11, and thus increasing the resolution.

PART FIVE SUMMARY

Optics

Optics is the study of light and its behavior. **Geometrical optics** is an approximation that holds when the objects with which light interacts are much larger than its wavelength. In this case, light generally travels in straight lines called **rays**. **Physical optics**, in contrast, treats light explicitly as a wave. Physical optics explains a host of phenomena that ultimately involve the interference of light waves.

When light rays are incident on an interface between two materials, they generally undergo **reflection** and, for transparent materials, **refraction**. The angles of incidence and reflection are equal. **Snell's law** relates the angles of incidence and refraction:

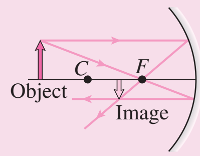


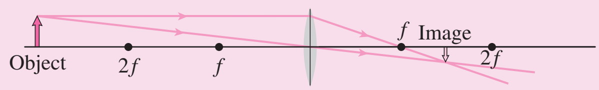
Reflection: $\theta'_1 = \theta_1$
Refraction (Snell's law):
$n_1 \sin\theta_1 = n_2 \sin\theta_2$

The **index of refraction** n relates light's speed in a medium to its speed c in vacuum:

$$n = \frac{c}{v}$$

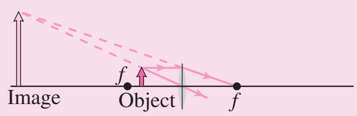
Lenses and curved mirrors use refraction and reflection, respectively, to form images.

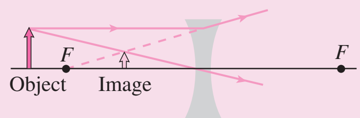




In both cases the object distance s, image distance s', and **focal length** f are related by $\frac{1}{s} + \frac{1}{s'} = \frac{1}{f}$.

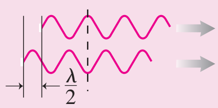
With **real images**, shown in both figures above, light actually comes from the image. With **virtual images**, shown in both figures below, light only *appears* to come from the image:



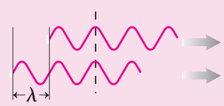


The wave nature of light becomes important when light interacts with objects comparable in size to its wavelength, or when light travels different paths and recombines to produce interference.

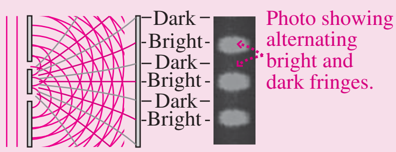
Destructive interference occurs when waves are out of step by an odd-integer multiple of a half-wavelength.



Constructive interference occurs when waves are out of step by an integer multiple of the wavelength.

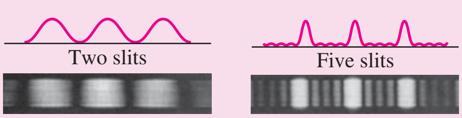


A system consisting of two narrow slits produces a pattern of **interference fringes** resulting from alternating regions of constructive and destructive interference:



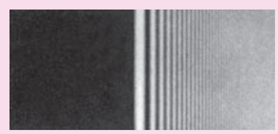
$$d \sin\theta = m\lambda \qquad \text{(bright fringes)}$$
$$d \sin\theta = \left(m + \tfrac{1}{2}\right)\lambda \qquad \text{(dark fringes)}$$

With multiple slits the bright fringes become narrower and brighter:

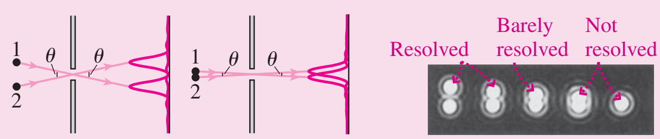


A multiple-slit system constitutes a **diffraction grating** and is used to separate different wavelengths in **spectroscopy**.

Huygens' principle explains the propagation of waves by stating that each part of a wavefront acts as a source of circular waves that spread out and interfere to propagate the wave. When light passes through small apertures or by sharp edges, Huygens' principle shows that the light **diffracts**, bending and producing interference fringes as waves from different points interfere.



Diffraction fundamentally limits our ability to resolve small objects or to see closely spaced but distant objects as separate.



For a circular aperture of diameter d (such as a telescope with d being its mirror diameter), the **diffraction limit** gives the smallest angular separation that can be resolved at a given wavelength λ:

$$\theta_{\min} = \frac{1.22\lambda}{d}$$

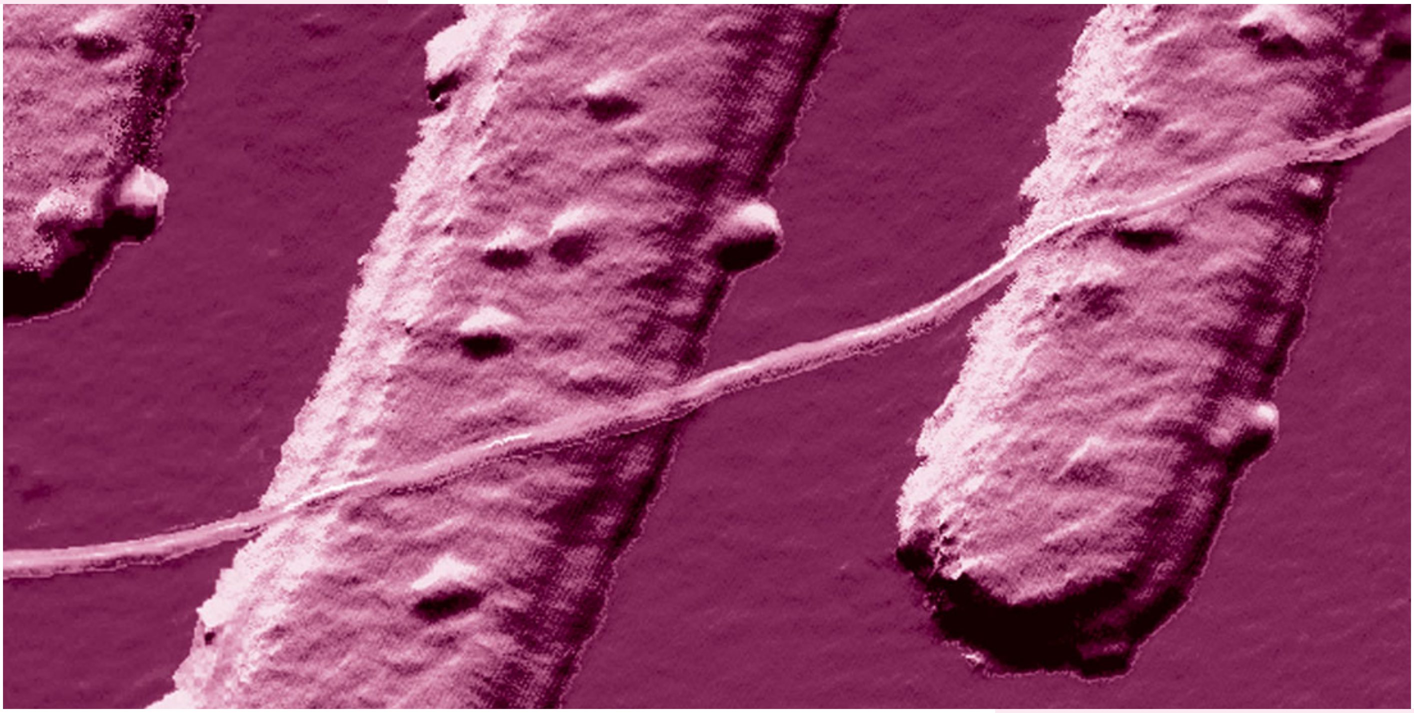
Part Five Challenge Problem

A double-slit system consists of two slits each of width a, with separation d between the slit centers $(d > a)$. Light of intensity S_0 and wavelength λ is incident on the system, perpendicular to the plane containing the slits. Find an expression for the outgoing intensity as a function of angular position θ, taking into account both the slit width and the separation. Plot your result for the case $d = 4a$, and compare with Fig. 32.22.

PART SIX

OVERVIEW

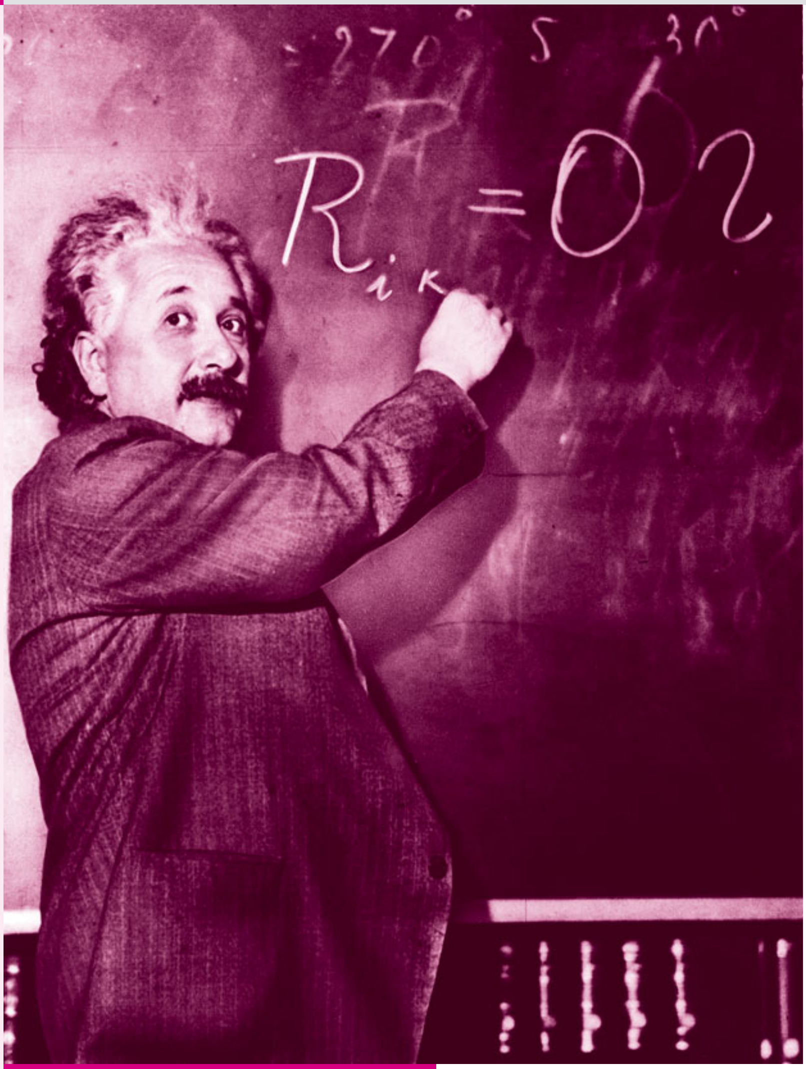
Modern Physics



The world's smallest electrical wire, a carbon nanotube, is only 10 atoms across. In this image made with an atomic force microscope, the nanotube wire runs across a backdrop of platinum electrodes. Our understanding of physics at the atomic and molecular level lets us construct an increasing variety of practical nanoscale devices.

What are the fundamental particles of matter? What holds them together to make protons, neutrons, nuclei, atoms, molecules, and solids? Is nature fundamentally predictable, or does uncertainty rule in the microscopic world? At the other extreme, how big is the universe? How did it begin, and how will it end? All these are questions for relativity and quantum physics—collectively called "modern physics" because they were developed after the turn of the 20th century. In Part 6 we give a brief account of Einstein's theory of relativity, followed by a glimpse at quantum physics and its applications. We end with an overview of the latest developments in fundamental physics, from the nature of elementary particles to surprising new findings about the origin and composition of the universe.

33 Relativity



Behind Einstein's theories is a profoundly simple principle that can be stated in a single statement. What is it?

New Concepts, New Skills

By the end of this chapter you should be able to

- Recount the historical progression from 19th-century physics to relativity (33.1, 33.2).
- State the principle of special relativity (33.3).
- Work quantitative problems involving time dilation and length contraction (33.4, 33.5).
- Explain why events simultaneous in one frame of reference need not be simultaneous in another (33.5).
- Use the Lorentz transformations to find the space and time coordinates of events in different reference frames (33.6).
- Discredit the common misconception that Einstein's theories imply that "everything is relative" (33.6).
- Describe quantitatively the relations among mass, energy, and momentum in relativity (33.7).
- Explain qualitatively how relativity connects electricity and magnetism (33.8).
- Explain qualitatively why a generalization of relativity leads to Einstein's theory of gravity (33.9).

Connecting Your Knowledge

- This chapter builds on your existing knowledge of physics, especially electromagnetism and electromagnetic waves (Chapter 29).
- We'll elaborate on the concept of reference frames, especially inertial reference frames (3.3, 4.2).
- Be prepared to look at old ideas in new ways: Relativity alters the meanings of and relations among physics concepts like mass, momentum, and energy, as well as the underlying concepts of space and time.

Maxwell's electromagnetic theory was a crowning achievement of 19th-century physics, providing an understanding of the nature of light and enabling a host of practical technologies. At the same time, Maxwell's electromagnetism led to baffling questions and contradictions that shook the roots of physical understanding and even of common sense.

The theory of relativity resolved these contradictions. It radically altered our fundamental understanding of the physical world, and its influence spilled over into all areas of human thought. Relativity stands as a monument to human intellect and imagination, and it reveals a universe far richer than earlier physicists could conceive of. We'll approach relativity historically, building on our understanding of electromagnetism. That way you'll get a sense of the questions that electromagnetism posed to 19th-century physicists and of how Einstein's answer to these questions was at once profoundly bold and sweepingly simple.

33.1 Speed *c* Relative to What?

Maxwell's equations show that electromagnetic waves in vacuum propagate with speed c. Speed c relative to what? In Chapter 14 we found the speed of waves on a string; clearly the speed in that case was relative to the string. Similarly, the 340-m/s speed of sound in air is clearly relative to the air. If you move through the air, the speed of sound *relative to you* won't be the same. In these and other cases of mechanical waves, it's clear what the wave speed means: It's the speed relative to the medium in which the wave is a disturbance.

The Ether Concept

What about light? Nineteenth-century physicists, their worldview built on the highly successful mechanical paradigm of Newtonian physics, supposed that light waves were like mechanical waves in requiring a medium. They postulated a tenuous substance called the **ether** that permeated the entire universe, allowing light from distant stars to reach us. Electric and magnetic fields were stresses in the ether, and electromagnetic waves were propagating disturbances moving through the ether at speed c.

The ether had to have some unusual properties. It must offer no resistance to material bodies, or the planets would lose energy and spiral into the Sun. It must be very stiff, to account for the high speed of light. And it had to be more jelly-like than fluid, to support transverse electromagnetic waves. These properties make ether a rather improbable substance, but to 19th-century physicists the ether was essential in understanding electromagnetic waves.

The speed of light c follows from Maxwell's equations. But in the 19th-century view, light has speed c only for an observer at rest with respect to the ether. Therefore, Maxwell's equations could be correct only in the ether frame of reference. This put electromagnetism in a rather different position from mechanics. In mechanics, the concept of absolute motion is meaningless. You can eat your dinner, toss a ball, or do any mechanical experiment as well on an airplane moving steadily at 1000 km/h as you can when the plane is at rest on the ground. This is the principle of **Galilean relativity**, which states that the laws of mechanics are valid in all inertial reference frames—that is, frames of reference in uniform motion (Section 4.2). But the laws of electromagnetism seemed valid only in the ether's reference frame because only in this frame was the prediction of electromagnetic waves moving at speed c correct.

So for 19th-century physicists, the laws for one branch of physics (mechanics) seemed to work in all inertial frames, while those of another branch (electromagnetism) could not. Despite this dichotomy, physicists had great faith in mechanical models and in the ether concept, for without the ether the question "speed c relative to what?" seemed impossible to answer. Thus the late 19th century saw a flurry of experiments to detect the ether. Ultimately they failed, paving the way for the new worldview of relativity that, in Einstein's own words, "arose from necessity, from serious and deep contradictions in the old theory from which there seemed no escape."

33.2 Matter, Motion, and the Ether

It was natural for 19th-century physicists to ask about Earth's motion relative to the ether. If Earth is moving through the ether, then the speed of light should be different in different directions. On the other hand, Earth might be at rest relative to the ether. Because other planets, stars, and galaxies move with respect to Earth, it's hard to imagine that the ether is everywhere fixed with respect to Earth alone: This violates the Copernican view that Earth doesn't occupy a privileged spot in the universe. But maybe Earth drags with it the ether in its immediate vicinity. If this "ether drag" occurs, then the speed of light must be independent of direction, but if there's no ether drag, then the speed of light measured on Earth must depend on direction. Through observation and experiment, 19th-century physicists sought to resolve the question of Earth's motion through the ether.

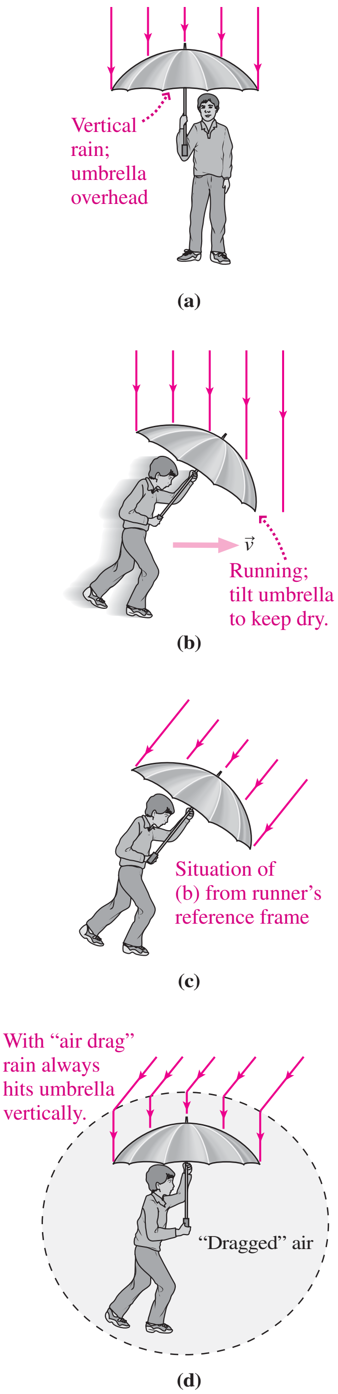


FIGURE 33.1 A rain/umbrella analogy for aberration of starlight.

Aberration of Starlight

Imagine standing in a rainstorm with rain falling vertically. To keep dry, you hold your umbrella with its shaft straight up, as shown in Fig. 33.1*a*. But if you run, as in Fig. 33.1*b*, you'll keep driest if you tilt your umbrella forward. Why? Because the direction of rainfall *relative to you* is at an angle, as shown in Fig. 33.1*c*. This assumes you don't drag with you a large volume of air. If such an "air drag" occurred, raindrops entering the region around you would be accelerated quickly in the horizontal direction by the air moving with you, so they would now fall vertically relative to you, as in Fig. 33.1*d*. No matter which way you ran, as long as you dragged air with you, you would point your umbrella vertically upward to stay dry.

This umbrella example is exactly analogous to the observation of light from stars, with the rain being starlight and the umbrella a telescope. If Earth doesn't drag ether, then the direction from which starlight comes will depend on Earth's motion relative to the ether. But if "ether drag" occurs in analogy with Fig. 33.1*d*, then light from a particular star will always come from the same direction.

In fact we do observe a tiny change in the direction of starlight. As Earth swings around in its orbit, we must first point a telescope one way to see a particular star. Then, six months later, Earth's orbital motion is in exactly the opposite direction, and we must point the telescope in a slightly different direction. This phenomenon is called **aberration of starlight** and shows that *Earth does not drag the ether*.

The Michelson–Morley Experiment

If we reject the pre-Copernican notion that Earth alone is at rest relative to the ether, then aberration of starlight forces us to conclude that Earth moves through the ether. Furthermore, the relative velocity of the motion must change throughout the year as Earth orbits the Sun.

In 1881–1887, American scientists Albert A. Michelson and Edward W. Morley attempted to determine Earth's velocity relative to the ether. They used Michelson's interferometer (Fig. 33.2), whose operation we described in Chapter 32. Recall that the interferometer produces a pattern of interference fringes that shifts if the round-trip travel time for light on one of its two perpendicular arms changes. The interference pattern reflects, among other things, possible differences in travel times that arise from differences in the speed of light in different directions—differences that should result from Earth's motion through the ether. Rotating the apparatus through 90° would interchange the directions of the arms and should therefore shift the interference pattern.

Now suppose Earth moves at speed v relative to the ether. Then to an observer on Earth, there's an "ether wind" blowing past Earth. Suppose the Michelson–Morley apparatus is oriented with one light path parallel to the wind and the other perpendicular. Consider a light beam moving the distance L at right angles to the wind. The beam must be aimed slightly upwind so that it will actually move perpendicular to the wind. The light moves in this direction at speed c relative to the ether, but the ether wind sweeps it back so its path

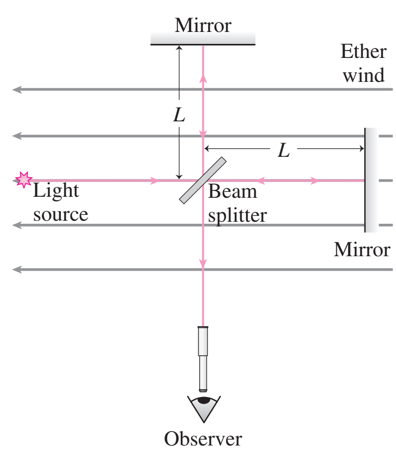


FIGURE 33.2 Simplified diagram of the Michelson–Morley experiment. Ether wind should result in a longer travel time for light on the horizontal arm.

in the Michelson–Morley apparatus is at right angles to the wind. From Fig. 33.3, we see that its speed relative to the apparatus is $u = \sqrt{c^2 - v^2}$, so the round-trip travel time is

$$t_{\text{perpendicular}} = \frac{2L}{u} = \frac{2L}{\sqrt{c^2 - v^2}} \tag{33.1}$$

Light sent a distance L "upstream"—against the ether wind—travels at speed c relative to the ether but at speed $c - v$ relative to Earth. It therefore takes time $t_{\text{upstream}} = L/(c - v)$. Returning, the light moves at $c + v$ relative to Earth, so $t_{\text{downstream}} = L/(c + v)$. The round-trip time parallel to the ether wind is then

$$t_{\text{parallel}} = \frac{L}{c - v} + \frac{L}{c + v} = \frac{2cL}{c^2 - v^2} \tag{33.2}$$

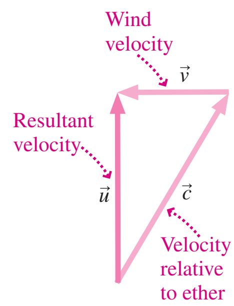


FIGURE 33.3 Vector diagram for light moving at right angles to an ether wind.

The two round-trip travel times differ, with the trip parallel to the ether wind always taking longer (see Exercises 13, 14, and 27). Light on the parallel trip slows when it moves against the ether wind, then speeds up when it moves with the wind. But slowing always dominates because the light spends more time moving against the wind than with it.

The Michelson–Morley experiment of 1887 was sensitive enough to detect differences in the speed of light an order of magnitude smaller than Earth's orbital speed. The experiment was repeated with the apparatus oriented in different directions, and at different times throughout the year, and the same simple but striking result always emerged: There was never any difference in the travel times for the two light beams. In terms of the ether concept, the Michelson–Morley experiment showed that *Earth does not move relative to the ether*.

A Contradiction in Physics

Aberration of starlight shows that Earth doesn't drag ether with it. Earth must therefore move relative to the ether. But the Michelson–Morley experiment shows that it doesn't. This contradiction is a deep one, rooted in the fundamental laws of electromagnetism and in the analogy between mechanical waves and electromagnetic waves. The contradiction arises directly in trying to answer the simple question: With respect to what does light move at speed c?

Physicists at the end of the 19th century made ingenious attempts to resolve the dilemma of light and the ether, but their explanations either were inconsistent with experiment or lacked sound conceptual bases.

33.3 Special Relativity

In 1905, at the age of 26, Albert Einstein (Fig. 33.4) presented his **special theory of relativity**, which resolved the dilemma but altered the very foundation of physical thought. Einstein declared simply that the ether is a fiction. But then with respect to what does light move at speed c? With respect, Einstein declared, to anyone who cares to observe it. This statement is at once simple, radical, and conservative. Simple, because its meaning is clear and obvious. Anyone who measures the speed of light in vacuum will get the value $c = 3.0 \times 10^8$ m/s. Radical, because it alters our commonsense notions of space and time. Conservative, because it asserts for electromagnetism what had long been true in mechanics: that the laws of physics don't depend on the motion of the observer. Einstein summarized his new ideas in the **principle of relativity**, which is expressed in this simple sentence:

> The laws of physics are the same in all inertial reference frames.

FIGURE 33.4 In 1905, when he formulated special relativity, Einstein was a 26-year-old father.

Recall that inertial frames are unaccelerated—that is, frames in which the laws of *mechanics* were already valid. Einstein's statement encompasses *all* laws of physics, including mechanics and electromagnetism. The prediction that electromagnetic waves move at speed c must, then, be a universal prediction that holds in *all* inertial reference frames. The *special* theory of relativity is special because it's valid only for the special case of inertial frames. Later we'll discuss the *general* theory of relativity, which removes this restriction.

Einstein's relativity explains the result of the Michelson–Morley experiment: No matter what Earth's speed is relative to anything, an observer on Earth should measure the same speed for light in all directions. But at the same time, relativity flagrantly violates our commonsense notions of space and time. We'll see just how in Sections 33.4–33.6.

33.4 Space and Time in Relativity

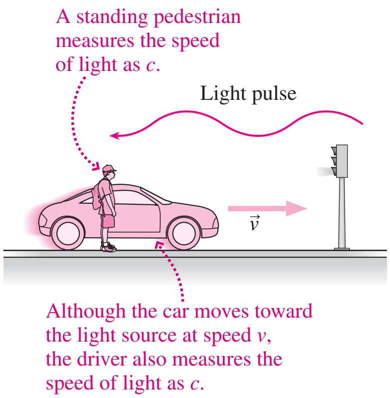


FIGURE 33.5 Both driver and pedestrian measure the same speed c for light, even though they're in relative motion.

A pedestrian stands by the roadside as a car drives by (Fig. 33.5). Driver and pedestrian each measure the speed of light from a blinking traffic signal. The theory of relativity says they'll both get the same value, $c = 3.0\times10^8$ m/s, even though the car is moving toward the light source. How is that possible? Consider how each observer might make the measurement. Each has a meter stick and an accurate stopwatch. Suppose a light flash passes the front end of each meter stick just as they coincide. Each observer measures the time for the flash to traverse the 1-m stick and calculates *speed* = *distance/time*. They get the same answer—even though common sense suggests that the light should pass the far end of the "moving" meter stick sooner.

How can this be? Maybe the car's motion affects the driver's stopwatch, making it inaccurate. But no; this suggestion violates relativity's assertion that *all* uniformly moving reference frames are equally good vantage points for doing physics. There can't be anything special about the "moving" reference frame; in fact, it's meaningless to talk about the car as "moving" and the pedestrian "at rest." This is the point of relativity: The concept of absolute motion is meaningless.

The only way out, consistent with relativity, is to let go of absolute space and time. Our two observers' instruments are measuring different quantities that depend on their reference frame—namely, the distance and travel time for the light flash. Those quantities differ in just the right way to make the speed of light come out the same for both observers. This is certainly not what common sense tells us about space and time. But in relativity it's the laws of physics, not measures of space and time, that must be the same for all. Keep in mind the principle of relativity, and you'll see how the rest of special relativity's remarkable consequences follow.

Time Dilation

Figure 33.6*a* shows a "light box," consisting of a box of length L with a light source at one end and a mirror at the other. A light flash leaves the source, reflects off the mirror, and returns to the source. We want the time between two events: the emission of the flash and its return to the source. An **event** is an occurrence specified by giving its position and its time.

For concreteness, we'll imagine that the light box is in a spaceship moving past Earth at a uniform velocity. But don't think there's something special about space or spaceships. The whole point of relativity is that all inertial frames are equivalent places for doing physics,

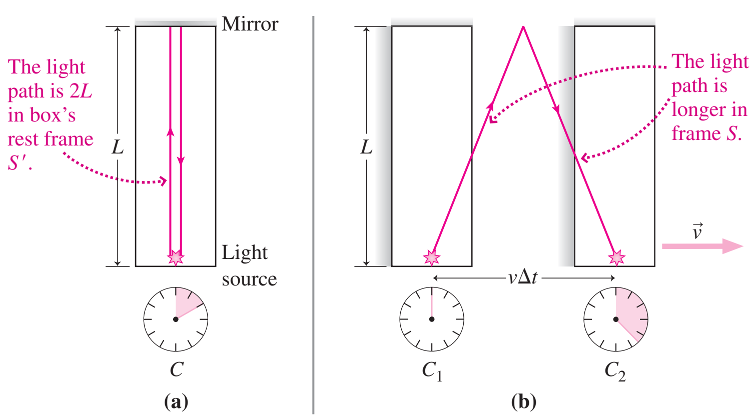


FIGURE 33.6 A "light box" to explore time dilation shown (a) in a reference frame S' at rest with respect to the box and (b) in a frame S where the box is moving to the right.

and our spaceship is just one inertial reference frame. We'll call that frame S'. There's also an accurate clock, C, in the spaceship, and C reads zero just as the light flash is emitted.

In Fig. 33.6*a* we consider the light-box experiment viewed in the spaceship's reference frame. Since the light box is at rest in this frame, the light travels a round-trip distance $2L$ from source to mirror and back, giving a round-trip travel time of $\Delta t' = 2L/c$. This is the time read on the spaceship's clock C.

Now consider the situation as viewed in Earth's reference frame, which we'll call S. In this frame, spaceship and light clock are moving to the right with speed v, as shown in Fig. 33.6*b*. Suppose there are two clocks in Earth's frame, positioned so the light box passes clock C_1 just as the flash is emitted, and passes C_2 just as the flash returns to the source. The clocks are synchronized, and C_1 reads zero just as the light flash is emitted. We want to know C_2's reading at the instant the light box passes it and the flash returns to its source; that will be the time, Δt, between the flash emission and return as measured in Earth's frame S.

Figure 33.6*b* shows that the box moves to the right a distance $v\,\Delta t$ in the time between emission and return of the light flash. Meanwhile the light takes a diagonal path up to the mirror of the moving box and then back down. The path length is twice the diagonal from source to mirror or, by the Pythagorean theorem, $2\sqrt{L^2 + (v\,\Delta t/2)^2}$. The time for light to go this distance is the distance divided by the speed of light, or $\Delta t = 2\sqrt{L^2 + (v\,\Delta t/2)^2}/c$. We explicitly used relativity here, assuming the speed of light remained c in Earth's frame. If we didn't believe relativity, we would have vectorially added light's velocity $\vec{c}$ and the box's velocity $\vec{v}$. But that would make the spaceship's frame the only one in which the speed of light was c—in violation of the relativity principle.

The unknown Δt appears on both sides of our expression; multiplying through by c and squaring give

$$c^2(\Delta t)^2 = 4L^2 + v^2(\Delta t)^2$$

We then solve for $(\Delta t)^2$ to get

$$(\Delta t)^2 = \frac{4L^2}{c^2 - v^2} = \frac{4L^2}{c^2}\left(\frac{1}{1 - v^2/c^2}\right)$$

Taking the square root of both sides, and noting that $2L/c$ is just the time $\Delta t'$ measured in the frame S' at rest with respect to the box, we have $\Delta t = \Delta t'/\sqrt{1 - v^2/c^2}$ or

$$\Delta t' = \Delta t\sqrt{1 - v^2/c^2} \qquad \text{(time dilation)} \tag{33.3}$$

Equation 33.3 describes the phenomenon of **time dilation**, in which the time between two events is shortest in a frame of reference in which the two events occur at the same place. In our example, the events are the emission and return of the light flash, and they occur at the same place in the spaceship frame S'—namely, at the bottom of the box. They don't occur at the same place in the Earth frame S because the box moves relative to Earth, so the bottom of the box isn't in the same place when the light flash returns as it was when the flash was emitted. Thus $\Delta t'$—the time interval measured in the spaceship frame S', where the events occur at the same place—is shorter than Δt, as you can see from Equation 33.3 and as illustrated in Fig. 33.7.

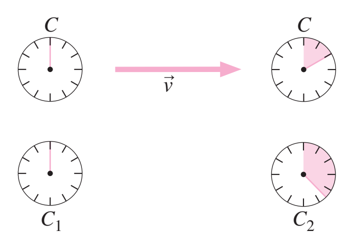


FIGURE 33.7 Clock C moves between clocks C_1 and C_2, which are at rest relative to each other and synchronized in their rest frame. C measures a shorter elapsed time.

The shorter time $\Delta t'$ measured by a single clock present at two events is called a **proper time**. Here "proper" doesn't mean "correct" or "right"—that would violate relativity, since a time measurement in any inertial frame has equal claim to being valid. Rather, "proper" is used in the sense of "proprietary" in that proper time is the time that *belongs* to this one particular clock. In Earth frame S there's no single clock present at both the emission and return of the light flash, so the time measurement in this frame doesn't "belong" to any one clock.

Time dilation is sometimes characterized by saying "moving clocks run slow," but this statement violates the spirit of relativity because it suggests that some frames are "really" moving and others aren't. The whole point of relativity is that all inertial frames are equally good for describing physical reality, so none can claim to be "at rest" while others are "moving." What the statement "moving clocks run slow" is trying to convey is what we've just seen: The time interval between two events is shortest in a reference frame where the two occur at the same place. There's no significance whatever to our putting the light box in a "moving" spaceship and comparing ship time with Earth time. We could equally well have put the light box and its clock on Earth, and two separate clocks in the ship. Then Earth with its light box would be moving past the ship at speed v—so the Earth clock would measure the time $\Delta t'$ in Equation 33.3, and the two clocks in the ship frame would measure the longer time Δt. (That may sound like a contradiction, but it can't be because *there's nothing special about any inertial frame, including Earth's*. We'll return to this point shortly.)

We used a light box to illustrate time dilation. But time dilation isn't something that happens only when we use light to determine time intervals. It's something that happens to *time itself*. Take away the light box in Fig. 33.6, and the clocks will show the same discrepancy. Don't look for a physical mechanism that slows things down. All manifestations of time—the oscillations of the quartz crystal in a digital watch, the swing of a pendulum clock, the period of vibration of atoms in an atomic clock, biological rhythms, and human lifetimes—are affected in the same way.

EXAMPLE 33.1 Calculating Time Dilation: Star Trek

A spaceship leaves Earth on a one-way star trip that earthbound observers judge will take 25 years. If the ship travels at $0.95c$ relative to Earth, how long will the trip take as judged by observers in the ship?

INTERPRET This is a problem about time dilation. Since the events of departure from Earth and arrival at the star occur at the same place in the ship's reference frame, we identify the ship time as the proper time $\Delta t'$ and the Earth time as Δt in our discussion of time dilation.

DEVELOP Equation 33.3, $\Delta t' = \Delta t\sqrt{1 - v^2/c^2}$, relates the two times. We're given the Earth time Δt, so we can use this equation to find $\Delta t'$.

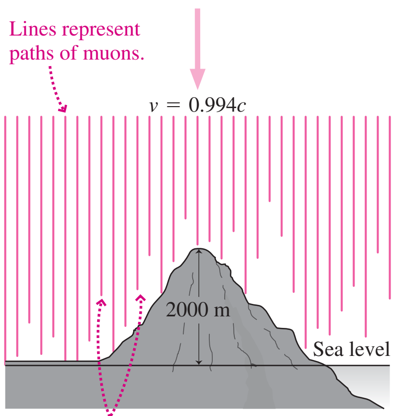
EVALUATE With $v = 0.95c$, $v/c = 0.95$ and the quantity v^2/c^2 in Equation 33.3 becomes 0.95^2. Then

$$\Delta t' = \Delta t\sqrt{1 - v^2/c^2} = (25\text{ y})\sqrt{1 - 0.95^2} = 7.8\text{ y}$$

ASSESS This time is considerably shorter than 25 years, confirming our statement that the time between events is shortest as measured in a reference frame where the events occur at the same place. We'll soon explore what happens if the ship turns around and returns to Earth. ■

We don't notice time dilation in our everyday lives because the factor v^2/c^2 is so small for even our fastest motion relative to Earth. Even in a jet airplane, the time difference amounts to a few milliseconds per century. This illustrates the important point that any results from relativity should agree with commonsense Newtonian physics when relative velocities are small compared with the speed of light. Since our intuition and common sense are built on experience at low relative velocities, it's not surprising that effects at high relative velocities seem counter to our common sense.

APPLICATION Mountains and Muons



Time dilation is obvious in experiments with subatomic particles moving, relative to us, at speeds near c. In a classic experiment, the "clocks" are the lifetimes of particles called muons, which are created by the interaction of cosmic rays with Earth's upper atmosphere and subsequently decay. The experiment consists in counting the number of muons incident each hour on the top of Mt. Washington in New Hampshire, about 2000 m above sea level. The measurement is then repeated at sea level. The figure shows the situation in Earth's reference frame.

Using a detector that records only those muons moving at about $0.994c$ at the mountaintop altitude, the experiment shows that an average of about 560 muons with this speed are incident on the mountaintop each hour. If the mountain weren't there, the muons would travel from the mountaintop altitude to sea level in a time given by

$$\Delta t = \frac{2000\text{ m}}{(0.994)(3.0\times 10^8\text{ m/s})} = 6.7\ \mu\text{s}$$

The muon's decay rate is such that one should expect only about 25 of the original 560 muons to remain after a 6.7-μs interval, so that's approximately the number we might expect to detect each hour at sea level. However, that 6.7-μs interval is measured in Earth's reference frame—not the muons'. In the muons' frame, time dilation should reduce that interval to

$$\Delta t' = (6.7\ \mu\text{s})\sqrt{1 - 0.994^2} = 0.73\ \mu\text{s}$$

The muons' decay is determined by *their* measure of time, and their decay rate is such that we should expect 414 muons to survive for 0.73 μs.

So what happens? Observers count just over 400 muons per hour at sea level. This is no subtle effect. The difference between 25 and 414 is dramatic. At $0.994c$, the nonrelativistic description is hopelessly inadequate, and time dilation is obvious.

The Twin Paradox

Time dilation lets us travel into the future! The famous "twin paradox" shows how. One twin boards a spaceship for a journey to a distant star; the other stays on Earth. There are clocks at both Earth and star, like clocks C_1 and C_2 in Fig. 33.7. There's a clock on the spaceship, like C in Fig. 33.7. Ship clock and Earth clock read the same time as the ship departs (Fig. 33.8*a*), but when it arrives at the star, time dilation means the ship clock reads less time than the star clock (Fig. 33.8*b*). Now the ship turns around and returns home. Again, the situation is just like Fig. 33.7 or our light box of Fig. 33.6, so less time elapses on the ship (Fig. 33.8*c*). The traveling twin arrives home younger than her earthbound brother! Depending on how far and how fast she goes, that age difference can be arbitrarily large. But this is a one-way trip to the future. If the traveling twin doesn't like what she finds in the future, there's no going back.

EXAMPLE 33.2 Time Dilation: The Twin Paradox

Earth and a star are 20 light years (ly) apart, measured in a frame at rest with respect to Earth and star. Twin A boards a spaceship, travels at $0.80c$ to the star, and then returns immediately to Earth at $0.80c$. Determine the round-trip travel times in Earth and ship reference frames.

INTERPRET This problem involves time dilation, here applied to the two separate legs of the round-trip journey. We identify the ship clock as C in Fig. 33.7, and the ship time as $\Delta t'$.

DEVELOP Earth–star time for the one-way journey follows from *distance* = *speed* × *time*, so that's how we'll find Δt. Then we can apply Equation 33.3, $\Delta t' = \Delta t\sqrt{1 - v^2/c^2}$, to get the ship time $\Delta t'$. We'll double both to get the round-trip times.

EVALUATE At $0.80c$, the time to go 20 ly is $\Delta t = (20\ \text{ly})/(0.80\ \text{ly/y}) = 25\ \text{y}$. Equation 33.3 then gives

$$\Delta t' = (25\ \text{y})\sqrt{1 - 0.80^2} = 15\ \text{y}$$

Doubling these values gives round-trip times of 50 years and 30 years for the Earth and ship, respectively.

ASSESS The traveling twin returns younger, by 20 years! We've marked the various times on the clocks in Fig. 33.8.

✓TIP Years, Light Years, and the Speed of Light

A light year (ly) is the distance light travels in one year. By definition, therefore, the speed of light is 1 ly/y. It's often easiest in relativity to work in units where the speed of light is 1, whether those units be light years and years, light seconds and seconds, or whatever.

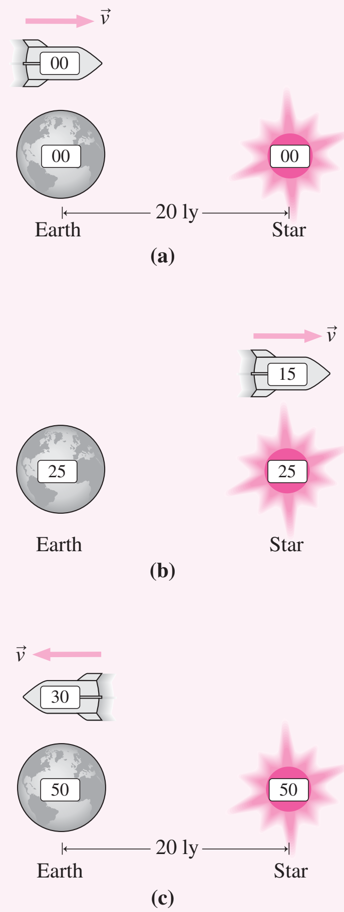


FIGURE 33.8 The twin's journey, drawn from the viewpoint of the Earth–star reference frame. Clock readings are in years, corresponding to Example 33.2.

Here's the seeming paradox: From the ship's viewpoint, it looks like Earth recedes, turns around, and returns. So why isn't the Earth twin younger? The answer lies in what's *special* about special relativity—namely, it applies only to reference frames in uniform motion. The traveling twin accelerates when turning around, so briefly she's in a noninertial reference frame. Although relativity precludes us from saying that one twin is moving and the other is not, we can say that one twin's motion *changes* and the other's doesn't. This is obvious to the traveling twin, who experiences forces associated with the turnaround. The earthbound twin, of course, doesn't feel anything unusual when the ship turns around. During the journey the ship occupies two different inertial frames, separated by the turnaround acceleration, while Earth remains in a single inertial frame. It's that asymmetry that resolves the paradox. The traveling twin really is younger!

What if the traveling twin didn't turn around? Then the situation would be symmetric, and each could argue that the other's clocks "run slow." But unless they get together again at the same place, there's no unambiguous way to compare their clock readings or ages. And they can't get together without accelerating. As we'll soon see, clocks that are synchronized in one reference frame aren't synchronized in another—and that takes the seeming contradiction out of two observers each finding that the other's clocks "run slow."

GOT IT? 33.1 Triplets A and B board spaceships and head away from Earth in opposite directions, each traveling the same distance at the same speed before returning to Earth. When they get back, how do their ages compare? How do they compare with triplet C, who remained on Earth?

Length Contraction

In Example 33.2 the 20-ly distance, 25-y time, and $0.8c$ speed are related through the expression $\Delta x = v\,\Delta t$, where Δx and Δt are measured in the Earth frame. But relativity tells us that this relation must hold in *all* inertial reference frames, so it's also valid in the spaceship frame except during the turnaround. From the ship's viewpoint, Earth and star are moving at $v = 0.80c$, and it takes $\Delta t' = 15$ y for the Earth–star system to pass the ship. Then $\Delta x' = v\,\Delta t' = (0.80\text{ ly/y})(15\text{ y}) = 12$ ly is the Earth–star distance as measured in the ship frame. Thus measures of space as well as time depend on one's reference frame. Equation 33.3 gives $\Delta t' = \Delta t\sqrt{1 - v^2/c^2}$, so we can write generally that

$$\Delta x' = v\,\Delta t' = v\,\Delta t\sqrt{1 - v^2/c^2} = \Delta x\sqrt{1 - v^2/c^2} \quad \text{(length contraction)} \quad (33.4)$$

for the distance between two objects measured in a reference frame in which they move with speed v. Here Δx is the distance in a reference frame at rest with respect to the two objects. Since $\sqrt{1 - v^2/c^2}$ is less than 1 for $v > 0$, Equation 33.4 shows that $\Delta x > \Delta x'$. Therefore, the distance is greatest in this so-called rest frame. The two points in question could be the ends of a single object, in which case Δx is the object's length. This phenomenon of **length contraction** is also called **Lorentz–Fitzgerald contraction**, after Dutch physicist H. A. Lorentz and Irish physicist George F. Fitzgerald, who independently proposed it as an ad hoc way of explaining the Michelson–Morley experiment. Only with Einstein's work did the contraction acquire a solid conceptual basis.

Length contraction shows that an object is longest in its own rest frame and is shorter to observers for whom it's moving. As with time dilation, don't go looking for a physical mechanism that squashes "moving" objects. That presupposes an absolute space with respect to which contraction occurs. Rather, it's space itself that's different for different observers. Accepting relativity means giving up notions of absolute space and time; length contraction and time dilation are necessary consequences.

EXAMPLE 33.3 Length Contraction and Time Dilation: SLAC

At the Stanford Linear Accelerator Center (SLAC), subatomic particles are accelerated to high energies over a straight path whose length, in Earth's reference frame, is 3.2 km. For an electron traveling at $0.9999995c$, how long does the trip take as measured (a) in Earth's frame and (b) in the electrons' frame? (c) What's the length of the linear accelerator in the electrons' frame?

INTERPRET We're being asked about time dilation and length contraction. The Earth frame here is like the Earth frame of Example 33.2, with the ends of the accelerator replacing Earth and star, and $\Delta x = 3.2$ km in place of the 20-ly Earth–star separation. In (a) we're therefore being asked for Δt, in (b) for $\Delta t'$, and in (c) for $\Delta x'$.

DEVELOP As always $\Delta x = v\,\Delta t$ relates distance, time, and speed in a single reference frame. We're given Δx and v, so we'll first solve for Δt. Then we can use Equation 33.3, $\Delta t' = \Delta t\sqrt{1 - v^2/c^2}$, for $\Delta t'$, and Equation 33.4, $\Delta x' = \Delta x\sqrt{1 - v^2/c^2}$, for $\Delta x'$.

EVALUATE The electrons' speed is so close to c that it suffices to calculate the travel time with $v = c$. (a) We have $\Delta t = \Delta x/c = (3.2\text{ km})/(3.0\times 10^8\text{ m/s}) = 11\ \mu\text{s}$. (b) Equation 33.3 gives $\Delta t' = \Delta t\sqrt{1 - v^2/c^2}$; with $v/c = 0.9999995$, the square root works out to be 10^{-3}, so $\Delta t' = 11$ ns. (c) Equation 33.4 shows that the length shrinks by the same factor, to 3.2 m.

ASSESS In this case of extremely relativistic speed, the relativistic factor $\sqrt{1 - v^2/c^2}$ is tiny, and the effects of time dilation and length contraction are dramatic. Note that we could approximate v as c in finding Δt, but not in working with the relativistic factor, where even the slightest difference from c is crucial. As a check on our answer, note that $\Delta x' = v\,\Delta t'$, as required by the principle of relativity. ■

Equations 33.3 and 33.4 show that relativistic effects are significant only at high relative speeds, with v^2/c^2 comparable to 1. We've no experience of such speeds in our everyday lives, so relativity seems counterintuitive. Had we grown up moving relative to our surroundings at speeds approaching c, the relativity of space and time would be as obvious as our commonsense notions seem now. For physicists working with high-energy particles or studying distant, rapidly moving galaxies, relativistic effects *are* obvious features of physical reality.

33.5 Simultaneity Is Relative

One remarkable consequence of relativity is that simultaneity of events and sometimes even their time order depend on one's reference frame. Here we explore how this comes about.

Figure 33.9*a* shows two identical rods approaching each other in a reference frame S where they have the same speed v. Figure 33.9*b* shows that the right end of rod A passes the right end of rod B at the same instant that the left ends pass. Call the passing of the two right ends event E_1 and the passing of the left ends E_2. Clearly E_1 and E_2 are **simultaneous**, meaning they occur at the same time in S.

Now consider a reference frame S' in which rod A is at rest and is therefore longer than it was in frame S. Rod B, meanwhile, is moving toward rod A with a greater speed relative to S' than it had relative to S. Therefore, it's shorter than it was in S. Figure 33.10 shows that, as a result of their different lengths, the right ends of the two rods coincide before the left ends; in other words, event E_1 precedes E_2. Now look at the situation from a reference frame in which rod B is at rest, and you'll see that E_2 precedes E_1 (Fig. 33.11). So events that are simultaneous in one reference frame aren't simultaneous in another frame.

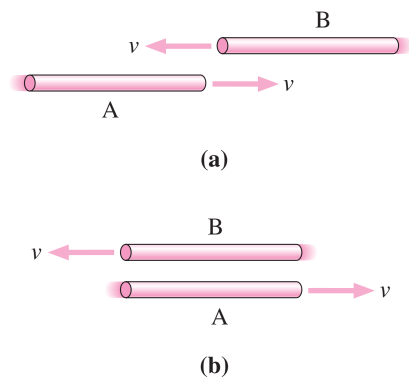


FIGURE 33.9 (a) In frame *S*, rods A and B have the same speed *v* and both are contracted by the same amount. (b) Their ends coincide at the same time.

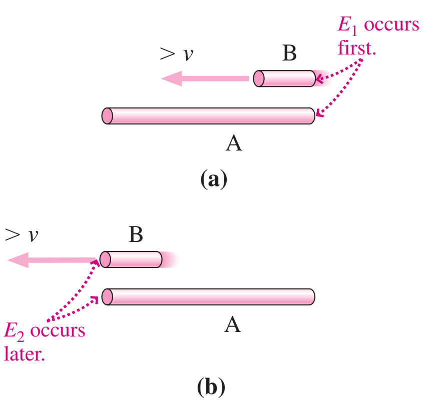


FIGURE 33.10 The passing rods viewed in a reference frame S' at rest with respect to rod A.

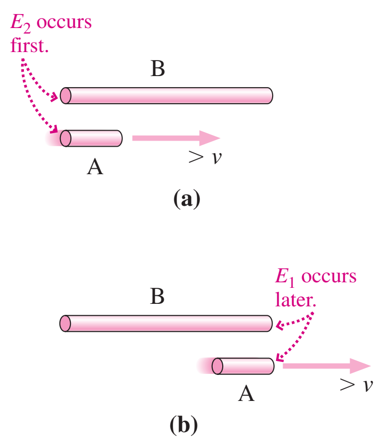


FIGURE 33.11 The passing rods viewed in a reference frame at rest with respect to rod B.

Isn't this just an illusion resulting from apparent length differences due to the rods' motion? Isn't the picture in frame S (Fig. 33.9) the "real" one? No! Relativity assures us that all inertial reference frames are equally valid for describing physical reality. The length differences and changes in the ordering of events aren't "apparent" and they aren't "illusions." They result from valid descriptions in different reference frames, and each has equal claim to reality. If you insist that one frame—say S—has more validity, then you're clinging to the 19th-century notion that there's one favored reference frame in which alone the laws of physics are valid.

But how can observers disagree about the *order* of events? After all, if one event causes another, we expect cause always to precede effect. As we'll soon show, the only events whose time order is different for different observers are those that are so far apart in space, and so close in time, that not even a light signal from one event could reach the location of the other event before it happened. There's no way for such events to influence each other, so they can't be causally related.

CONCEPTUAL EXAMPLE 33.1 "Running Slow": A Contradiction?

In Example 33.2, the outbound trip from Earth to star took 25 years in the Earth–star reference frame but only 15 years in the spaceship's frame. Thus, observers in the Earth–star frame can say that clocks on the ship "run slow." What do passengers on the ship say about clocks in the Earth–star frame?

EVALUATE During the outbound trip, the spaceship is in a perfectly good inertial reference frame. So the laws of physics are the same for the ship's passengers as they are for earthbound observers, and they can make exactly the same argument: They see Earth and star moving, and they conclude that clocks in the Earth–star frame must "run slow."

Were you inclined to give the seemingly more logical answer that, since the ship clocks "run slow," the Earth–star clocks "run fast"? If so, you haven't applied the principle of relativity: The laws of physics are the same in all inertial reference frames. There's nothing special about the spaceship's frame that would make it see Earth's clocks "running fast" when observers on Earth, in an exactly analogous situation, see the ship's clocks "running slow."

ASSESS How is this not a contradiction? The answer lies in the relativity of simultaneity. Earth and star clocks are synchronized in the Earth–star frame, but not in the ship's frame. From the ship's viewpoint, Earth–star clocks are "running slow" by the factor $\sqrt{1 - v^2/c^2} = 0.6$ that we found in Example 33.2. So the 15-year trip time in the ship frame takes, from the ship's perspective, only (0.6)(15 years) $=$ 9 years on the Earth–star clocks. The Earth clock read zero when the ship left Earth, so in the ship frame it reads 9 years when the ship reaches the star. Yet the star clock reads 25 years at that event. So, from the ship's viewpoint, Earth and star clocks are out of sync by 16 years. Figure 33.12 shows the situation from the ship's frame. Compare with Figs. 33.8*a* and *b*: Each observer thinks the other's clocks "run slow," yet there's no contradiction!

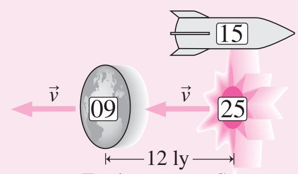


FIGURE 33.12 The situation in the ship's frame. Note that Earth, star, and the distance between them are contracted, while the ship appears longer than in Fig. 33.8.

MAKING THE CONNECTION For the star trek of Example 33.1, how do Earth and the star clock readings differ as judged in the spaceship's reference frame?

EVALUATE The ship sees Earth–star clocks "running slow" by the factor $\sqrt{1 - 0.95^2} = 0.312$. Given the 7.8-year time for the trip on the ship clock, observers on the ship judge that the elapsed time in the Earth–star frame is only $(7.8 \text{ years})(0.312) = 2.4$ years. But we know the star clock reads 25 years when the ship arrives, so the Earth clock is behind by 25 years $-$ 2.4 years $=$ 22.6 years as judged in the ship's reference frame.

33.6 The Lorentz Transformations

Events are determined by where (three spatial coordinates) and when (time) they occur. Our work with time dilation, length contraction, and the ordering of events suggests that these coordinates depend on one's frame of reference. Here we develop general expressions, called the **Lorentz transformations**, that relate the time and space coordinates of events in different reference frames.

Consider coordinate axes in a reference frame S and in another frame S' moving in the x-direction with speed v relative to S. The origins of the two systems coincide at time $t = t' = 0$. Given an event with coordinates x, y, z, t in S, what are its coordinates x', y', z', t' in S'? Were it not for relativity, we'd expect y, z, and t to remain unchanged, while the relative motion along x means a given x in S would correspond to $x' = x - vt$ in S' (Fig. 33.13).

At $t = 2$ s, the x'-axis has moved 2 m to the right, so x' is 2 m less than x.

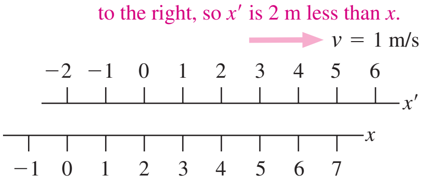


FIGURE 33.13 Nonrelativistic picture of two coordinate axes in relative motion, shown at time $t = 2$ s. In general, $x' = x - vt$.

Relativity should alter the transformations for both time and the spatial coordinate, x, along the direction of relative motion. However, our results must reduce to the nonrelativistic results $x' = x - vt$ and $t' = t$ in the limit $v \ll c$. A simple form with this property is $x' = \gamma(x - vt)$, where γ is a factor, still to be determined, that reduces to 1 as $v \to 0$. We could also transform the other way; the only difference is that the x-axis is moving in the negative direction relative to x'. Therefore, we should have a similar form, with the sign of v reversed: $x = \gamma(x' + vt')$. Now suppose a light flash goes off just as the origins of the two coordinate systems coincide. The coordinates of this event, E_1, are $x = 0, t = 0$ in S, and $x' = 0, t' = 0$ in S'. At some later time t, an observer at position x in S observes the light flash; call this event E_2. Since light travels with speed c, $x = ct$. In frame S', E_2 has coordinates x', t'. But relativity requires that light travel with speed c in *all* inertial reference frames, so we must have $x' = ct'$. Putting these expressions for x and x' in our proposed transformation equations gives $ct' = \gamma t(c - v)$ and $ct = \gamma t'(c + v)$. Multiplying these two equations yields $c^2 = \gamma^2(c - v)(c + v) = \gamma^2(c^2 - v^2)$. Therefore, $\gamma = 1/\sqrt{1 - v^2/c^2}$. Taking $v \to 0$ in this expression shows that $\gamma \to 1$ in the nonrelativistic limit, as required. So we have our transformation equations for x.

What about y, z, and t? The y- and z-axes are perpendicular to the direction of motion, so there's no length contraction and therefore $y' = y$ and $z' = z$. The fact of time dilation makes clear that measures of time differ in different reference frames, so it's not surprising that $t' \neq t$. You can derive the transformation equations for t from those for x (see Problem 42). The results, along with the equations we found for x, y, and z, are summarized in Table 33.1.

Table 33.1 The Lorentz Transformations

S to S'	S' to S	
$y' = y$	$y = y'$	
$z' = z$	$z = z'$	where $\gamma = \dfrac{1}{\sqrt{1 - v^2/c^2}}$
$x' = \gamma(x - vt)$	$x = \gamma(x' + vt')$	
$t' = \gamma(t - vx/c^2)$	$t = \gamma(t' + vx'/c^2)$	

PROBLEM-SOLVING STRATEGY 33.1 Lorentz Transformations

INTERPRET Make sure the problem involves space and time coordinates of events as measured in two different frames of reference. Identify the frames, which generally are those of specific objects introduced in the problem statement. Identify also the events and the particular coordinates you're interested in.

DEVELOP Establish coordinate systems in the two reference frames, arbitrarily designated S and S'. Choosing the direction of relative motion along the x-axis will let you use the Lorentz transformations as they appear in Table 33.1. Remember that time is also a coordinate, and take the two coordinate systems to coincide at time $t = t' = 0$. You'll make the math simpler by choosing the origin in space and time to occur at one of the events in the problem. Then determine any other event coordinates that are implicit in the problem statement.

EVALUATE Apply the appropriate Lorentz transformations from Table 33.1 to calculate the unknown coordinates.

ASSESS Ask whether your results make sense. If your calculated order of events differs in different frames, be sure you're dealing with events that are far enough apart in space and close enough in time that they can't be causally related.

EXAMPLE 33.4 Galactic Fireworks: Using the Lorentz Transformations

Our Milky Way and the Andromeda Galaxy are approximately at rest with respect to each other and are 2 million light years (Mly) apart. Supernova explosions occur simultaneously in both galaxies, as judged in the galaxies' reference frame. A spacecraft is traveling at $0.8c$ from the Milky Way toward Andromeda. Find the time between the supernova events as measured in the spacecraft's reference frame.

INTERPRET We're given two distant events—both supernova explosions—that are simultaneous in a particular reference frame. We're asked to find the time between them in a different reference frame. So this problem is about using the Lorentz transformations for time coordinates.

DEVELOP Following our strategy, we establish coordinate system S, the galaxy reference frame. We take the origin at the Milky Way, with the x-axis extending through Andromeda (here we're treating the galaxies as points). We take $t = 0$ to be the time of the supernova explosions, which are simultaneous in S. Then the coordinates of the two supernovas in S are $x_{MW} = 0$, $t_{MW} = 0$ and $x_A = 2$ Mly, $t_A = 0$. Here the subscripts designate not just the galaxies but the specific events of the supernova explosions. We show the situation at time $t = 0$ in Fig. 33.14. The other reference frame is that of the spaceship, which we designate S'; it's moving at speed $v = 0.8c$ along the x-axis

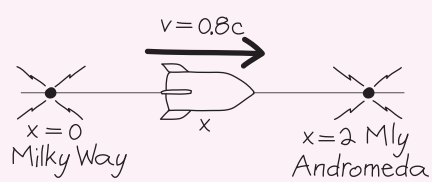


FIGURE 33.14 Our sketch for Example 33.4, drawn in frame S at time $t = 0$, when observers in S judge both supernova explosions to occur. This picture does not apply in frame S'!

(continued)

relative to frame S. Taking the two coordinate systems to coincide at time $t = 0$ gives $t'_{\text{MW}} = 0$, and we're in the situation we used to derive the Lorentz transformations of Table 33.1. So our plan is to apply the transformation equation that gives t', namely, $t' = \gamma(t - vx/c^2)$.

EVALUATE We first evaluate the factor γ, finding that $\gamma = 1/\sqrt{1 - 0.8^2} = 5/3$. Then we apply the transformation equation for t' to get the time of the Andromeda supernova in the spacecraft's reference frame:

$$t'_{\text{A}} = \gamma\left(t_{\text{A}} - \frac{vx_{\text{A}}}{c^2}\right) = \left(\frac{5}{3}\right)\left(0 - \frac{(0.8 \text{ ly/y})(2 \text{ Mly})}{(1 \text{ ly/y})^2}\right) = -2.7 \text{ My}$$

ASSESS Since the Milky Way supernova goes off at $t'_{\text{MW}} = 0$, our *negative* answer means the Andromeda event occurs 2.7 million years *earlier* in the spacecraft's reference frame. Again, there's no problem with causality; since the distant events are simultaneous in *some* frame (the galaxy frame), they can't possibly be cause and effect. You can easily show that a spacecraft observer going the other way would judge the Andromeda supernova to occur 2.7 My later. Problems 39–41 explore the case when the supernovas occur far enough apart in time that they *could* be causally related. ■

Relativistic Velocity Addition

If you're in an airplane going 1000 km/h relative to the ground and you walk toward the front of the plane at 5 km/h, common sense suggests that you move at 1005 km/h relative to the ground. But measures of time and distance vary among frames of reference in relative motion. For this reason the velocity of an object with respect to one frame doesn't simply add to the relative velocity between frames to give the object's velocity with respect to another frame. In the airplane your speed with respect to the ground is actually a little less than 1005 km/h as you stroll down the aisle, although the difference is insignificant at such a low speed.

The correct expression for **relativistic velocity addition** follows from the Lorentz transformations. Consider a reference frame S and another frame S' moving in the positive x-direction with speed v relative to S. Let their origins coincide at time $t = t' = 0$, so the Lorentz transformations of Table 33.1 apply.

Suppose an object moves with velocity u' along the x'-axis in S'. We seek the velocity u of the object relative to the frame S. (We're using u and u' for the object because we've already used v for the relative velocity of the reference frames.)

In either frame, velocity is the ratio of change in position to change in time, or $u = \Delta x/\Delta t$. Designating the beginning of the interval Δt by the subscript 1 and the end by 2, we can use the Lorentz transformations to write

$$\Delta x = x_2 - x_1 = \gamma[(x'_2 - x'_1) + v(t'_2 - t'_1)] = \gamma(\Delta x' + v\,\Delta t')$$

and

$$\Delta t = t_2 - t_1 = \gamma[(t'_2 - t'_1) + v(x'_2 - x'_1)/c^2] = \gamma(\Delta t' + v\,\Delta x'/c^2)$$

Forming the ratio of these quantities, we have

$$\frac{\Delta x}{\Delta t} = \frac{\Delta x' + v\,\Delta t'}{\Delta t' + v\,\Delta x'/c^2} = \frac{(\Delta x'/\Delta t') + v}{1 + v(\Delta x'/\Delta t')/c^2}$$

But $\Delta x'/\Delta t'$ is the velocity u' of the object in frame S', and $\Delta x/\Delta t$ is the velocity u, so

$$u = \frac{u' + v}{1 + u'v/c^2} \quad \text{(relativistic velocity addition)} \tag{33.5a}$$

The numerator of this expression is just what we would expect from common sense. But this simple sum of two velocities is altered by the second term in the denominator, which is significant only when both the object's velocity u' and the relative velocity v are comparable with c. Solving Equation 33.5a for u' in terms of u, v, and c gives the inverse transformation:

$$u' = \frac{u - v}{1 - uv/c^2} \quad \text{(relativistic velocity addition)} \tag{33.5b}$$

EXAMPLE 33.5 Relativistic Velocity Addition: Collision Course

Two spacecraft approach Earth from opposite directions, each moving at $0.80c$ relative to Earth, as shown in Fig. 33.15*a*. How fast do the spacecraft move relative to each other?

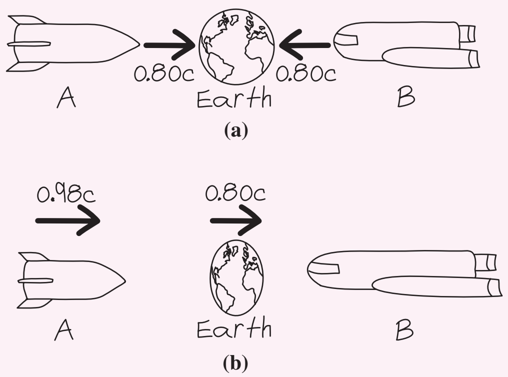


FIGURE 33.15 Sketch for Example 33.5 (a) in Earth's frame S' and (b) in spacecraft B's frame S.

INTERPRET The naïve answer, $1.6c$, isn't consistent with relativity. Instead, this is a problem involving relativistic velocity addition. We identify the Earth frame of reference as S' and ship B's frame as S. Ship A is moving at $u' = 0.8c$ relative to the Earth frame S', and we want to know its speed u relative to ship B's frame S. Ship B is also moving toward Earth at $0.8c$, or, equivalently, Earth is moving toward ship B at this speed, so we identify $v = 0.8c$ as the relative velocity between frames.

DEVELOP With all terms identified, our plan is to apply Equation 33.5a to find the velocity u of ship A relative to ship B.

EVALUATE $u = \dfrac{u' + v}{1 + u'v/c^2} = \dfrac{0.80c + 0.80c}{1 + (0.80c)(0.80c)/c^2} = \dfrac{1.6c}{1.64} = 0.98c$

ASSESS The relative speed is less than the $1.6c$ we get from a naïve addition and also less than the speed of light. This result is quite general: Equations 33.5 imply that as long as an object moves with speed $u < c$ relative to some frame, then its speed relative to any other frame is less than c. And if you set $u = c$ to describe a light beam, you'll find that Equations 33.5 give $u' = c$ as well—reaffirming the relativistic point that the speed of light is the same in all inertial reference frames. ■

GOT IT? 33.2 You're driving down the highway, and your speedometer reads exactly 30 km/h. A car passes you, going in the same direction at exactly 20 km/h relative to you. Does its speedometer—which measures the car's speed relative to the road—read more or less than 50 km/h?

Is Everything Relative?

You already know the answer: The laws of physics aren't relative—that's the fundamental principle of relativity. Neither is the speed of light, whose existence and value follow from laws of physics—specifically, Maxwell's equations. And there are a host of other **relativistic invariants**, independent of reference frame. One such invariant is the **spacetime interval**, a kind of four-dimensional "distance" between two events in space and time. The spacetime interval is given by an expression that looks like a modified Pythagorean theorem:

$$(\Delta s)^2 = c^2(\Delta t)^2 - [(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2] \qquad (33.6)$$

where the Δ quantities are the differences between the time and space coordinates of the events. The invariance of Δs follows directly from the Lorentz transformations, as you can show in Problem 61.

The spacetime interval describes a relation between two events that's independent of reference frame. The invariance of the spacetime interval suggests that something absolute underlies the shifting sands of relativistic space and time. That something is **spacetime**—a four-dimensional framework linking space and time into a single continuum. The spacetime interval is the magnitude of a four-dimensional vector—a **4-vector**—whose components involve the three spatial distances Δx, Δy, Δz and the time Δt between two events. The individual space and time components differ in different reference frames, but they always conspire to give the same invariant interval. This is analogous to the vectors of ordinary two- and three-dimensional space, where the vector components depend on your choice of coordinate system. But the actual vector quantity—for example, a force—has a reality independent of your coordinate choices, and its magnitude doesn't depend on the coordinate system (Fig. 33.16). This analogy isn't perfect because of the negative sign in Equation 33.6's expression for the spacetime interval. That sign reflects the fact that the underlying geometry of spacetime isn't the Euclidean geometry you learned in high school.

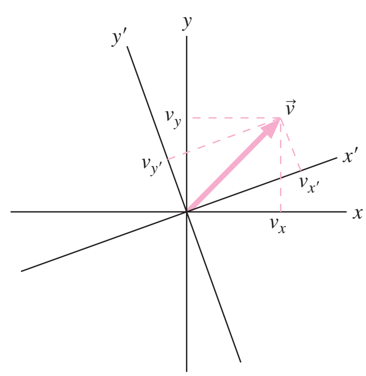


FIGURE 33.16 Although the x- and y-components of an ordinary vector depend on the choice of coordinate system, the magnitude of the vector does not.

Other 4-vectors play a role in more advanced treatments of relativity. These include a four-dimensional electric-current density, whose components involve charge density and the three components of ordinary current density; a four-dimensional "wave vector" that links frequency and wavelength and whose invariant magnitude yields the Doppler effect in its correct relativistic form; and a 4-potential that yields both the electric and magnetic fields. A particularly important example is the energy–momentum 4-vector. Its invariant magnitude is famously related to mass, as we'll see in the next section.

33.7 Energy and Momentum in Relativity

Conservation of momentum and conservation of energy are cornerstones of Newtonian mechanics, where they hold in any inertial reference frame. But momentum and energy are functions of velocity, and we've just seen that relativity alters the Newtonian picture of how velocities transform from one reference frame to another. How, then, can momentum and energy be conserved in all reference frames?

Momentum

In Newtonian mechanics the momentum of a particle with mass m and velocity $\vec{u}$ is $m\vec{u}$. (Here we use $\vec{u}$ for particle velocities, reserving $\vec{v}$ for the relative velocity between reference frames.) But if a system's momentum—the sum of its individual particles' momenta $m\vec{u}$—is conserved in one frame of reference, then relativistic velocity addition suggests that it won't be conserved in another. The problem here lies not with momentum conservation but with the Newtonian expression for momentum. The expression $m\vec{u}$ is actually an approximation valid only for speeds u much less than c. The measure of momentum valid at any speed is

$$\vec{p} = \frac{m\vec{u}}{\sqrt{1 - u^2/c^2}} = \gamma m\vec{u} \qquad \text{(relativistic momentum)} \tag{33.7}$$

where γ is the familiar relativistic factor. The momentum in Equation 33.7 is conserved in all reference frames, and at low velocities it reduces to the Newtonian expression $\vec{p} = m\vec{u}$.

As $u \to c$, the factor γ grows arbitrarily large, and so does the relativistic momentum (Fig. 33.17). Since force is the rate of change of momentum, that means a very large force is required to produce even the slightest change in the velocity of a rapidly moving particle. This helps answer a common question about relativity: Why is it impossible to accelerate an object to the speed of light? The answer is that the object's momentum would approach infinity, and no matter how close to c it was moving, it would still require infinite force to give it the last bit of speed.

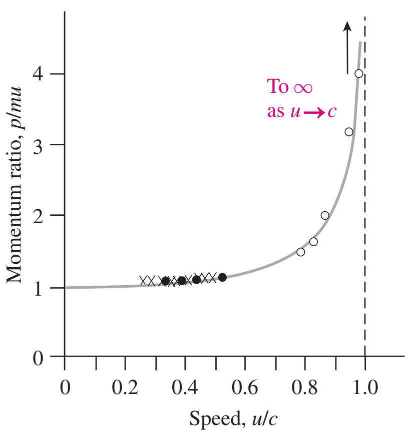


FIGURE 33.17 The ratio of relativistic momentum to Newtonian momentum mu. The curve follows Equation 33.7; crosses and circles mark experimental data.

Energy and Mass

The most widely known result of relativity is the famous equation $E = mc^2$. Here we develop a general expression for relativistic energy and show just what $E = mc^2$ means. In the process we'll gain new insights into energy, momentum, and mass.

In Chapter 6 we derived the work–energy theorem, and thus developed the concept of kinetic energy, by calculating the work needed to accelerate a mass m from rest to some final speed. We did that by integrating force over distance, using Newton's law $F = dp/dt$ to write force in terms of momentum. We'll do the same thing here, now using our relativistic expression for momentum. So we have

$$\frac{dp}{dt} = \frac{d}{dt}\left[\frac{mu}{\sqrt{1 - u^2/c^2}}\right] = \frac{m(du/dt)}{(1 - u^2/c^2)^{3/2}}$$

Then the kinetic energy gained as a particle accelerates from rest in our reference frame to a final speed u is

$$K = \int \frac{dp}{dt}dx = \int \frac{dp}{dt}u\,dt = \int \frac{m(du/dt)}{(1 - u^2/c^2)^{3/2}}u\,dt = \int_0^u \frac{mu}{(1 - u^2/c^2)^{3/2}}du$$

where we used $u = dx/dt$ to replace dx with $u\,dt$. The integral is readily evaluated using the fact that $u\,du = \frac{1}{2}d(u^2)$, giving

$$K = \frac{mc^2}{\sqrt{1 - u^2/c^2}} - mc^2 = \gamma mc^2 - mc^2 \tag{33.8}$$

for a particle's kinetic energy in a reference frame where the particle has speed u. Once again, γ here is the relativistic factor $1/\sqrt{1 - u^2/c^2}$.

For speeds low compared with c, Equation 33.8 reduces to the Newtonian $K = \frac{1}{2}mu^2$, as you can show in Problem 59. But Equation 33.8 suggests, more generally, that kinetic energy is the difference of two energies—the velocity-dependent quantity γmc^2 and the term mc^2 that depends only on mass, not velocity. Pursuing this interpretation, we identify γmc^2 as the particle's **total energy** and mc^2 as its **rest energy**. Rearranging Equation 33.8 lets us write the total energy as the sum of the kinetic energy and the rest energy: $E = K + mc^2$ or, more simply

$$E = \gamma mc^2 = \frac{mc^2}{\sqrt{1 - u^2/c^2}} \qquad (\text{total energy}) \tag{33.9}$$

What does all this mean? Put $u = 0$ in Equation 33.9 and $E = mc^2$—showing that the total energy of a stationary particle isn't zero but is directly proportional to its mass. Thus that particle has energy simply by virtue of having mass or, as Einstein first recognized, *mass and energy are equivalent*. The proportionality between mass and energy is a whopping big c^2—about 9×10^{16} J/kg in SI units.

Although we developed Equation 33.9 by considering kinetic energy, the mass–energy equivalence $E = mc^2$ is universal. Energy, like mass, exhibits inertia. A hot object is slightly harder to accelerate than an otherwise identical cold one because of the inertia of its thermal energy. A stretched spring is more massive than an otherwise identical unstretched one because of its extra potential energy. When a system loses energy, it loses mass as well.

To the public, $E = mc^2$ is synonymous with nuclear energy. The equation does describe mass changes in nuclear reactions, but it applies equally well to chemical reactions and all other energy conversions. Weigh a nuclear power plant and weigh it again a month later, and you'll find it weighs slightly less. Weigh a coal-burning power plant and all the coal and oxygen that go into it for a month, then weigh all the carbon dioxide and other combustion products, and you'll find a difference. If both plants produce the same amount of energy, the mass difference is the same for both. The distinction lies in the amount of mass released as energy in individual reactions. Fission of a single uranium nucleus involves about 50 million times as much energy, and therefore mass, as the reaction of a carbon atom with oxygen to make carbon dioxide. That's why a coal-burning power plant consumes many hundred-car trainloads of coal each week, while a nuclear plant needs only a few truckloads of uranium every year or so. Incidentally, neither process converts very much of the fuel mass to energy; if we could convert *all* the mass in a given object to energy, ordinary matter would be an almost limitless source of energy. Such conversion is in fact possible, but only in the annihilation of matter and antimatter.

EXAMPLE 33.6 Mass–Energy Equivalence: Annihilation

A positron is an antimatter particle with the same mass as the electron but the opposite electric charge. When an electron and positron meet, they annihilate and produce a pair of identical gamma rays (bundles of electromagnetic energy). Find the energy of each gamma ray.

INTERPRET This is a problem about mass–energy equivalence—in this case all the mass of the electron–positron pair changing to gamma-ray energy. We'll assume the pair has negligible kinetic energy K, so the total energy available is just their rest energy.

DEVELOP With $K = 0$, Equation 33.9 reduces to $E = mc^2$. Since there are two particles, each of mass m, and two identical gamma rays, each gets energy mc^2. So our plan is to evaluate this quantity. We'll need the electron mass m, which is given inside the front cover of this book.

EVALUATE We have $E = mc^2 = (9.1\times10^{-31}\,\text{kg})(3.0\times10^{8}\,\text{m/s})^2 = 82\,\text{fJ}$.

ASSESS High-energy physicists usually work in electronvolts, and our 82-fJ answer is equivalent to 511 keV. Detection of 511-keV gamma rays from laboratory or astrophysical sources is a sure sign of electron–positron annihilation. The medical technique PET (positron emission tomography) uses these annihilation gamma rays to image processes occurring inside the body.

Given the fame of $E = mc^2$, it's easy to overlook the fact that the rest energy mc^2 is generally only part of a particle's total energy. For a particle moving at velocity u that's small compared with c, the total energy is only slightly greater than mc^2; the extra is very nearly the Newtonian kinetic energy, $\frac{1}{2}mu^2$. (Here "at rest" and "moving" are, of course, relative to some inertial reference frame.) But when a particle moves with nearly the speed of light, the relativistic factor $\gamma = 1/\sqrt{1 - u^2/c^2}$ becomes much greater than 1, and the total energy γmc^2 is many times the rest energy. Such a particle is termed **relativistic**.

EXAMPLE 33.7 Total Energy: A Relativistic Electron

An electron has total energy 2.50 MeV. Find (a) its kinetic energy and (b) its speed.

INTERPRET Here we're given the electron's total energy and asked how much of that is the kinetic energy associated with its speed, which we're also asked for.

DEVELOP Equation 33.9 gives total energy, which is the sum of kinetic energy K and rest energy mc^2. So our plan is to subtract the rest energy to find the kinetic energy. Equation 33.9 also expresses the total energy as γmc^2, so we can find $\gamma = 1/\sqrt{1 - u^2/c^2}$ and then solve for the speed u. We don't need to calculate mc^2 because we found it in Example 33.6: For the electron, $mc^2 = 511$ keV or 0.511 MeV.

EVALUATE (a) We have

$$K = E - mc^2 = 2.50\text{ MeV} - 0.511\text{ MeV} = 1.99\text{ MeV}$$

(b) Since $E = \gamma mc^2$, $\gamma = E/mc^2 = 2.50\text{ MeV}/0.511\text{ MeV} = 4.89$. But $\gamma = 1/\sqrt{1 - u^2/c^2}$, which we solve to get

$$u = c\sqrt{1 - 1/\gamma^2} = 2.94\times 10^8\text{ m/s}$$

ASSESS Our answer for kinetic energy is considerably greater than the electron's rest energy, and our speed u is close to c, both confirming that this is a relativistic electron. ■

GOT IT? 33.3 The rest energy of a proton is 938 MeV. Without doing any calculations, quickly estimate the speed of a proton with total energy 1 TeV (10^{12} eV).

The Energy–Momentum Relation

In Newtonian physics the equations $p = mu$ and $K = \frac{1}{2}mu^2$ yield $p^2 = 2K/m$. Similarly, Problem 60 shows that in relativity we can combine the equations $p = \gamma mu$ and $E = \gamma mc^2$ to get

$$E^2 = p^2c^2 + (mc^2)^2 \qquad \text{(energy–momentum relation)} \qquad (33.10)$$

which involves E rather than K because in relativity the energy includes both kinetic and rest energies. For a particle at rest, $p = 0$ and Equation 33.10 shows that the total energy is just the rest energy. For highly relativistic particles, the rest energy is negligible and the total energy becomes very nearly $E = pc$. Some "particles"—like the photons that, in quantum physics, are "bundles" of electromagnetic energy—have no mass. These particles exist only in motion at the speed of light, and for them Equation 33.10 gives the exact relation $E = pc$.

Rearranging Equation 33.10 gives $(mc^2)^2 = E^2 - p^2c^2$. This should remind you of Equation 33.6 for the spacetime interval, whose square is the difference between the squares of the time component $c\,\Delta t$ and the spatial separation $\sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}$. Similarly, our rearranged Equation 33.10 gives the square of the rest energy mc^2 as the difference between the squares of the total energy and the magnitude of the momentum multiplied by c. We can therefore think of energy and momentum as the time and space components of a 4-vector. Your frame of reference determines how this energy–momentum 4-vector breaks out into time and space components—that is, into energy and momentum. In a particle's rest frame, for example, $p = 0$ and the vector has only a time component equal to the rest energy. But no matter what frame you're in, the magnitude of the 4-vector is the same, and it's equal to the rest energy mc^2. Therefore, mass—the rest energy divided by the constant c^2—is a relativistic invariant.

33.8 Electromagnetism and Relativity

Historically, relativity arose from deep questions about the propagation of electromagnetic waves. We've seen that relativity alters concepts, such as space, time, energy, and momentum, that are fundamental to Newtonian physics. For that reason, Newtonian physics becomes an approximation valid at low speeds. But relativity is built on the premise that Maxwell's equations of electromagnetism are correct in all reference frames—including the prediction of electromagnetic waves propagating at speed c. Even the title of Einstein's famous 1905 paper—"On the Electrodynamics of Moving Bodies"—shows how intimately related are electromagnetism and relativity. Maxwell's equations are relativistically correct and require no modification.

Although electric and magnetic fields in any frame of reference obey the same Maxwell equations, the fields themselves aren't invariant. Sit in the rest frame of a point charge, and you see a spherically symmetric point-charge field. Move relative to the charge, and you see a magnetic field as well, associated with the moving charge. So electric and magnetic fields aren't absolutes; what one observer sees as a purely electric field another may see as a mix of electric and magnetic fields, and vice versa. You can think of the electric and magnetic fields as components of a more fundamental electromagnetic field; how that field breaks out into electric and magnetic fields depends on your frame of reference. Although the fields are different in different reference frames, there's an important electromagnetic quantity that's invariant—namely, electric charge.

We'll illustrate the deep relationship that relativity imposes on electricity and magnetism by considering the force on a positively charged particle moving relative to a current-carrying wire. For simplicity, assume the wire contains equal line-charge densities of positive and negative charge, moving in opposite directions with the same speed v relative to the wire (Fig. 33.18*a*). The resulting current produces a magnetic field that encircles the wire, and the charged particle moving to the right with velocity $\vec{u}$ as shown in Fig. 33.18*a* experiences a *magnetic* force $\vec{F}_B = q\vec{u} \times \vec{B}$ toward the wire. Because the positive and negative line-charge densities are equal, the wire is neutral, so there's no *electric* force.

Now look at the situation in the particle's reference frame. The positive charges in the wire have a lower speed relative to the particle than the negative charges, so in the particle's frame the distances between negative charges are contracted *more* than the distances between positive charges, as shown in Fig. 33.18*b*. But charge is invariant, so that means the charge per unit length is *greater* for the negative charges. In the frame of the charged particle, *the wire carries a net negative charge!* That results in an *electric* field pointing toward the wire, and therefore the charged particle experiences an *electric* force $\vec{F}_E = q\vec{E}$ toward the wire. Of course, there's still a magnetic field as well, but since the particle is at rest in its own frame, the magnetic force $q\vec{u} \times \vec{B}$ is zero.

We've given two quite different descriptions of the force on the charged particle. In the wire's reference frame, there's a purely magnetic force that we could determine knowing the magnetic force $q\vec{u} \times \vec{B}$ and how magnetic fields arise from currents. In contrast, describing the force in the particle's reference frame requires no knowledge of magnetism whatever. We need to know only the electric force $q\vec{E}$ and how electric fields arise from charges. This illustrates a profound point: Electricity and magnetism aren't independent phenomena that happen to be related. Rather, they're two aspects of a single phenomenon—electromagnetism. Given the relativity principle, it's impossible to have electricity without magnetism, or vice versa. Relativity provides the complete unification of electromagnetism that we've hinted at throughout our study of these phenomena.

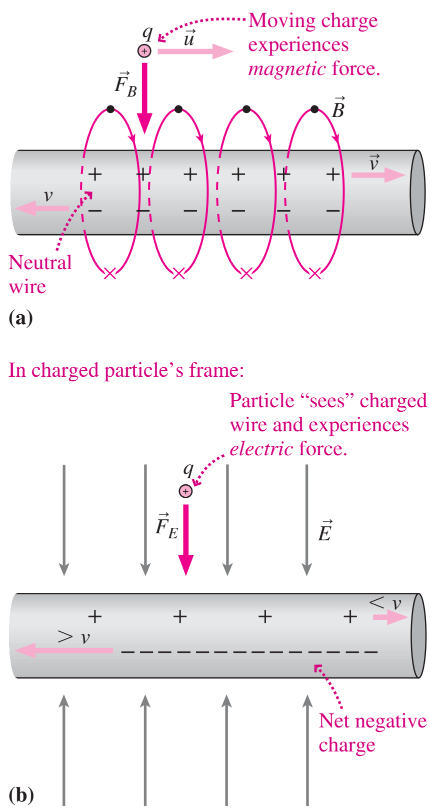


FIGURE 33.18 The force on a charged particle is magnetic or electric, depending on the reference frame.

33.9 General Relativity

The *special* theory of relativity is restricted to reference frames in uniform motion. Following special relativity, Einstein attempted to formulate a theory that would encompass observers in accelerated motion. But he recognized that it's impossible to distinguish the effects of uniform acceleration from those of a uniform gravitational field (Fig. 33.19). Consequently, Einstein's 1916 **general theory of relativity** became a theory of gravity. General relativity describes gravity as the geometrical curvature of four-dimensional

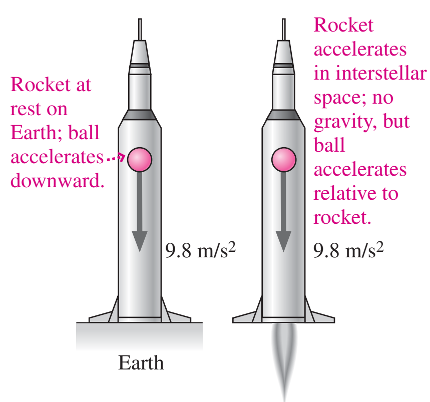


FIGURE 33.19 It's impossible to distinguish the effects of uniform gravitation from acceleration, which is why general relativity is about gravity.

spacetime. In this description, matter and energy curve spacetime in their vicinity, and objects moving through the curved spacetime follow the straightest possible paths—which aren't the straight lines of Euclidean geometry. Figure 33.20 shows a two-dimensional analogy for particles in curved spacetime.

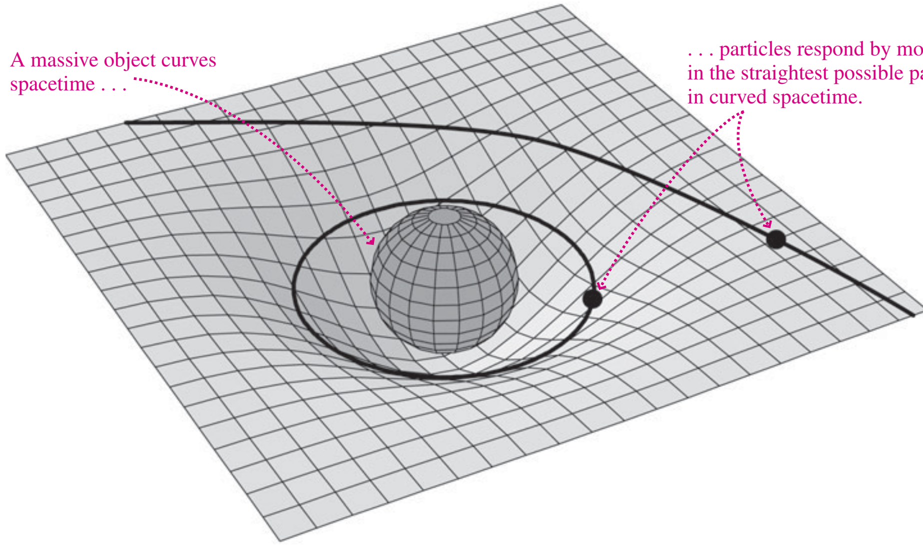


FIGURE 33.20 A two-dimensional analog of curved spacetime in general relativity.

General relativity's predictions differ substantially from those of Newton's theory of gravity only in regions of very strong gravitational fields—far stronger than those found anywhere in our solar system. For that reason general relativity has become a cornerstone of modern astrophysics, describing the physics of such bizarre objects as neutron stars, black holes, and so-called gravitational lenses that can produce multiple images of astrophysical objects (Fig. 33.21). General relativity also addresses cosmological questions of the origin and ultimate fate of the universe. But we're not without terrestrial uses for general relativity; the Global Positioning System (GPS) would be off by several kilometers if the effects of general relativity on the GPS satellites' clocks weren't taken into account.

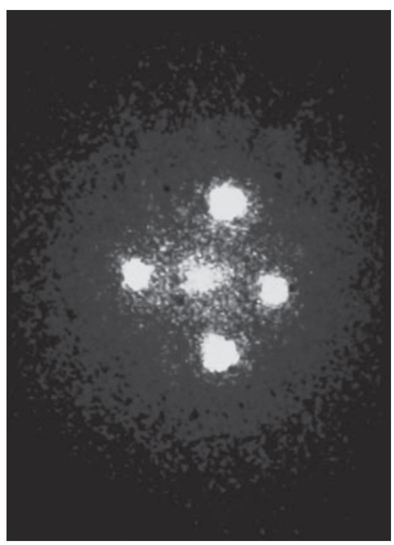


FIGURE 33.21 The "Einstein Cross" comprises four images of the same quasar, formed as light follows different paths in the curved spacetime surrounding a massive galaxy (visible at center).

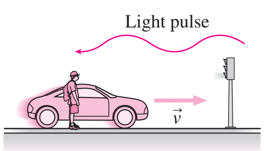
CHAPTER 33 SUMMARY

Big Picture

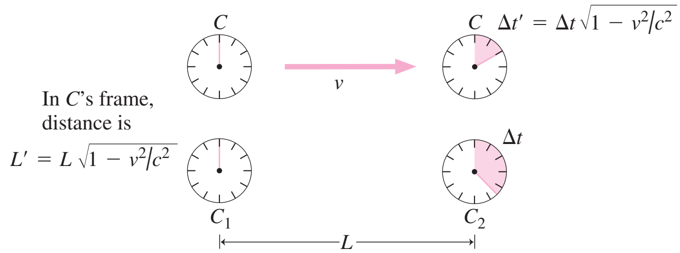
The big idea here is a simple one: The laws of physics are the same for all observers and don't depend on one's state of motion. That's the **principle of relativity. Special relativity** is restricted to inertial reference frames; **general relativity** removes that restriction and in so doing becomes a theory of gravity. Both relativity theories radically alter our commonsense notions of space and time. In relativity, measures of space and time—but not the laws of physics—become dependent on one's reference frame.

Key Concepts and Equations

Among the laws of physics are Maxwell's equations, with their prediction of electromagnetic waves (e.g., light) propagating at the speed of light c. Therefore, the speed of light is the same in all inertial reference frames.



Invariance of c leads to **time dilation** and **length contraction**:



Additionally, events simultaneous in one reference frame may not be simultaneous in another frame.

Time dilation and length contraction are specific instances of the **Lorentz transformations** of space and time coordinates of events observed in different reference frames. The transformations here apply to relative motion in the x-direction.

S to S'	**S' to S**
$y' = y$	$y = y'$
$z' = z$	$z = z'$
$x' = \gamma(x - vt)$	$x = \gamma(x' + vt')$
$t' = \gamma(t - vx/c^2)$	$t = \gamma(t' + vx'/c^2)$

where $\gamma = \dfrac{1}{\sqrt{1 - v^2/c^2}}$

Underlying the changing measures of space and time is four-dimensional **spacetime**, in which exist **4-vectors** whose magnitude is independent of reference frame.

Invariant spacetime interval:

$$(\Delta s)^2 = c^2(\Delta t)^2 - [(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2]$$

Invariant particle mass:

$$(mc^2)^2 = E^2 - p^2c^2$$

Energy, momentum, and mass are closely related in relativity:

Momentum: $\vec{p} = \dfrac{m\vec{u}}{\sqrt{1 - u^2/c^2}} = \gamma m\vec{u}$

Energy: $E = \dfrac{mc^2}{\sqrt{1 - u^2/c^2}} = \gamma mc^2 = K + mc^2 = \sqrt{(pc)^2 + (mc^2)^2}$

Kinetic energy   Rest energy

Applications

The **Michelson–Morley experiment** of 1887 failed to detect any motion of Earth relative to the **ether**, the medium in which 19th-century physicists believed light propagated with speed c. This result helped pave the way for relativity, with no ether and with c the same in all inertial reference frames.

Relativity shows that velocities don't simply add; rather,

$$u = \frac{u' + v}{1 + u'v/c^2}$$

where u' is an object's velocity relative to a reference frame S', u its velocity relative to frame S, and v the relative velocity between S and S'.

Einstein's famous $E = mc^2$ describes a universal interchangeability between matter and energy; contrary to common opinion, it isn't just about nuclear energy.

For Thought and Discussion

1. Why was the Michelson–Morley experiment a more sensitive test of motion through the ether than independent measurements of the speed of light in two perpendicular directions?
2. Why was it necessary to repeat the Michelson–Morley experiment throughout the year?
3. What's *special* about the *special* theory of relativity?
4. Does relativity require that the speed of sound be the same for all observers? Why or why not?
5. Time dilation is sometimes described by saying that "moving clocks run slow." In what sense is this true? In what sense does the statement violate the spirit of relativity?
6. If you're in a spaceship moving at $0.95c$ relative to Earth, do you perceive time to be passing more slowly than it would on Earth? Think! Is your answer consistent with the relativity principle?
7. The Andromeda Galaxy is 2 million light years from the Milky Way. Although nothing can go faster than light, it would still be possible to travel to Andromeda in much less than 2 million years. How?
8. Is matter converted to energy in a nuclear reactor? In a burning candle? In your body?
9. If you took your pulse while traveling in a high-speed spacecraft, would it be faster than, slower than, or the same as on Earth?
10. The rest energy of an electron is 511 keV. What's the approximate speed of an electron whose total energy is 1 GeV? (*Note:* No calculations needed!)
11. An atom in an excited state emits a burst of light. What happens to the atom's mass?
12. The quantity $\vec{E} \cdot \vec{B}$ is invariant. What does this say about how different observers will measure the angle between $\vec{E}$ and $\vec{B}$ in a light wave?

Exercises and Problems

Exercises

Section 33.2 Matter, Motion, and the Ether

13. An airplane makes a round trip between two points 1800 km apart, flying with airspeed 800 km/h. What's the round trip flying time (a) if there's no wind, (b) with wind at 130 km/h perpendicular to a line joining the two points, and (c) with wind at 130 km/h along a line joining the two points?
14. Consider a Michelson–Morley experiment with 11-m light paths perpendicular and parallel to the ether wind. What would be the difference in light travel times on the two paths if Earth moved relative to the ether at (a) its orbital speed (Appendix E); (b) $0.01c$; (c) $0.5c$; and (d) $0.99c$?

Section 33.4 Space and Time in Relativity

15. Two stars are 50 ly apart, measured in their common rest frame. How far apart are they to a spaceship moving between them at $0.75c$?
16. How long would it take a spacecraft traveling at $0.65c$ to get from Earth to Pluto according to clocks (a) on Earth and (b) on the spacecraft? Assume Earth and Pluto are on the same side of the Sun.
17. A spaceship passes by you at half the speed of light, and you determine that it's 35 m long. Find its length as measured in its rest frame.
18. An extraterrestrial spacecraft whizzes through the solar system at $0.80c$. How long does it take to go the 8.3 light minute distance from Earth to Sun (a) according to an observer on Earth and (b) according to an alien aboard the ship?
19. How fast would you have to move relative to a meter stick for it to be 99 cm long in your reference frame?
20. BIO A hospital's linear accelerator produces electron beams for cancer treatment. The accelerator is 1.6 m long and the electrons reach a speed of $0.98c$. How long is the accelerator in the electrons' reference frame?

Section 33.7 Energy and Momentum in Relativity

21. By what factor does an object's momentum change if you double its speed when its original speed is (a) 25 m/s and (b) 100 Mm/s?
22. At what speed will the momentum of a proton (mass 1 u) equal that of an alpha particle (mass 4 u) moving at $0.5c$?
23. At what speed will the Newtonian expression for momentum be in error by 1%?
24. A particle is moving at $0.90c$. If its speed increases by 10%, by what factor does its momentum increase?
25. Find (a) the total energy and (b) the kinetic energy of an electron moving at $0.97c$.
26. At what speed will the relativistic and Newtonian expressions for kinetic energy differ by 10%?

Problems

27. Show that the time of Equation 33.2 is longer than that of Equation 33.1 when $0 < v < c$.
28. You're designing a Michelson interferometer in which a speed-of-light difference of 100 m/s in two perpendicular directions is supposed to shift the interference pattern so a bright fringe of 550-nm light ends up where the adjacent dark fringe would be in the absence of a speed difference. How long should you make the interferometer's arms?
29. Earth and Sun are 8.3 light minutes apart, as measured in their rest frame. (a) What's the speed of a spacecraft that makes the trip in 5.0 min according to its on-board clocks? (b) What's the trip time as measured by clocks in the Earth–Sun reference frame?
30. You're the communications officer on a fast spaceship that takes 50 years in ship time to reach the Andromeda Galaxy, 2 million light years from Earth in the common rest frame of Earth and Andromeda. As soon as you reach Andromeda, your captain orders you to send a radio message to Earth announcing your arrival; he claims the message will reach Earth about a century after you left. You claim it will be much later when the message arrives. Who's right?
31. You wish to travel to a star N light years from Earth. How fast must you go if the one-way journey is to occupy N years of your life?
32. The nearest star beyond our solar system is about 4 light years away. If a spaceship can get to the star in 5 years, as measured on Earth, (a) how long would the ship's pilot judge the journey to take? (b) How far from Earth would the pilot find the star to be?
33. Twins A and B live on Earth. On their 20th birthday, twin B climbs into a spaceship and makes a round-trip journey at $0.95c$ to a star 30 light years distant, as measured in the Earth–star reference frame. What are their ages when twin B returns to Earth?
34. Radioactive oxygen-15 decays at such a rate that half the atoms in a given sample decay every 2 min. If a tube containing 1000 O-15 atoms is moved at $0.80c$ relative to Earth for 6.67 min according to clocks on Earth, how many atoms will be left at the end of that time?

35. Two distant galaxies are receding from Earth at $0.75c$ in opposite directions. How fast does an observer in one galaxy measure the other to be moving?
36. Two spaceships are racing. The "slower" one passes Earth at $0.70c$, and the "faster" one moves at $0.40c$ relative to the slower one. What's the faster ship's speed relative to Earth?
37. Use relativistic velocity addition to show that if an object moves at speed $v < c$ relative to some inertial reference frame, then its speed relative to any other inertial frame must also be less than c.
38. Earth and Sun are 8.33 light minutes apart. Event A occurs on Earth at time $t = 0$ and event B on the Sun at $t = 2.45$ min, as measured in the Earth–Sun frame. Find the time order and time difference between A and B for observers (a) moving on a line from Earth to Sun at $0.750c$, (b) moving on a line from Sun to Earth at $0.750c$, and (c) moving on a line from Earth to Sun at $0.294c$.
39. You're writing a galactic history involving two civilizations that evolve on opposite sides of a 10^5-ly-diameter galaxy. In the galaxy's reference frame, civilization B launched its first spacecraft 50,000 years after civilization A. You and your readers, from a more advanced civilization, are traveling through the galaxy at $0.99c$ on a line from A to B. Which civilization do you record as having first achieved interstellar travel, and how much in advance of the other?
40. Repeat Problem 39, now assuming that civilization B lags A by 1 million years in the galaxy's reference frame.
41. Could there be observers who would judge the events in Problems 39 and 40 to be simultaneous? If so, how must such observers be moving relative to the galaxy?
42. Derive the Lorentz transformations for time from the transformations for space.
43. In the light box of Fig. 33.6, let event A be the emission of the light flash and event B its return to the source. Assign suitable space and time coordinates to these events in the frame in which the box moves with speed v. Apply the Lorentz transformations to show that the time between the two events in the box frame is given by Equation 33.3.
44. You're a consultant for the director of a sci-fi movie. The film starts with two spaceships, each measuring 25 m long in its rest frame, approaching Earth in opposite directions with speeds shown in Fig. 33.22. The director wants to know how long to make ship B for scenes shot (a) in Earth's reference frame and (b) in ship A's frame. Your answers?

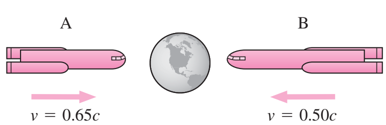


FIGURE 33.22 Problem 44; the drawing is in Earth's reference frame

45. How fast would you have to go to reach a star 200 light years away in a 75-year human lifetime?
46. An advanced civilization has developed a spaceship that goes, with respect to the galaxy, only 50 km/s slower than light. (a) According to the ship's crew, how long does it take to cross the galaxy's 100,000-ly diameter? (b) What's the galactic diameter measured in the ship's reference frame?
47. A spaceship travels at $0.80c$ from Earth to a star 10 light years distant, as measured in the Earth–star reference frame. Let event A be the ship's departure from Earth and event B its arrival at the star. (a) Find the distance and time between the two events in the Earth–star frame. (b) Repeat for the ship's frame. (*Hint:* The distance in the ship frame is the distance an observer has to move *with respect to that frame* to be at both events—not the same as the Lorentz-contracted distance between Earth and star.) (c) Compute the square of the spacetime interval in both frames to show explicitly that it's invariant.
48. Use Equation 33.6 to calculate the square of the spacetime interval between the events (a) of Problem 39 and (b) of Problem 40. Comment on the signs of your answers in relation to the possibility of a causal relationship between the events.
49. A light beam is emitted at event A and arrives at event B. Show that the spacetime interval between the two events is zero.
50. Compare the momentum changes needed to boost a spacecraft (a) from $0.1c$ to $0.2c$ and (b) from $0.8c$ to $0.9c$.
51. Event A occurs at $x = 0$ and $t = 0$ in reference frame S. Event B occurs at $x = 3.8$ light years and $t = 1.6$ years in S. Find (a) the distance and (b) the time between A and B in a frame moving at $0.80c$ along the x-axis of S.
52. When a particle's speed doubles, its momentum increases by a factor of 3. What was its original speed?
53. Find (a) the speed and (b) the momentum of a proton with kinetic energy 500 MeV.
54. Among the most energetic cosmic rays ever detected are protons with energies around 10^{20} eV. Find the momentum of such a proton, and compare with that of a 25-mg insect crawling at 2 mm/s.
55. A large city consumes electrical energy at the rate of 1 GW. If you converted all the rest mass in a 1-g raisin to electrical energy, for how long could it power the city?
56. In a nuclear-fusion reaction, two deuterium nuclei combine to make a helium nucleus plus a neutron, releasing 3.3 MeV of energy in the process. By how much do the combined masses of the helium nucleus and the neutron differ from the combined masses of the original deuterium nuclei?
57. Find the kinetic energy of an electron moving at (a) $0.0010c$, (b) $0.60c$, and (c) $0.99c$. Use suitable approximations where possible.
58. Find the speed of an electron with kinetic energy (a) 100 eV, (b) 100 keV, (c) 1 MeV, and (d) 1 GeV. Use suitable approximations where possible.
59. Use the binomial approximation (Appendix A) to show that Equation 33.8 reduces to the Newtonian expression for kinetic energy in the limit $u \ll c$.
60. Show that Equation 33.10 follows from the expressions for relativistic momentum and total energy.
61. Show from the Lorentz transformations that the spacetime interval of Equation 33.6 has the same value in all reference frames.
62. How fast would you have to travel to reach the Crab Nebula, 6500 light years from Earth, in 20 years? Give your answer to seven significant figures.
63. At what speed are a particle's kinetic and rest energies equal?
64. A cosmic-ray proton with energy 20 TeV is heading toward Earth. What's Earth's diameter measured in the proton's reference frame?
65. When an object's speed increases by 5%, its momentum increases by a factor of 5. What was its original speed?
66. Use the Lorentz transformations to show that if two events are separated in space and time so that a light signal leaving one event cannot reach the other, then there is an observer for whom the two events are simultaneous. Show that the converse is also true: If a light signal can get from one event to the other, then no observer will find them simultaneous.
67. A source emitting light with frequency f moves toward you at speed u. By considering both time dilation and the effect of

wavefronts "piling up" as shown in Fig. 14.34, show that you measure a Doppler-shifted frequency given by

$$f' = f\sqrt{\frac{c+u}{c-u}}$$

Use the binomial approximation (Appendix A) to show that this result can be written in the form of Equation 14.13 when $u \ll c$.

68. Equation 33.5a transforms the velocity $\bar{u}$ of an object moving in the x-direction—the same direction as the relative velocity $\bar{v}$ of the two reference frames. Now suppose the object's velocity also has a component u_y perpendicular to the two frames' relative velocity $\vec{v}$. Find the transformation from u_y' to u_y.
69. Consider a relativistic particle of mass m moving along a straight line. Use Equation 33.7 to find an expression for the force on the particle, defined as $F = dp/dt$, in terms of its acceleration $a = du/dt$.
70. Find the speed of a particle whose relativistic kinetic energy is 50% greater than the Newtonian value calculated for the same speed.
71. It's the 24th century, and you're a curator at the Starfleet Museum of Ancient Technology. Archaeologists have unearthed a "TV tube," an ancient device for displaying moving images. Your job is to get it working. One reference says the device accelerated electrons, which then bombarded a screen to produce images; to the electrons, the tube was 57 cm long. You measure the tube and find it's 60 cm long. To get it working, you need to know the electrons' speed and the potential difference needed to accelerate them. The electron's rest energy is 511 keV. Your answers?

Passage Problems

You've been named captain of NASA's first interstellar mission since the Voyager robotic spacecraft. You board your spaceship, accelerate quickly to $0.8c$, and cruise at constant speed toward Proxima Centauri, the closest star to our Sun. Proxima Centauri is 4 light-years distant as measured in the two stars' common rest frame. On the way, you conduct various medical experiments to determine the effects of a long space voyage on the human body.

72. Taking your pulse, you find
    a. it's significantly slower than when you're on Earth.
    b. it's the same as when you're on Earth.
    c. it's significantly faster than when you're on Earth.
73. How much do you age during your interstellar journey?
    a. 3 years
    b. just under 4 years
    c. just over 4 years
    d. 5 years
74. Back on Earth, Mission Control judges that your shipboard clocks run slow. What do you judge about clocks at Mission Control?
    a. They run fast.
    b. They keep time at the same rate as your clocks.
    c. They run slow.
    d. You can't tell anything about their clocks.
75. In your spaceship's reference frame, the distance from the Sun to Proxima Centauri is
    a. 2.4 light years.
    b. just under 4 light years.
    c. 4 light years.
    d. 5 light years.

Answers to Chapter Questions

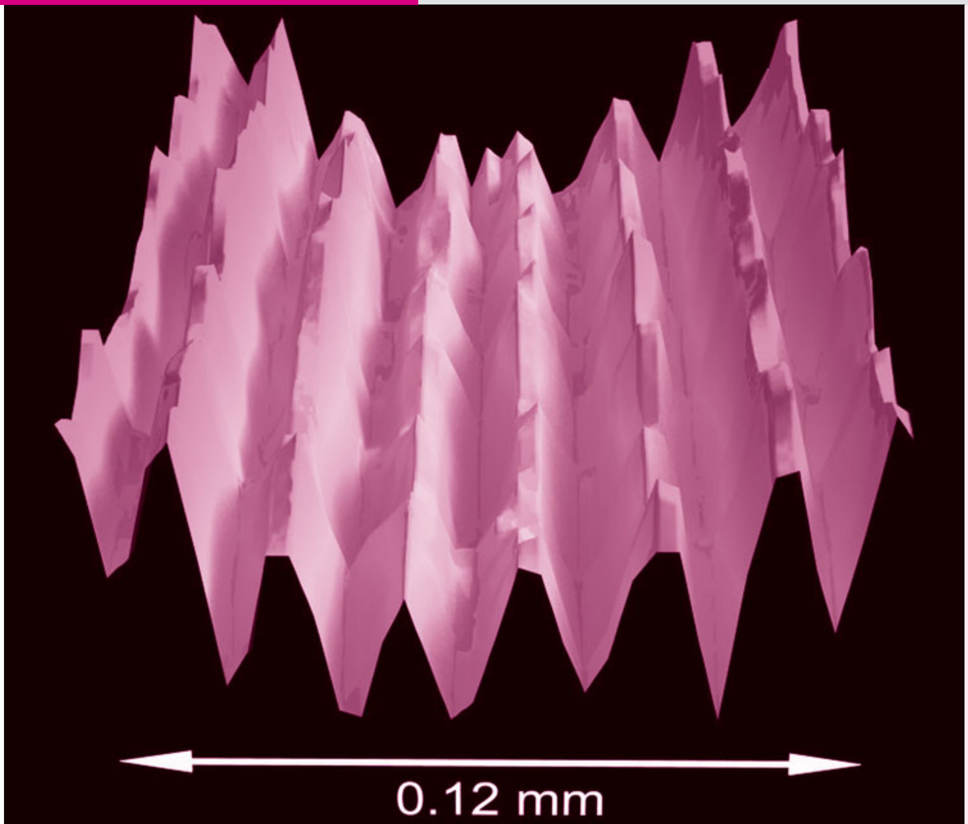
Answer to Chapter Opening Question

The laws of physics are the same for all observers, regardless of their state of motion.

Answers to GOT IT? Questions

33.1. A and B are the same age, both younger than C.
33.2. Less, although insignificantly so at such low relative speeds.
33.3. Very nearly c.

34 Particles and Waves



Ridges and valleys represent bright and dark fringes in the interference pattern produced by two beams of ultracold sodium atoms. What does this picture tell us about the nature of matter?

New Concepts, New Skills

By the end of this chapter you should be able to

- Explain how blackbody radiation, the photoelectric effect, the Compton effect, and atomic spectra contradict classical physics and thus led to the development of quantum physics (34.2, 34.3, 34.4).
- Describe the quantization hypotheses of Planck, Einstein, and Bohr, including the role of Planck's constant h (34.2, 34.3, 34.4).
- Assess quantitatively the wavelengths of radiation for blackbodies at different temperatures (34.2).
- Use the quantization condition $E = hf$ to determine photon energies (34.3).
- Calculate photon energies and wavelengths associated with electron transitions in Bohr's model for the atom (34.4).
- Describe wave–particle duality, and calculate the wavelengths of matter waves given mass and velocity (34.5).
- Explain the uncertainty principle and Bohr's complementarity principle (34.6, 34.7).

Connecting Your Knowledge

- This chapter introduces radical new ideas at the heart of quantum physics. Nevertheless, we'll build on some already-familiar concepts, including energy and momentum (Chapters 4 and 6).
- We'll use wave concepts, including wavelength, frequency, interference, and diffraction (14.1, 14.2, 14.5, 32.5).

Newtonian mechanics and Maxwell's electromagnetism constitute the core of **classical physics**, providing a deep understanding of physical reality. Although these theories were firmly established by the middle of the 19th century, they remain central to the work of many contemporary scientists and engineers.

Nevertheless, at the end of the 19th century a few seemingly minor phenomena defied classical explanation. Most physicists felt that it was only a matter of time before these, too, came under the classical umbrella. But that was not to be. We've seen how questions about light led to a radical restructuring of our concepts of space and time. Other questions, especially those concerning matter at the atomic scale, brought about an even more radical transformation of physical thought.

This chapter explores some phenomena that led to quantum physics and recounts early attempts to explain them. The next chapter gives a fuller account of quantum theory, and subsequent chapters explore its application to atoms, molecules, nuclei, and quantum-based technologies.

34.1 Toward Quantum Theory

Are matter and energy continuously divisible? The essential difference between classical and quantum physics is that the former answers this question "in general, yes," while the latter says definitively "no." Most physical quantities are **quantized**, coming in only certain discrete values.

The idea that physical quantities might come in discrete "chunks" is not new. Some 2400 years ago the Greek philosopher Democritus proposed that all matter consists of indivisible atoms. By the start of the 20th century a more scientifically grounded atomic theory was widely accepted. J. J. Thomson's discovery of the electron in 1897 showed that atoms might be divisible after all, but at the same time it revealed a finer division of matter into discrete "chunks." Robert A. Millikan's 1909 oil-drop experiment showed that electric charge is similarly quantized. Discovery of the proton and later the neutron further solidified the notion that matter comprises fundamental building blocks with definite values for their physical properties.

Quantization of matter into particles with discrete properties is not incompatible with classical physics as long as those particles behave according to classical laws—in particular, that they move continuously through space and can have *any* amount of energy. Add electromagnetism to the picture and the classical viewpoint requires that the fields be continuous, exerting forces on charged particles and changing, in a gradual and continuous way, the particles' energies.

The startling fact of quantum physics is that this classical behavior does not occur at the atomic scale; instead, energy itself is often quantized. Reconciling the implications of that fact with our commonsense notions of matter and motion has proved impossible; instead, the quantum world speaks a different language, one in which deeply ingrained ideas about causality and the solid reality of matter seem no longer to apply. Here we look at three distinct phenomena that force us to accept the idea that energy can be quantized.

34.2 Blackbody Radiation

Heat an object hot enough and it glows, emitting electromagnetic radiation in the form of light. As we saw in Section 16.3, the total power radiated is proportional to the fourth power of the temperature. There's also a change in wavelength with increasing temperature: The first visible glow is a dull red, changing with higher temperatures to orange and then yellow colors corresponding to ever-shorter wavelengths.

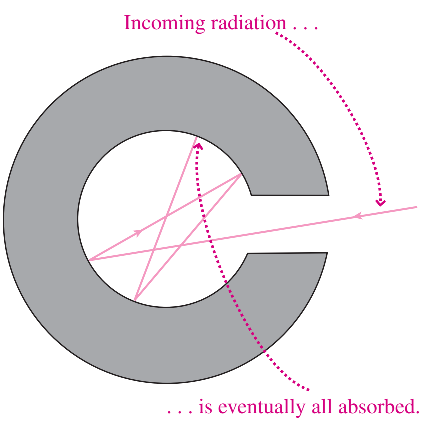


FIGURE 34.1 A cavity with a small hole absorbs nearly all incident radiation and hence is a near-perfect blackbody.

A perfect absorber of electromagnetic radiation is called a **blackbody** because it absorbs all light and thus appears black. When a blackbody is heated, it emits **blackbody radiation** at a range of wavelengths. Many objects—such as the Sun or an electric-stove burner—behave approximately like blackbodies. An excellent approximation to a blackbody is a hollow piece of *any* material with a small hole. As Fig. 34.1 shows, any radiation entering the hole undergoes multiple reflections and is eventually absorbed. The hole, therefore, is a nearly perfect absorber, so when the material is heated, the radiation emerging is blackbody radiation.

Experimental study of blackbody radiation shows three characteristic features:

1. The radiation covers a continuous range of wavelengths, with the total power radiated at all wavelengths combined given by the Stefan–Boltzmann law that we introduced in Chapter 16:
$$P_{\text{blackbody}} = \sigma A T^4 \tag{34.1}$$
where A is the area of the radiating surface, T is its absolute (kelvin) temperature, and $\sigma = 5.67 \times 10^{-8}\ \text{W}/(\text{m}^2 \cdot \text{K}^4)$ is the Stefan–Boltzmann constant.
2. The radiation peaks at a wavelength that's inversely proportional to the temperature; this is known as **Wien's law**.
3. The distribution of wavelengths depends only on temperature, not on the material of which the blackbody is made.

A blackbody's **radiance** measures the radiated power as a function of wavelength. Because the blackbody emits a continuous spectrum, we have to express radiance as power per unit spectral interval. If we choose intervals in wavelength, then the relation implied in feature 2 above gives a peak radiance at wavelength λ_{peak} such that

$$\lambda_{\text{peak}} T = 2.898\ \text{mm} \cdot \text{K} \qquad (\text{Wien's law}) \tag{34.2a}$$

We emphasize, however, that the choice of fixed wavelength intervals is arbitrary. If we had chosen fixed frequency intervals, then the constant in Equation 34.2a would be different, and a plot of radiance versus wavelength would peak at a different wavelength (see Problem 74). So λ_{peak} in Equation 34.2a is not some absolute measure of the wavelength at which the blackbody emits the "most" radiation, but rather the wavelength of the maximum radiation if you choose to keep track of power in intervals of fixed wavelength. A more physically based quantity is the median wavelength, below and above which half the power is radiated; it's given by

$$\lambda_{\text{median}}T = 4.11 \text{ mm} \cdot \text{K} \qquad (34.2b)$$

Whatever measure one chooses, though, the important point is that the peak wavelength is inversely proportional to temperature. In our subsequent discussion we'll adopt a definition of radiance as the power emitted per unit area per unit wavelength interval; then Equation 34.2a describes the peak wavelength. Figure 34.2 plots this measure of blackbody radiance at three temperatures.

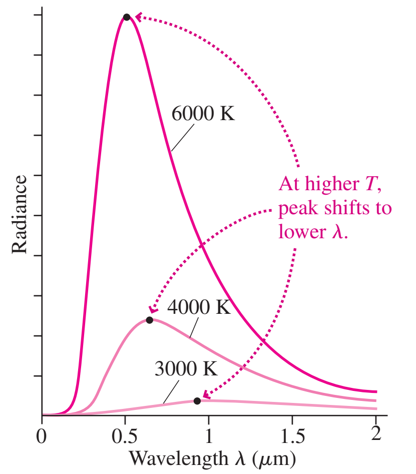


FIGURE 34.2 Blackbody radiance—energy per unit wavelength interval—as a function of wavelength.

Microscopically, blackbody radiation is associated with the thermal motions of atoms and molecules, so it's not surprising that the radiation increases with temperature. In the late 1800s physicists tried to apply the laws of electromagnetism and statistical mechanics to explain the experimental observations of blackbody radiation. They met with some success in describing such aspects as the T^4 dependence of the total energy radiated, and the shifting of the radiation distribution toward shorter wavelengths with increasing temperature, but they could not reproduce the actual observed distribution at all wavelengths.

In 1900 the German physicist Max Planck formulated an equation that fit the observed radiance curves for blackbody radiation at all wavelengths:

$$R(\lambda, T) = \frac{2\pi hc^2}{\lambda^5(e^{hc/\lambda kT} - 1)} \qquad (34.3)$$

Two familiar quantities here are Boltzmann's constant $k = 1.38\times10^{-23}$ J/K, introduced in Chapter 17, and the speed of light c. A new quantity is the constant h, whose value Planck chose in order to make the equation fit experimental data.

Planck first presented his law as a purely empirical equation describing blackbody experiments. Later he showed that his equation had a remarkable physical interpretation:

The energy of a vibrating molecule is quantized, meaning it can have only certain discrete values. Specifically, if f is the vibration frequency, then the energy must be an integer multiple of the quantity hf:

$$E = nhf, \qquad n = 0, 1, 2, 3, \ldots \qquad (34.4)$$

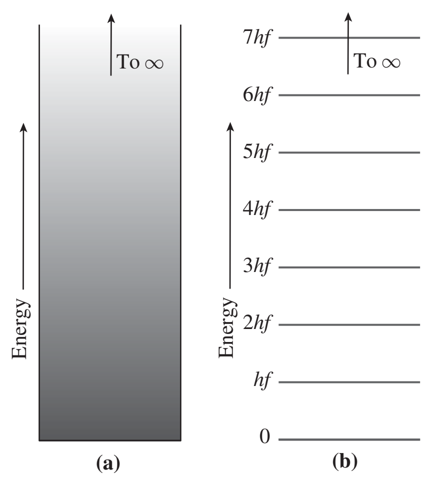


FIGURE 34.3 (a) In classical physics, a vibrating molecule can have any energy. (b) Allowed energies in Planck's theory are integer multiples of *hf*. Energy-level diagrams like this are used frequently in quantum physics, and usually the horizontal axis has no physical significance.

where h is the constant Planck introduced in Equation 34.3. Today we know h as one of the fundamental constants of nature and call it **Planck's constant**. Its value is approximately 6.63×10^{-34} J·s, and it's because h is so small that quantum phenomena are usually obvious only in the atomic and molecular realm. Planck's quantization of the energy of vibrating molecules implies further that a molecule can absorb or emit energy only in discrete "bundles" of size hf, and that in doing so it jumps abruptly from one of its allowed energy levels to another (Fig. 34.3). (Later developments showed that Planck was correct about the size of the energy jumps but that the factor n in Equation 34.4 should actually be $n + \frac{1}{2}$.)

Planck himself was very conservative and reluctant to accept or elaborate on his theory's evident disagreement with classical physics; nevertheless, his revolutionary work won Planck the 1918 Nobel Prize. Other physicists subsequently emphasized the contrast between Planck's work and the classical treatment of blackbody radiation. That earlier treatment, based on the assumption that energy is shared equally among

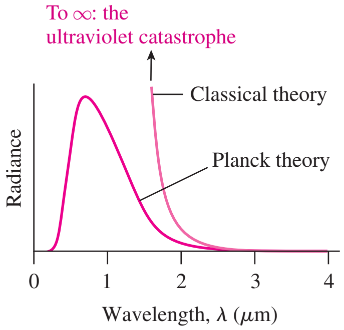


FIGURE 34.4 Radiance versus wavelength for blackbody radiation at 6000 K, showing also the incorrect classical prediction.

all possible vibrational modes, had led to the **Rayleigh–Jeans law** for the radiance of a blackbody:

$$R(\lambda, T) = \frac{2\pi ckT}{\lambda^4} \tag{34.5}$$

Not only did the Rayleigh–Jeans law contradict experimental measurements, but it led to the absurd conclusion that every object, at every nonzero temperature, should emit electromagnetic energy at an infinite rate, with that energy concentrated at the shortest wavelengths (Fig. 34.4). Since the shortest wavelength known at the time was ultraviolet, this phenomenon was called the **ultraviolet catastrophe**. In Planck's equation the exponential term in the denominator grows rapidly with decreasing wavelength, diminishing the radiance and averting the ultraviolet catastrophe. Problems 68, 72, and 75 show that Planck's law reduces to the Rayleigh–Jeans law for longer wavelengths, and that it also leads to Wien's displacement law and the Stefan–Boltzmann law.

EXAMPLE 34.1 Blackbody Radiation: Lightbulb Efficiency

A standard incandescent lightbulb's filament temperature is about 3000 K. (a) Find the wavelength of peak radiance, and (b) compare the radiance at 550 nm—the approximate center of the visible spectrum—with the peak radiance.

INTERPRET This problem involves radiation from an object at a known temperature, and we identify the filament as a blackbody. We're asked for both the peak wavelength and a comparison of radiances at two different wavelengths. We're implicitly adopting our definition of radiance as power emitted per unit area per unit wavelength, and our answers will reflect this choice.

DEVELOP Equation 34.2a gives the peak wavelength, and Equation 34.3 gives the radiance as a function of wavelength. So our plan for (a) is to solve Equation 34.2a, $\lambda_{\text{peak}}T = 2.898\ \text{mm}\cdot\text{K}$, for $T = 3000$ K. For (b) we'll form a ratio of radiances from Equation 34.3, using the result of (a) and the given 550-nm visible wavelength.

EVALUATE (a) Equation 34.2a gives $\lambda = 2.898\ \text{mm}\cdot\text{K}/3000\ \text{K} = 966$ nm. (b) To compare radiances, we form a ratio of the right-hand sides of Equation 34.3 evaluated at the wavelengths $\lambda_2 = 550$ nm and $\lambda_1 = 966$ nm. The numerators cancel, giving

$$\frac{R(\lambda_2, T)}{R(\lambda_1, T)} = \frac{\lambda_1^5(e^{hc/\lambda_1 kT} - 1)}{\lambda_2^5(e^{hc/\lambda_2 kT} - 1)} = 0.38$$

ASSESS Our 966-nm answer for (a) lies in the infrared, suggesting that incandescent lightbulbs aren't very efficient at producing visible light; that's the reason they're gradually being phased out. Our answer to (b) confirms this point, showing there's much less radiance in the visible than at the infrared peak. And remember that we've defined radiance as power per unit area per unit wavelength interval. If we adopt the more physical median wavelength given by $\lambda_{\text{median}}T = 4.11\ \text{mm}\cdot\text{K}$, we find $\lambda_{\text{median}} = 1.4\ \mu\text{m}$, well into the infrared. Since half the radiation occurs at wavelengths longer than this median, the bulb emits far more than half its energy as invisible infrared. ■

GOT IT? 34.1 Two identical blackbodies are heated until A's temperature is twice B's. Compare their total radiated power and their wavelengths of peak radiance.

34.3 Photons

Planck showed that vibrating molecules could exchange energy with electromagnetic radiation only in quantized bundles of size hf. Is the radiation's energy similarly quantized?

The Photoelectric Effect

In 1887 Heinrich Hertz observed that metals emit electrons when struck by light. Observations of this **photoelectric effect** continued with experiments involving metal electrodes in evacuated glass containers (Fig. 34.5). Illuminating one electrode causes it to emit electrons. Making the second electrode positive attracts the electrons, and the resulting current measures the rate at which electrons are ejected. Make the second electrode sufficiently negative, on the other hand, and the electron energy isn't great enough to overcome the repulsive potential; then the current ceases. This so-called *stopping potential* provides a measure of the ejected electrons' maximum kinetic energy $K_{\text{max}} = eV_s$.

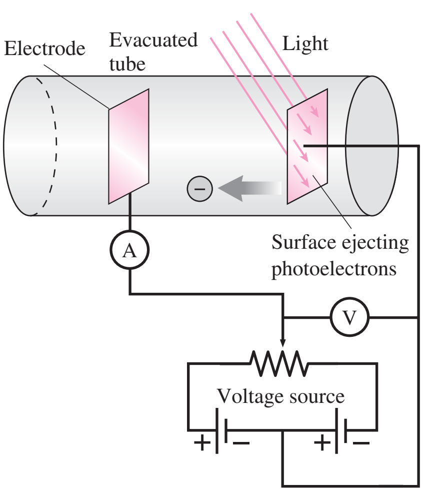


FIGURE 34.5 Apparatus for studying the photoelectric effect.

Classical physics suggests that the photoelectric effect should occur because an electron experiences a force in the oscillating electric field of a light wave. As the electron absorbs energy from the wave, the amplitude of its motion should grow until eventually it has enough energy to escape from the metal. Because the energy in a wave is spread throughout the entire wave, it might take a while for a single tiny electron to absorb enough energy. Increasing the light intensity should increase the electric field, resulting in the electron being ejected sooner and with more energy. Changing the wave frequency should have little effect.

The photoelectric effect does occur, but not in the way classical physics suggests. Figure 34.6 shows results from a photoelectric experiment, in the form of current versus voltage as read by the meters in Fig. 34.5. These results, along with observations made by varying the frequency of the incident light, show three major disagreements with the classical prediction:

1. Current begins immediately, showing that electrons are ejected immediately, even in dim light.
2. The maximum electron energy, as measured by the stopping potential V_s, is independent of the light intensity.
3. Below a certain cutoff frequency *no* electrons are emitted, no matter how intense the light. Above the cutoff frequency electrons are emitted with a maximum energy that increases in proportion to the light-wave frequency.

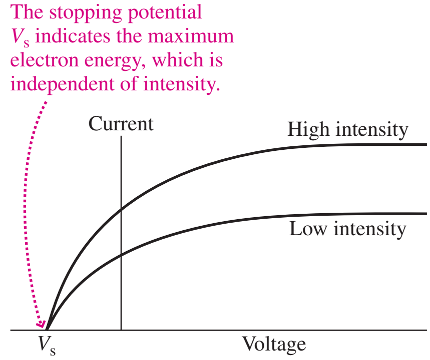


FIGURE 34.6 Current versus voltage for the photoelectric experiment of Fig. 34.5, shown for two light intensities at the same frequency.

In 1905, the same year he formulated the special theory of relativity, Albert Einstein proposed an explanation for the photoelectric effect. Einstein suggested that an electromagnetic wave's energy is concentrated in "bundles" called **quanta** or **photons**. Einstein applied to these photons the same energy-quantization condition that Planck had already proposed for molecular vibrations: that photons in light with frequency f have energy hf, where again h is Planck's constant:

$$E = hf \qquad \text{(photon energy)} \tag{34.6}$$

The more intense the light, the more photons—but the energy of each photon is unrelated to the light intensity.

Einstein's idea explains all three nonclassical aspects of the photoelectric effect. Each material has a minimum energy—called the *work function*, ϕ—required to eject an electron. (Table 34.1 lists work functions for selected elements.) Since the energy in a photon of light with frequency f is hf, the photons in low-frequency light have less energy than the work function and are therefore unable to eject electrons—no matter how many photons there are. At the cutoff frequency, the photon energy equals the work function, and the photons have just enough energy to eject electrons. As the frequency increases still further, the electrons emerge with maximum kinetic energy K equal to the difference between the photon energy and the work function:

$$K_{max} = hf - \phi \tag{34.7}$$

Thus, the electrons' maximum kinetic energy depends only on the photon energy—that is, on the light frequency but not on its intensity (Fig. 34.7). Finally, the immediate ejection of electrons follows because an individual photon delivers its entire bundle of energy to an electron all at once. Einstein received the 1921 Nobel Prize primarily for his explanation of the photoelectric effect rather than for his more controversial relativity theories. In 1914 Millikan, who had earlier demonstrated the quantization of electric charge, carried out meticulous photoelectric experiments that confirmed Einstein's hypothesis and helped earn Millikan the 1923 Nobel Prize.

Table 34.1 Work Functions

Element (Symbol)	ϕ (eV)
Silver (Ag)	4.26
Aluminum (Al)	4.28
Cesium (Cs)	2.14
Copper (Cu)	4.65
Potassium (K)	2.30
Sodium (Na)	2.75
Nickel (Ni)	5.15
Silicon (Si)	4.85

GOT IT? 34.2 If you replot Fig. 34.7 for a substance with a different work function, (a) will the slope of the line change? (b) Will the point at which it intersects the horizontal axis change?

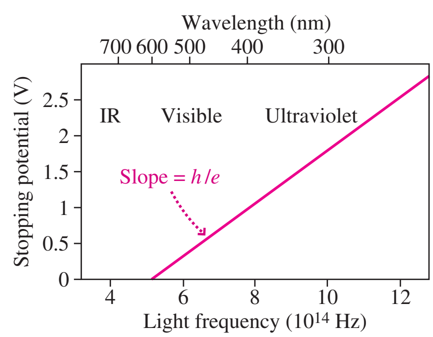


FIGURE 34.7 Results of a photoelectric experiment, showing stopping potential as a function of light frequency and wavelength. The stopping potential in volts is a direct measure of the electron energy in electronvolts.

Today, the photoelectric effect is used in extremely sensitive light detectors called **photomultipliers**. In these devices, electrons dislodge additional electrons in a series of electrodes called dynodes, resulting in a cascade of as many as a billion electrons for each incident photon (Fig. 34.8).

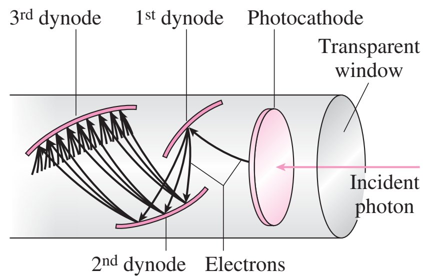


FIGURE 34.8 A photomultiplier produces a large pulse of electrons from a single incident photon.

EXAMPLE 34.2 The Photoelectric Effect: Designing a Photomultiplier

(a) Choose a suitable material from Table 34.1 for the light-sensitive surface in a photomultiplier that will respond to visible light at wavelengths of 575 nm and shorter. (b) Find the maximum kinetic energy of electrons ejected with the shortest-wavelength visible light, around 400 nm.

INTERPRET This problem is about the photoelectric effect. In (a) we're asked for a material in which 575-nm light can eject electrons. That means the work function can be no greater than the energy of a 575-nm photon. In (b) we need to find the excess electron energy, over the work function, for 400-nm light.

DEVELOP Equation 34.6, $E = hf$, relates the quantized energy of a photon to its frequency. Since $f\lambda = c$, we can rewrite Equation 34.6 as $E = hc/\lambda$. We'll use this to find the photon energy, and then we'll consult Table 34.1 for an appropriate material. Finally, we can use $f = c/\lambda$ in Equation 34.7, $K_{max} = hf - \phi$, to get the maximum kinetic energy of electrons when $\lambda = 400$ nm.

EVALUATE (a) At 575 nm, $E = hc/\lambda = 3.46\times10^{-19}$ J, or 2.16 eV, where 1 eV $= 1.6\times10^{-19}$ J. This energy must be enough for the electron to overcome the work function; the only material in Table 34.1 for which this is possible is cesium, with $\phi = 2.14$ eV. (b) At $\lambda = 400$ nm and $f = c/\lambda$, Equation 34.7 gives $K_{max} = hc/\lambda - \phi = 0.96$ eV. Here we converted the SI value of hc/λ to electronvolts before subtracting ϕ.

ASSESS Make sense? The work function we've chosen is just under the 2.16-eV photon energy at 575 nm, so electrons ejected with photons of this wavelength have negligible kinetic energy. The 400-nm minimum visible wavelength corresponds to roughly 50% higher frequency and therefore energy, or roughly 3 eV. It takes about 2 eV to overcome the work function, leaving about 1 eV of electron kinetic energy.

✓TIP Working with Electronvolts

Recall that 1 electronvolt (eV) is the *energy* gained by an electron across a 1-volt potential. So electronvolts are a unit of energy, but *not* the standard SI unit, which is the joule. We computed an energy $E = hc/\lambda$ in SI units and then converted to electronvolts using the factor 1 eV $= 1.6\times10^{-19}$ J. In general it's safest to work in SI units and then convert to eV as needed.

Waves or Particles?

In positing the existence of photons, Einstein gave the first inklings of the **wave–particle duality**—the seemingly dual nature of light, which acts in some situations like a wave and in others, as in the photoelectric effect, more like a localized particle. We now turn to another phenomenon that demonstrates light's particle-like aspect. Later we'll see how the wave–particle duality encompasses not only light but matter as well.

The Compton Effect

In 1923 the American physicist Arthur Holly Compton, at Washington University in St. Louis, did an experiment that dramatically confirmed the particle-like aspect of electromagnetic radiation. Although Compton's work came much later in the history of quantum theory than Einstein's, we include it here because it so strongly corroborates Einstein's photon hypothesis.

Compton was studying the interaction of X rays with electrons. Classically, an electron subject to an electromagnetic wave should undergo oscillatory motion, driven by the wave's oscillating electric field. Since accelerated charge is the source of electromagnetic waves, the electron should itself produce electromagnetic waves *of the same frequency as the incident waves* (Fig. 34.9*a*). As we saw in Section 29.7, the electron should radiate in all directions, with maximum radiation perpendicular to its oscillatory motion.

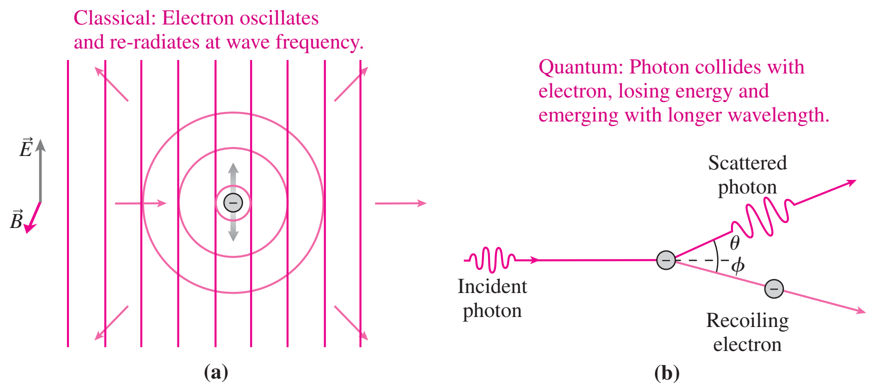


FIGURE 34.9 Classical and quantum descriptions of the interaction between electromagnetic waves and free electrons.

Compton and his coworkers measured the intensity of scattered X rays as a function of wavelength for different scattering angles. Remarkably, they found the greatest concentration of scattered X rays at a wavelength *longer* than that of the incident radiation (Fig. 34.10). They interpreted their results as implying that particle-like photons had collided with electrons, losing energy to the electrons and therefore, since $E = hf$, emerging with lower frequency and correspondingly longer wavelength (Fig. 34.9*b*).

We can understand this **Compton effect** by treating the interaction as an elastic collision between the incident photon and a stationary electron. The photon moves at c, so it's necessary to use relativistic expressions for energy and momentum. You can work out the details in Problem 73; the result gives the **Compton shift**: $\Delta\lambda = \lambda - \lambda_0$—that is, the change from the photon's original wavelength λ_0:

$$\Delta\lambda = \frac{h}{mc}(1 - \cos\theta) \qquad \text{(Compton shift)} \qquad (34.8)$$

Figure 34.10 shows that this equation is in excellent agreement with experimental data.

The term h/mc in Equation 34.8 is the **Compton wavelength** of the electron and gives the wavelength shift for a photon scattering at $\theta = 90°$. Its value is $\lambda_C = h/mc = 0.00243$ nm, or 2.43 pm. Equation 34.8 shows that the largest wavelength shift will be $2\lambda_C$, occurring at $\theta = 180°$. For the shift to be noticeable it should be a significant fraction of the incident wavelength, which therefore can't be too many times the Compton wavelength. For X rays, λ is in the range from approximately 10 pm to 10 nm, and therefore, detection of the Compton shift in X rays is already difficult. It would be totally impossible with visible light.

Today, Compton scattering with gamma rays is a widely used technique for studying the structure of matter. For example, abnormalities in human bone can be detected through Compton scattering of gamma rays emitted by a radioactive source embedded in bone. And the inverse Compton effect—the scattering of a rapidly moving electron off a photon—is a common process in high-energy astrophysical systems and is used in the laboratory to produce beams of gamma radiation.

The wavelength shift in Compton scattering admits no classical explanation. Coming after a decade of experimental and theoretical work that pointed increasingly to quantization as the essence of the atomic world, Compton's experimental results were for many physicists the convincing evidence for the reality of quanta.

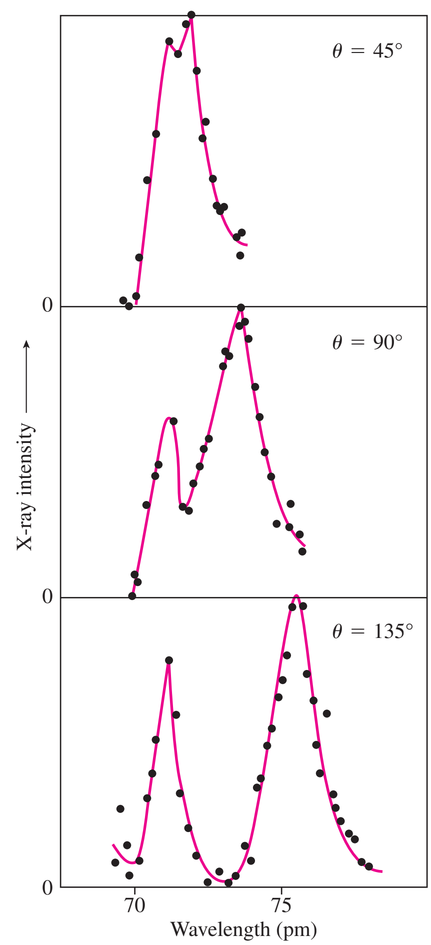


FIGURE 34.10 Compton's results for scattering of X rays with $\lambda = 71$ pm. Right-hand peak shows the wavelength shift of the Compton effect. The unshifted left-hand peak is from photons scattering off tightly bound atomic electrons, which don't absorb significant energy. Solid curves are theoretical predictions.

GOT IT? 34.3 Will the Compton wavelength shift be greater or less for photons of a given wavelength scattering off protons rather than electrons?

34.4 Atomic Spectra and the Bohr Atom

In Chapter 29 we found that accelerated charges are the source of electromagnetic radiation. By 1900 it was known that atoms contain negative electrons as well as regions of positive charge; by 1911 experiments by Ernest Rutherford and his colleague Hans Geiger and student Ernest Marsden had localized the positive charge in the tiny but relatively massive nucleus. According to classical physics, electrons should orbit the nucleus under the influence of the electric force, radiating electromagnetic wave energy as they accelerate in their orbits. In fact, a classical calculation shows that atomic electrons will quickly radiate away all their energy and spiral into the nucleus. Thus, the very existence of atoms is at odds with classical physics.

The Hydrogen Spectrum

A more subtle problem involving atoms dates to 1804, when William Wollaston noticed lines between some of the colors dispersed by a prism. Ten years later the German optician Josef von Fraunhofer dispersed the solar spectrum sufficiently that he could see hundreds of narrow, dark lines against the otherwise continuous spectrum. Studies of light emitted by diffuse gases excited by electric discharges show similar **spectral lines**, these bright against an otherwise dark background (recall Fig. 30.13). Such **emission spectra** result when atoms emit light of discrete frequencies. **Absorption spectra**, in contrast, arise when atoms in a diffuse gas absorb discrete frequencies of light from a continuous source. We emphasize the word "diffuse": Discrete spectra generally arise only when a gas is sufficiently diffuse that light from one atom stands a strong chance of escaping the gas before it interacts with other atoms. In dense gases multiple interactions result in the continuous spectrum of blackbody radiation.

Every element produces its own unique spectral lines, so analysis of spectra, even from the remote reaches of the cosmos, allows us to identify and characterize the material emitting the light. Spectral analysis led to the discovery of helium in the Sun's atmosphere before that element had been identified on Earth—hence the name, from the Greek word *helios* for Sun. Measuring the Doppler shift of spectral lines lets us "see" stars orbiting black holes in distant galaxies, and also gives direct evidence for the expansion of the universe. Back on Earth, the technique of atomic absorption spectroscopy uses spectral lines to determine the elemental composition of substances, helping identify pollutants or trace the flow of elements in biological samples.

In 1884, a Swiss schoolteacher named Johann Balmer realized that the wavelengths of the first four lines in the visible spectrum of hydrogen (see Fig. 30.13) were related by the equation

$$\frac{1}{\lambda} = R_{\mathrm{H}}\left(\frac{1}{2^2} - \frac{1}{n^2}\right)$$

where $n = 3, 4, 5, 6, \ldots$ and R_{H} is the **Rydberg constant** for hydrogen, with the approximate value $1.097\times10^7\ \mathrm{m}^{-1}$. Other series of lines in the hydrogen spectrum were soon found, and Balmer's equation was generalized to

$$\frac{1}{\lambda} = R_{\mathrm{H}}\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right) \tag{34.9}$$

where $n_1 = n_2 + 1, n_2 + 2, \ldots$. The Balmer series of lines has $n_2 = 2$; the Lyman series, in the ultraviolet, has $n_2 = 1$; and the infrared Paschen series has $n_2 = 3$. There are in fact infinitely many such series, corresponding to $n_2 = 1, 2, 3, \ldots$.

But why should atoms emit discrete spectral lines? And why should the hydrogen lines form patterns with the simple regularity of Equation 34.9?

The Bohr Atom

In 1913 the great Danish physicist Niels Bohr proposed an atomic theory that accounted for the spectral lines of hydrogen. In the **Bohr atom** the electron moves in a circular orbit about the nucleus, held by the electric force. Classically, any orbital radius and correspondingly any energy and angular momentum should be possible. But Bohr quantized the atom, stating that the only possible orbits were those with angular momentum an integer multiple of Planck's constant divided by 2π. Angular momentum quantization implies energy quantization, which, as we'll show, leads to Equation 34.9 for the hydrogen spectral lines.

Bohr asserted that an electron in an allowed orbit does not radiate energy, in contradiction to the predictions of classical electromagnetism. But an electron can jump from one orbit to another, emitting or absorbing a photon whose energy is equal to the energy difference between the two orbital levels. We can therefore find the expected photon energies—and the corresponding wavelengths—if we know the allowed energy levels.

To find the quantized atomic energy levels in Bohr's model, consider a hydrogen atom consisting of a fixed proton and an electron in circular orbit. Treating the proton as fixed is a good approximation because its mass is nearly 2000 times the electron's. We consider only electron speeds much less than that of light, which is an excellent approximation in hydrogen.

In Example 11.1 we found that the angular momentum of a particle with mass m and speed v, moving in a circular path of radius r, is mvr. Thus, Bohr's quantization condition reads

$$mvr = n\hbar \qquad \text{(quantization, Bohr atom)} \tag{34.10}$$

where $n = 1, 2, 3, \ldots$ and where we define $\hbar \equiv h/2\pi$ (read "h bar"). We need to relate the electron's angular momentum to its energy so we can find out what Equation 34.10 implies about energy quantization.

We studied circular orbits for the inverse-square force of gravity in Chapter 8, where we found that kinetic and potential energies in a circular orbit are related by $K = -\frac{1}{2}U$, with the zero of potential energy at infinity. The total energy $K + U$ is therefore $\frac{1}{2}U$. These results hold for any $1/r^2$ force, including the electric force. In the electric case the potential energy U is the point-charge potential of the proton, ke/r, multiplied by the electron charge, $-e$. Then the total energy is $E = \frac{1}{2}U = -ke^2/2r$. The minus sign means the electron is *bound* to the proton, in that it would take energy to separate them. Solving this equation for r then gives

$$r = -\frac{ke^2}{2E} \tag{34.11}$$

Since the kinetic energy is $K = -\frac{1}{2}U = -E$, we also have $\frac{1}{2}mv^2 = -E$ or $v = \sqrt{-2E/m}$. Using our expressions for r and v in the quantization condition 34.10 gives $m\sqrt{-2E/m}\left(-ke^2/2E\right) = n\hbar$. Solving for the energy E, we find

$$E = -\frac{k^2e^4m}{2\hbar^2n^2}$$

It's convenient to define the **Bohr radius**, a_0, as

$$a_0 = \frac{\hbar^2}{mke^2} = 0.0529\ \text{nm}$$

With this definition the energy becomes

$$E = -\frac{ke^2}{2a_0}\left(\frac{1}{n^2}\right) \qquad \text{(energy levels, Bohr atom)} \tag{34.12a}$$

Equation 34.12a gives us the allowed energy levels under Bohr's quantization condition. Evaluating this expression for the case $n = 1$ gives $E_1 = -2.18\times10^{-19}\ \text{J} = -13.6\ \text{eV}$; it's then convenient to write Equation 39.12a numerically in the form

$$E = -\frac{13.6\ \text{eV}}{n^2} \tag{34.12b}$$

where in both forms $n = 1, 2, 3, \ldots$. The lowest energy state, $n = 1$, is called the **ground state**; the others are **excited states**.

Now we have the allowed energy levels. What about spectra? When an electron jumps between energy levels, it emits or absorbs a photon whose energy hf is equal to the energy difference between the levels. So imagine an electron going from a higher level n_1 to a lower level n_2. The energy difference, according to Equation 34.12a, is

$$\Delta E = -\frac{ke^2}{2a_0}\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right) = \frac{ke^2}{2a_0}\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right)$$

and this is equal to the energy of the emitted photon. But the photon energy is $\Delta E = hf = hc/\lambda$, and therefore $1/\lambda = \Delta E/hc$ or, using our expression for ΔE,

$$\frac{1}{\lambda} = \frac{ke^2}{2a_0hc}\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right)$$

This looks just like Equation 34.9 for the hydrogen spectral lines, except that $ke^2/2a_0hc$ replaces the Rydberg constant R_H. Evaluating this quantity gives $ke^2/2a_0hc = 1.09\times10^7\ \mathrm{m}^{-1}$, which is very close to the experimentally observed Rydberg constant. The small discrepancy results from our approximation that the proton is stationary.

Bohr's theory of quantized angular momentum thus accounts brilliantly for the observed spectrum of hydrogen. We can understand the origin of the various spectral line series using Fig. 34.11, an **energy-level diagram** for the Bohr model of hydrogen. Allowed energy levels are shown as horizontal lines, and various possible transitions among levels as vertical arrows. Transitions with a common final state are grouped, and each group represents a different series of spectral lines.

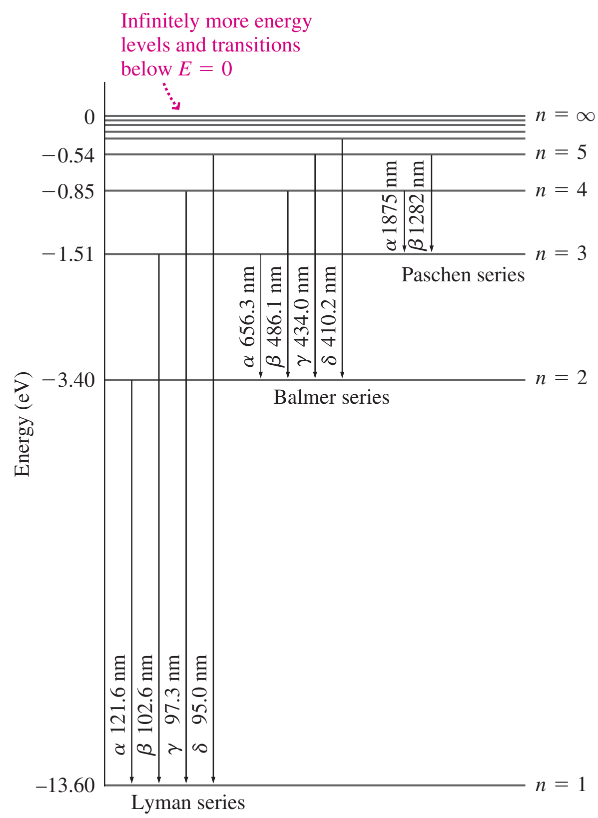


FIGURE 34.11 Energy-level diagram for the Bohr model of the hydrogen atom, showing transitions responsible for the first three series of spectral lines. Each series arises from jumps to a common final state.

Knowing the energy levels of Equation 34.12, we can also find the radii of the allowed electron orbits, as given by Equation 34.11:

$$r = -\frac{ke^2}{2E} = \left(\frac{ke^2}{2}\right)\left(\frac{2a_0n^2}{ke^2}\right) = n^2a_0 \qquad (34.13)$$

Thus, the lowest energy orbit has a radius of one Bohr radius, with higher-energy orbits growing rapidly with increasing n. A hydrogen atom in its ground state—$n = 1$— has a diameter of two Bohr radii, or about 0.1 nm. As we'll see in Chapter 35, the Bohr model's precise electron orbits aren't compatible with the fully developed theory of quantum mechanics; nevertheless, Equation 34.13 does give the approximate size of atoms.

EXAMPLE 34.3 The Bohr Model: Big Atoms

Hydrogen atoms are normally in their ground state, with diameter approximately 0.1 nm. But in the diffuse gas of interstellar space, atoms exist in highly excited states with sizes approaching a fraction of a millimeter. Such **Rydberg atoms** can also be produced temporarily in the lab. Transitions among Rydberg states result in photons at radio wavelengths. One of the longest wavelengths observed corresponds to a transition from $n = 273$ to $n = 272$. (a) What's the diameter of a hydrogen atom in the $n = 273$ state? (b) At what wavelength should a radio telescope be set to observe this transition?

INTERPRET This is a problem about electron transitions in hydrogen atoms, albeit of unusual size. The Bohr model applies.

DEVELOP We'll use Equation 34.13, $r = n^2 a_0$, to find the atomic diameter, with $n = 273$. Equation 34.9, $1/\lambda = R_H(1/n_2^2 - 1/n_1^2)$, will give the transition wavelength with $n_1 = 273$ and $n_2 = 272$.

EVALUATE (a) The diameter is twice the radius, so Equation 34.13 gives $d = (2)(273^2)a_0 = 7.9\ \mu\text{m}$. (b) Inverting Equation 34.9 to get the wavelength gives

$$\lambda = \left[R_H\left(\frac{1}{272^2} - \frac{1}{273^2}\right)\right]^{-1} = 92\ \text{cm}$$

with $R_H = 1.097\times10^7\ \text{m}^{-1}$.

ASSESS Our atomic diameter is some 75,000 times that of ground-state hydrogen and about the size of a red blood cell! A wavelength of 92 cm corresponds to a frequency $f = c/\lambda$ of about 300 MHz, which happens to lie in a gap between VHF TV channel 13 and UHF channel 14. ■

Equation 34.12 shows, and Fig. 34.11 suggests, that there are infinitely many electron energy levels between the ground state at -13.6 eV and zero energy. It's possible to give an atomic electron enough energy to bring it above the $E = 0$ level, but then it's no longer bound to the proton. Removing an electron is **ionization**, and Equation 34.12b and Fig. 34.11 show that it takes 13.6 eV to ionize a hydrogen atom in its ground state. This quantity is the **ionization energy**.

Limitations of the Bohr Model

Bohr's theory proved astoundingly successful in explaining the hydrogen spectrum. It also explains the spectra of hydrogen-like ions—atoms with all but one of their electrons removed—with the appropriate change in the value of the nuclear charge. And it has some success in predicting the spectra of atoms such as lithium and sodium that have a single valence electron beyond a group of more tightly bound electrons. But it can't account for the spectra of more complicated atoms, even two-electron helium. And with hydrogen, there are subtle spectral details that the Bohr model doesn't address. Furthermore, like Planck's original quantum hypothesis, Bohr's quantization of atomic energy levels lacked a convincing theoretical basis. We'll see in the next chapter how the much more comprehensive theory of quantum mechanics overcomes these limitations.

34.5 Matter Waves

In classical physics, light is purely a wave phenomenon. Einstein's photons gave light a particle-like quality as well. In 1923, ten years after Bohr's atomic theory, a French prince named Louis de Broglie (pronounced "de Broy") set forth a remarkable hypothesis in his doctoral thesis. If light has both wave-like and particle-like properties, he reasoned, why shouldn't matter also exhibit both properties?

We saw in Chapter 29 that light with energy E also carries momentum $p = E/c$. Combined with Equation 34.6, that means a photon of light with frequency f has

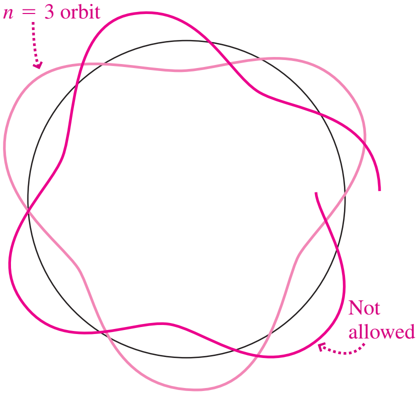


FIGURE 34.12 The allowed electron orbits in the Bohr atom are those that can fit an integral number of de Broglie wavelengths around the circular Bohr orbit.

momentum $p = hf/c$. Since $f\lambda = c$, the photon's momentum and wavelength are therefore related by

$$\lambda = \frac{h}{p} \qquad \text{(de Broglie wavelength)} \tag{34.14}$$

De Broglie proposed that this same relation should hold for particles of matter; at nonrelativistic speed, for example, an electron should have associated with it a **de Broglie wavelength** given by h/mv.

De Broglie used his matter-wave hypothesis to explain why atomic electron orbits are quantized. He proposed that Bohr's allowed orbits were those in which standing waves could exist (Fig. 34.12), in much the same way that a violin string can support only certain frequencies of standing waves. Suppose that n full wavelengths of a de Broglie electron wave fit around the circumference of the electron's circular orbit. Then we must have $n\lambda = n(h/p) = n(h/mv) = 2\pi r$, with r the orbit radius. Multiplying both sides by $mv/2\pi$ then gives $mvr = nh/2\pi = n\hbar$, which is Bohr's quantization condition. Thus, de Broglie's hypothesis provides a natural explanation for the quantization of atomic energy levels.

CONCEPTUAL EXAMPLE 34.1 The de Broglie Wavelength: Large and Small

If matter has wave properties, why don't we observe baseballs, cars, and people undergoing quantum interference?

EVALUATE Planck's constant h is tiny, and the masses of macroscopic objects are large. That makes the de Broglie wavelength (Equation 34.14) of macroscopic objects minuscule if they have any velocity whatsoever. Since wave behavior is evident only when waves interact with systems comparable in size to the wavelength, the wave aspect of macroscopic objects isn't evident. Even with subatomic particles, wave behavior isn't obvious at high velocities (i.e., high values of momentum p in the denominator of Equation 34.14). In the atom, though, it's a different story, as Making the Connection shows.

ASSESS Couldn't we make a macroscopic object's wavelength large by making its momentum mv small? Yes—but at normal temperatures, thermal agitation always means a significant random velocity. Only at very low temperatures can macroscopic systems exhibit quantum interference.

MAKING THE CONNECTION Find the de Broglie wavelength of (a) a 150-g baseball pitched at 45 m/s and (b) an electron moving at 1 Mm/s. Compare your results with the sizes of home plate and an atom, respectively.

EVALUATE Given mass and speed, Equation 34.14 becomes $\lambda = h/mv$. This gives $\lambda_{\text{baseball}} \simeq 10^{-34}$ m, unimaginably smaller than home plate. But $\lambda_{\text{electron}} \simeq 0.7$ nm, several times the size of an atom. Therefore, wave effects dominate this electron's interactions with atoms.

APPLICATION The Electron Microscope

In Chapter 32 we found that light can't sharply image objects whose size is on the order of the wavelength or smaller—a factor that limits the resolving power of conventional microscopes. But Equation 34.14 shows that we can control the wavelength of electrons by adjusting their speed—and therefore we can achieve electron wavelengths much shorter than that of light. The **electron microscope** exploits this effect, providing resolutions down to about 1 nm and magnifications of 10^6.

Electron microscopes accelerate electron beams to energies of 50–100 keV, with corresponding wavelengths of about 0.005 nm. Magnetic fields act as focusing lenses, forming an image of whatever object is placed in the beam path. An electronic detector reads the image, which is then displayed on a screen.

Electron microscopes are indispensable tools in biology, chemistry, and materials science. A related device, the scanning electron microscope, produces dramatic three-dimensional images at magnifications of $10-10^5$, as shown in the photo of an ant carrying a microelectronic chip.

Electron Diffraction and Matter-Wave Interference

In 1927 the American physicists Clinton Davisson and Lester Germer gave a convincing verification of de Broglie's matter-wave hypothesis. Davisson and Germer were studying the interaction of an electron beam with a nickel crystal, and they noticed regular intensity peaks reminiscent of X-ray diffraction. Shortly afterward, the Scottish physicist George Thomson observed electron diffraction directly, further evidence of the electron's wave nature (Fig. 34.13). Thomson was the son of J. J. Thomson, who had discovered the electron as a particle in 1897. Together their work captured the electron's wave–particle duality. Today, experiments with entire atoms and even larger clusters of matter exhibit wave interference—as shown in this chapter's opening photo.

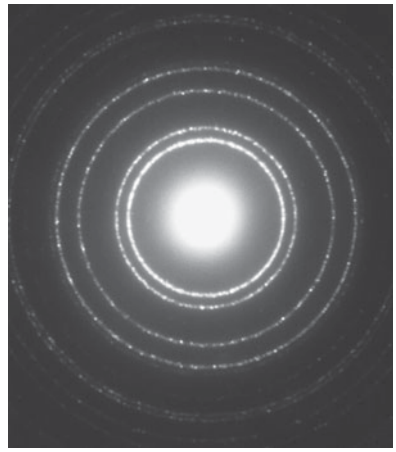


FIGURE 34.13 Diffraction produced by passing an electron beam through a circular aperture shows that electrons have a wave-like character.

34.6 The Uncertainty Principle

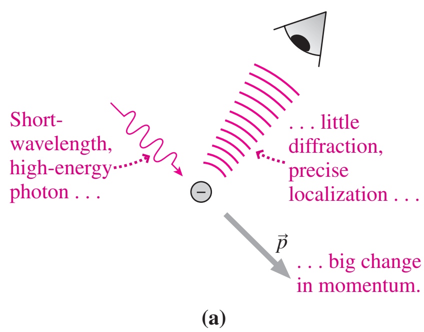
In classical physics it's possible, in principle, to know the exact position and velocity of a particle and therefore to predict with certainty its future behavior. But not so in quantum physics! In 1927 the German physicist Werner Heisenberg presented his famous **uncertainty principle**, which states that some pairs of quantities cannot be measured simultaneously with arbitrary precision. Position and momentum constitute one such pair; if we measure a particle's position to within an uncertainty Δx, then we can't simultaneously determine its momentum to an accuracy better than Δp, where

$$\Delta x \, \Delta p \geq \hbar \qquad (\text{uncertainty principle}) \tag{34.15}$$

Why this limitation? The fundamental reason is quantization. To measure some property of a system requires interacting with the system—for example, shining light on it. Interaction involves energy, and the interaction energy disturbs the system slightly. As a result, values inferred from the measurement are no longer quite right. In classical physics the energy can be arbitrarily small, resulting in a negligible disturbance. But in quantum theory the minimum energy is a single quantum, like a photon of light, and thus the disturbance can't be arbitrarily small.

So why not use lower-frequency light, whose photon energy hf is lower? Because lower frequency means longer wavelength and, as we found in Chapter 32, diffraction effects limit resolution at longer wavelengths. Heisenberg summarized this dilemma with the "thought experiment" illustrated in Fig. 34.14, which uses a single photon to observe an electron. A short-wavelength photon allows precise localization of the electron (Fig. 34.14*a*). But short wavelength means high frequency and thus high photon energy. The high-energy photon imparts considerable momentum to the electron, and thus the very act that fixes the electron's position degrades our knowledge of its momentum. We can decrease this disturbance with a lower-energy, longer-wavelength photon (Fig. 34.14*b*). But now diffraction precludes precisely determining the electron's position. So we can measure the electron's position accurately, at the expense of knowing its momentum. Or we can measure its momentum accurately, but then we can't know its position. With a photon of intermediate wavelength we could measure both quantities, but neither precisely. The uncertainty principle, Equation 34.15, quantifies this trade-off.

The uncertainty principle is intimately connected with de Broglie's wave hypothesis. Suppose we pass an electron beam through a slit, as shown in Fig. 34.15 (next page). Then we know the electrons' vertical position to within the slit width. If the slit is much wider than the electrons' de Broglie wavelength, there's minimal diffraction. The electrons follow straight lines and we're quite sure of their vertical momentum, in this case zero (Fig. 34.15*a*). But with a wide slit we don't know much about the electrons' vertical position. Making the slit smaller gives a more precise position, but then diffraction spreads the beam, increasing the uncertainty in the electrons' vertical momentum (Fig. 34.15*b*). So the wave nature of matter ultimately imposes a trade-off: The more we know of a particle's position, the less we know of its momentum, and vice versa.



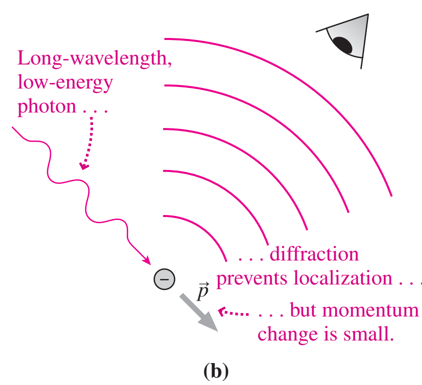


FIGURE 34.14 Heisenberg's "quantum microscope" thought experiment.

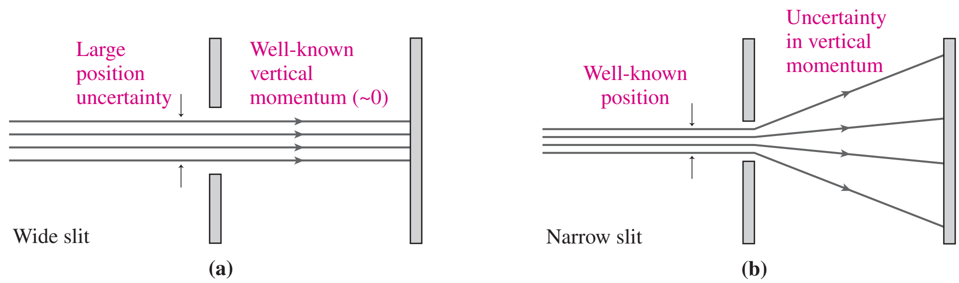


FIGURE 34.15 The wave nature of matter is intimately related to the uncertainty principle, as shown here for beams of electrons passing through wide and narrow slits. In (b), diffraction introduces the uncertainty in vertical momentum.

EXAMPLE 34.4 The Uncertainty Principle: Microelectronics

A beam of aluminum atoms is used to dope a semiconductor chip to set its electrical properties. If the atoms' velocity is known to within 0.2 m/s, how accurately can they be positioned?

INTERPRET This is a question about simultaneously knowing the atoms' position and velocity—paired quantities governed by the uncertainty principle.

DEVELOP We're given the velocity uncertainty Δv, from which we'll find the momentum uncertainty $\Delta p = m\,\Delta v$. Then we can use the uncertainty principle, Equation 34.15, $\Delta x\,\Delta p \geq \hbar$, to find the uncertainty Δx in position. To find the mass, we'll need aluminum's atomic weight from Appendix D, along with Appendix C's conversion from unified mass units (u) to kilograms.

EVALUATE We have

$$\Delta p = m\,\Delta v = (26.98\,\text{u})(1.66 \times 10^{-27}\,\text{kg/u})(0.20\,\text{m/s})$$
$$= 9\times10^{-27}\,\text{kg}\cdot\text{m/s}$$

Then Equation 34.15 gives the position uncertainty:

$$\Delta x = \frac{\hbar}{\Delta p} = 12\,\text{nm}$$

where, again, $\hbar = h/2\pi$.

ASSESS Our 12-nm answer is about 100 atomic diameters and shows that the uncertainty principle constrains our ability to fabricate very small microelectronic structures. ■

It sounds like the uncertainty principle only limits our knowledge. But in fact it proves useful in estimating the size and energies of atomic-scale systems, as the next example shows.

EXAMPLE 34.5 The Uncertainty Principle: Estimating Atomic and Nuclear Energies

Use the uncertainty principle to estimate the minimum energy possible for (a) an electron confined to a region of atomic dimensions, about 0.1 nm, and (b) a proton confined to a region of nuclear dimensions, about 1 fm.

INTERPRET We're given the uncertainty in position; that's the width of the region in which the particles are confined. The particles can't be at rest, or we'd know that their momentum was exactly zero—in violation of the uncertainty principle. So they must have a minimum momentum and therefore energy. We're asked to find that energy.

DEVELOP We need to find the minimum momentum consistent with the uncertainty principle, and from it the energy. Suppose a particle has momentum of magnitude p, but we don't know its direction. It could be going one way, with momentum p, or the other way, with momentum $-p$. Then the momentum itself is uncertain by $\Delta p = p - (-p) = 2p$. The uncertainty principle says $\Delta p \geq \hbar/\Delta x$, so there's a minimum magnitude for the momentum given by $p \geq \hbar/2\,\Delta x$. Using $p = mv$ and $K = \frac{1}{2}mv^2$ gives $K = p^2/2m$, and therefore the uncertainty principle requires

$$K \geq \frac{1}{2m}\left(\frac{\hbar}{2\,\Delta x}\right)^2$$

EVALUATE Evaluating this constraint for an electron with $\Delta x = 0.1$ nm and for a proton with $\Delta x = 1$ fm gives minimum energies of about 1 eV and 5 MeV, respectively.

ASSESS Energies in electronvolts are typical of atomic-scale systems, as we saw in Fig. 34.11. Our result shows that nuclear energies are some 5 million times greater—indicating the dramatic difference between chemical and nuclear energy sources. ■

Energy–Time Uncertainty

A second pair of variables that defy simultaneous measurement are the energy of a system and the time it remains at that energy. The energy uncertainty ΔE is related to the time Δt through the inequality

$$\Delta E\,\Delta t \geq \hbar \tag{34.16}$$

One effect of energy–time uncertainty is to render atomic and nuclear energy levels inexact and therefore to broaden spectral lines. If an atom were forever in a fixed energy state, we could take infinitely long to measure its energy and therefore make ΔE arbitrarily small. But excited states of atoms have characteristic lifetimes (typically $\sim 10^{-8}$ s), which limit the measurement time and therefore set a minimum uncertainty in the energy level. Problem 66 and Passage Problems 79–82 explore energy–time uncertainty.

Observers, Uncertainty, and Causality

The uncertainty principle moves the observer from a passive onlooker to an active participant in physical events. To observe is necessarily to disturb, and quantum theory is therefore concerned with the role of the observer and the process of measurement. The uncertainty principle is fundamentally a statement about what can and cannot be learned through measurement.

Position and momentum cannot be measured simultaneously with perfect accuracy. Surely, though, a particle has well-defined values of both, even though we can't know them? The answer seems to be no. The standard interpretation of quantum mechanics suggests that it makes no sense to talk about what can't be measured, and recent experiments have ruled out "hidden variables" that might be active at a lower level to guide the particle in a deterministic path. Its wave aspect makes a particle a "fuzzy" thing, and it really makes no sense to think of it as a tiny ball with definite momentum and position. For that reason it also makes no sense to think of the particle's future as being fully determined in the sense that Newton's laws determine the future path of, say, a baseball. We're left with uncertainty—or indeterminacy, as Heisenberg's word also translates—as a fundamental fact of our universe.

34.7 Complementarity

One of the most disturbing aspects of quantum theory is the wave–particle duality—the seeming contradiction that matter and light have both wave-like and particle-like properties. If this bothers you, you're in good company: Heisenberg himself expressed frustration in trying to understand the quantum world:

> I remember discussions with Bohr which went through many hours till very late at night and ended almost in despair; and when at the end of the discussion I went alone for a walk in the neighboring park I repeated to myself again and again the question: Can nature possibly be as absurd as it seems to us in these atomic experiments?*

Bohr dealt with the wave–particle duality through his **principle of complementarity**. The wave and particle pictures, he said, are complementary aspects of the same reality. If we do an experiment to measure a wave-like property—for example, the diffraction of electrons—then we find wave properties but not particle properties. If we do an experiment to measure a particle-like property—for example, localizing an electron—then we won't find wave properties. The two measurements require different experiments, and we can't perform both simultaneously on the same entity. So we'll never catch wave and particle in an outright contradiction, and the answer to the question "Which is it, wave or particle?" has to be that it's both, and which you find depends on what experiment you choose to perform.

Bohr articulated a second principle that helps reconcile the seeming contradiction between classical and quantum physics. His **correspondence principle** states that the predictions of classical and quantum physics should agree in situations where the size of individual quanta is negligible. Taking $h \to 0$ in Planck's law, for example, gives the classical Rayleigh–Jeans law (see Problem 68). Or, for large n, the energies of adjacent atomic states in the Bohr model become so close that the levels appear essentially as a continuum of allowed energies—as expected in classical physics. Or consider a 1000-W radio beam; the photon energy hf is so low that the beam contains an enormous number of photons per unit beam length, and we can consider the energy distributed essentially continuously over the beam. But in a 1000-W X-ray beam, the photon energy is much higher and the number of photons correspondingly fewer; it's therefore difficult to avoid the fundamental fact of energy quantization. Visible light lies somewhere in between; we can often treat its energy as being continuously distributed, except when it interacts with systems as small as individual atoms.

*Werner Heisenberg, *Physics and Philosophy: The Revolution in Modern Science* (New York: Harper & Brothers, 1962).

CHAPTER 34 SUMMARY

Big Picture

The big ideas here are at the heart of quantum physics—a radically different view of reality at the atomic scale. **Quantization** means that some physical quantities—often including energy—come only in discrete values. Another fundamental aspect of quantum reality is **wave–particle duality**, wherein light and matter exhibit both wave-like and particle-like aspects. Bohr's **complementarity principle** precludes these ever being in direct conflict. Finally, quantization and wave–particle duality lead to the **uncertainty principle**, which states that it's impossible to measure simultaneously and with arbitrary precision a particle's position and momentum.

Key Concepts and Equations

Planck's constant,

$$h = 6.63\times10^{-34}\ \mathrm{J\cdot s}$$

sets the fundamental scale of quantization.

It's also expressed as "h bar":

$$\hbar = h/2\pi$$

The energy of electromagnetic radiation with frequency f is quantized in **photons** with energy

$$E = hf$$

Electron energies in the **Bohr model** of hydrogen are quantized according to

$$E = -\frac{ke^2}{2a_0}\left(\frac{1}{n^2}\right) \simeq -\frac{13.6\ \mathrm{eV}}{n^2}$$

where n is an integer and $a_0 = 0.0529$ nm is the **Bohr radius**.

The **de Broglie wavelength** of a particle with momentum p is

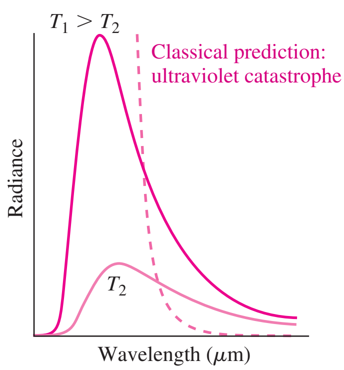
$$\lambda = \frac{h}{p}$$

The **uncertainty principle** relates uncertainties in position and momentum:

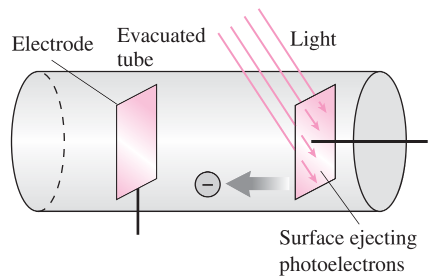
$$\Delta x\, \Delta p \geq \hbar$$

Applications

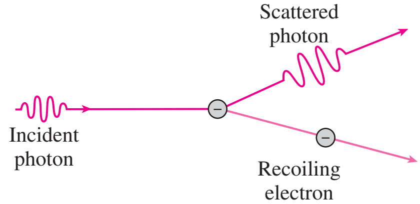
A correct description of **blackbody radiation** requires Planck's quantization hypothesis. The peak radiance—energy radiated per unit wavelength interval—from a blackbody at temperature T occurs at a wavelength given by $\lambda T = 2.898\ \mathrm{mm\cdot K}$.



The **photoelectric effect** involves the ejection of electrons from a metal surface illuminated with electromagnetic waves. Explanation of the effect led Einstein to propose **photons** as the quanta of electromagnetic-wave energy.



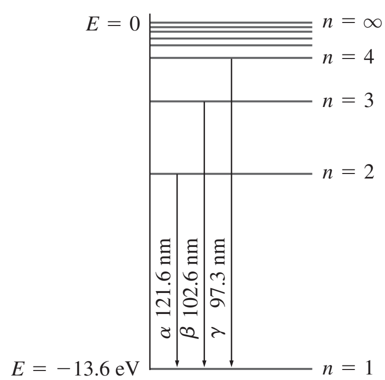
The **Compton effect** shows that photons interact with free electrons exactly like colliding particles, losing energy and emerging with longer wavelength.



Quantization of atomic energy levels leads directly to **atomic spectra**. In the Bohr model of hydrogen, the spectral line produced in a transition from the n_1 to the n_2 energy level is given by

$$\frac{1}{\lambda} = R_{\mathrm{H}}\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right)$$

where $R_{\mathrm{H}} = 1.097\times10^{7}\ \mathrm{m}^{-1}$.



For Thought and Discussion

1. Why does classical physics predict that atoms should collapse?
2. Looking at the night sky, you see one star that appears red, another yellow, and another blue. Compare their temperatures.
3. Imagine an atom that, unlike hydrogen, had only three energy levels. If these levels were evenly spaced, how many spectral lines would result? How would their wavelengths compare?
4. What colors of visible light have the highest-energy photons?
5. Why is the immediate ejection of electrons in the photoelectric effect surprising from a classical viewpoint?
6. Suppose the Compton effect were significant at radio wavelengths. What problems might this present for radio and TV broadcasting?
7. How are the uncertainty principle and wave–particle duality related?
8. How many spectral lines are in the entire Balmer series?
9. Why are the lines of the Lyman series in the ultraviolet while some Balmer lines are in the visible?
10. Why does the photoelectric effect suggest that light has particle-like properties?
11. Energy–time uncertainty limits the precision with which we can know the mass of unstable particles (those that decay after a finite time). Why?
12. If you measure a particle's position with perfect accuracy, what do you know about its momentum?
13. How might our everyday experience be different if Planck's constant had the value 1 J·s?
14. Why are the energies given by Equations 34.12 negative?

Exercises and Problems

Exercises

Section 34.2 Blackbody Radiation

15. If you double a blackbody's temperature, by what factor does its radiated power increase?
16. The surface temperature of the star Rigel is 10^4 K. Find (a) the power radiated per square meter of its surface, (b) its λ_{peak}, and (c) its λ_{median}.
17. Find λ_{peak} and λ_{median} for Earth, considered a 288-K blackbody.
18. Spacecraft instruments measure the radiation from an asteroid, and the data show that the power per unit wavelength peaks at 40 μm. Assuming the asteroid is a blackbody, find its surface temperature.
19. The Sun approximates a blackbody at 5800 K. (a) Find the wavelength of peak radiance on the per-unit-wavelength basis implicit in Equation 34.2a. (b) Find the median wavelength, below which half the radiation is emitted (Equation 34.2b). Identify the spectral region of each.

Section 34.3 Photons

20. Find the energy in electronvolts of (a) a 1.0-MHz radio photon, (b) a 5.0×10^{14}-Hz optical photon, and (c) a 3.0×10^{18}-Hz X-ray photon.
21. BIO The human eye is sensitive to wavelengths from about 400 to 700 nm. What's the corresponding range of photon energies?
22. A microwave oven uses electromagnetic radiation at 2.4 GHz. (a) What's the energy of each microwave photon? (b) At what rate does a 900-W oven produce photons?
23. A red laser at 650 nm and a blue laser at 450 nm emit photons at the same rate. How do their total power outputs compare?
24. Find the maximum work function for a surface to emit electrons when illuminated with 900-nm infrared light.

Section 34.4 Atomic Spectra and the Bohr Atom

25. Calculate the wavelengths of the first three lines in the Lyman series for hydrogen.
26. Which spectral line of the hydrogen Paschen series $(n_2 = 3)$ has wavelength 1282 nm?
27. What's the maximum wavelength of light that can ionize hydrogen in its ground state? In what spectral region is this?
28. At what energy level does the Bohr hydrogen atom have diameter 5.18 nm?

Section 34.5 Matter Waves

29. Find the de Broglie wavelength of (a) Earth, orbiting the Sun at 30 km/s, and (b) an electron moving at 10 km/s.
30. How slowly must an electron be moving for its de Broglie wavelength to be 1 mm?
31. A proton and electron have the same de Broglie wavelength. How do their speeds compare, assuming $v \ll c$ for both?
32. Find the de Broglie wavelength of electrons with kinetic energies (a) 10 eV, (b) 1.0 keV, and (c) 10 keV.

Section 34.6 The Uncertainty Principle

33. A proton is confined to a space 1 fm wide (about the size of an atomic nucleus). What's the minimum uncertainty in its velocity?
34. Is it possible to determine an electron's velocity accurate to ±1 m/s while simultaneously finding its position to within ±1 μm? What about a proton?
35. A proton has velocity $v = (1500 \pm 0.25)$ m/s. What's the uncertainty in its position?
36. An electron is moving in the $+x$-direction with speed measured at 50 Mm/s, accurate to $\pm10\%$. What's the minimum uncertainty in its position?
37. Find the minimum energy for a neutron in a uranium nucleus whose diameter is 15 fm.

Problems

38. Find the power per unit area emitted by a 3000-K incandescent lamp filament in the wavelength interval from 500 nm to 502 nm.
39. Treating the Sun as a 5800-K blackbody, compare its UV radiance at 200 nm with its visible radiance at its 500-nm peak wavelength.
40. For a 2.0-kK blackbody, by what percentage is the Rayleigh–Jeans law in error at wavelengths of (a) 1.0 mm, (b) 10 μm, and (c) 1.0 μm?
41. The radiance of a blackbody peaks at 660 nm. (a) What's its temperature? (b) How does its radiance at 400 nm compare with that at 700 nm?
42. (a) Find the Compton wavelength for a proton. (b) Find the energy in electronvolts of a gamma ray whose wavelength equals the proton's Compton wavelength.
43. Find the rate of photon production by (a) a radio antenna broadcasting 1.0 kW at 89.5 MHz, (b) a laser producing 1.0 mW of 633-nm light, and (c) an X-ray machine producing 0.10-nm X rays with total power 2.5 kW.
44. Electrons in a photoelectric experiment emerge from an aluminum surface with maximum kinetic energy 1.3 eV. Find the wavelength of the illuminating radiation.

45. (a) Find the cutoff frequency for the photoelectric effect in copper. (b) Find the maximum energy of the ejected electrons if the copper is illuminated with light of frequency 1.8×10^{15} Hz.
46. The stopping potential in a photoelectric experiment is 1.8 V when the illuminating radiation has wavelength 365 nm. Determine (a) the work function of the emitting surface and (b) the stopping potential for 280-nm radiation.
47. Chlorophyll is a photosynthetic molecule common in green **BIO** plants. On a per-unit-wavelength basis, its ability to absorb visible light has two peaks, at 430 nm and 662 nm. (a) Find the corresponding photon energies. (b) Use these peak wavelengths to explain why plants are green.
48. Find the initial wavelength of a photon that loses half its energy when it Compton-scatters from an electron and emerges at 90° to its initial direction.
49. When light shines on potassium, the photoelectrons' maximum speed is 4.2×10^5 m/s. Find the light's wavelength.
50. The maximum electron energy in a photoelectric experiment is 2.8 eV. When the wavelength of the illuminating radiation is increased by 50%, the maximum energy drops to 1.1 eV. Find (a) the work function of the emitting surface and (b) the original wavelength.
51. A 150-pm X-ray photon Compton-scatters off an electron and emerges at 135° to its original direction. Find (a) the wavelength of the scattered photon and (b) the electron's kinetic energy.
52. Find the kinetic energy of an initially stationary electron after a 0.10-nm X-ray photon scatters from it at 90°.
53. A photocathode ejects electrons with maximum energy 0.85 eV when illuminated with 430-nm blue light. Will it eject electrons when illuminated with 633-nm red light? If so, what will be the maximum electron energy?
54. (a) Find the highest possible energy for a photon emitted as the electron jumps between two adjacent energy levels in the Bohr hydrogen atom. (b) Which energy levels are involved?
55. Find (a) the wavelength and (b) the energy in electronvolts of the photon emitted when a Rydberg hydrogen atom drops from the $n = 180$ level to the $n = 179$ level.
56. The wavelengths of a spectral line series tend to a limit as $n_1 \to \infty$. Evaluate the series limit for (a) the Lyman series and (b) the Balmer series in hydrogen.
57. A Rydberg hydrogen atom makes a downward transition to the $n = 225$ state, emitting a 9.32-μeV photon. What was the original state?
58. A hydrogen atom is in its ground state when its electron absorbs a 48-eV photon. What's the energy of the resulting free electron?
59. How much energy does it take to ionize a hydrogen atom in its first excited state?
60. Ultraviolet light with wavelength 75 nm shines on hydrogen atoms in their ground states, ionizing some of the atoms. What's the energy of the electrons freed in this process?
61. Helium with one of its two electrons removed acts very much like hydrogen, and the Bohr model successfully describes it. Find (a) the radius of the ground-state electron orbit and (b) the photon energy emitted in a transition from the $n = 2$ to the $n = 1$ state in this singly ionized helium.
62. Through what potential difference should you accelerate an electron from rest so its de Broglie wavelength will be the size of a hydrogen atom, about 0.1 nm?
63. Find the minimum electron speed that would make an electron microscope superior to an optical microscope using 450-nm light.
64. You're a cell biologist who wants to image microtubules that **BIO** form the "skeletons" of living cells. The microtubules are 25 nm in diameter, and, as Chapter 32 shows, you need to image with waves whose wavelength is at least this small. You can use either an inexpensive electron microscope that accelerates electrons to kinetic energies of 40 keV, or a more expensive unit that produces 100-keV electrons. Will the less expensive microscope work?
65. An electron is trapped in a "quantum well" 20 nm wide. Find its minimum possible speed.
66. Typically, an atom remains in an excited state for about 10^{-8} s before it drops to a lower state, emitting a photon in the process. What's the uncertainty in the energy of this transition?
67. An electron is moving at 10^6 m/s and you wish to measure its energy to an accuracy of $\pm 0.01\%$. What's the minimum time necessary for this measurement?
68. Use the series expansion for e^x (Appendix A) to show that Planck's law (Equation 34.3) reduces to the Rayleigh–Jeans law (Equation 34.5) when $\lambda \gg hc/kT$.
69. A photon's wavelength is equal to the Compton wavelength of a particle with mass m. Show that the photon's energy is equal to the particle's rest energy.
70. Show that the frequency range of the hydrogen spectral line series involving transitions ending at the nth level is $\Delta f = cR_{\mathrm{H}}/(n + 1)^2$.
71. A photon undergoes a 90° Compton scattering off a stationary electron, and the electron emerges with *total* energy $\gamma m_e c^2$, where γ is the relativistic factor introduced in Chapter 33. Find an expression for the initial photon energy.
72. Show that Wien's law (Equation 34.2a) follows from Planck's law (Equation 34.3). (*Hint:* Differentiate Planck's law with respect to wavelength.)
73. Consider an elastic collision between a photon with initial wavelength λ_0 moving in the x-direction and a stationary electron, as depicted in Fig. 34.9*b*. Use relativistic expressions for energy and momentum from Chapter 33 to show that conservation of energy and momentum yield the equations $hc/\lambda_0 + mc^2 = hc/\lambda + \gamma mc^2$, $h/\lambda_0 = (h/\lambda)\cos\theta + \gamma mu\cos\phi$, and $0 = (h/\lambda)\sin\theta - \gamma mu\sin\phi$, where λ is the post-collision photon wavelength and the angles θ and ϕ are as shown in Fig. 34.9*b*. Solve these equations to find the Compton shift (Equation 34.8).
74. What would the constant in Equation 34.2a be if blackbody radi- **C** ance were defined for fixed intervals of frequency rather than wavelength? (*Hint:* Use $\lambda = c/f$ to express the radiance as $R(f, T)$, then differentiate to find the maximum, and solve the resulting relation numerically. Express your answer in a form like Equations 34.2a and b.)
75. Integrate Equation 34.3 over all wavelengths to get the total power radiated per unit area. Show that your result is equivalent to Equation 34.1, with the Stefan–Boltzmann constant given by $\sigma = 2\pi^5k^4/15c^2h^3$. (*Hint:* Use $hc/\lambda kT$ as the integration variable.)
76. Perform a numerical integration of Equation 34.3 to the wave- **C** length given by Equation 34.2b. Divide by the result of Problem 75, and thus verify that Equation 34.2b gives the wavelength above and below which a blackbody radiates half its energy.
77. Use the momentum conservation equations in Problem 73 and Equation 34.8 for the Compton shift to show that the electron's recoil angle in Fig. 34.9*b* is given by $\tan\phi = \sin\theta/(1 + \lambda_{\mathrm{C}}/\lambda_0(1 - \cos\theta))$.
78. Show that in the Bohr model, the frequency of a photon emitted in a transition between levels $n + 1$ and n, in the limit of large n, is equal to the electron's orbital frequency. (This is an example of Bohr's correspondence principle.)

Passage Problems

Particle physicists use the energy–time uncertainty relation to estimate the lifetimes of unstable particles produced in high-energy particle accelerators (Chapter 39). Some particles have lifetimes of 10^{-24} s and shorter—impossible to measure directly. However, physicists can measure particle masses, and they do so for many instances of the same particle to get a distribution of masses. By Einstein's $E = mc^2$, that corresponds to a distribution of energies (Fig. 34.16). Measuring the distribution's width at half its peak (see Fig. 34.16) gives an estimate of the energy uncertainty, and the corresponding Δt from inequality 34.16 provides the particle's lifetime.

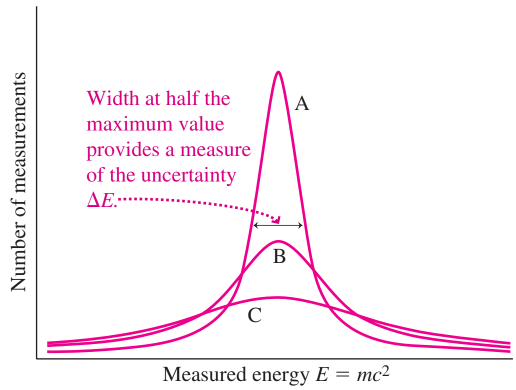


FIGURE 34.16 Mass distributions for high-energy particles (Passage Problems 79–82). The vertical axis gives the number of measurements that yield a given value on the horizontal axis

79. Which of the curves in Fig. 34.16 represents the particle with the shortest lifetime?
    a. A
    b. B
    c. C
    d. You can't tell from the graph.
80. An energy uncertainty of 1 MeV corresponds to a particle lifetime closest to
    a. 10^{-34} s.
    b. 10^{-21} s.
    c. 10^{-9} s.
    d. 1 μs.
81. The converse approach is used for particles with longer lifetimes: Direct measurement of the lifetime yields, through energy–time uncertainty, a range of expected values for particle energies or masses. The longer the lifetime,
    a. the wider the mass range and the narrower the energy range.
    b. the wider the mass and energy ranges.
    c. the narrower the mass range and the wider the energy range.
    d. the narrower the mass and energy ranges.
82. For a particle with lifetime 10^{-7} s, the corresponding mass range is closest to
    a. 10^{-44} u.
    b. 10^{-27} u.
    c. 10^{-17} u.
    d. 1 u.

Answers to Chapter Questions

Answer to Chapter Opening Question

That matter, like light, behaves as waves under some circumstances.

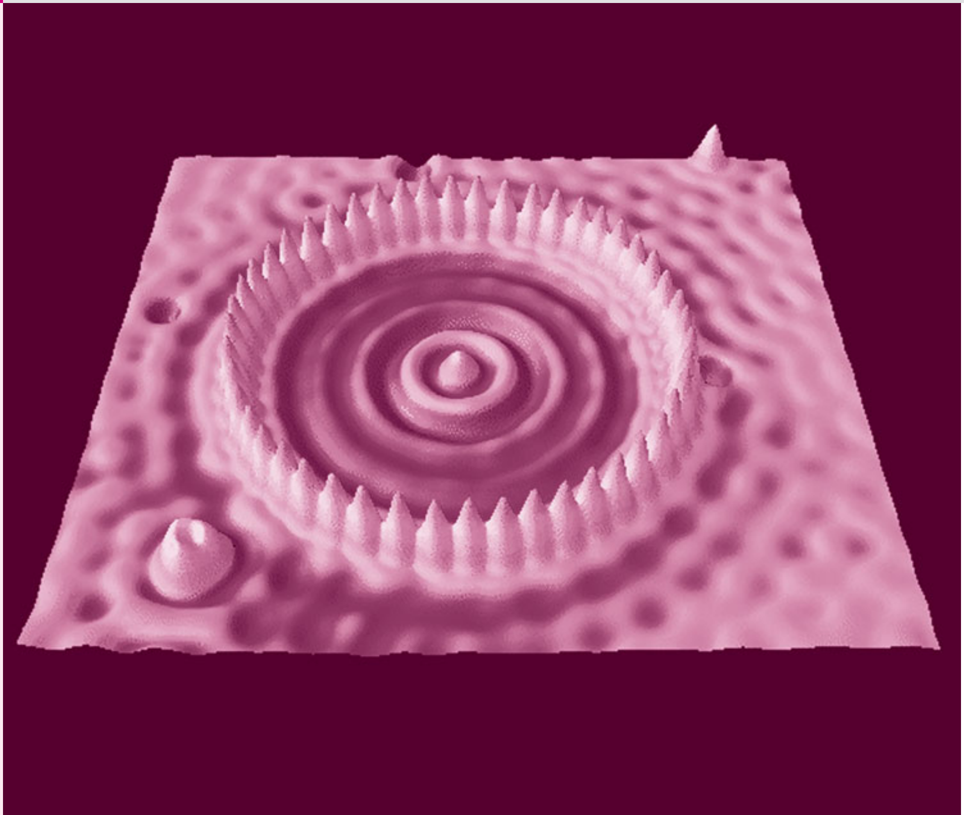
Answers to GOT IT? Questions

34.1. A emits 16 times as much radiation, with a peak wavelength half that of B.

34.2. (a) No, the slope remains h/e. (b) Yes; the horizontal intercept is the cutoff frequency, which depends on the work function.

34.3. Much less because λ_C is inversely proportional to mass.

35 Quantum Mechanics



This scanning-tunneling microscope image shows a "quantum corral" of 48 iron atoms on a copper surface. What unusual quantum phenomenon enables this type of microscopy?

New Concepts, New Skills

By the end of this chapter you should be able to

- Describe quantitatively the relation between the wave and particle descriptions of quantum systems (35.1).
- Solve the Schrödinger equation for one-dimensional square-well potentials (35.2).
- Apply Schrödinger-equation solutions for quantum harmonic oscillators (35.3).
- Describe quantum tunneling (35.3).
- Explain the complexities, especially quantum degeneracy, that arise in two and three dimensions (35.4).
- Describe qualitatively the basis of antimatter and spin in the application of special relativity to quantum physics.

Connecting Your Knowledge

- Our quantitative description of quantum physics builds on the fundamental ideas introduced in Chapter 34.
- We'll use the mathematical description of waves in terms of frequency, wavelength, and wave number k (14.2).
- The concept of potential energy and potential-energy curves will be vital in describing confined particles (7.2, 7.4).
- We'll develop the quantum analog of the simple harmonic oscillator, and we'll use both differential and integral calculus in solving quantum systems (13.2).

The ideas developed in the preceding chapter are at the core of the **old quantum theory**. The old quantum theory introduced the basic concepts of quantum physics and was successful in explaining a number of quantum phenomena—for example, blackbody radiation, the photoelectric effect, and the hydrogen spectrum. On the other hand, it couldn't treat even the simplest multielectron atoms, and it left some subtle spectral features unexplained. Furthermore, the old quantum theory was a hodgepodge of separate but loosely related ideas, each developed to explain a particular phenomenon; it lacked coherence and clear guiding principles.

Is there a more coherent theory that predicts the behavior of systems at the atomic and subatomic scales, and that offers a satisfying description of how such systems really work? The answer is at once an emphatic yes and a disappointing no. Yes, because **quantum mechanics**, developed in the 1920s, predicts with remarkable precision the observed properties of atomic systems, including their energies, the wavelengths of spectral lines, and the lifetimes of excited atoms. No, because quantum mechanics doesn't give a satisfying visual *picture* of the atomic and subatomic worlds. The uncertainty principle and wave–particle duality are essential aspects of quantum mechanics. Any picture we formulate of electrons or photons whizzing around like miniature balls with precise positions and momenta is inappropriate. But quantum mechanics does provide a self-consistent description that lets us explore and predict the behavior of atoms, the organization of chemical elements, the physics of semiconductors and superconductors, the extraordinary behavior of matter at low temperature, the formation of white dwarf stars, the operation of lasers, and a host of other phenomena for which classical physics is at best inaccurate and at worst totally inadequate. In this chapter we explore the mathematical structure and physical interpretation of quantum mechanics. In Chapters 36 and 37 we'll apply quantum mechanics first to the atom and then to more complex systems that involve quantum-mechanical interactions among many atoms.

35.1 Particles, Waves, and Probability

Photons and Light Waves

In Maxwell's electromagnetic theory we had a seemingly complete description of light as an electromagnetic wave. Now we find, through the photoelectric and Compton effects, that light sometimes manifests itself as particles. What's the connection between wave and particle descriptions?

In a photoelectric experiment, the rate at which electrons are ejected depends on the light's intensity. Since an electron is ejected when it absorbs a photon, we conclude that the number of photons in the incident light is proportional to light intensity. Now, the intensity of an electromagnetic wave depends on the square of the electric or magnetic field (Equations 29.20b, c). The fields, in turn, obey Maxwell's equations, so one aspect of a photoelectric experiment—namely, the rate of electron ejection—relates to Maxwell's description of light as an electromagnetic wave.

We can quantify the relation between waves and photons, but only in a statistical sense. The ejection of individual electrons in a photoelectric experiment is quite random. The uncertainty principle prevents us from following a photon trajectory and predicting when and where an electron will be ejected. All we can say is that electrons are more likely to be ejected where the wave intensity is greater. Specifically, the probability that an electron will be ejected is directly proportional to the intensity of the incident electromagnetic waves—that is, to the square of the wave fields. More generally, the probability of finding a photon in a beam of electromagnetic waves is directly proportional to the wave intensity (Fig. 35.1).

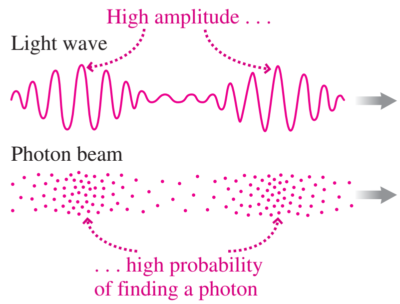
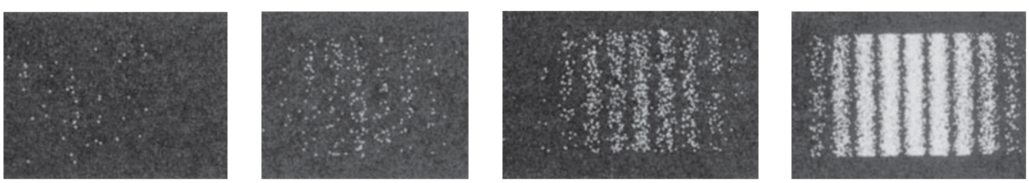


FIGURE 35.1 The probability of finding a photon is directly proportional to the intensity of the electromagnetic wave. The figure is only suggestive because we can't depict photons as localized particles.

In this quantum-mechanical description, the fields still evolve according to Maxwell's equations. For example, the fields of an electromagnetic wave undergoing double-slit interference develop regions of maximum and minimum wave intensity—the bright and dark bands of the interference pattern. But the wave fields determine only the *probability* that individual photons will be detected in the interference pattern. That's why a very short exposure or a low-intensity beam results not in a weak version of the interference pattern but in a seemingly arbitrary pattern. Only with large numbers of photons does the statistical pattern emerge (Fig. 35.2).



(a) (b) (c) (d)

FIGURE 35.2 Development of a two-slit interference pattern from random photon events: (a) 50 photons, (b) 250 photons, (c) 1000 photons, (d) 10,000 photons.

In quantum mechanics, then, the relation between the wave and particle aspects of light is this: As long as we don't try to detect the light, it propagates as a wave governed by Maxwell's equations. But when we detect the light, we do so through interactions involving individual photons. Those interactions are random events whose probability depends on the wave intensity—that is, on the square of the wave fields.

Electrons and Matter Waves

In Chapter 34 we introduced de Broglie's remarkable hypothesis that matter, as well as light, exhibits both wave and particle properties. The wave–particle duality puts matter

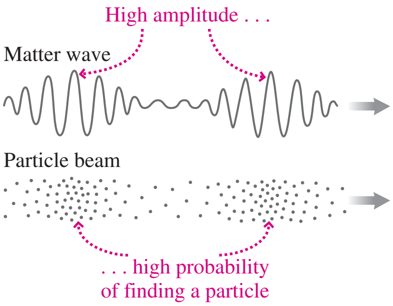


FIGURE 35.3 A beam of particles and its associated matter wave.

and light on essentially the same footing, and the statistical interpretation is the same for each. Figure 35.3 shows a beam of particles and its associated de Broglie matter wave. Just as the probability of finding a photon is proportional to the wave intensity—that is, the *square* of the electromagnetic-field amplitude—so the probability of finding a particle is directly proportional to the square of the matter-wave amplitude. And as with light, the particle nature of matter manifests itself only when we try to detect a particle; leave it alone, and the particle's behavior is governed by its wave nature.

Maxwell's equations determine the behavior of light waves, but what equation describes matter waves? In 1926 the Austrian physicist Erwin Schrödinger answered this question with his **Schrödinger wave equation**. In the same year, Schrödinger showed that his wave theory was equivalent to a matrix-based theory that Heisenberg, Max Born, and Pascual Jordan had formulated in 1925. Heisenberg received the 1932 Nobel Prize in physics, and Schrödinger shared the 1933 Nobel Prize with Paul Dirac for their contributions to quantum theory.

35.2 The Schrödinger Equation

The Schrödinger equation describes matter waves in terms of a **wave function**, ψ (Greek psi), which depends on both space and time. The solution of differential equations in two variables is beyond the mathematical level of this text, so here we'll consider only spatial variations, and for now we'll further restrict ourselves to one dimension.

We can understand the Schrödinger equation by considering a sinusoidal wave of the form $\psi(x) = A\sin kx$, where, as usual in describing waves, $k = 2\pi/\lambda$, with λ the wavelength. Differentiating this expression twice gives

$$\frac{d^2\psi(x)}{dx^2} = -Ak^2\sin kx = -k^2\psi(x)$$

But $k = 2\pi/\lambda$ and, for matter waves, de Broglie's hypothesis gives $\lambda = h/p$, with h Planck's constant and p the particle's momentum. Thus we can write k in terms of momentum as $k = 2\pi p/h = p/\hbar$. Now in classical physics a particle of mass m has kinetic energy K and momentum p related by $K = p^2/2m$. Furthermore, kinetic energy is the difference between the total energy E and the potential energy U; thus $E - U = p^2/2m$. Putting this all together, we can write the quantity k^2 in our differentiated wave expression as

$$k^2 = \frac{p^2}{\hbar^2} = \frac{2m(E-U)}{\hbar^2}$$

Make this substitution and do a little algebra; the result is

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + U(x)\psi(x) = E\psi(x) \qquad \left(\begin{array}{c}\text{time-independent}\\ \text{Schrödinger equation}\end{array}\right) \tag{35.1}$$

This is the **time-independent Schrödinger equation**, giving the spatial variation of matter waves in one dimension. A solution of the full time-dependent equation consists of a solution to Equation 35.1 multiplied by a sinusoidal oscillation with frequency $f = E/h$, where E is the particle energy. We developed the time-independent Schrödinger equation by merging de Broglie's matter-wave hypothesis $\lambda = h/p$ with the Newtonian relation $K = p^2/2m$; for that reason, we expect the equation to hold only for nonrelativistic particles.

The Schrödinger equation provides a description of physical reality in remarkable agreement with experiments. As we'll see, Schrödinger's equation goes a long way toward explaining the structure of atoms, their chemical properties, and indeed the entire science of chemistry. Furthermore, the Schrödinger description obeys the correspondence principle, agreeing with Newtonian mechanics for macroscopic systems where quantum effects are small.

The Meaning of ψ

What's the meaning of the wave function ψ? That's a deep question that physicists and philosophers continue to debate. In the standard interpretation, ψ is not an observable quantity. It manifests itself only in the statistical distributions of particle detections. More specifically, the probability per unit volume—also called the **probability density**—that we'll find a particle is given by ψ^2. For a particle confined to one dimension, the probability density becomes the probability per unit length, and we interpret ψ to mean that the probability $P(x)$ of finding the particle in a small interval dx at position x is

$$P(x) = \psi^2(x)\,dx \qquad \text{(probability and the wave function)} \tag{35.2}$$

We can interpret Equation 35.2 in two ways. At face value, it gives the probability that a single experiment, with a detector at position x set up to find particles in an interval of width dx, will detect the particle (Fig. 35.4). Or, if we do many such experiments, the equation gives the fraction of the experiments in which we'll find a particle in our detector.

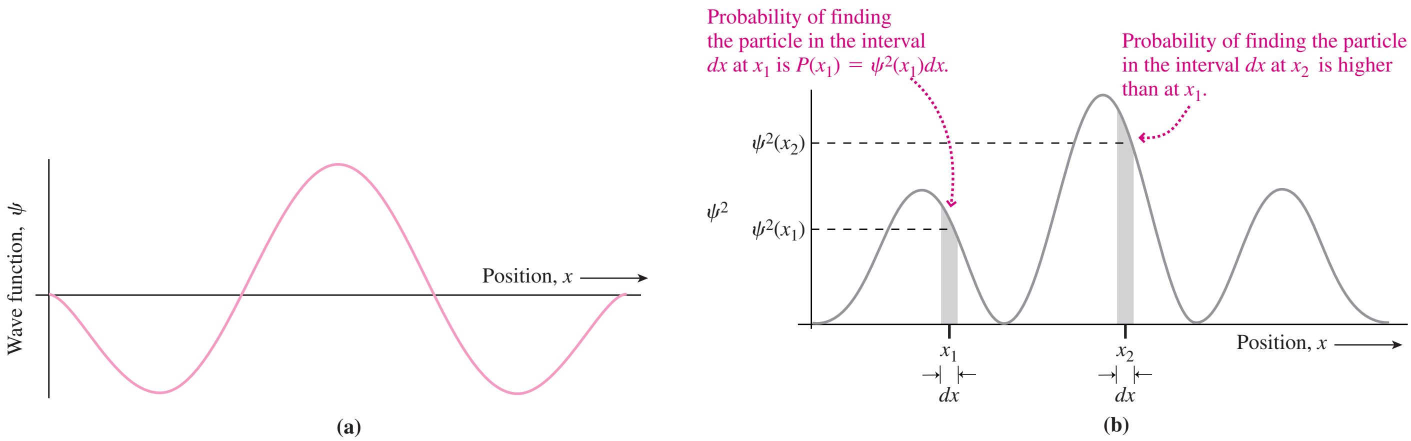


FIGURE 35.4 The meaning of the probability density $\psi^2(x)$. (a) A wave function and (b) its square, which gives the probability density.

But *what is* ψ? How can it be unobservable yet govern the behavior of matter? There can't be a direct causal link between the wave function and individual particles, since ψ determines only the *probability* that a particle will behave in a certain way. Think about this! In quantum mechanics the outcome of an experiment isn't fully determined. The Schrödinger equation describes only the probability of a given outcome. The quantum world is so different, according to the standard interpretation, that our macroscopic language, concepts, and pictorial models are simply inadequate. In particular, macroscopic causality gives way to microscopic indeterminancy in which quantum events are truly random; physical laws govern only the statistical pattern of events.

The philosophical implications of quantum mechanics have been debated since the theory was formulated. A central theme in this debate is the possibility of "hidden variables," physical quantities that might be hidden from us by the uncertainty principle but that might nevertheless govern the microscopic world in a fully deterministic way. Experiments of the early 1980s placed severe restrictions on such hidden-variable theories, but fascinating discussions on the interpretation of quantum mechanics continue. Here, however, we turn to the Schrödinger equation to see how it's used in analyzing quantum-mechanical systems.

Normalization and Other Constraints on the Wave Function

In one dimension, the quantity $\psi^2 dx$ represents the probability of finding the particle in the interval dx. But the particle *must* be *somewhere*. Therefore, if we sum the probabilities of finding the particle in all such intervals dx, the result must be 1; there must be a 100% chance that we'll find the particle somewhere. Since the probability density may vary with position, that sum becomes an integral:

$$\int_{-\infty}^{+\infty} \psi^2 dx = 1 \qquad \text{(normalization condition)} \tag{35.3}$$

Once we have a solution $\psi(x)$ to the Schrödinger equation 35.1, this **normalization condition** sets the overall amplitude of the function ψ.

The Schrödinger equation contains the second derivative of ψ. In order that this term be well defined, both ψ itself and its first derivative must be continuous. (An exception to the continuity condition on $d\psi/dx$—possible only in unrealistic example situations—occurs if the potential energy U becomes infinite.)

35.3 Particles and Potentials

The Infinite Square Well

We first solve the Schrödinger equation for a particularly simple system—a particle trapped in one dimension between two perfectly rigid walls. Although unrealistic in some respects, this system nevertheless is a surprisingly good approximation to some real quantum systems, including some electronic devices and simple nuclei. More important, its analysis illustrates the general procedure for applying the Schrödinger equation and shows how energy quantization emerges from Schrödinger's theory.

In classical physics, a particle trapped between rigid walls moves back and forth with constant speed. In the absence of friction or other losses, the particle's energy remains constant at its initial value. And in classical physics, that value can be anything.

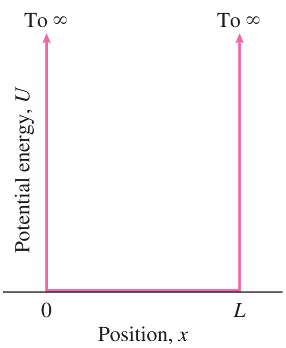


FIGURE 35.5 Infinite square well potential-energy curve describes a particle constrained to move in one dimension between rigid walls separated by a distance L.

We can describe the particle's situation using its potential-energy curve. Since the particle experiences no forces while it's between the walls, its potential energy U is constant in this region, and we can fix the arbitrary zero of potential energy by setting $U = 0$. If the walls are perfectly rigid, then the particle can't penetrate them, no matter what its energy. This means that the potential energy becomes abruptly infinite at the walls. Then the potential-energy curve for our particle looks like Fig. 35.5; you can see from the figure why this curve is called an **infinite square well**. In this case the well extends from $x = 0$ to $x = L$.

We now consider the quantum-mechanical description of a particle in the infinite square well. The particle has a wave function whose time-independent part is given by the Schrödinger equation (Equation 35.1):

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + U(x)\psi = E\psi$$

where the potential energy $U(x)$ is that of the square well in Fig. 35.5:

$$U = 0 \text{ for } 0 < x < L$$
$$U = \infty \text{ for } x < 0 \text{ or } x > L$$

Since there's no chance that the particle can penetrate the rigid walls, the function ψ must be exactly zero in the region where $U = \infty$. All we need to calculate, then, is ψ inside the well, where $0 \le x \le L$. To ensure that the particle is confined to the well, our solution must satisfy so-called **boundary conditions** $\psi = 0$ at $x = 0$ and at $x = L$.

Within the well, $U = 0$ and the Schrödinger equation becomes

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} = E\psi \qquad (35.4)$$

To find solutions, recall de Broglie's hypothesis that the allowed orbits in the Bohr atom are those for which standing waves just "fit" around the orbit. We have a similar situation with the infinite square well, in which the allowed solutions should be standing waves with nodes at the ends of the well—exactly analogous to standing waves on a string with both ends clamped that we discussed in Chapter 14. So we want a sinusoidal wave for $\psi(x)$, subject to the boundary conditions $\psi(0) = 0$ and $\psi(L) = 0$. The first condition is satisfied if we take a wave of the form $\psi = A\sin kx$, with A and k both constants. The second condition requires that $k = n\pi/L$, where n is any integer—a condition equivalent to saying that an integer number of half-wavelengths fit in the well. So we propose a solution of the form

$$\psi(x) = A\sin\left(\frac{n\pi x}{L}\right)$$

with the constant A still undetermined. This equation represents standing waves with nodes at the ends of the square well, but does it satisfy the Schrödinger equation? We can find out by substituting into Equation 35.4. We need not only ψ but also its second derivative; twice differentiating our proposed solution gives

$$\frac{d^2\psi}{dx^2} = -A\frac{n^2\pi^2}{L^2}\sin\left(\frac{n\pi x}{L}\right)$$

Substituting ψ and $d^2\psi/dx^2$ into Equation 35.4 gives

$$\left(-\frac{\hbar^2}{2m}\right)\left[-A\frac{n^2\pi^2}{L^2}\sin\left(\frac{n\pi x}{L}\right)\right] = EA\sin\left(\frac{n\pi x}{L}\right)$$

which reduces to

$$E = \frac{n^2\pi^2\hbar^2}{2mL^2} = \frac{n^2h^2}{8mL^2} \quad \text{(energy levels for an infinite square well potential)} \quad (35.5)$$

Equation 35.5 says that our proposed solution can indeed satisfy the Schrödinger equation—provided the particle energy E has a value given by Equation 35.5, with n an integer.

Our standing-wave solutions show how the quantization of energy arises naturally from the Schrödinger equation. Physically, the reason for quantization remains as de Broglie had postulated: Matter waves in a confined system must be standing waves with an integer number of half-wavelengths. Although de Broglie's hypothesis and the Schrödinger equation lead to exactly the same conclusion for the infinite square well, we'll see that with more complicated potential-energy functions only the Schrödinger equation can give us the full story.

The integer n that appears in Equation 35.5 is the **quantum number** for the particle in the square well. The physical state of a quantum-mechanical system is its **quantum state**. Here one quantum number suffices to specify the quantum state, which then tells us everything quantum mechanics has to say about the situation. As far as the Schrödinger equation is concerned, it looks like all integer values of n are allowed. The choice of negative or positive n has no physical significance, since ψ^2 has the same value with either sign of ψ; for this reason, negative n's are redundant. But $n = 0$ implies $\psi = 0$ everywhere, giving no chance of finding the particle anywhere. So we're left with positive integer values of n.

With only nonzero n's allowed, Equation 35.5 shows that the particle's energy is always positive; zero energy isn't allowed. The lowest possible energy is $E_1 = h^2/8mL^2$, obtained with $n = 1$. This is the **ground-state energy**; the corresponding wave function is the **ground-state wave function**. A nonzero ground-state energy is a common feature of quantum systems and one with no classical counterpart. Figure 35.6 is an energy-level diagram for the infinite square well.

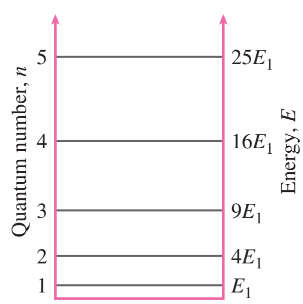


FIGURE 35.6 Energy-level diagram for a particle in an infinite square well. Energy is proportional to n^2, so the levels aren't evenly spaced.

CONCEPTUAL EXAMPLE 35.1 Ground-State Energy

Why can't the ground-state energy of the square well be zero?

EVALUATE Consider the uncertainty principle: If the ground-state energy were zero, then we would know precisely the particle's kinetic energy $p^2/2m$—zero—and therefore we would know that its momentum p was also zero. But we know that the particle is within the well, so the uncertainty in its position is at most the well width L. The product $\Delta p\,\Delta x$ would then be zero, in violation of the uncertainty principle.

ASSESS We used the uncertainty principle in the preceding chapter to estimate the minimum energies of confined particles. The ground-state energy for the square well is a specific instance of this so-called *zero-point energy*.

MAKING THE CONNECTION An electron drops from the $n = 2$ state to the ground state of a 0.75-nm-wide infinite square well, emitting a photon in the process. Find the photon's energy.

EVALUATE Equation 35.5 gives the square-well energies. Here the photon's energy is the difference between E_2 and the ground-state energy E_1: $\Delta E = 3.2\times10^{-19}$ J, or 2.0 eV.

GOT IT? 35.1 Electron A is confined to a square well 1 nm wide; electron B to a similar well only 1 pm wide. How do their ground-state energies compare?

Normalization, Probabilities, and the Correspondence Principle

We still don't know the constant A in our solution for the infinite square well. We find this using the normalization condition 35.3: $\int_{-\infty}^{\infty} \psi^2 dx = 1$. Inside the well, $\psi = A \sin(n\pi x/L)$; outside, $\psi = 0$. So we can write the normalization condition as an integral over $0 < x < L$:

$$\int_0^L A^2 \sin^2\left(\frac{n\pi x}{L}\right) dx = 1$$

If we divided $\int_0^L \sin^2(n\pi x/L)\, dx$ by the well width L, we would have the average of sine squared over an integer number of half-cycles—or just $\frac{1}{2}$. So the integral of $\sin^2(n\pi x/L)$ from 0 to L is $\frac{1}{2}L$, and therefore $A^2(L/2) = 1$, or $A = \sqrt{2/L}$. The normalized wave function is then

$$\psi_n = \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right) \tag{35.6}$$

where the subscript n refers to the function associated with the nth quantum state. Figure 35.7 shows wave functions for the ground state and three excited states.

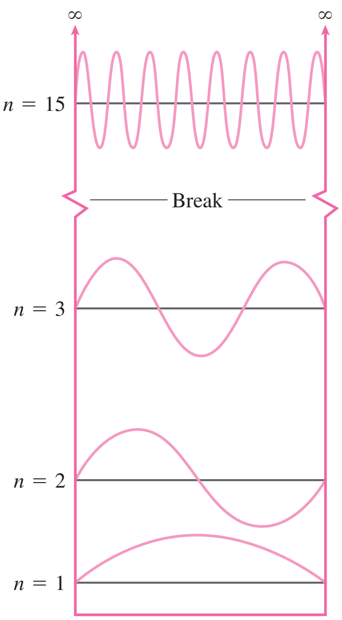


FIGURE 35.7 Wave functions for a particle in an infinite square well, each centered on the corresponding energy level.

Where are we likely to find the particle? Classically, it would move back and forth at constant speed and therefore would be equally likely to be anywhere in the well. Quantum-mechanically, the probability of finding it at some position x is proportional to the probability density ψ^2 at that point. Figure 35.8 shows the probability densities given by squaring the wave functions of Fig. 35.7. For $n = 1$ we're clearly most likely to find the particle near the middle of the well—in marked contrast to the classical prediction of equal probability everywhere. For other low-n states there are obvious regions of high and low probability. But as the quantum number increases, the maxima and minima of the probability density get closer together. Any instrument we use to detect the electron has a finite resolution, and once the periodicity of the wave function drops below that resolution, we measure an average probability, which is essentially constant over the interval (Fig. 35.8).

This is a manifestation of Bohr's correspondence principle: For large quantum numbers n, the interval between adjacent energy levels becomes small compared with the energy itself, and a measurement of the electron's position gives results in agreement with classical physics. But classical physics is totally inadequate at low n, where the nonclassical zero-point energy and quantization are most evident.

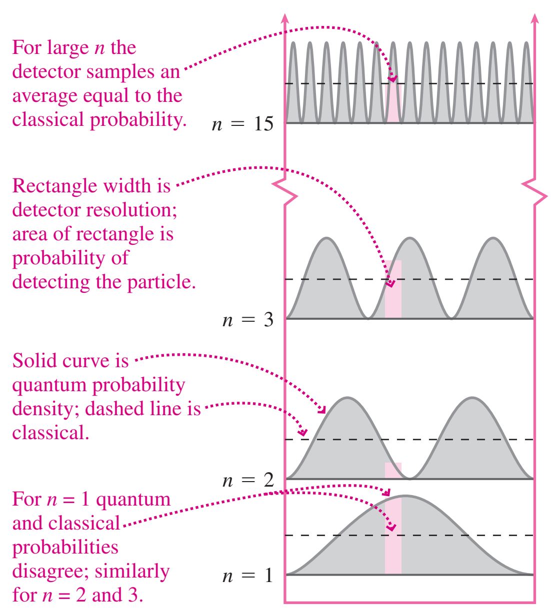


FIGURE 35.8 Classical (dashed) and quantum (solid) probability densities for a particle in an infinite square well. The shaded area under each curve is 1, indicating that the particle must be somewhere in the well. Width of the colored rectangle is the resolution of an instrument used to detect the particle.

EXAMPLE 35.1 **Quantum Probability: The Square-Well Ground State**

A particle is in the ground state of an infinite square well. Find the probability that it will be found in the left-hand quarter of the well.

INTERPRET This is a question about probability, and we know that the probability density is the square of the wave function. So our solution is going to involve ψ^2.

DEVELOP The ground-state wave function from Equation 35.6 is $\psi_1 = \sqrt{2/L}\,\sin(\pi x/L)$. We normalized the wave function so the integral of ψ_1^2 over the entire well is 1, showing that the particle is *somewhere* in the well. That is, the area under the entire plot of ψ_1^2 is 1. We sketched ψ_1^2 in Fig. 35.9, showing that the probability of finding the particle in some region is the area under the curve in that region. So to find the probability that the particle is in the left-hand quarter of the well, we'll evaluate $\int \psi^2\,dx$ from 0 to $\frac{1}{4}L$.

EVALUATE The probability becomes

$$P = \frac{2}{L}\int_0^{L/4} \sin^2\!\left(\frac{\pi x}{L}\right) dx$$

We can integrate using the table at the end of Appendix A; the result is

$$P = \frac{2}{L}\left(\frac{x}{2} - \frac{\sin(2\pi x/L)}{4\pi/L}\Bigg|_0^{L/4}\right) = \frac{2}{L}\left(\frac{L}{8} - \frac{L}{4\pi}\right) = 0.091$$

ASSESS This is considerably lower than the probability $P = 0.25$ we would expect classically for finding the particle in any quarter of the well, and reflects the lower value of ψ^2 nearer the well ends. Problem 52 repeats the calculation of this example for arbitrary quantum numbers, showing that classical and quantum probabilities agree at large n. ■

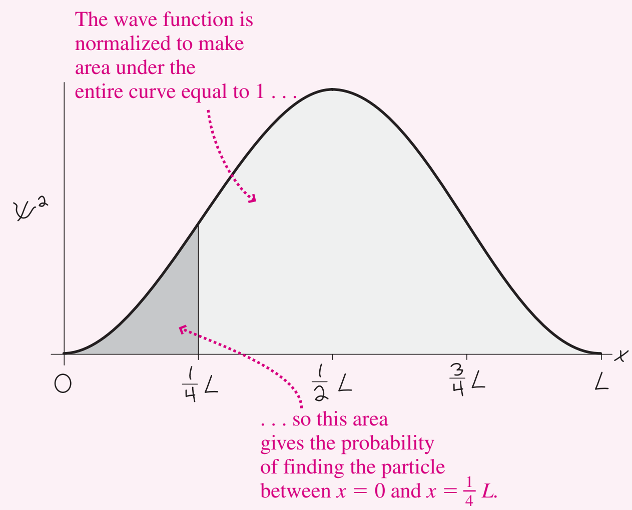


FIGURE 35.9 Sketch for Example 35.1.

GOT IT? 35.2 Which of the following would be a reasonable answer if Example 35.1 had asked for the probability that the particle would be found in the central quarter of the well: (a) 0.091, (b) 0.25, (c) 0.475, (d) 0.90?

The infinite square well gives insights into important quantum phenomena shared by more realistic systems such as atoms. These include quantized energy levels, nonzero ground-state energy, nonclassical probabilities, and agreement with classical physics at large quantum numbers. In Chapter 36 we'll apply the Schrödinger equation to atoms, where we'll find many of the same phenomena. First, though, we look at some other simple systems that exhibit additional quantum behaviors.

The Harmonic Oscillator

In Chapter 13 we studied simple harmonic motion, which occurs when a particle is subject to a restoring force that's directly proportional to the displacement from equilibrium. Such a *linear* restoring force implies a *quadratic* potential-energy function, and conversely, as we pointed out in Section 13.5, any system with a quadratic potential-energy function is a harmonic oscillator. That includes many systems at the atomic and molecular scale. Understanding the quantum-mechanical harmonic oscillator is therefore crucial in describing the behavior of matter on small scales.

A mass–spring system has potential energy $U = \frac{1}{2}kx^2$ and oscillates with angular frequency given by Equation 13.7: $\omega = \sqrt{k/m}$. Combining these equations gives $U = \frac{1}{2}m\omega^2x^2$, providing a potential-energy function suitable for an electron or atom vibrating at the end of a molecular bond. Solving the Schrödinger equation for this potential requires advanced math techniques, and shows that normalizable wave functions exist only for discrete values of the energy E:

$$E_n = \left(n + \tfrac{1}{2}\right)\hbar\omega \tag{35.7}$$

where now $n = 0$ is the ground state. Figure 35.10 shows an energy-level diagram for the harmonic oscillator; note the even spacing implied by Equation 35.7. The additive factor $\frac{1}{2}$

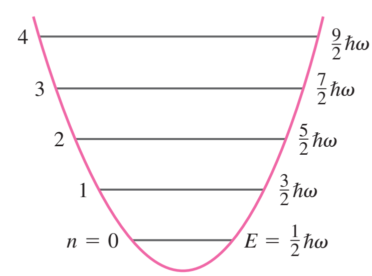


FIGURE 35.10 Energy-level diagram for a quantum-mechanical harmonic oscillator, superposed on its quadratic (i.e., parabolic) potential-energy curve.

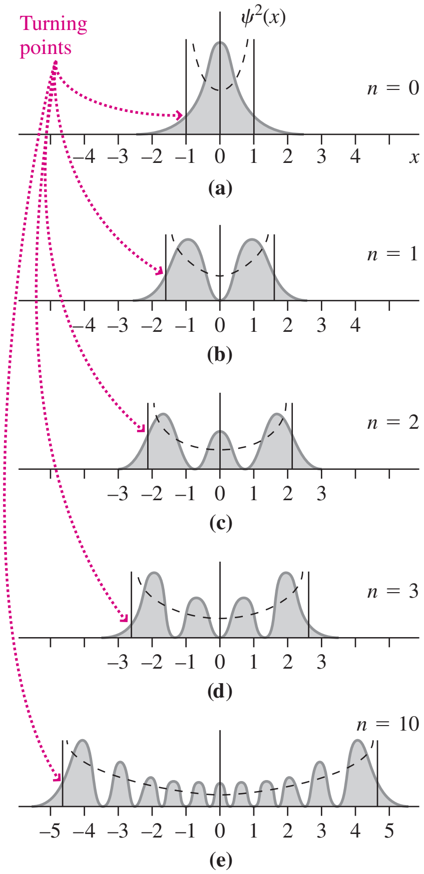


FIGURE 35.11 Probability densities $\psi^2(x)$ for some states of the harmonic oscillator. Dashed curves are classical predictions. Increasing spread in the turning points reflects the higher energy of the higher-n states.

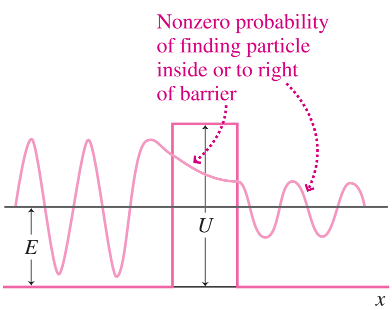


FIGURE 35.12 A potential barrier of height U, showing the wave function for a particle incident from the left with energy E lower than the barrier energy U.

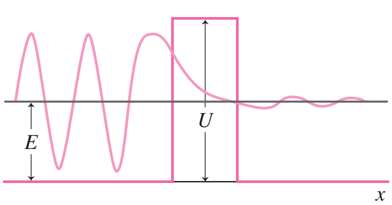


FIGURE 35.13 For a massive particle, the wave function drops rapidly in the barrier, giving negligible probability of penetration.

in Equation 35.7 shows that Planck wasn't quite right in suggesting that the allowed harmonic-oscillator energies should be multiples of $hf(=\hbar\omega)$. Planck's spectral distribution (Equation 34.3) is nevertheless correct, but he did not foresee the existence of nonzero ground-state energy.

The even spacing between the energy levels of the harmonic oscillator is in marked contrast to the situation in atoms (Fig. 34.11) or in the infinite square well (Fig. 35.6). A quantum harmonic oscillator emits or absorbs photons as it makes transitions among adjacent levels, and the even spacing means that all transitions between adjacent levels of a pure harmonic oscillator involve photons of the same energy.

A classical harmonic oscillator moves slowest near its turning points, so it's most likely to be found at the extremes of its motion. It's least likely to be at its equilibrium position, where it's moving fastest. As with the square well, the harmonic oscillator in low-n states exhibits unclassical behavior; in the ground state it's *most* likely to be found at its equilibrium position! Figure 35.11 shows classical and quantum probability densities for the harmonic oscillator; note that for larger n the two begin to agree, once again showing Bohr's correspondence principle at work.

Quantum Tunneling

One remarkable feature of Fig. 35.11 is the nonzero probability of finding a quantum-harmonic oscillator beyond its classical turning points—the points at which its kinetic energy has been converted entirely to potential energy. This unusual situation, which seems to violate energy conservation, has no counterpart in the classical description of matter.

Another example of penetration into a classically forbidden region is a particle encountering a potential barrier (Fig. 35.12). Examples of such barriers include electric potential differences associated with atomic nuclei, gaps between solid materials, and insulating layers in some semiconductor devices. Classically, a particle whose total energy is lower than the barrier energy is confined to one side of the barrier. If we solve the Schrödinger equation for this potential-energy curve, however, we find oscillatory solutions on either side of the barrier, joined according to the continuity conditions on ψ and $d\psi/dx$ by exponential functions within the barrier. Such a solution is shown superimposed on the barrier in Fig. 35.12. The probability density ψ^2 associated with this solution remains nonzero through the barrier and continues to give a nonzero probability of finding the particle on the far side—implying that a particle initially on one side of the barrier may later be found on the other side.

How likely is this phenomenon, called **quantum tunneling**? That depends on the relation of the particle energy E to the barrier energy U, and also on the width of the barrier. As you can show in Problem 47, the ψ function inside the barrier involves exponential functions of the form $e^{\pm\sqrt{2m(U-E)}x/\hbar}$. In general, these exponentials drop very rapidly across the barrier width unless the particle energy E is close to the barrier energy or the particle mass m is small. The probability that a particle will be found on the far side of the barrier is therefore very low when the mass m is large, so quantum tunneling is a microscopic phenomenon (Fig. 35.13).

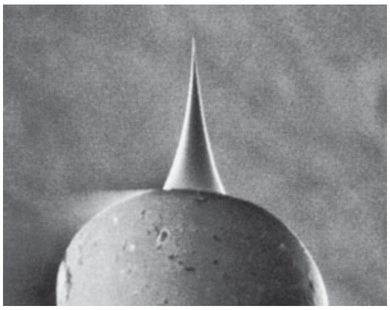
It looks as if tunneling violates energy conservation. But we're saved by the uncertainty principle. If we catch the particle within the barrier, the uncertainty in its position is no greater than the barrier width. We know from Example 34.5 that this implies a minimum energy. A quantitative analysis shows that minimum to be such that we can no longer be sure the particle energy is lower than the barrier energy. If we don't try to detect a particle within the barrier, its penetration is a purely wave phenomenon to which our particulate energy considerations don't apply. Again we see the wave–particle duality at work: If we don't observe the particle, its behavior is governed by the associated waves and may result in most unparticle-like phenomena such as tunneling. If we do try to catch it in the act of such behavior, it ceases to be wave-like and the surprising phenomena cease.

Tunneling is important in a number of quantum-mechanical phenomena and technological devices. That the Sun shines—and therefore that we're alive—is a consequence of quantum tunneling of nuclei in the Sun's core. Classically, those nuclei don't have sufficient energy to get close enough to overcome their mutual electric repulsion. But they can tunnel through this "Coulomb barrier" and fuse to release the enormous energy that powers the

Sun. An opposite process, alpha decay, occurs as alpha particles tunnel through a potential barrier that traps them inside large nuclei like uranium. Measurement of the alpha particles' energy shows it to be lower than the barrier energy, confirming that tunneling occurs. Semiconductor devices involving quantum tunneling hold the promise of a new generation of much faster electronic circuits. Finally, tunneling is the basis of the scanning-tunneling microscope (STM), a remarkable device that lets us image individual atoms.

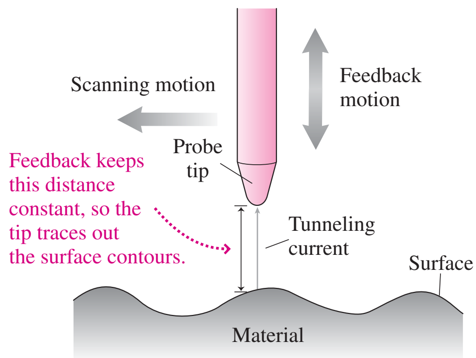
GOT IT? 35.3 A proton and an electron approach a barrier. Both have the same energy E, which is lower than the barrier potential U. Which is more likely to get through?

APPLICATION **Scanning Tunneling Microscope**



Developed in the 1980s by Heinrich Rohrer and Gerd Binnig of IBM Zurich Research Laboratory, the Scanning Tunneling Microscope (STM) has become a vital tool for semiconductor engineers, biologists, chemists, and nanotechnologists. The STM works by quantum tunneling between an extraordinarily fine conducting tip and the surface under study. The photo shows a scanning electron microscope image of an STM tip, which may be only one atom wide. As in the barrier of Fig. 35.12, the electron wave function tapers off exponentially in the space outside the surface. Place a conducting tip near but not touching the surface, and there's a nonzero probability that electrons will tunnel through the gap to reach the tip, resulting in an electric current. The exponential falloff of the wave function means this **tunneling current** is extremely sensitive to the tip-to-surface gap, and therefore changes significantly with surface irregularities.

A practical STM scans the tip over the surface, and feedback devices move the tip to keep the tunneling current constant despite surface irregularities, as shown in the figure. Therefore, the tip traces out surface irregularities, and this information is used to construct an image of the surface (see this chapter's opening photo).



Finite Potential Wells

Both the infinite square well and the harmonic oscillator have potential wells of infinite depth. No matter what its energy a particle is bound in such a well; it can't escape to large distances. Its quantized energy states are therefore all **bound states**. Provided they aren't too shallow, wells of finite depth also exhibit quantized bound states whose wave functions resemble those of the infinite square well (Fig. 35.14), although they show a small but nonzero probability of tunneling into the classically forbidden region outside the well.

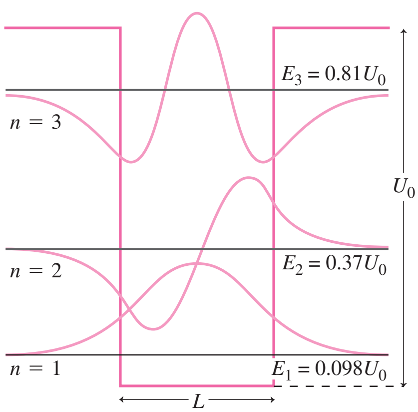


FIGURE 35.14 Bound-state wave functions for a finite square well, superposed on the associated energy levels. For this combination of well depth, well width, and particle mass there are only three bound states.

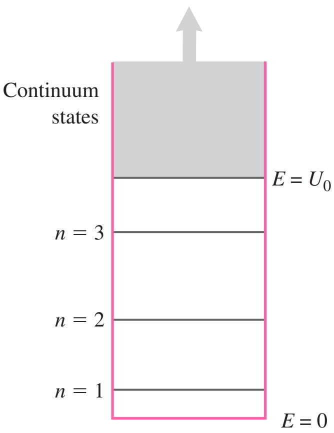


FIGURE 35.15 Energy-level diagram for a finite square well shows discrete bound states and a continuum of unbound states.

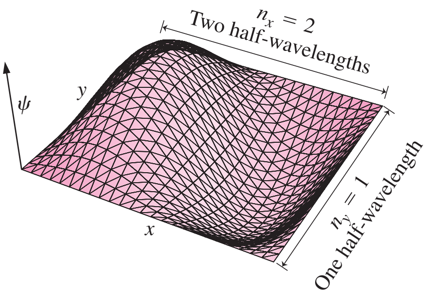


FIGURE 35.16 The wave function for a particle confined to a square region in two dimensions. The function is $\sin(n_x\pi x/L)\,\sin(n_y\pi y/L)$, with $n_x = 2$ and $n_y = 1$.

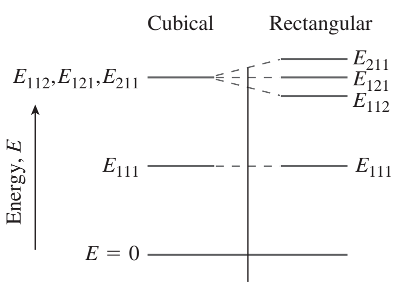


FIGURE 35.17 Energy-level diagrams showing the ground state and first excited state for a particle in a three-dimensional box. Making the sides different lengths removes the degeneracy.

Quantized bound states represent particles with energy lower than the well height. Particles with higher energy are free to move anywhere, and their wave functions are everywhere oscillatory. Furthermore, particles in these **unbound states** can have any energy whatsoever as long as it exceeds the well height; unbound energies aren't quantized. Rather, there's a **continuum** of allowed energies above the well top, in contrast to the discrete, quantized levels below (Fig. 35.15). We'll find both bound and unbound states again in the next chapter when we study the atom.

35.4 Quantum Mechanics in Three Dimensions

One-dimensional quantum systems show important features of the quantum world, like energy quantization and tunneling. But atoms and most other quantum systems are three-dimensional. The wave function then depends on all three spatial variables, and the Schrödinger equation reflects this complexity. You can explore the three-dimensional Schrödinger equation in Problem 49; here we just point out some new features of three-dimensional quantum systems.

A single quantum number n characterizes quantum states in one dimension. With the infinite square well, for example, an integer number of half-wavelengths can fit in the well, and n is that number. Each n is associated with a distinct energy level. In two or three dimensions, similar considerations lead to independent quantum numbers for each dimension (Fig. 35.16). For each set of quantum numbers there's an associated energy. For a particle of mass m confined to a cubical box of side L, for example, a generalization of the one-dimensional square well leads to the energy levels

$$E = \frac{h^2}{8mL^2}\left(n_x^2 + n_y^2 + n_z^2\right) \qquad (35.8)$$

where the n's are the quantum numbers associated with each spatial dimension. As in one dimension, the allowed values for the n's are positive integers. Thus, the ground state has $n_x = n_y = n_z = 1$. But what's the first excited state? It could be $n_x = 2, n_y = n_z = 1$. But it could equally well be $n_x = n_y = 1, n_z = 2$, or $n_x = n_z = 1, n_y = 2$, since all three of these combinations give the same energy.

Two or more quantum states with the same energy are termed **degenerate**. The first excited state of a particle confined to a cubical box is threefold degenerate, meaning there are three distinct states with the same energy. Degeneracy is often associated with symmetry of the quantum-mechanical system. In the cubical box, the equal-length sides result in different combinations of quantum numbers with the same energy. Making the sides different would remove the degeneracy, splitting a single energy level into three (Fig. 35.17). The same thing happens in more realistic quantum systems. For example, imposing a magnetic field on an otherwise spherically symmetric atom breaks the symmetry and may split energy levels that were previously degenerate (Fig. 35.18). Detection of this splitting in optical spectra allows measurement of magnetic fields on the Sun and in other remote objects.

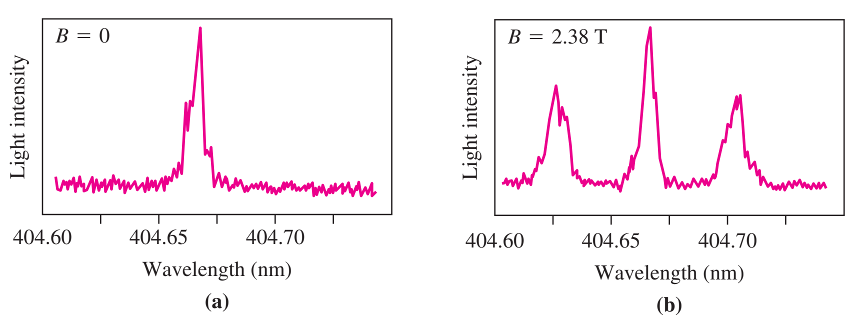


FIGURE 35.18 (a) Spectral line at 404.66 nm produced by mercury atoms undergoing transitions from $n = 7$ to $n = 6$. The upper level is actually threefold degenerate. (b) A magnetic field breaks the symmetry and removes the degeneracy, splitting the spectral line.

35.5 Relativistic Quantum Mechanics

Like Newtonian physics, quantum mechanics based on the Schrödinger equation is not consistent with special relativity's requirement that the laws of physics be the same in all inertial reference frames. It's therefore an approximation valid for particle speeds v much lower than c. For most applications in atomic, molecular, and condensed-matter physics, $v \ll c$ so the Schrödinger equation applies. But when particle speeds are a significant fraction of c, the Schrödinger equation becomes inadequate and must be replaced with a relativistic wave equation. And even for slowly moving particles, the requirement of relativistic invariance leads to some surprising new phenomena.

The Dirac Equation and Antiparticles

In 1928 the English physicist Paul Dirac formulated a relativistic wave equation for electrons. In the process he encountered several unexpected mathematical requirements with deep physical significance.

Dirac replaced the Newtonian energy–momentum relation $K = p^2/2m$ with the relativistic expression $E^2 = (mc^2)^2 + p^2c^2$ that we saw in Chapter 33. But this expression implies two values for E, depending on which sign one chooses in taking the square root. Dirac argued that both roots are meaningful and that the negative root implies the existence of a particle identical in mass to the electron but carrying positive charge. The 1932 discovery of this **positron** vindicated Dirac's brilliant idea. Today we know that every elementary particle has a corresponding **antiparticle**, identical in mass but opposite in electric, magnetic, and other properties.

Einstein's energy–mass equivalence implies that **pair creation** of a particle–antiparticle pair is possible, given energy $2mc^2$ equivalent to the mass of the pair. The opposite process, annihilation, occurs as particle and antiparticle meet and disappear to form a pair of photons. Although pair creation is rare today, it was commonplace in the hot, early universe, where thermal energy alone was high enough to create particle–antiparticle pairs. In those early times Einstein's mass–energy equivalence would have been obvious, and the number of particles in a closed volume wouldn't have remained constant.

Electron Spin

Another unexpected mathematical result of Dirac's work was that the wave function had to involve matrices. This, Dirac showed, implied physically that the electron must possess an intrinsic angular momentum—something physicists had already inferred from experiments, but without any theoretical grounding. This angular momentum, called **spin**, has enormous significance in quantum mechanics and particularly in atomic structure, as we'll see in the next chapter.

CHAPTER 35 SUMMARY

Big Picture

The big idea here is the description of particles in the quantum realm using **wave functions**, whose square relates to the probability of finding a particle. Thus the link between the most thorough description physics can provide—the wave function—and the behavior of an individual particle is only statistical. The **Schrödinger equation** gives the wave function for nonrelativistic particles and leads to energy quantization for confined particles.

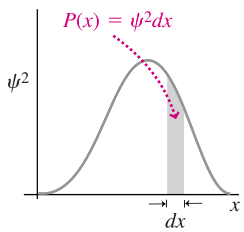
Key Concepts and Equations

The **time-independent Schrödinger equation** gives the wave function ψ for a particle of mass m with total energy E and potential energy U:

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + U(x)\psi = E\psi$$

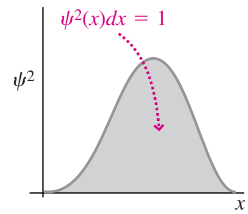
The square of the wave function is the **probability density**. In one dimension the probability of finding the particle in some small interval dx at position x is

$$P(x) = \psi^2(x)\,dx$$



Normalization: A particle must be *somewhere*, so

$$\int_{-\infty}^{+\infty} \psi^2(x)dx = 1$$



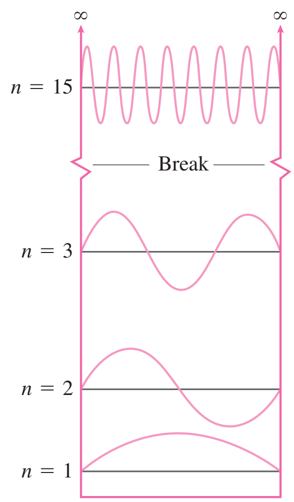
Applications

Infinite square well

Wave functions: $\psi_n = \sqrt{\frac{2}{L}}\sin\left(\frac{n\pi x}{L}\right)$

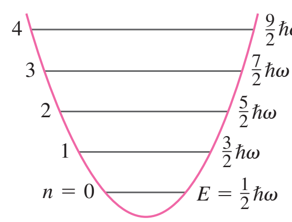
Energy levels: $E_n = \frac{n^2h^2}{8mL^2}$

3-D well: $E = \frac{h^2}{8mL^2}(n_x^2 + n_y^2 + n_z^2)$

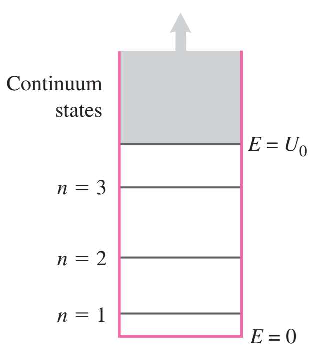


Harmonic oscillator

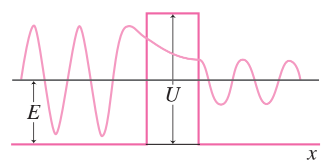
Energy levels: $E_n = \left(n + \frac{1}{2}\right)\hbar\omega$



Finite well
Discrete bound states; continuum of unbound states



Quantum tunneling
Nonzero probability of finding a quantum particle in a region forbidden by classical energy conservation leads to the possibility of barrier penetration:



For Thought and Discussion

1. Explain qualitatively why a particle confined to a finite region cannot have zero energy.
2. Does quantum tunneling violate energy conservation? Explain.
3. Bohr's correspondence principle states that quantum and classical mechanics must agree in a certain limit. Give an example.
4. The ground-state wave function for a quantum harmonic oscillator has a single central peak. Why is this at odds with classical physics?
5. What's the essential difference between the energy-level structures of infinite and finite square wells?
6. In terms of de Broglie's matter-wave hypothesis, how does making the sides of a box different lengths remove the degeneracy associated with a particle confined to that box?
7. A particle is confined to a two-dimensional box whose sides are in the ratio 1:2. Are any of its energy levels degenerate? If so, give an example. If not, why not?
8. What did Einstein mean by his remark, loosely paraphrased, that "God does not play dice"?
9. Some philosophers argue that the strict determinism of classical physics is inconsistent with human free will, but that the indeterminacy of quantum mechanics does leave room for free will. Others claim that physics has no bearing on the question of free will. What do you think?

Exercises and Problems

Exercises

Section 35.2 The Schrödinger Equation

10. What are the units of the wave function $\psi(x)$ in a one-dimensional situation?
11. A particle's wave function is $\psi = Ae^{-x^2/a^2}$, where A and a are constants. (a) Where is the particle most likely to be found? (b) Where is the probability per unit length half its maximum value?
12. The solution to the Schrödinger equation for a particular potential is $\psi = 0$ for $|x| > a$ and $\psi = A\sin(\pi x/a)$ for $-a \leq x \leq a$, where A and a are constants. In terms of a, what value of A is required to normalize ψ?

Section 35.3 Particles and Potentials

13. What's the quantum number for a particle in an infinite square well if the particle's energy is 25 times the ground-state energy?
14. A particle in an infinite square well makes a transition from a higher to a lower energy state; the corresponding energy decrease is 33 times the ground-state energy. Find the quantum numbers of the initial and final states.
15. Determine the ground-state energy for an electron in an infinite square well of width 10.0 nm.
16. Find the width of an infinite square well in which a proton's minimum energy is 100 eV.
17. A *quantum wire* is a conducting structure so thin that quantum effects are evident. Electron energies in a quantum wire are quantized and so, therefore, are electrical properties such as resistivity. A particular quantum wire is made from carbon nanotubes 1.0 nm in diameter. Approximating the structure as a one-dimensional infinite square well, find the energies (in eV) of an electron in (a) the ground state and (b) the first excited state.
18. One reason we don't notice quantum effects in everyday life is that Planck's constant h is so small. Treating yourself as a particle (mass 60 kg) in a room-sized one-dimensional infinite square well (width 2.6 m), how big would h have to be if your minimum possible energy corresponded to a speed of 1.0 m/s?
19. A particle is confined to a 1.0-nm-wide infinite square well. If the energy difference between the ground state and the first excited state is 1.13 eV, is the particle an electron or a proton?
20. A 3-g snail crawls at 0.5 mm/s between two rocks 15 cm apart. Treating this system as an infinite square well, determine the approximate quantum number. Does the correspondence principle permit the use of the classical approximation in this case?
21. An alpha particle (mass 4 u) is trapped in a uranium nucleus with diameter 15 fm. Treating the system as a one-dimensional square well, what would be the minimum energy for the alpha particle?
22. A quantum harmonic oscillator has ground-state energy 0.14 eV. What would be the system's classical oscillation frequency f?
23. Find the ground-state energy for a particle in a harmonic oscillator potential whose classical angular frequency ω is $1.0\times10^{17}\ \text{s}^{-1}$.
24. A harmonic oscillator emits a 1.1-eV photon as it undergoes a transition between adjacent states. Find its classical oscillation frequency f.
25. The ground-state energy of a harmonic oscillator is 4.0 eV. Find the energy separation between adjacent quantum states.
26. Your roommate is taking Newtonian physics, while you've moved on to quantum mechanics. He claims that QM can't be right, because he didn't see any evidence of quantized energy levels in a mass–spring harmonic oscillator experiment. You reply by calculating the spacing between energy levels of this system, which consists of a 1-g mass on a spring with $k = 80$ N/m. What is that spacing, and how does this help your argument?

Section 35.4 Quantum Mechanics in Three Dimensions

27. If all sides of a cubical box are doubled, what happens to the ground-state energy of a particle in that box?
28. A very crude model for an atomic nucleus is a cubical box 1 fm on a side. What would be the energy of a gamma ray emitted if a proton in such a nucleus made a transition from its first excited state to the ground state?
29. An electron is confined to a cubical box. For what box width will a transition from the first excited state to the ground state result in emission of a 950-nm infrared photon?

Problems

30. Find an expression for the normalization constant A for the wave function given by $\psi = 0$ for $|x| > b$ and $\psi = A(b^2 - x^2)$ for $-b \leq x \leq b$.
31. Suppose ψ_1 and ψ_2 are solutions of the Schrödinger equation for the same energy E. Show that the linear combination $a\psi_1 + b\psi_2$ is also a solution, where a and b are arbitrary constants.
32. An electron is trapped in an infinite square well 25 nm wide. Find the wavelengths of the photons emitted in these transitions: (a) $n = 2$ to $n = 1$; (b) $n = 20$ to $n = 19$; (c) $n = 100$ to $n = 1$.
33. An electron drops from the $n = 7$ to the $n = 6$ level of an infinite square well 1.5 nm wide. Find (a) the energy and (b) the wavelength of the photon emitted.
34. Show explicitly that the difference between adjacent energy levels in an infinite square well becomes arbitrarily small compared with the energy of the upper level, in the limit of large quantum number n.
35. An electron is in a narrow molecule 4.4 nm long, a situation that approximates a one-dimensional infinite square well. If the

electron is in its ground state, what is the maximum wavelength of electromagnetic radiation that can cause a transition to an excited state?

36. The ground-state energy for an electron in infinite square well A is equal to the energy of the first excited state for an electron in well B. How do the wells' widths compare?
37. Electrons in an ensemble of 10-nm-wide square-well systems are initially in the $n = 4$ state. Find the wavelengths of all spectral lines emitted as the electrons cascade to the ground state through all possible downward transitions.
38. Sketch the probability density for the $n = 2$ state of an infinite square well extending from $x = 0$ to $x = L$, and determine where the particle is most likely to be found.
39. An infinite square well extends from $-L/2$ to $L/2$. (a) Find expressions for the normalized wave functions for a particle of mass m in this well, giving separate expressions for even and odd quantum numbers. (b) Find the corresponding energy levels.
40. A particle is in the ground state of an infinite square well. What's the probability of finding the particle in the left-hand third of the well?
41. A laser emits 1.96-eV photons. If this emission is due to electron transitions from the $n = 2$ to $n = 1$ states of an infinite square well, what's the well width?
42. What's the probability of finding a particle in the central 80% of an infinite square well, assuming it's in the ground state?
43. Is quantization significant for macromolecules confined to biological cells? To find out, consider a protein of mass 250,000 u confined to a 10 μm-diameter cell. Treating this as a particle in a one-dimensional square well, find the energy difference between the ground state and the first excited state. Given that biochemical reactions typically involve energies on the order of 1 eV, what do you conclude about the role of quantization? BIO
44. In your physical chemistry course, you model hydrogen chloride as a hydrogen atom on a spring; the other end of the spring is attached to a rigid wall (the massive chlorine atom). In order to determine the spring constant in your model, you measure the minimum photon energy that will promote HCl molecules to their first excited state. The result is 0.358 eV. What do you calculate for the effective k?
45. A particle detector has a resolution 15% of the width of an infinite square well. What's the probability that the detector will find a particle in the ground state of the square well if the detector is centered on (a) the midpoint of the well and (b) a point one-fourth of the way across the well?
46. Find the probability that a particle in an infinite square well is located in the central one-fourth of the well for the quantum states $n =$ (a) 1, (b) 2, (c) 5, and (d) 20. (e) What's the classical probability in this situation?
47. A particle of mass m is in a region where its total energy E is less than its potential energy U. Show that the Schrödinger equation has nonzero solutions of the form $Ae^{\pm\sqrt{2m(U-E)}x/\hbar}$. Such solutions describe the wave function in quantum tunneling, beyond the turning points in a quantum harmonic oscillator, or beyond the well edges in a finite potential well.
48. (a) Use Equation 35.8 to draw an energy-level diagram for the first six energy levels of a particle in a cubical box, in terms of $h^2/8mL^2$, and (b) give the degeneracy of each.
49. The generalization of the Schrödinger equation to three dimensions is

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}\right)+U(x, y, z)\psi = E\psi$$

(a) For a particle confined to the cubical region $0 \le x \le L$, $0 \le y \le L$, $0 \le z \le L$, show by direct substitution that the equation is satisfied by wave functions of the form $\psi(x, y, z) = A\sin(n_x\pi x/L)\sin(n_y\pi y/L)\sin(n_z\pi z/L)$, where the n's are integers and A is a constant. (b) In the process of working part (a), verify that the energies E are given by Equation 35.8.

50. A 9-W laser beam shines on an ensemble of 10^{24} electrons, each in the ground state of a one-dimensional infinite square well 0.72 nm wide. The photon energy is just high enough to raise an electron to its first excited state. How many electrons can be excited if the beam shines for 10 ms?
51. A large number of electrons are confined to infinite square wells 1.2 nm wide. They're undergoing transitions among all possible states. How many visible lines (400 nm to 700 nm) are in the spectrum emitted by this ensemble of square-well systems?
52. A particle is in the nth quantum state of an infinite square well. (a) Show that the probability of finding it in the left-hand quarter of the well is

$$P = \frac{1}{4} - \frac{\sin(n\pi/2)}{2n\pi}$$

(b) Show that for odd n, the probability approaches the classical value $\frac{1}{4}$ as $n \to \infty$.

53. (a) Using the potential energy $U = \frac{1}{2}m\omega^2x^2$ discussed on page 635, develop the Schrödinger equation for the harmonic oscillator. (b) Show by substitution that $\psi_0(x) = A_0e^{-\alpha^2x^2/2}$ satisfies your equation, where $\alpha^2 = m\omega/\hbar$ and the energy is given by Equation 35.7 with $n = 0$. (c) Find the normalization constant A_0. You then have the ground-state wave function for the harmonic oscillator.
54. You're trying to convince a friend that nuclear energy represents a much more concentrated energy source than fossil fuels, whose combustion involves rearranging atomic electrons. For a rough comparison, you calculate the ground-state energy of a proton confined to 1-fm-diameter atomic nucleus and that of an electron confined to a 0.1-nm-diameter atom. Approximating each system as a one-dimensional infinite square well, what's the ratio of their ground-state energies?

Passage Problems

BIO *Quantum dots*, or *qdots*, are nanoscale crystals of semiconductor material that trap electrons in a potential well closely resembling the three-dimensional square well discussed in Section 35.4. Physicists, materials scientists, and semiconductor engineers have been studying qdots for their potential to miniaturize electronic components. More recently, qdots have been used in biology and medicine to "tag" individual molecules, helping scientists follow cellular processes (Fig. 35.19). Qdots also facilitate high-resolution imaging within the cell, and they show promise for medical diagnostics and targeting tumors for the delivery of anticancer agents. In the biomedical context, qdots work as replacements for traditional fluorescent dyes. Illuminating qdots promotes their electrons to higher energy levels; as they drop back, they emit photons of precise wavelength. A dot's size and structure determine this wavelength.

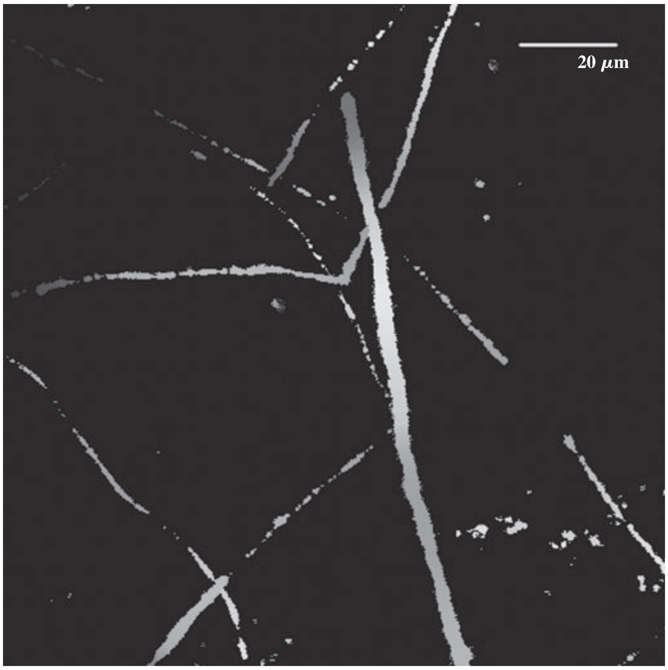


FIGURE 35.19 In this microscopic photo, motor protein molecules called dynein have been tagged with quantum dots, allowing their paths to be tracked (Passage Problems 55–58).

55. If a qdot's size is decreased, what happens to the wavelength of the photon emitted in a transition from the dot's first excited state to the ground state?
    a. The wavelength increases.
    b. The wavelength decreases.
    c. The wavelength is unchanged.

56. If the dot behaves as a perfectly cubical 3-D square well, the first excited state is
    a. nondegenerate.
    b. twofold degenerate.
    c. threefold degenerate.
    d. You can't tell without knowing the energy.

57. If the dot behaves as a perfectly cubical 3-D square well, the ground state is
    a. nondegenerate.
    b. twofold degenerate.
    c. threefold degenerate.
    d. You can't tell without knowing the energy.

58. If all three sides of a qdot are halved, its ground-state energy
    a. is halved.
    b. drops to one-fourth its original value.
    c. doubles.
    d. quadruples.

Answers to Chapter Questions

Answer to Chapter Opening Question

Quantum tunneling, the ability of particles to penetrate a barrier that classical physics says they don't have sufficient energy to overcome.

Answers to GOT IT? Questions

35.1. B's ground-state energy is 10^6 times higher than A's.
35.2. (c), which is almost twice the classical prediction.
35.3. The electron.

36 Atomic Physics

How do the principles of quantum physics explain the different chemical elements?

New Concepts, New Skills

By the end of this chapter you should be able to

- Evaluate the energies and angular momenta of hydrogen-like atoms (36.1).
- Calculate the probability of locating an electron in hydrogen-like atoms (36.1).
- Explain the role of electron spin, and evaluate spin angular momentum (36.2).
- Couple orbital and spin angular momenta to find allowed values for total angular momentum (36.2).
- Explain the exclusion principle and how it accounts for the periodic table of the elements (36.3, 36.4).
- Determine the electronic structure of multielectron atoms (36.4).
- Explain the origin of atomic spectra, and interpret energy-level diagrams for complex atoms (36.5).
- Explain how lasers work (36.5).

Connecting Your Knowledge

- This chapter applies the Schrödinger equation to hydrogen and hydrogen-like atoms (35.2).
- We'll build on the concepts of wave function and probability density, and we'll treat atomic electrons as quantum particles trapped in the potential well associated with the electric force (35.1, 35.3, 20.2, 22.2, 23.1).
- We'll revisit angular momentum (11.3) and find that it plays a major role in atomic physics.
- Your knowledge of chemistry will help you appreciate how quantum physics explains the periodic table of the elements.

In Chapter 35 we applied the Schrödinger equation to simplified quantum systems. Here we turn to the more realistic case of the atom, and explore how quantum mechanics explains atomic structure and the periodic table of the elements. We'll deal most thoroughly with the simplest atom, hydrogen, and we'll be more qualitative in describing multielectron atoms.

36.1 The Hydrogen Atom

Like a particle in a three-dimensional box, the electron in hydrogen is confined to a three-dimensional potential well. For the electron, the well results from the proton's electrostatic attraction. From Chapter 22 we know that the electric potential of the proton, treated as a point charge e, is $V(r) = ke/r$, with r the distance to the proton and the zero of potential at infinity. Electric potential is energy per unit charge, so we multiply by the electron charge $-e$ to get the potential energy of an electron in the presence of a proton:

$$U(r) = -\frac{ke^2}{r} \qquad (36.1)$$

We treat the massive proton as being at rest at the origin, so Equation 36.1 gives the electron's potential energy as a function of radial position r. We can therefore use Equation 36.1 as the potential energy in the Schrödinger equation for the hydrogen atom.

The Schrödinger Equation in Spherical Coordinates

Because the electron's potential energy depends on radial distance r, it's best to work in spherical coordinates, where the position of a point is given by its distance r from the origin along with two angles θ and ϕ that specify its orientation (Fig. 36.1). Converting the Schrödinger equation to spherical coordinates is straightforward but tedious; the result is

$$-\frac{\hbar^2}{2mr^2}\left[\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)+\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right)+\frac{1}{\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2}\right]-\frac{ke^2}{r}\psi=E\psi \quad \left(\begin{array}{l}\text{Schrödinger equation,}\\ \text{spherical coordinates}\end{array}\right) \tag{36.2}$$

where we've used Equation 36.1 for the potential-energy function.

Although Equation 36.2 looks forbidding, it can be solved readily using advanced techniques. For total energy E less than zero, corresponding to bound states in hydrogen's potential well, most solutions become infinite at large r and therefore aren't normalizable. As a result, only certain values of the energy E give acceptable bound-state solutions. For total energy greater than zero, the electron is unbound and any energy proves possible, as with the finite square well.

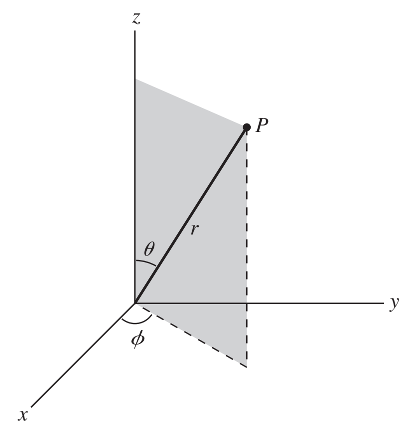


FIGURE 36.1 Spherical coordinates r, θ, ϕ provide an alternative to rectangular coordinates x, y, z.

The Hydrogen Ground State

In general, solutions to Equation 36.2 depend on all three variables r, θ, ϕ. But some solutions, including the ground state, are spherically symmetric—they depend only on r. Here we show that the ground state has the form of an exponential, and in the process derive the ground-state energy. Consider the function

$$\psi = Ae^{-r/a_0} \tag{36.3}$$

where A and a_0 are as yet undetermined constants, the latter with the units of length. For this spherically symmetric function, nothing depends on the angular variables θ and ϕ, so derivatives with respect to those variables are strictly zero. We're then dealing with a function of only one variable, so we can write total instead of partial derivatives. Equation 36.2 then becomes

$$-\frac{\hbar^2}{2mr^2}\frac{d}{dr}\left(r^2\frac{d\psi}{dr}\right)-\frac{ke^2}{r}\psi=E\psi \tag{36.4}$$

Substituting the proposed solution 36.3 for ψ gives

$$-\frac{\hbar^2}{2ma_0^2}+\frac{\hbar^2}{mra_0}-\frac{ke^2}{r}=E_1$$

(see Problem 42), where E_1 is the ground-state energy. This equation must be satisfied for all values of r, so the two r-dependent terms must cancel:

$$\frac{\hbar^2}{mra_0}=\frac{ke^2}{r}$$

or

$$a_0=\frac{\hbar^2}{mke^2}=5.29\times10^{-11}\text{ m}=0.0529\text{ nm}$$

This is precisely the **Bohr radius** that we introduced in Chapter 34.

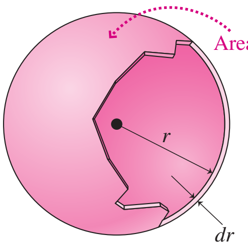


FIGURE 36.2 A thin shell has volume $dV = 4\pi r^2 dr$; thus the probability per unit radial distance is $4\pi r^2$ times the probability per unit volume.

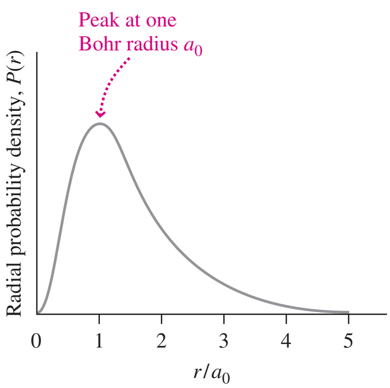


FIGURE 36.3 Radial probability density for the hydrogen ground state.

With the r terms gone, our expression for the ground-state energy becomes $E_1 = -\hbar^2/2ma_0^2 = -13.6\text{ eV}$, where the minus sign shows that the atom is a bound system. Thus Equation 36.3 is indeed a solution to the Schrödinger equation for hydrogen, with energy $E_1 = -13.6\text{ eV}$.

In deriving expressions for a_0 and E_1, we've shown how Schrödinger's theory gives two fundamental parameters of atomic physics: the Bohr radius and the hydrogen ground-state energy. Both agree with the values we found in Chapter 34 using the simpler Bohr model. But Bohr's theory still clings to the notion of classical orbits, with a_0 the ground-state orbital radius. Schrödinger's theory is truly quantum mechanical, representing the electron with its wave function ψ and associated probability distribution. The Bohr radius is no longer an actual orbital radius but instead determines atomic size only in a statistical sense.

The Radial Probability Distribution

Because the ground-state wave function falls off exponentially as e^{-r/a_0}, we're unlikely to find the electron at distances far greater than the Bohr radius. But where are we most likely to find it? Although ψ is greatest at $r = 0$, that's not the answer. In three dimensions the probability density ψ^2 is the probability *per unit volume* of finding the electron. In asking where we're most likely to find the electron, we want the probability *per unit radial distance*. Figure 36.2 shows a thin spherical shell with radius r and therefore area $4\pi r^2$. It has thickness dr, so its volume is $dV = 4\pi r^2\,dr$. Then the probability of finding the electron in this shell is $\psi^2\,dV = 4\pi r^2\psi^2\,dr$. The **radial probability density**, $P(r)$, is the probability per unit radius, or

$$P(r) = 4\pi r^2\psi^2 \qquad (\text{radial probability density}) \qquad (36.5)$$

For the hydrogen ground state, we use Equation 36.3 for ψ to get $P_1 = 4\pi r^2 A^2 e^{-2r/a_0}$. Figure 36.3 is a plot of this probability density, which peaks at $r = a_0$. Thus the single most likely place to find the electron in ground-state hydrogen is one Bohr radius from the proton.

EXAMPLE 36.1 The Hydrogen Atom: Normalization and the Probability Distribution

(a) Determine the normalization constant A in Equation 36.3. (b) Use the resulting wave function to find the probability that the electron in the hydrogen ground state will be found beyond the Bohr radius.

INTERPRET The wave function 36.3 contains an undetermined constant A. This problem is asking us to apply the normalization condition to find A and then use the concept of radial probability density to determine the probability of finding the electron beyond $r = a_0$.

DEVELOP The electron must be *somewhere* in the range $r = 0$ to $r = \infty$. Since $P(r)\,dr$ is the probability of finding the electron in a region of width dr, the normalization condition becomes $\int_0^\infty P(r)\,dr = 1$. So our plan is to evaluate this integral using the ground-state probability density $P_1 = 4\pi r^2 A^2 e^{-2r/a_0}$. We'll then solve for the unknown A. Then we can integrate again, this time from $r = a_0$ to $r = \infty$, to get the probability of finding the electron beyond a_0.

EVALUATE (a) Using the probability density P_1, the normalization condition becomes

$$\int_{r=0}^{r=\infty} 4\pi r^2 A^2 e^{-2r/a_0}\,dr = 1$$

We could evaluate using integration by parts; however, the result is in the integral table at the end of Appendix A. Replacing x by r and a by $-2/a_0$ in the table's expression for $\int x^2 e^{ax}\,dx$, we have

$$\int_0^\infty 4\pi A^2 r^2 e^{-2r/a_0}\,dr$$
$$= 4\pi A^2\left\{\frac{r^2 e^{-2r/a_0}}{(-2/a_0)} - \frac{2}{(-2/a_0)}\left[\frac{e^{-2r/a_0}}{(-2/a_0)^2}\left(-\frac{2}{a_0}r - 1\right)\right]\right\}\Bigg|_0^\infty = 1$$

The expression in curly brackets vanishes at $r = \infty$, and at $r = 0$ the exponentials are just 1, so we have $4\pi A^2\left[0 - \left(-\tfrac{1}{4}a_0^3\right)\right] = 1$, or $A = 1/\sqrt{\pi a_0^3}$.

(b) Now we change the lower limit on the integral from 0 to a_0. The result is

$$P(r > a_0) = \int_a^\infty 4\pi r^2 A^2 e^{-2r/a_0}\,dr$$
$$= 4\pi A^2 a_0^3\left(\tfrac{1}{2}e^{-2} + \tfrac{3}{4}e^{-2}\right) = 5\pi A^2 a_0^3 e^{-2}$$

With $A^2 = 1/\pi a_0^3$, this becomes $P(r > a_0) = 5e^{-2} \simeq 0.677$.

ASSESS Our result shows that about two-thirds of the time, the electron will be found beyond the Bohr radius. So although it's reasonable to say that the atom's radius is roughly the Bohr radius, both Fig. 36.3 and our result here show that there's no sharp cutoff that marks the "size" of the atom. ■

Excited States of Hydrogen

So far we've examined only the ground state of hydrogen. But Equation 36.2 admits many more normalizable solutions, corresponding to the excited states of hydrogen.

In general, each energy level is associated with one spherically symmetric wave function and a number of nonsymmetric ones. For historical reasons, the spherically symmetric states are called ***s* states**. The distinct energy levels are labeled by the quantum number n, called the **principal quantum number**. The ground state, for example, is the 1*s* state. The energy of the nth level, derivable from the Schrödinger equation, turns out to agree exactly with the earlier Bohr theory:

$$E_n = -\frac{1}{n^2}\frac{\hbar^2}{2ma_0^2} = \frac{E_1}{n^2} = \frac{-13.6\ \text{eV}}{n^2} \qquad \text{(hydrogen energy levels)} \tag{36.6}$$

The spherically symmetric state with energy E_2—that is, the 2*s* state—has wave function given by

$$\psi_{2s} = \frac{1}{4\sqrt{2\pi a_0^3}}\left(2 - \frac{r}{a_0}\right)e^{-r/2a_0} \tag{36.7}$$

By substituting this function into Equation 36.4, you can verify that the energy E_2 is given by Equation 36.6 (see Problem 60). The radial probability densities for the first three spherically symmetric states are plotted in Fig. 36.4; notice that the excited states correspond to larger, more "smeared-out" atoms.

Although we're discussing hydrogen, our results generalize to any single-electron atom—that is, to an atom of atomic number Z ionized $Z - 1$ times. For such an atom the potential-energy function becomes $-kZe^2/r$, and our calculations go through as before except that the factor e^2 is replaced by Ze^2. Then the energy levels become

$$E_n = -\frac{Z^2}{n^2}\frac{\hbar^2}{2ma_0^2} = \frac{Z^2E_1}{n^2} = -\frac{(13.6\ \text{eV})Z^2}{n^2} \tag{36.8}$$

reflecting the tighter binding of the more highly charged nucleus (see Fig. 36.5 and Problem 61).

GOT IT? 36.1 Which is the most appropriate estimate for the radial "size" of a hydrogen atom in its 2*s* state: (a) a_0, (b) $2a_0$, (c) $5a_0$, (d) $15a_0$?

Orbital Quantum Numbers and Angular Momentum

In the spherically symmetric *s* states, it turns out that the **orbital angular momentum** associated with the electron's motion is zero. This is at odds with Bohr's prediction that angular momentum should be an integer multiple of $\hbar$. And it makes clear that we can't be talking here about classical orbits, since motion in an elliptical or circular path entails angular momentum. But there are other solutions to the Schrödinger equation for hydrogen, solutions that aren't spherically symmetric and that have nonzero angular momentum.

For a given principal quantum number n, there are in fact n distinct solutions with different angular momenta. The **orbital quantum number** l distinguishes these states and ranges from 0 to $n - 1$. Thus the ground state $(n = 1)$ corresponds to the single value $l = 0$. Higher energy levels, however, are degenerate, meaning there's more than one l value for each $n > 1$. The orbital quantum number determines the magnitude L of the electron's orbital angular momentum:

$$L = \sqrt{l(l+1)}\hbar \qquad \text{(quantization of orbital angular momentum)} \tag{36.9}$$

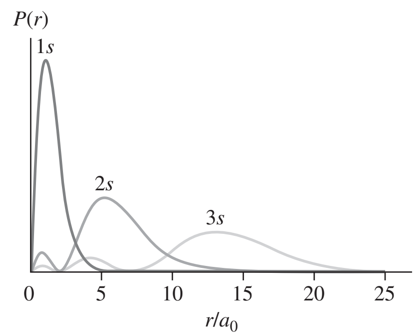


FIGURE 36.4 Radial probability densities for the spherically symmetric 1*s*, 2*s*, and 3*s* states of hydrogen.

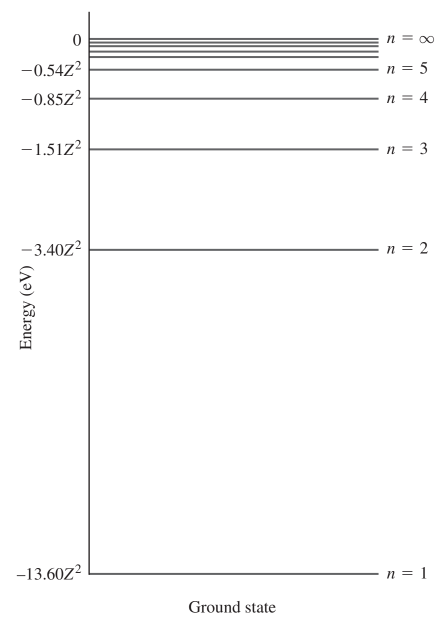


FIGURE 36.5 Energy-level diagram for a one-electron atom with atomic number Z. Energies scale as Z^2.

EXAMPLE 36.2 Orbital Angular Momentum: An Excited State

Find the possible values for the orbital angular momentum of an electron in the $n = 3$ state of hydrogen.

INTERPRET We're asked about the orbital angular momentum L, whose value follows from the orbital quantum number l. Thus we'll need the possible l values for $n = 3$.

DEVELOP For any n, there are n distinct l values, from 0 to $n - 1$. For $n = 3$, that means $l = 0$, 1, or 2. So our plan is to evaluate L using Equation 36.9, $L = \sqrt{l(l+1)}\hbar$, for these three l values.

EVALUATE With $l = 0$, Equation 36.9 gives $L = 0$; for $l = 1$, $L = \sqrt{2}\hbar$; and for $l = 2$, $L = \sqrt{6}\hbar$.

ASSESS $l = 0$ is the spherically symmetric 3s state, which we've seen has zero angular momentum. The higher-l states have increasing angular momentum. ■

States with l values 0, 1, 2, 3, 4, 5, . . . are given the letter labels $s, p, d, f, g, h, \ldots$. These combine with the principal quantum number n to specify both the energy and angular momentum of a state. Thus, the ground state is 1s, and the $n = 2$ state with $l = 1$ is the 2p state. (The lowercase letters $s, p, d, \ldots$ are used in discussing individual electrons, while the corresponding capital letters denote orbital angular momentum states of an entire atom. For one-electron hydrogen, the two are the same.)

Quantization of orbital angular momentum is another nonclassical aspect of quantum mechanics. In classical physics, an electron of a given energy can have any angular momentum, up to the maximum of a circular orbit (Fig. 36.6). At high n, the number of l values is so large that we don't notice quantization—another manifestation of Bohr's correspondence principle. But for low n, the quantum-mechanical discreteness of both energy and angular momentum is clearly evident.

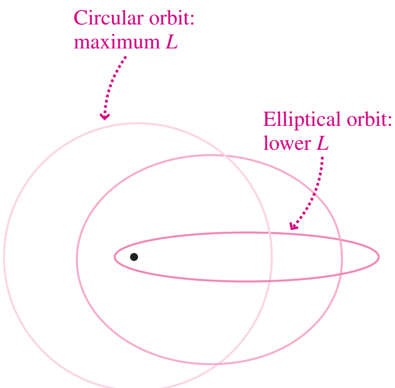


FIGURE 36.6 Classical electron orbits with the same energy but different angular momenta.

Space Quantization

Angular momentum is a vector, and the angular momentum vector is quantized not only in magnitude but also in direction—a phenomenon called **space quantization**. Space quantization of orbital angular momentum gives rise to a third quantum number, m_l. Space quantization becomes evident when an atom is in a magnetic field that establishes a preferred axis along which the angular momentum component can be measured; for this reason m_l is called the **orbital magnetic quantum number**.

Space quantization requires that the component L_z of orbital angular momentum along any chosen axis have only values given by

$$L = m_l\hbar \qquad \text{(space quantization)} \tag{36.10}$$

where m_l takes integer values from $-l$ to l. Thus an $l = 1$ state can have one of three possible m_l values: -1, 0, or $+1$, corresponding to angular momentum components $-\hbar$, 0, or $+\hbar$ along some axis. Since the magnitude of the angular momentum in an $l = 1$ state is $\sqrt{2}\hbar$ (see Example 36.2), none of these values corresponds to full alignment with the axis. Instead, we can think geometrically of the angular momentum vectors as being constrained to lie at angles $\cos^{-1}(L_z/L)$ to the axis; for $l = 1$ these angles are $\pm 45°$ and $90°$ (Fig. 36.7). Although the angle is useful for diagramming the angular momentum vector, we emphasize that the quantum numbers l and m_l tell everything there is to know about quantized orbital angular momentum. Quantum physicists, therefore, aren't usually concerned with the orientation of angular momentum vectors.

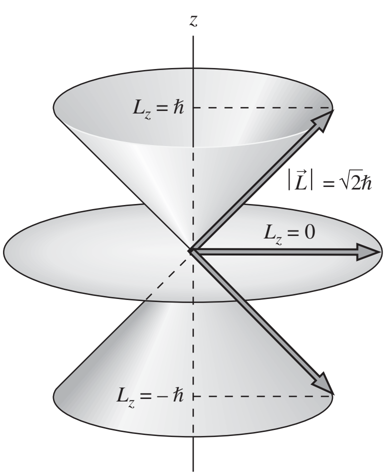


FIGURE 36.7 The three possible orientations for the angular momentum vector in the $l = 1$ state, where $L = \sqrt{2}\hbar$. Only the z-component is fixed; the x- and y-components are uncertain.

36.2 Electron Spin

Detailed observation of the hydrogen spectrum shows that spectral lines exhibit a fine splitting; where a lower-resolution spectrum shows one spectral line, at higher resolution there appears a closely spaced pair of lines. This splitting could not be explained using the three quantum numbers n, l, and m_l. In 1925 the Austrian physicist Wolfgang Pauli suggested that a fourth quantum number, capable of taking only two values, might be needed.

Soon Samuel Goudsmit and George Uhlenbeck realized that the spectral splitting could be explained if this fourth quantum number were associated with an intrinsic angular momentum, or **spin**, carried by the electron. Later, as we indicated in Chapter 35, Paul Dirac showed that electron spin follows from the requirement of relativistic invariance. Spin is an inherently quantum-mechanical property with no classical analog. Although it can be visualized crudely by imagining the electron to be a small sphere spinning about an axis, this classical picture is really inappropriate.

Spin angular momentum is quantized similarly to orbital angular momentum. But unlike the orbital quantum number l that takes a range of integer values, the electron spin quantum number s has only the single value $s = \frac{1}{2}$. The electron is therefore a **spin-$\frac{1}{2}$** particle. The magnitude of the spin angular momentum is related to the spin quantum number in the same way that the magnitude of orbital angular momentum is related to the orbital quantum number l:

$$S = \sqrt{s(s+1)}\hbar \qquad \text{(quantization of spin angular momentum)} \tag{36.11}$$

Since s takes only the value $\frac{1}{2}$, the magnitude of the electron spin angular momentum is $S = \frac{\sqrt{3}}{2}\hbar$.

Spin angular momentum also exhibits space quantization. That is, the component of spin along a chosen axis takes only the values

$$S_z = m_s\hbar \tag{36.12}$$

where the quantum number m_s has the two possible values $-\frac{1}{2}$ and $+\frac{1}{2}$. Figure 36.8 shows space quantization of electron spin.

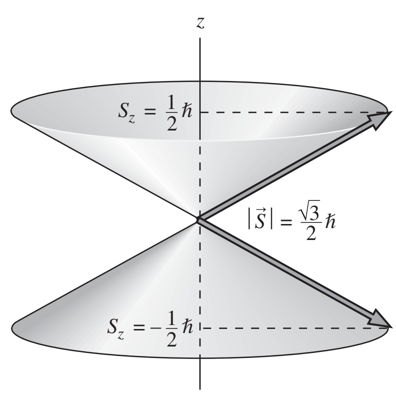


FIGURE 36.8 Space quantization of electron spin.

Magnetic Moment of the Electron

Together, the electron's spin and electric charge mean the electron behaves like a miniature current loop, with an intrinsic magnetic dipole moment. The dipole moment vector $\vec{M}$ associated with the spin angular momentum vector $\vec{S}$ is given by

$$\vec{M} = -\frac{e}{m}\vec{S} \tag{36.13}$$

with e/m the electron's charge-to-mass ratio (see Problem 67). Since the component of $\vec{S}$ on any axis can take only the values $\pm\frac{1}{2}\hbar$, the components of the magnetic moment can be only

$$M_z = \pm\frac{e\hbar}{2m} \tag{36.14}$$

The quantity $\mu_B = e\hbar/2m$ is a fundamental unit for measuring magnetic moments called the **Bohr magneton**; its value is $9.27\times10^{-24}\,\text{A}\cdot\text{m}^2$.

The ratio of magnetic moment to spin angular momentum is twice what we would expect classically for a charged particle in circular motion. Like spin itself, the factor of 2 is a relativistic effect first explained by Dirac. Actually, the factor is not quite 2 but approximately 2.00232, a result that follows from the theory of quantum electrodynamics.

The Stern–Gerlach Experiment

In 1922, Otto Stern and Walther Gerlach at the University of Hamburg demonstrated the quantization of atomic angular momentum vectors. The **Stern–Gerlach experiment** used a nonuniform magnetic field to separate a beam of silver atoms according to the component of their angular momentum along the field direction. T. E. Phipps and J. B. Taylor repeated the experiment in 1927, giving unambiguous verification of quantized electron spin. They used hydrogen atoms in the ground state; as we've seen, this state has zero orbital angular momentum, so the only angular momentum effects are due to electron spin. Classically a beam of hydrogen should be spread into a continuous band corresponding to angular momentum components from $\frac{-\sqrt{3}}{2}\hbar$ to $\frac{+\sqrt{3}}{2}\hbar$. But in fact the beam always splits in two, corresponding to the two angular momentum components $\pm\frac{1}{2}\hbar$. Figure 36.9 shows the experiment.

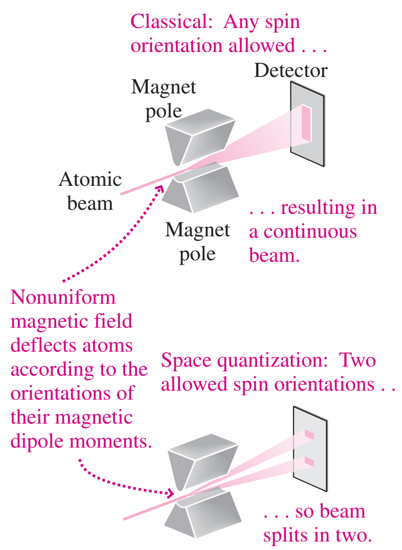


FIGURE 36.9 The Stern–Gerlach experiment.

GOT IT? 36.2 The nucleus of oxygen-17 has spin $\frac{5}{2}$. How many possible orientations are there for its spin angular momentum vector?

Total Angular Momentum and Spin-Orbit Coupling

Orbital and spin angular momenta combine through the process of **spin-orbit coupling** to give an atom's total angular momentum, $\vec{J}$:

$$\vec{J} = \vec{L} + \vec{S} \tag{36.15}$$

The magnitude J is quantized similarly to orbital and spin angular momenta:

$$J = \sqrt{j(j+1)}\hbar \qquad \text{(quantization of total angular momentum)} \tag{36.16}$$

For an atom with a single electron, the quantum number j takes the values

$$j = l \pm \tfrac{1}{2} \qquad \text{for } l \neq 0 \tag{36.17a}$$

$$j = \tfrac{1}{2} \qquad \text{for } l = 0 \tag{36.17b}$$

The state of an atom with total angular momentum J is specified by the principal quantum number, the capital letter designating the orbital angular momentum $(S, P, D, F, G, \ldots)$, and, as a subscript, the j value. Thus a hydrogen atom with $n = 3$, $l = 2$, and $j = \frac{3}{2}$ is designated $3D_{3/2}$.

Total angular momentum also exhibits space quantization, with the component of $\vec{J}$ on some axis given by

$$J_z = m_j\hbar \tag{36.18}$$

Here the quantum number m_j takes the values $(-j, -j+1, \ldots, j-1, j)$.

Derivation of these so-called **angular momentum coupling rules** is not easy, but we can understand them in terms of simple vector diagrams like those shown in Fig. 36.10.

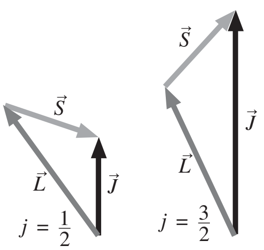


FIGURE 36.10 Spin-orbit coupling with $l = 1$, for which Equation 36.17a gives $j = \frac{1}{2}$ or $j = \frac{3}{2}$.

EXAMPLE 36.3 Spin-Orbit Coupling: Finding the Angular Momentum

(a) Find the possible magnitudes for the total angular momentum of hydrogen in the $l = 2$ state. (b) For each possible J, how many values are there for the component of $\vec{J}$ on a given axis?

INTERPRET We're asked about total angular momentum, which results from spin and orbital angular momentum contributions as shown in Fig. 36.10, and about the space quantization of $\vec{J}$.

DEVELOP (a) Equations 36.16, $J = \sqrt{j(j+1)}\hbar$, and 36.17a, $j = l \pm \frac{1}{2}$, determine J. With $l = 2$, our plan is to use Equation 36.17a to find the possible j values. Then we can apply Equation 36.16 to get each corresponding J. (b) Equation 36.18, $J_z = m_j\hbar$, determines the J_z values in terms of m_j, which ranges from $-j$ to j. So once we have the j's, we can determine the number of J_z values.

EVALUATE (a) With $l = 2$, Equation 36.17a gives $j = \frac{3}{2}$ or $j = \frac{5}{2}$. For $j = \frac{3}{2}$, Equation 36.16 yields $J = \sqrt{\frac{3}{2}\left(\frac{3}{2}+1\right)}\hbar = \frac{\sqrt{15}}{2}\hbar$; similarly, $j = \frac{5}{2}$ gives $J = \frac{\sqrt{35}}{2}\hbar$. (b) For $j = \frac{3}{2}$ there are four possible m_j values from $-j$ to j: $-\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}$, and correspondingly four J_z values, given by Equation 36.18: $-\frac{3}{2}\hbar, -\frac{1}{2}\hbar, \frac{1}{2}\hbar, \frac{3}{2}\hbar$. Similar counting gives six values for $j = \frac{5}{2}$.

ASSESS Figure 36.11 shows how the spin and orbit angular momenta combine to give the two possible j values in this example. ■

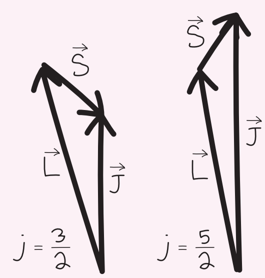


FIGURE 36.11 Vector diagrams for spin-orbit coupling with $l = 2$ (Example 36.3).

The two j values for a given l correspond to two distinct quantum states, and these states have slightly different energies. This energy difference is associated with the orientation of the electron's magnetic moment in a magnetic field that, in the electron's reference frame, results from the apparent motion of the positively charged nucleus around the electron—a field whose direction is that of the electron's orbital angular momentum $\vec{L}$. As Equation 36.13 shows, the electron's negative charge means its spin $\vec{S}$ and magnetic moment $\vec{M}$ have opposite directions. Because a magnetic dipole has the highest energy when it's oriented opposite the magnetic field, this means that the more nearly parallel alignment of $\vec{S}$ and $\vec{L}$—corresponding to $j = l + \frac{1}{2}$—has the higher energy.

In hydrogen, the magnitude of the energy difference between the $j = \frac{1}{2}$ and $j = \frac{3}{2}$ states of the first excited level is only 5×10^{-5} eV, far smaller than the 10.2-eV separation between this level and the ground state. Because the $n = 2, l = 1$ state is actually two states of slightly different energy, hydrogen atoms undergoing transitions from these states to the ground state emit two spectral lines slightly separated in wavelength. The term **fine structure** describes this and related spectral-line splittings. In the present example, the split spectral line is called a **doublet**. Figure 36.12 is an energy-level diagram showing the effect of spin-orbit splitting in hydrogen. Other fine-structure effects also alter energy levels; in hydrogen, for example, both levels shown in Fig. 36.12 actually drop below the Schrödinger prediction, and the $2S_{1/2}$ level is degenerate with $2P_{1/2}$.

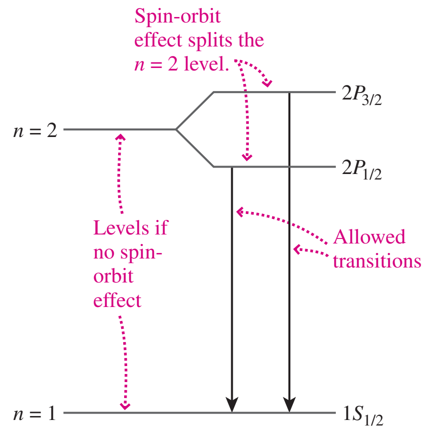


FIGURE 36.12 Energy-level diagram showing spin-orbit splitting of the $2P$ levels in hydrogen. Other fine-structure effects lower the energies of both $2P$ states and of the ground state, and make the $2S$ state degenerate with $2P_{1/2}$.

The spin-orbit effect results from a magnetic field internal to the atom itself. But splitting of energy levels also occurs in an external magnetic field and is called the **Zeeman effect**. We showed an example of Zeeman splitting in Fig. 35.18.

Since it has zero orbital angular momentum, the ground state does not exhibit spin-orbit splitting. But interaction of the electron spin with the magnetic dipole moment of the nucleus results in an even finer splitting known as **hyperfine structure**. The transition between the two hyperfine levels of the hydrogen ground state—corresponding physically to a change in the orientation of the electron spin vector—involves a photon of 21-cm wavelength. Radio astronomers use the 21-cm hydrogen radiation to map interstellar hydrogen in the cosmos.

36.3 The Exclusion Principle

In trying to understand why atomic electrons distributed themselves as they did, Pauli in 1924 developed his **exclusion principle**, which, loosely, states that **two electrons cannot be in the same quantum state**. Since an electron's quantum state includes its spin orientation specified by m_s, the exclusion principle means that at most two electrons can occupy a state whose other quantum numbers n, l, and m_l are the same.

The Pauli exclusion principle has profound implications for multielectron systems, requiring that most electrons remain in high-energy states (Fig. 36.13). If the exclusion principle didn't hold, atomic electrons would collapse to the ground state and there would be no such thing as chemistry or life! The exclusion principle even manifests itself at the cosmic scale, as the Application shows.

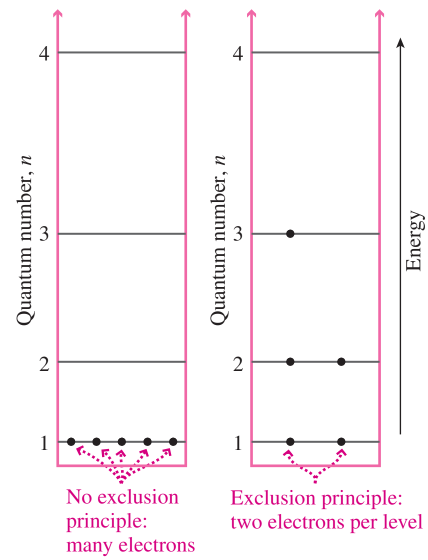
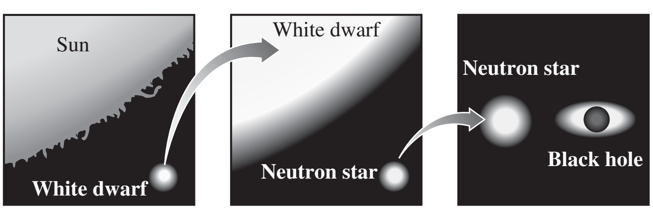


FIGURE 36.13 Particles in a square well, showing the exclusion principle's effect on electron distribution.

APPLICATION **White Dwarfs and Neutron Stars**



When a star exhausts its nuclear fuel, it collapses because there's no longer pressure to counter gravity. For a star with more than several times the Sun's mass, there's no force strong enough to halt the collapse, and the star becomes a black hole from which nothing can escape. But in less massive stars the collapse eventually halts because of a quantum-mechanical pressure associated with the exclusion principle.

When the Sun collapses some 5 billion years from now, its electrons will drop into the lowest available energy states. But as with the square well in Fig. 36.13, the exclusion principle requires that most of the Sun's 10^{57} electrons will end up in high-energy states. The associated **degenerate electron pressure**—independent of temperature, unlike the pressure of an ordinary gas—will stabilize the Sun as a **white dwarf**, about the size of Earth. For stars more massive than about 1.4 Sun masses, collapse proceeds until the protons and electrons merge to form neutrons. The neutrons, too, develop a degenerate pressure that stabilizes the resulting **neutron star** with a mass exceeding the Sun's crammed into a 20-km sphere! The figure compares the sizes of these stellar endpoints.

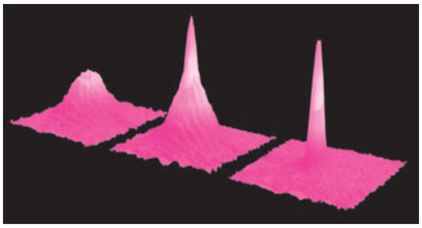


FIGURE 36.14 Velocity distribution of atoms in a Bose–Einstein condensate shows a large peak at the near-zero velocity of atoms all in their common ground state. The three peaks show the evolution from a normal gas to the condensate.

The Pauli exclusion principle quickly became fundamental to the developing quantum mechanics of the late 1920s. But physicists remained dissatisfied invoking this seemingly ad hoc rule with no theoretical basis. Late in the 1930s, following detailed analysis of relativistic quantum theories, Pauli finally showed that the exclusion principle, like spin, is ultimately grounded in the requirement of relativistic invariance. Pauli found that particles whose spin quantum number s is a half-integer (collectively called **fermions**) must necessarily obey the exclusion principle. On the other hand, particles with integer spin (called **bosons**) *do not* obey the exclusion principle. Photons, for example, are spin-1 particles and therefore an arbitrarily large number of them can occupy exactly the same quantum state. The laser, with its intense, coherent beam of light, is possible because the many photons that make up the beam are essentially all in the same state. In 1995 physicists at the University of Colorado first succeeded in producing an assemblage of bosonic matter all in the same quantum state (Fig. 36.14). This so-called **Bose–Einstein condensate** had been a goal of physicists since 1924, when the Indian physicist Satyendra Nath Bose first suggested the possibility. The Bose–Einstein condensate represents a truly new state of matter, in which thousands of atoms join quantum mechanically to behave as a single entity. Today physics labs around the world are experimenting with Bose–Einstein condensates, probing the fundamentals of quantum physics and developing applications including atom-beam analogs of the optical laser.

36.4 Multielectron Atoms and the Periodic Table

Our modern understanding of the chemical elements developed in the late 18th century, when chemists first distinguished compounds, such as water, from elements, defined as substances that couldn't be decomposed by chemical means. From the formulas for various compounds, chemists could determine the relative atomic masses of the elements. The first attempts to organize the elements systematically used atomic mass, but a breakthrough occurred in 1869, when the Russian chemist Dmitri Mendeleev set up a table with the approximately 60 elements then known. He left blanks where necessary to maintain the periodic occurrence of similar chemical properties. Elements filling the blanks were soon discovered, validating Mendeleev's periodic table and suggesting an underlying order in the composition of atoms. Then, early in the 20th century, studies of X-ray spectra led to a table organized by **atomic number** Z, the number of protons in the nucleus. When this was done, a number of elements missing from earlier periodic tables were identified. The modern periodic table is shown in Fig. 36.15 and is printed with atomic weights inside the back cover.

Explaining the Periodic Table

The orderly arrangement of elements in the periodic table enhances our understanding of chemistry and our ability to formulate new and useful compounds. But why does nature exhibit this regularity? The answer lies in the Schrödinger equation and the exclusion principle.

Solution of the Schrödinger equation for multielectron atoms is complicated by the interactions among the electrons; analytic solutions like those for hydrogen aren't generally available. But qualitatively, we still find energy levels characterized by the principal quantum number n. Each such level is called a **shell**; for historical reasons, the shells $n = 1, 2, 3, \ldots$ are also labeled with the letters $K, L, M, \ldots$. As with hydrogen, an electron at the nth energy level can have any of the n values $l = 0, 1, 2, \ldots, n - 1$ for the orbital quantum number. The different angular momentum states within a shell are termed **subshells**; subshells with the values $l = 0, 1, 2, 3, \ldots$ are labeled with the letters $s, p, d, f, \ldots$. Finally, for each subshell there are $2l + 1$ possible values of the magnetic orbital quantum number m_l, ranging from $-l$ to l. A state characterized by all three quantum numbers n, l, and m_l is called an **orbital**. Table 36.1 summarizes shell-structure notation; for completeness, the table also lists the spin quantum number m_s.

The structure of a multielectron atom is determined by the quantum states of its constituent electrons—their distribution among the shells, subshells, and orbitals. According

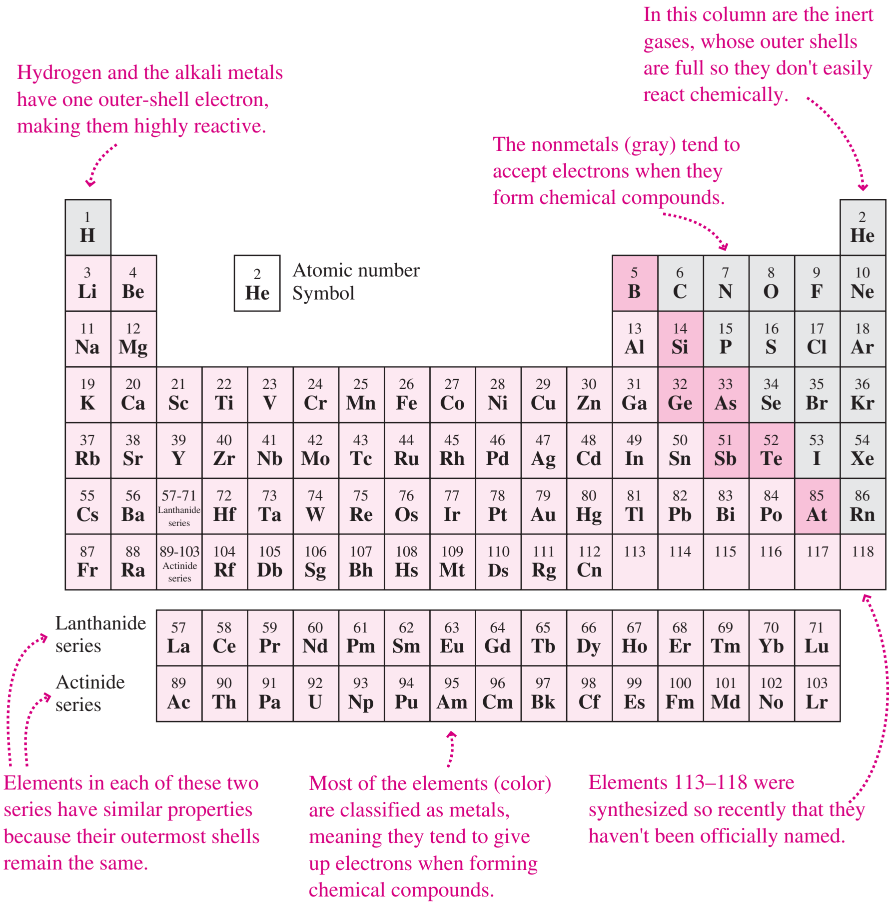


FIGURE 36.15 The periodic table. A larger version, with atomic weights, is inside the back cover; the names of the elements are in Appendix D.

Table 36.1 Atomic Shell Structure

Quantum Number	Shell Notation	Allowed Values	Letter Labels	Number of States
n	Shell	$1, 2, 3, \ldots$	$K, L, M, \ldots$	Infinite
l	Subshell	$0, 1, 2, \ldots, n-1$	$s, p, d, f, \ldots$	n
m_l	Orbital	$-l, -l+1, \ldots, l-1, l$	—	$2l+1$
m_s	—	$-\frac{1}{2}, +\frac{1}{2}$		2

to the exclusion principle, no two electrons can be in exactly the same quantum state; that is, they can't have the same values for all four quantum numbers n, l, m_l, and m_s. Since an atomic orbital is characterized by the three quantum numbers n, l, and m_l, the exclusion principle implies that at most two electrons can occupy a single orbital.

We're now ready to understand the ground-state electronic structure of multielectron atoms. The simplest is helium (He), with two electrons. The K shell $(n = 1)$ is the lowest possible energy level. As Table 36.1 shows, only the zero-angular-momentum s subshell is permitted within the K shell, and within that subshell there's only the single orbital corresponding to $m_l = 0$. But that orbital can accommodate two electrons. So in the ground state of helium, both electrons occupy the s subshell of the K shell. We describe this with the notation $1s^2$, where 1 stands for the principal quantum number n, s for the subshell, and the superscript 2 for the number of electrons in that subshell. The corresponding notation for hydrogen is $1s^1$.

After helium comes lithium (Li), with three electrons. From our analysis of helium, we know that the K shell is full with two electrons. So the third electron goes into the L shell, or $n = 2$ energy level. Of the subshells in the L shell, the s subshell turns out to have slightly lower energy than the others, so the third electron occupies the s subshell. Then the electronic configuration of lithium is $1s^2 2s^1$—that is, a helium-like core with a single outer electron in the s subshell of the $n = 2$ level.

Beryllium (Be), with four electrons, fills the $1s$ and $2s$ subshells; its designation is $1s^2 2s^2$. The fifth electron of boron (B) then goes into the $2p$ subshell, giving the structure $1s^2 2s^2 2p^1$. Table 36.1 shows that a p subshell $(l = 1)$ allows three m_l values—that is, three orbitals, capable of holding a total of six electrons. As we advance in atomic number, electrons continue to fill the p subshell. Finally, at neon $(Z = 10)$, the $2p$ subshell is full. Only with the next element, sodium $(Z = 11)$, does the $n = 3$ shell begin to fill. Table 36.2 lists electronic configurations for the elements hydrogen $(Z = 1)$ through argon $(Z = 18)$, along with their ionization energies (the energy required to remove the outermost electron).

Table 36.2 Electronic Configurations and Ionization Energies of Elements 1–18

Atomic Number, z	Element	Electronic Configuration	Ionization Energy (eV)
1	H	$1s^1$	13.60
2	He	$1s^2$	24.60
3	Li	$1s^2 2s^1$	5.390
4	Be	$1s^2 2s^2$	9.320
5	B	$1s^2 2s^2 2p^1$	8.296
6	C	$1s^2 2s^2 2p^2$	11.26
7	N	$1s^2 2s^2 2p^3$	14.55
8	O	$1s^2 2s^2 2p^4$	13.61
9	F	$1s^2 2s^2 2p^5$	17.42
10	Ne	$1s^2 2s^2 2p^6$	21.56
11	Na	$1s^2 2s^2 2p^6 3s^1$	5.138
12	Mg	$1s^2 2s^2 2p^6 3s^2$	7.644
13	Al	$1s^2 2s^2 2p^6 3s^2 3p^1$	5.984
14	Si	$1s^2 2s^2 2p^6 3s^2 3p^2$	8.149
15	P	$1s^2 2s^2 2p^6 3s^2 3p^3$	10.48
16	S	$1s^2 2s^2 2p^6 3s^2 3p^4$	10.36
17	Cl	$1s^2 2s^2 2p^6 3s^2 3p^5$	13.01
18	Ar	$1s^2 2s^2 2p^6 3s^2 3p^6$	15.76

Gaps mark ends of periodic-table rows.

An atom's chemical behavior is determined primarily by its outermost electrons, because these electrons interact most directly with nearby atoms and because they're most weakly bound to their nuclei. Table 36.2 shows that the outer-electron configurations for lithium through neon (Ne) are the same as the corresponding configurations for sodium (Na) through argon (Ar). The chemical properties of the corresponding atoms are therefore similar. Both lithium and sodium, for example, have a single electron in their outermost shell. As their relatively low ionization energies suggest, this electron is loosely bound, so it readily interacts with other atoms, giving these elements their extreme reactivity. Neon and argon, in contrast, both have completely filled outermost shells. All the outer-shell electrons have essentially the same energy; the corresponding ionization energy is high; and there's little tendency for these electrons to interact with other atoms. As a result, argon and neon are **inert**; they don't readily form chemical compounds, and at normal temperatures they're gases. Other element pairs also share similar properties. Fluorine and chlorine, for instance, each need only one more electron to achieve the energetically favorable inert-gas configuration. Consequently these elements readily accept electrons. Materials such as common salt, NaCl, owe their high melting points to the strong bond that results when reactive sodium gives up its outer electron to electron-accepting chlorine, and the resulting positive and negative ions bind strongly by the electrostatic force. We'll consider molecular bonding in the next chapter.

Beyond argon $(Z = 18)$, shielding effects of the inner electrons result in the $4s$ states having lower energy than the $3d$ states. Potassium $(Z = 19)$ thus has the electronic configuration $1s^2 2s^2 2p^6 3s^2 3p^6 4s^1$ rather than $1s^2 2s^2 2p^6 3s^2 3p^6 3d^1$. After potassium comes calcium, with two electrons in its single $4s$ orbital. But the $4p$ orbitals do have higher energy than the $3d$, so elements beyond calcium begin filling the $3d$ orbitals. The next ten elements, scandium through zinc, have chemical properties that vary only slightly because their two outermost electrons remain $4s$ electrons; collectively, they're **transition elements**. (Chromium, $Z = 24$, and copper, $Z = 29$, are minor exceptions; in these an extra electron goes into the $3d$ orbitals, leaving only one $4s$ electron.) Finally, elements 31 (gallium) through 36 (krypton) repeat the pattern of aluminum through argon shown in Table 36.2, as their $4p$ orbitals fill with electrons. Krypton, with its outer p subshell full, is again an inert gas.

CONCEPTUAL EXAMPLE 36.1 The Periodic Table

Explain the general structure of the periodic table's first five rows.

EVALUATE Each row of the periodic table starts with an element whose outermost shell contains a single s electron; these elements include hydrogen and the highly reactive alkali metals. Each row ends with an inert gas, its outermost p subshell full. The first row involves filling the $1s$ orbital only; since this orbital holds at most two electrons, there are only two elements in the first row. The second row has eight elements, associated with the filling of the $2s$ and $2p$ orbitals, as shown in Table 36.2. The third row is like the second, but with the $3s$ and $3p$ orbitals filling. Because the $4s$ orbitals have lower energy than the $3d$ orbitals, the third row ends with an inert gas whose $3p$ orbitals are full, and the fourth row begins as the $4s$ orbital begins to fill. Then come the elements $Z = 21$ through $Z = 30$, in which the $3d$ orbitals are filling; these make for ten additional elements in the fourth row. The fifth row is a repeat of the fourth, as first the $5s$ orbitals fill, then the higher-energy $4d$ orbitals, then the remaining $5p$ orbitals.

ASSESS Our explanation shows why elements in each column of the periodic table have similar chemical properties, while those properties generally change as we move across the table's rows.

MAKING THE CONNECTION Determine the electronic configuration of iron.

EVALUATE $Z = 26$ for iron, so we need to accommodate eight more electrons outside iron's argon-like core. Since iron is a transition element, its $4s$ orbitals fill before $3d$. So two electrons go into $4s$ and the remaining six into $3d$. Iron's electronic configuration is therefore $1s^2 2s^2 2p^6 3s^2 3p^6 3d^6 4s^2$.

The sixth and seventh rows don't quite fit our analysis in Conceptual Example 36.1. At element 57, lanthanum, the $4f$ orbitals begin filling while the outermost electron remains $6s$. This continues through element 71, lutetium, giving elements 57–71 similar chemical properties. These elements constitute the **lanthanide series**, and they're printed separately below the main table. Row seven repeats this pattern, with the **actinide series**. Seventh-row elements beyond uranium (element 92) are radioactive with half-lives that are short compared to Earth's age. They're not found naturally but are produced in particle accelerators, fission reactors, and nuclear explosions. Elements beyond the actinides are very short-lived.

Note the crucial role the exclusion principle plays in our discussion of the chemical elements. Without that principle, every atom in its ground state would have all its electrons in the $1s$ orbital. There would be no qualitative distinction among the elements, and the science of chemistry would not exist. Nor would there be any chemists or physicists; life itself would be impossible without the rich diversity of chemical compounds formed from the different elements.

36.5 Transitions and Atomic Spectra

Emission and absorption of photons with specific energies provide the most direct manifestation of quantized atomic-energy levels, and give rise to the spectral lines that permit precise analysis of atomic systems from the laboratory to distant astrophysical objects. Even simple hydrogen exhibits myriad quantum states. In multielectron atoms, the possibilities for electronic excitation are even more numerous. The spectra of atoms reflect this rich array of available quantum states.

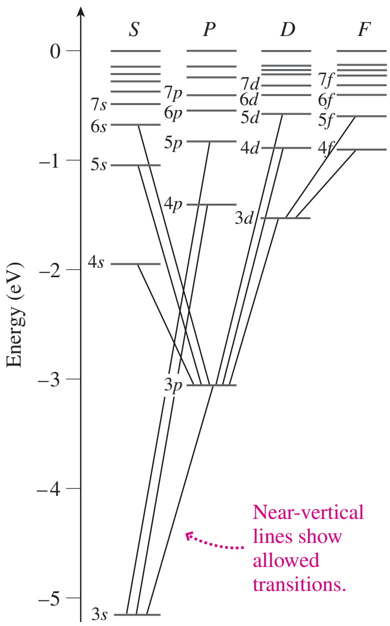


FIGURE 36.16 Energy-level diagram for sodium, neglecting spin-orbit splitting. Note the widely separated 4*s* and 4*p* levels, with 3*d* between them; this explains why the 4*s* orbital fills before 3*d*.

Selection Rules

Not all energy-level transitions are equally likely. So-called **selection rules** determine which are **allowed transitions**—those most likely to occur. One rule limits allowed transitions to those for which the orbital quantum number l changes by $\Delta l = \pm 1$; this and other selection rules are related to conservation of angular momentum. Quantum mechanics also provides a way of calculating transition probabilities, and from them the mean lifetimes of excited states. For outer electrons, excited states that de-excite by allowed transitions have typical lifetimes on the order of 10^{-9} s.

Transitions that are not allowed by selection rules are called **forbidden transitions**; most are not strictly impossible but just extremely unlikely. States that can lose energy only by forbidden transitions are **metastable states**; their lifetimes are many orders of magnitude longer than the nanosecond timescale for allowed transitions. "Glow in the dark" phosphorescent materials emit light through the slow de-excitation of metastable states. Forbidden spectral lines are valuable probes of low-density astrophysical gases in which collisions are rare, and atoms can therefore remain in metastable states.

Optical Spectra

Spectral lines in or near the visible involve transitions among the incompletely filled outer atomic shells. The alkali metals, with a single outer *s* electron, therefore produce spectra qualitatively similar to that of hydrogen. However, the more complicated structure of a multielectron atom shifts some energy levels (Fig. 36.16). Many of the transitions in Fig. 36.16 are actually doublets or triplets resulting from spin-orbit splitting (Fig. 36.17). The energy-level structure is even more complicated for atoms with more than one outer-shell electron.

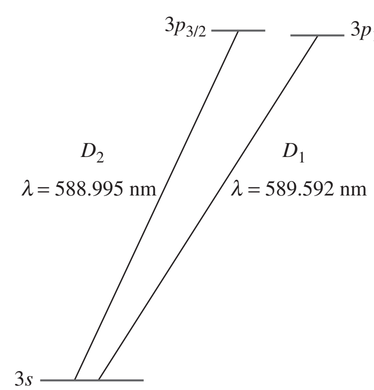


FIGURE 36.17 Magnified portion of sodium's energy-level diagram, showing spin-orbit splitting of the 3*p* level. Transitions from the two states result in the slightly different wavelengths of the sodium *D* doublet.

GOT IT? 36.3 Which of the transitions shown in Fig. 36.16 results in a photon of the shortest wavelength?

EXAMPLE 36.4 Atomic Spectra: The Sodium Doublet

Use Fig. 36.17 to determine the energy difference between the 3*p* states of sodium.

INTERPRET We're asked about the energy difference between two atomic states ($3p_{1/2}$ and $3p_{3/2}$), and we're given the wavelengths of photons emitted in transitions from those states to a common end state (3*s*). We know that those photons carry off energy equal to the difference between the energies of the starting and ending states.

DEVELOP The quantization condition $E = hf$ relates photon energy and frequency; since $f\lambda = c$, we also have $E = hc/\lambda$. Our plan is to use this expression for the energies for the two transitions shown in Fig. 36.17, and then subtract to get the energy difference between the 3*p* levels.

EVALUATE We have

$$\Delta E_{3p} = \frac{hc}{588.995\text{ nm}} - \frac{hc}{588.592\text{ nm}} = 3.42\times 10^{-22}\text{ J.}$$

ASSESS Our answer is about 2 meV, much lower than the eV-range energies associated with optical transitions themselves. That's expected, given the small separation between the 3*p* states evident in Fig. 36.17. In sodium, states below 3*s* are all full, so 3*s* is the lowest end state for optical transitions. ■

Spontaneous and Stimulated Transitions

What makes an electron jump between energy levels? In an upward transition, the electron must absorb the appropriate amount of energy. Generally, that energy is supplied by a photon whose energy is equal to the energy difference between the two levels; the process is called **stimulated absorption** (Fig. 36.18*a*). (Upward transitions can result from other processes as well, as, for example, in an energetic collision between two atoms or the interaction of a free electron with atomic electrons.)

For most downward transitions, however, there's no specific cause. An electron spontaneously jumps from a higher to a lower energy level and a photon is emitted; this is **spontaneous emission** (Fig. 36.18*b*). Although an individual spontaneous emission is a random event, quantum mechanics gives the probability per unit time for that event to occur; the inverse of that probability is the mean lifetime of the excited state.

In 1917 Einstein recognized a third possibility: Excited atoms can be stimulated to drop into lower energy states by the mere presence of a photon, again of energy appropriate to the transition. A second photon is emitted in the process, with the same energy and phase as the stimulating photon, and in the same direction. This process, **stimulated emission**, is the reverse of stimulated absorption (Fig. 36.18*c*).

Spontaneous emission, stimulated absorption, and stimulated emission play major roles in the transfer of radiation through gases. And stimulated emission is responsible for an important technological development: the laser.

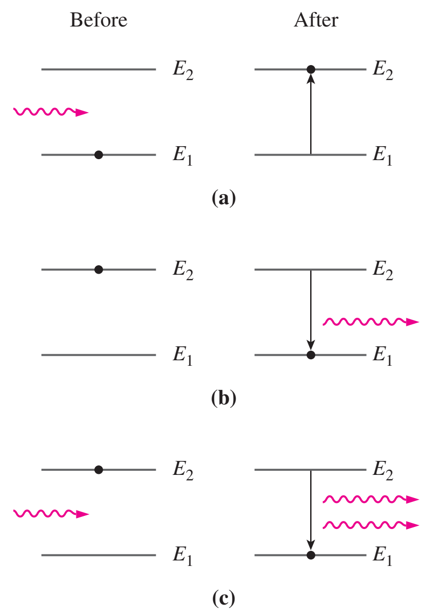
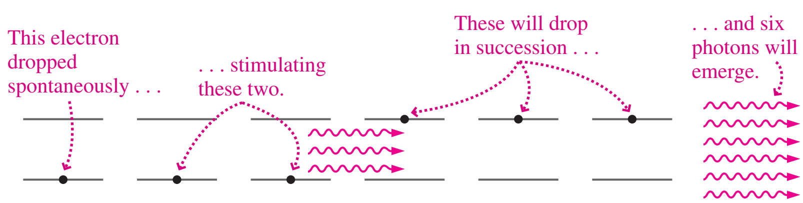


FIGURE 36.18 Interaction of photons with atomic electrons. Horizontal lines denote two atomic energy levels, and the wave is a photon with energy equal to the difference between the two levels. (a) Stimulated absorption; (b) spontaneous emission; (c) stimulated emission.

APPLICATION **The Laser**



As Fig. 36.18*c* suggests, stimulated emission is a way to multiply photons with identical wavelength and phase. The **laser**, whose name derives from **l**ight **a**mplification by **s**timulated **e**mission of **r**adiation, exploits this effect to produce an intense beam of coherent light. The key to laser action is a **population inversion**, an unusual situation with many atoms in an excited state. The excited state is usually metastable, to prevent spontaneous emission from de-exciting the atoms. Atoms are first excited to a higher state from which they quickly drop by spontaneous emission to the metastable state, where they're "stuck" by the lack of allowed transitions downward. The excitation process is called **pumping**, and the excitation energy source is the **pump**. Laser pumps include flash lamps, sunlight, other lasers, electric currents, chemical reactions, and even nuclear explosions.

With a large number of excited atoms, it isn't long before one randomly de-excites even from the metastable state. It emits a photon that passes by other excited atoms, causing stimulated emission. That makes more photons and still more stimulated emission, as shown in the figure. The process snowballs, resulting in an intense beam of photons with the same wavelength and phase. In a laser, the radiating medium sits in a cavity with mirrors at the ends; as the photons reflect off the mirrors and traverse the medium, more and more stimulated emission results, building up the beam intensity. One mirror is only partially reflective to allow the laser beam to emerge. Some lasers produce a short burst of radiation before being pumped to prepare for another burst. Others are pumped continuously, resulting in a continuous beam.

The first laser, built in 1960, used a ruby rod as the lasing medium, surrounded by a coiled flash lamp for the pump. Since then a myriad of laser types have been developed. Almost anything can be used as the lasing medium, provided a population inversion is possible. Laser media include gases, solids, liquids, semiconductors, and ionized plasmas. Natural laser action occurs even in interstellar gas clouds. Some lasers, especially those using chemical dyes or temperature-sensitive semiconductors, are tunable over a range of wavelengths.

Laser light is monochromatic, since all photons have essentially the same energy. It's coherent because the photons all have the same phase. Coherence allows the beam to travel long distances with minimal spreading and enables very precise focusing. Finally, laser light can be made extremely intense, since stimulated emission extracts energy from many atoms simultaneously. Since photons are spin-1 particles that don't obey the exclusion principle, there's no limit to the number of photons in a laser beam. Small lasers like those used in laser pointers have power outputs in the sub-milliwatt range, while large continuous lasers in excess of 1 MW are available, and pulsed outputs of 100 MW and higher have been achieved. Multiple laser systems, like those used in nuclear fusion experiments, reach peak powers of 10^{14} W—100 times the output of all the world's electric generating plants.

Today lasers are ubiquitous. They're used in commonplace applications like bar-code scanners and CD/DVD drives. Medical lasers correct vision, whiten teeth, and perform bloodless surgery. Biologists use laser beams as "optical tweezers" to manipulate microscopic structures within cells. Lasers have replaced older technologies in surveying, leveling, and measuring instruments used in construction. Industrial lasers cut metal, shape gears, and harden surfaces. Semiconductor lasers drive the optical fibers that carry communications signals and Internet traffic. Military lasers lock on targets for precise weapons guidance. Ultrafast lasers probe chemical reactions that occur on femtosecond timescales. Lasers halt atoms' thermal motion, cooling materials to nanokelvin levels and enabling Bose–Einstein condensates. Laser beams reflected from the Moon measure its distance to within a few centimeters, testing Einstein's general relativity. Holograms capture interfering laser wavefronts, creating three-dimensional images. Future laser applications may include laser-driven spaceflight and the use of lasers to clear space debris.

CHAPTER 36 SUMMARY

Big Picture

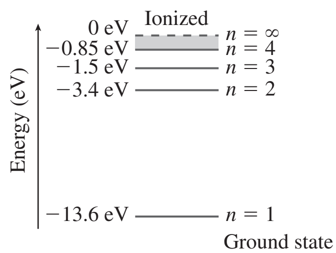
The big idea here is that atomic electrons are quantum particles trapped in the 3-dimensional potential well associated with the electric force. Solving the Schrödinger equation then leads to quantized energy levels. Considerations of electron spin and orbital angular momentum introduce subtle details into the atom's energy-level structure. The **exclusion principle** permits only one atomic electron per quantum state, and this fact underlies the shell structure of atoms and the periodic table of the elements.

Key Concepts and Equations

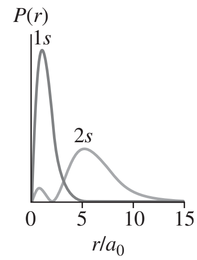
The **principal quantum number** n determines the energy levels in hydrogen:

$$E_n = -\frac{1}{n^2}\frac{\hbar^2}{2ma_0^2} = \frac{E_1}{n^2} = \frac{-13.6\text{ eV}}{n^2}$$

For $n = 1$ the electron is most likely to be found one Bohr radius a_0 from the nucleus; in higher-energy states it's likely to be farther away.



Energy-level diagram



Radial probability density

The orbital quantum number l determines the angular momentum:

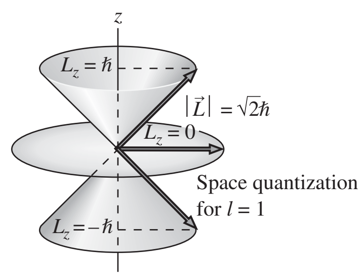
$$L = \sqrt{l(l+1)}\hbar$$

where l ranges from 0 to $n - 1$.

The orbital magnetic quantum number m_l determines the component of the angular momentum along any given axis:

$$L_z = m_l\hbar$$

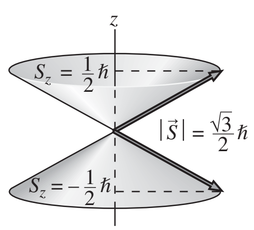
This is **space quantization,** where m_l ranges from $-l$ to l.



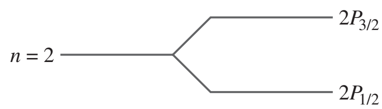
Electrons are **spin-$\frac{1}{2}$** particles or **fermions**; the component of their spin angular momentum on a given axis is $\pm\frac{1}{2}\hbar$.

Electron spin gives rise to the electron's intrinsic magnetic dipole moment, characterized by the **Bohr magneton**:

$$\mu_B = e\hbar/2m = 9.27\times10^{-24}\text{ A}\cdot\text{m}^2$$



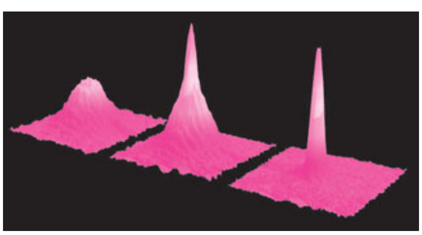
Spin-orbit coupling results in **fine-structure** splitting of atomic-energy levels.



Spin angular momentum $\vec{S}$ and total angular momentum $\vec{J}$ obey quantization rules similar to those of orbital angular momentum.

Applications

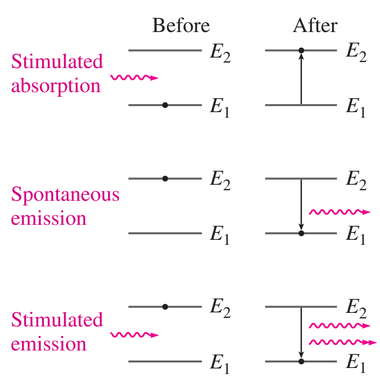
Bosons are particles with integer spin. They don't obey the exclusion principle, allowing many particles to be in the same state, as happens in a **Bose–Einstein condensate** or a laser beam.



Forming a Bose–Einstein condensate

In stimulated absorption, an electron absorbs a photon and jumps to a higher energy level.

Electrons in excited states can drop to lower energy states by either **spontaneous** or **stimulated emission**.



For Thought and Discussion

1. The electron in a hydrogen atom is somewhat like a particle confined to a three-dimensional box. In the atom, what plays the role of the confining box?
2. A friend who hasn't studied physics asks you the size of a hydrogen atom. How do you answer?
3. How many quantum numbers are required to specify fully the state of a hydrogen atom?
4. Both the Bohr and Schrödinger theories predict the same ground-state energy for hydrogen. Do they agree about the angular momentum in the ground state? Explain.
5. Is it possible for a hydrogen atom to be in the $2d$ state? Explain.
6. Can the component of a quantized angular momentum measured on a given axis ever equal the magnitude of the angular momentum vector? Explain.
7. The electron is a spin-$\frac{1}{2}$ particle. Does this mean the electron's intrinsic angular momentum is $\frac{1}{2}\hbar$? Explain.
8. How does the Stern–Gerlach experiment provide convincing evidence for space quantization?
9. Why is there no spin-orbit splitting in hydrogen's ground state?
10. How does the exclusion principle explain the diversity of chemical elements?
11. Helium and lithium exhibit very different chemical behavior, yet they differ by only one unit of nuclear charge. Explain.
12. Why is stimulated emission essential for laser action?
13. What distinguishes a Bose–Einstein condensate from ordinary matter?

Exercises and Problems

Exercises

Section 36.1 The Hydrogen Atom

14. Using physical constants accurate to four significant figures (see inside front cover), verify the numerical values of the Bohr radius a_0 and the hydrogen ground-state energy E_1.
15. A group of hydrogen atoms is in the same excited state, and photons with at least 1.5-eV energy are required to ionize these atoms. What's the quantum number n for the initial excited state?
16. Find the maximum possible magnitude for the orbital angular momentum of an electron in the $n = 7$ state of hydrogen.
17. Which of the following is not a possible value for the magnitude of the orbital angular momentum in hydrogen: (a) $\sqrt{12}\hbar$; (b) $\sqrt{20}\hbar$; (c) $\sqrt{30}\hbar$; (d) $\sqrt{40}\hbar$; (e) $\sqrt{56}\hbar$?
18. The orbital angular momentum of the electron in a hydrogen atom has magnitude 2.585×10^{-34} J·s. Find its minimum possible energy.
19. What's the orbital quantum number for an electron whose orbital angular momentum has magnitude $L = \sqrt{30}\hbar$?
20. A hydrogen atom is in the $6f$ state. Find (a) its energy and (b) the magnitude of its orbital angular momentum.
21. Give a symbolic description for the state of the electron in a hydrogen atom with total energy -1.51 eV and orbital angular momentum $\sqrt{6}\hbar$.

Section 36.2 Electron Spin

22. Verify the value of the Bohr magneton μ_B in Equation 36.14.
23. Theories of quantum gravity predict a spin-2 particle called the *graviton.* What would be the magnitude of the graviton's spin angular momentum?
24. Some very short-lived particles known as delta resonances have spin $\frac{3}{2}$. Find (a) the magnitude of their spin angular momentum and (b) the number of possible spin states.
25. What are the possible j values for a hydrogen atom in the $3D$ state?

Section 36.3 The Exclusion Principle

26. An infinite square well contains nine electrons. Find the energy of the highest-energy electron in terms of the ground-state energy E_1.
27. A quantum harmonic oscillator with frequency ω contains 21 electrons. What's the energy of the highest-energy electron?

Section 36.4 Multielectron Atoms and the Periodic Table

28. Use shell notation to characterize rubidium's outermost electron.
29. Write the full electronic structure of scandium.
30. Write the full electronic structure of bromine.

Section 36.5 Transitions and Atomic Spectra

31. Show that the wavelength λ in nm of a photon with energy E in eV is $\lambda = 1240/E$.
32. The $4f \to 3p$ transition in sodium produces a spectral line at 567.0 nm. Find the energy difference between these two levels.
33. The $4p \to 3s$ transition in sodium produces a double spectral line at 330.2 and 330.3 nm. What's the energy splitting of the $4p$ level?

Problems

34. Adapt part (b) of Example 36.1 to find the probability that an electron in the hydrogen ground state will be found beyond two Bohr radii.
35. Determine the principal and orbital quantum numbers for a hydrogen atom whose electron has energy -0.850 eV and orbital angular momentum $L = \sqrt{12}\hbar$.
36. Find (a) the energy and (b) the magnitude of the orbital angular momentum for an electron in the $5d$ state of hydrogen.
37. Assuming the Moon's orbital angular momentum is quantized, estimate its orbital quantum number l.
38. The maximum possible angular momentum for a hydrogen atom in a certain state is $30\sqrt{11}\hbar$. Find (a) the principal quantum number and (b) the energy.
39. A hydrogen atom is in an $l = 2$ state. What are the possible angles its orbital angular momentum vector can make with a given axis?
40. A hydrogen atom has energy $E = -0.850$ eV. Find the maximum possible values for (a) its orbital angular momentum and (b) the component of that angular momentum on a given axis.
41. An electron in hydrogen is in the $5f$ state. What possible values, in units of $\hbar$, could a measurement of the orbital angular momentum component on a given axis yield?
42. Substitute Equation 36.3 into Equation 36.4 and carry out the differentiations to show that you get the first unnumbered equation following Equation 36.4.
43. Differentiate the radial probability density for the hydrogen ground state, and set the result to zero to show that the electron is most likely to be found at one Bohr radius.
44. Repeat Exercise 25 for the case where you know only that the principal quantum number is 3; that is, l might have any of its possible values.
45. A hydrogen atom is in the $4F_{5/2}$ state. Find (a) its energy in units of the ground-state energy, (b) its orbital angular momentum in

units of $\hbar$, and (c) the magnitude of its total angular momentum in units of $\hbar$.

46. Suppose you put five electrons into an infinite square well of width L. Find an expression for the minimum energy of this system, consistent with the exclusion principle.
47. A harmonic oscillator potential of natural frequency ω contains eight electrons and is in its lowest-energy state. (a) What is its energy? (b) What would the lowest energy be if the electrons were replaced by spin-1 particles of the same mass?
48. You work for a nanotechnology company developing a new quantum device that operates essentially as a one-dimensional infinite square well of width 2.5 nm. You're asked to specify the maximum number of electrons in the device before the total electron energy exceeds 25 eV. Your answer?
49. Determine the electronic configuration of copper.
50. An electron in a highly excited state of hydrogen $(n_1 \gg 1)$ drops into the state $n = n_2$. Find the lowest value of n_2 for which the emitted photon will be in the infrared.
51. A solid-state laser made from lead–tin selenide has a lasing transition at a wavelength of 30 μm. If its power output is 2.0 mW, how many lasing transitions occur each second?
52. For hydrogen, fine-structure splitting of the $2p$ state is only about 50 μeV. What percentage is this difference of the photon energy emitted in the $2p \rightarrow 1s$ transition? Your answer shows why it's hard to observe spin-orbit splitting in hydrogen.
53. Find the probability that the electron in the hydrogen ground state will be found in the radial-distance range $r = a_0 \pm 0.1a_0$.
54. **BIO** You've acquired a laser for your dental practice. It produces 400-mJ pulses at 2.94-μm wavelength. A patient wonders about the number of photons in each pulse, and where they lie in the EM spectrum. Your answer?
55. What's the most orbital angular momentum that could be added to an atomic electron initially in the $6d$ state without changing its principal quantum number? What would be the new state?
56. A hydrogen atom is in an F state. (a) Find the possible values for its total angular momentum. (b) For the state with the greatest angular momentum, find the number of possible values for the component of $\vec{J}$ on a given axis.
57. A hydrogen atom is in the $2s$ state. Find the probability that its electron will be found (a) beyond one Bohr radius and (b) beyond 10 Bohr radii.
58. Show that the maximum number of electrons in an atom's nth shell is $2n^2$.
59. Form the radial probability density $P_2(r)$ associated with the ψ_{2s} state of Equation 36.7, and find the electron's most probable radial position.
60. Substitute the wave function ψ_2 of Equation 36.7 into Equation 36.4 to verify that the equation is satisfied and that the energy is given by Equation 36.6 with $n = 2$.
61. (a) Verify Equation 36.8 by considering a single-electron atom with nuclear charge Ze instead of e. (b) Calculate the ionization energies for single-electron versions of helium, oxygen, lead, and uranium.
62. **BIO** Excimer lasers for vision correction generally use a combination of argon and fluorine to form a molecular complex that can exist only in an excited state. Stimulated de-excitation produces 6.42-eV photons, which form the laser's intense beam. What's the corresponding photon wavelength, and where in the spectrum does it lie?
63. A selection rule for the infinite square well allows only those transitions in which n changes by an odd number. Suppose an infinite square well of width 0.200 nm contains an electron in the $n = 4$ state. (a) Draw an energy-level diagram showing all allowed transitions that could occur as this electron drops toward the ground state, including transitions from lower levels that could be reached from $n = 4$. (b) Find all the possible photon energies emitted in these transitions.
64. An ensemble of one-electron square-well systems of width 1.17 nm all have their electrons in highly excited states. They undergo all possible transitions in dropping toward the ground state, obeying the selection rule that Δn must be odd. (a) What wavelengths of visible light are emitted? (b) Is there any infrared emission? If so, how many spectral lines lie in the infrared?
65. Use the radial probability density from Equation 36.5 and the normalized ground-state hydrogen wave function from Equation 36.3 and Example 36.1 to calculate the average radial distance r_{av} for an electron in the ground state. (*Note:* Because the probability-density curve isn't symmetric, the average radial distance isn't the same as the most probable distance shown in Fig. 36.3.)
66. Follow the procedure in Problem 65 to calculate the average radial distance for an electron in the $2s$ state of hydrogen.
67. The ratio of the magnetic moment, in units of the Bohr magneton μ_B, to the angular momentum, in units of $\hbar$, is called the g-factor. (a) Show that the classical orbital g-factor for an atomic electron in a circular Bohr orbit is $g_L = 1$. (b) Show that Equation 36.13 gives $g_S = 2$ for the g-factor associated with electron spin.
68. You work for a company that makes red helium–neon lasers widely used in physics experiments. Figure 36.19 shows an energy-level diagram for this laser. An electric current excites helium to a metastable level E_1 at 20.61 eV above the ground state. Collisions transfer energy to neon atoms, exciting them to $E_2 = 20.66$ eV. The lasing transition drops the atoms to E_3, emitting a 632.8-nm photon in the process. You're asked to find the maximum possible efficiency for this laser—that is, the light energy emitted as a percentage of the energy supplied to excite the atoms. Your answer?

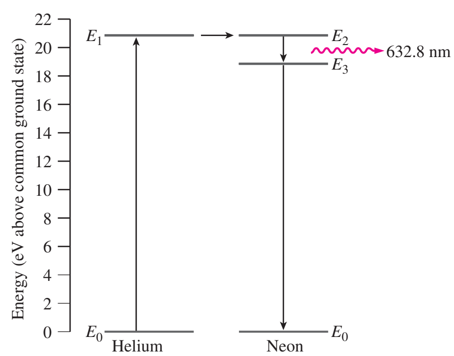


FIGURE 36.19 Energy-level diagram for the helium–neon laser (Problem 68).

Passage Problems

With sufficient energy, it's possible to eject an electron from an inner atomic orbital. A higher-energy electron will then drop into the unoccupied state, emitting a photon with energy equal to the difference between the two levels. For inner-shell electrons, photon energies are in the keV range, putting them in the X-ray region of the spectrum. These **characteristic X rays** are labeled with the letter indicating the shell to which the electron drops, followed by a Greek letter indicating the higher level from which it drops; thus $K\alpha$ designates a transition from the L shell to the K shell.

Characteristic X rays provide scientists and physicians with an important diagnostic tool. Environmental scientists bombard pollution samples with high-energy electrons, knocking out inner-shell electrons and thus producing X-ray spectra that help identify contaminants (Fig. 36.20*a*). Geologists do the same with rocks. Medical radiologists reverse the process, exploiting the fact that X rays cause inner-shell transitions as well as complete ejection of inner-shell electrons. In particular, radiologists use the element barium in this way to produce high-contrast X-ray images of the intestinal tract (Fig. 36.20*b*).

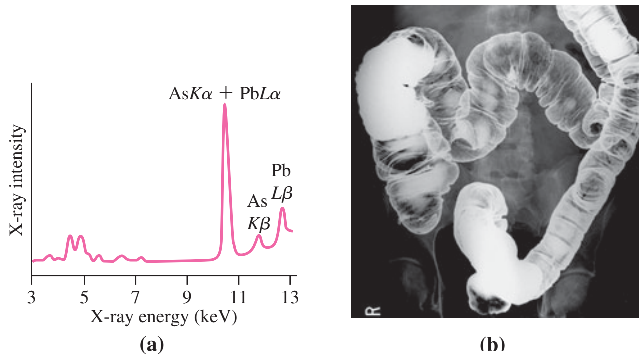


FIGURE 36.20 Passage Problems 69–72. (a) An X-ray spectrum from air pollutants trapped on a filter. The labeled peaks show the presence of lead (Pb) and arsenic (As), as evidenced by $K\alpha$, $K\beta$, $L\alpha$, and $L\beta$ characteristic X rays. (b) X-ray of an intestinal tract, made by coating the intestinal wall with X-ray-opaque barium

69. Molybdenum's X-ray spectrum has its $K\alpha$ peak at 17.4 keV. The corresponding X-ray wavelength is closest to
    a. 1 pm.
    b. 100 pm.
    c. 1 nm.
    d. 100 nm.

70. In general, how should the energy of an element's $L\alpha$ X rays compare with the energy of its $K\alpha$ X rays?
    a. They have less energy.
    b. They have the same energy.
    c. They have greater energy.
    d. You can't tell without knowing the element.

71. Elements A and B have atomic numbers Z_A and $Z_B = 2Z_A$. How do you expect element B's $K\alpha$ X-ray energy to compare with that of element A?
    a. B's $K\alpha$ energy should be about one-fourth that of A.
    b. B's $K\alpha$ energy should be about half that of A.
    c. B's $K\alpha$ energy should be about twice that of A.
    d. B's $K\alpha$ energy should be about four times that of A.

72. Emission of characteristic X rays occurs in the context of multielectron atoms that generally have all but one of their electrons present. You should therefore expect the X-ray energies to be described
    a. quite accurately by Bohr's atomic theory.
    b. through hydrogen-like solutions to the Schrödinger equation.
    c. only approximately by Bohr's or hydrogenic solutions to the Schrödinger equation.

Answers to Chapter Questions

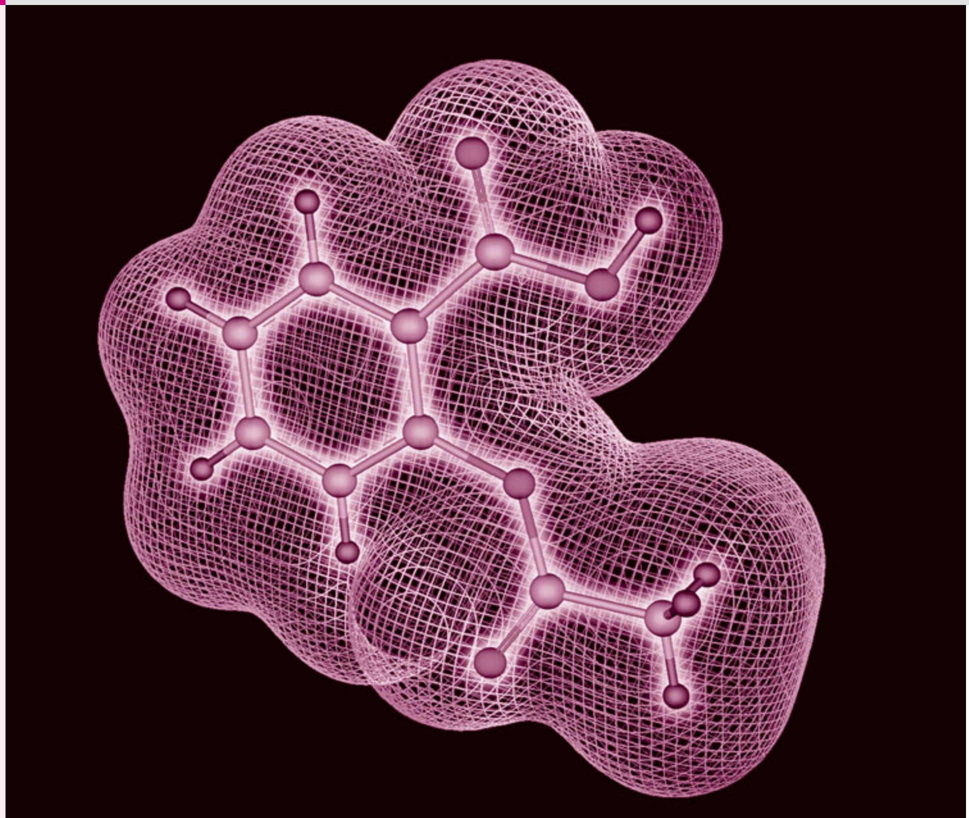
Answer to Chapter Opening Question

Only one electron is allowed in a given atomic quantum state, leading to the shell structure of atoms and to chemical properties based on the outermost atomic electrons.

Answers to GOT IT? Questions

36.1. (c).
36.2. Six.
36.3. $5p \rightarrow 3s$.

37 Molecules and Solids



What equation determines the structure of a complex molecule, like the aspirin molecule shown here?

New Concepts, New Skills

By the end of this chapter you should be able to

- Describe the principal kinds of molecular bonding (37.1).
- Evaluate energy levels associated with molecular rotation and vibration (37.2).
- Predict and analyze molecular spectra (37.2).
- Describe quantitatively the structure of ionic solids (37.3).
- Describe the electrical properties of insulators, conductors, and semiconductors using band theory (37.3).
- Explain the role of the Fermi energy in metallic conduction and its temperature dependence (37.3).
- Distinguish between type I and type II superconductors (37.4).

Connecting Your Knowledge

This chapter applies quantum mechanics to molecules and solids. It's necessarily more qualitative than preceding chapters, but it draws on and elaborates the following quantitative notions:

- Potential-energy curves (7.4)
- The quantum harmonic oscillator (35.3)
- Angular momentum and rotational inertia (10.3, 11.3)
- Rotational energy (10.4)
- Quantization of angular momentum (36.2, 36.3)
- The interaction of electric dipoles (20.5) as the origin of the van der Waals force between molecules (17.1)
- Mechanisms of electrical conduction, including semiconductors (24.2)
- The specific heats of gases (18.3)

In principle, we could apply the Schrödinger equation to all the particles making up a molecule or even a solid, but in practice that's difficult for all but the simplest molecules. Ever-increasing computer power has brought more complex molecules within reach of structural calculations based in the Schrödinger equation, but in many cases it's appropriate to describe molecular and solid structure more qualitatively.

37.1 Molecular Bonding

The binding of atoms into molecules involves both electric forces and quantum-mechanical effects associated with the exclusion principle. Although individual atoms are electrically neutral, the distribution of charge within them gives rise to attractive or repulsive forces. When atoms are squeezed closely together, interactions involving the spins of their outermost electrons also result in attractive or repulsive interactions. For atoms with unfilled outer shells, it's energetically favorable for electrons to pair with opposite spins; this causes an attractive interaction. When the outer atomic shells are filled, the exclusion principle forces electrons from separate atoms into different states as the atoms are brought together. One or more electrons go into higher energy states, raising the overall energy and resulting in a repulsive interaction. Finally, if atoms get very close, the electrical repulsion of the nuclei becomes important. Ultimately, the balance of attractive and repulsive interactions determines the equilibrium configuration of a molecule. In energy terms, we can think of a stable molecule as a minimum-energy configuration of the electrons and nuclei making up two or more atoms (Fig. 37.1). Although such force and energy considerations ultimately govern all molecules, we distinguish several molecular binding mechanisms based on which interactions are most important.

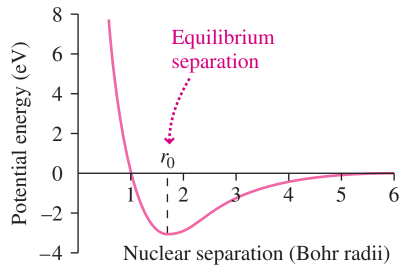


FIGURE 37.1 Potential energy of a pair of hydrogen atoms as a function of the distance between their nuclei.

Ionic Bonding

As we saw in Chapter 36, elements near the left side of the periodic table have few electrons in their outermost shells and correspondingly low ionization energies. In contrast, elements near the right side of the table have nearly filled shells and consequently strong affinities for electrons. When atoms from these different regions of the periodic table come together, it takes relatively little energy to transfer electrons between them. Sodium, for example, has an ionization energy of 5.1 eV, meaning it takes this much energy to make an Na^+ ion. Chlorine, at the opposite end of the periodic table, has such a strong electron affinity that the energy of a Cl^- ion is actually 3.8 eV below that of a neutral Cl atom. Thus an expenditure of only 1.3 eV $(5.1\text{ eV} - 3.8\text{ eV})$ is required to transfer the outermost electron from a sodium to a chlorine atom. The resulting ions are strongly attracted and come together until they reach equilibrium at an internuclear separation of about 0.24 nm. The total energy of the pair is then 4.2 eV below that of neutral chlorine and sodium atoms at large separation (Fig. 37.2). Since it would therefore take 4.2 eV to separate the atoms, this quantity is called the **dissociation energy**.

Because the minimum-energy sodium–chlorine structure consists of ions bound by the electrostatic force, the binding mechanism is termed **ionic bonding**. Ionic bonding generally occurs in crystalline solids. Because the building blocks of an ionically bound substance are electrically charged, each can bind to several of the opposite charge, resulting in a regular crystal pattern (Fig. 37.3). Because the electrostatic force is strong, ionic solids are tightly bound and therefore have high melting points (801°C for NaCl). And because all electrons are bound to individual nuclei, there are no free electrons and therefore ionic solids are electrical insulators.

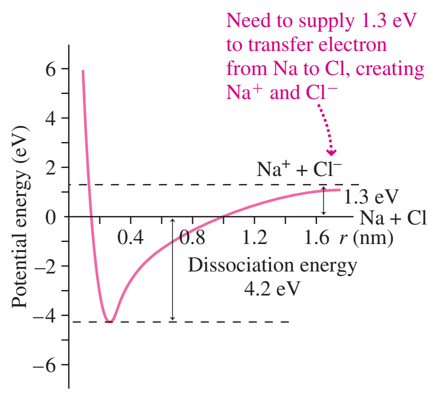


FIGURE 37.2 Potential-energy curve for Na^+ and Cl^- ions, with zero energy corresponding to infinite separation of neutral Na and Cl atoms.

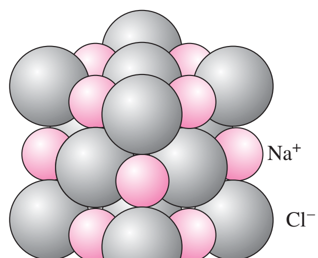


FIGURE 37.3 A sodium chloride crystal is a regular array of sodium and chlorine atoms, bound by the electrostatic force.

Covalent Bonding

In an ionic bond, each electron is associated with only one ion. In a **covalent bond**, on the other hand, electrons are shared among atoms. Covalent bonds occur between atoms whose outermost shells aren't full, and whose outer electrons can therefore pair with opposite spins. The simplest example of a covalent bond is the hydrogen molecule, H_2. Since each hydrogen atom has a single $1s$ electron, each could accommodate in its $1s$ shell a second electron with opposite spin. When two hydrogen atoms join, quantum mechanics predicts a molecular ground state in which both electrons share a single orbital, with the highest probability of finding the electrons between the nuclei (Fig. 37.4). Dissociation energies for covalent bonds are, like those of ionic bonds, on the order of a few electronvolts.

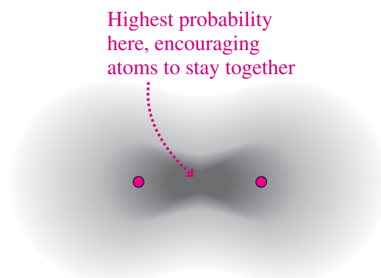


FIGURE 37.4 Probability density for finding electrons in the ground state of molecular hydrogen (H_2).

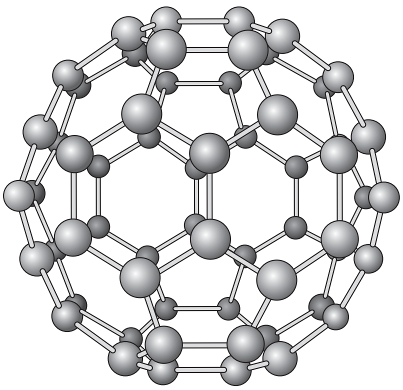


FIGURE 37.5 Buckminsterfullerene, C_{60}, is a symmetric arrangement of 60 carbon atoms. Discovered in the 1980s, C_{60} and related fullerenes hold promise in a wide range of technological applications.

With their outermost molecular orbitals full, covalently bonded molecules often have no room for another electron in their structures. For example, adding a third hydrogen to H_2 is impossible because the ground-state orbital already contains two electrons with opposite spins, so the exclusion principle requires that a third electron go into a higher energy state. The energy of that state is higher than that of an H_2 molecule and a distant H atom; for this reason H_3 isn't a stable molecule. Because their outermost molecular orbitals are full, covalent molecules interact only weakly, and as a result many common covalent materials—for example, H_2, CO, N_2, and H_2O—are either gases or liquids at ordinary temperatures. In other cases covalent bonds can form crystalline structures. A simple example is diamond, a pure-carbon solid formed when each carbon atom bonds covalently to its four nearest neighbors. A more dramatic covalent molecule is buckminsterfullerene, C_{60}, a soccer-ball configuration of 60 carbon atoms (Fig. 37.5).

Hydrogen Bonding

If water consists of covalently bonded molecules, why is it ever a solid? The answer lies in **hydrogen bonds** that form when the tiny, positively charged proton of a hydrogen nucleus nestles close to the negative parts of other molecules. In ice, hydrogen bonds link a proton in one H_2O molecule to the oxygen in another. Because covalent bonding within the water molecule leaves the oxygen only slightly negative and the hydrogen only slightly positive, hydrogen bonds are much weaker than ionic or covalent bonds; a typical hydrogen-bond energy is 0.1 eV. Hydrogen bonds are important in determining the overall configuration of complicated molecules. In DNA, for example, covalent bonds join atoms to form long chains; hydrogen bonds then link two chains into the double-helix structure.

Van der Waals Bonding

In Section 20.5 we mentioned the **van der Waals force** that arises from electrostatic interactions between induced dipole moments of otherwise nonpolar molecules. In gases, the van der Waals force causes deviations from the ideal-gas law that are most pronounced at high densities. As temperature drops, this weakly attractive force becomes effective in binding molecules into liquids and solids. Liquid and solid oxygen (O_2) and nitrogen (N_2), for example, are held together by van der Waals bonds.

Metallic Bonding

In a metal, the outermost atomic electrons aren't bound to individual nuclei, but move throughout the material. The metal forms a crystal lattice of positive ions, bound by this "electron gas." The free electrons give a metal its high electrical and thermal conductivities.

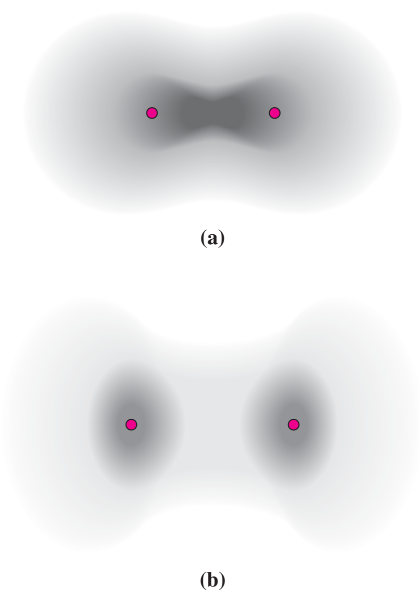


FIGURE 37.6 Electron probability densities for (a) the ground state and (b) the first excited state of hydrogen. Nuclei are farther apart in the excited state.

37.2 Molecular Energy Levels

In a molecule, electric forces bind electrons and nuclei into a single structure. Like any quantum-mechanical bound system, the energy levels of a molecule are quantized. As in atoms, differences among molecular energy levels are associated with different electronic configurations (Fig. 37.6). But molecules are more complex than atoms, and molecular energy can take additional forms.

In Chapter 18, we found that a complete description of the specific heats of gases required that we consider the rotational and vibrational motions of individual molecules. We hinted at quantum mechanics, pointing out that each of these modes of molecular motion could absorb only certain discrete amounts of energy (see Fig. 18.17). Here, in a quantum-mechanical treatment of molecular energetics, we consider rotational and vibrational energy states as well as electronic configuration.

Rotational Energy Levels

If a molecule is rotating, it has angular momentum L whose magnitude, from Equation 11.4, is $L = I\omega$, where I is the rotational inertia and ω the angular speed. The quantization conditions that we found in Chapter 36 for atomic angular momenta also hold for the angular momentum of molecular rotation, so we have

$$L = \sqrt{l(l+1)}\hbar \qquad (37.1)$$

where the quantum number l takes on integer values 0, 1, 2, 3, ... But then the rotational energy, which from Equation 10.17 is $E_{rot} = \frac{1}{2}I\omega^2$, must also be quantized. Solving the equation $L = I\omega$ for ω allows us to write the energy as

$$E_{rot} = \tfrac{1}{2}I\left(\frac{L}{I}\right)^2 = \frac{L^2}{2I}$$

Applying the quantization condition 37.1 for L, we then have the quantized rotational energy levels:

$$E_{rot} = \frac{\hbar^2}{2I}l(l+1) \qquad \text{for } l = 0, 1, 2, 3, \ldots \qquad (37.2)$$

EXAMPLE 37.1 Molecular Rotation: Computing Molecular Size

A gas of HCl molecules shows spectral lines that result from transitions between pairs of adjacent rotational energy levels. The energy of the transition increases by 2.63 meV from one spectral line to the next. (a) Use this experimental result to determine the rotational inertia of the HCl molecule. (b) Approximating the more massive chlorine as being essentially fixed, find an expression for the rotational inertia in terms of the hydrogen mass and the separation of the two atomic nuclei. (c) Use the results of parts (a) and (b) to determine the internuclear separation in HCl.

INTERPRET We're asked to infer molecular properties from spectroscopic observations. We're given not the wavelength or energy of a given spectral line, but the energy difference associated with adjacent lines.

DEVELOP Spectral lines result from photons emitted as a molecule drops from one of the energy levels of Equation 37.2 to the next lower level. So the energy of such a photon is

$$\Delta E_{l\to l-1} = \frac{\hbar^2}{2I}[l(l+1) - (l-1)l] = \frac{\hbar^2 l}{I}$$

An adjacent spectral line would result from the transition $l - 1 \to l - 2$, giving $\Delta E_{l-1\to l-2} = \hbar^2(l-1)/I$. So our plan for (a) is to take the difference between these two transition energies, equate it to the observed energy difference of 2.63 meV between adjacent spectral lines, and solve for I. For (b) we can treat the molecule as having a fixed center (Cl) with a particle (H) in circular motion. Its rotational inertia is then given by Equation 10.12: $I = mR^2$. Equating the two expressions for I will let us solve for the internuclear separation R in (c).

EVALUATE (a) The energy difference $\Delta(\Delta E)$ between adjacent transition energies is

$$\Delta(\Delta E) = \frac{\hbar^2 l}{I} - \frac{\hbar^2(l-1)}{I} = \frac{\hbar^2}{I}$$

Setting this to the observed 2.63-meV spacing and solving for I give $I = 2.65\times10^{-47}\,\text{kg}\cdot\text{m}^2$, where we had to convert meV to J to get the result in SI. (b) We've already shown that $I = mR^2$. (c) Equating our algebraic and numerical expressions for I and solving give the internuclear distance:

$$R = \sqrt{\frac{I}{m}} = \sqrt{\frac{2.65\times10^{-47}\,\text{kg}\cdot\text{m}^2}{1.67\times10^{-27}\,\text{kg}}} = 0.126\text{ nm}$$

where we approximated the hydrogen mass, m, with the proton mass listed inside the front cover.

ASSESS Our answer makes sense, since it's slightly larger than an isolated hydrogen atom. However, it's approximate because we took the chlorine as perfectly fixed and we also ignored the quantum-mechanical ground-state energy of molecular vibration, which stretches the molecule. Photon energies for low l values are on the order of that 2.63-meV spacing, corresponding to a wavelength of about 0.5 mm. This is in the microwave region of the spectrum, with much lower energy and longer wavelength than we've seen for transitions between atomic energy levels. ■

Vibrational Energy Levels

The equilibrium configuration of a molecule corresponds to the minimum of the molecular potential-energy curve. In the vicinity of that minimum the curve is well approximated by a parabola (Fig. 37.7). In Chapter 13 we saw that parabolic potential-energy curves result in simple harmonic motion, and in Chapter 35 we used a parabolic potential-energy curve in the Schrödinger equation for the harmonic oscillator. There, we found that quantized vibrational energy levels are given by

$$E_{vib} = \left(n + \tfrac{1}{2}\right)\hbar\omega \qquad (37.3)$$

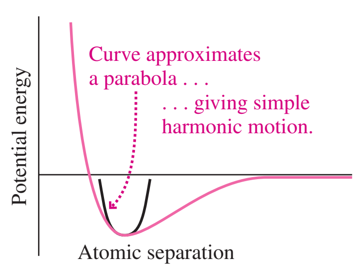


FIGURE 37.7 Near its minimum, the molecular potential-energy curve approximates a parabola.

where the quantum number n takes on integer values 0, 1, 2, 3, . . . , and where ω is the natural frequency for classical harmonic oscillations of the molecule. The selection rule for harmonic oscillators limits allowed transitions to those with $\Delta n = \pm 1$, so $\hbar\omega$ is the energy of photons emitted or absorbed in allowed transitions among vibrational energy levels. (Actually, the small-amplitude approximation is often justified for only the lower quantum states, so Equation 37.3 and the selection rule $\Delta n = \pm 1$ may apply to only these states.) For typical diatomic molecules, ω is on the order of $10^{14}\,\text{s}^{-1}$, in the infrared region of the spectrum. Consequently, study of molecular vibrations involves infrared spectroscopy.

As we found in Chapter 35, the minimum energy of a quantum harmonic oscillator is the ground-state energy $E_0 = \frac{1}{2}\hbar\omega$. Thus a molecule can never have zero vibrational energy, although Equation 37.2 shows that it *can* have zero rotational energy.

EXAMPLE 37.2 Molecular Energies: Rotational and Vibrational

An HCl molecule is in its vibrational ground state. Its classical vibration frequency is $f = 8.66\times10^{13}$ Hz. If its rotational and vibrational energies are nearly equal, what are its rotational quantum number and angular momentum?

INTERPRET We're being asked to compare energies associated with two different processes: vibration and rotation. We're told the vibrational energy state (the ground state, $n = 0$), and we need to find the rotational state with comparable energy.

DEVELOP Equation 37.3, $E_{\text{vib}} = \left(n + \frac{1}{2}\right)\hbar\omega$, with $n = 0$ gives the ground-state vibrational energy $E_{\text{vib}} = \frac{1}{2}\hbar\omega = \frac{1}{2}hf$, where we used $\hbar = h/2\pi$ and $\omega = 2\pi f$. Our plan is to equate this to the rotational energy of Equation 37.2, $E_{\text{rot}} = (\hbar^2/2I)l(l + 1)$, and solve for the quantity $l(l + 1)$. Then we'll use Equation 37.1 to get the angular momentum.

EVALUATE Equating the vibrational and rotational energies gives

$$\frac{\hbar^2}{2I}l(l + 1) = \tfrac{1}{2}\hbar\omega = \hbar\pi f$$

where we used $\omega = 2\pi f$. So $l(l + 1) = 2\pi f I/\hbar$, and Equation 37.1, $L = \sqrt{l(l + 1)}\hbar$, gives

$$L = \sqrt{\frac{2\pi f I}{\hbar}}\hbar = \sqrt{2\pi f I\hbar} = \sqrt{fIh} = 1.23\times10^{-33}\,\text{J}\cdot\text{s}$$

where we used the rotational inertia $I = 2.65\times10^{-47}\,\text{kg}\cdot\text{m}^2$ that we found in Example 37.1.

ASSESS Our angular momentum is about 12 times $\hbar$; approximating $l(l + 1)$ by l^2 in Equation 37.1 then shows that l is about 12. So we need a fairly high rotational quantum state to give the same energy as the vibrational ground state. That's consistent with transitions between adjacent rotational states involving microwave photons, while vibrational transitions involve infrared photons. ■

Molecular Spectra

A molecule with vibrational quantum number n and rotational quantum number l can undergo transitions obeying the selection rules $\Delta n = \pm 1$ and $\Delta l = \pm 1$. If molecules couldn't rotate, the molecular spectrum would consist of a single line at the classical vibration frequency, corresponding to transitions among adjacent vibrational states. But each vibrational level corresponds to an infinite number of rotational states. The resulting energy-level diagram is shown in Fig. 37.8. At typical temperatures, only the ground and first vibrational levels are significantly populated, but with energy distributed among many rotational levels. As a result, molecular spectra show a rich structure, with many lines corresponding to the different transitions of Fig. 37.8. Figure 37.9 is a spectrum of HCl, taken with a high-resolution infrared spectrometer that resolves the individual spectral lines. At lower resolution, the pattern shows up as a broad band, and we often speak of infrared absorption bands in describing the effect of molecules on infrared radiation. For example, absorption bands of atmospheric carbon dioxide limit the escape of infrared radiation from Earth, causing the global warming that we described in Chapter 16. Molecular energy levels are therefore at the heart of today's most global environmental concern.

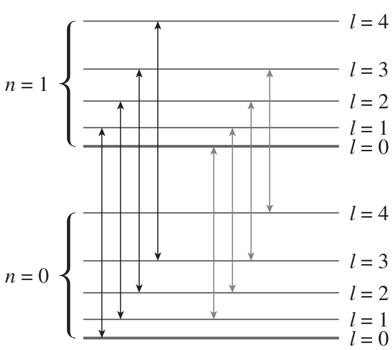


FIGURE 37.8 Energy-level diagram showing the ground state and first vibrational excited state of a diatomic molecule; also shown are four of the infinitely many rotational states for each n.

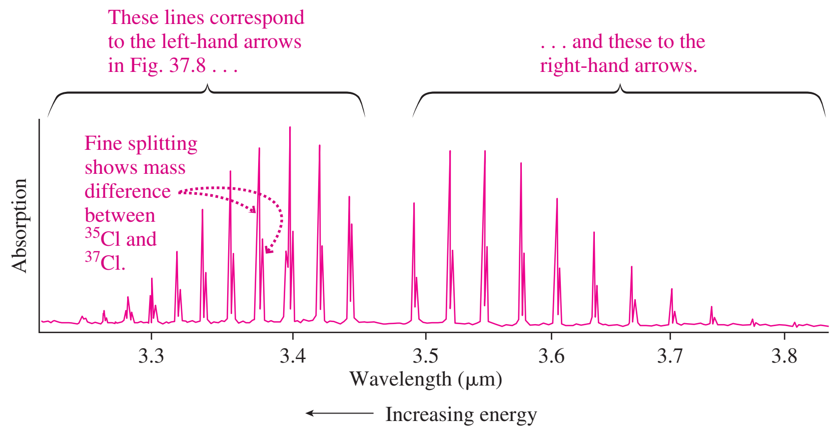


FIGURE 37.9 Absorption spectrum of HCl, showing lines that result from transitions between the $n = 1$ and $n = 0$ vibrational states.

GOT IT? 37.1 You meet a scientist who uses microwave technology to study molecular structure. What form of molecular energy is she most concerned with?

37.3 Solids

Bonding mechanisms can join relatively few atoms to form a molecule, or many to form a solid. In the lowest energy state, the atoms of a solid are arranged in a regular, repeating pattern; the solid is then **crystalline**. Sometimes solids form without their atoms having the opportunity to achieve a crystal structure; such solids are termed **amorphous**. Glass is a common amorphous solid. Amorphous materials are difficult to analyze due to their inherent randomness, so we concentrate here on crystalline solids.

Crystal Structure

The hallmark of a crystalline solid is the regular arrangement of atoms. Looking closely shows that a basic pattern repeats throughout the crystal (Fig. 37.10). This basic arrangement is the **unit cell**. Different crystalline materials have different unit cells (Fig. 37.10*a*, *c*). Sometimes the same underlying matter may assume different structures, depending on how the solid was formed; this is the case with diamond and graphite, both crystalline forms of carbon.

As with individual molecules, properties like atomic separation in a crystalline solid are determined by the interplay of attractive and repulsive interactions. The situation is complicated, however, because an individual atom experiences forces from many other atoms in the crystal. With ionic bonding, those forces are electrical attraction and repulsion as

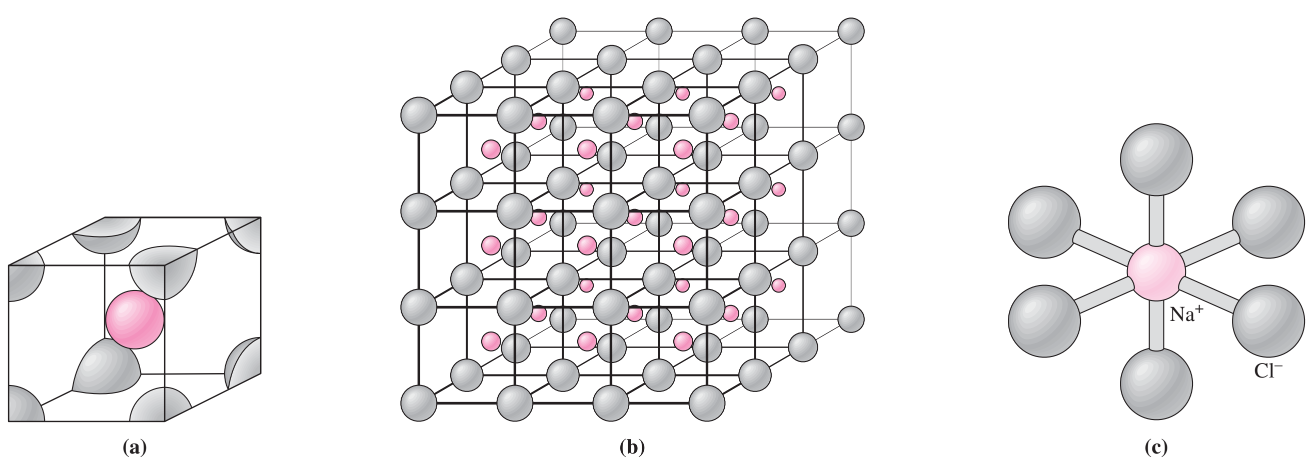


FIGURE 37.10 (a) The unit cell of cesium chloride has eight chlorine ions surrounding each cesium ion. (b) A cesium chloride crystal is a periodic array of unit cells. (c) Sodium chloride is different; here each ion is surrounded by only six nearest neighbors of the opposite type.

described by Coulomb's law; that makes ionic crystals most amenable to simple mathematical treatment.

For ionic crystals, we can take individual ions to be point charges. Consider the NaCl structure of Fig. 37.10*c*. Each sodium ion is surrounded by six nearest chlorine ions, each some distance r away. The potential energy of a singly ionized positive sodium ion in the potential of each negative chlorine ion is $-ke^2/r$. So the contribution to the potential energy of the six nearest chlorines is $-6ke^2/r$, with the minus indicating an attractive interaction. But then there are 12 sodium ions a distance $\sqrt{2}r$ from the sodium; they give rise to a repulsive force and consequently a positive potential energy $+12ke^2/\sqrt{2}r$. At a distance of $\sqrt{3}r$ there are eight more chlorines, giving potential energy $-8ke^2/\sqrt{3}r$. The result is that the electrostatic potential energy of the sodium ion can be written as $U_1 = -\alpha(ke^2/r)$, where $\alpha = 6 - 12/\sqrt{2} + 8/\sqrt{3} - \cdots$; α is called the **Madelung constant**. Many terms in the series are required to compute α accurately, showing that the effect of distant ions is significant in determining the energy of an ion in the crystal. For the NaCl structure, α is approximately 1.748.

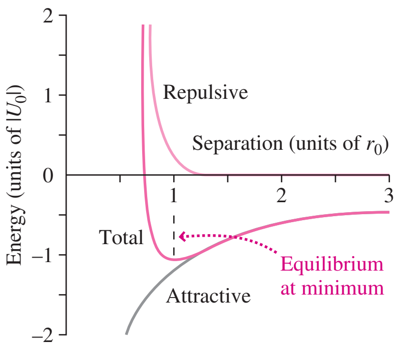


FIGURE 37.11 Potential-energy function for an ionic crystal, showing separate contributions of the attractive and repulsive terms.

As ions are brought closer together, they experience the repulsive effect of the exclusion principle, as we discussed in Section 37.1. This repulsion is described approximately by a potential energy of the form $U_2 = A/r^n$, where A and n are constants. So the total potential energy of an ion in the crystalline solid is

$$U = U_1 + U_2 = -\alpha\frac{ke^2}{r} + \frac{A}{r^n}$$

At equilibrium the potential energy is a minimum (Fig. 37.11), corresponding to zero net force on the ion. Differentiating the potential energy with respect to r and setting dU/dr to zero to find the minimum, we have

$$0 = \frac{\alpha ke^2}{r_0^2} - \frac{nA}{r_0^{n+1}}$$

where r_0 designates the equilibrium separation. Solving for A gives $A = \alpha ke^2 r_0^{n-1}/n$, so the potential energy becomes

$$U = -\alpha\frac{ke^2}{r_0}\left[\frac{r_0}{r} - \frac{1}{n}\left(\frac{r_0}{r}\right)^n\right] \qquad (37.4)$$

The value of U at the equilibrium separation r_0 is designated U_0 and is called the **ionic cohesive energy**. The magnitude of U_0 represents the energy needed to remove an ion entirely from the crystal. The cohesive energy is sometimes given in kcal/mol, in which case its magnitude is the energy per mole needed to break an entire crystal into its constituent ions (see Exercise 23).

EXAMPLE 37.3 Potential Energy in a Solid: The NaCl Crystal

The ionic cohesive energy for NaCl is -7.84 eV. The equilibrium separation, which follows from the measured density, is 0.282 nm (see Exercise 22). Use these values and the Madelung constant $\alpha = 1.748$ to find the exponent n in Equation 37.4 for NaCl.

INTERPRET Here we're given all but one of the quantities in the expression for a crystal's potential energy, and we're asked to solve for the one unknown, n.

DEVELOP Equation 37.4 for the potential energy U looks formidable, with n in two places, including an exponent. But we're given the ionic cohesive energy U_0, which is the value of U when $r = r_0$. So the two terms r_0/r become 1, and since $1^n = 1$, the n in the exponent drops out. With $r = r_0$ and $U = U_0$, Equation 37.4 simplifies to

$$U_0 = -\alpha\frac{ke^2}{r_0}\left(1 - \frac{1}{n}\right)$$

Our plan is to solve this equation for n.

EVALUATE Solving gives

$$n = \left(1 + \frac{U_0 r_0}{\alpha ke^2}\right)^{-1} = 8.22$$

with k the constant in Coulomb's law and e the elementary charge; other quantities are given in the problem statement.

ASSESS The large value of this exponent shows that the NaCl crystal is strongly resistant to compression. In Problem 46 you can calculate the associated repulsive force. ■

Band Theory

Quantum-mechanical analysis of a solid containing 10^{23} atoms or so might seem a hopeless task. But the regularity of a crystalline solid makes that problem, while not easy, at least amenable to mathematical treatment. The physical regularity of the solid is reflected mathematically in the properties of the wave function; specifically, the wave function for a crystalline solid in equilibrium is itself periodic. That's because equivalent points in different unit cells have exactly the same physical properties.

We won't solve the Schrödinger equation for a crystal, or even write the solutions. But we can see what some properties of those solutions must be. Consider two identical atoms, initially widely separated, as they're brought closer together. When the atoms are far apart, they're described by identical wave functions and associated energy-level diagrams; a given electron state, for example, has exactly the same energy in each atom. But as the atoms move closer, their wave functions begin to overlap to form a single wave function that characterizes the entire composite system. Because of the exclusion principle, two electrons that were in identical states in the two widely separated atoms can no longer be in the same state. This effect manifests itself as a separation of what were originally identical energy levels (Fig. 37.12*a*). As more and more atoms come together, initially identical energy levels split into ever more finely spaced levels (Fig. 37.12*b*). In a crystalline solid, there are so many atoms that each level splits into an essentially continuous **band** of allowed energies (Fig. 37.12*c*). **Band gaps** separate the bands arising from distinct single-atom states, as shown in Fig. 37.12*c*. An electron can have any energy between the top and bottom of a band, but energies in the band gaps are forbidden. The situation is like a single atom, where electrons are allowed only certain discrete energies, except now the discrete levels have broadened into bands.

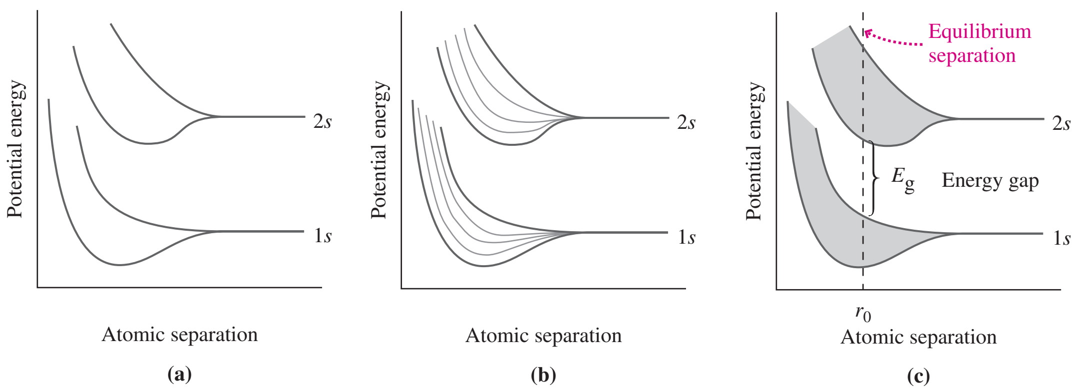


FIGURE 37.12 (a) Energy levels of the 1*s* and 2*s* states as a pair of atoms are brought close together. (b) With five atoms, each level splits into a group of five closely spaced levels. (c) In a crystalline solid, the large number of atoms results in essentially continuous energy bands, separated by gaps.

We're usually interested in the properties of a solid at or near its equilibrium state, designated r_0 in Fig. 37.12*c*. There the solid is characterized by an energy-level diagram in which the energy levels are those of Fig. 37.12*c* at the value $r = r_0$ (Fig. 37.13).

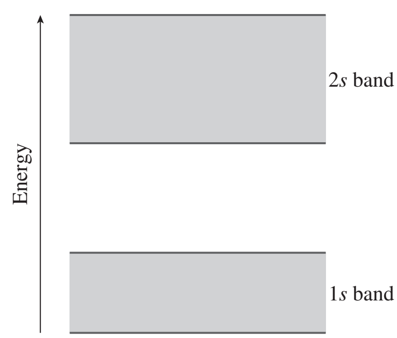


FIGURE 37.13 Energy-level diagram for the equilibrium separation of Fig. 37.12*c*.

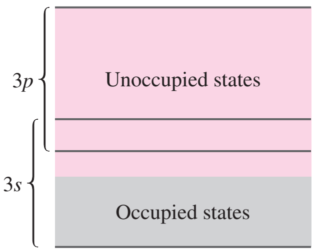


FIGURE 37.14 Band structure of metallic sodium, with gray indicating filled states and color unfilled states. Bands lower than 3*s* aren't shown; they correspond to inner electrons, whose levels aren't split significantly.

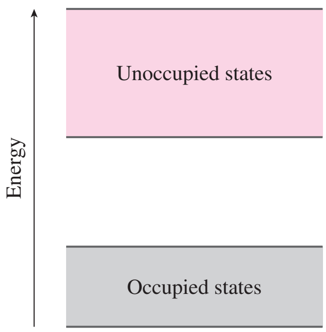


FIGURE 37.15 Band structure for an insulator.

Conductors, Insulators, and Semiconductors

Sometimes the splitting and shifting depicted in Fig. 37.12 result in overlapping bands. Figure 37.14 shows the band structure of sodium, in which the 3*s* and 3*p* bands overlap. Note that the high-energy band containing electrons—here the 3*s*/3*p* band—is not completely full, meaning that energy levels in the upper portion of the band aren't occupied by electrons.

We can determine which of the allowed energy levels of a solid will be occupied in the same way we established the electronic structure of atoms: by placing a given total number of electrons in the lowest possible levels consistent with the exclusion principle. In some materials, like sodium, that filling process results in the highest-energy occupied band being only partially full. But in others, like the material shown in Fig. 37.15, the highest-energy occupied band is completely full.

Figures 37.14 and 37.15 show the essential difference between conductors and insulators. A conductor is a material in which charges are free to move in response to an electric field. Classically, there's no problem with this: Apply an electric field, and if an electron is "free," it will accelerate and gain energy. But quantum mechanically, an electron can gain energy only by jumping into a higher allowed energy level. So there needs to be a higher unoccupied level available.

In sodium, the 3*s* atomic level contains a single electron, although it has room for two. Put N sodium atoms together to form a crystal, and the 3*s* band contains only N of the total $2N$ electrons it could hold. So the 3*s* band is only half full, as Fig. 37.14 shows, and therefore electrons near the top of the filled portion have available unoccupied states with only a little more energy. That makes it easy for an electric field to promote electrons to unoccupied levels. For that reason sodium is an electrical conductor.

In the material of Fig. 37.15, in contrast, one band is completely full and the next higher one empty. An electron in the filled band can't gain energy unless it's enough to jump the band gap. Electric fields of reasonable magnitude can't provide this energy, so the electrons are stuck in the filled bands. That makes the material an insulator.

Metallic Conductors

We found in Chapter 24 that classical physics can't account for the details of metallic conduction, in particular the temperature dependence of conductivity. Quantum mechanically, the conduction electrons in a metal are like electrons in the three-dimensional box of Section 35.4. They're free to move about inside the metal, but not to leave it. The number of states available to the electrons, per unit energy interval, turns out to increase with energy. You can see the beginnings of this trend in Fig. 35.18, which shows the first few states of the three-dimensional box. We won't do this count; the result, however, is given by

$$g(E) = \left(\frac{2^{7/2}\pi m^{3/2}}{h^3}\right)\sqrt{E} \tag{37.5}$$

where m is the electron mass and $g(E)$ is the **density of states**—the number of states per unit volume per unit energy interval centered on the energy E.

At absolute zero, electrons fill the lowest available states according to the exclusion principle. The energy of the highest filled level at absolute zero is the **Fermi energy**, E_F. At temperature $T = 0$, all states below the Fermi energy are full, and all those above are empty, as shown in Fig. 37.16*a*.

For $T > 0$, thermal energy promotes some electrons to levels above the Fermi energy, leaving some levels just below E_F vacant (Fig. 37.16*b*). Now, the Fermi energy in most metals is about 1–10 eV, much higher than the thermal energy at typical temperatures (0.025 eV at room temperature). So the electron distribution changes only slightly—and that means electrons near the Fermi energy carry essentially all the electric current, regardless of temperature. The mean electron speed is therefore quite different from the classical thermal speed (see Problem 50), and that makes the temperature dependence of electrical conductivity in metals very different from the classical prediction.

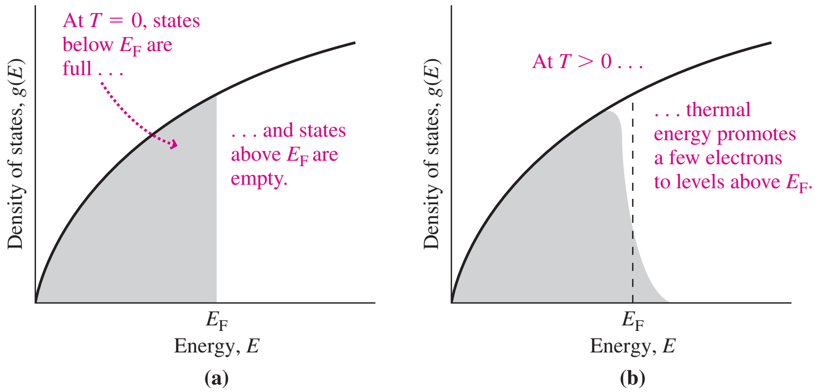


FIGURE 37.16 Density of states given by Equation 37.5, with shaded region indicating occupied energy levels. (a) $T = 0$; (b) $T > 0$.

GOT IT? 37.2 Both parts of Fig. 37.16 describe the same piece of metal. How do the shaded areas compare? Why? What's your interpretation of this shaded area?

Semiconductors

In Chapter 24 we gave a classical description of semiconductors—the materials at the heart of our modern electronic world. Here we see how band theory gives a quantum-mechanical explanation of semiconductors.

Our band diagram for an insulator (Fig. 37.15) is strictly correct only at absolute zero. Here the highest occupied band—the **valence band**—is full, and above it the **conduction band** is empty. At temperatures above absolute zero, though, random thermal energy may give an occasional electron enough energy to jump the gap into the conduction band, where it has plenty of nearby states available and can thus respond freely to an electric field. In good insulators, the band gap is many electronvolts and this effect is negligible. But in some materials, notably silicon and germanium, the band gap is on the order of 1 eV (see Table 37.1). At room temperature, thermal excitation promotes enough electrons into the conduction band that these materials conduct electricity, although their conductivity is much lower than in metallic conductors. Such a material is a **semiconductor**. Figure 37.17 compares the band structures for conductors, insulators, and semiconductors.

Table 37.1 Band-Gap Energies for Selected Semiconductors (at 300 K)

Semiconductor	Band-Gap Energy (eV)
Si	1.14
Ge	0.67
InAs	0.35
InP	1.35
GaP	2.26
GaAs	1.43
CdS	2.42
CdSe	1.74
ZnO	3.2
ZnS	3.6

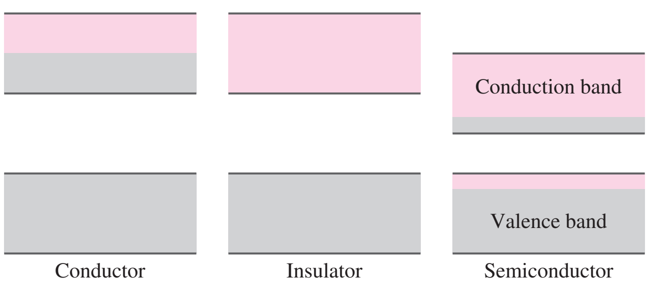


FIGURE 37.17 Band structures for a conductor, an insulator, and a semiconductor. Gray indicates occupied states.

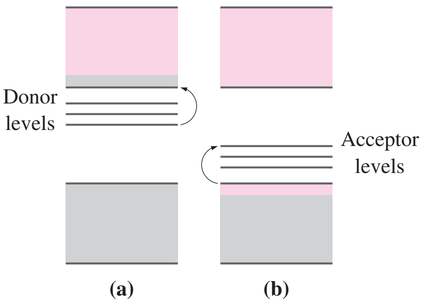


FIGURE 37.18 Band structures in doped semiconductors: (a) *N*-type; (b) *P*-type.

In Chapter 24 we showed how doping—adding small quantities of impurities—could radically alter the electrical properties of semiconductors. In terms of band theory, a dopant such as phosphorus, with five valence electrons, adds **donor levels** just below the conduction band (Fig. 37.18*a*). Thermal energy readily promotes electrons from these levels into the conduction band, greatly increasing the conductivity. This makes an ***N*-type semiconductor** because its predominant charge carriers are electrons. A dopant like boron, in contrast, creates **acceptor levels** just above the valence band (Fig. 37.18*b*). Electrons promoted to these levels leave behind **holes** that act as positive charge carriers. The result is a ***P*-type semiconductor**.

In Chapter 24 we gave a classical explanation of how a junction of *P*- and *N*-type semiconductors conducts electric current in only one direction. From there we went on to describe the operation of the transistor—the semiconductor device at the heart of all modern electronics. We can also understand the *PN* junction in terms of band structure. In Fig. 24.10 we showed how electrons and holes diffuse across a *PN* junction, depleting the junction of charge carriers and making it a poor conductor. Diffusion of electrons also gives the *P*-type side of the junction a net negative charge, and diffusion of the holes makes the *N*-type side positive. This charge separation creates an electric field pointing from *N* to *P*, as shown in Fig. 37.19*a*. The field opposes further diffusion of charge and thus establishes an equilibrium in which there's no net charge flow across the junction. Because they've moved with the electric field, the electrons that have diffused into the *P*-type region have higher potential energy than those that remained behind in the *N*-type region. (Remember that electrons are negative, so their potential energy increases when they move in the *same* direction as an electric field.) As a result, the band-structure diagram for electrons in the *PN* junction takes the form shown in Fig. 37.19*b*.

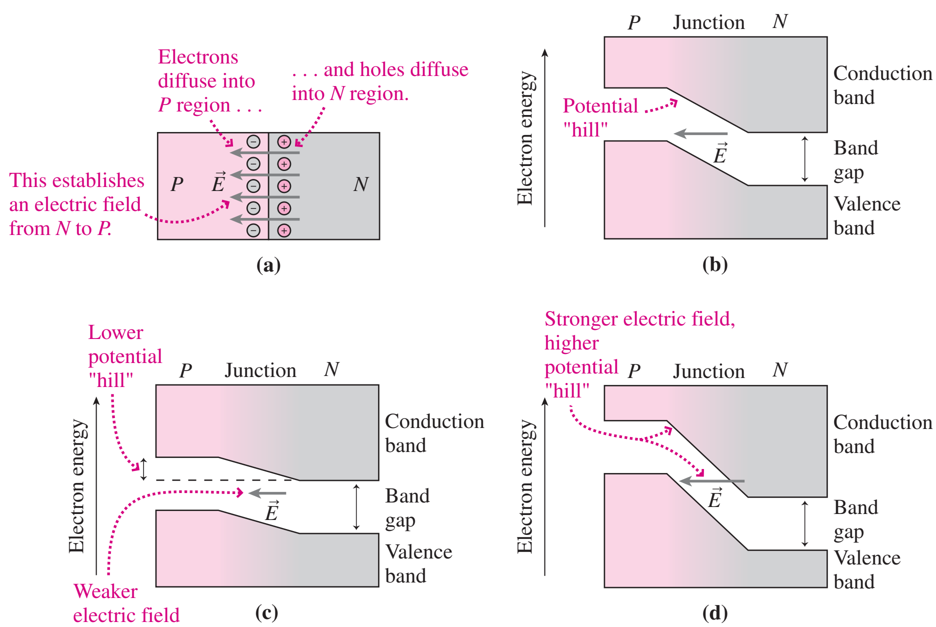


FIGURE 37.19 (a) Physical picture of an unbiased *PN* junction and (b) the corresponding band structure. (c) Band structure of a forward-biased junction and (d) a reverse-biased junction.

Now suppose we connect a battery to our *PN* junction, with the positive terminal to the *P*-type side of the junction. This condition is called **forward bias**. The effect is to make the *P*-type material less negative, the *N*-type less positive, and thus weaken the electric field and lower the potential "hill" that separates the two regions (Fig. 37.19*c*). It becomes easier for electrons to move from *N* to *P* and, as we could show with analogous diagrams for holes, it's easier for holes to move from *P* to *N*. So a current flows in the *P*-to-*N* direction, and the forward-biased *PN* junction becomes a good conductor. If, on the other hand, we connect the battery's positive terminal to the N-type material, then we strengthen the internal electric field and steepen the potential "hill," making it hard for charges to cross the junction (Fig. 37.19d). Now the PN junction is **reverse biased**, and it's a poor conductor.

As electrons and holes pour across a forward-biased junction, many recombine; that is, they drop from the conduction band into the valence band, releasing energy in the process. In light-emitting diodes (LEDs) and diode lasers, this energy appears as photons whose energy is close to that of the band gap. Because $E = hf$, the band gap determines the frequency and, equivalently, the wavelength and color of the emitted light. Development of semiconductor lasers with ever-larger band gaps enabled the evolution from CD to DVD to Blu-ray discs that we outlined in Chapter 32. Conversely, a material whose band gap corresponds to visible-light photons can absorb light energy, promoting electrons to the conduction band and driving current through an external circuit. Such **photovoltaic cells** have long been used to generate electricity on spacecraft and in remote terrestrial applications. As their cost continues to drop, photovoltaics are increasingly used for electric power generation on individual buildings and in larger-scale solar power plants (Fig. 37.20).

FIGURE 37.20 Japan's Sanyo Solar Ark incorporates some 5000 photovoltaic panels and generates a peak power of 630 kW.

CONCEPTUAL EXAMPLE 37.1 CD to Blu-ray: Engineering the Band Gap

The amount of information stored on CDs, DVDs, and Blu-ray discs is limited in part by diffraction effects associated with the wavelength of the laser light used to "read" the disc (see Chapter 32's Application "Movies on Disc"). The lasers used in optical drives are semiconductor lasers, with wavelengths set by the semiconductors' band gaps. Compare the band gaps of lasers used for reading CDs and Blu-ray discs.

EVALUATE A CD holds 74 minutes of audio, yet a Blu-ray disc of the same physical size holds several hours of high-definition video. So Blu-ray data are stored at a smaller spatial scale and thus require a shorter wavelength to "read" the data. Since $E = hf = hc/\lambda$, that means higher photon energy and therefore a larger band gap.

ASSESS As Making the Connection shows, the band gap for Blu-ray is nearly twice that of a CD laser. In fact, the trade name "Blu-ray" comes from the blue wavelength used. Multiple-layer storage and better compression algorithms also contribute to Blu-ray's much greater capacity.

MAKING THE CONNECTION The lasers that "read" CDs, DVDs, and Blu-ray discs operate at 780 nm, 650 nm, and 405 nm, respectively. Find the corresponding band gaps.

EVALUATE Photon energy quantization $E = hf = hc/\lambda$ gives the photon energy and therefore the required band gap. Working in electronvolts gives 1.59 eV for CD, 1.91 eV for DVD, and 3.07 eV for Blu-ray.

37.4 Superconductivity

In Chapter 24 we introduced **superconductivity**—the complete loss of electrical resistance in some materials at low temperature. First discovered in mercury in 1911, superconductivity was for decades limited to a few elements and alloys below about 20 K. A breakthrough in 1986 brought a new class of metal-oxide superconductors with superconducting **transition temperatures** of about 100 K; today the highest transition temperatures exceed 160 K. The ultimate goal of a room-temperature superconductor, once thought beyond reach, may yet be achieved.

Superconductors find use in an ever-increasing range of applications, including high-strength electromagnets for MRI scanners, particle accelerators, materials separation, and research; compact, efficient motors for vehicle and marine propulsion; high-Q filters for cell-phone base stations; sensitive magnetic-field sensors for brain-wave imaging and physics research; underground power transmission in crowded cities; and so-called synchronous condensers for optimizing the power factor in AC power transmission (see Section 28.5). Other applications include superconducting electronic devices that promise orders-of-magnitude increases in computer speed, and magnetically levitated vehicles for ground transportation at speeds up to 500 km/h (see the Application on p. 675).

Superconductivity and Magnetism

The hallmark of a superconductor is zero electrical resistance. Another distinguishing characteristic is the **Meissner effect**, wherein a superconductor excludes magnetic flux from its interior (Fig. 37.21). Figure 37.21*c* shows why: Currents in the superconductor create their own magnetic field that exactly cancels the field within the material. As we pointed out in Section 27.6, a superconductor's exclusion of magnetic flux makes it perfectly diamagnetic. The magnetic levitation shown in Fig. 27.34 is a manifestation of the Meissner effect, wherein a magnet is supported over a superconductor by mutual repulsion between the magnet and currents in the superconductor.

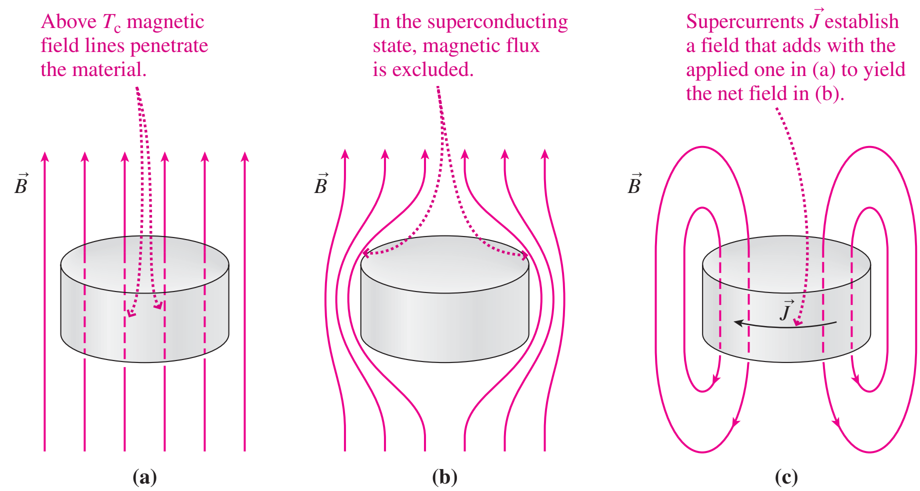


FIGURE 37.21 The Meissner effect.

As the strength of an applied magnetic field increases, so do the currents and resulting magnetic field of the superconductor. But beyond a **critical field**, the external magnetic field alters the superconducting state, and the superconductor no longer excludes magnetic flux. In **type I superconductors**, superconductivity ceases abruptly at the critical field (Fig. 37.22*a*). **Type II superconductors**, in contrast, have upper and lower critical fields, between which superconductivity gradually diminishes (Fig. 37.22*b*). At the lower critical field the material begins to allow flux penetration, and a regular array of nonsuperconducting regions forms, centered on magnetic field lines. These grow with increasing field, until at the upper critical field the superconducting regions vanish altogether.

Because electric currents generate magnetic fields, the critical field can limit the current-carrying capability of superconductors. Fortunately type II superconductors have high enough upper critical fields to permit substantial currents. Type IIs tend to be alloys or complex mixtures, and include all the high-T superconductors. Critical fields of high-T superconductors are as high as 100 T; however, these materials are brittle ceramics and present engineering challenges to the fabrication of wires and other flexible conductors.

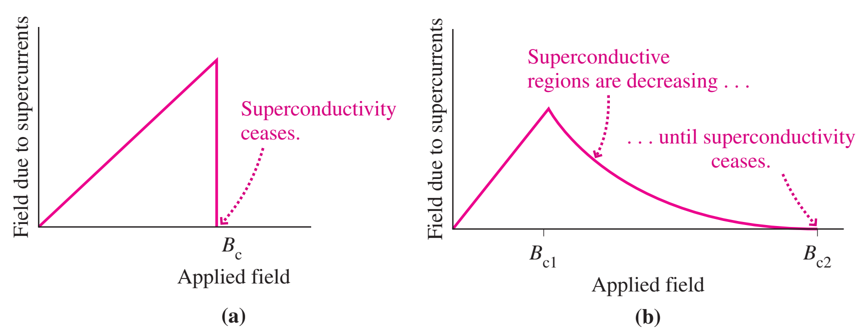


FIGURE 37.22 Responses of (a) type I and (b) type II superconductors to applied magnetic fields. B_c denotes the critical field.

APPLICATION **Maglev!**

Passengers arriving at Shanghai's Pudong International Airport make the 30-km trip to a city metro station in only 7 minutes, "flying" on a magnetically levitated vehicle—a maglev—at speeds exceeding 400 km/h. The Shanghai system uses conventional electromagnets and electronic feedback circuits to keep the vehicle levitated a mere 1 cm above its guideway. But others are developing maglev systems that rely on superconducting magnets for both levitation and propulsion. Coils in the guideway carry alternating current, alternately pushing and pulling the vehicle's onboard magnet. In effect, vehicle and guideway become a linear electric motor, much like a conventional motor that's been "unwound" to produce straight-line motion. In superconducting systems, any deviation from perfect alignment with the guideway results in induced currents and, correspondingly, magnetic forces that act to keep the maglev vehicle centered in its guideway. Today's superconducting maglevs require onboard refrigeration systems, so development of a room-temperature superconductor would make maglev a much more attractive transportation alternative. The photo shows a Japanese superconducting maglev that has achieved speeds of 450 km/h.

Theories of Superconductivity

Superconductivity is a purely quantum-mechanical phenomenon; classical physics is totally inadequate to explain its existence. A successful theory of conventional low-T_c superconductors, called the **BCS theory** after its originators, was formulated in 1957 by John Bardeen, Leon Cooper, and John Robert Schrieffer; the trio shared the 1972 Nobel Prize in physics.

In BCS theory, superconductivity results from a quantum-mechanical pairing of electrons that leads to a lower-energy state in which electron pairs move through the crystal lattice with no energy loss to the ions, resulting in zero electrical resistance. The electron pairing involves one electron slightly deforming the ion lattice, with the second electron attracted by the slight positive charge of the deformed lattice (Fig. 37.23*a, b*). But the paired electrons aren't physically close; typically, a million other electrons, each paired with another distant electron, may lie between the two (Fig. 37.23*c*). The result of this long-range pairing is coherent motion of the conduction electrons that extends throughout the superconductor. Like well-choreographed dancers, the electrons all move together in a way that precludes energy loss to the ion lattice.

High-temperature superconductors aren't fully understood, although they almost certainly involve quantum-mechanical pairing of charge carriers. The mechanism of the pairing is less clear; one promising candidate involves magnetic interactions, although other mechanisms are under investigation. Superconductivity presents a continuing challenge to both theorists and experimentalists.

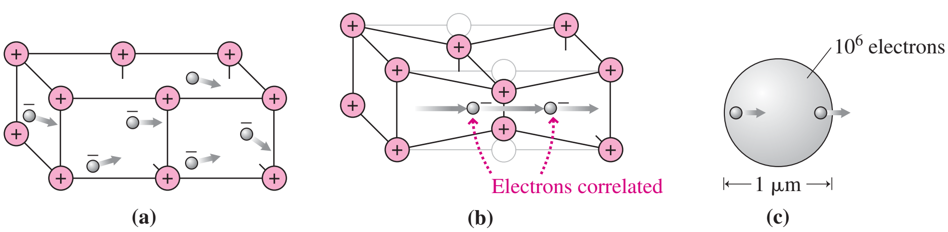


FIGURE 37.23 Electron pairing in BCS theory. (a) A normal conductor, with uncorrelated electrons. (b) In a superconductor, one electron passing through the lattice deforms it slightly. About 10^{-12} s later, a second electron passes through and experiences the potential of the deformed lattice. The two electrons are therefore correlated. (c) Paired electrons are typically 1 μm apart, with a million others in their vicinity. The coherent motion of all the paired electrons results in superconductivity.

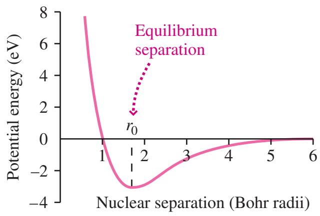
CHAPTER 37 SUMMARY

Big Picture

The big idea here is that quantum mechanics can explain the structure of molecules and solids as well as the atoms treated in Chapter 36. At this level we can't solve the Schrödinger equation for these many-particle systems, but we've argued—using energy and angular momentum quantization and the exclusion principle—that quantum effects are important in molecules and solids.

Key Concepts and Equations

Common types of molecular bonding include **ionic, covalent, hydrogen, van der Waals,** and **metallic** bonding. Whatever the bonding mechanism, a stable molecule is at the minimum of its potential-energy curve, shown below for H_2.



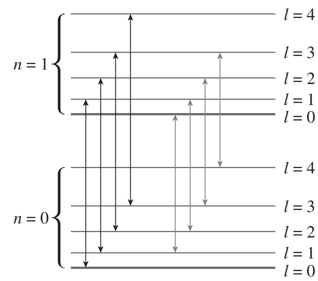
Molecules exhibit both rotational and vibrational energy, giving rise to a rich structure of quantized energy levels and the resulting spectra.

- Quantization of angular momentum leads to quantized rotational energy levels:

$$E_{\text{rot}} = \frac{\hbar^2}{2I}l(l+1),\ l = 0, 1, 2, \ldots$$

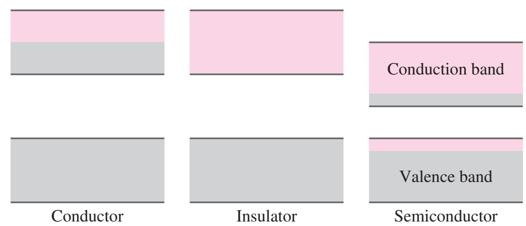
- The vibrational energy levels are those of the harmonic oscillator:

$$E_{\text{vib}} = \left(n + \tfrac{1}{2}\right)\hbar\omega,\ n = 0, 1, 2, \ldots$$



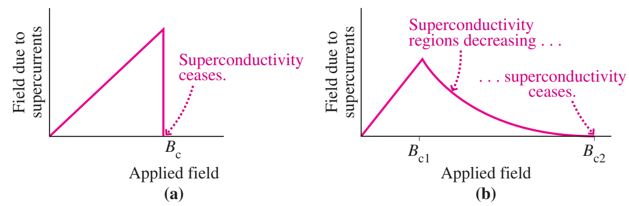
When atoms join to make solids, individual atomic energy levels separate to form **bands**. **Band theory** distinguishes conductors from insulators depending on whether the uppermost occupied band is partially or completely full, respectively. **Semiconductors** are like insulators, but with a much smaller band gap that permits thermal excitation of electrons into the conduction band.

In a metallic conductor, the energy of the highest occupied state at absolute zero is the **Fermi energy**.

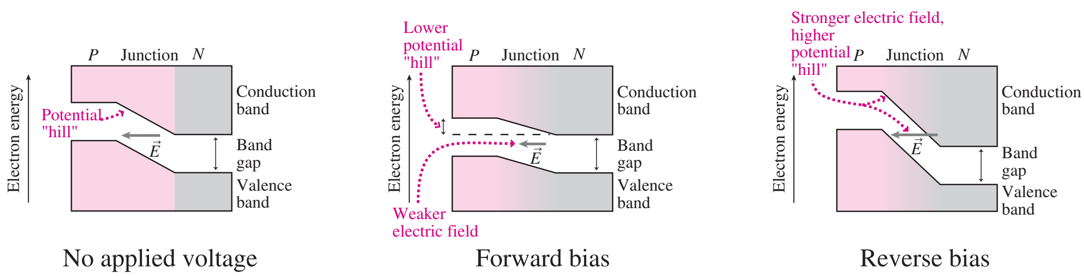


Applications

Superconductivity is a quantum-mechanical phenomenon that occurs at low temperatures and admits no classical explanation. Paired electrons move coherently through a superconductor without energy loss to the ion lattice, resulting in zero electrical resistance. Superconductors exclude magnetic fields—the **Meissner effect**—up to a **critical field** that destroys superconductivity, abruptly in **type I superconductors** and gradually in **type II superconductors**.



The band structure of doped semiconductors helps explain the one-way conduction of a *PN* junction, a phenomenon at the heart of modern electronics.



For homework assigned on MasteringPhysics, go to www.masteringphysics.com | BIO Biology and/or medical-related problems C Computer problems

For Thought and Discussion

1. Why is the exclusion principle crucial to the existence of stable molecules?
2. Why do ionically bonded materials have high melting points?
3. Ionic bonds clearly result from electrostatic attraction between ions. In what way do covalent bonds also involve electrostatic attraction?
4. Does it make sense to distinguish individual NaCl molecules in a salt crystal? What about individual H_2O molecules in an ice crystal? Explain.
5. Is it useful to think of the highest-energy electrons as "belonging" to individual atoms in an ionically bonded molecule? In a covalently bonded molecule?
6. What are the approximate relative magnitudes of the energies associated with electronic excitation of a molecule, with molecular vibration, and with molecular rotation?
7. Radio astronomers have discovered many complex organic molecules in interstellar space. Why were these discoveries made with radio telescopes and not optical telescopes?
8. In Fig. 18.17, why are rotational states excited at lower temperatures than vibrational states?
9. Would you expect solid hydrogen to conduct electricity? Why or why not?
10. The Fermi energy in metals is much higher than the thermal energy at typical temperatures. Why does this make the mean speed of conduction electrons nearly independent of temperature?
11. Why does the size of the band gap determine whether a material is an insulator or a semiconductor?
12. How would you expect the conductivity of an undoped semiconductor to depend on temperature? Why?
13. Name some technological innovations that might result from a room-temperature superconductor.
14. Suppose a room-temperature superconductor were discovered, but it had a very low critical field. In what way would this limit its practical applicability?
15. How do type I and type II superconductors differ?

Exercises and Problems

Exercises

Section 37.2 Molecular Energy Levels

16. Find the energies of the first four rotational states of the HCl molecule described in Example 37.1.
17. The rotational inertia of oxygen (O_2) is 1.95×10^{-46} kg·m^2. Find the wavelength of electromagnetic radiation needed to excite oxygen molecules to their first rotational excited state.
18. Find the wavelength of a photon emitted in the $l = 5$ to $l = 4$ transition of a molecule whose rotational inertia is 1.75×10^{-47} kg·m^2.
19. Photons of wavelength 1.68 cm excite transitions from the rotational ground state to the first rotational excited state in a gas. What's the rotational inertia of the gas molecules?
20. The classical vibration frequency for diatomic hydrogen (H_2) is 1.32×10^{14} Hz. Find the spacing between its vibrational energy levels.
21. The energy between adjacent vibrational levels in diatomic nitrogen is 0.293 eV. What's the classical vibration frequency of N_2?

Section 37.3 Solids

22. Use the 2.16-g/cm^3 density of NaCl to calculate the ionic spacing r_0 in the NaCl crystal. (*Hint:* Consult Appendix D.)
23. Express the 7.84-eV ionic cohesive energy of NaCl in kilocalories per mole of ions.
24. Lithium fluoride, LiF, has the same crystal structure as NaCl and therefore has essentially the same Madelung constant α. Its ionic cohesive energy is -10.5 eV and the value of n in Equation 37.4 is 6.25. Find the equilibrium ionic separation in LiF.
25. Find the wavelength of light emitted by a gallium phosphide (GaP) light-emitting diode. (*Hint:* See Table 37.1.)
26. What's the shortest wavelength of light that could be produced by electrons jumping the band gap in a material from Table 37.1? What is that material?
27. Which material in Table 37.1 would provide the longest wavelength of light in a light-emitting diode? What's that wavelength?
28. A common light-emitting diode is made with gallium arsenide phosphide (GaAsP) and emits red light at 650 nm. What's its band gap?

Problems

29. A molecule drops from the $l = 2$ to the $l = 1$ rotational level, emitting a 2.50-meV photon. If the molecule then drops to the rotational ground state, what energy photon will it emit?
30. A molecule absorbs a photon and jumps to the next higher rotational state. If the photon energy is three times what would be needed for a transition from the rotational ground state to the first rotational excited state, between what two levels is the transition?
31. Find an expression for the energy of a photon required for a transition from the $(l - 1)$th level to the lth level in a molecule with rotational inertia I.
32. A molecule with rotational inertia I undergoes a transition from the lth rotational level to the $(l - 1)$th level. Show that the wavelength of the emitted photon is $\lambda = 4\pi^2 Ic/hl$.
33. The rotational spectrum of diatomic oxygen shows spectral lines spaced 0.356 meV apart in energy. Find O_2's atomic separation. (*Hint:* See Example 37.1, and remember that the oxygen atoms have equal mass.)
34. Use the result given in Problem 59 to find the bond length in carbon monoxide (CO), given that excitation of the first rotational state requires photons of wavelength 2.59 mm.
35. For the HCl molecule of Example 37.2, determine (a) the energy of the vibrational ground state and (b) the energies of photons emitted in transitions among adjacent vibrational states, for the cases $\Delta l = +1$ and $\Delta l = -1$.
36. Diatomic deuterium has classical vibration frequency 9.35×10^{13} Hz and rotational inertia 9.17×10^{-48} kg·m^2. Find (a) the energy and (b) the wavelength of a photon emitted in a transition between the $n = 1, l = 1$ state and the $n = 0, l = 2$ state.
37. Carbon dioxide contributes to global warming because the triatomic CO_2 molecule exhibits many vibrational and rotational excited states, and transitions among them occur in the infrared region where Earth emits most of its radiation. Among the strongest IR-absorbing transitions is one that takes CO_2 from its ground state to the first excited state of a "bending" vibration and sets the molecule rotating in its first rotational excited state. The energy required for this transition is 82.96 meV. What IR wavelength does this transition absorb?

38. An oxygen molecule is in its vibrational and rotational ground states. It absorbs a photon of energy 0.19653 eV and jumps to the $n = 1, l = 1$ state. It then drops to $n = 0, l = 2$, emitting a 0.19546-eV photon. Find (a) the classical vibration frequency and (b) the rotational inertia of the molecule.
39. The internuclear spacing in diatomic hydrogen (H_2) is 74 pm. Find the energy of a photon emitted in a transition from the first rotational excited state to the ground state.
40. BIO Biological macromolecules are complex structures that exhibit many more vibrational modes than the diatomic molecules considered in this chapter. DNA has a low-frequency "breathing" mode whose associated photon wavelength is 330 μm. Find the corresponding (a) frequency and (b) photon energy in eV.
41. What wavelength of infrared radiation is needed to excite a transition between the $n = 0, l = 3$ state and the $n = 1, l = 2$ state in KCl, for which the rotational inertia is 2.43×10^{-45} kg·m^2 and the classical vibration frequency is 8.40 THz?
42. Find the wavelengths emitted in all allowed transitions between the first three rotational states in the $n = 1$ level to any states in the $n = 0$ level in H_2, whose rotational inertia and classical vibration frequency are 4.60×10^{-48} kg·m^2 and 3.69×10^{14} Hz, respectively.
43. Determine the constant n in Equation 37.4 for potassium chloride (KCl), which has the same crystal structure as NaCl and for which $r_0 = 0.315$ nm and $U_0 = -7.21$ eV.
44. A salt crystal contains 10^{21} sodium–chlorine pairs. How much energy would it take to compress the crystal to 90% of its normal size?
45. Lithium chloride, LiCl, has the same structure and therefore the same Madelung constant as NaCl. The equilibrium separation in LiCl is 0.257 nm, and $n = 7$ in Equation 37.4. Find the ionic cohesive energy of the LiCl crystal.
46. You're researching the possibility of storing radioactive waste in underground salt formations. In support of this idea, you'd like to demonstrate that salt is extremely resistant to compression. You differentiate Equation 37.4 to obtain an expression for the force on an ion in an ionic crystal, and then use your result to find the force on an ion in NaCl if the crystal were compressed to half its equilibrium spacing (see Example 37.3 for relevant parameters). You compare this with the electrostatic attraction at this compression. What do you find?
47. Integrating Equation 37.5 over all energies gives the total number of states per unit volume in a metal. Therefore, integrating from $E = 0$ to $E = E_F$—that is, over the occupied states only—gives the number of conduction electrons per unit volume. Carry out this integration to show that the electron number density is given by
$$n = \left(\frac{2^{9/2}\pi m^{3/2}}{3h^2}\right)E_F^{3/2}$$
48. The Fermi energy in aluminum is 11.6 eV. Use the result of Problem 47 to find the density of conduction electrons in aluminum.
49. Use the result of Problem 47 to determine the Fermi energy for calcium, which has 4.6×10^{28} conduction electrons per cubic meter.
50. You're trying to explain to your classmates how classical and quantum descriptions of electrical conduction in metals differ. Using copper's Fermi energy (7.0 eV), you calculate the associated electron speed, then compare your result with the classical thermal speed for an electron at room temperature (300 K). What do you find, and how does this help with your explanation?
51. The *Fermi temperature* is defined by equating the thermal energy kT to the Fermi energy, where k is Boltzmann's constant. Calculate the Fermi temperature for silver, for which $E_F = 5.48$ eV, and compare it with room temperature.
52. Photons with energy lower than a semiconductor's band gap aren't readily absorbed by the material, so a measurement of absorption versus wavelength gives the band gap. An absorption spectrum for silicon shows no absorption for wavelengths longer than 1090 nm. Use this information to calculate the band gap in silicon, and verify its value in Table 37.1.
53. Calculate the median wavelength λ_{median} for sunlight, treating the Sun as a 5800-K blackbody (see Equation 34.2b). Use your result to decide whether zinc selenide, with band gap 3.6 eV, would make a good photovoltaic cell.
54. Pure aluminum, which superconducts below 1.20 K, exhibits a critical field of 9.57 mT. Find the maximum current that can be carried in a 30-gauge (0.255-mm-diameter) aluminum superconducting wire without the field from that current exceeding the critical field. (*Hint:* Where is the field greatest? Consult Example 26.7.)
55. The critical field in a niobium–titanium superconductor is 15 T. What current in a 5000-turn solenoid 75 cm long will produce a field of this strength?
56. The transition from the ground state to the first rotational excited state in diatomic oxygen (O_2) requires 356 μeV. At what temperature would the thermal energy kT be sufficient to set diatomic oxygen into rotation? Would you ever find diatomic oxygen exhibiting the specific heat of a monatomic gas at normal pressure?
57. BIO *Green fluorescent protein* (GFP) is a substance that was first extracted from jellyfish; variants are used to "tag" biological molecules for study. The original "wild" GFP absorbs 395-nm light, undergoing an upward transition to an excited state. Movement of a proton within the protein then excites it to 2.44 eV above the ground state. Photons emitted in the subsequent downward transition to the ground state provide a visual indication of the GFP's location as seen in a microscope. What's the wavelength of these photons?
58. The density of rubidium iodide (RbI) is 3.55 g/cm^3, and its ionic cohesive energy is -145 kcal/mol. Determine (a) the equilibrium separation and (b) the exponent n in Equation 37.4 for RbI.
59. You're troubled that Example 37.1 neglects the mass of the hydrogen, and you wonder how much error this introduces. So you consider a diatomic molecule consisting of different atoms with masses m_1 and m_2, separated by a distance R, and derive an expression for the molecule's rotational inertia about its center of mass. You then calculate a more accurate value for the HCl bond length in Example 37.1. Your results?
60. What fraction of conduction electrons in a metal at absolute zero have energies less than half the Fermi energy?
61. The Madelung constant (Section 37.3) is notoriously difficult to calculate because it's the sum of an alternating series of nearly equal terms. But it can be calculated for a hypothetical one-dimensional crystal consisting of an evenly spaced line of alternating positive and negative ions (Fig. 37.24). Show that the potential energy of an ion in this "crystal" can be written as
$$U = -\alpha\frac{ke^2}{r_0}$$
where the Madelung constant α has the value $2\ln 2$.

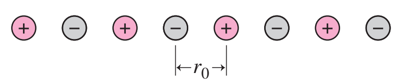


FIGURE 37.24 Problem 61

62. The lower-energy states in a covalently bound diatomic molecule can be found approximately from the so-called Morse potential $U(r) = U_0(e^{2(r-r_0)/a} - e^{-2(r-r_0)/a})$, where r is the atomic separation and U_0, r_0, and a are constants determined from experimental data. Calculate dU/dr and d^2U/dr^2 to show that U has a minimum, and find expressions for (a) U_{min} and (b) the separation r_{min} at the minimum energy.
63. (a) Count the number of electron states $N(E)$ with energy equal to or less than E in Equation 35.8 by finding the volume available to such states in the space with Cartesian coordinate axes n_x, n_y, n_z. (*Hint*: Consider each set of positive integers, at the corner of a unit cube, and that lies inside a radius $\sqrt{n_x^2 + n_y^2 + n_z^2}$, and remember that there are two spin values per state.) (b) Differentiate $N(E)$ with respect to E to obtain Equation 37.5.
64. Use Equation 37.5 to calculate the average energy of a conduction electron at $T = 0$ in terms of the Fermi energy.
65. BIO You're designing a new medical MRI imager, which calls for a long solenoid wound with 75 turns per meter of niobium–titanium superconductor. The upper critical field for your particular Nb-Ti alloy is 12 T. To avoid a disastrous loss of superconductivity (see Example 27.9), you want to limit the actual field to half the upper critical field. What maximum current do you specify for your device?

Passage Problems

Photovoltaic (PV) cells convert sunlight energy directly into electricity, with no moving parts (recall Fig. 37.20). In a PV cell, photons incident on a semiconductor *PN* junction promote electrons to the conduction band, producing electron-hole pairs and driving current through an external circuit (Fig. 37.25). Commercially available PV cells are 15–20% efficient, meaning they convert this fraction of incident sunlight into electrical energy; the theoretical maximum efficiency is around 33% for silicon-based PV cells. An important limitation on PV efficiency is the relation between the solar spectrum and PV cells' semiconductor band-gap energy. For silicon, the band gap is 1.14 eV; photons with less energy can't promote electrons to the conduction zone and are thus unavailable for the PV energy conversion. Conversely, photons with more than the band-gap energy give up their excess energy as heat, also reducing PV efficiency.

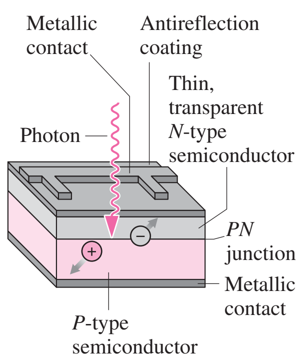


FIGURE 37.25 Operation of a photovoltaic cell, showing a solar photon producing an electron-hole pair at the *PN* junction (Passage Problems 66–69).

66. Problem 53 shows that the median wavelength in the solar spectrum is 710 nm, at the visible-IR boundary. What percentage of the incident solar energy can a silicon PV cell absorb? (*Hint*: See Exercise 36.31.)
   a. about 25%
   b. about 50%
   c. about 75%
67. How does the percentage of the number of incident solar photons that a PV cell absorbs compare with the energy percentage in the preceding problem?
   a. It's less than the energy percentage.
   b. It's the same as the energy percentage.
   c. It's more than the energy percentage.
68. Making PV cells with a semiconductor whose band gap is lower than silicon's will
   a. increase the fraction of solar energy absorbed while decreasing the amount of absorbed energy lost as heat.
   b. increase both the fraction of solar energy absorbed and the amount of absorbed energy lost as heat.
   c. decrease the fraction of solar energy absorbed while increasing the amount of absorbed energy lost as heat.
   d. decrease both the fraction of solar energy absorbed and the amount of absorbed energy lost as heat.
69. One way to improve PV efficiency is to make multi-layer cells with several *PN* junctions using semiconductors with different band gaps. For a multi-layer PV cell to be effective,
   a. the junction with the largest band gap should be closest to the top of the PV cell.
   b. the junction with the largest band gap should be closest to the bottom of the PV cell.
   c. the largest band gap should correspond to infrared wavelengths.
   d. the smallest band gap should correspond to ultraviolet wavelengths.

Answers to Chapter Questions

Answer to Chapter Opening Question

The Schrödinger equation.

Answers to GOT IT? Questions

37.1. Rotational energy.
37.2. The shaded areas are the same; they represent the number of conduction electrons per unit volume.

38 Nuclear Physics

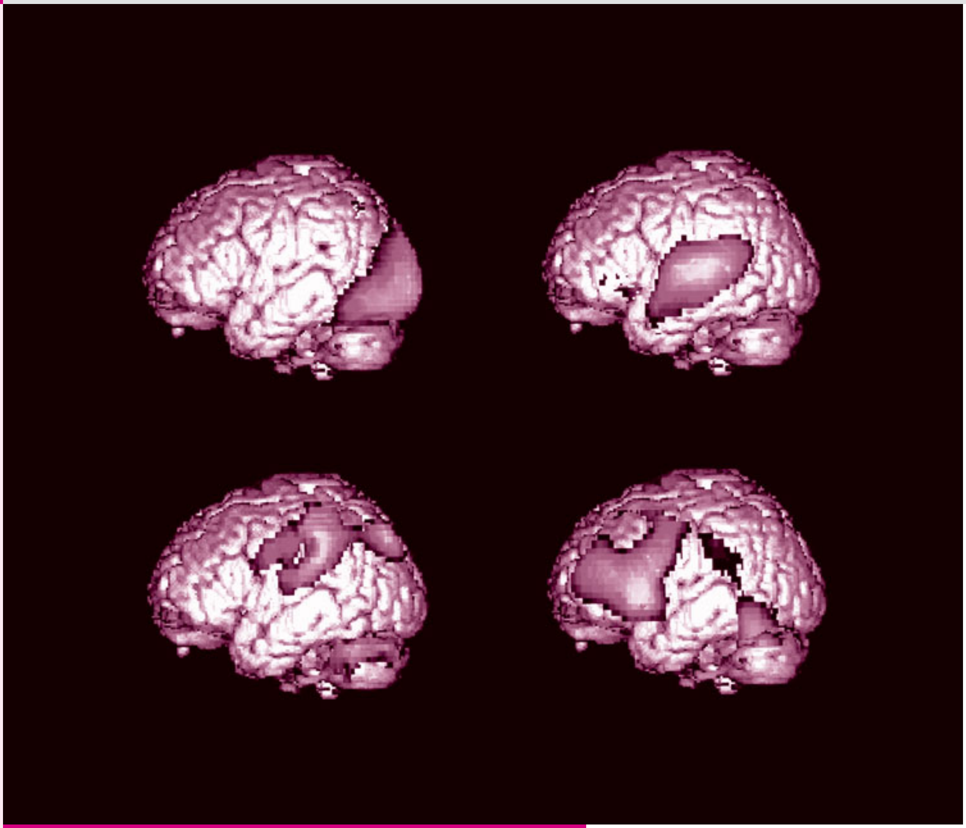
New Concepts, New Skills

By the end of this chapter you should be able to

- Characterize nuclei by atomic number and mass number, and explain the difference between isotopes of the same element (38.1).
- Distinguish stable from unstable nuclei, and explain the chart of the nuclides (38.1).
- Calculate nuclear sizes and spin angular momenta (38.1).
- Describe qualitatively models of nuclear structure (38.1).
- Describe the three common types of radiation, and write equations describing each (38.2).
- Quantify radioactivity, and describe its time dependence using half-life (38.2).
- Describe the curve of binding energy and how it explains energy release in nuclear fission and fusion (38.3).
- Explain nuclear fission and its role as an energy source, including several types of nuclear reactors (38.4).
- Describe nuclear fusion, explaining how it powers the Sun and stars and outlining prospects for terrestrial fusion (38.5).

Connecting Your Knowledge

- This chapter takes us into the nucleus at the heart of the atom. Although much of this material is new, it builds on the idea of atomic structure (34.4).
- It also draws on earlier physics concepts, including energy and the fundamental forces (4.3, Chapters 6 and 7).



Changing regions of mental activity are evident in these PET (positron emission tomography) scans of a human brain. What is a PET scan actually "seeing," and why does PET require hospitals to have on-site cyclotrons?

In Chapters 36 and 37 we explored atomic structure and saw how atoms join to form molecules and solids. Now we turn inward, to the atomic nucleus. Since Ernest Rutherford and his colleagues discovered the nucleus in 1911, we've known that all the atom's positive charge and nearly all its mass are concentrated in a tiny nuclear region only about 10^{-5} of the atom's diameter. By 1920 Rutherford had proposed that nuclei beyond hydrogen contain neutral as well as positive particles, and today we know the nucleus is a composite of positive protons and neutral neutrons—collectively called **nucleons**. As we've seen, the uncertainty principle implies high minimum energies for particles confined in small regions, so we can infer that the nucleus is a huge energy repository. We'll conclude this chapter with a look at humankind's attempts to harness that energy.

38.1 Elements, Isotopes, and Nuclear Structure

We saw in Chapter 36 how the number of electrons determines an atom's shell structure and therefore its chemical behavior. It's the number of protons in the nucleus—the **atomic number**, Z—that, in turn, determines the number of electrons in a neutral atom. That means all nuclei with the same Z belong to the same element.

Isotopes and Nuclear Symbols

Nuclei of the same element can, however, have different numbers of neutrons. That's because neutrons don't affect the nuclear charge and therefore have negligible influence on chemical behavior. Nuclei of the same element with different numbers of neutrons are distinct **isotopes**. We call the total number of nucleons the **mass number**, A. Specification of the atomic number Z and mass number A then fully describes a nucleus. Figure 38.1 shows the conventional symbolism used in describing nuclei: the element symbol with a preceding subscript for Z and superscript for A. Actually, the atomic number and symbol are redundant. To be helium (He), for example, *means* to have two nuclear protons and therefore $Z = 2$; to be uranium *means* $Z = 92$. Sometimes, therefore, we write helium-4, He-4, or $^{4}\mathrm{He}$ to mean the same thing as $^{4}_{2}\mathrm{He}$.

Elements typically have several naturally occurring isotopes; a few are shown in Fig. 38.1. Most hydrogen has a single proton in its nucleus, but about one in 6500 hydrogen atoms is deuterium $\left(^{2}_{1}\mathrm{H}\right)$, whose nucleus contains a proton and a neutron. Most oxygen is $^{16}_{8}\mathrm{O}$, but O-17 and O-18 also occur naturally; their ratios in polar ice cores provide valuable information about past climates. Most uranium is $^{238}_{92}\mathrm{U}$, but 0.7% is the U-235 that's used in fission reactors and weapons—hence the great concern about the proliferation of uranium-enrichment facilities to increase the proportion of U-235. Incidentally, the atomic masses listed in the periodic table are averages that reflect the natural abundances of an element's several isotopes. Most elements also have short-lived radioactive isotopes that don't usually occur naturally but can be produced through nuclear reactions; more on these later.

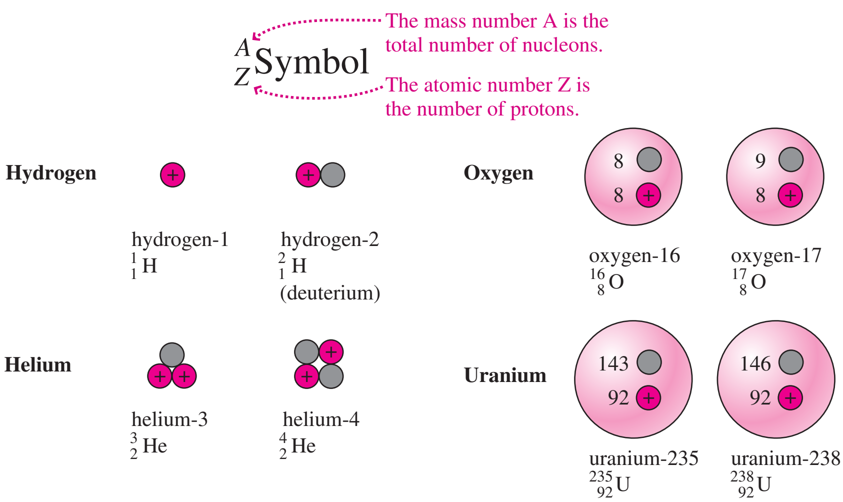


FIGURE 38.1 Isotopes of a given element have the same number of protons but different numbers of neutrons.

GOT IT? 38.1 Determine the number of protons and neutrons in these nuclei: (a) $^{12}_{6}\mathrm{C}$; (b) $^{15}_{8}\mathrm{O}$; (c) $^{57}_{26}\mathrm{Fe}$; (d) $^{239}_{94}\mathrm{Pu}$.

The Nuclear Force

Given the electrical repulsion of the protons, there must be another force acting attractively to bind the nuclear constituents. Throughout much of the 20th century, this **nuclear force** was considered fundamental, but we now recognize it as a manifestation of a more fundamental force between the quarks that make up neutrons and protons. We'll explore quarks and their interactions in Chapter 39.

The attractive nuclear force acts between all nucleons—neutrons and protons, protons and protons, neutrons and neutrons. It's very strong at distances of less than a few femtometers $\left(10^{-15}\,\mathrm{m}\right)$, but falls approximately exponentially with distance—more rapidly than the inverse-square falloff of the electric force. The attractive nuclear force therefore dominates between two neighboring protons, but electrical repulsion becomes dominant for more widely separated protons. The structure of the nucleus is determined, to a first approximation, by the interplay between the weaker but long-range electric force and the stronger but shorter-range nuclear force.

Stable Nuclei

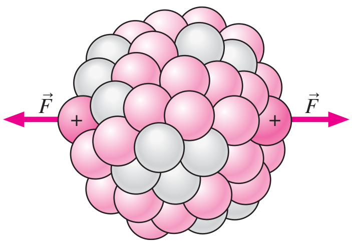


FIGURE 38.2 Two widely separated protons in a large nucleus experience significant electrical repulsion and negligible nuclear attraction.

Not every combination of protons and neutrons will stick together indefinitely. Too many protons, and electrical repulsion wins out; sooner or later the nucleus decays by emitting a chunk of nuclear material (more details in Section 38.2). In larger nuclei most protons are far apart and therefore experience electrical repulsion more strongly than nuclear attraction (Fig. 38.2). To hold these nuclei together therefore requires more neutrons, which contribute attractive nuclear force but not electrical repulsion. So larger nuclei tend to have a higher ratio of neutrons to protons. Even this effect has its limits, though, and the result is that there are no stable nuclei for $Z > 83$.

Too many neutrons also make a nucleus unstable. That's because the exclusion principle requires extra neutrons to go into higher energy states, making individual particles more likely to escape the nucleus. Furthermore, the neutron itself is an unstable particle; an isolated neutron decays spontaneously into a proton, an electron, and an elusive particle called a neutrino. This decay is suppressed in stable nuclei, but occurs if there are too many neutrons.

The delicate balance between neutrons and protons results in about 400 known stable nuclei, collectively called **nuclides**. Figure 38.3 is a **chart of the nuclides**, showing the stable nuclei, along with many unstable ones, on a chart of atomic number Z versus neutron number $N = A - Z$. The chart shows that lighter nuclei tend to have equal numbers of protons and neutrons, but that heavier nuclei invariably have more neutrons to compensate for the increasing electrical repulsion of their widely separated protons.

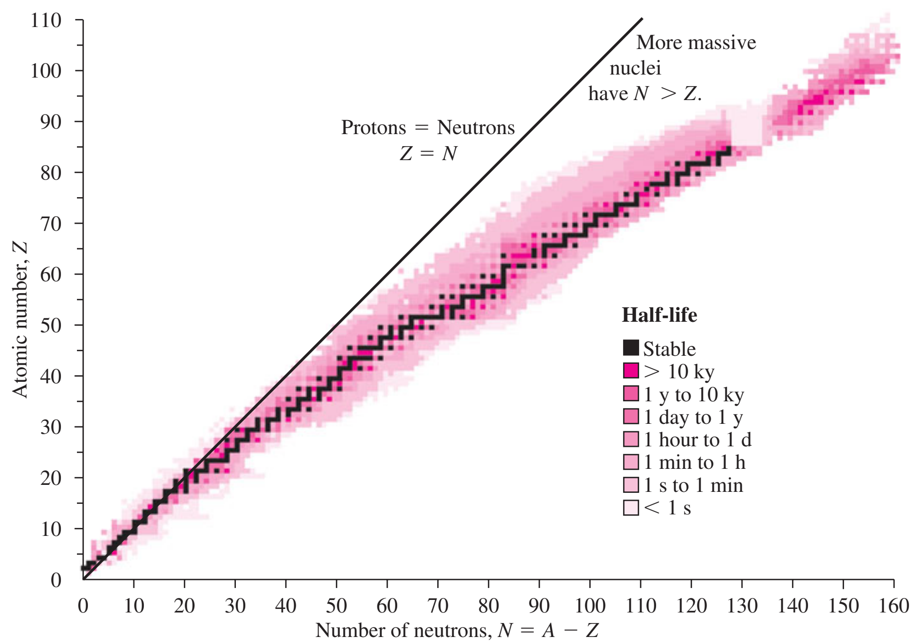


FIGURE 38.3 A chart of the nuclides, color-coded by half-life.

Nuclear Size

Unlike atomic electrons in their widely separated orbitals, nucleons pack tightly into the nucleus. Studies show that most nuclei are spherical, with the **nuclear radius**—defined as the radius at which the density has fallen to half its central value—given approximately by

$$R = R_0 A^{1/3} \tag{38.1}$$

where $R_0 = 1.2$ fm and A is the mass number. This cube-root dependence is what we should expect for a tightly packed sphere whose volume is proportional to the number A of its constituent particles, as suggested in Fig. 38.2. This tight packing also suggests that all nuclei have approximately the same density, on the order of 10^{17} kg/m^3. A teaspoon of nuclear matter has a mass roughly equal to the mass of the Rock of Gibraltar! That absurdly high density reaffirms our picture of the complete atom as mostly empty space with its mass concentrated in a tiny nucleus.

Nuclear Spin

In Chapter 36 we noted the important role of electron spin in atomic structure. Protons and neutrons are, like electrons, spin-$\frac{1}{2}$ particles. The spins of individual nucleons, combined with any angular momentum associated with their motions within the nucleus, give the nucleus a quantized spin angular momentum I that obeys the same rules we've seen for other quantized angular momenta:

$$I = \sqrt{i(i+1)}\hbar \tag{38.2}$$

Here i, the nuclear spin quantum number, is a multiple of one-half. The component of I on a given axis is also quantized, just like other angular momenta, according to $I_z = m_i\hbar$, where m_i ranges from $-i$ to i in steps of 1.

The spin quantum number i is an even or odd multiple of one-half depending on whether the number of nucleons is even or odd. That makes even-A nuclei bosons, particles with integer spin that don't obey the exclusion principle. Odd-A nuclei, in contrast, have half-integer spin and are fermions that obey the exclusion principle. This distinction can lead to profound differences in physical behavior between isotopes of the same element. Helium-4, for example, becomes superfluid at low temperatures, meaning it flows without any viscosity. That's possible because helium-4 nuclei are bosons that can all occupy the same quantum state. Similar superfluidity doesn't occur in fermionic helium-3, although at extremely low temperatures He-3 nuclei themselves pair to form spin-1 particles that do make a superfluid.

The angular momentum of the nucleus results in a nuclear magnetic dipole moment, usually expressed in units of the **nuclear magneton**, $\mu_N = e\hbar/2m_p = 5.05\times10^{-27}$ J/T, where m_p is the proton mass. The proton itself has a magnetic moment whose component on a given axis takes either of the values $\pm2.793\ \mu_N = \pm1.41\times10^{-26}$ J/T—a value that's usually listed as "the magnetic moment of the proton" although it's actually the component. Interaction of the nuclear magnetic moment with magnetic fields alters very slightly the energy levels of the atom—although the effect is much smaller than with atomic electrons because the higher proton mass makes for a much smaller magnetic moment. In hydrogen, for example, the proton can have either of two spin orientations relative to the magnetic field due to the electron, and the result is **hyperfine splitting** of the ground state into two levels a mere 5.9 μeV apart (Fig. 38.4). Transitions between these levels result in a spectral line at a radio wavelength of 21 cm. Radio astronomers use this line to detect interstellar clouds of neutral hydrogen.

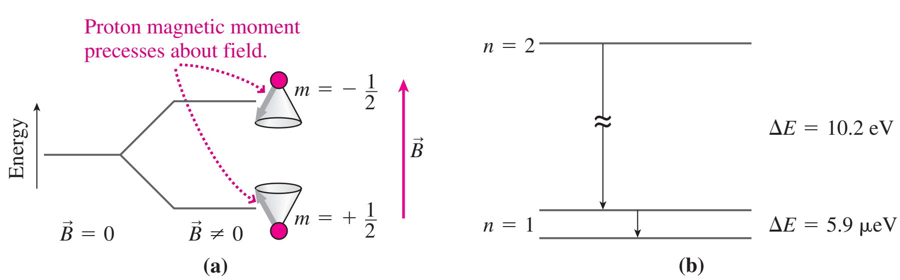
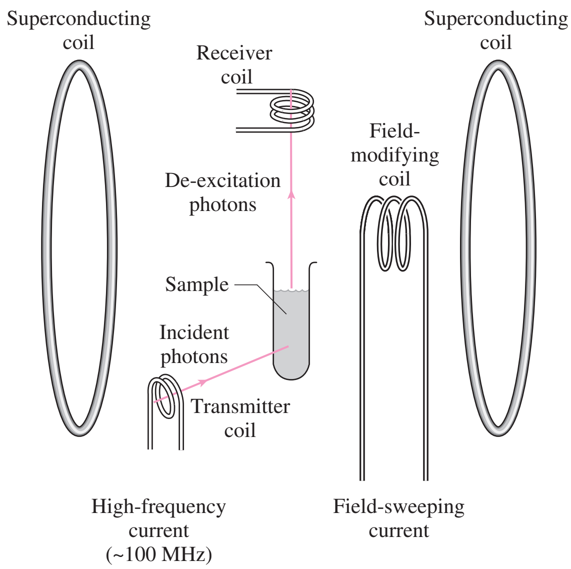


FIGURE 38.4 (a) A nonzero magnetic field $\vec{B}$ splits the energy level of the spin-$\frac{1}{2}$ proton into two levels. (b) The two possible orientations of the proton in the magnetic field of the electron split the hydrogen ground state into two levels 5.9 μeV apart.

APPLICATION Nuclear Magnetic Resonance

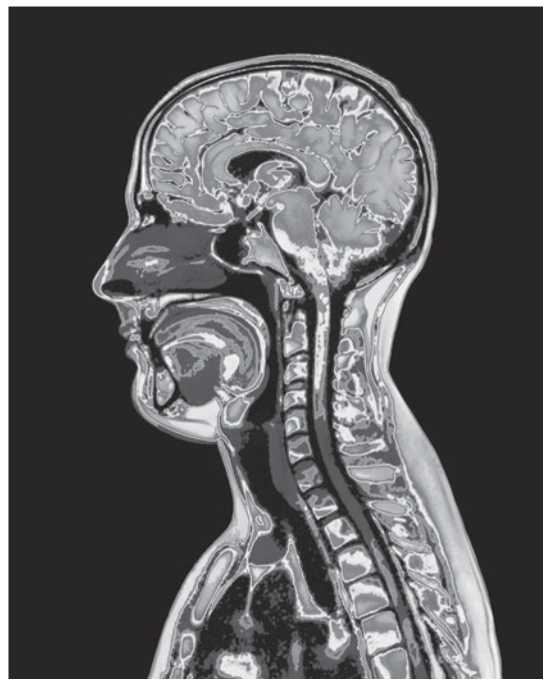


Putting nuclei in an external magnetic field creates two possible energy states, as suggested in Fig. 38.4*a*, depending on whether the nuclear magnetic moments are more nearly parallel or antiparallel to the field. Applying electromagnetic radiation with the appropriate photon energy will flip nuclei into the higher energy state. But because nuclei also experience magnetic fields from the electrons moving around them, the exact energy required is extremely sensitive to the details of the electron distribution—that is, to the surrounding molecular structure.

Nuclear magnetic resonance (NMR) uses this nuclear spin flipping to determine the structure of chemical compounds. In an NMR spectrometer, shown in the figure, the sample under analysis is placed in a uniform magnetic field B, usually from superconducting coils. A smaller coil carries AC current at a frequency f corresponding to photon energy hf that would flip the spin of an isolated nucleus in the field B. The coil emits electromagnetic waves, and if the nuclei absorb the associated photons, then they flip into their higher states and drop back, emitting radiation of frequency f in the process. A receiver coil detects this radiation.

Because of the extra magnetic effect of the surrounding electrons, nuclei won't generally flip at the exact frequency and field B. So the field is varied until the superposition of the applied field and the electron-generated field is exactly right. This condition of magnetic resonance then produces the up/down spin flips that generate a signal in the receiver coil. Scanning the field through a range of values detects nuclei in different electron environments, and from this information scientists can deduce the molecular structure.

Nuclear magnetic resonance with protons (H nuclei) is the basis of **magnetic resonance imaging** (MRI), a widely used medical procedure. In MRI, a person is placed inside a large solenoid whose field varies slightly with position. That makes the magnetic resonance frequency a function of position, and thus the resonance signal can be used to localize the protons undergoing magnetic resonance. A computer then uses the resonance information to construct an image. Most of the MRI signal comes from fat and water, making MRI especially good at imaging soft tissue that doesn't show well in X rays. The photo shows an MRI image of a human head and upper torso; soft-tissue structures including the brain are clearly visible.



EXAMPLE 38.1 Nuclear Spins: Finding the MRI Frequency

The MRI solenoid of Example 26.9 produces a 1.5-T magnetic field. What frequency should be used to drive the transmitter coil in this MRI device?

INTERPRET MRI is an implementation of nuclear magnetic resonance using protons (see the Application), so we're being asked for the frequency corresponding to photons that will flip a proton in a 1.5-T magnetic field.

DEVELOP We need to calculate the necessary photon energy and then use $E = hf$ to find the corresponding frequency. We've just seen that the proton acts like a magnetic dipole whose component along the field is $\mu_p = \pm 1.41 \times 10^{-26}$ J/T. Equation 26.16 gives the energy of a magnetic dipole: $U = -\vec{\mu} \cdot \vec{B}$. Here we're given the component of the magnetic moment $\vec{\mu}$ along the field, so our two energies become $U = \pm\mu_p B$, where the signs correspond to the two possible spin orientations. A spin flip changes a proton's energy from $+\mu_p B$ to $-\mu_p B$, so our plan is to find the energy difference between these levels, equate it to the photon energy hf, and solve for f.

EVALUATE We have $E = \mu_p B - (-\mu_p B) = 2\mu_p B$, so

$$f = \frac{E}{h} = \frac{2\mu_p B}{h} = \frac{(2)(1.41\times10^{-26}\ \mathrm{J/T})(1.5\ \mathrm{T})}{6.63\times10^{-34}\ \mathrm{J\cdot s}} = 63.8\ \mathrm{MHz}$$

ASSESS This frequency is in the radio region of the electromagnetic spectrum, consistent with the diagram in the Application showing the use of coils and currents, and the approximate transmitter coil frequency of 100 MHz. ■

Models of Nuclear Structure

We've seen how the right ratio of neutrons to protons is essential for stable nuclei, and why that ratio increases for larger nuclei. Figure 38.3, the chart of the nuclides, summarizes this information. But take a closer look at that figure: There are more stable nuclei for even values of Z, and some—those with the so-called **magic numbers** 2, 8, 20, 28, 50, 82, and 126 protons or neutrons—have many more stable nuclei. Why?

Answering this question and explaining the decay mechanisms and lifetimes of unstable nuclei require a theory of nuclear structure. There still is no complete nuclear theory, analogous to the atomic theory of Chapter 36, that explains all aspects of all nuclei. Our still-imperfect knowledge of the nuclear force, and the tight packing of nucleons, render useless a simple two-particle model like the one we used for hydrogen. Instead, nuclear physicists resort to several models to explain different aspects of nuclear structure. Together, these models provide a good understanding of the nucleus and accurately predict nuclear properties, although not with the precision available in atomic physics.

The **liquid-drop model** provides a reasonable approximation for heavier nuclei, whose many nucleons behave somewhat like the molecules in a drop of liquid. A liquid-drop nucleus can rotate, vibrate, and change shape as long as its volume doesn't change, and the resulting quantized energy levels predict nuclear gamma-ray spectra that are in good agreement with observation. The liquid-drop model also helps explain nuclear fission, as we'll explore in Section 38.4. But it can't account for the dramatic effects of small changes in nucleon number, particularly the role of the magic numbers.

The **nuclear shell model**, advanced in the late 1940s by physicists Maria Goeppert Mayer and J. Hans Jensen, gives the nucleus a shell structure similar to that of atoms. The shells occur because neutrons and protons obey the exclusion principle, and the magic numbers correspond to closed-shell configurations analogous to the electronic structure of inert gases. Closed-shell nucleons are tightly bound, making a magic nucleus particularly stable. Additional nucleons beyond a closed shell stay largely on the outskirts of the nucleus, where they're more readily excited to higher energy levels. Neutrons and protons behave independently in the shell model, and each has its own set of quantum numbers. Closed-shell structure therefore occurs with magic numbers of either protons or neutrons. Some nuclei, like $^{40}_{20}\mathrm{Ca}\ (Z = 20, N = 20)$, are "doubly magic" and show exceptional stability.

The **collective model**, advanced by Niels Bohr's son Aage, combines aspects of the liquid-drop and shell models, emphasizing the collective quantum-mechanical behavior of the nucleons. One remarkable prediction of the collective model is that larger, nonmagical nuclei may be more stable if they take nonspherical shapes.

Active areas of nuclear-structure research involve the creation and exploration of exceptionally heavy or neutron-rich nuclei. The creation of elements 115 and 116 in the early 2000s suggests that physicists are approaching a region of longer-lived nuclei dubbed the "island of stability," which may be associated with a new magic neutron number of 184. And experiments in 2005 created silicon-42, whose relative stability implies that its atomic number $Z = 14$ becomes magic in this neutron-bloated $(N = 28)$ species. Until we have a complete nuclear theory, experiments like these will continue to challenge physicists with nuclear surprises.

38.2 Radioactivity

In 1896 Henri Becquerel of Paris noticed that a photographic plate stored near uranium compounds became fogged, as though exposed to invisible rays. Becquerel had discovered **radioactivity**, wherein some substances spontaneously emit high-energy particles or photons. Marie and Pierre Curie promptly began a thorough exploration of the phenomenon, for which Marie Curie coined the name "radioactivity." The Curies shared the 1903 Nobel Prize in physics with Becquerel, and Marie Curie won the 1911 Nobel Prize in chemistry for her discovery of polonium and radium.

Decay Rate and Half-Life

Radioactivity results from the decay of unstable nuclei, a process that occurs at vastly differing rates in different isotopes. The number of decays per unit time is the **activity** of a radioactive sample; the SI unit of activity is the **becquerel** (Bq), equal to one decay per second. An older unit, the **curie** (Ci), is 3.7×10^{10} Bq and is approximately the activity of 1 gram of radium-226. For a given isotope, activity is proportional to the number N of nuclei present. N decreases as nuclei decay, so we can write

$$\frac{dN}{dt} = -\lambda N$$

where λ is the **decay constant**. As we've seen with discharging capacitors and decaying inductor currents, this differential equation is a prescription for exponential decay. We solve it the same way, multiplying both sides by dt/N and integrating:

$$\int_{N_0}^{N} \frac{dN}{N} = -\lambda \int_0^t dt$$

where N_0 is the initial number of nuclei at time $t = 0$. Evaluating the integrals gives $\ln(N/N_0) = -\lambda t$ or, exponentiating each side and using $e^{\ln x} = x$:

$$N = N_0 e^{-\lambda t} \tag{38.3a}$$

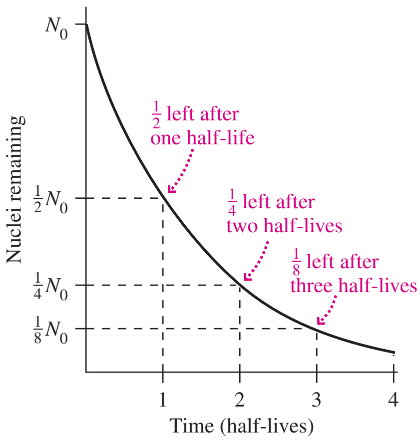


FIGURE 38.5 Exponential decay of a radioactive sample.

Equation 38.3a shows that the decay constant λ is a measure of the exponential decay rate. We can also interpret λ as the probability that a given atom will decay in a 1-s time interval. Another convenient measure of exponential decay is the **half-life**, $t_{1/2}$, defined as the time for half the nuclei in a given sample to decay. If we start with N_0 nuclei at time $t = 0$, then at a later time t the number of nuclei remaining will be

$$N = N_0 2^{-t/t_{1/2}} \quad \text{(radioactive decay)} \tag{38.3b}$$

You can quickly show that $t_{1/2}$ and λ are related by $t_{1/2} = \ln 2/\lambda \simeq 0.693/\lambda$ (see Problem 48). Figure 38.5 is a graph of Equation 38.3b. Since activity and number of nuclei are proportional, both decline with the same half-life, as described in Equation 38.3b. Table 38.1 lists some significant radioisotopes and their half-lives.

Table 38.1 Selected Radioisotopes

Isotope	Half-life	Decay Mode	Comments
Carbon-14 ($^{14}_{6}C$)	5730 years	β^-	Used in radiocarbon dating
Iodine-131 ($^{131}_{53}I$)	8.04 days	β^-	Fission product abundant in fallout from nuclear weapons and reactor accidents; damages thyroid gland
Oxygen-15 ($^{15}_{8}O$)	2.03 minutes	β^+	Short-lived oxygen isotope produced with cyclotrons and used for medical diagnosis in PET
Potassium-40 ($^{40}_{19}K$)	1.25×10^9 years	β^-	Comprises 0.012% of natural potassium; dominant radiation source within the normal human body; used in radioisotope dating
Plutonium-239 ($^{239}_{94}Pu$)	24,110 years	α	Fissile isotope used in nuclear weapons; produced by neutron capture in $^{238}_{92}U$
Radium-226 ($^{226}_{88}Ra$)	1600 years	α	Highly radioactive isotope discovered by Marie and Pierre Curie; results from decay of $^{238}_{92}U$
Radon-222 ($^{222}_{86}Rn$)	3.82 days	α	Radioactive gas formed naturally in decay of $^{226}_{88}Ra$; seeps into buildings, where it may cause serious radiation exposure
Strontium-90 ($^{90}_{38}Sr$)	29 years	β^-	Fission product that behaves chemically like calcium; readily absorbed into bones
Tritium ($^{3}_{1}H$)	12.3 years	β^-	Hydrogen isotope used in biological studies and to enhance yields of nuclear weapons
Uranium-235 ($^{235}_{92}U$)	7.04×10^8 years	α	Fissile isotope comprising 0.72% of natural uranium; used as reactor fuel and in simple nuclear weapons
Uranium-238 ($^{238}_{92}U$)	4.46×10^9 years	α	Predominant uranium isotope; cannot sustain a chain reaction

EXAMPLE 38.2 Radioactive Decay: Fallout from Chernobyl

The 1986 nuclear accident at Chernobyl in Ukraine spread radioactive fallout over Eastern Europe and Scandinavia. Iodine-131 was of great concern, since it's absorbed by the thyroid gland and can cause thyroid cancer. Following the accident, I-131 levels in Romanian milk rose to 2900 Bq per liter. How long did Romanians have to wait before I-131 levels were below their government's 185-Bq/L safety standard?

INTERPRET This is a problem about radioactive decay. We're given the initial activity per liter of milk, and we need to find how long it takes for that to decay to the given safety level. Using Table 38.1, we identify I-131's half-life as 8.04 days.

DEVELOP Equation 38.3b, $N = N_0 2^{-t/t_{1/2}}$, describes equally well the decline in the number of radioactive nuclei and their activity. The equation shows that after n half-lives, activity drops to $1/2^n$ of its original level. So our plan is to find the number of half-lives n that will lower the milk's I-131 activity from 2900 Bq/L to 185 Bq/L. Mathematically, we want $1/2^n = 185/2900$. We'll solve for n and then use the known half-life to get an actual time.

EVALUATE Inverting our expression $1/2^n = 185/2900$ and taking logarithms of both sides give

$$\ln(2^n) = \ln(2900/185)$$

But $\ln(2^n) = n \ln 2$, so

$$n = \frac{\ln(2900/185)}{\ln 2} = 3.97 \text{ half-lives}$$

With $t_{1/2} = 8.04$ days, this amounts to 32 days.

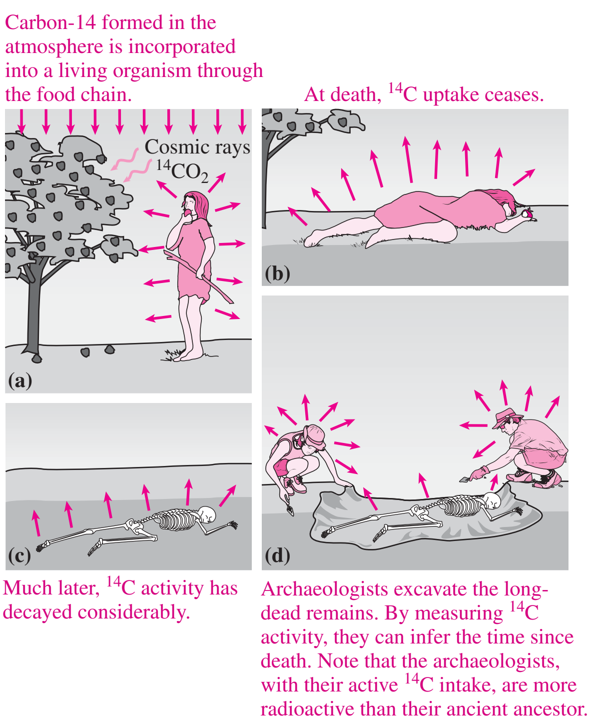
ASSESS A quick check shows that our answer is about right: After one half-life the activity has dropped in half, to $2900/2 = 1450$ Bq/L. Another half-life, and it's half of this, or 725 Bq/L. A third, and it's about 360 Bq/L. A fourth, and it's down to about 180 Bq/L. So to reach 185 Bq/L must take just under 4 half-lives. Romania's 32-day wait time was dependent on both the government safety standard and the initial contamination level. Problem 49 explores the situation in three other affected countries.

✓TIP Half-Life and Powers of 2

After n half-lives, activity has dropped by a factor of $1/2^n$. When we estimate activity levels, it's useful to note that $2^{10} = 1024$, or very nearly 1000. Thus activity drops by a factor of about 1000 every 10 half-lives—and therefore by about 1 million in 20 half-lives.

GOT IT? 38.2 A PET-scan patient is injected with radioactive oxygen-15, whose half-life is 2 min. Approximately what fraction of the original ^{15}O remains undecayed an hour later?

APPLICATION Radiocarbon Dating



Archaeologists, art historians, geologists, and others use radioactive decay to date ancient objects. For ages up to a few tens of thousands of years, the 5730-year isotope carbon-14 is especially useful. ^{14}C forms continuously in the atmosphere through reactions of cosmic rays with nitrogen. Living things take in ^{14}C and maintain a steady concentration through the balance between uptake and radioactive decay. At death, uptake ceases and the level of ^{14}C begins to drop. Measuring the ratio of ^{14}C to stable ^{12}C in a sample of once-living matter and comparing with the ratio found in living material then provides the age (see the figure and Example 38.3).

The cosmic-ray flux at Earth varies with solar activity, and so, therefore, does the atmospheric $^{14}C/^{12}C$ ratio. Scientists correct for this effect with data from growth rings in ancient trees, which provide an independent measure of age. Measuring the actual radioactivity takes a fairly large sample, so today the most sophisticated dating is done instead by counting individual C-14 atoms, separating them from ordinary C-12 using a mass spectrometer—a device we described in Example 26.2.

Radiocarbon dating is quite accurate to about 20,000 years and can be used back to about 50,000 years. For longer time spans, up to the billions of years characterizing the ages of rocks, ratios of longer-lived isotopes provide age information. Much knowledge of our own past, and our planet's and our solar system's, comes from radioisotope dating.

EXAMPLE 38.3 Radioactive Decay: Archaeology

Archaeologists unearth charcoal from an ancient campfire and find its carbon-14 activity per unit mass to be 7.4% of the activity measured in living wood. Find the charcoal's age.

INTERPRET This is a problem about using the decay of carbon-14 to date a once-living material. We want the time it takes for ^{14}C activity to decline to 7.4% of its original level. From Table 38.1, we identify the half-life of ^{14}C as 5730 years.

DEVELOP Equation 38.3b, $N = N_0 2^{-t/t_{1/2}}$, shows that activity drops by $1/2^n$ in n half-lives, so our plan is to find the number of half-lives that makes the factor $1/2^n$ equal to 0.074. Then we can multiply by the half-life to get the actual time.

EVALUATE Solving as we did in Example 38.2 gives

$$n \ln 2 = \ln(1/0.074)$$

which gives $n = 3.76$ half-lives. With $t_{1/2} = 5730$ y, the age is then 21,500 years.

ASSESS Again a quick check suffices: One half-life drops activity to 50%; two half-lives to 25%, three to 12.5%, and four to just over 6%. So it must take a little less than four half-lives to get down to 7.4% of the original activity level.

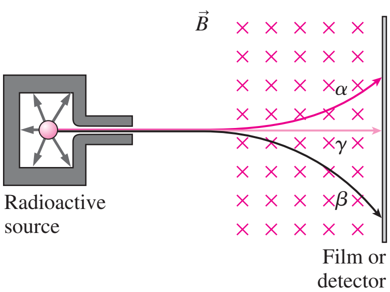


FIGURE 38.6 The three types of radiation go separate ways in a magnetic field.

Types of Radiation

Passing nuclear radiation through a magnetic field shows that there are three types: one positively charged, one negative, and one neutral (Fig. 38.6). Early researchers named these alpha, beta, and gamma radiation, respectively. Today we know that alpha radiation consists of He-4 nuclei, beta radiation consists of high-energy electrons (or positrons), and gamma rays are high-energy photons. They differ in penetrating power: A sheet of paper can stop alpha particles, several centimeters of matter stop most betas, and gamma rays can penetrate substantial thicknesses of concrete or lead. Different radioisotopes emit not only different types of radiation but also radiation of different energies.

Alpha Decay

Alpha emitters are nuclei with too much positive charge. They shed charge, and mass, by emitting a bundle of two protons and two neutrons—an alpha particle, ^4_2He. Symbolically,

$$^A_ZX \longrightarrow {}^{A-4}_{Z-2}Y + {}^4_2He \qquad (\text{alpha decay}) \tag{38.4}$$

Here X is the original or **parent** nucleus, and Y is the **daughter**. Note that the sums of the atomic numbers on both sides of this equation are equal, as are the mass numbers. Most of the energy released in the reaction appears as kinetic energy of the alpha particle. The alpha particle actually emerges with less energy than needed to overcome the nuclear potential barrier, and this provides one of the most direct confirmations of quantum tunneling—which is the only way the alpha particle can escape the nucleus.

Beta Decay

Beta emitters have too many neutrons, one of which decays into an electron, a proton, and an elusive neutral particle called a neutrino (symbol ν). The electron exits at high energy to form beta radiation, leaving a nucleus with essentially the same mass but its atomic number increased because it has one more unit of positive charge:

$$^A_ZX \longrightarrow {}_{Z+1}^{\;\;A}Y + e^- + \bar{\nu} \qquad (\text{beta decay}) \tag{38.5a}$$

In ordinary beta decay the neutrino is, in fact, an antineutrino—hence the bar over its symbol ν.

Beta decay is a manifestation of the weak nuclear force, and in the Sun it produces a steady stream of neutrinos that provide direct information on conditions in the solar core. That's because neutral, nearly massless neutrinos interact only rarely with matter; for example, they pass through the entire Earth with little probability of interaction. We'll see in Chapter 39 how neutrinos nonetheless are opening a new window on distant astrophysical events and the early universe.

A second type of beta decay converts a proton into a neutron, emitting both a positron (an anti-electron, e^+) and a neutrino:

$$ {}^{A}_{Z}X \longrightarrow {}^{\;\;\,A}_{Z-1}Y + e^+ + \nu \qquad (\text{beta decay, positron emission}) \qquad (38.5b)$$

This reaction occurs in some short-lived isotopes of lighter elements like carbon and oxygen, and gamma rays from the subsequent annihilations of positrons are used in the medical imaging procedure known as positron emission tomography (PET).

A third beta-decay process is **electron capture**, in which a nucleus captures an inner-shell atomic electron, converting a proton to a neutron and ejecting a neutrino:

$$ {}^{A}_{Z}X + e^- \longrightarrow {}^{\;\;\,A}_{Z-1}Y + \nu \qquad (\text{electron capture}) \qquad (38.5c)$$

Gamma Decay

A nucleus in an excited state decays by emitting a photon, just like an atom. But the much higher energy associated with nuclear processes puts such photons in the gamma-ray region of the spectrum. Since the gamma-ray photon is neutral and massless, it doesn't change the type of nucleus; therefore, we write

$$ {}^{A}_{Z}X^* \longrightarrow {}^{A}_{Z}X + \gamma \qquad (\text{gamma decay}) \qquad (38.6)$$

where X^* designates the excited state.

Decay Series and Artificial Radioactivity

A few radioisotopes, like ^{40}K and ^{238}U, have half-lives comparable to Earth's age, so it's not surprising to find these in nature. But we also find shorter-lived species. Some, like cosmic-ray-produced ^{14}C, result from naturally occurring nuclear reactions. Many others arise in the decay of long-lived isotopes, while some we produce in particle accelerators, nuclear reactors, and nuclear explosions.

Figure 38.7 shows the **decay series** for uranium-238, whose 4.46-billion-year half-life ensures that there's still plenty of it around. The shorter-lived daughter products in this series are present wherever there's natural uranium. A balance between formation and decay establishes the abundance of each product in the decay series. One of the uranium daughters is radon-222, a radioactive gas that can be a serious health hazard in closed spaces.

In 1930 Marie Curie's daughter Irène and her husband Frédéric Joliot-Curie were the first to induce artificial radioactivity, by bombarding stable isotopes with alpha particles. Today we produce radioisotopes with particle beams or with neutrons from nuclear reactors, or by extracting them from the by-products of nuclear fission.

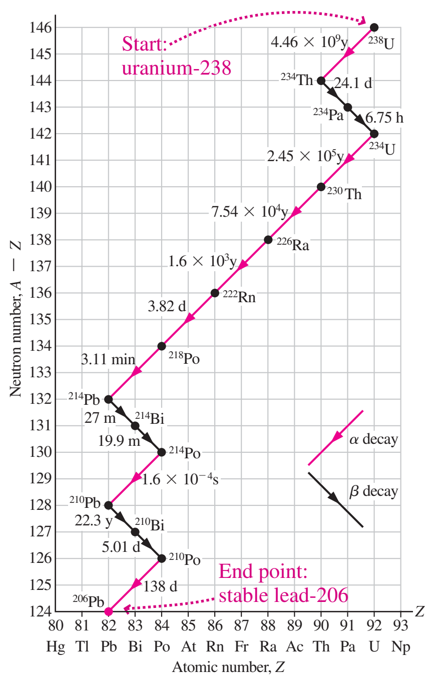


FIGURE 38.7 The decay of uranium-238 results in a series of shorter-lived nuclei. Times shown are half-lives.

Uses of Radioactivity

Nuclear radiation has numerous beneficial uses in our technological society. Here we survey just a few:

- **Radioactive Tracers** "Tagging" molecules with radioactive atoms makes it easy to trace their flows through biological and physical systems. Biologists use radioactive tracers routinely to study the uptake and distribution of chemicals. Engineers use radioisotopes to study wear in mechanical parts. Physicians "tag" bone-seeking compounds with radioisotopes to image the skeletal system; the resulting "bone scans" reveal cancer and other diseases.

- **Cancer Treatment** Radiation destroys living cells, especially fast-dividing cancer cells. Early radiation treatment used gamma radiation; today, particle beams deliver radiation with less effect on surrounding tissue. Alternatively, radioisotope "seeds" are embedded directly into a tumor.
- **Food Preservation** High radiation doses destroy bacteria and enzymes that cause food spoilage, providing longer shelf life and a safer food supply. Though controversial, food irradiation is becoming increasingly common.
- **Insect Control** Radiation preferentially damages reproductive cells and can therefore cause sterility. Sterilizing large numbers of pest insects with radiation causes populations to collapse when the sterile insects mate with normal ones. The Mediterranean fruit fly, a serious pest of citrus crops, has been controlled in this way.
- **Fire Safety** Common smoke detectors contain americium-241, whose alpha radiation ionizes air, allowing it to carry electric current. Smoke particles interfere with the current, triggering the alarm. Exit signs containing radioactive tritium (^3H) glow without the need for electricity, providing another measure of fire safety in public buildings.
- **Activation Analysis** Bombarding materials with neutrons or other particles results in excited states or the production of unstable isotopes. Analyzing the resulting radiation helps identify unknown materials. Art historians use this technique to detect forgeries; environmental scientists identify the constituents of pollution; and airport luggage scanners search for the radiation "fingerprint" of chemical explosives.

Biological Effects of Radiation

Nuclear radiation has sufficiently high energy to ionize or otherwise disrupt biological molecules. Results include cell death, loss of biological functions, and mutations that lead to cancer or to genetic changes in future generations. Many early nuclear scientists, including Marie Curie and her daughter Irène, succumbed to leukemia and other cancers that undoubtedly resulted from radiation exposure.

The energy absorbed in a radiation dose is a rough measure of its biological danger. The SI unit of absorbed dose is the **gray** (Gy), defined as 1 J of energy per kg of absorbing material. A more appropriate measure is the **sievert** (Sv), which is weighted by the biological effectiveness of particular radiation types. Alpha particles, for example, cause more damage per unit energy than do gamma rays, so 1 Gy of alpha radiation is more harmful than 1 Gy of gamma radiation. But 1 Sv of alphas and 1 Sv of gammas cause essentially the same damage.

The biological effects of high radiation doses are well known; exposure to 4 Sv, for example, causes death in about half its human victims. But doses in the 0.1-Sv range and lower are more controversial. There are only a few cases of well-quantified exposures to populations large enough that small effects can be determined accurately. Even less certain are the effects of very low doses, such as the 10^{-5}-Sv average dose to people living near the 1979 Three Mile Island nuclear accident. On the one hand, biological repair mechanisms may limit damage at low doses. On the other, even low doses may disproportionately affect the young and the unborn. A 2005 study by the U.S. National Academy of Sciences (NAS) suggests that the risk of cancer—the dominant health effect of low-level radiation—should scale linearly with dose. For a one-time dose of 1 mSv, the NAS study estimates a lifetime cancer risk of 1 in 10,000. This compares with a 42% lifetime chance of developing cancer from all causes.

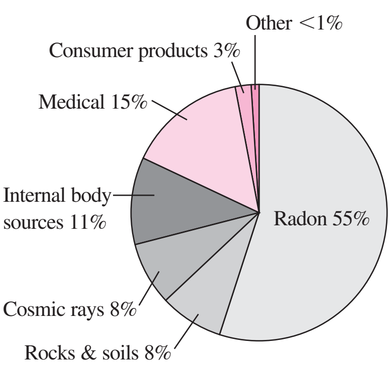


FIGURE 38.8 Natural (gray) and artificial (color) sources of radiation, as percentages of the U.S. average yearly dose of 3.6 mSv. "Other" includes nuclear power, radioactive waste, and weapons tests.

The average U.S. citizen receives about 3.6 mSv of radiation per year, most of it from natural sources (Fig. 38.8). The dominant source, at 55%, is the uranium decay product radon-222, which seeps into buildings from the decay of naturally occurring uranium in the ground and in building materials. Our own bodies account for some 11%, mostly from natural potassium-40. About 19% of our average exposure in the United States comes from artificial sources, mainly medical procedures.

Consumer products—mostly tobacco, drinking water, and building materials—account for about 3%. Less than 1% of our yearly radiation dose is from nuclear power and weapons. Radiation doses vary greatly with location and occupation; for example, residents of high-altitude Denver have greater exposure to cosmic rays, and airline flight crews' cosmic-radiation dose often exceeds the allowed dose for nuclear-plant workers. No matter what your exact dose, though, the risks to your health from radiation exposure pale compared with other risks you knowingly take.

38.3 Binding Energy and Nucleosynthesis

Disassembling a nucleus requires energy to overcome the strong nuclear force. The more tightly bound the nucleus, the higher this **binding energy**. The energies involved in nuclear interactions are high enough that Einstein's mass–energy equivalence is clearly evident, so accounting for energy conservation requires us to consider the rest energy of the particles. Then we can write

$$m_N c^2 + E_b = Z m_p c^2 + (A - Z) m_n c^2 \qquad (38.7)$$

where the terms on the left are the rest energy of the nucleus, whose mass is m_N, and the binding energy E_b. The terms on the right are the rest energies of the Z individual protons and $A - Z$ neutrons that make up the nucleus. So Equation 38.7 shows that we can disassemble a nucleus into its constituent nucleons if we supply additional energy equal to the binding energy. Equivalently, E_b is the energy released if we assemble a nucleus from isolated nucleons.

Equation 38.7 shows that the nuclear mass m_N is *not* the sum of the constituent particles' masses; rather, it's less by the amount E_b/c^2. This is clear evidence for mass–energy equivalence. Again, as in Chapter 33, we emphasize that there's nothing uniquely nuclear about this so-called **mass defect**. The mass of a water molecule is also less than the sum of its constituent hydrogen and oxygen atoms—but with chemical binding the effect is so small as to be virtually immeasurable. It's the strength of the nuclear force that makes mass–energy equivalence more obvious in nuclear interactions.

It's convenient to measure nuclear and particle masses in **unified mass units**, u, defined as one-twelfth the mass of a neutral carbon-12 atom. The unified mass unit is very nearly 1.66054×10^{-27} kg, slightly less than the mass of the proton or neutron. High-energy physicists, ever cognizant of mass–energy equivalence, often express masses in MeV/c^2—a value numerically equal to the rest energy in MeV. Table 38.2 lists selected particle masses in kg, u, and MeV/c^2. In practice one often knows atomic rather than nuclear masses, but the difference is generally negligible because the extra mass of the electrons is so small.

Table 38.2 Selected Masses

	Mass (kg)	Mass (u)	Mass (MeV/c^2)
Electron	$9.109\,39 \times 10^{-31}$	0.000 548 579	0.510 999
Proton	$1.672\,62 \times 10^{-27}$	1.007 276	938.272
Neutron	$1.674\,93 \times 10^{-27}$	1.008 665	939.566
$^{1}_{1}$H atom	$1.673\,53 \times 10^{-27}$	1.007 825	938.783
α particle ($^{4}_{2}$He nucleus)	$6.644\,66 \times 10^{-27}$	4.001 506	3727.38
$^{12}_{6}$C atom	$1.992\,65 \times 10^{-26}$	12	11 177.9
Unified mass unit (u)	$1.660\,54 \times 10^{-27}$	1	931.494

EXAMPLE 38.4 Mass Defect in Helium: Powering the Sun

Use the appropriate masses from Table 38.2 to find the binding energy of ${}^{4}_{2}\text{He}$.

INTERPRET This is a question about binding energy—the energy difference between separate constituents of helium-4 and the helium-4 nucleus. We identify the constituent particles from the symbol ${}^{4}_{2}\text{He}$: $Z = 2$ protons and $N = A - Z = 2$ neutrons.

DEVELOP Equation 38.7 determines the binding energy in terms of the various masses:

$$E_\text{b} = Zm_\text{p}c^2 + (A - Z)m_\text{n}c^2 - m_\text{N}c^2$$

EVALUATE Using our values for Z and $N = A - Z$, along with the proton, neutron, and alpha-particle (He-4 nucleus) masses from Table 38.2, gives

$$E_\text{b} = 2(938.272\ \text{MeV}/c^2)c^2 + 2(939.566\ \text{MeV}/c^2)c^2 - (3727.38\ \text{MeV}/c^2)c^2 = 28.3\ \text{MeV}$$

ASSESS Notice how easy it was to work with mass in units of MeV/c^2; the factor c^2 canceled and we didn't need to use the speed of light explicitly. The formation of helium through a sequence of nuclear reactions is what powers the Sun, and our 28.3-MeV result is very close to the actual 26.7 MeV released for each He-4 nucleus formed through the solar process. ■

The Curve of Binding Energy

Binding energy plays a crucial role in the formation of the elements and in nuclear energy. Figure 38.9 shows the **curve of binding energy**, a plot of binding energy *per nucleon* as a function of mass number A. The higher this quantity, the more tightly bound is the nucleus. The broad peak in the vicinity of $A = 60$ shows that nuclei with mass numbers around this value are most tightly bound. That means it's energetically favorable for two lighter nuclei to join through the process of **nuclear fusion**, making a middle-weight nucleus. But heavier nuclei can reach a lower energy state if they split or **fission** into two middle-weight nuclei. We'll discuss fission and fusion later in the chapter.

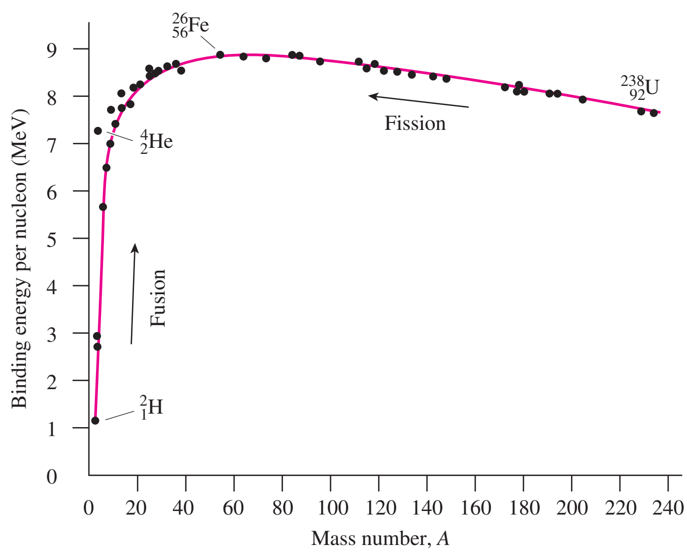


FIGURE 38.9 The curve of binding energy, showing how fusion and fission can result in the release of nuclear energy.

GOT IT? 38.3 Rank order these nuclei from the most to the least tightly bound: ${}^{4}_{2}\mathrm{He}$, ${}^{238}_{92}\mathrm{U}$, ${}^{57}_{26}\mathrm{Fe}$, ${}^{2}_{1}\mathrm{H}$, ${}^{132}_{54}\mathrm{Xe}$.

Nucleosynthesis and the Origin of the Elements

Since it's energetically favorable for light nuclei to fuse together, they'll do so if they have enough energy to overcome their electrical repulsion. This condition held in the high-temperature early universe, particularly from about 1 minute to 30 minutes after the start of the Big Bang. During that time, protons fused to form helium, leaving the universe with approximately its present composition of about 75% hydrogen and 25% helium, with traces of deuterium, lithium, beryllium, and boron. Hundreds of millions of years later the first stars formed, and in the interiors of more massive stars conditions were ripe for a two-step process that fused three helium nuclei to make carbon-12. From there fusion reactions led to the formation of isotopes up to those near the $A = 60$ peak in the curve of binding energy. In fact, the nuclei of essentially all the elements with $A < 60$—including most of the materials in our own bodies—were formed in the interiors of massive stars (Fig. 38.10). Some nuclei with $A > 60$ also formed inside massive stars; others formed in the violent supernova explosions that end such stars' lives. Those explosions spewed fusion-synthesized elements into the interstellar medium where, eons later, they're incorporated into new stars, planets, and even living things.

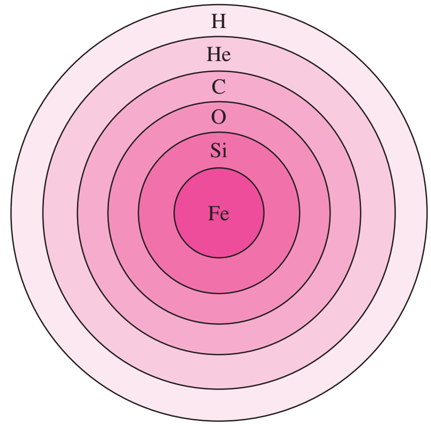


FIGURE 38.10 Onionlike structure of a massive star before it goes supernova. Successive stages of fusion reactions produce the elements shown, which accounts for their relative abundance.

38.4 Nuclear Fission

Neutrons, first discovered in 1932, make excellent probes of the nucleus because they don't have to overcome electrical repulsion. In 1938 the German chemists Otto Hahn and Fritz Strassmann bombarded uranium with neutrons. They were puzzled to find among the reaction products radioactive versions of the much lighter elements barium and lanthanum. Physicist Lise Meitner and her nephew Otto Frisch interpreted these results to mean that uranium had split or, in their words, **fissioned** (Fig. 38.11). It was the eve of World War II, and the military implications were obvious and ominous: Nuclear fission represented an energy source orders of magnitude more potent than chemical reactions. The race for nuclear weapons was on. With the help of the international physics community, many of whom had fled fascism, the U.S. effort succeeded. A team led by the Italian Enrico Fermi built the first nuclear reactor under the stands of the University of Chicago stadium; it became operational in 1942. Three years later came the first nuclear weapons test, at Trinity Site in New Mexico, followed quickly by the nuclear destruction of Hiroshima and Nagasaki.

Although fission can occur spontaneously, it's much more likely when a neutron strikes a nucleus. Figure 38.11 shows U-235 absorbing a neutron to become U-236. This unstable nucleus undergoes dumbbell-shaped oscillations until electrical repulsion tears it apart. The resulting **fission products** are generally a pair of middle-weight nuclei with unequal masses; typically two to three neutrons are also released in fission. Skipping the intermediate U-236 nucleus, neutron-induced fission of U-235 takes the form

$$ {}^{1}_{0}n + {}^{235}_{92}\mathrm{U} \longrightarrow X + Y + b\,{}^{1}_{0}n \qquad \text{(fission)} \tag{38.8}$$

Here ${}^{1}_{0}n$ is the neutron, with 0 charge and 1 mass unit; X and Y are the fission products; and b is the number of neutrons released immediately. A specific example of Equation 38.8 is ${}^{235}\mathrm{U}$ fission that produces barium and krypton: ${}^{1}_{0}n + {}^{235}_{92}\mathrm{U} \longrightarrow {}^{141}_{56}\mathrm{Ba} + {}^{92}_{36}\mathrm{Kr} + 3\,{}^{1}_{0}n$. Note how the equation balances: The total charge (subscripts) is the same on both sides, and the mass numbers (superscripts) also agree.

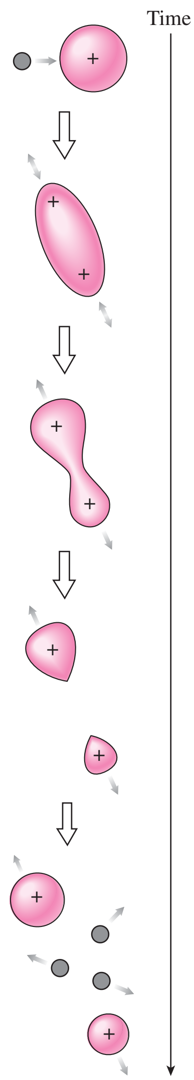


FIGURE 38.11 Neutron-induced fission of ${}^{235}\mathrm{U}$, showing three neutrons (gray) released in the process.

CONCEPTUAL EXAMPLE 38.1 Radioactive Waste!

Use Figure 38.3 to explain why fission products are necessarily radioactive.

EVALUATE Figure 38.3 shows that more massive nuclei need higher ratios of neutrons to protons in order to overcome the protons' electrical repulsion. When uranium fissions, the resulting nuclei have the same neutron-to-proton ratio as the original uranium. But that gives them way too many neutrons, making them highly radioactive via beta decay. Figure 38.12 shows a simplified chart of the nuclides to help make this point.

ASSESS Highly radioactive materials decay rapidly, giving them relatively short half-lives. The longer-lived fission products have half-lives measured typically in decades.

MAKING THE CONNECTION Neutron-induced fission of ^{235}U yields $^{102}_{42}Mo$, three neutrons, and another fission product. What's that product?

EVALUATE This reaction is a specific instance of Equation 38.8: $^1_0n + ^{235}_{92}U \rightarrow ^{102}_{42}Mo + ^Z_AX + 3^1_0n$, with X being the unknown fission product. Balancing atomic and mass numbers gives $A = 131$ and $Z = 50$. The periodic table shows that $Z = 50$ is iodine, so X is I-131, a dangerous contaminant discussed in Example 38.2.

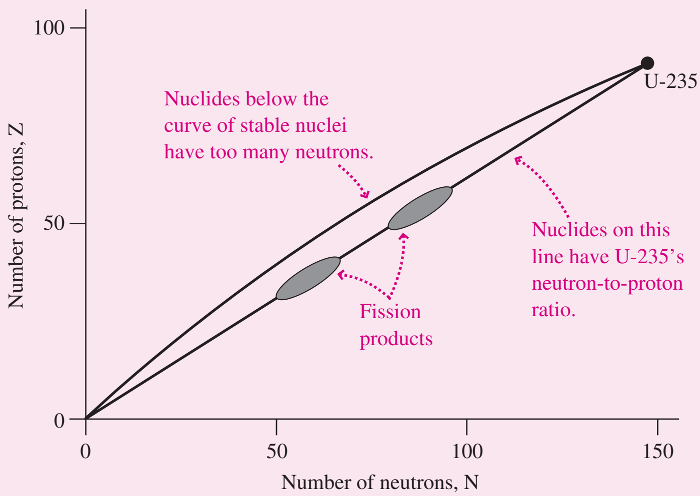


FIGURE 38.12 This chart of nuclides, simplified from Fig. 38.3, shows that fission products lie below the stable nuclei because they have too many neutrons.

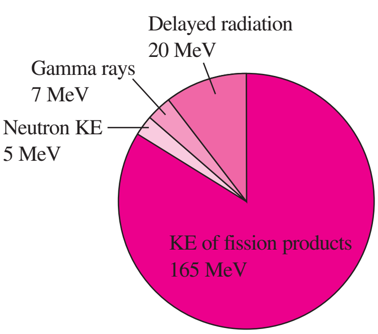


FIGURE 38.13 Fission energy is distributed among fission products, neutrons, and radiation.

Energy from Fission

Fission of a uranium nucleus releases about 200 MeV of energy, as shown in Fig. 38.13. Spontaneous fission is rare because of the energy barrier associated with forces on the outermost nucleons; rather, fission usually results when a nucleus absorbs a neutron, initiating the process shown in Fig. 38.11. Many heavy nuclei, including ^{238}U and ^{235}U, are **fissionable**, meaning they can undergo neutron-induced fission. **Fissile** nuclei will fission with neutrons of *any* energy, including thermal energy. The three important fissile nuclei are uranium-233, uranium-235, and plutonium-239.

Uranium-235 presently constitutes only about 0.7% of natural uranium; nearly all the rest is ^{238}U. For most uses, uranium must be enriched in ^{235}U, to several percent for commercial power reactors and 80% or more for weapons. **Uranium enrichment** is difficult and expensive; since the isotopes ^{235}U and ^{238}U are chemically similar, enrichment techniques make use of their very slight mass difference. The technique of choice today involves spinning uranium hexafluoride gas in a sequence of high-speed centrifuges. Enrichment technology is highly sensitive because a nation possessing it can produce weapons-grade uranium.

Plutonium-239, with a 24,110-year half-life, does not occur in nature. It's produced by neutron bombardment of ^{238}U. The reaction forms ^{239}U, which undergoes two beta decays to produce first ^{239}Np and then the fissile ^{239}Pu:

$$\begin{aligned} ^1_0n + ^{238}_{92}U &\longrightarrow ^{239}_{92}U \\ ^{239}_{92}U &\longrightarrow ^{239}_{93}Np + e^- + \bar{\nu} \\ ^{239}_{93}Np &\longrightarrow ^{239}_{94}Pu + e^- + \bar{\nu} \end{aligned}$$

Although ^{239}Pu is produced copiously in nuclear reactors (see Problem 74), **reprocessing** spent reactor fuel to extract plutonium is difficult and dangerous. Contamination with other plutonium isotopes further complicates the process. Like uranium enrichment, plutonium reprocessing is a sensitive technology, and the decision of several European countries and Japan to engage in commercial reprocessing for reactor fuel has made Pu-239 a commercial commodity.

EXAMPLE 38.5 Nuclear Fission: Comparing with Coal

Assuming 200 MeV per fission, estimate the amount of pure ^{235}U that would provide the same energy as 1 metric ton (1000 kg) of coal.

INTERPRET We're asked to compare the energies of nuclear fission and the chemical burning of coal.

DEVELOP For coal, we can look up the energy released per unit mass in Appendix C's "Energy Content of Fuels" table. We can then find the number of fission events, at 200 MeV per fission, needed to release the same energy as 1000 kg of coal. Finally, we'll use the mass of a U-235 nucleus to find the corresponding mass of uranium.

EVALUATE Appendix C gives an energy content of 29 MJ/kg for coal, so burning 1000 kg of coal releases 29 GJ of energy. With 1.6×10^{-19} J/eV, each 200-MeV fission releases about 3.2×10^{-11} J. Then we need a total of

$$29\ \text{GJ}/3.2\times10^{-11}\ \text{J/fission} = 9.1\times10^{20}\ \text{fission events}$$

Each of the 9.1×10^{20} U-235 nuclei has a mass of approximately 235 u, so the total mass required is

$$(9.1\times10^{20}\ \text{nuclei})(235\ \text{u})(1.66\times10^{-27}\ \text{kg/u}) = 0.35\ \text{g}$$

ASSESS That's about one-hundredth of an ounce of ^{235}U, packing as much energy as a ton of coal! Our result shows that U-235 contains about three million times as much energy as the same amount of coal. That's the reason nuclear power plants are fueled only about once a year, with a truckload or so of nuclear fuel, while coal-burning plants burn many 110-car trainloads of coal each week. It's also the reason for the immense destructive power of nuclear weapons. ■

The Chain Reaction

Neutrons induce fission, and fission itself releases more neutrons. This makes possible a **chain reaction**, in which each fission event results in more fission. To sustain a chain reaction, each nucleus that fissions must, on average, cause at least one more fission event; otherwise, the reaction will fizzle to a halt. In a piece of material that's too small, most neutrons will escape without causing additional fission. For that reason there's a **critical mass** of nuclear fuel necessary to sustain a chain reaction. More than that amount is **supercritical** and results in an exponentially growing chain reaction (Fig. 38.14).

The size of the critical mass depends on the purity of the fissile material, its configuration, and surrounding materials. For plutonium it can be less than 5 kg, and as low as 15 kg for uranium. Those numbers are frighteningly small, and they show why we worry about city-destroying "suitcase bombs."

The **multiplication factor**, k, is the average number of neutrons from a fission event that cause additional fission. A critical mass has $k = 1$, and a supercritical mass has $k > 1$. The average time between successive fissions is the **generation time**. In a supercritical mass this can be as short as 10 ns, leading to the entire mass fissioning in about 1 μs.

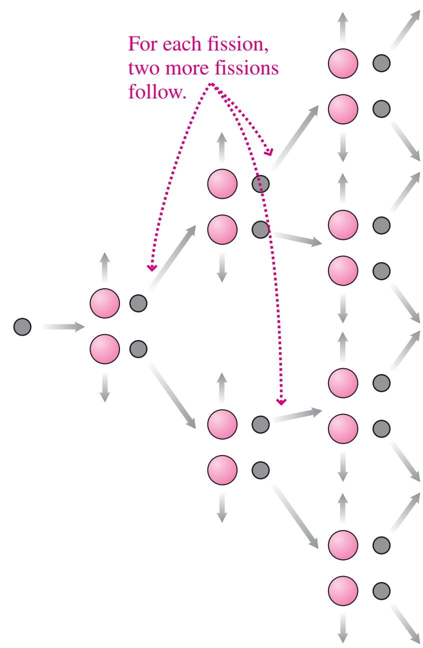


FIGURE 38.14 A supercritical chain reaction with multiplication factor $k = 2$.

Fission Weapons

A rapidly fissioning supercritical mass is a nuclear explosive. The major technological difficulty in producing a fission weapon is to assemble a supercritical mass so rapidly that the chain reaction consumes enough fissile material before it blows apart. With highly enriched uranium that's not an insurmountable challenge. The crude bomb that destroyed Hiroshima contained about 50 kg of enriched uranium, of which only about 1 kg actually fissioned. So confident were its developers that they never tested this design. Plutonium weapons present a greater challenge; neutrons from spontaneous fission make it more likely that the weapon will "pre-ignite" and blow itself apart.

Construction of a simple fission weapon is distressingly straightforward, but acquisition of weapons-grade fissile material is not. Again, that's why uranium enrichment and plutonium reprocessing technologies are so sensitive. We live in a dangerous and unstable world, and it's going to get more dangerous if fissile materials become widely available.

Nuclear Power

A **nuclear reactor** uses a controlled fission chain reaction with $k = 1$ to release energy at a steady rate. Since the average number of neutrons emitted in U-235 fission is about 2.5, reactors require that most neutrons don't cause fission. Commercial power reactors limit k in part by keeping the concentration of fissile U-235 low—typically a few percent—so that many neutrons are absorbed by U-238 instead of causing fission. **Control rods** made of neutron-absorbing material provide additional control over k; these can be moved into and out of the nuclear fuel to provide precise control of the power level. A small fraction—about 0.65%—of fission-produced neutrons are emitted with delays from about 0.2 s to 1 min, and these **delayed neutrons** allow for relatively slow mechanical control of nuclear reactors. The next example explores this point.

EXAMPLE 38.6 Nuclear Fission: Delayed Neutrons and Reactor Control

A change in operating conditions makes a nuclear reactor slightly supercritical, with $k = 1.001$. Determine the time it would take the reactor power to double (a) if delayed neutrons establish a generation time $\tau = 0.1$ s, and (b) if prompt neutrons—those released immediately—sustain the reaction to give $\tau = 10^{-4}$ s.

INTERPRET We're asked to calculate the time until the reaction rate doubles, given the multiplication factor k and two different values for the generation time.

DEVELOP A multiplication factor $k = 1.001$ means the rate of fissioning increases by a factor of 1.001 with each generation time; after two generations it will have increased by k^2, and so forth. So our plan is to find the number n of generations that gives $k^n = 2$. Then we can multiply by the two different τ values to find the actual times: $t = n\tau$.

EVALUATE We set $k^n = 2$ and take the logarithm of both sides. With $\ln(k^n) = n \ln k$, we have $n \ln k = \ln 2$, or $n = \ln 2/\ln k = 693$ with $k = 1.001$. With $\tau = 0.1$ s that gives $t = n\tau = 69.3$ s or just over 1 min, but with $\tau = 10^{-4}$ s it's only 0.07 s.

ASSESS With delayed neutrons the doubling time is long enough for the reactor operators and their mechanical controls to take corrective action; with prompt neutrons there isn't time to prevent a serious nuclear accident. Delayed neutrons are crucial to reactor control! Loss of control to a reaction governed by prompt neutrons alone was a key factor in the 1986 Chernobyl nuclear accident. ■

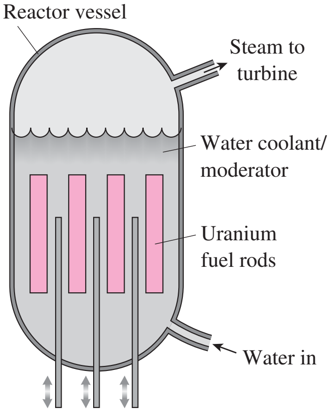


FIGURE 38.15 A boiling-water reactor, one of two types commonly used in the United States.

High-energy fission neutrons aren't very effective at causing additional fission events, so in most reactor designs they must be slowed to roughly the mean thermal speed. A substance called the **moderator** effects this slowing through elastic collisions between neutrons and the moderator nuclei. In Chapter 9 we found that the maximum energy transfer occurs when colliding particles have equal mass; therefore, the best moderators have low-mass nuclei. The choice of moderator is among the most significant distinguishing features of different reactor designs. Another important choice is the **coolant**, which carries off fission-generated heat.

Power reactors in the United States are **light-water reactors** (LWRs), using ordinary water with the protons of its hydrogen serving as the moderator nuclei. The same water acts as coolant and circulates through a pressure vessel containing uranium fuel rods and control rods. About one-third of the United States' roughly 100 power reactors are **boiling-water reactors** (BWRs), in which water boils in the reactor vessel to make steam that drives a turbine-generator (Fig. 38.15). The remainder are **pressurized-water reactors** (PWRs), in which liquid water under pressure transfers its energy to a secondary loop where water boils to make steam (Fig. 38.16). An advantage of this more complex system is that the steam loop doesn't

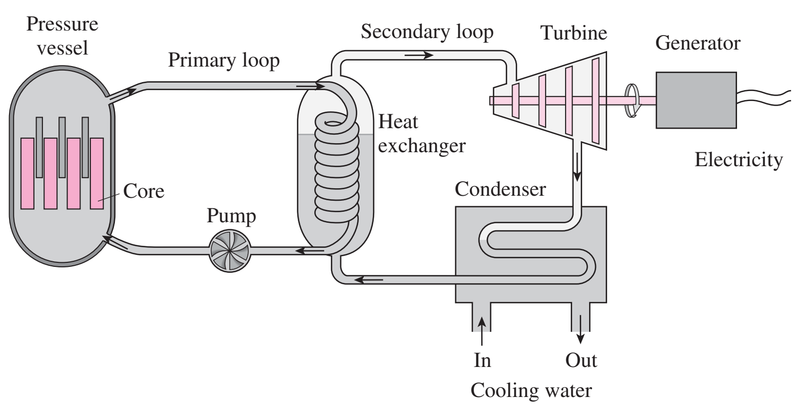


FIGURE 38.16 A complete power plant using a pressurized-water reactor, the most common type of power reactor in the United States.

become radioactive. Both types of light-water reactors have an intrinsic safety feature, in that a loss of coolant also means loss of moderator, and that brings the chain reaction to a halt. But light water has the disadvantage that ${}^1_1H$ readily absorbs neutrons, and therefore light-water reactor fuel must be enriched in ${}^{235}U$ in order to sustain the chain reaction. Refueling a LWR is also a big operation: The reactor must be shut down and the lid removed from the pressure vessel—a process that can take a month or longer.

The Canadian CANDU design uses heavy water (${}^2_1H_2O$, or deuterium oxide) as moderator and coolant. Low neutron absorption means CANDU reactors can operate on natural uranium, eliminating the need for sensitive enrichment technology. And the CANDU design allows continuous refueling, although that increases another proliferation risk by making it easier to extract plutonium.

An older Soviet-era design is the graphite-moderated, water-cooled RBMK reactor. Often built to provide both electric power and plutonium for weapons, this design suffered from the safety defect that loss of coolant not only didn't shut down the chain reaction but could actually accelerate it due to loss of neutron-absorbing hydrogen in the H_2O coolant. The disastrous 1986 Chernobyl accident involved an RBMK reactor. During a test of the emergency cooling system, operators inadvertently put the reactor in an unstable state where an increase in power boiled away more cooling water, resulting in a further increase. The power level soared by a factor of 4000 in 5 seconds, causing a steam explosion that blew the top off the reactor and ignited the flammable graphite moderator. Heavy smoke carried radioactive materials into the atmosphere, resulting in widespread contamination, which we explored in Example 38.2. Today, thousands of square miles surrounding Chernobyl remain officially uninhabitable.

Other reactor designs include gas-cooled reactors that can operate at higher temperatures and therefore greater thermodynamic efficiencies, and **breeder reactors** designed specifically to "breed" plutonium from U-238 and therefore turn most of the nonfissile U-238 into fissile Pu-239. Breeders have no moderator, use liquid sodium coolant, and are critical with fast neutrons alone. Breeders are therefore less stable than so-called thermal reactors using slow neutrons, and widespread adoption of breeder technology entails international trafficking in fissile plutonium.

Nuclear Waste

We've seen that fission products are highly radioactive because they contain too many neutrons for stable middle-weight nuclei. Because of their high activity, fission products have relatively short half-lives, typically measured in decades. That makes fission-product waste dangerous for centuries to a few millennia. However, neutron absorption in fission reactors also produces plutonium and a host of other **transuranic** isotopes—those heavier than uranium—with much longer lifetimes. It's these substances that mean we'll have to safeguard nuclear waste for tens of thousands of years.

As fission proceeds, the concentration of fission products in the fuel increases. Before a reactor's ${}^{235}U$ is exhausted, fission products begin absorbing enough neutrons to interfere with the chain reaction. In U.S. LWRs, that requires about one-third of the fuel rods to be replaced annually. Older fuel is also rich in fissile plutonium, and at the end of a fuel rod's 3 years in the reactor, more than half the energy generation comes from fissioning plutonium rather than uranium. Figure 38.17 shows the evolution of nuclear fuel in a U.S. LWR.

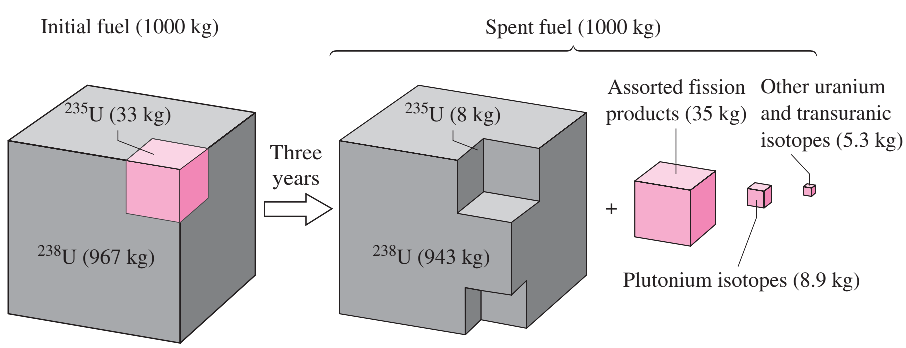


FIGURE 38.17 Evolution of 1000 kg of 3.3% enriched uranium over its 3-year stay in a light-water reactor.

The disposal of nuclear waste is a thorny issue, mixing political and scientific considerations. To date, the United States has no repository for commercial nuclear waste, which continues to accumulate at reactor sites. Lest you picture mountains of nuclear garbage, however, remember that factor-of-10^7 difference between nuclear and chemical energy sources. That translates into far less fuel needed for nuclear power plants, and far less waste produced. A 1-GW power reactor produces some 20 tons of high-level nuclear waste annually, while a comparable coal plant produces 1000 tons of carbon dioxide and 30 tons of solid waste *every hour*.

GOT IT? 38.4 Transportation and mining accidents involving coal are much more frequent than those involving uranium fuel. What's the fundamental reason for this?

Prospects for Nuclear Power

Today, nuclear power supplies some 15% of the world's electrical energy; in nuclear-intensive France the figure is nearly 80%, and in the United States it's 20%. Dozens of new reactors are under construction, predominantly in Asia; most use advanced versions of light-water reactor designs. Recently, the United States has seen nuclear-plant license applications for the first time in 30 years. But worldwide, hundreds of older reactors are nearing the end their lifetimes, and without massive new construction, nuclear's share of the world's energy supply is unlikely to increase significantly.

Concern over climate change from fossil-fuel combustion has spurred a renewed interest in nuclear power, even among some environmentalists. New reactor designs promise greater reliability, economic viability, and, most important, safety based on intrinsic reactor design rather than complicated safety systems. Most physicists agree that the public's concern over nuclear power is exaggerated. We regularly accept much greater risks from other technologies—for example, some 24,000 premature deaths each year in the United States due to pollution from coal-fired power plants. Comparable estimates for nuclear plants range from about 10 to fewer than 1000 from even the most vigorously antinuclear groups.

Nevertheless, ongoing uncertainties about the risk of catastrophic nuclear accidents, long-term waste storage, terrorism, and weapons proliferation continue to haunt the nuclear power industry. Proliferation, especially, is a very real concern. Although nuclear power and weapons development are different enterprises, they share infrastructure and an educated technological elite that can be put to either purpose. If nuclear power is to advance, it will need to do so under strict international guarantees against diversion of materials and expertise to weapons production.

38.5 Nuclear Fusion

The curve of binding energy (Fig. 38.9) shows that fusion of light nuclei provides another approach to nuclear energy production. The curve is steepest at its left end, indicating that the most energy per nucleon comes from the fusion of hydrogen. Indeed, the fusion reactions powering the Sun and many other stars begin with the fusion of hydrogen to form deuterium. Also emitted in the process are a positron, a neutrino, and 0.42 MeV of energy:

$$ {}^1_1\mathrm{H} + {}^1_1\mathrm{H} \longrightarrow {}^2_1\mathrm{H} + e^+ + \nu \qquad (0.42\ \mathrm{MeV}) \tag{38.9a}$$

Deuterium then fuses with hydrogen to form helium-3 and a gamma ray:

$$ {}^2_1\mathrm{H} + {}^1_1\mathrm{H} \longrightarrow {}^3_2\mathrm{He} + \gamma \qquad (5.49\ \mathrm{MeV}) \tag{38.9b}$$

Two helium-3 nuclei then react to form helium-4 and a pair of protons $\left({}^1_1\text{H}\right)$, releasing 12.86 MeV:

$$ {}^3_2\text{He} + {}^3_2\text{He} \longrightarrow {}^4_2\text{He} + 2\,{}^1_1\text{H} \qquad (12.86\text{ MeV}) \qquad (38.9\text{c}) $$

In addition, the positron from reaction 38.9a annihilates with an electron, forming two gamma rays with a total energy of $2mc^2$ or 1.022 MeV. Together, these reactions constitute the **proton–proton cycle**. In the full cycle, reactions 38.9a and b occur twice for each occurrence of reaction 33.9c. The net effect is to convert four protons and two electrons to a single He-4, releasing 26.7 MeV (Fig. 38.18). In massive stars, ${}^4_2\text{He}$ then becomes a building block for still heavier elements, as we discussed earlier.

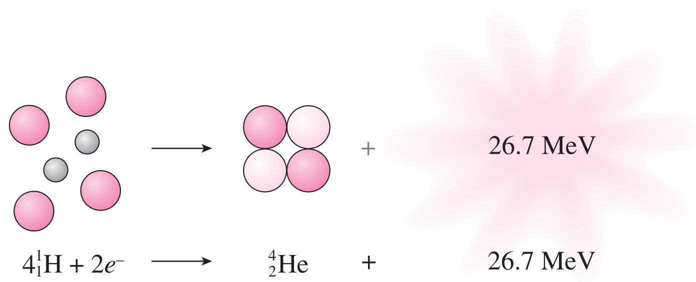


FIGURE 38.18 Net result of the proton–proton cycle of Equations 38.9.

Reaction 38.9a does not occur readily, and terrestrial fusion research has therefore focused on reactions involving the heavier hydrogen isotopes. Of immediate interest are deuterium–tritium (D-T) and deuterium–deuterium (D-D) reactions, listed below with the energy released in each:

$$ {}^2_1\text{H} + {}^3_1\text{H} \longrightarrow {}^4_2\text{He} + {}^1_0n \qquad (17.6\text{ MeV; D-T reaction}) \qquad (38.10\text{a}) $$
$$ {}^2_1\text{H} + {}^2_1\text{H} \longrightarrow {}^3_2\text{He} + {}^1_0n \qquad (3.27\text{ MeV; D-D reaction}) \qquad (38.10\text{b}) $$
$$ {}^2_1\text{H} + {}^2_1\text{H} \longrightarrow {}^3_1\text{H} + {}^1_1\text{H} \qquad (4.03\text{ MeV; D-D reaction}) \qquad (38.10\text{c}) $$

The two outcomes of the D-D reaction have nearly equal probability.

The electrical repulsion between nuclei makes it difficult to get them close enough to fuse. Although quantum tunneling helps, it still takes very high nuclear speeds—corresponding to high temperatures—to initiate fusion. At fusion temperatures, atoms are stripped of their electrons and the fusing material constitutes a plasma. It's necessary somehow to contain this hot plasma. Stars achieve both ends with their immense gravity, which compresses stellar material to fusion temperatures and simultaneously provides confinement. In the Sun's core, for example, the temperature is some 15 MK, and fusing nuclei approach with energies on the order of 1 keV—although even under these conditions the process isn't particularly efficient.

Terrestrial fusion requires still higher temperature, as high-energy particles undergo large accelerations that result in the plasma losing energy by radiation. The temperature at which fusion-generated power exceeds radiation loss is the **critical ignition temperature**. For the D-D reactions of Equations 38.10b and c, Fig. 38.19 shows that the ignition temperature is about 600 MK; for D-T it's a lower 50 MK. Net fusion-energy production requires not only high temperature but also confinement for long enough that the fusion energy produced exceeds the energy required to heat the plasma. The heat required depends on the number of nuclei or, on a volume basis, on the number density n. However, the rate of fusion-energy production depends on the *square* of the density. That's because doubling n doubles *both* the number of nuclei available to strike other nuclei and the number of nuclei available to be struck; the result is a quadrupling of the fusion rate. The total energy released therefore scales as $n^2\tau$, where τ is the **confinement time**. Meanwhile the

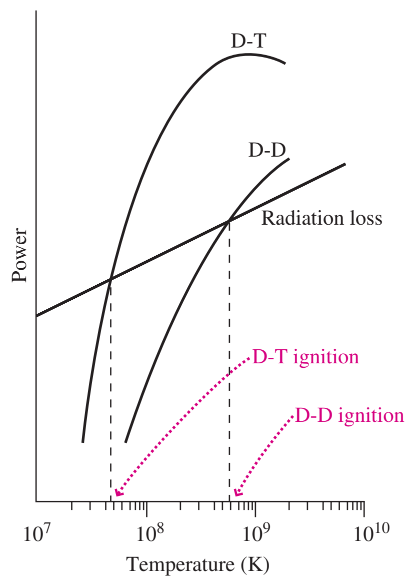


FIGURE 38.19 Power loss by radiation and power produced by D-D and D-T fusion reactions, as functions of temperature on a log-log plot.

radiation energy loss depends linearly on n, and as a result there's a minimum value of the product $n\tau$ necessary in an energy-producing fusion device. This condition is the **Lawson criterion**, given approximately by

$$\begin{aligned} n\tau &> 10^{22}\,\mathrm{s/m^3} \qquad (\text{Lawson criterion, D-D fusion}) \\ n\tau &> 10^{20}\,\mathrm{s/m^3} \qquad (\text{Lawson criterion, D-T fusion}) \end{aligned} \tag{38.11}$$

The factor-of-100 difference here shows that D-T fusion will be much easier to achieve.

Fusion technologies use two distinct approaches to the Lawson criterion. **Inertial confinement** strives for very high densities with short confinement times—so short that the particles' inertia alone is sufficient to prevent them from leaving the fusion site during the brief time needed. **Magnetic confinement** holds lower-density plasma in a "magnetic bottle" whose magnetic-field configuration minimizes the chance of escape during a relatively long confinement time. Neither approach has yet produced a sustained energy yield from fusion.

Inertial Confinement Fusion

Although peaceful fusion devices still elude us, inertial confinement has been used successfully since the 1950s in thermonuclear weapons—often called "hydrogen bombs" to distinguish them from fission explosives (incorrectly called "atomic bombs"). Thermonuclear weapons aren't pure fusion devices, though. They use a fission explosion to achieve the high temperatures needed to ignite fusion, and a clever arrangement for focusing the fission energy on a mixture of lithium deuteride and plutonium-239. The mixture is compressed to fusion temperatures, and fission neutrons convert lithium to helium and tritium. D-T fusion then occurs, providing the device with approximately half its explosive yield. The remainder comes from fission in an outer layer of natural uranium, whose nonfissile U-238 nevertheless fissions under the bombardment of high-energy neutrons. There's essentially no limit to the yield of a thermonuclear weapon, and devices as large as 58 megatons (Mt) TNT equivalent have been tested. That's 5000 times the energy of the fission bomb that destroyed Hiroshima. Today's missile-based thermonuclear weapons range from 40 kt to about 1 Mt, and there are still thousands of them in the world's arsenals.

Inertial confinement schemes for controlled fusion focus high-power laser beams on tiny deuterium–tritium targets, producing sequences of miniature thermonuclear explosions. Most advanced is the National Ignition Facility (NIF) at California's Lawrence Livermore National Laboratory, which began experiments in 2009. NIF uses 192 laser beams to focus a 1.8-MJ pulse of several nanoseconds duration on the millimeter-diameter D-T target (Fig. 38.20). During this brief interval the laser power—energy per time—is about 500 TW, or 1000 times the output of all U.S. electric power plants. Fusion energy is only one of NIF's three broad purposes; the others are to explore matter under extreme conditions, and to simulate nuclear weapons explosions without carrying out actual nuclear tests.

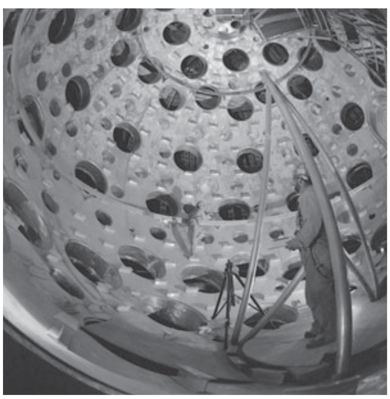


FIGURE 38.20 Target chamber of the National Ignition Facility is 11 m in diameter and weighs 130 tonnes. Holes are ports for the 192 laser beams that converge on the millimeter-size target.

Magnetic Confinement Fusion

In Chapter 26 we saw how charged particles in highly conducting plasma are essentially "frozen" to the magnetic field lines. Trapping of charged particles on magnetic field lines is the essence of magnetic confinement fusion schemes. The first job of magnetic confinement is to create a magnetic configuration that keeps plasma away from the relatively cool walls of the device. Plasma can escape to the walls in three general ways, as shown in Fig. 38.21.

The most promising magnetic fusion device is the **tokamak**, a Russian invention now used worldwide in fusion research. The tokamak has a toroidal configuration whose magnetic field lines never penetrate the walls, eliminating the end loss shown in Fig. 38.21*a*. Making the machine larger reduces field-line curvature, and with it the cross-field drift of Fig. 38.21*b*. Additional field components enhance confinement and suppress the

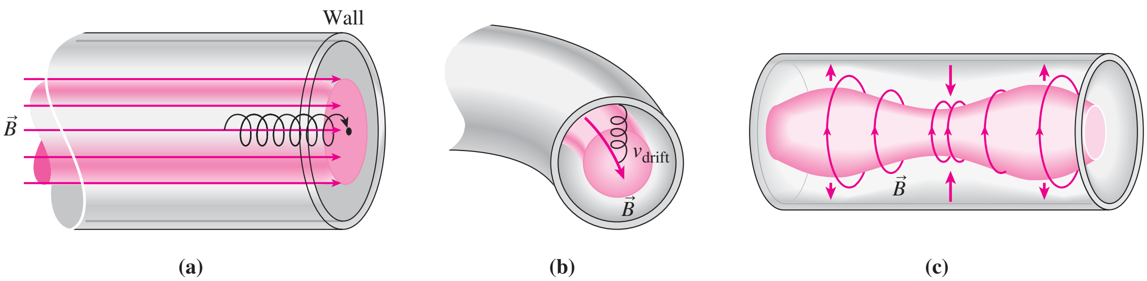


FIGURE 38.21 Plasma loss in magnetic confinement. (a) End losses occur when field lines intersect device walls. (b) Curvature of field lines results in cross-field drifts. (c) Instabilities distort the plasma and magnetic field. In (a) and (b) the spiral represents the path of a charged particle.

instabilities of Fig. 38.21*c*. After smaller tokamaks paved the way, an international consortium is constructing the ITER device in France for operation in about 2019 (Fig. 38.22). ITER is expected to be the first magnetic fusion system to produce net energy exceeding the energy used for plasma heating. ITER will operate at a plasma temperature higher than 100 MK and should generate 400 MW of fusion power from its 840 cubic meters of D-T plasma. ITER will use deuterium and lithium as fuel, with tritium $\left({}^3_1\text{H}\right)$ "bred" right in the reactor by neutron bombardment: ${}^6_3\text{Li} + {}^1_0n \longrightarrow {}^4_2\text{He} + {}^3_1\text{H}$.

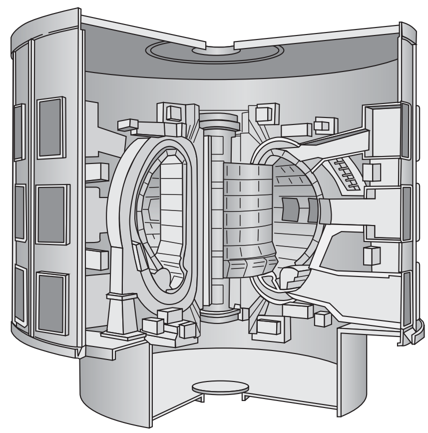


FIGURE 38.22 Cutaway diagram of the ITER fusion reactor. D-shaped structures are cross sections of the toroidal plasma chamber.

Prospects for Fusion Energy

When work on controlled fusion began in the 1950s, scientists confidently predicted that limitless fusion energy would be available in a few decades. More than half a century later, controlled fusion still appears decades away. But it's a goal worth pursuing: Problem 66 shows that with D-D fusion, a gallon of seawater is equivalent to some 300 gallons of gasoline, and Problem 67 shows that fusion energy resources could last far longer than the Sun will continue to shine!

Once controlled fusion proves scientifically feasible, there will be formidable engineering challenges in the design of a practical fusion power plant. Intense neutron fluxes from D-T fusion degrade materials that form the reaction chamber. Neutron-capture reactions produce radioactive isotopes within the walls, although the associated radioactivity and that of the tritium fuel are much less than the radioactivity of fission waste.

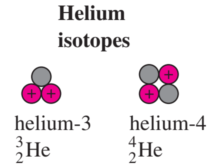
The first practical fusion plants will likely use D-T fusion because its ignition temperature and Lawson criterion are lower than for D-D fusion; the resulting energy will probably run a conventional steam cycle. But the D-D reaction promises cleaner and more efficient power production. Because D-D reaction 38.10c produces charged protons $\left({}^1_1\text{H}\right)$ rather than neutral neutrons, there's the possibility of using a magnetohydrodynamic (MHD) generator, in which electromagnetic induction converts charged-particle kinetic energy directly to electricity. MHD generators would bypass the conventional steam cycle and greatly increase the thermodynamic efficiency of a fusion power plant.

There's one caveat to this rosy fusion future: Although fusion itself is relatively clean and produces no greenhouse gases, the availability of unlimited cheap energy would likely spur industrial growth at a level our planet might not tolerate. And ultimately, if fusion energy use grew exponentially, the waste heat from fusion power could itself have climate-changing consequences.

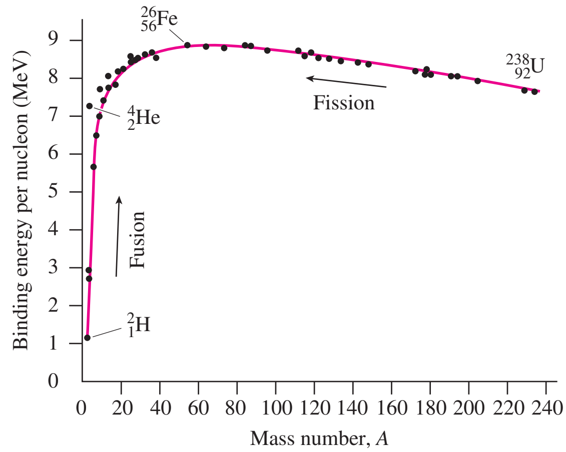
CHAPTER 38 SUMMARY

Big Picture

The big idea here is that the tiny but massive atomic nucleus is a repository of vast energy—on the order of 10^7 times the energy released in chemical reactions. The protons and neutrons that make up the nucleus can take many configurations, with the **atomic number**, Z, determining the element and the **mass number**, A, determining the particular **isotope**. We write ${}^{A}_{Z}X$ to describe the isotope with mass number A of the element whose atomic number is Z and whose symbol is X.

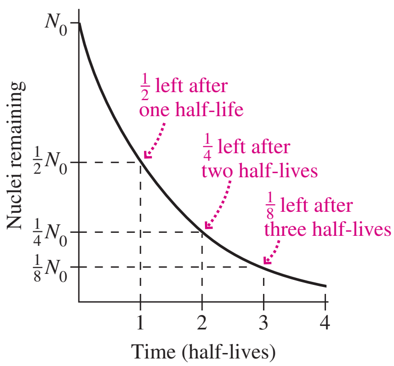


Stable isotopes require a delicate balance between protons and neutrons, with near equal numbers for lighter stable nuclei and more neutrons for heavier nuclei. **Unstable isotopes** are **radioactive** and decay by shedding particles. The **curve of binding energy** shows that energy can be released by either **fusion** of lighter nuclei or **fission** of heavier nuclei.



Key Concepts and Equations

Radioactive isotopes decay with a characteristic **half-life**, $t_{1/2}$: $N = N_0 2^{-t/t_{1/2}}$.



Alpha decay emits a helium-4 nucleus:

$$ {}^{A}_{Z}X \longrightarrow {}^{A-4}_{Z-2}Y + {}^{4}_{2}\mathrm{He} $$

Beta decay emits an electron or a positron and an antineutrino or a neutrino:

$$ {}^{A}_{Z}X \longrightarrow {}_{Z+1}^{A}Y + e^{-} + \bar{\nu} $$

Gamma decay emits a high-energy photon (gamma ray) as an excited nucleus drops to a lower energy state:

$$ X^{*} \longrightarrow X + \gamma $$

Applications

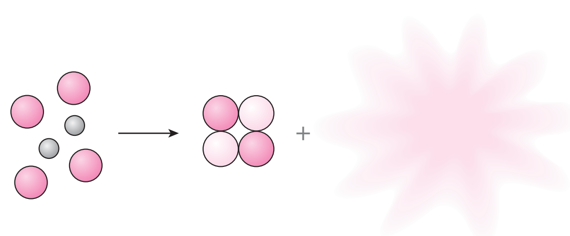
Radioactivity is measured in **becquerels**, with 1 Bq equal to one decay per second. **Sieverts** (Sv) measure the biological effects of radiation. Residents of the United States receive an average yearly radiation dose of about 3.6 mSv from both natural and artificial sources.

For fission, the most important isotopes are the **fissile** ${}^{235}_{92}\mathrm{U}$ and ${}^{239}_{94}\mathrm{Pu}$, which can fission when struck by low-energy neutrons:

$$ {}^{1}_{0}n + {}^{235}_{92}\mathrm{U} \longrightarrow X + Y + b\,{}^{1}_{0}n $$

The extra neutrons produced in fission can sustain a **chain reaction** provided there's a **critical mass** of fissile material. Exponentially growing chain reactions power fission weapons, while controlled fission occurs in **nuclear reactors** used for power generation.

Fusion powers the Sun and stars but has proved elusive on Earth except in thermonuclear weapons. **Inertial confinement** or **magnetic confinement** fusion may one day provide us with nearly limitless energy.



For Thought and Discussion

1. Why do nuclei contain neutrons?
2. Why are there no stable nuclei for sufficiently high atomic numbers?
3. Why might future archaeologists have problems dating samples from the second half of the 20th century?
4. Beta decay by positron emission is soon followed by a pair of 511-keV gamma rays. Why?
5. Why would it have been easier to make bombs fueled with uranium-235 a few billion years ago?
6. Why are iodine-131 and strontium-90 particularly dangerous radioisotopes?
7. Which model, liquid-drop or nuclear shell, does a better job explaining (a) nuclear fission and (b) gamma-ray spectra?
8. On an energy-release-per-unit-mass basis, by approximately what factor do nuclear reactions exceed chemical reactions?
9. Explain and distinguish the roles of the control rods and moderator in a nuclear reactor.
10. Why is a water-moderated reactor intrinsically safer in a loss-of-coolant accident than a graphite-moderated reactor?
11. Is ^{238}U fissionable? Is it fissile? Explain the distinction.
12. Why are fission fragments necessarily radioactive?

Exercises and Problems

Exercises

Section 38.1 Elements, Isotopes, and Nuclear Structure

13. Three radon isotopes have 125, 134, and 136 neutrons. Write the symbol for each.
14. Write the symbol for the germanium isotope with 44 neutrons.
15. How do (a) the number of nucleons and (b) the nuclear charge compare in the two nuclei $^{35}_{17}$Cl and $^{35}_{19}$K?
16. Compare the radius of the proton (the $A = 1$ nucleus) with the Bohr radius of the hydrogen atom.
17. A uranium-235 nucleus splits into two roughly equal-size fragments. Find their common radius.

Section 38.2 Radioactivity

18. How many half-lives will it take for the activity of a radioactive sample to diminish to 10% of its original level?
19. Copper-64 can decay by any of the three beta-decay processes. Write the equation for each decay.
20. Referring to Fig. 38.7, write equations describing the decays of (a) radon-222 and (b) lead-214.
21. A milk sample shows iodine-131 activity of 450 pCi/L. What's its activity in Bq/L?
22. BIO Carbon-11–labeled acetate shows promise in PET scans for determining the extent of metastasized prostate cancer. (a) Given C-11's 20.4-min half-life, how long will it take an initial dose of 2.0×10^9 Bq to decay to 7 kBq (roughly the natural radioactivity of the human body)? (b) What nucleus remains after C-11 decays by positron emission?
23. Nuclear bomb tests of the 1950s deposited a layer of strontium-90 over Earth's surface. How long will it take from the time of the bomb tests for (a) 99% and (b) 99.9% of this radioactive contaminant to decay?

Section 38.3 Binding Energy and Nucleosynthesis

24. Find the total binding energy of oxygen-16, given its nuclear mass of 15.9905 u.
25. Determine the nuclear mass of nickel-60, given that its binding energy is very nearly 8.8 MeV/nucleon.
26. Find the nuclear mass of plutonium-239, given its atomic mass of 239.052157 u.
27. The mass of a lithium-7 nucleus is 7.01435 u. Find the binding energy per nucleon.

Section 38.4 Nuclear Fission

28. A ^{235}U nucleus undergoes neutron-induced fission, yielding ^{141}Cs, three neutrons, and another nucleus. What's that nucleus?
29. Neutron-induced fission of ^{235}U yields fission products iodine-139 and yttrium-95. How many neutrons are released?
30. Write a complete equation for neutron-induced fission of plutonium-239 that yields barium-143, two neutrons, and another nucleus.
31. Assuming 200 MeV per fission, determine the number of fission events occurring each second in a reactor whose thermal power output is 3.2 GW.

Section 38.5 Nuclear Fusion

32. Verify from Equations 38.9 that the proton–proton cycle yields net energy of 26.7 MeV.
33. In a magnetic-confinement fusion device with confinement time 0.5 s, what density is required to meet the Lawson criterion for D-T fusion?
34. Inertial confinement schemes generally involve confinement times on the order of 0.1 ns. What's the corresponding density needed to meet the Lawson criterion for D-T fusion?
35. What confinement time is required for the D-T Lawson criterion in the ITER fusion reactor, given its plasma density of 10^{19} particles per cubic meter?

Problems

36. To what diameter would Earth have to collapse to be at nuclear density?
37. Find the energy needed to flip the spin state of a proton in Earth's magnetic field, whose magnitude is about 30 μT.
38. An NMR spectrometer is described as a "300-MHz instrument," meaning 300 MHz is the frequency supplied to its transmitter coil to flip the spin states of bare protons. What's the strength of its unperturbed magnetic field?
39. Iron-56, with nuclear mass 55.9206 u, is among the most tightly bound nuclei. Find the binding energy per nucleon, and check your answer against Fig. 38.9.
40. Find the atomic mass of iridium-193, whose binding energy is 7.94 MeV/nucleon.
41. As a geologist, you're assessing the feasibility of determining the ages of Earth's earliest rocks using radioactive dating. You estimate the number of half-lives that have passed for three different isotopes during Earth's 4.5-billion-year lifetime, and from that you determine the number of atoms remaining today from 10^6 atoms present at Earth's formation. The isotopes you consider are carbon-14, uranium-238, and potassium-40. What are your estimates, and which isotopes do you conclude are suitable for radioactive dating?
42. You measure the activity of a radioactive sample at 2.4 MBq. Thirty minutes later, the activity level is 1.9 MBq. Find the material's half-life.
43. You're a home inspector, and you find radon-222 activity of 23 pCi/L in the air inside a house, well above the EPA's "action"

limit of 4 pCi/L. If radon infiltration were stopped but there were no significant ventilation, how long would it take for the radon activity to drop below the action limit?

44. Nitrogen-13 is a 10-min-half-life isotope used to "tag" ammonia **BIO** for PET scans, including quantification of myocardial infarction. Consider an intravenous injection incorporating 20 mCi of N-13. Plot a graph of N-13 activity versus time, with your vertical axis logarithmic and your horizontal axis linear. Why is the graph a straight line? What's the significance of its slope?
45. Thorium-232 is an α emitter with 14-billion-year half-life. Radium-228 is a β^- emitter with 5.75-year half-life. Actinium-228 is a β^- emitter with 6.13-hour half-life. (a) What's the third daughter in the thorium-232 decay series? (b) Make a chart similar to Fig. 38.7 showing the first three decays in the thorium series.
46. How much cobalt-60 ($t_{1/2} = 5.24$ years) must be used to make a laboratory source whose activity will exceed 1 GBq for 2 years?
47. Archaeologists unearth a bone and find its carbon-14 content is 34% of that in a living bone. How old is the archaeological find?
48. Show that the decay constant and half-life are related by $t_{1/2} = \ln 2/\lambda \simeq 0.693/\lambda$.
49. The table below lists reported levels of iodine-131 contamination in milk in three countries affected by the 1986 Chernobyl accident, along with each country's safety guideline. Given I-131's half-life of 8.04 days, how long did each country have to wait for I-131 levels to decline to a level deemed safe by its standards?

	Level (Bq/L)	
Country	**Reported**	**Safety Guideline**
Poland	2000	1000
Austria	1500	370
Germany	1184	500

50. How many atoms are in a radioactive sample with activity 12 Bq and half-life 15 days?
51. Analysis of a Moon rock shows that 82% of its initial K-40 has decayed to Ar-40, a process with a half-life of 1.2×10^9 years. How old is the rock?
52. You're assessing the safety of an airport bomb-detection system in which neutron activation of the stable nitrogen isotope ${}^{15}_{7}\mathrm{N}$ turns it into unstable ${}^{16}_{7}\mathrm{N}$. The N-16 decays by beta emission with 7.13-s half-life. How long after activation will the N-16 activity have dropped by a factor of 1 million?
53. *Brachytherapy* is a cancer treatment involving implantation of ra- **BIO** dioactive "seeds" at the tumor site. Iridium-192, often used for cancers of the head and neck, undergoes beta decay by electron capture with 74.2-day half-life. Inner-shell electrons drop to the orbital occupied by the captured electron, resulting in emission of gamma rays that kill surrounding tumor cells. What percentage of initial Ir-192 activity will remain one year after implant?
54. Today, uranium-235 comprises only 0.72% of natural uranium; essentially all the rest is U-238. Use the half-lives in Table 38.1 to determine the percentage of uranium-235 when Earth formed about 4.5 billion years ago.
55. You're a geologist assessing underground sites for nuclear waste storage. A recent ruling by the U.S. Environmental Protection Agency suggests that waste-storage facilities should be designed for a million years of radiation protection. You're asked for the fraction of plutonium-239 initially in nuclear waste that would remain after that time. Your answer?
56. Oxygen-15 $(t_{1/2} = 2.0\ \mathrm{min})$ is produced in a hospital's cyclotron. **BIO** What should the initial activity concentration be if it takes 3.5 min to get the O-15 to a patient undergoing a PET scan requiring 0.5 mCi/L of activity?
57. How much ${}^{235}\mathrm{U}$ would be needed to fuel the reactor of Exercise 31 for 1 year? (*Note:* Your answer is an overestimate because fission of ${}^{239}\mathrm{Pu}$ also contributes to the power output.)
58. How much uranium-235 would be consumed in a fission bomb with a 20-kt explosive yield?
59. A neutron collides elastically head-on with a stationary deuteron in a reactor moderated by heavy water. How much of its kinetic energy is transferred to the deuteron? (*Hint:* Consult Chapter 9.)
60. A buildup of fission products "poisons" a reactor, dropping the multiplication factor to 0.992. How long will it take the reactor power to decrease by half, assuming a generation time of 0.10 s?
61. The total thermal power generated in a nuclear power reactor is 1.5 GW. How much U-235 does it consume in a year?
62. New Hampshire's Seabrook nuclear power plant produces electrical energy at the rate of 1.2 GW and consumes 1311 kg of U-235 each year. Assuming the plant operates continuously, find (a) its thermal power output and (b) its efficiency.
63. In the dangerous situation of prompt criticality in a fission reactor, the generation time drops to 100 μs as prompt neutrons alone sustain the chain reaction. If a reactor goes prompt critical with $k = 1.001$, how long does it take for a 100-fold increase in reactor power?
64. How much heavy water (deuterium oxide, ${}^{2}\mathrm{H_2O}$ or $\mathrm{D_2O}$) would be needed to fuel a 1000-MW D-D fusion power plant for 1 year?
65. The proton–proton cycle consumes four protons while producing about 27 MeV of energy. (a) At what rate must the Sun consume protons to produce its power output of about 4×10^{26} W? (b) The present phase of the Sun's life will end when it has consumed about 10% of its original protons. Estimate how long this phase will last, assuming the Sun's 2×10^{30}-kg mass was initially 71% hydrogen.
66. You're enthusiastic about fusion energy, and you want to convince others of the enormous fuel resource represented by the 0.015% of hydrogen nuclei that are actually deuterium. Using an average of 7.2 MeV per deuteron, you calculate the energy that would be released if all the deuterium in a gallon of seawater underwent fusion, and you compare your result with the energy in a gallon of gasoline (see Appendix C). What do you find for the gasoline equivalent of a gallon of seawater?
67. In a further effort to convince others of the benefits of fusion energy, you use the data from Problem 66 to estimate how long the deuterium in the world's oceans (average depth 3 km) could supply humanity's energy needs at the current consumption rate of about 15 TW. You then compare this with the Sun's remaining lifetime, about 5 billion years. What do you find?
68. The atomic masses of uranium-238 and thorium-234 are 238.050784 u and 234.043593 u, respectively. Find the energy released in the alpha decay of U-238.
69. Bismuth-209 and chromium-54 combine to form a heavy nucleus plus a neutron. Identify the heavy nucleus.
70. It's possible, though difficult, to realize alchemists' dreams of synthesizing gold. One reaction involves bombarding mercury-198 with neutrons to produce, for each neutron captured, a gold-197 nucleus and another particle. Write the equation for this reaction.
71. Nickel-65 beta decays by electron emission with decay constant $\lambda = 0.275\ \mathrm{h}^{-1}$. (a) Identify the daughter nucleus. (b) In a sample of initially pure Ni-65, find the time when there are twice as many daughter nuclei as parents.

72. The dominant naturally occurring radioisotopes in the typical human body include 16 mg of ^{40}K and 16 ng of ^{14}C. Using half-lives from Table 38.1, estimate the body's natural radioactivity. **BIO**
73. A laser-fusion fuel pellet has mass 5.0 mg and consists of equal parts (by mass) of deuterium and tritium. (a) If half the deuterons and an equal number of tritons participate in D-T fusion, how much energy is released? (b) At what rate must pellets be fused in a power plant with 3000-MW thermal power output? (c) What mass of fuel would be needed to run the plant for 1 year? Compare your answer with the 3.6×10^6 tons of coal needed to fuel a comparable coal-burning power plant.
74. Of the neutrons emitted in each fission event in a light-water reactor, an average of 0.6 neutron is absorbed by ^{238}U, leading to the formation of ^{239}Pu. (a) Assuming 200 MeV per fission, how much ^{239}Pu forms each year in a 30%-efficient nuclear plant whose electric power output is 1.0 GW? (b) With careful design, a fission explosive can be made from 5 kg of ^{239}Pu. How many potential bombs are produced each year in the power plant of part (a)?
75. In the liquid-drop model, the mass of a nucleus with mass number A can be expressed as a quadratic in Z: $M(A, Z) = c_1A - c_2Z + (c_2A^{-1} + c_3A^{-1/3})Z^2$, where the cs are constants determined from experimental data. Show that the value of Z that gives the minimum mass (not necessarily an integer) is $Z_{\text{min}} = (A/2)/[1 + (c_3/c_2)A^{2/3}]$. (*Note:* A plot of Z_{min} versus $N = A - Z$ gives the line of greatest nuclear stability in Figs. 38.3 and 38.12.)
76. The probability that a radioactive nucleus will have lifetime t is the probability that it will survive from time 0 to time t multiplied by the probability that it will decay in the interval from t to $t + dt$. Use this to show that the average lifetime of a nucleus is equal to the inverse of the decay constant in Equation 38.3a.
77. Nucleus A decays into B with decay constant λ_A and B decays into a stable product C with decay constant λ_B. A pure sample starts with N_0 nuclei A at $t = 0$. Find an expression for the total activity of the sample at time t.
78. (a) Example 38.6 explains that the number of fission events in a chain reaction increases by a factor k with each generation. Show that the total number of fission events in n generations is $N = (k^{n+1} - 1)/(k - 1)$. (b) In a typical nuclear explosive, k is about 1.5 and the generation time is about 10 ns. Use the result from (a) to calculate the time for all the nuclei in a 10-kg mass ^{235}U to fission. (*Hint*: Sum a series in part (a), and neglect 1 compared with N in part (b).)
79. A family member is about to have a brain scan using technetium-99*, an excited isotope with 6.01-hour half-life. The hospital makes Tc-99* from the decay of molybdenum-99 ($t_{1/2} = 2.7$ days), then delivers it to the nuclear medicine department. You're told that the Tc-99* will arrive 90 minutes after production, and that there must be 10 mg of it. The technician says she will produce 12 mg of Tc-99*. Is that sufficient? **BIO**

Passage Problems

In 1972, a worker at a nuclear fuel plant in France discovered that uranium from a mine in Oklo, in the African Republic of Gabon, had less U-235 than the normal 0.7%—a quantity known from meteorites and Moon rocks to be constant throughout the solar system. Further analysis showed the presence of isotopes that would result from the decay of fission products. Scientists drew the remarkable conclusion that a natural nuclear fission reaction had occurred some 2 billion years ago, lasting for about 100,000 years. Water, mixing with rich uranium ore, provided the moderator that made the chain reaction possible. More significantly, U-235's 700-million-year half-life means that 2 billion years ago there was a higher abundance of U-235 in natural uranium.

80. At the time of the Oklo fission reaction, the actual *amount* of U-235 present was
    a. about the same as today.
    b. about twice as much as today.
    c. about four times as much as today.
    d. about eight times as much as today.
81. Given U-238's 4.5-billion-year half-life, the percentage of U-235 in natural uranium 2 billion years ago was
    a. about 1%.
    b. about 4%.
    c. about 10%.
    d. nearly 100%.
82. The power output of the fission reactions at Oklo was 10 to 100 kW. If at some point that power had been sufficient to boil away the water at the reaction site, the chain reaction would have
    a. ceased.
    b. continued, but more slowly.
    c. been unaffected.
    d. sped up.
83. At the Oklo site today, you would expect to find measurable amounts of
    a. strontium-90.
    b. cesium-137.
    c. plutonium-239.
    d. none of the above.

Answers to Chapter Questions

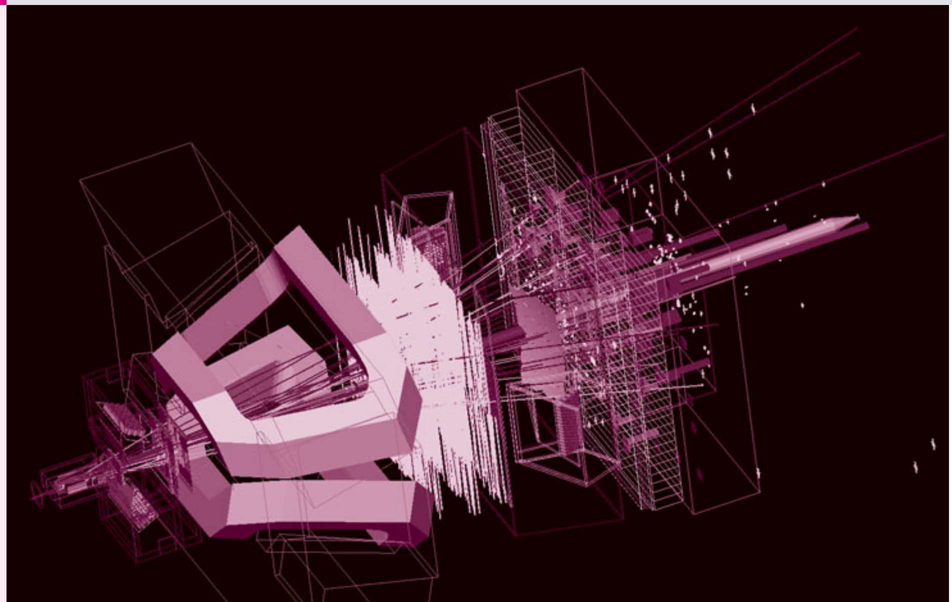
Answer to Chapter Opening Question

PET imaging "sees" the gamma rays produced in electron–positron annihilation, using multiple detectors to locate the annihilation site precisely. The positrons themselves are emitted by short-lived radioactive isotopes of elements such as carbon, oxygen, or fluorine. Because the lifetimes are so short (typically minutes), the isotopes must be made on site. Particles accelerated in the cyclotron slam into target materials, causing nuclear reactions that produce the desired isotopes.

Answers to GOT IT? Questions

38.1. (a) $Z = 6, N = 6$; (b) $Z = 8, N = 7$; (c) $Z = 26, N = 31$; (d) $Z = 94, N = 145$.
38.2. About one-billionth.
38.3. $^{57}_{26}Fe$, $^{132}_{54}Xe$, $^{238}_{92}U$, $^{4}_{2}He$, $^{2}_{1}H$.
38.4. There is a much higher concentration of energy in nuclear fuels (by a factor of around 10^7), which means far less nuclear fuel is mined and transported.

39 From Quarks to the Cosmos



A collision between two high-energy protons creates a spray of particles in the Large Hadron Collider. How do such experiments help us understand the origin of the universe?

New Concepts, New Skills

By the end of this chapter you should be able to

- Tell how particles mediate forces in the quantum description of force (39.1).
- Evaluate conserved quantities in particle interactions (39.2).
- Describe the standard model of particles and fields (39.3).
- Determine the quark composition of baryons and mesons (39.3).
- Describe the current status of physicists' attempts to unify the fundamental forces (39.4).
- Use Hubble's law to find cosmic distances and times (39.5).
- Give evidence for the Big Bang, describing especially the importance of the cosmic microwave background radiation (39.5).
- Outline the overall history of the universe (39.5).

Connecting Your Knowledge

This chapter is necessarily more descriptive than earlier chapters and introduces a host of new ideas at the forefront of physics. It builds on a range of earlier concepts, including:

- The fundamental forces: the strong force, the electroweak force, and gravity (4.3)
- Photons (34.3)
- The uncertainty principle (34.6)
- Blackbody radiation (34.2)
- The Doppler effect (14.8)
- Gravitational potential energy (8.4)

The past five chapters have extended the realm of physics to the scales of atoms and molecules and then down to the atomic nucleus. Here we go further still, probing the structure of nucleons themselves and trying to make sense of the host of subatomic particles nature reveals. We'll be asking questions about the ultimate nature of matter at the smallest scales, but in the process we'll find a remarkable connection with questions of the largest scale—questions about the origin and ultimate fate of the entire universe.

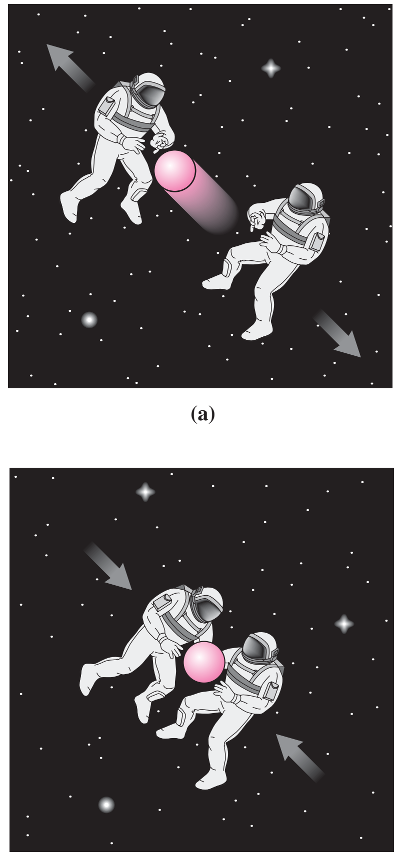
39.1 Particles and Forces

By 1932 four "elementary" particles of matter were known: the electron, the proton, the neutron, and the neutrino. In addition, there were the positron, antiparticle to the electron, and the photon of electromagnetic radiation. There were also the seemingly fundamental forces—gravity, the electromagnetic force, the nuclear force, and the weak force that manifests itself in beta decay.

In Chapter 34 we saw how the interaction of electromagnetic waves with matter ultimately involves individual photons—the quanta of the electromagnetic field. In the quantum-mechanical view of electromagnetism, the force between two charged particles also involves photons, now exchanged between the interacting particles. Imagine two astronauts tossing a ball back and forth (Fig. 39.1*a*). Catching or throwing the ball, one astronaut gains momentum in a direction away from the other, so the exchange results in a net average repulsive force. If the two astronauts struggle for possession of the ball, then the ball mediates what appears as an attractive interaction (Fig. 39.1*b*). Figure 39.1 gives classical analogs for the attractive and repulsive electrical interactions involving photon exchange.

We know that photons are emitted when a particle jumps into a lower energy state, with the photon carrying off energy equal to the energy difference between the two states. The process obviously conserves energy. But now we're saying that a single, free electron emits photons that it exchanges with another particle to produce what we call the electromagnetic force. How can that process conserve energy? The energy–time uncertainty relation (Equation 34.16) says that an energy measured in a time Δt is necessarily uncertain by an amount $\Delta E \geq \hbar/\Delta t$. The photon exchanged by two particles lasts only a short time, and therefore its energy is uncertain. So we can't really say that energy conservation is violated. A photon created in this way and lasting for only the short time it takes to exchange with another particle is called a **virtual photon**. We never "see" the virtual photon, since it's emitted by one particle and absorbed by the other.

The quantum theory of the electromagnetic interaction is called **quantum electrodynamics** (QED). Although begun by Paul Dirac, it was brought to consistent form in 1948 by Richard Feynman, Sin-Itiro Tomonaga, and Julian Schwinger. The fundamental event in QED is the interaction of a photon with an electrically charged particle. Two such events joined by a common virtual photon give the quantum electrodynamical description of the electromagnetic force (Fig. 39.2). The predictions of quantum electrodynamics have been confirmed experimentally to a remarkably high precision, and today QED is our best-verified theory of physical reality.



(a)

(b)

FIGURE 39.1 Analogs for particle-mediated forces: (a) repulsive and (b) attractive. Ball represents a photon.

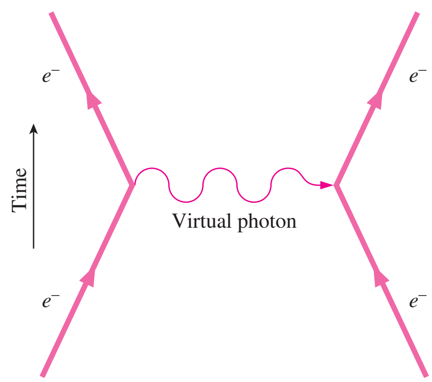


FIGURE 39.2 A Feynman diagram, showing the interaction of two electrons through the exchange of a virtual photon. The diagram provides a quantum description of the electrons' Coulomb repulsion.

Mesons

In 1935 the Japanese physicist Hideki Yukawa proposed that the nuclear force should, like the electromagnetic force, be mediated by exchange of a particle. Yukawa called his hypothetical particle a **meson**. Because the range of the nuclear force is limited, Yukawa argued, the meson should have nonzero mass. The reason for this connection between mass and range again lies in the energy–time uncertainty relation.

The electromagnetic force falls off as $1/r^2$ and thus has an infinite range. Two particles can be very far apart and still interact electromagnetically. Since photons travel at the finite speed of light, the time Δt for a photon interaction can be arbitrarily long. The energy–time uncertainty relation $\Delta E\,\Delta t \geq \hbar$ thus shows that the energy uncertainty ΔE can be arbitrarily small. Thus, the possible energies for virtual photons must extend downward toward zero—and that can happen only if photons have zero rest energy.

The nuclear force, in contrast, has a finite range of about 1.5 fm. At close to the speed of light, the longest a particle mediating this force would need to exist is a time Δt given by

$$\Delta t = \frac{\Delta x}{c} = \frac{1.5\times10^{-15}\ \text{m}}{3.0\times10^{8}\ \text{m/s}} = 5.0\times10^{-24}\ \text{s}$$

Then energy–time uncertainty gives

$$\Delta E \geq \frac{\hbar}{\Delta t} = \frac{1.05\times10^{-34}\ \text{J}\cdot\text{s}}{5.0\times10^{-24}\ \text{s}} = 2.1\times10^{-11}\ \text{J} = 130\ \text{MeV}$$

Yukawa therefore proposed a new particle with mass 130 MeV/c^2, about 250 times that of the electron. Yukawa's prediction was eventually confirmed—but not before physicists found yet another particle.

39.2 Particles and More Particles

In the 1930s the most available source of high-energy particles was cosmic radiation—high-energy protons and other particles of extraterrestrial origin. In 1937 the American physicist Carl Anderson (who had earlier discovered the positron) and his colleagues identified in cosmic rays a particle with a mass 207 times that of the electron. Now called the **muon**, this particle had the same charge and spin as the electron and seemed to behave much like a heavier version of the electron. Two muons were found: the negatively charged μ^- and its antiparticle μ^+. Although the muon mass was close to Yukawa's prediction, the muon interacted only weakly with nuclei and, therefore, could not be the mediator of the nuclear force.

The real Yukawa particle was discovered 10 years later in 1947, again in cosmic rays, and turned out to have a mass about 270 times that of the electron. This time there were three related particles, now called **pions**: positive π^+, negative π^-, and neutral π^0.

The new particles are all unstable, undergoing decays that ultimately result in well-known stable particles. The negative pion, for example, decays with a mean lifetime of 26 ns to a negative muon and an antineutrino:

$$\pi^- \longrightarrow \mu^- + \bar{\nu}$$

The muon then decays with a 2.2-μs lifetime to an electron and a neutrino–antineutrino pair:

$$\mu^- \longrightarrow e^- + \nu + \bar{\nu}$$

APPLICATION **Particle Detection**

Despite their small size, we can, remarkably, follow the trajectories of individual subatomic particles. Early particle detectors included the **cloud chamber** and the **bubble chamber**, in which particles ionize vapor or liquid, causing visible condensation or bubble formation along the particle tracks. More recent is the **multiwire proportional chamber**, in which crisscrossed wire grids record current pulses from electrons liberated as particles pass through a gas-filled chamber (see the photo, which shows a multiwire chamber at the Stanford Linear Accelerator Center). Analyzing the pulse distribution then reveals particle trajectories. Applying a magnetic field in a particle detector curves the trajectories, enabling scientists to determine the particles' charge-to-mass ratios. **Scintillation detectors** give off light flashes as particles pass through them, and the flash intensity provides a measure of particle energy. **Calorimeters**, consisting of layers of scintillators and energy-absorbing material, analyze the showers of secondary particles produced by a single high-energy particle to determine the original particle's energy. Modern detectors are huge agglomerations of several basic detector types, arranged to extract the maximum information from particle interactions. Computer analysis of detector output allows the identification of events so rare they may occur only once in a million interactions.

Classifying Particles

The availability of increasingly powerful particle accelerators led to an upsurge in particle discoveries. By 1980 there were more than 100 "elementary" particles. Early attempts to classify particles distinguished them by mass, but a more enlightening approach is based on the fundamental forces. We'll now use this approach to outline three particle classes.

Leptons are particles that don't experience the strong force. They include the familiar electron, the muon, a more massive particle called the tau, and three types of neutrinos, one associated with each of the charged leptons. The neutrinos were long thought to be massless, but recent experiments show that neutrinos have small nonzero mass and that they "oscillate" among the three types. Each of the leptons has an antiparticle as well. There are thus a total of six lepton–antilepton pairs, and experimental evidence strongly suggests that no others can exist. The leptons all have spin $\frac{1}{2}$ and are therefore **fermions**, which obey the Pauli exclusion principle. Leptons are believed to be true elementary point particles with zero size and no internal structure.

Hadrons are particles that do experience the strong force. They fall into two subclasses: mesons and baryons. **Mesons** have integer spin and are therefore **bosons** that don't obey the exclusion principle. Mesons include Yukawa's pions and a host of others; all are unstable. **Baryons** have half-integer spins and are therefore fermions. They include the familiar proton and neutron and similar but more massive particles. The baryons are grouped into pairs, triplets, and higher-multiple groupings of closely related particles. The neutron and proton, for example, form a pair that differ in charge and very slightly in mass. Each baryon has an antiparticle, as do most mesons, but some neutral mesons are their own antiparticles.

The third class of particles comprises the **field particles** or **gauge bosons**, quanta of the different force fields and "carriers" of those forces. These include the familiar photon for the electromagnetic force; three particles called the W^+, W^-, and Z for the weak force; a particle called the gluon; and a hypothetical graviton that would carry the gravitational force in an as yet incomplete theory of quantum gravity. All the field particles are bosons, carrying spin 1 or, for the graviton, spin 2. You might think Yukawa's meson should be in this category in its role as carrier of the nuclear force. That it doesn't appear here is a hint that the nuclear force isn't really fundamental; as we'll soon see, it's the gluon that plays the more fundamental role.

Table 39.1 lists some of the particles known even before full confirmation of today's elementary particle theories.

Table 39.1 Selected Particles*

Category/ Particle	Symbol (Particle/ Antiparticle)	Spin	Mass (MeV/c^2)	Baryon Number, B	Lepton Number, L	Strangeness, s	Lifetime (s)
Field particles							
Photon	γ, γ	1	0	0	0	0	Stable
Z^0	Z^0, Z^0	1	91,188	0	0	0	$\sim 10^{-25}$
Leptons							
Electron	e^-, e^+	$\frac{1}{2}$	0.511	0	+1	0	Stable
Muon	μ^-, μ^+	$\frac{1}{2}$	105.7	0	+1	0	2.2×10^{-6}
Tau	τ^-, τ^+	$\frac{1}{2}$	1777	0	+1	0	2.9×10^{-13}
Electron neutrino	$\nu_e, \overline{\nu}_e$	$\frac{1}{2}$	$<3\times10^{-6}$	0	+1	0	Stable
Muon neutrino	$\nu_\mu, \overline{\nu}_\mu$	$\frac{1}{2}$	<0.19	0	+1	0	Stable
Tau neutrino	$\nu_\tau, \overline{\nu}_\tau$	$\frac{1}{2}$	<18.2	0	+1	0	Stable
Hadrons							
Mesons							
Pion	π^+, π^-	0	139.6	0	0	0	2.6×10^{-8}
Pion	π^0, π^0	0	135.0	0	0	0	8.4×10^{-17}
Eta	η^0, η^0	0	547.8	0	0	0	$\sim 5\times10^{-19}$
Rho	ρ^0, ρ^0	1	775.8	0	0	0	$\sim 4\times10^{-24}$
Kaon	K^+, K^-	0	493.7	0	0	1	1.2×10^{-8}
Kaon	K^0, K^0	0	497.6	0	0	1	0.895×10^{-10} $5.18\times10^{-8\dagger}$
Baryons							
Proton	$p, \overline{p}$	$\frac{1}{2}$	938.3	1	0	0	Stable
Neutron	$n, \overline{n}$	$\frac{1}{2}$	939.6	1	0	0	885.7
Lambda	$\Lambda^0, \overline{\Lambda}^0$	$\frac{1}{2}$	1115.7	1	0	−1	2.6×10^{-10}
Sigma	$\Sigma^+, \overline{\Sigma}^-$	$\frac{1}{2}$	1189.4	1	0	−1	0.80×10^{-10}
Sigma	$\Sigma^0, \overline{\Sigma}^0$	$\frac{1}{2}$	1192.6	1	0	−1	7.4×10^{-20}
Sigma	$\Sigma^-, \overline{\Sigma}^+$	$\frac{1}{2}$	1197.4	1	0	−1	1.5×10^{-10}
Omega	$\Omega^-, \overline{\Omega}^+$	$\frac{3}{2}$	1672.45	1	0	−3	0.82×10^{-10}

*Does not include any hadrons with c, b, or t quarks.
†The neutral kaon exists as a quantum-mechanical superposition of states with two different lifetimes.

Particle Properties and Conservation Laws

Many new particles can be characterized by known properties such as mass, spin, and electric charge. Of these, spin and charge are associated with important conservation laws—conservation of angular momentum and conservation of electric charge. Allowed interactions among particles *must* conserve these quantities. The annihilation of an electron–positron pair, for example, is allowed because the initial particles have no net charge and neither do the resulting photons. Similarly, beta decay of the neutron produces an electron, a proton, and a neutral antineutrino and thus conserves charge:

$$n \longrightarrow p + e^- + \bar{\nu}_e$$

Here the subscript on the antineutrino indicates that it's an electron antineutrino, as opposed to the muon or tau variety.

Other particle properties appear to be conserved as well. Associated with each baryon or antibaryon is its **baryon number**, assigned the value $+1$ for a baryon and -1 for an antibaryon. All experimental evidence to date points to conservation of baryon number: In all observed particle reactions, the sums of the baryon numbers before and after the reaction are equal. An example is, again, beta decay of the neutron: The process starts with a neutron of baryon number $B = 1$ and ends with a proton $(B = 1)$, an electron, and an antineutrino. The last two are leptons, so their baryon numbers are zero, and thus baryon number is conserved. Some theories, which we describe shortly, suggest that baryon number is only approximately conserved. If that's so, then the proton itself is an unstable particle with a mean lifetime in excess of 10^{35} years.

Lepton number seems also to be conserved. Again, beta decay provides an example: The neutron and proton, being baryons, have lepton number zero, while the resulting electron and *anti*neutrino have lepton numbers $+1$ and -1, respectively.

In the late 1950s particles called K, Λ, Σ, and Ξ were discovered. Strange characteristics of these particles' decays could be explained by introducing a new fundamental property, called **strangeness**, with the new particles having $s = \pm 1$ or ± 2. Strangeness is conserved in strong and electromagnetic interactions, but in weak interactions its value can change. We'll soon see that several other new properties are needed to characterize matter fully.

CONCEPTUAL EXAMPLE 39.1 Conservation Laws: Evaluating a Particle Interaction

A pion collides with a proton to produce a neutral kaon and a lambda particle:

$$\pi^- + p \longrightarrow K^0 + \Lambda^0$$

(a) Which of the following are conserved: electric charge, baryon number, lepton number, strangeness? (b) Could another possible result of a pion–proton collision be an electron and a proton?

EVALUATE (a) Table 39.1 shows that all the particles are hadrons, so the lepton number is zero on both sides of the equation. On the left we have a positive and a negative particle and on the right two neutrals, so the electric charge is zero on both sides. The pion is a meson and the proton a baryon, so the baryon number on the left is 1; similarly, the kaon is a meson and the Λ^0 a baryon, so the baryon number is conserved. Finally, neither pion nor proton is strange, so the total strangeness on the left is zero. Table 39.1 lists the K^0 with $s = +1$ and the Λ^0 with $s = -1$, so strangeness, too, is conserved. (b) Having an electron and a proton as the final state would conserve charge (0), strangeness (0), and baryon number (1). But it wouldn't conserve lepton number, which was originally 0 but would become 1.

ASSESS This example shows how conservation laws restrict the possible outcomes of particle interactions.

MAKING THE CONNECTION What minimum kinetic energy is required for the pion and proton together in order for this reaction to occur?

EVALUATE Table 39.1 gives the rest masses of the pion as 139.6 MeV/c^2 and the proton as 938.3 MeV/c^2; therefore, their rest energies mc^2 are 139.6 MeV and 938.3 MeV, for a total of 1078 MeV. But the tabulated masses of the K^0 and Λ^0 show that their total rest energy is 1613 MeV, so we need an additional $1613 - 1078 = 535$ MeV of energy to drive the reaction. This is the minimum kinetic energy for the pion and proton together.

Symmetries

Watch a physical process in a mirror, and you expect the image to be a possible physical process; that is, the laws of physics should exhibit **symmetry** with respect to mirror reflection. At the subatomic level, the statement that a process and its mirror image are equally likely is called **conservation of parity**. Mathematically, a system has parity $+1$ if its wave function is unchanged on reflection through the origin—that is, on a coordinate change $x \to -x$, $y \to -y$, $z \to -z$. If the wave function changes sign, then the parity is -1. Parity is conserved if its value is unchanged in a particle interaction.

In 1957 theoretical physicists Tsung-dao Lee and Chen Ning Yang pointed out that parity conservation had not been tested for the weak force. They made the revolutionary suggestion that parity need not be conserved—tantamount to suggesting that nature can distinguish right- from left-handed systems that are otherwise identical. A group led by Chien-Shiung Wu took up the challenge. Wu studied the beta decay of cobalt-60 in a magnetic field that established a left–right symmetry. Her experiments showed a preferential beta emission opposite the field direction—a clear violation of parity conservation (Fig. 39.3).

Although parity might not be conserved, theorists held that a combination of parity reversal (P) and charge conjugation (C)—changing particles into antiparticles—would result in indistinguishable physical behavior. But in 1964 a violation of this CP conservation was found in a rare decay of the neutral kaon to a pion–antipion pair. The Russian physicist Andrei Sakharov suggested that this asymmetric decay might account for the preponderance of matter over antimatter in today's universe.

It still appears that CPT conservation holds; that is, a combination of mirror reflection, charge conjugation, and reversal of the time coordinate makes a new physical process indistinguishable from the original. There may be a deep philosophical connection here with the direction of time, but the full implications of CPT symmetry and the failure of its individual components aren't fully understood.

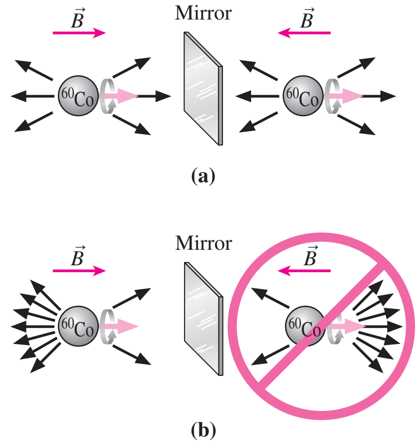


FIGURE 39.3 Experimental evidence for nonconservation of parity. At left of the mirror, a ^{60}Co nucleus has its spin aligned with a magnetic field. (a) Reflected in the mirror, the spin vector still points to the right, even though the magnetic field is reversed. If the mirror image were equally likely, beta emission (arrows) would occur with equal probability along and opposite the spin direction. (b) Experiment shows that beta emission occurs preferentially opposite the spin direction, so the mirror-image situation at the right in (b) does not occur.

39.3 Quarks and the Standard Model

The proliferation of particles distressed physicists used to finding an underlying simplicity in nature. Were all those particles really "elementary," or was there a more fundamental, simpler level? In 1961 physicists Murray Gell-Mann and Yuval Ne'eman independently noticed patterns in the then-known particles. They called their patterns the **Eightfold Way**, after Buddhist principles for right living. An empty spot in one pattern led Gell-Mann to predict a new particle with strangeness -3. Experimentalists soon found the particle, now known as the Ω^-, in bubble-chamber photographs from earlier experiments.

Quarks

Success of the Eightfold Way convinced physicists that many "elementary" particles weren't really elementary. In 1964 Gell-Mann and his colleague George Zweig independently proposed a set of three particles called **quarks** that combined to form the then-known hadrons. These became known as the **up quark**, the **down quark**, and the **strange quark**. For each there was a corresponding antiquark.

One surprising thing about quarks is that they carry fractional electric charges. The two least massive quarks, the up and the down, carry $+\frac{2}{3}e$ and $-\frac{1}{3}e$, respectively; their antiparticles have the opposite charges. The quarks combine in pairs or triplets to make the two classes of hadrons. Baryons, like the proton and the neutron, consist of three quarks (Fig. 39.4). Mesons contain quark–antiquark pairs (Fig. 39.5). The quarks all have spin $\frac{1}{2}$, which explains why the three-quark baryons all have odd half-integer spin, and why the two-quark mesons have integer spin.

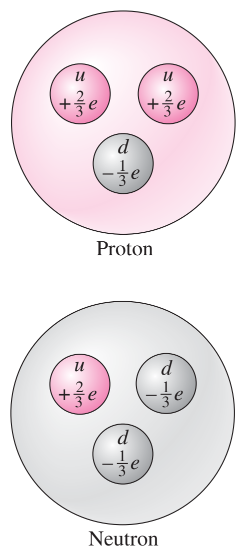


FIGURE 39.4 Protons and neutrons consist of the quark combinations *uud* and *udd*, respectively.

The Pauli exclusion principle precludes three particles having the same quantum numbers, so there must be an additional property distinguishing quarks. Called **color**, this property is a kind of "charge"—not to be confused with electric charge—that can take on any of three values called, whimsically, red, green, and blue. The force binding quarks of different colors is the **strong force**, and the quark theory is known as **quantum chromodynamics (QCD)**. In QCD, **gluons** are the field particles that play the role of photons in quantum electrodynamics, binding particles subject to the strong force. Particles formed from quarks—the mesons and the baryons—are always **colorless**. This is true of mesons because they contain a quark of one color and another of its anticolor. It's true of baryons because they contain three quarks of different colors, which combine to give the baryon

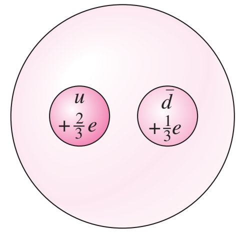


FIGURE 39.5 Mesons consist of a quark and an antiquark. The meson here is the π^+, made from an up quark and a down antiquark.

zero net color charge. The nuclear force, once thought to be fundamental, is actually a residual manifestation of the strong force, acting between the quarks in colorless particles—in much the same way that the van der Waals force between neutral gas molecules is a "residue" of the stronger electric force among the particles that make up the molecules.

Photons mediate the electromagnetic force between charged particles but are themselves uncharged. In contrast, gluons, like the quarks they bind, carry color charge. There are eight different gluons; six carry combinations like red–antiblue $(\mathrm{R}\overline{\mathrm{B}})$, green–antired $(\mathrm{G}\overline{\mathrm{R}})$, and so on; the other two are colorless. Exchange of a colored gluon, unlike photon exchange in quantum electrodynamics, thus changes the colors of the particles involved.

Another surprising aspect of quarks is that the strong force doesn't decrease with separation. For that reason it appears impossible to isolate a single quark (Fig. 39.6). As a result we never see individual free particles with fractional electric charge.

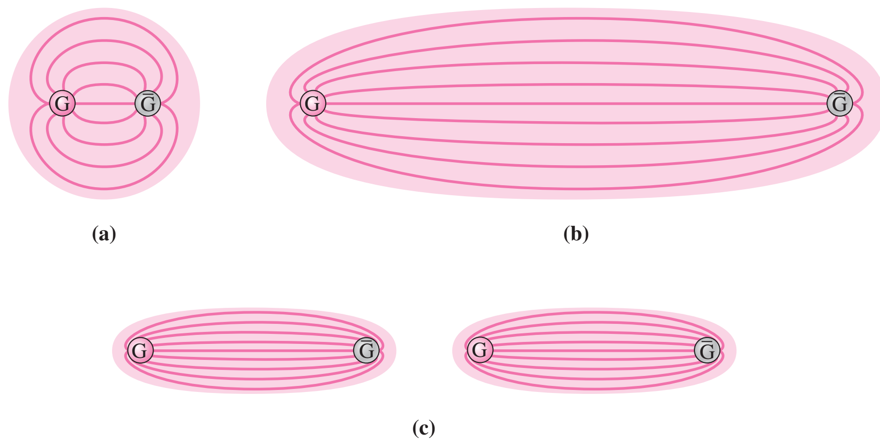


FIGURE 39.6 Quark confinement in a meson consisting of a green quark G and its antiquark $\overline{\mathrm{G}}$. (a) Field lines represent the "color field" that joins the two. (b) The field remains confined as the quarks are moved apart, so the field strength stays essentially constant. (c) Pulling the quarks far apart builds up enough energy to create another quark–antiquark pair, rather than isolating individual parts.

The up, down, and strange quarks soon proved insufficient to account for all the observed particles. Theorist Sheldon Glashow argued for a fourth quark, called the **charmed quark**. Ten years later, following intensive searches, experimental teams at Brookhaven National Laboratory and the Stanford Linear Accelerator Center announced the discovery of a particle that implies the existence of the charmed quark. The charmed and strange quarks form a related pair, similar to the up/down quark pair.

There's still one more quark pair. A 1977 experiment confirmed the existence of the **bottom quark**, and in 1995 Fermilab announced the discovery of the **top quark**. The more exotic quarks are more massive, and therefore, through mass–energy equivalence, it takes more energy to produce particles containing them. This need for higher energy is what drives the push for ever more powerful and expensive particle accelerators. (See Application, p. 715.) Table 39.2 lists some properties of the six quarks.

Table 39.2 Matter Particles of the Standard Model

Quark Name	Symbol	Approximate Mass (MeV/c^2)	Charge	Corresponding Leptons (Symbol, Mass in MeV/c^2)
Down	d	6	$-\frac{1}{3}e$	Electron $(e, 0.511)$, electron neutrino (ν_e)
Up	u	3	$+\frac{2}{3}e$	
Strange	s	100	$-\frac{1}{3}e$	Muon $(\mu, 106)$, muon neutrino (ν_μ)
Charmed	c	1300	$+\frac{2}{3}e$	
Bottom	b	4300	$-\frac{1}{3}e$	Tau $(\tau, 1777)$, tau neutrino (ν_τ)
Top	t	1.7×10^5	$+\frac{2}{3}e$	

EXAMPLE 39.1 **Quarks: Particle Composition and Properties**

Given that the strange quark has strangeness $s = -1$, find the charge and strangeness of the Λ^0 particle, which has quark composition *uds*.

INTERPRET We're asked to find how the charge and strangeness of three individual quarks combine in a composite particle, the Λ^0.

DEVELOP Our plan is to sum the values of the quark's charge and strangeness to get the parameters of the Λ^0.

EVALUATE Table 39.2 gives the charges of the u, d, and s quarks as $+\frac{2}{3}e$, $-\frac{1}{3}e$, and $-\frac{1}{3}e$, respectively. These sum to zero net charge, so the Λ^0 is electrically neutral. The up and down quarks aren't strange, so they have $s = 0$. With $s = -1$ for the strange quark, the Λ^0 must have strangeness -1.

ASSESS Table 39.1 shows we're right about the strangeness of the Λ^0, and its superscript 0 implies that this is a neutral particle, as we've found. ■

The Standard Model

We now have six **flavors** of quarks—up, down, strange, charmed, top, bottom—that seem to be truly elementary constituents of matter. Quarks join to form the hadrons—baryons and mesons. But other particles—namely, leptons and field particles—aren't made from quarks. They, like quarks, seem to be truly indivisible, elementary particles.

In this "zoo" of elementary particles, physicists recognize three distinct "families." The up and the down quarks make the neutron and proton; together with the electron and its related neutrino, they account for the properties of ordinary matter. A second family consists of the strange and charmed quarks, the electron-like muon, and the muon neutrino. The quarks of this family are more massive than the up and down quarks, and the muon is more massive than the electron. More massive still are the particles of the third family, consisting of the top and bottom quarks, the electron-like tau particle, and the tau neutrino. Table 39.2 shows these three families, from which all known matter is constructed.

You may be expecting that future editions of this book will tell of additional quarks, and thus of additional families of matter. Whether such additional families exist was an open question until physicists at the Large Electron Positron Collider in Geneva, Switzerland, examined more than half a million particle-decay events and concluded that the number of different types of neutrinos that can exist is 2.99 ± 0.06. Since there's presumably a neutrino type for each family, this result seems to preclude the existence of additional families.

The theory that currently describes elementary particles and their interactions is called the **standard model**. In addition to the particles shown in Table 39.2, the standard model includes the photons that mediate the electromagnetic interaction, the gluons of the strong force, the W and Z particles that mediate the weak force, and an as yet undetected particle called the Higgs boson, believed responsible for other particles' masses (Table 39.3). The standard model is successful in explaining the phenomena of particle physics, but it leaves many fundamental questions unanswered. Why, for example, do the quarks and leptons

Table 39.3 Field Particles and Forces

Particle	Mass (GeV/c^2)	Electric and Color Charges*	Force Mediated	Range	Approximate Strength at 1 fm (Relative to Strong Force)
Graviton	0	0, 0	Gravity	Infinite	10^{-38}
$W^{\pm}$	80.2	± 1, 0	Weak	$<2.4\times 10^{-18}$ m	10^{-13}
Z^0	91.2	0, 0			
Photon, γ	0	0, 0	Electromagnetic	Infinite	10^{-2}
Gluon, g (8 varieties)	0	0, 6 color–anticolor combinations, 2 colorless	Strong	Infinite†	1
Higgs boson, H^0	114–185?	0, 0	The Higgs boson is an as yet undetected particle needed to account for the masses of the other particles.		

*Color is a quark property analogous to, but more complicated than, electric charge.

†The nuclear force is the residual strong force between colorless particles and has a range of about 1 fm (10^{-15} m).

have the particular masses they do? Why are there only three families of elementary particles? Why are leptons and quarks distinct? Are these particles really elementary, or are there even smaller structures at hitherto unexplored scales? Continuing theoretical work and experiments at ever-higher energies may someday answer these questions.

39.4 Unification

We introduced the fundamental forces of nature in Chapter 4: gravity, the electroweak force, and the strong force. It was not always that simple, though. In Chapters 20–29 we studied the electric and magnetic forces, first separately but then with the realization that they fall under the single umbrella of electromagnetism. The unification of electricity and magnetism was a major step forward in our understanding of physical reality. Physicists continue to strive for further unification, with the ultimate hope that someday all the forces will be understood as a manifestation of a single common interaction.

Electroweak Unification

FIGURE 39.7 Particle tracks from the decay of a *Z* particle help confirm electroweak unification.

In the 1960s and early 1970s, a century after Maxwell formalized the unification of electromagnetism, physicists Steven Weinberg, Abdus Salam, and Sheldon Glashow proposed that the electromagnetic force and the weak force are really aspects of the same thing. Their theory predicted the existence of the particles W^+, Z^0, and W^-, the "carriers" of the unified electroweak force. In 1983 a huge international consortium headed by Carlo Rubbia discovered the W and Z particles, using advances in accelerator technology developed by Simon van der Meer (Fig. 39.7). That discovery confirmed the electroweak unification, and Rubbia and van der Meer joined a long list of physicists who had won the Nobel Prize for contributions to our understanding of the structure of matter.

Further Unification

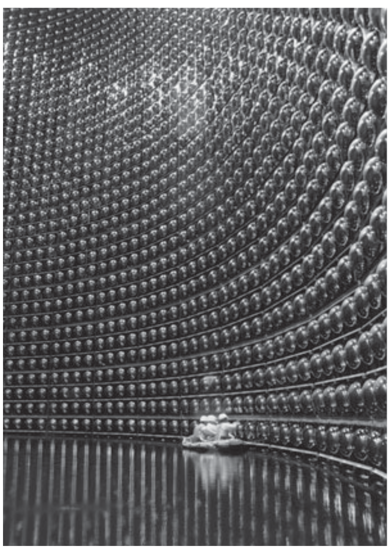


FIGURE 39.8 Japan's Super Kamiokande experiment consists of 50,000 tons of pure water in an underground chamber, surrounded by some 10,000 photomultiplier tubes to detect flashes from rare nuclear reactions including neutrino interactions and hypothetical proton decays.

Electroweak unification led to the present situation in physics, with the strong force, the electroweak force, and gravity comprising the **fundamental forces** that describe all interactions of matter. A further step, the **grand unification theories** (GUTs), attempts to merge the electroweak and strong forces. Some versions of GUT predict that the proton should decay on the very long timescale of some 10^{36} years. We can't wait that long, but we can put 10^{34} protons together in the form of tens of thousands of tons of water and watch for proton decay (Fig. 39.8). Such experiments have not found the predicted decays, but another GUT prediction—that neutrinos have mass—was verified, also in the device of Fig. 39.8. Many physicists believe that some form of grand unification will soon be achieved.

Even grand unification would still leave two forces, one of them gravity. Attempts to reconcile our current theory of gravity—Einstein's general theory of relativity—with quantum mechanics have so far made little progress. Yet such a reconciliation is a necessary prerequisite for a final unification of all known forces. A possible candidate is **string theory**, which pictures elementary particles as vibration modes on stringlike structures that may be as short as 10^{-35} m (Fig. 39.9). String theory is set not in the four-dimensional spacetime to which we're accustomed, but in a spacetime with 10 or more dimensions. The

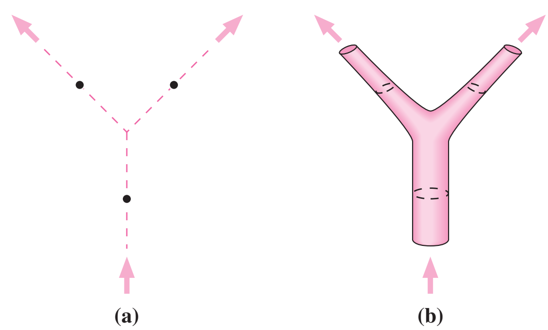


FIGURE 39.9 (a) A particle decay involving point particles in the standard model. (b) A similar decay in string theory. Note that there's no single point in spacetime at which the decay takes place.

extra dimensions are "compactified" in a way that makes them undetectable in normal interactions. To some physicists, string theories hold the promise of a "theory of everything," explaining all our observations about the behavior of the universe. To others they're another in a long line of unsuccessful attempts at a comprehensive explanation of physical phenomena. Only further research will tell.

Symmetry Breaking

Unification theories predict that phenomena that appear distinct under normal conditions will be seen as one at sufficiently high energies. The observed unification represents a kind of symmetry that's "broken" as the energy level drops. Figure 39.10 shows a mechanical analogy for such **symmetry breaking**. With high energy, a ball sits atop a potential "hill," and the situation is symmetric. But when the ball drops to a low-energy state, it ends up at a particular angular position, and the symmetry is broken. Analogously, at energies above 100 GeV, what we call the electromagnetic and weak forces are one and the same. But at lower energies the symmetry is broken, and we see two distinct forces. Particle accelerators now being planned will exceed the energy of electroweak symmetry breaking, allowing us to explore that interaction in its fundamental simplicity. But the energy at which symmetry breaking occurs increases to some 10^{15} GeV as we move from electroweak to grand unification, making it unlikely that we'll achieve that energy in the foreseeable future. And the energy at which gravity, too, would join a single unified force is an even more remote 10^{19} GeV.

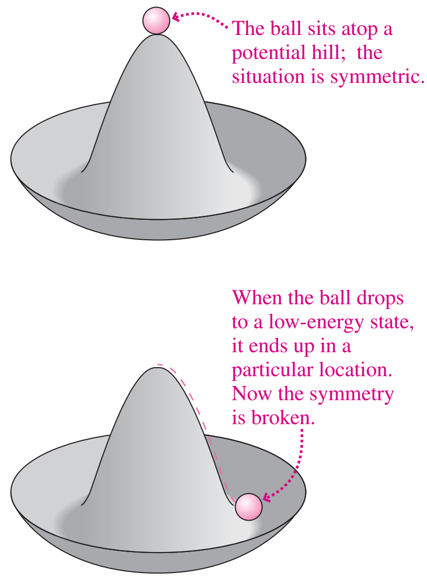
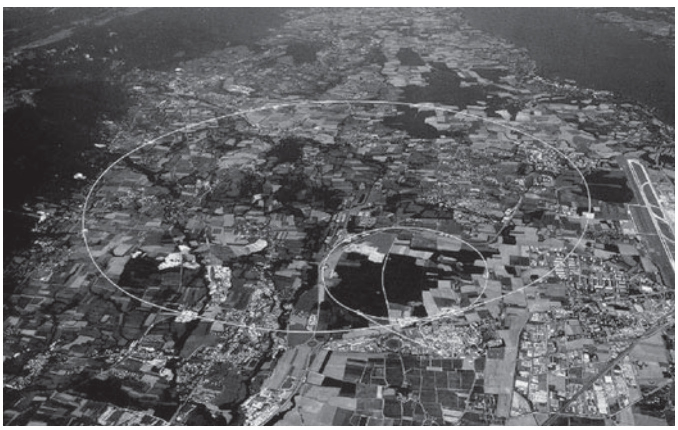


FIGURE 39.10 A mechanical analogy for symmetry breaking, showing a ball subject to a hat-shaped potential-energy curve.

APPLICATION **Particle Accelerators**



Aerial view showing the location of the 4.3-km-diameter Large Hadron Collider on the Swiss–French border. The accelerator itself is in a tunnel 50–175 m below ground. The device accelerates protons to 7 TeV. The smaller ring accelerates the protons to 450 GeV before they're injected into the main accelerator.

Most particles are more massive than the proton, and some, like the weak-force mediators $W^{\pm}$ and Z, are extremely massive. Since the more massive particles are all unstable, discovering them involves first creating them—and that requires energy of at least mc^2, with m the particle mass. That energy requirement, along with the hint of new phenomena such as force unification, drives particle physicists' seemingly insatiable desire for particle accelerators of ever-higher energies.

The earliest accelerators were electrostatic devices that established large potential differences between conducting electrodes, and they used the associated electric field to accelerate charged particles. But such accelerators are limited to maximum energies of about 20 MeV because of the difficulties of handling high voltages. We saw in Chapter 26 how this problem is cleverly circumvented in the cyclotron, an accelerator that uses a magnetic field to keep particles in circular orbits so they can gain energy on each orbit from a modest electric field. But cyclotrons work for only nonrelativistic particles, for which the cyclotron frequency is independent of particle energy. Today's high-energy experiments call for ultrarelativistic particles, whose speeds differ only minutely from the speed of light. As a result, today's accelerators are primarily variations on the **synchrotron**, a device in which the magnetic field increases with the particle energy to maintain particles in a circular orbit of fixed radius. An alternative to the synchrotron is the **linear accelerator**, the largest of which is the 3-km-long Stanford Linear Accelerator.

A head-on collision between two cars is much more damaging than a collision of a moving car with a stationary one, since in the former case all the energy goes into damaging the cars while in the latter a great deal of energy goes into accelerating the initially stationary "target" car. For the same reason head-on collisions of high-energy particles make much more energy available for creation of new particles. As a result, most of the highest-energy accelerators today are colliders, with particle beams accelerated in opposite directions and brought to collide inside elaborate detectors. The largest accelerator today is the Large Hadron Collider (LHC) at CERN, the European Laboratory for Particle Physics, at Geneva, Switzerland. With a 27-km circumference, LHC collides proton beams at energies up to 14 TeV. Both LHC and the Relativistic Heavy Ion Collider at Brookhaven National Laboratory in the United States create conditions that existed when the universe was only 10^{-12} s old; more on this in the next section.

39.5 The Evolving Universe

We come at the end to the broadest possible questions about physical reality: How did the universe begin? What is its overall structure? What will its future bring? Remarkably, these cosmological questions are closely linked with the questions of particle physics.

Expansion of the Universe

Early in the 20th century, astronomers argued about the nature of certain fuzzy patches visible in telescope photographs. Many thought they were gas clouds scattered among the visible stars, but others made a more radical proposal: that some of these "nebulae" were gravitationally bound systems containing billions of stars and that they were almost inconceivably distant.

In the 1920s, the opening of the 2.5-m telescope at California's Mt. Wilson Observatory finally resolved the issue. There, astronomer Edwin Hubble proved that some nebulae were indeed distant galaxies like our own Milky Way, each containing billions of stars. Today cosmologists think of galaxies as individual "point particles" whose distribution traces the overall structure of the universe.

Hubble continued to study the galaxies throughout the 1920s, and by analyzing their spectra he made a remarkable discovery: Spectral lines from distant galaxies are shifted toward the red, with the amount of shift dependent on the distance to the galaxies. The most reasonable and widely held explanation is that the redshift is caused by the Doppler effect (see Section 14.8). Then the implication of Hubble's work is that the distant galaxies are receding from us at speeds proportional to their distances. This result is known as **Hubble's law**:

$$v = H_0 d \tag{39.1}$$

where v is the recession speed, d the distance, and H_0 the **Hubble constant**, whose value is now known to be very nearly 22.7 kilometers per second per million light years of distance. Astronomers now use the Hubble relation to find the distances to remote galaxies, measuring their redshifts and using Equation 39.1 to infer their distances. Although the Hubble relation is written in terms of velocity, a more sophisticated view describes the Hubble expansion not as galactic motion but as a stretching of space itself—a process that also stretches light waves, giving the observed wavelength increase.

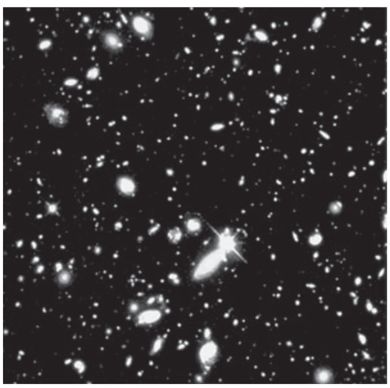


FIGURE 39.11 A portion of the Hubble Deep Field, a Hubble Space Telescope image showing distant galaxies whose redshifts provide information about cosmic expansion.

It may sound like Hubble's law puts us right in the center of things, in grotesque violation of modern science's view that Earth and its inhabitants don't occupy a favored position in the universe. But actually the inhabitants of any other galaxy would observe the same thing: All the distant galaxies would be receding from them at speeds proportional to their distances. As long as the universe is infinite in extent, none can claim to be at the center. And if it's not infinite, then Einstein's general theory of relativity gives it a closed-curve shape that still has no center.

Extrapolating Hubble's law backward in time suggests there was a time when all the galaxies were together. This implies that the universe had a definite beginning, in the form of a colossal explosion that flung matter into an expansion that continues today. Based on additional evidence that we'll describe shortly, scientists are quite certain that the universe began with such a **Big Bang**. Some of that evidence comes from the Hubble Space Telescope, named to honor Hubble's pioneering studies (Fig. 39.11).

EXAMPLE 39.2 Hubble's Law: Calculating the Age of the Universe

Using $H_0 = 22.7$ km/s/Mly and assuming the expansion rate has been constant, find out how long the universe has been expanding.

INTERPRET We're given the Hubble constant H_0 and asked to extrapolate back in time until all the galaxies were together.

DEVELOP If a galaxy has been moving with constant speed for time t, then its distance from us today is $d = vt$. But Hubble's law gives its speed in terms of distance: $v = H_0 d$. So our plan is to substitute $H_0 d$ for v and solve for the time t.

EVALUATE We have $t = d/v = d/H_0 d = 1/H_0$. To evaluate this expression we need to convert from the mixed units used for the Hubble constant:

$$t = \frac{1}{H_0} = \frac{1}{(22.7\ \text{km/s/Mly})/[3.00\times10^5(\text{km/s})/(\text{ly/y})]} = 13.2\ \text{Gy}$$

ASSESS Our calculation shows that the universe is about 13 billion years old—on the assumption that the expansion rate hasn't changed. We'll soon see that this assumption isn't quite correct, but our result remains a good estimate for the age of the universe (the actual number is closer to 14 Gy). Note that we used $c = 3.00\times10^5$ km/s in converting from km/s to ly/y so our answer would come out in years.

The Cosmic Background Radiation

In 1965 Arno Penzias and Robert Wilson at Bell Laboratories found a faint "noise" of microwave radiation in a satellite communications antenna they were testing. The noise seemed to come from all directions in the sky. Theorists at Princeton identified Penzias and Wilson's "noise" as radiation dating to a much earlier era in the universe. This **cosmic microwave background radiation** is the strongest evidence yet for the Big Bang.

The Big Bang theory suggests that the universe started very hot and then cooled as it expanded, doing work against its own gravitation. At first it was so hot that any atoms that formed would be dashed apart by collisions at the high thermal energy prevailing. Thus, in its early times the universe was populated by individual charged particles. These interacted readily with electromagnetic radiation, making the universe opaque. But by about 400,000 years the temperature had dropped to some 3000 K, and at that point atoms of hydrogen and helium could form. Since neutral atoms interact much more weakly with electromagnetic radiation, the universe became transparent, and photons emitted as the atoms formed could travel throughout the universe with little chance of being subsequently absorbed. Those photons became the cosmic background radiation, permeating the entire universe.

Measurements of the cosmic microwave background show a near-perfect fit to a 2.7-K blackbody spectrum (Fig. 39.12). Applying the "stretching of space" interpretation of the Hubble expansion shows that the universe has expanded about 1000-fold since the background radiation formed, dropping the temperature from 3000 K to about 3 K, and stretching the radiation's wavelength by the same factor. That's why radiation that initially had μm wavelengths now peaks in the microwave region with mm wavelengths. Thus the cosmic microwave background is a direct reflection of the conditions when it formed 400,000 years after the universe began.

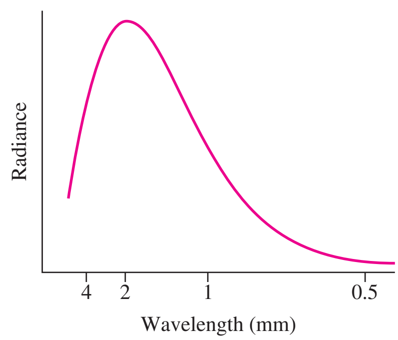


FIGURE 39.12 Spectrum of the cosmic microwave background matches perfectly that of a blackbody at 2.726 K.

The cosmic microwave background (CMB) is remarkably uniform, but not perfectly so. Today, satellite instruments record minute spatial variations in the microwave background, and analysis of these variations reveals a wealth of information about the early universe (Fig. 39.13). For example, the CMB variations are consistent with a universe whose overall geometry is flat, and the variations themselves represent the "seeds" of galaxies, galaxy clusters, and even larger structures that dominate the universe today.

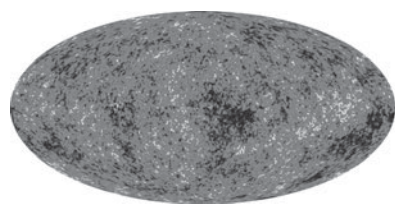


FIGURE 39.13 All-sky map of the cosmic microwave background shows minute spatial variations in the radiation intensity. Data were taken with the Wilkinson Microwave Anisotropy Probe (WMAP) satellite.

The Earliest Times

The cosmic microwave background shows us the universe as it was some 400,000 years after the Big Bang. Nuclear physics takes us back even further, to the time from about 1 second to 30 minutes when the lightest nuclei were forming. The first composite nuclei were the simplest: deuterium, consisting of a proton and a neutron. The rate of deuterium formation is critically sensitive to the expansion rate of the universe. Measurements of deuterium abundance, based on spectral lines from interstellar deuterium, therefore provide direct evidence for conditions early in the Big Bang.

Evidence for still earlier times comes from particle physics, which predicts the interactions and particle populations under the hot, high-energy conditions that existed a fraction of a second after the Big Bang. In 2005, experiments using the Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Laboratory produced a so-called quark–gluon plasma similar to the state of the universe just microseconds after the Big Bang. Thus RHIC and other high-energy particle accelerators can act as "time machines," allowing us to study conditions that prevailed in the very early universe. Figure 39.14 (next page) summarizes our understanding of the universe's evolution, showing that the phenomena of particle physics and unification are inextricably tied with cosmic expansion.

The Inflationary Universe

The original Big Bang theory has difficulty explaining several features of the observed universe. Why, for example, do we find only matter but not antimatter? Why does the universe seem homogeneous and in thermodynamic equilibrium on the largest scales, when the most distant regions are so far apart that light could not have traveled between them in the time since the beginning? And why does the overall geometry of the universe appear flat, when general relativity allows for curved space?

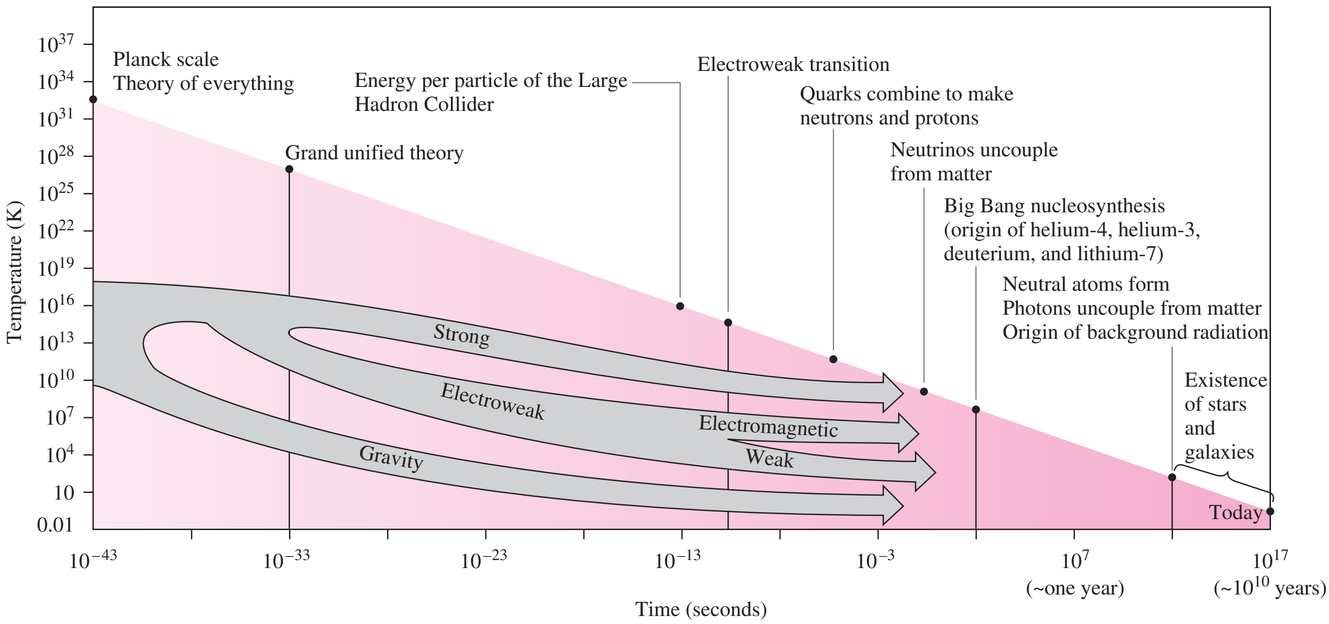


FIGURE 39.14 Evolution of the universe from the earliest times to the present. Note the highly logarithmic scales.

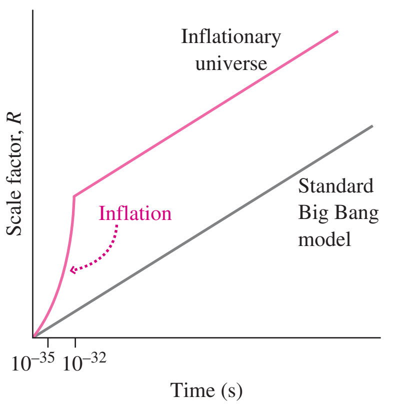


FIGURE 39.15 Expansion in the standard and inflationary Big Bang theories. The scale factor *R* measures the amount of expansion.

The solution to these conundrums is **inflation**, an idea first advanced by Alan Guth of MIT. Guth's theory holds that the universe underwent a period of exponential expansion beginning at about 10^{-35} s and lasting until about 10^{-32} s (Fig. 39.15). The expansion was the result of a delay in the symmetry breaking that made the fundamental forces appear distinct. Because of the tremendous expansion, now-distant locations would once have been close enough to reach the thermodynamic equilibrium that we now observe. Furthermore, the inflationary expansion would "flatten out" any overall curvature, giving us the flat universe we see today.

Dark Matter, Dark Energy, and the Future of the Universe

Will the universe expand forever, or will the expansion eventually stop and reverse? That's like asking whether a spacecraft will escape Earth forever or ultimately return, and the answer is the same: If the system's kinetic energy exceeds the magnitude of its (negative) potential energy, then expansion will proceed forever. Thus a single parameter—designated Ω (capital Greek omega)—determines the fate of the universe:

$$\Omega = \frac{|\text{potential energy of the universe}|}{\text{kinetic energy of the universe}}$$

where $\Omega > 1$ implies eventual collapse, and $\Omega < 1$ means continued expansion. An alternative way of expressing the same dichotomy is in terms of the average density. A simple Newtonian calculation based on kinetic and gravitational potential energies of particles in an expanding universe gives Ω in terms of the average density ρ and the Hubble constant: $\Omega = 8\pi G\rho/3H_0^2$—a result that turns out to be identical to the correct general relativistic calculation. Setting $\Omega = 1$ gives the **critical density** that divides eternal expansion from eventual collapse:

$$\rho_c = \frac{3H_0^2}{8\pi G} \qquad \text{(critical density)} \qquad (39.2)$$

Analysis of the cosmic microwave background fluctuations strongly suggests that the actual universe is at the critical density, with $\Omega = 1$. But when we total the visible matter in the universe, we come up with far less than the critical density. Elementary particle theories corroborate this observational result, suggesting that there can be only enough ordinary matter—protons, neutrons, nuclei, and atoms—to make up about 4% of the critical density. Furthermore, the motions of stars in galaxies and of galaxies in clusters suggest the presence of a great deal more mass than is visible. All this implies the existence of

dark matter whose composition is unknown and which can't be the sort of matter—made mostly from quarks—with which we're familiar.

Another approach to the cosmic density is to study the most distant galaxies, whose light has taken so long to reach us that we're seeing them as they were in the early universe. One might expect that cosmic expansion was faster in earlier times, and slowed as the galaxies did work against their mutual gravitational attraction. Comparing the recession speeds of ancient, distant galaxies with the speeds of nearer ones should then give the rate of cosmic deceleration—which, in turn, should depend on cosmic density. But observations in 1998 gave a surprising and unexpected result: The cosmic expansion is actually accelerating!

Cosmic acceleration implies a kind of "antigravity" operating on the largest scales. Ironically, Einstein had proposed just such a phenomenon in his original formulation of general relativity. Einstein needed this **cosmological constant** in his theory in order to keep the universe static—which, in 1916, astronomers thought it was. When Hubble then showed that the universe is expanding, Einstein dropped the cosmological constant and called it "the greatest blunder of my life." Now it appears that Einstein had the right idea in the first place.

The source of the cosmic acceleration is **dark energy**, which may be just another name for Einstein's cosmological constant or may be a different phenomenon with the same "antigravity" effect. At this point we don't know quite what it is. But we do know how much of it there is: Dark energy is fully 73% of the "stuff" that makes up our universe. Another 23% is dark matter. That means only 4% of the universe is in the form of familiar matter made from baryons and thus ultimately from quarks. These numbers come from a confluence of recent observations of the cosmic microwave background, distant supernovae, and surveys of distant galaxies; together, they tightly constrain the relative amounts of matter and dark energy (Fig. 39.16). Here, at the end of a long physics course, it's sobering to realize that much of our universe consists of "stuff" about which we know so very little!

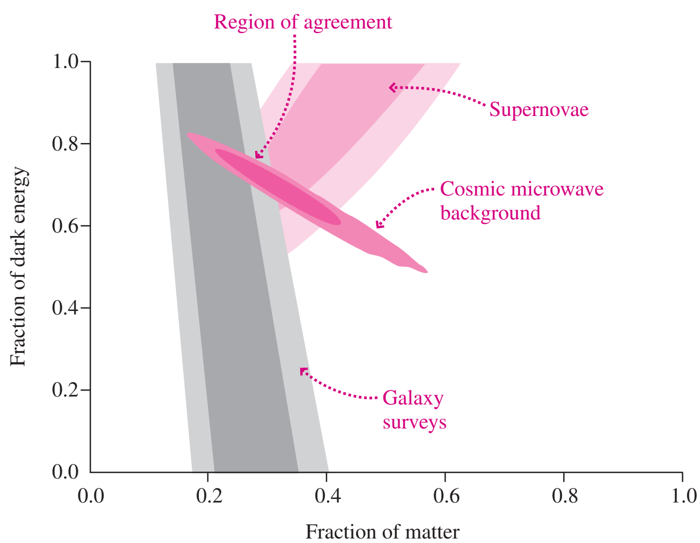


FIGURE 39.16 Constraints on the relative abundances of dark energy and of matter in the universe, from three independent sets of observations. All three converge to agreement on approximately 73% dark energy and 27% matter; the latter includes 23% dark matter and 4% ordinary matter.

Understanding the Universe

With this chapter's brief survey, we've reached the limits of present understanding of the universe. On the way we've seen that particle physics and cosmology are inextricably linked. To understand our universe, we need to understand all its aspects from the largest to the smallest—and that means understanding all the forces, from gravity to the weak force; all the physical laws, from Newton's and Maxwell's to the laws of quantum mechanics; and the nature of the elementary particles. In this text we've touched on all these topics, and we hope we've given you a foundation for further understanding and appreciation of the richness and diversity of the physical universe.

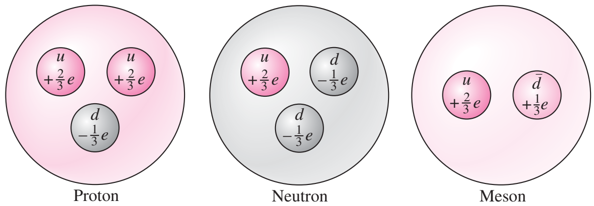
CHAPTER 39 SUMMARY

Big Picture

The big idea here is that the structure of ordinary matter and its interactions can be explained in terms of a handful of particles: **quarks**, which make up the familiar proton and neutron and a host of other **hadrons**; **leptons**, including the familiar electrons and elusive neutrinos; and **gauge bosons**, which mediate the fundamental forces and include the photon for the electromagnetic force. The **standard model** describes these elementary particles and their interactions; it sheds light not only on the structure of matter we see today but on the early universe as well. But there's a humbling caveat: Only about 4% of the universe consists of familiar matter. The rest is **dark energy** and **dark matter**, about which we know very little.

Key Concepts and Equations

Quarks, which come in six flavors and three colors, join in threes to form **baryons**, including protons and neutrons, and in twos to form **mesons**. Quarks carry fractional electric charges, but because the **strong force** between quarks doesn't decrease with distance, it appears impossible to have an isolated quark.



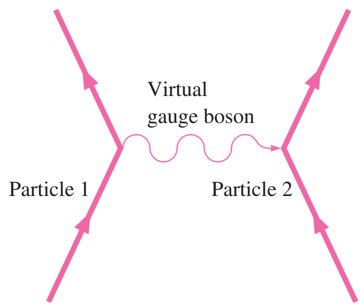
Leptons are the other class of elementary matter particles, and include the electron and its more massive cousins, the muon and tau, and the three types of neutrinos.

The three **fundamental forces**—strong, electroweak, and gravity—are believed to be manifestations of the same fundamental interactions that appear unified at high enough energies.

At typical energies in today's universe, though, the electroweak force separates into the electromagnetic and weak forces. The forces differ greatly in strength:

Force	Relative strength at 1 fm (approximate)
Gravity	10^{-38}
Weak	10^{-13}
Electromagnetic	10^{-2}
Strong	1

Exchange of **gauge bosons** explains the forces between particles.



Gauge Bosons

Electromagnetic force: photon

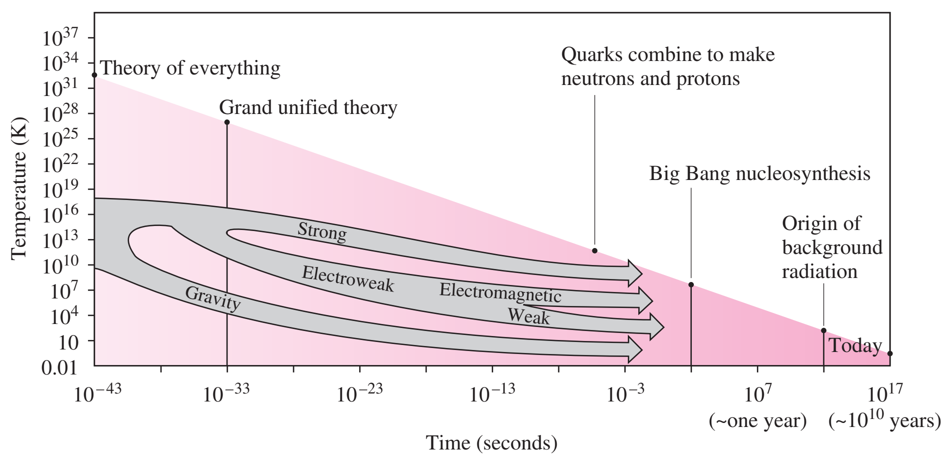
Weak force: $W^{\pm}$, Z^0

Strong force: gluon (8 varieties)

Gravity: graviton

Applications

Our knowledge of particles and their interactions, combined with general relativity and with observations of the **Hubble expansion** and the **cosmic microwave background radiation**, gives us a picture of the origin and structure of the universe. The universe began some 14 billion years ago in a hot **Big Bang**. The simplest nuclei formed within the first half-hour, and the first atoms at about 400,000 years; at that point the universe became transparent and the cosmic microwave background formed. We've learned recently that the universe is overall flat and that its expansion is accelerating under the influence of mysterious **dark energy**. There's still much that we don't know!



For homework assigned on MasteringPhysics, go to www.masteringphysics.com | BIO *Biology and/or medical-related problems* C *Computer problems*

For Thought and Discussion

1. Why did Yukawa conclude that the particle mediating the strong force should have nonzero mass?
2. How can we follow the tracks of individual particles?
3. How are baryons fundamentally different from leptons?
4. What coordinates are changed under the inversion processes P and T?
5. Why are we unlikely to observe an isolated quark?
6. Describe the relation between the strong force and the nuclear force.
7. What's the role of gluons?
8. Classify (a) mesons and (b) baryons as fermions or bosons, and relate your classification to the particles' quark compositions.
9. Name the fundamental force involved in (a) binding of a proton and a neutron to make a deuterium nucleus; (b) decay of a neutron to a proton, an electron, and a neutrino; (c) binding of an electron and a proton to make a hydrogen atom.
10. What forces are unified in the electroweak theory?
11. What forces would be unified by GUTs?
12. Why do we need higher-energy particle accelerators to explore fully the standard model?
13. How can Hubble's law hold without the universe having a center?
14. Is it possible for a charged particle to be its own antiparticle?
15. Describe the origin of the cosmic microwave background.
16. Explain how particle accelerators can help us understand the early universe.
17. BIO What medical diagnostic procedure makes use of the fact that every particle has an antiparticle? What particle/antiparticle pair is involved?
18. The radiation that we observe as the cosmic microwave background started out largely as infrared. Why is it now the *microwave* background?

Exercises and Problems

Exercises

Section 39.1 Particles and Forces

19. How long could a virtual photon of 633-nm red laser light exist without violating conservation of energy?
20. Some scientists have speculated on a possible "fifth force," with a range of about 100 m. Following Yukawa's reasoning, what would be the mass of the field particle mediating such a force?

Section 39.2 Particles and More Particles

21. Write the equation for the decay of a positive pion to a muon and a neutrino, being sure to label the type of neutrino. (*Hint:* The positive muon is an antiparticle.)
22. Use Table 39.1 to find the total strangeness before and after the decay $\Lambda^0 \rightarrow \pi^- + p$, and use your answer to determine which force is involved in this reaction.
23. The η^0 particle is a neutral nonstrange meson that can decay to a positive pion, a negative pion, and a neutral pion. Write the reaction for this decay, and verify that it conserves charge, baryon number, and strangeness.
24. Are either or both of these decay schemes possible for the tau particle: (a) $\tau^- \rightarrow e^- + \bar{\nu}_e + \nu_\tau$; (b) $\tau^- \rightarrow \pi^- + \pi^0 + \nu_\tau$?
25. Is the interaction $p + p \rightarrow p + \pi^+$ allowed? If not, what conservation law does it violate?

Section 39.3 Quarks and the Standard Model

26. Determine the quark composition of the π^-.
27. The Eightfold Way led Gell-Mann to predict a baryon with strangeness -3. Determine this particle's quark composition.
28. The Σ^+ and Σ^- have quark compositions *uus* and *dds*, respectively. Are the Σ^+ and Σ^- each other's antiparticles? If not, give the quark compositions of their antiparticles.

Section 39.4 Unification

29. Estimate the volume of the 50,000 tons of water used in the Super Kamiokande experiment shown in Fig. 39.8.
30. Estimate the temperature in a gas of particles such that the thermal energy kT is high enough to make electromagnetism and the weak force appear as a single phenomenon.
31. Repeat Exercise 30 for the 10^{15}-GeV energy of grand unification.

Section 39.5 The Evolving Universe

32. Express the Hubble constant in SI units.
33. Find the distance to a galaxy that is receding from us at 2×10^4 km/s.
34. Find the recession speed of a galaxy 300 Mly from Earth.
35. What would be the age of the universe, assuming constant expansion rate, if the Hubble constant were 25 km/s/Mly?

Problems

36. The mass of the photon is assumed to be zero, but experiments put only an upper limit of 5×10^{-63} kg on the photon mass. What would the range of the electromagnetic force be if the photon mass were actually at this upper limit?
37. Which of the reactions (a) $\Lambda^0 \rightarrow \pi^+ + \pi^-$ and (b) $K^0 \rightarrow \pi^+ + \pi^-$ is not possible, and why not?
38. Both the neutral kaon and the neutral ρ meson can decay to a pion–antipion pair. Which of these decays is mediated by the weak force? How can you tell?
39. Some grand unification theories suggest that the decay $p \rightarrow \pi^0 + e^+$ may be possible, in which case all matter may eventually become radiation. Are (a) baryon number and (b) electric charge conserved in this hypothetical proton decay?
40. Consider systems described by wave functions that are proportional to the terms (a) xy^2z, (b) x^2yz, and (c) xyz, where x, y, and z are the spatial coordinates. Which pairs of these systems could be transformed into each other under a parity-conserving interaction?
41. The J/ψ particle is an uncharmed meson that nevertheless includes charmed quarks. Determine its quark composition.
42. List all the possible quark triplets formed from any combination of up, down, and charmed quarks, along with the charge of each.
43. The Tevatron at Fermilab accelerates protons to energy of 1 TeV. (a) How much is this in joules? (b) How far would a 1-g mass have to fall in Earth's gravitational field to gain this much energy?
44. (a) What's the relativistic factor γ for a 7-TeV proton in the Large Hadron Collider? (b) Find an accurate value for the proton's speed.
45. How long, as measured in the lab frame, does it take a 7-TeV proton to circle the 27-km circumference of the Large Hadron Collider?
46. Estimate the critical density of the universe.
47. Estimate the diameter to which the Sun would have to be expanded for its average density to be the critical density found in Problem 46.

48. A baryon called the neutral lambda particle has mass 1116 MeV/c^2. Find the minimum speed necessary for the particles in a proton–antiproton collider to produce lambda–antilambda pairs.
49. A so-called muonic atom is a hydrogen atom with the electron replaced by a muon; the muon's mass is 207 times the electron's. Find (a) the size and (b) the ground-state energy of a muonic atom.
50. (a) By what factor must the magnetic field in a proton synchrotron be increased as the proton energy increases by a factor of 10? Assume the protons are highly relativistic, so $\gamma \gg 1$. (b) By what factor must the diameter of the accelerator be increased to raise the energy by a factor of 10 without changing the magnetic field?
51. A galaxy's hydrogen-β spectral line, normally at 486.1 nm, appears at 495.4 nm. (a) Use the Doppler shift of Chapter 14 to find the galaxy's recession speed, and (b) infer the distance to the galaxy. Is it appropriate to use Chapter 14's nonrelativistic Doppler formulas in this case?
52. At the time the cosmic microwave background radiation originated, the temperature of the universe was about 3000 K. What were (a) the median wavelength of the newly formed radiation (Equation 34.2b) and (b) the corresponding photon energy?
53. Many particles are far too short-lived for their lifetimes to be measured directly. Instead, tables of particle properties often list "width," measured in energy units and indicating the width of the distribution of measured rest energies. For example, the Z^0 has mass 91.18 GeV and width 2.5 GeV. Use the energy–time uncertainty relation to estimate its corresponding lifetime.
54. A mix of particles starts with equal numbers of the three types of sigma particles listed in Table 39.1. Find the relative portion of each after (a) 5×10^{-20} s and (b) 5×10^{-10} s. Give your answer in a reference frame in which the particles are at rest.
55. You pick up an old astronomy book and read that the Hubble constant is 17 km/s/Mly. You know that today's more accurate value is 22.7 km/s/Mly. Use the simplified reasoning of Example 39.2 to compare the ages for the universe implied by these two values of H_0.
56. A friend believes the universe is very young, less than 10,000 years old. Based on Hubble's law, how would you argue that the universe is older? What would the Hubble constant be for a million-year-old universe?
57. Your roommate is writing a science-fiction novel set very far in the future, 60 Gy after the Big Bang. One of the characters is a cosmologist, and your roommate wants to know what the cosmologist will measure for the Hubble constant. What's your answer, assuming a steady expansion rate?

Passage Problems

Pions are the lightest mesons, with mass some 270 times that of the electron. Charged pions decay typically into a muon and a neutrino or antineutrino. This makes pion beams useful for producing beams of neutrinos, which physicists use to study those elusive particles. In a medical application during the late 20th century, accelerator centers installed "biomedical beam lines" to test pions for cancer therapy. In these experiments, pions attached themselves to atomic nuclei within cancer cells. The nuclei would literally explode, delivering a "pion star" of cancer-killing nuclear debris. Unfortunately, results were not as encouraging as hoped, and enthusiasm for this technique has waned.

58. The negative pion usually decays into a negative muon and one other particle. The other particle could be
    a. a proton.
    b. an antineutrino.
    c. a neutrino.
    d. an up quark.
59. In the cancer-treatment experiments described in the passage, for which pions is it energetically easiest to be captured by a nucleus?
    a. π^+
    b. π^0
    c. π^-
    d. Energetically, capture is equally likely for all three pions.
60. The lifetime of charged pions is 26 ns. The length of an accelerator's biomedical beam line, from the point where pions are created to the patient, could be at most about
    a. 800 m long.
    b. 80 m long.
    c. 8 m long.
    d. 80 cm long.
61. The quark composition of the negative pion is
    a. uud.
    b. $d\bar{u}$.
    c. ud.
    d. $c\bar{c}$.

Answer to Chapter Question

Answer to Chapter Opening Question

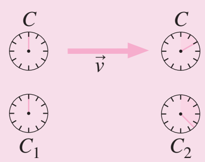
The Large Hadron Collider creates conditions similar to those in the first microseconds of the universe, providing physicists a direct look at the behavior of matter in these early times.

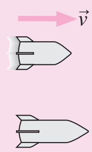
Modern Physics

Modern physics, developed since the year 1900, contrasts with the **classical physics** that came before. Modern physics is essential in understanding physical reality at the atomic scale, at very low temperatures, at very high relative velocities, in regions of very strong gravity, and in the evolution and large-scale structure of the universe.

The two big ideas in modern physics are **relativity** and **quantum mechanics**. Relativity is based on a simple principle but drastically alters our commonsense notions of space and time, matter and energy, and the nature of gravity. Quantum mechanics replaces Newtonian determinism with a statistical description in which matter and energy exhibit both wave-like and particle-like behaviors.

Einstein's **special theory of relativity** is based in the statement that the laws of physics are the same for all observers in uniform motion. Therefore, Maxwell's prediction that there should be electromagnetic waves propagating at the speed of light c is valid in all uniformly moving reference frames. So measures of space and time cannot be absolute, but depend on one's reference frame.





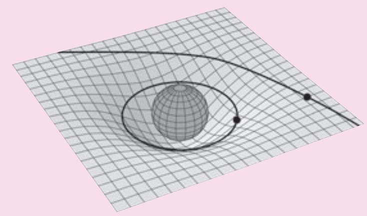
The time between two events is shortest in a reference frame where events occur at the same place; here that's the reference frame of clock C.

An object's length is longest in a reference frame in which it's at rest.

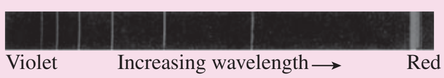
Energy E, momentum p, and mass m in relativity are related by $E^2 = p^2c^2 + (mc^2)^2$.

For an object at rest with respect to an observer, this gives $E = mc^2$, showing the relativistic equivalence of mass and energy.

General relativity is Einstein's theory of gravity, which explains gravity as the geometric curvature of spacetime. General relativity is central to modern astrophysics and cosmology, describing phenomena from black holes to the overall structure of the universe.



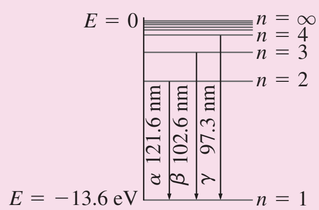
Quantum physics arose from attempts to explain several phenomena observed around the turn of the 20th century. These include blackbody radiation from hot objects, the photoelectric effect, and the existence and spectra of atoms.



Hydrogen spectrum

Wave–particle duality is at the heart of quantum physics. The energy in electromagnetic radiation of frequency f is concentrated in particle-like "bundles" called photons. Thus electromagnetic energy is quantized, with each photon carrying energy $E = hf$, where $h = 6.63 \times 10^{-34}$ J·s is Planck's constant. Conversely, matter exhibits wave-like behavior. The de Broglie wavelength associated with a particle of momentum p is $\lambda = h/p$.

Quantization of atomic angular momentum leads to the **Bohr model** for the atom, with quantized energy levels that help explain atomic spectra.



Hydrogen energy levels

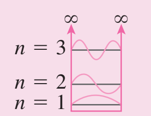
Wave–particle duality is closely related to the **Heisenberg uncertainty principle**, which states that it's impossible to measure a particle's position and momentum simultaneously with perfect accuracy. Rather, the uncertainties Δx and Δp in position and momentum must obey the inequality

$$\Delta x\, \Delta p \geq \hbar \text{ (uncertainty principle), where } \hbar = h/2\pi$$

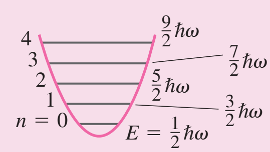
Quantum mechanics describes phenomena at the atomic scale. The **Schrödinger equation** gives the **wave function**, ψ, for a particle of mass m with potential energy U and total energy E:

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + U(x)\psi = E\psi(x)$$

ψ^2 is the **probability density** for finding the particle. Application of the Schrödinger equation to bound systems results in quantized energy levels.



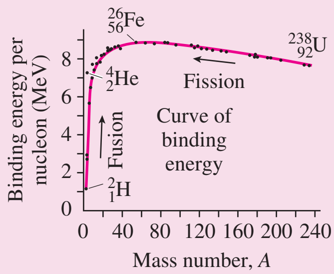
Infinite square well



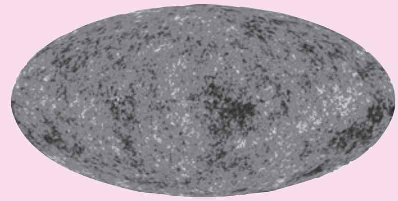
Harmonic oscillator

Application of the Schrödinger equation to atoms and molecules explains atomic and molecular structure, the organization of the chemical elements in the periodic table, and the behavior of crystalline solids.

Nuclear physics plunges into the heart of the atom and shows that larger nuclei require more neutrons than protons in order for the strong nuclear force to overcome the repulsive electric interaction between protons. Nuclear physics describes such phenomena as radioactivity and the production of energy by nuclear fission and fusion.



Applying the principles of physics from the subatomic scale of quantum and particle physics to the largest scales described by general relativity gives us our modern-day picture of the origin and evolution of the universe.



The cosmic microwave background radiation, shown here in an image from the WMAP satellite.

Part Six Challenge Problem

Derive Equation 39.2 for the critical density of the universe on the assumption that the universe is spherically symmetric, is homogeneous, and obeys Newton's laws on large scales.

Mathematics

APPENDIX A

A-1 Algebra and Trigonometry

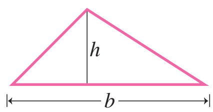
Quadratic Formula

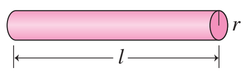
If $ax^2 + bx + c = 0$, then $x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

Circumference, Area, Volume

Where $\pi \simeq 3.14159\ldots$:

circumference of circle	$2\pi r$
area of circle	πr^2
surface area of sphere	$4\pi r^2$
volume of sphere	$\frac{4}{3}\pi r^3$
area of triangle	$\frac{1}{2}bh$
volume of cylinder	$\pi r^2 l$





Trigonometry

definition of angle (in radians): $\theta = \dfrac{s}{r}$

2π radians in complete circle

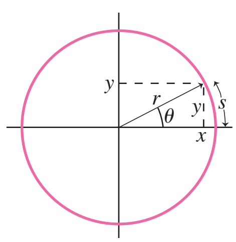
1 radian $\simeq 57.3°$

Trigonometric Functions

$$\sin\theta = \frac{y}{r}$$

$$\cos\theta = \frac{x}{r}$$

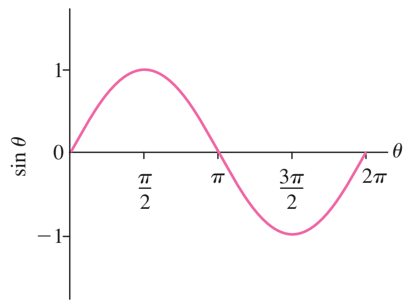
$$\tan\theta = \frac{\sin\theta}{\cos\theta} = \frac{y}{x}$$

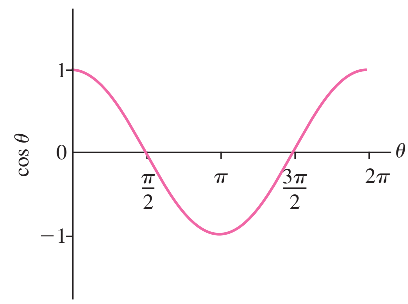


Values at Selected Angles

$\theta \rightarrow$	**0**	$\frac{\pi}{6}$ **(30°)**	$\frac{\pi}{4}$ **(45°)**	$\frac{\pi}{3}$ **(60°)**	$\frac{\pi}{2}$ **(90°)**
$\sin\theta$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1
$\cos\theta$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0
$\tan\theta$	0	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$	∞

Graphs of Trigonometric Functions





Trigonometric Identities

$$\sin(-\theta) = -\sin\theta$$

$$\cos(-\theta) = \cos\theta$$

$$\sin\left(\theta \pm \frac{\pi}{2}\right) = \pm\cos\theta$$

$$\cos\left(\theta \pm \frac{\pi}{2}\right) = \mp\sin\theta$$

$$\sin^2\theta + \cos^2\theta = 1$$

$$\sin 2\theta = 2\sin\theta\cos\theta$$

$$\cos 2\theta = \cos^2\theta - \sin^2\theta = 1 - 2\sin^2\theta = 2\cos^2\theta - 1$$

$$\sin(\alpha \pm \beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta$$

$$\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$$

$$\sin\alpha \pm \sin\beta = 2\sin\left[\tfrac{1}{2}(\alpha \pm \beta)\right]\cos\left[\tfrac{1}{2}(\alpha \mp \beta)\right]$$

$$\cos\alpha + \cos\beta = 2\cos\left[\tfrac{1}{2}(\alpha + \beta)\right]\cos\left[\tfrac{1}{2}(\alpha - \beta)\right]$$

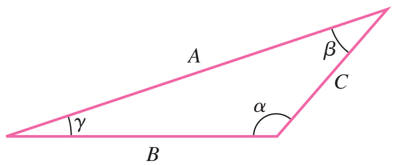
$$\cos\alpha - \cos\beta = -2\sin\left[\tfrac{1}{2}(\alpha + \beta)\right]\sin\left[\tfrac{1}{2}(\alpha - \beta)\right]$$

Laws of Cosines and Sines

Where A, B, C are the sides of an arbitrary triangle and α, β, γ the angles opposite those sides:

Law of cosines

$$C^2 = A^2 + B^2 - 2AB\cos\gamma$$



Law of sines

$$\frac{\sin\alpha}{A} = \frac{\sin\beta}{B} = \frac{\sin\gamma}{C}$$

Exponentials and Logarithms

$e^{\ln x} = x, \quad \ln e^x = x \quad e = 2.71828\ldots$

$a^x = e^{x\ln a} \qquad \ln(xy) = \ln x + \ln y$

$a^x a^y = a^{x+y} \qquad \ln\left(\frac{x}{y}\right) = \ln x - \ln y$

$(a^x)^y = a^{xy} \qquad \ln\left(\frac{1}{x}\right) = -\ln x$

$\log x \equiv \log_{10} x = \ln(10)\ln x \simeq 2.3\ln x$

Approximations

For $|x| \ll 1$, the following expressions provide good approximations to common functions:

$e^x \simeq 1 + x$

$\sin x \simeq x$

$\cos x \simeq 1 - \frac{1}{2}x^2$

$\ln(1 + x) \simeq x$

$(1 + x)^p \simeq 1 + px \quad$ (binomial approximation)

Expressions that don't have the forms shown may often be put in the appropriate form. For example:

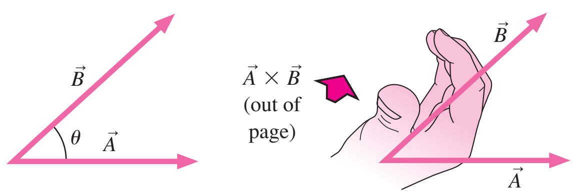
$$\frac{1}{\sqrt{a^2 + y^2}} = \frac{1}{a\sqrt{1 + \dfrac{y^2}{a^2}}} = \frac{1}{a}\left(1 + \frac{y^2}{a^2}\right)^{-1/2} \simeq \frac{1}{a}\left(1 - \frac{y^2}{2a}\right) \quad \text{for} \quad y^2/a^2 \ll 1, \text{or } y^2 \ll a^2$$

Vector Algebra

Vector Products

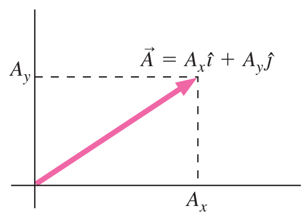
$\vec{A}\cdot\vec{B} = AB\cos\theta$

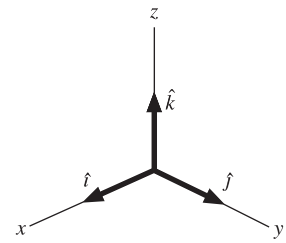
$|\vec{A}\times\vec{B}| = AB\sin\theta$, with direction of $\vec{A}\times\vec{B}$ given by the right-hand rule:



Unit Vector Notation

An arbitrary vector $\vec{A}$ may be written in terms of its components A_x, A_y, A_z and the unit vectors $\hat{\imath}$, $\hat{\jmath}$, $\hat{k}$ that have length 1 and lie along the x-, y-, and z-axes:





In unit vector notation, vector products become

$$\vec{A} \cdot \vec{B} = A_xB_x + A_yB_y + A_zB_z$$

$$\vec{A} \times \vec{B} = (A_yB_z - A_zB_y)\hat{\imath} + (A_zB_x - A_xB_z)\hat{\jmath} + (A_xB_y - A_yB_x)\hat{k}$$

Vector Identities

$$\vec{A} \cdot \vec{B} = \vec{B} \cdot \vec{A}$$

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

$$\vec{A} \cdot (\vec{B} \times \vec{C}) = \vec{B} \cdot (\vec{C} \times \vec{A}) = \vec{C} \cdot (\vec{A} \times \vec{B})$$

$$\vec{A} \times (\vec{B} \times \vec{C}) = (\vec{A} \cdot \vec{C})\vec{B} - (\vec{A} \cdot \vec{B})\vec{C}$$

A-2 Calculus

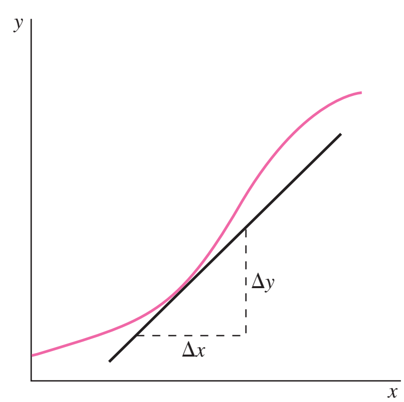
Derivatives

Definition of the Derivative

If y is a function of x, then the **derivative of y with respect to x** is the ratio of the change Δy in y to the corresponding change Δx in x, in the limit of arbitrarily small Δx:

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

Algebraically, the derivative is the rate of change of y with respect to x; geometrically, it is the slope of the y versus x graph—that is, of the tangent line to the graph at a given point:



Derivatives of Common Functions

$$\frac{da}{dx} = 0 \quad (a \text{ is a constant})$$

$$\frac{dx^n}{dx} = nx^{n-1} \quad (n \text{ need not be an integer})$$

$$\frac{d}{dx}\sin x = \cos x$$

$$\frac{d}{dx}\cos x = -\sin x$$

$$\frac{d}{dx}\tan x = \frac{1}{\cos^2 x}$$

$$\frac{de^x}{dx} = e^x$$

$$\frac{d}{dx}\ln x = \frac{1}{x}$$

Derivatives of Sums, Products, and Functions of Functions

1. **Derivative of a constant times a function**

$$\frac{d}{dx}[af(x)] = a\frac{df}{dx} \qquad (a \text{ is a constant})$$

2. **Derivative of a sum**

$$\frac{d}{dx}[f(x) + g(x)] = \frac{df}{dx} + \frac{dg}{dx}$$

3. **Derivative of a product**

$$\frac{d}{dx}[f(x)g(x)] = g\frac{df}{dx} + f\frac{dg}{dx}$$

Examples

$$\frac{d}{dx}(x^2\cos x) = \cos x\frac{dx^2}{dx} + x^2\frac{d}{dx}\cos x = 2x\cos x - x^2\sin x$$

$$\frac{d}{dx}(x\ln x) = \ln x\frac{dx}{dx} + x\frac{d}{dx}\ln x = (\ln x)(1) + x\left(\frac{1}{x}\right) = \ln x + 1$$

4. **Derivative of a quotient**

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{1}{g^2}\left(g\frac{df}{dx} - f\frac{dg}{dx}\right)$$

Example

$$\frac{d}{dx}\left(\frac{\sin x}{x^2}\right) = \frac{1}{x^4}\left(x^2\frac{d}{dx}\sin x - \sin x\frac{dx^2}{dx}\right) = \frac{\cos x}{x^2} - \frac{2\sin x}{x^3}$$

5. **Chain rule for derivatives**
If f is a function of u and u is a function of x, then

$$\frac{df}{dx} = \frac{df}{du}\frac{du}{dx}$$

Examples

a. Evaluate $\frac{d}{dx}\sin(x^2)$. Here $u = x^2$ and $f(u) = \sin u$, so

$$\frac{d}{dx}\sin(x^2) = \frac{d}{du}\sin u\frac{du}{dx} = (\cos u)\frac{dx^2}{dx} = 2x\cos(x^2)$$

b. $\frac{d}{dt}\sin\omega t = \frac{d}{d\,\omega t}\sin\omega t\frac{d}{dt}\omega t = \omega\cos\omega t \qquad (\omega \text{ is a constant})$

c. Evaluate $\dfrac{d}{dx}\sin^2 5x$. Here $u = \sin 5x$ and $f(u) = u^2$, so

$$\frac{d}{dx}\sin^2 5x = \frac{d}{du}u^2\frac{du}{dx} = 2u\frac{du}{dx} = 2\sin 5x\frac{d}{dx}\sin 5x$$
$$= (2)(\sin 5x)(5)(\cos 5x) = 10\sin 5x\cos 5x = 5\sin 2x$$

Second Derivative

The second derivative of y with respect to x is defined as the derivative of the derivative:

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right)$$

Example

If $y = ax^3$, then $dy/dx = 3ax^2$, so

$$\frac{d^2y}{dx^2} = \frac{d}{dx}3ax^2 = 6ax$$

Partial Derivatives

When a function depends on more than one variable, the partial derivatives of that function are the derivatives with respect to each variable, taken with all other variables held constant. If f is a function of x and y, then the partial derivatives are written

$$\frac{\partial f}{\partial x} \quad \text{and} \quad \frac{\partial f}{\partial y}$$

Example

If $f(x, y) = x^3 \sin y$, then

$$\frac{\partial f}{\partial x} = 3x^2\sin y \quad \text{and} \quad \frac{\partial f}{\partial y} = x^3\cos y$$

Integrals

Indefinite Integrals

Integration is the inverse of differentiation. The **indefinite integral**, $\int f(x)\,dx$, is defined as a function whose derivative is $f(x)$:

$$\frac{d}{dx}\left[\int f(x)\,dx\right] = f(x)$$

If $A(x)$ is an indefinite integral of $f(x)$, then because the derivative of a constant is zero, the function $A(x) + C$ is also an indefinite integral of $f(x)$, where C is any constant. Inverting the derivatives of common functions listed in the preceding section gives the integrals that follow (a more extensive table appears at the end of this appendix).

$$\int a\,dx = ax + C \qquad \int \cos x\,dx = \sin x + C$$

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + C, \quad n \neq -1 \qquad \int e^x\,dx = e^x + C$$

$$\int \sin x\,dx = -\cos x + C \qquad \int x^{-1}\,dx = \ln x + C$$

Definite Integrals

In physics we're most often interested in the **definite integral**, defined as the sum of a large number of very small quantities, in the limit as the number of quantities grows arbitrarily large and the size of each arbitrarily small:

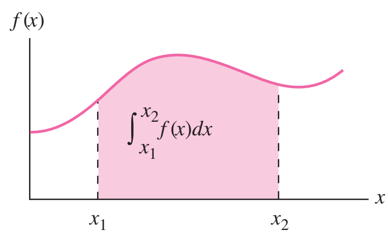
$$\int_{x_1}^{x_2} f(x)\,dx \equiv \lim_{\substack{\Delta x \to 0 \\ N \to \infty}} \sum_{i=1}^{N} f(x_i)\,\Delta x$$

where the terms in the sum are evaluated at values x_i between the limits of integration x_1 and x_2; in the limit $\Delta x \to 0$, the sum is over all values of x in the interval.

The key to evaluating the definite integral is provided by the **fundamental theorem of calculus**. The theorem states that, if $A(x)$ is an *indefinite* integral of $f(x)$, then the *definite integral* is given by

$$\int_{x_1}^{x_2} f(x)\,dx = A(x_2) - A(x_1) \equiv A(x)\Big|_{x_1}^{x_2}$$

Geometrically, the definite integral is the area under the graph of $f(x)$ between the limits x_1 and x_2:



Evaluating Integrals

The first step in evaluating an integral is to express all varying quantities within the integral in terms of a single variable; Chapter 9 outlines a general strategy for setting up an integral. Once you've set up an integral, you can evaluate it yourself or look it up in tables. Two common techniques can help you evaluate integrals or convert them to forms listed in tables:

1. **Change of variables**
   An unfamiliar integral can often be put into familiar form by defining a new variable. For example, it is not obvious how to integrate the expression

   $$\int \frac{x\,dx}{\sqrt{a^2 + x^2}}$$

   where a is a constant. But let $z = a^2 + x^2$. Then

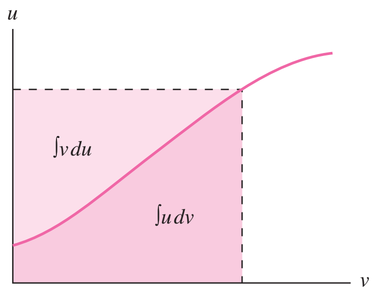
   $$\frac{dz}{dx} = \frac{da^2}{dx} + \frac{dx^2}{dx} = 0 + 2x = 2x$$

   so $dz = 2x\,dx$. Then the quantity $x\,dx$ in our unfamiliar integral is just $\frac{1}{2}dz$, while the quantity $\sqrt{a^2 + x^2}$ is just $z^{1/2}$. So the integral becomes

   $$\int \frac{1}{2} z^{-1/2}\,dz = \frac{\frac{1}{2}z^{1/2}}{\frac{1}{2}} = \sqrt{z}$$

   where we have used the standard form for the integral of a power of the independent variable. Substituting back $z = a^2 + x^2$ gives

   $$\int \frac{x\,dx}{\sqrt{a^2 + x^2}} = \sqrt{a^2 + x^2}$$



2. Integration by parts

The quantity $\int u\,dv$ is the area under the curve of u as a function of v between specified limits. In the figure, that area can also be expressed as the area of the rectangle shown minus the area under the curve of v as a function of u. Mathematically, this relation among areas may be expressed as a relation among integrals:

$$\int u\,dv = uv - \int v\,du \qquad \text{(integration by parts)}$$

This expression may often be used to transform complicated integrals into simpler ones.

Example

Evaluate $\int x\cos x\,dx$. Here let $u = x$, so $du = dx$. Then $dv = \cos x\,dx$, so we have $v = \int dv = \int \cos x\,dx = \sin x$. Integrating by parts then gives

$$\int x\cos x\,dx = (x)(\sin x) - \int \sin x\,dx = x\sin x + \cos x$$

where the $+$ sign arises because $\int \sin x\,dx = -\cos x$.

Table of Integrals

More extensive tables are available in many mathematical and scientific handbooks; see, for example, ***Handbook of Chemistry and Physics*** (Chemical Rubber Co.) or Dwight, ***Tables of Integrals and Other Mathematical Data*** (Macmillan). Some math software, including *Mathematica* and *Maple*, can also evaluate integrals symbolically. Wolfram Research provides *Mathematica*-based integration at *http://integrals.wolfram.com*.

In the expressions below, a and b are constants. An arbitrary constant of integration may be added to the right-hand side.

$$\int e^{ax}\,dx = \frac{e^{ax}}{a}$$

$$\int \sin ax\,dx = -\frac{\cos ax}{a}$$

$$\int \cos ax\,dx = \frac{\sin ax}{a}$$

$$\int \tan ax\,dx = -\frac{1}{a}\ln(\cos ax)$$

$$\int \sin^2 ax\,dx = \frac{x}{2} - \frac{\sin 2ax}{4a}$$

$$\int \cos^2 ax\,dx = \frac{x}{2} + \frac{\sin 2ax}{4a}$$

$$\int x\sin ax\,dx = \frac{1}{a^2}\sin ax - \frac{1}{a}x\cos ax$$

$$\int x\cos ax\,dx = \frac{1}{a^2}\cos ax + \frac{1}{a}x\sin ax$$

$$\int \frac{dx}{\sqrt{a^2 - x^2}} = \sin^{-1}\left(\frac{x}{a}\right)$$

$$\int \frac{dx}{\sqrt{x^2 \pm a^2}} = \ln\left(x + \sqrt{x^2 \pm a^2}\right)$$

$$\int \frac{dx}{x^2 + a^2} = \frac{1}{a}\tan^{-1}\left(\frac{x}{a}\right)$$

$$\int \frac{x\,dx}{\sqrt{a^2 - x^2}} = -\sqrt{a^2 - x^2}$$

$$\int \frac{x\,dx}{\sqrt{x^2 \pm a^2}} = \sqrt{x^2 \pm a^2}$$

$$\int \frac{dx}{(x^2 \pm a^2)^{3/2}} = \frac{\pm x}{a^2\sqrt{x^2 \pm a^2}}$$

$$\int xe^{ax}\,dx = \frac{e^{ax}}{a^2}(ax - 1)$$

$$\int x^2e^{ax}\,dx = \frac{x^2e^{ax}}{a} - \frac{2}{a}\left[\frac{e^{ax}}{a^2}(ax - 1)\right]$$

$$\int \frac{dx}{a + bx} = \frac{1}{b}\ln(a + bx)$$

$$\int \frac{dx}{(a + bx)^2} = -\frac{1}{b(a + bx)}$$

$$\int \ln ax\,dx = x\ln ax - x$$

The International System of Units (SI)

APPENDIX B

This material is from the U.S. edition of the English translation of the seventh edition of "Le Système International d'Unités (SI)," the definitive publication in the French language issued in 1991 by the International Bureau of Weights and Measures (BIPM). The year the definition was adopted is given in parentheses.

length (meter): The meter is the length of the path traveled by light in vacuum during a time interval of 1/299 792 458 of a second. (1983)

mass (kilogram): The kilogram is equal to the mass of the international prototype of the kilogram. (1889)

time (second): The second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom. (1967)

electric current (ampere): The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross section, and placed 1 meter apart in vacuum, would produce between these conductors a force equal to 2×10^{-7} newton per meter of length. (1948)

temperature (kelvin): The kelvin, unit of thermodynamic temperature, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water. (1967)

amount of substance (mole): The mole is the amount of substance of a system that contains as many elementary entities as there are atoms in 0.012 kilogram of carbon-12. (1971)

luminous intensity (candela): The candela is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540×10^{12} hertz and that has a radiant intensity in that direction of $(1/683)$ watt per steradian. (1979)

SI Base and Supplementary Units

	SI Unit	
Quantity	**Name**	**Symbol**
Base Unit		
Length	meter	m
Mass	kilogram	kg
Time	second	s
Electric current	ampere	A
Thermodynamic temperature	kelvin	K
Amount of substance	mole	mol
Luminous intensity	candela	cd
Supplementary Units		
Plane angle	radian	rad
Solid angle	steradian	sr

SI Prefixes

Factor	Prefix	Symbol
10^{24}	yotta	Y
10^{21}	zetta	Z
10^{18}	exa	E
10^{15}	peta	P
10^{12}	tera	T
10^{9}	giga	G
10^{6}	mega	M
10^{3}	kilo	k
10^{2}	hecto	h
10^{1}	deka	da
10^{0}	—	—
10^{-1}	deci	d
10^{-2}	centi	c
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f
10^{-18}	atto	a
10^{-21}	zepto	z
10^{-24}	yocto	y

Some SI Derived Units with Special Names

	SI Unit			
Quantity	**Name**	**Symbol**	**Expression in Terms of Other Units**	**Expression in Terms of SI Base Units**
Frequency	hertz	Hz		s^{-1}
Force	newton	N		$m \cdot kg \cdot s^{-2}$
Pressure, stress	pascal	Pa	N/m^2	$m^{-1} \cdot kg \cdot s^{-2}$
Energy, work, heat	joule	J	$N \cdot m$	$m^2 \cdot kg \cdot s^{-2}$
Power	watt	W	J/s	$m^2 \cdot kg \cdot s^{-3}$
Electric charge	coulomb	C		$s \cdot A$
Electric potential, potential difference, electromotive force	volt	V	J/C	$m^2 \cdot kg \cdot s^{-3} \cdot A^{-1}$
Capacitance	farad	F	C/V	$m^{-2} \cdot kg^{-1} \cdot s^4 \cdot A^2$
Electric resistance	ohm	Ω	V/A	$m^2 \cdot kg \cdot s^{-3} \cdot A^{-2}$
Magnetic flux	weber	Wb	$V \cdot s$	$m^2 \cdot kg \cdot s^{-2} \cdot A^{-1}$
Magnetic field	tesla	T	Wb/m^2	$kg \cdot s^{-2} \cdot A^{-1}$
Inductance	henry	H	Wb/A	$m^2 \cdot kg \cdot s^{-2} \cdot A^{-2}$
Radioactivity	becquerel	Bq	1 decay/s	s^{-1}
Absorbed radiation dose	gray	Gy	J/kg, 100 rad	$m^2 \cdot s^{-2}$
Radiation dose equivalent	sievert	Sv	J/kg, 100 rem	$m^2 \cdot s^{-2}$

Conversion Factors

APPENDIX C

The listings below give the SI equivalents of non-SI units. To convert from the units shown to SI, multiply by the factor given; to convert the other way, divide. For conversions within the SI system, see the table of SI prefixes in Appendix B, Chapter 1, or the inside front cover. Conversions that are not exact by definition are given to, at most, four significant figures.

Length

1 inch (in) $= 0.0254$ m
1 foot (ft) $= 0.3048$ m
1 yard (yd) $= 0.9144$ m
1 mile (mi) $= 1609$ m
1 nautical mile $= 1852$ m
1 angstrom (Å) $= 10^{-10}$ m
1 light year (ly) $= 9.46 \times 10^{15}$ m
1 astronomical unit (AU) $= 1.5 \times 10^{11}$ m
1 parsec $= 3.09 \times 10^{16}$ m
1 fermi $= 10^{-15}$ m $=$ 1 fm

Mass

1 slug $= 14.59$ kg
1 metric ton (tonne; t) $= 1000$ kg
1 unified mass unit (u) $= 1.661 \times 10^{-27}$ kg

Force units in the English system are sometimes used (incorrectly) for mass. The units given below are actually equal to the number of kilograms multiplied by g, the acceleration of gravity.

1 pound (lb) $=$ weight of 0.454 kg
1 ton $= 2000$ lb $=$ weight of 908 kg
1 ounce (oz) $=$ weight of 0.02835 kg

Time

1 minute (min) $= 60$ s
1 hour (h) $= 60$ min $= 3600$ s
1 day (d) $= 24$ h $= 86\,400$ s
1 year (y) $= 365.2422$ d* $= 3.156 \times 10^{7}$ s

Area

1 hectare (ha) $= 10^{4}$ m^2
1 square inch (in^2) $= 6.452 \times 10^{-4}$ m^2
1 square foot (ft^2) $= 9.290 \times 10^{-2}$ m^2
1 acre $= 4047$ m^2
1 barn $= 10^{-28}$ m^2
1 shed $= 10^{-30}$ m^2

Volume

1 liter (L) $= 1000$ cm^3 $= 10^{-3}$ m^3
1 cubic foot (ft^3) $= 2.832 \times 10^{-2}$ m^3
1 cubic inch (in^3) $= 1.639 \times 10^{-5}$ m^3
1 fluid ounce $= 1/128$ gal $= 2.957 \times 10^{-5}$ m^3
1 barrel $= 42$ gal $= 0.1590$ m^3
1 gallon (U.S.; gal) $= 3.785 \times 10^{-3}$ m^3
1 gallon (British) $= 4.546 \times 10^{-3}$ m^3

Angle, Phase

1 degree (°) $= \pi/180$ rad $= 1.745 \times 10^{-2}$ rad
1 revolution (rev) $= 360° = 2\pi$ rad
1 cycle $= 360° = 2\pi$ rad

* The length of the year changes very slowly with changes in Earth's orbital period.

Speed, Velocity

$1\ \text{km/h} = (1/3.6)\ \text{m/s} = 0.2778\ \text{m/s}$

$1\ \text{mi/h (mph)} = 0.4470\ \text{m/s}$

$1\ \text{ft/s} = 0.3048\ \text{m/s}$

$1\ \text{ly/y} = 3.00\times10^{8}\ \text{m/s}$

Angular Speed, Angular Velocity, Frequency, and Angular Frequency

$1\ \text{rev/s} = 2\pi\ \text{rad/s} = 6.283\ \text{rad/s}\ (\text{s}^{-1})$

$1\ \text{Hz} = 1\ \text{cycle/s} = 2\pi\ \text{s}^{-1}$

$1\ \text{rev/min (rpm)} = 0.1047\ \text{rad/s}\ (\text{s}^{-1})$

Force

$1\ \text{dyne} = 10^{-5}\ \text{N}$

$1\ \text{pound (lb)} = 4.448\ \text{N}$

Pressure

$1\ \text{dyne/cm}^2 = 0.10\ \text{Pa}$

$1\ \text{atmosphere (atm)} = 1.013\times10^{5}\ \text{Pa}$

$1\ \text{torr} = 1\ \text{mm Hg at } 0°\text{C} = 133.3\ \text{Pa}$

$1\ \text{bar} = 10^{5}\ \text{Pa} = 0.987\ \text{atm}$

$1\ \text{lb/in}^2\ \text{(psi)} = 6.895\times10^{3}\ \text{Pa}$

$1\ \text{in}\ H_2O\ (60°\text{F}) = 248.8\ \text{Pa}$

$1\ \text{in Hg}\ (60°\text{F}) = 3.377\times10^{3}\ \text{Pa}$

Energy, Work, Heat

$1\ \text{erg} = 10^{-7}\ \text{J}$

$1\ \text{calorie}^{*}\ \text{(cal)} = 4.184\ \text{J}$

$1\ \text{electronvolt (eV)} = 1.602\times10^{-19}\ \text{J}$

$1\ \text{foot-pound}\ (\text{ft}\cdot\text{lb}) = 1.356\ \text{J}$

$1\ \text{Btu}^{*} = 1.054\times10^{3}\ \text{J}$

$1\ \text{kWh} = 3.6\times10^{6}\ \text{J}$

1 megaton (explosive yield; Mt) $= 4.18\times10^{15}\ \text{J}$

Power

$1\ \text{erg/s} = 10^{-7}\ \text{W}$

$1\ \text{horsepower (hp)} = 746\ \text{W}$

$1\ \text{Btu/h (Btuh)} = 0.293\ \text{W}$

$1\ \text{ft}\cdot\text{lb/s} = 1.356\ \text{W}$

Magnetic Field

$1\ \text{gauss (G)} = 10^{-4}\ \text{T}$

$1\ \text{gamma}\ (\gamma) = 10^{-9}\ \text{T}$

Radiation

$1\ \text{curie (ci)} = 3.7\times10^{10}\ \text{Bq}$

$1\ \text{rad} = 10^{-2}\ \text{Gy}$

$1\ \text{rem} = 10^{-2}\ \text{Sv}$

Energy Content of Fuels

Energy Source	Energy Content
Coal	$29\ \text{MJ/kg} = 7300\ \text{kWh/ton} = 25\times10^{6}\ \text{Btu/ton}$
Oil	$43\ \text{MJ/kg} = 39\ \text{kWh/gal} = 1.3\times10^{5}\ \text{Btu/gal}$
Gasoline	$44\ \text{MJ/kg} = 36\ \text{kWh/gal} = 1.2\times10^{5}\ \text{Btu/gal}$
Natural gas	$55\ \text{MJ/kg} = 30\ \text{kWh/100 ft}^3 = 1000\ \text{Btu/ft}^3$
Uranium (fission)	
Normal abundance	$5.8\times10^{11}\ \text{J/kg} = 1.6\times10^{5}\ \text{kWh/kg}$
Pure U-235	$8.2\times10^{13}\ \text{J/kg} = 2.3\times10^{7}\ \text{kWh/kg}$
Hydrogen (fusion)	
Normal abundance	$7\times10^{11}\ \text{J/kg} = 3.0\times10^{4}\ \text{kWh/kg}$
Pure deuterium	$3.3\times10^{14}\ \text{J/kg} = 9.2\times10^{7}\ \text{kWh/kg}$
Water	$1.2\times10^{10}\ \text{J/kg} = 1.3\times10^{4}\ \text{kWh/gal} = 340\ \text{gal gasoline/gal}$
H_2O	
100% conversion, matter to energy	$9.0\times10^{16}\ \text{J/kg} = 931\ \text{MeV/u} = 2.5\times10^{10}\ \text{kWh/kg}$

*Values based on the thermochemical calorie; other definitions vary slightly.

The Elements

APPENDIX D

The atomic weights of stable elements reflect the abundances of different isotopes; values given here apply to elements as they exist naturally on Earth. For stable elements, parentheses express uncertainties in the last decimal place given. For elements with no stable isotopes (indicated in **boldface**), at most three isotopes are given; for elements 99 and beyond, only the longest-lived isotope is given. (Exceptions are the unstable elements thorium, protactinium, and uranium, for which atomic weights reflect natural abundances of long-lived isotopes.) See also the periodic table inside the back cover.

Atomic Number	Names	Symbol	Atomic Weight
1	Hydrogen	H	1.00794 (7)
2	Helium	He	4.002602 (2)
3	Lithium	Li	6.941 (2)
4	Beryllium	Be	9.012182 (3)
5	Boron	B	10.811 (5)
6	Carbon	C	12.011 (1)
7	Nitrogen	N	14.00674 (7)
8	Oxygen	O	15.9994 (3)
9	Fluorine	F	18.9984032 (9)
10	Neon	Ne	20.1797 (6)
11	Sodium (Natrium)	Na	22.989768 (6)
12	Magnesium	Mg	24.3050 (6)
13	Aluminum	Al	26.981539 (5)
14	Silicon	Si	28.0855 (3)
15	Phosphorus	P	30.973762 (4)
16	Sulfur	S	32.066 (6)
17	Chlorine	Cl	35.4527 (9)
18	Argon	Ar	39.948 (1)
19	Potassium (Kalium)	K	39.0983 (1)
20	Calcium	Ca	40.078 (4)
21	Scandium	Sc	44.955910 (9)
22	Titanium	Ti	47.88 (3)
23	Vanadium	V	50.9415 (1)
24	Chromium	Cr	51.9961 (6)
25	Manganese	Mn	54.93805 (1)
26	Iron	Fe	55.847 (3)
27	Cobalt	Co	58.93320 (1)
28	Nickel	Ni	58.69 (1)
29	Copper	Cu	63.546 (3)
30	Zinc	Zn	65.39 (2)
31	Gallium	Ga	69.723 (1)
32	Germanium	Ge	72.61 (2)
33	Arsenic	As	74.92159 (2)

(continued)

Atomic Number	Names	Symbol	Atomic Weight
34	Selenium	Se	78.96 (3)
35	Bromine	Br	79.904 (1)
36	Krypton	Kr	83.80 (1)
37	Rubidium	Rb	85.4678 (3)
38	Strontium	Sr	87.62 (1)
39	Yttrium	Y	88.90585 (2)
40	Zirconium	Zr	91.224 (2)
41	Niobium	Nb	92.90638 (2)
42	Molybdenum	Mo	95.94 (1)
43	**Technetium**	**Tc**	**97, 98, 99**
44	Ruthenium	Ru	101.07 (2)
45	Rhodium	Rh	102.90550 (3)
46	Palladium	Pd	106.42 (1)
47	Silver	Ag	107.8682 (2)
48	Cadmium	Cd	112.411 (8)
49	Indium	In	114.82 (1)
50	Tin	Sn	118.710 (7)
51	Antimony (Stibium)	Sb	121.75 (3)
52	Tellurium	Te	127.60 (3)
53	Iodine	I	126.90447 (3)
54	Xenon	Xe	131.29 (2)
55	Cesium	Cs	132.90543 (5)
56	Barium	Ba	137.327 (7)
57	Lanthanum	La	138.9055 (2)
58	Cerium	Ce	140.115 (4)
59	Praseodymium	Pr	140.90765 (3)
60	Neodymium	Nd	144.24 (3)
61	**Promethium**	**Pm**	**145, 147**
62	Samarium	Sm	150.36 (3)
63	Europium	Eu	151.965 (9)
64	Gadolinium	Gd	157.25 (3)
65	Terbium	Tb	158.92534 (3)
66	Dysprosium	Dy	162.50 (3)
67	Holmium	Ho	164.93032 (3)
68	Erbium	Er	167.26 (3)
69	Thulium	Tm	168.93421 (3)
70	Ytterbium	Yb	173.04 (3)
71	Lutetium	Lu	174.967 (1)
72	Hafnium	Hf	178.49 (2)
73	Tantalum	Ta	180.9479 (1)
74	Tungsten (Wolfram)	W	183.85 (3)
75	Rhenium	Re	186.207 (1)
76	Osmium	Os	190.2 (1)
77	Iridium	Ir	192.22 (3)
78	Platinum	Pt	195.08 (3)
79	Gold	Au	196.96654 (3)
80	Mercury	Hg	200.59 (3)
81	Thallium	Tl	204.3833 (2)
82	Lead	Pb	207.2 (1)
83	Bismuth	Bi	208.98037 (3)

Atomic Number	Names	Symbol	Atomic Weight
84	Polonium	Po	209, 210
85	Astatine	At	210, 211
86	Radon	Rn	211, 220, 222
87	Francium	Fr	223
88	Radium	Ra	223, 224, 226
89	Actinium	Ac	227
90	Thorium	Th	232.0381 (1)
91	Protactinium	Pa	231.03588 (2)
92	Uranium	U	238.0289 (1)
93	Neptunium	Np	237, 239
94	Plutonium	Pu	239, 242, 244
95	Americium	Am	241, 243
96	Curium	Cm	245, 247, 248
97	Berkelium	Bk	247, 249
98	Californium	Cf	249, 250, 251
99	Einsteinium	Es	252
100	Fermium	Fm	257
101	Mendelevium	Md	258
102	Nobelium	No	259
103	Lawrencium	Lr	262
104	Rutherfordium	Rf	263
105	Dubnium	Db	268
106	Seaborgium	Sg	266
107	Bohrium	Bh	272
108	Hassium	Hs	277
109	Meitnerium	Mt	276
110	Darmstadtium	Ds	281
111	Roentgenium	Rg	280
112	Copernicium	Cn	285
113	—	—	284
114	—	—	289
115	—	—	288
116	—	—	292
117	—	—	294
118	—	—	294

APPENDIX E

Astrophysical Data

Sun, Planets, Principal Satellites

Body	Mass (10^{24} kg)	Mean Radius (10^6 m Except as Noted)	Surface Gravity (m/s^2)	Escape Speed (km/s)	Sidereal Rotation Period* (days)	Mean Distance from Central Body† (10^6 km)	Orbital Period	Orbital Speed (km/s)
Sun	1.99×10^6	696	274	618	36 at poles 27 at equator	2.6×10^{11}	200 My	250
Planets								
Mercury	0.330	2.44	3.70	4.25	58.6	57.6	88.0 d	48
Venus	4.87	6.05	8.87	10.4	−243	108	225 d	35
Earth	5.97	6.37	9.81	11.2	0.997	150	365.3 d	30
Moon	0.0735	1.74	1.62	2.38	27.3	0.385	27.3 d	1.0
Mars	0.642	3.38	3.74	5.03	1.03	228	1.88 y	24.1
Phobos	9.6×10^{-9}	9–13 km	0.001	0.008	0.32	9.4×10^{-3}	0.32 d	2.1
Deimos	2×10^{-9}	5–8 km	0.001	0.005	1.3	23×10^{-3}	1.3 d	1.3
Jupiter	1.90×10^3	69.1	26.5	60.6	0.414	778	11.9 y	13.0
Io	0.0888	1.82	1.8	2.6	1.77	0.422	1.77 d	17
Europa	0.479	1.57	1.3	2.0	3.55	0.671	3.55 d	14
Ganymede	0.148	2.63	1.4	2.7	7.15	1.07	7.15 d	11
Callisto	0.107	2.40	1.2	2.4	16.7	1.88	16.7 d	8.2
and 13 smaller satellites								
Saturn	569	56.8	11.8	36.6	0.438	1.43×10^3	29.5 y	9.65
Tethys	0.0007	0.53	0.2	0.4	1.89	0.294	1.89 d	11.3
Dione	0.00015	0.56	0.3	0.6	2.74	0.377	2.74 d	10.0
Rhea	0.0025	0.77	0.3	0.5	4.52	0.527	4.52 d	8.5
Titan	0.135	2.58	1.4	2.6	15.9	1.22	15.9 d	5.6
and 12 smaller satellites								
Uranus	86.6	25.0	9.23	21.5	−0.65	2.87×10^3	84.1 y	6.79
Ariel	0.0013	0.58	0.3	0.4	2.52	0.19	2.52 d	5.5
Umbriel	0.0013	0.59	0.3	0.4	4.14	0.27	4.14 d	4.7
Titania	0.0018	0.81	0.2	0.5	8.70	0.44	8.70 d	3.7
Oberon	0.0017	0.78	0.2	0.5	13.5	0.58	13.5 d	3.1
and 11 smaller satellites								
Neptune	103	24.0	11.9	23.9	0.768	4.50×10^3	165 y	5.43
Triton	0.134	1.9	2.5	3.1	5.88	0.354	5.88 d	4.4
and 7 smaller satellites								
Dwarf Planets								
Ceres	0.00095	0.487	0.27	0.51	0.38	416	4.60 y	17.9
Pluto	0.013	1.15	0.58	1.2	−6.39	5.92×10^3	248 y	4.7
Charon	0.001	0.6			−6.39	0.018	6.39 d	0.2
and 2 smaller satellites, Nix and Hydra								
Eris	0.0167	1.2	0.8	1.3	0.3	1.0×10^4	557 y	3.44
and 1 small satellite, Dysnomia								

* Negative rotation period indicates retrograde motion, in opposite sense from orbital motion. Periods are sidereal, meaning the time for the body to return to the same orientation relative to the distant stars rather than the Sun.

†Central body is galactic center for Sun, Sun for planets, and planet for satellites.

Answers to Odd-Numbered Problems

Chapter 1

11. 10^5
13. T = 108.783 ps
15. 10^8
17. 0.62 rad = 35°
19. 30 g
21. 10^6
23. 8.6 m^2/L
25. 3.6 km/h
27. 57.3°C
31. 4×10^6
33. 41 m
35. (a) 5.18 (b) 5.20
37. 3×10^6
39. ~0.15%
41. (a) $\sim3\times10^3$ m^3 (b) ~100 days
43. 10^5
45. ~250 μm
47. (a) 0.1 μm (b) 2×10^5 calculations per second
49. $\Delta = 100(\pm0.05/N)\%$
51. 12¢ more in Canada than in the U.S.
53. (a) 1.0 m (b) 0.001 m^2 (c) 0.0 m (d) 1.0
57. c
59. c

Chapter 2

13. (a) 372 yd/min (b) 5 min, 46 s
15. 21 h
17. (a) 3.0×10^4m/s (b) 19 mi/s
21. (a) $v = b - 2ct$ (b) 8.4 s
23. 0.35 m/s^2
25. Falling: 9.82 m/s^2, stopping: −84.0 m/s^2
27. 17 m/s^2
29. $v = dx/dt = d/dt(x_0 + v_0t + at^2/2) = v_0 + at$
31. (a) 46 m/s^2 (b) 61 s
33. 27 ft/s^2
35. 1 × 104 m/s
37. 95 m
39. (a) 250 m (b) 39 m/s, 40 m (c) 9.8 m/s, 100 m; (d) −20 m/s, 100 m
41. 11 m/s
43. 48 mi/h
45. (a) 80 km/h (b) 50 h
47. 2.6 h, 2800 km from San Francisco or 2000 km from New York
49. $\bar{v} = \Delta x/\Delta t = bt^4/t = bt^3 = v(t)/4$
51. (a) 28.1 m/s (b) 22.5 m/s^2 (c) 9.38 m/s (d) 11.3 m/s^2
53. 55%
55. (a) 0.014 s (b) 51 cm
57. 0.89 km
59. (a) 25 m/s (b) 180 m
61. 0.0051 m/s^2
63. 11 m/s
65. 270 m
67. $-\frac{1}{2}\sqrt{hg}$
69. (a) −7.67 m/s (b) 0.162 s
71. 3.9 s, 6.2 m/s
73. 36 km/h
75. (a) $\bar{v} = (v_1 + v_2)/2$ (b) $\bar{v} = (2v_1v_2)/(v_1 + v_2)$
77. 70.7 %
79. −0.3 m/s
81. $\frac{h}{4}\left(\frac{2h}{g\Delta t^2}\right)\left(\frac{g\Delta t^2}{2h} - 1\right)^2$
83. 15 s^{-1}
87. c
89. c

Chapter 3

11. 270 m, 150°
13. 700 km, 110°
15. $105\hat{\imath} + 58\hat{\jmath}$ km
17. 1.414, $\theta = 45°$
19. (−14 m/s, −12 m/s)
21. $3ct^2\hat{\imath}$
23. (a) $\vec{v} = (-2.2\times10^{-6}\text{ m/s})\hat{\jmath}$ (b) $\vec{a} = (-3.2\times10^{-10}\text{ m/s}^2)\hat{\imath}$
25. $\vec{v}_2 = (1.3\text{ m/s})\hat{\imath} + (2.3\text{ m/s})\hat{\jmath}$
33. (a) 1.3 s (b) 15 m
35. 34 nm
37. 1090 m
39. 2.8 mm/s^2
41. (a) $A\sqrt{5}$ (b) $A\sqrt{10}$
43. $\vec{C} = -15\hat{\imath} + 9\hat{\jmath} - 18\hat{k}$
45. (a) $4c/3d$ (b) $c/3d$
47. −5.7 m/s^2
49. (a) $0.22\hat{\imath} + 0.13\hat{\jmath}$ m/s (b) $(-4.4\hat{\imath} + 7.6\hat{\jmath})\times10^{-4}$ m/s^2
51. $A = B$
53. 0.50 m/s^2
55. 5.7 m/s
57. (a) $x_1 = x_2$ implies $y_1 = h\left(1 - \frac{gh}{v_0^2}\right) = y_2$ (b) $v_0 \geq \sqrt{gh}$
59. 8.3 m/s, 61°
65. Yes
67. 66°
69. 2.3 km
71. $2h$
73. 19 m
75. $dx/d\theta_0 = 2v_0^2/g\cos(2\theta_0) = 0 \Rightarrow \theta_0 = 45°$.
79. $\frac{1}{2}\cos^{-1}(1/(1 + v_0^2/gh))$
81. c
83. c

Chapter 4

13. (a) 2.0 m/s^2 (b) 0.082 m/s^2
15. −13 kN
17. 2.0×10^6 m/s^2
19. 4.1 cm
23. 210 kg
25. 9000 kg
27. 490 N
29. 380 N
33. 55 kN
35. 130 N
37. 1.2 cm
39. $(-2.94\text{ m/s}^2)\hat{\imath}$
41. 4.9 m/s^2
43. 0.53 s
45. 6.0 N to the right
47. (a) 5.26×10^3 N (b) $(-1.08\times10^3\text{ N})\hat{\imath}$ (c) $(-494\text{ N})\hat{\imath}$ (d) $(590\text{ N})\hat{\imath}$
49. 680 m
51. 0.96 m
53. 950 N
55. (a) −0.40 mg (b) 2.40 mg (c) 1.40 mg
57. F-16: yes, at 1.2 m/s^2; A-380: no
61. 1.96 m/s^2
63. 11.8 m/s^2
65. 0.92 kg, 1.4 kg
67. $\omega F_0/M$
69. a
71. b

Chapter 5

13. $(4.0\text{ N})\hat{\imath} + (1.7\text{ N})\hat{\jmath}$
15. 26 s
17. 6.4 kN in each muscle
19. $m_R/m_L = 2.5$
21. (a) 3.9 m/s^2 (b) 530 N
25. The train exceeded the speed limit by 22 km/h.

27. 490 km/h
29. 0.18
31. 0.53
33. 0.43 m
35. $T = 530$ N ~ 3 times the monkey's weight
37. $T = m_2 g, \tau = 2\pi\sqrt{(m_1 R)/(m_2 g)}$
39. 310 N downward (b) $-m_{SB}v^2/R$ (c) nothing
41. 8.5 km
43. 0.21
45. 25 s
47. Yes
49. $0.23 \le \mu_s \le 0.30$
51. 4.2 m/s^2
53. 0.62
55. (a) 10 cm (b) no
57. 100 km/h
61. 17 min^{-1}
63. Brake, don't swerve
65. 28 cm
67. $T' = u_k/\sqrt{1 + \mu_k^2}$
71. Yes
73. 7.6 km
75. a
77. c

Chapter 6

11. 900 J
13. 150 kJ
15. 190 MN
17. $\vec{A}\cdot(\vec{B}+\vec{C}) = AB\cos(\theta_{AB}) + AC\cos(\theta_{AC}) = \vec{A}\cdot\vec{B} + \vec{A}\cdot\vec{C}$
19. 1.9 m
21. (a) 1 J (b) 3 J
23. 30 cm
25. 7.5 GJ
27. ±120 km/h
29. 110 m/s
31. 5.53 W
33. (a) 60 kW (b) 1 kW (c) 41.7 W
35. 9.4×10^6 J
37. 0 W
39. 22 s
41. (a) 400 J (b) 31 kg
43. (a) 76,000 (b) 14 kW
45. 25°
47. (a) 0 (b) 90°
49. 622 J
51. $k_B = 8k_A$
53. $W = F_0\left(x - \dfrac{x^2}{2L_0} + \dfrac{L_0^2}{L_0 + x} - L_0\right)$
55. $v_2 = \pm 2v_1$
57. $W = F_0 x_0/2$
59. 70.5°
61. 490×10^{12} gal/day
63. 26 m/s
65. 0.60
67. 8.0 kJ
69. 42 kJ
71. 6.0 years
75. $W_{x_1\to x_2} = 2b(\sqrt{x_2} - \sqrt{x_1})$, $W(x_1 = 0) = 2b\sqrt{x_2}$
77. (a) $\frac{1}{2}kL_0^2 + \frac{1}{3}bL_0^3 + \frac{1}{4}cL_0^4 + \frac{1}{5}dL_0^5$ (b) 12 kJ
81. Stopping force is 35 times the weight of leg
83. c
85. c

Chapter 7

11. $W_a = W_b = -mgL$
13. (a) 1.3 MJ (b) −59 kJ
15. 840 m
17. 55 cm
19. ±22 m/s, ±35 m/s
21. 92 m/s
23. 2.3 kN/m
25. ±2.0 m
27. (a) 67 N (b) 0.0 N (c) 45 N
29. $W_a = 0; W_b = F_0 a$
31. (a) 1.07 J (b) 1.12 J
33. 778 J, 4.90%
35. 2.5 J
37. $U(x) = -\frac{1}{3}ax^3 - bx$
39. $r = \dfrac{kx^2}{2mg\sin(\theta)}$
43. (a) −11 cm (b) ±4 m/s
45. $h \ge 5R/2$
49. (a) $U(x) = -\dfrac{a}{2}x^2 + \dfrac{b}{4}x^4$ (b) 0.7 and 2 m
51. 20 m/s, 30 m/s
53. 1.4 m
55. 62.5 cm
57. 2.9 m
59. 14 m
61. $v = 2x^{3/4}\sqrt{\dfrac{a}{3m}}$
63. Yes, the block reaches top with 1.6 J of kinetic energy.
65. 2.6×10^7 N/m
67. $\frac{1}{2}kx_2^2 + \frac{1}{2}ka(x_2^2 - x_1^2)$
69. d
71. b

Chapter 8

11. $R_P = R_E/\sqrt{2}$
13. 57.5%
15. 8.6 kg
17. 442 m
19. 3070 m/s
21. 1.77 d
23. 0.28×10^6 m
25. 3.17 GJ
27. 4.29 km/s
29. -2.64×10^{33} J
31. (a) 2.44 km/s (b) 2.10×10^8 m/s
33. 10 m/s^2
35. $g(h)/g(0) = 0.414$
37. 2.73×10^{-3} m/s^2, $a_c/g = 2.78\times 10^{-4}$
39. 60.5 min
41. 2.6×10^{41} kg
43. $T^2 = \dfrac{4\pi^2 L^3}{3GM}$
45. 2.47 AU
47. 4.48 km/s
51. The comet is going faster than the escape velocity from the Sun, so it will not return to Earth's vicinity.
55. (a) 2.06×10^6 m (b) 0.805×10^6 m
57. (a) 4.59 km/s (b) 14.2 km/s
59. 4.17 km/s
61. 4.60×10^{10} m
63. 1.42×10^3 km
65. 41.9 cm/c
67. No danger, since the puck needs at least 6100 km/h to go into orbit.
69. 1.5×10^6 km
71. d
73. d

Chapter 9

13. $2m$
15. $(0, 0.289L)$
17. $\vec{v}_2 = (-67 \text{ cm/s})\hat{\imath}$
19. $\vec{v}_{U^{235}} = (-2.68\times 10^5 \text{ m/s})\hat{\imath}$
21. 1.21 J
23. The impulse imparted by gravity is 0.08% of the collision impulse.
25. 41.8 s
31. 46 m/s
33. 11 Mm/s and −11 Mm/s
35. $(0, 0.115a)$
37. $\vec{r}_{cm} = \left(t^2 + \frac{10}{3}t + \frac{7}{3}\right)\hat{\imath} + \left(\frac{2}{3}t + \frac{8}{3}\right)\hat{\jmath}$; $\vec{v}_{cm} = \left(2t + \frac{10}{3}\right)\hat{\imath} + \left(\frac{2}{3}\right)\hat{\jmath}$; $\vec{a}_{cm} = (2)\hat{\imath}$
39. $m_b = 4m_m$
41. $(0, 0, h/4)$
43. (a) 0.99 m (b) 3.9 m/s
45. (a) $\vec{a}_c = \dfrac{v_0}{M}\left(\dfrac{dm}{dt}\right)\hat{\imath}$ (b) v_0
49. (a) (0, 0, 13 m) (b) (0, 0, 11 m)
51. $\vec{v}_3 = (4.4 \text{ m/s})\hat{\imath} + (3.0 \text{ m/s})\hat{\jmath}$
53. 9.3 m/s
55. $\frac{2}{5}v; \frac{7}{5}v$
57. (a) 37.7° (b) −65.8 cm/s
59. 5.8 s
61. 0.92 m/s
65. 120°
67. 5.83
71. 18.6%
73. $J = 2F_0/a$
75. (a) 8.10 m (b) −12.6 m/s
77. $v_1 = v/6, v_2 = 5v/6$
79. 8.3 kg
81. The peak force of 327 kN occurs at 165 ms.
85. The center of mass lies along line through the middle of the slice, at a distance of $(4R/3\theta)\sin\left(\frac{1}{2}\theta\right)$ from the tip.
87. 3.75 min
89. (a) $\dfrac{M}{1+a}$; (b) $\dfrac{1+a}{2+a}L$ (c) M and $\frac{1}{2}L$
91. b
93. a

Chapter 10

13. (a) $7.27\times10^{-5}\ s^{-1}$ (b) $1.75\times10^{-3}\ s^{-1}$ (c) $1.45\times10^{-4}\ s^{-1}$ (d) $31.4\ s^{-1}$
15. (a) 75 rad/s (b) 2.4×10^{-4} rad/s (c) 6×10^{3} rad/s (d) 2×10^{-7} rad/s
17. $7\times10^{-2}\ s^{-1}$
19. (a) 0.16 rev (b) 0.07 rad/s
21. 1.2 m
23. 7.9×10^{-2} N·m
25. (a) $2mL^2$ (b) mL^2
27. (a) $4.4\times10^{-4}\ kg\cdot m^2$ (b) 1.5×10^{-3} N·m
29. (a) mL^2 (b) $\frac{1}{2}mL^2$
31. (a) $1\times10^{38}\ kg\cdot m^2$ (b) -6×10^{33} N·m
33. 20 min
35. 750,000 y
37. (a) 1.6×10^{8} J (b) 16 MW
39. 1/2
41. (a) 6.9 rad/s (b) 3.7 s
43. (a) 1.1 rad/s (b) 1.1 m/s
45. (a) $170\ s^{-2}$ (b) $2.9\ m/s^2$ (c) 150 revolutions
47. 570 rev
49. (a) $2ML^2/3$ (b) $2ML^2/3$ (c) $4ML^2/3$
51. $Ma^2/3$
53. 33 pN
55. (a) 7.2 h (b) 1900 rev
57. 0.36
59. ±2.1 rad/s
61. $v = \sqrt{\frac{6}{5}gd\sin\theta}$
63. 17%
65. 0.494 MR^2
67. 33 m
69. (a) $M = \dfrac{2\pi\rho_0 wR^2}{3}$ (b) $I = 3MR^2/5$
73. $3MR^2/10$
75. $\tau = \frac{1}{2}MGL\sin\theta$
77. The specs are incorrect. The storage capacity is 3 MJ below what's claimed.
79. a
81. b
83. c

Chapter 11

13. $\vec{\omega} = 63\ s^{-1}$ west
15. (a) $1.1\times10^{6}\ s^{-2}$ (b) $-37°$
17. (a) $(-12\ N\cdot m)\hat{k}$ (b) $(36\ N\cdot m)\hat{k}$ (c) $(12\ N\cdot m)\hat{\imath} + (36\ N\cdot m)\hat{\jmath}$
19. 3.1 N·m, out of the page
21. $4.1\ kg\cdot m^2\cdot s^{-1}$
23. 2.3 J·s along axis
25. 17.4 rpm
27. 2.5 days.
29. $(-9.0\ N\cdot m)\hat{k}$
31. 1600 N·m
33. 37 J·s
35. 3.1×10^{-16} J·s
37. $0.21\ kg\cdot m^2$
39. 63%
41. 5.5 m/s
43. 3.1 rpm
45. (a) $d\omega\left(\frac{1}{2} - I/2md^2\right)$ (b) $d\omega$ (c) $d\omega(2 + I/md^2)$
47. 2.8%, orbital angular momentum
49. (a) 140 rpm (b) 27%
51. (a) $2\omega_0/7$ (b) $t = \dfrac{2R\omega_0}{\mu_k g}$
55. d
57. d

Chapter 12

17. (a) $\tau = mgL/2$ (b) $\tau = 0$ (c) $\tau = -mgL/2$
19. 16 m relative to the wall
21. (a) 0.61 m from left end (b) 1.42 m from left end
23. 480 N
25. -0.797 m, unstable; 1.46 m, stable
27. (a) 40 N·m (b) 1.3 kN
31. 79 kg
33. 1.4 W
37. 1.0×10^{3} kg
39. 63.4°, unstable
41. Slide
43. 74 kg
45. 0.366 mgs
49. $F_{app} = Mg\tan(\theta/2)$
51. $\mu_s < \tan\alpha = 1/2$
55. $\mu \geq \dfrac{\tan\theta}{2 + \tan^2\theta}$
57. 840 N
61. (a) $F = G\dfrac{M_E m}{R_E^2}(1.229)$, 21.3° (b) $\tau = G\dfrac{M_E m}{R_E}(-0.0356)$
63. The tie beam will not hold under the 10 kN of tension.
65. a
67. b

Chapter 13

17. 2.27×10^{-3} s
19. (a) $x(t) = (10\ cm)\cos[(10\pi s^{-1})t]$ (b) $x(t) = (2.5\ cm)\sin[(5\ s^{-1})t]$
21. 22 ms
23. 0.59 Hz; 1.7 s
25. (a) $4.3\ s^{-1}$ (b) 0.92 N/m (c) 0.82 m
27. (a) 2.2 rad/s (b) 2.8 s (c) 0.63 m
29. 1.21 s
31. 1.6 s
33. 7 oscillations in x direction for 4 oscillations in the y direction
35. ±1.7 rad, ±15 rad/s
37. 0.25 s
39. 65 km/h
41. 0.70 s
43. (a) $t = \pi\sqrt{m/k}$ (b) $A = v_0\sqrt{m/k}$
45. 50 min
47. (a) 67 μN/m (b) 3.4×10^{-10} kg
51. 800 kg
53. (a) $|\vec{r}| = A$ (b) $\vec{v} = \omega A\cos\omega t\hat{\imath} - \omega A\sin\omega t\hat{\jmath}$. (c) $|\vec{v}| = \omega A$ (d) ω
55. 0.147%
57. (a) 1.3 N/m (b) 0.80 kg
59. $\omega = \sqrt{(k_1 + k_2)/m}$
65. 34
67. (a) 6.5 cm (b) 0.51 s
69. $f = \dfrac{1}{2\pi}\sqrt{2a/m}$
71. (a) $E_1 = 4E_2$ (b) $a_{max,1} = 4a_{max,2}$
73. 27°
75. $T = 2\pi\sqrt{7/(10ga)}$
77. 0.54 Hz; 22 cm; -0.11 rad
83. $2.1\ m/s^2$
85. c
87. c

Chapter 14

17. (a) 0.19 s (b) 6.5 cm
19. (a) 300 m (b) 1.58 m (c) 3.0 cm (d) 8 μm (e) 500 nm (f) 3.0 Å
21. (a) 0.19 mm (b) 0.43 mm
23. (a) 1.3 cm (b) 9.1 cm (c) $0.20\ s^{-1}$ (d) 45 cm/s (e) $-x$ direction
25. (a) $14\ s^{-1}$ (b) $0.39\ cm^{-1}$ (c) $(2.5\ cm)\cos[(0.39\ cm^{-1})x + (14\ s^{-1})t]$
27. 250 m/s
29. 7.6 N
31. 9.9 W
33. 343 m/s
35. 420 m/s
37. 940 Hz
39. 5.4 m
41. (a) 280 Hz (b) 70 Hz (c) 210 Hz
43. 14 cm
45. 93 Hz
47. Galaxy receding
49. 30 m/s
51. $\overline{E} = \dfrac{4\pi^2 FA^2}{\lambda}$
53. 1.0×10^{2} W
57. $v = \sqrt{\dfrac{kL(L - L_0)}{m}}$
59. 10 m.
61. $L_0 = 5L_1/7$
63. 440 mph
67. 6.3 m
73. 90 km/h
75. 7.3 km
77. 41 m/s
81. Not sufficient: The minimum measurable speed is 5.4 km/h.
83. 3.9 kg
85. a
87. d

Chapter 15

15. 1.2 kg
17. (a) $180\ kg/m^3$ (b) $7.3\ m^3$
19. 249 kPa
21. 600 kPa
23. $1.7\times10^{3}\ kg/m^3$
25. 92 m

27. 7.4 kPa
29. 46 kg
31. 0.75%
33. 2.8 m/s
35. (a) $1.8 \times 10^4\ \mathrm{m^3/s}$ (b) 1.5 m/s
37. 1.8 cm/s
39. 830 $\mathrm{cm^2}$
41. (a) 620 Pa (b) 1.2 kPa
43. 3.6 mm
45. 8100 kg
47. The accused apparently drank 51 oz.
49. 27 m
51. (a) 49 kg (b) 2500 kg
53. 14 kPa
55. 14 m
57. (a) 1.5 m/s (b) 0.47 L/s
59. 70%
61. (a) 98% less (b) 17 cm
63. 15 kg
65. Yes, the wind farm could produce 1-GW of power.
67. $t = \dfrac{A_0}{A_1}\sqrt{\dfrac{2h}{g}}$
69. (b) 5.8 km
70. (a) $p(r, 0) = p_a + \rho g h_0 + \rho\omega^2 r^2/2$
    (b) $h(r) = h_0 + \omega^2 r^2 = 2g$
73. 2.1×10^{12} N·m
75. Yes
77. $\rho_{H_2O} L \tan\dfrac{\theta}{2}(h_0^2 - h_1^2)$
79. c
81. e

Chapter 16

15. 20°C
17. $-40°\mathrm{C} = -40°\mathrm{F}$
19. 102°F
21. 32 kJ
23. 100 W
25. (a) 170 J/K (b) 480 J/(kg·K)
27. 0.293 W
29. 55 kW
31. 4 W
33. $\mathcal{R}_{\text{air}} = 0.98\ \mathrm{m^2{\cdot}K/W}$,
    $\mathcal{R}_{\text{concrete}} = 0.03\ \mathrm{m^2{\cdot}K/W}$,
    $\mathcal{R}_{\text{fiberglass}} = 0.60\ \mathrm{m^2{\cdot}K/W}$,
    $\mathcal{R}_{\text{glass}} = 0.03\ \mathrm{m^2{\cdot}K/W}$,
    $\mathcal{R}_{\text{Styrofoam}} = 0.88\ \mathrm{m^2{\cdot}K/W}$,
    $\mathcal{R}_{\text{pine}} = 0.23\ \mathrm{m^2{\cdot}K/W}$
35. 2.2 kW
37. $2 \times 10^{-5}\ \mathrm{m^2}$
39. (a) 138 kPa (b) 33.4 kPa (c) 233 kPa
41. 263 K = −10°C
43. 364 g
45. (a) 23.2 kJ (b) 337 kJ (c) 65.2 kJ
47. 138 s
49. 0.56 kg
51. 1.8 kg
53. 9.2 K
55. 0.20 kg
57. 2.0×10^2 W
59. The house will remain at a comfortable 19°C
61. (a) 1200 K (b) 700 K
63. 24°C
65. 1200 K
67. (a) $200/month (b) $24/month
69. 44 K
71. The solar increase accounts for only 4% of recent warming.
73. The hutch temperature will be −2.5°C, so the water will freeze.
77. 37°F for min; 60°F for max
79. c
81. c

Chapter 17

17. 1.8 $\mathrm{m^3}$
19. 1.8×10^6 Pa
21. (a) 27 L (b) 330 K
23. 3.16 km/s
25. 22 kJ
27. 3.9 kg
29. 6.0 MJ
31. 0.987 L
33. 263°C
35. $1 \times 10^{15}\ \mathrm{m^{-3}}$, which is over 10 billion times less dense than Earth's atmosphere
37. (a) 235 mol (b) 5.65 $\mathrm{m^3}$
39. (a) 1.27 atm (b) 0.980 mol
    (c) 0.786 atm
41. 27.6 min
43. 14 min
45. 43.9 min
47. 10°C
49. 46.1°C
51. 177 g
53. 4.9°C
55. 19 kW
57. 56 min
59. 251 K
61. 307 K
63. $d = \dfrac{L_0}{2}\sqrt{2\alpha\Delta T + \alpha^2\Delta T^2}$
65. (a) 61 h (b) 52 h
67. 3.97°C
69. 34.1 km
73. b
75. a

Chapter 18

15. 18 kJ
17. 250 J
19. −14 kW
21. $2p_1V_1$
23. (a) 4/3 (b) 220 J
25. 0.177
27. 57.7%
29. (a) 200 K (b) 120 K
31. 380 W
33. (a) 1.49 mm (b) 10.7 μJ
35. 1.35
37. (a) 300 kPa (b) 240 J
39. 440°C
41. (a) 810 K (b) 25.8 atm
43. (a) −1.5 kJ, 300 K (b) 0 J, 336 K
    (c) 430 J, 326 K
45. (a) 40 kPa (b) 83 kPa (c) 80 kJ
47. 930 J
51. The temperature rises 75°C, missing the criteria.
53. 57 kJ
55. 330 K
57. (a) 202 J (b) 500 J transferred out of the gas
59. 20 mol
61. 140 atm
67. 2.0 MJ
69. $4p_1V_1/3$
71. Yes
73. d
75. d

Chapter 19

13. (a) 26.8% (b) 7.05% (c) 77.0%
15. 0.948 K
17. 9.10
19. No
21. 8.8 kJ/K
23. 21.9 kg
25. (a) 1/64 (b) 20/64
27. 52.1% (winter), 47.7% (summer)
29. (a) 1.75 GW (b) 43.0%
    (c) 232°C
31. 2×10^{11} kg/s
33. (a) 8.53 (b) 1.10×10^3 kg
35. $58
37. 2.83
39. (a) 5.7 (b) 3.5 kW (c) pump: 54¢/h; oil furnace; $1.73/h
41. (a) 17.4% (b) 83.3%
43. 140 MJ/K
47. (a) 53 J/K (b) 74 J/K (c) −8.5 J/K
51. 160 J/K
63. 36.2 J/K
65. d
67. b

Chapter 20

13. 3 C, or about 0.05 C/kg
15. (a) *uud* (b) *udd*
17. (a) $\hat{j}$ (b) $-\hat{\imath}$ (c) $0.316\hat{\imath} + 0.949\hat{j}$
19. (a) $\hat{j}$ (b) $-\hat{\imath}$ (c) $0.316\hat{\imath} + 0.949\hat{j}$
21. 3.8×10^9 N/C
23. (a) 2.2×10^6 N/C (b) 77 N
25. $-1.6\hat{\imath}$ pN
27. (a) $(-2.6\ \mathrm{GN/C})\hat{\imath}$ (b) $(0.52\ \mathrm{GN/C})\hat{\imath}$
    (c) $(-5.8\ \mathrm{GN/C})\hat{\imath}$
29. 1.1 kN/C
31. $E = kQ/(\sqrt{8}a^2)$
33. 5.1×10^4 N/C
35. 980 N/C
37. $\pm 22\ \mu$C
39. $16\hat{\imath} - 9.0\hat{j}$ N

41. $4q = (-a, 0), -q = (0, 0), 4q = (a, 0)$
43. (a) 20 μC (b) $(-1.6)\hat{\imath}$ N
45. (a) $2.3\hat{\imath}$ MN/C (b) $0.82\hat{\imath} + 0.82\hat{\jmath}$ MN/C (c) $\vec{E} = 0.30\hat{\imath} - 0.89\hat{\jmath}$ MN/C
47. $-4e$
49. (a) (8.0 GN/C)$\hat{\jmath}$ (b) (190 MN/C)$\hat{\jmath}$ (c) (220 kN/C)$\hat{\jmath}$
51. 0
53. $q_1 = \pm 40$ mC, $q_2 = \mp 6.9$ μC
57. The device doesn't work because its two halves depend on charge-to-mass ratio in the same way.
59. -14 μC/m
61. 1.3×10^{-30} C·m
63. (a) $2kQqa/x^2$ (b) $2kQqa/x^3$ (c) upward
65. 0.4 e, 0.03 e
67. (a) $\vec{E}(x) = 2kqa^2\dfrac{(3x^2 - a^2)}{x^2(x^2 - a)^2}(\hat{\imath})$ (b) $\vec{E}(x) \approx \dfrac{6kqa^2}{x^4}(\hat{\imath})$
69. (a) 2.5 μC/m (b) 300 kN/C (c) 1.8 N/C
75. $y = a/\sqrt{2}$
77. $E = -\dfrac{k\lambda_0\hat{\imath}}{L}\left[\dfrac{1}{2} + 2\ln(2)\right]$
79. mdv^2/qL^2
81. a
83. a

Chapter 21

17. 3 μC
19. $Q_C = 2Q = -Q_B$.
21. 650 kN/C
23. ± 1.5 kN·m^2/C
25. (a) $-q/\epsilon_0$ (b) $-2q/\epsilon_0$ (c) 0 (d) 0
27. 49 kN·m^2/C
29. (a) 1.2 MN/C (b) 2.0 MN/C (c) 50×10^4 N/C
31. Line symmetry
33. 49×10^3 N/C
35. (a) 5.1×10^6 N/C (b) 34 N/C
37. (a) 2.0×10^6 N/C (b) 7.2×10^3 N/C
39. (a) 0 (b) 4.0×10^{-3} C/m^2
41. 1.8 MN/C
43. $\pm E_0a^2/2$
45. 7.0 MN/C; 17 MN/C
47. (a) 2.8 cm (b) 3.5 nC
51. (a) (3.6 MN/C)$\hat{r}$ (b) (3.8 MN/C)$\hat{r}$ (c) (7.8 MN/C)$\hat{r}$
53. (a) (20 kNC)$\hat{r}$ (b) (1.7 kN/C)$\hat{r}$
55. 6.3 μC/m^3
57. (a) $\rho x/\epsilon_0$ (b) $\rho d/2\epsilon_0$
59. 18 N/C
61. (b) $-Q$
65. (a) $Q = \pi\rho_0a^3$ (b) $E(r) = \rho_0r^2/(4\epsilon_0a)$
67. $a = 5\rho_0/(3R^2)$
69. $R^3\dfrac{\rho_0}{\epsilon_0}(e - 2)$
71. a
73. b

Chapter 22

15. 15 μJ
17. 3.0 kV
19. 910 V
21. Proton, ionized He atom: 1.6×10^{-17} J, proton: 3.2×10^{-17} J
23. $-E_0y$
25. 53 nC
27. (a) 440 kV, 9.2×10^6 m/s
31. (a) 4 V (b) $E_x = 1$ V/m, $E_y = -12$ V/m, $E_z = 3$ V/m
33. 3 kV
35. 5.6 kV/m
37. 4.5 V
41. 6.1 μC
43. $\sqrt{2keQ/(mR)}$
45. kQ/R
47. $-ax^2/2$
49. -52 nC/m
51. $-a/2$, $a/4$
53. (a) 2.6 kV (b) 1.8 kV (c) 0
55. $V = 2kQ/R$
57. $2\pi k\sigma(\sqrt{x^2 + b^2} - \sqrt{x^2 + a^2})$
61. $(V/R)\hat{r}$
63. (a) 43 kV (b) 1.7 MN/C (c) 540 V (d) 0
65. $-E_0R/3$
67. (a) 7.2 kV (b) 14 kV
69. 14 cm, 1.7 nC
73. $-\dfrac{k\lambda_0}{L^2}\left[Lx + x^2\ln\left(\dfrac{2x - L}{2x + L}\right)\right]$
75. 8.0 mm
77. a
79. b

Chapter 23

13. 4.9 kJ
15. 0
17. -48.5 eV
19. (a) 1.4 J (b) 4.2 J
21. 22 nF
25. 740 pF
27. 1.5 J
29. 3.0 μF, 2/3 μF
31. (a) 0.012 mF (b) $Q_1 = 0.12$ mC, $Q_2 = 0.040$ mC, $Q_3 = 0.080$ mC (c) $V_1 = 60$ V, $V_2 = V_3 = 40$ V
33. 8.2×10^5 V/m
35. No
37. $Q_y = 4Q_0/(\sqrt{2} + 1) \approx 1.66Q_0$
39. 2.8 μC
41. $C = \dfrac{4\pi\epsilon_0ab}{b - a}$
43. (a) 2.5 mJ (b) 8.0 μF
45. 50 V
47. (a) 4.4 kV (b) 120 kW
49. 5C/3
51. 6/7 μF
55. (a) 4.1 nF (b) 1.3 kV
57. 2.7 nm
59. 24 μJ
61. $U = kQ^2/(2R)$
63. 6.0×10^{-4} J
65. 13 min
67. $\dfrac{1}{6}$
69. (a) $C_0\dfrac{\kappa + 1}{2}$ (b) $C_0V_0^2/(\kappa + 1)$ (c) $\dfrac{2C_0V_0^2(\kappa - 1)}{L(\kappa + 1)^2}$
71. $C_L = \pi\epsilon_0\ln\left(\dfrac{b - a}{a}\right)^{-1}$
73. (a) $Q^2x/(2A\epsilon_0)$ (b) $-Q^2/(2A\epsilon_0)$
75. d
77. d

Chapter 24

13. 9.4×10^{18}
15. 1.9×10^{11}
17. 3.2×10^6 A/m^2
19. 6.8 cm
21. (a) $5.95\times10^7(\Omega\cdot\text{m})^{-1}$ (b) $4.55\ (\Omega\cdot\text{m})^{-1}$
23. 360 V
25. 32 mΩ
27. $4R$
29. (a) 6.0 V (b) 8.0 Ω
31. 230 V
33. 300 Ω
35. (a) 0.12 mA (b) no
37. (a) 420 A/mm^2 (b) 0.24 A/mm^2
39. (a) 7.6 (b) 4
41. 9.7 μC
43. (a) 5.8 MA/m^2 (b) 97 mV/m
45. Ge
47. 50 ft
49. $R_1 = 388$ $\mu\Omega$, $R_2 = 0.971$ $\mu\Omega$, and $R_3 = 0.243$ $\mu\Omega$.
51. $d_1 = \sqrt{2}d_2$.
53. 0.63 A
55. Aluminum at \$3.30/m is more economical than copper at \$14/m.
57. 2.5 A
59. 250°C
63. $2\pi J_0a^2/3$
65. 19°
67. a
69. c

Chapter 25

17. 1.4 h
19. 43 kΩ
21. 10 V
23. 50 Ω
25. $I_1 = 2$ A, $I_2 = 0.2$ A, $I_3 = 2$ A
27. 0 A
29. -0.66%
35. $\epsilon R_2/(R_1 + R_2)$
37. 1.5 A
39. 30 A
41. 14 W

43. 120 mA, so yes, possibly fatal
45. (a) $\mathcal{E}R_1/(R + 2R_1)$
47. 2 W
49. $7R/5$
51. (a) 48 V (b) 57 V (c) 60 V
53. (a) 0.992 A (b) 0.83%
55. 360 μF; 1200 V
57. 3.4 μJ
59. (a) $V_C = 0$, $I_1 = 25$ mA, $I_2 = 0$
 (b) $V_C = 60$ V, $I_1 = I_2 = 10$ mA
 (c) $V_C = 60$ V, $I_1 = 0$, $I_2 = 10$ mA
 (d) $V_C = 0$, $I_1 = I_2 = 0$
61. 4.51 V, 35.2 Ω
62. 12 kΩ
63. 80 μs
65. 8 Ω; 89 W
67. (a) R_1 (b) R_1 (c) R_1
71. (a) 9 V (b) 1.5 ms (c) 0.3 μF
73. 220 mV
75. $\tau = \dfrac{R_1R_2C}{R_1 + R_2}$
77. Yes
79. a
81. b

Chapter 26

15. (a) 16 G (b) 23 G
17. (a) 2.0×10^{-14} N (b) 1.0×10^{-14} N (c) 0
19. 400 km/s
21. 360 ns
23. (a) 86 mT (b) 1.0 keV
25. 0.38 N
27. 12,000 lb, so clamping down the bar is a good idea.
29. 12 cm
31. 1.2 mT
33. 5 mN/m
35. 1.1×10^{-3} A·m^2 (b) 1.0×10^{-3} N·m
37. 7.0 A
39. (a) 4.0 G (b) 20 G
41. 17 T
43. 2.3×10^{27}A·m^2
45. 3.8 GA
49. (a) 71 μm (b) 440 μm
51. 0.53 A
53. 8.5×10^{22} cm^{-3}
55. (a) 0.35 A·m^2 (b) 4.2×10^{-2} N·m
57. 0.021 N, 45° above horizontal
59. $(1 + \pi)\dfrac{\mu_0 I}{2\pi a}\hat{k}$
61. $\dfrac{\mu_0 I}{4a}\hat{k}$
63. 7.1×10^{-6} N
65. (a) 0 (b) $B = \mu_0 I/(2\pi r)$
67. (a) 2.3×10^3 (b) 3.3 kW
71. (a) 8.0 μT (b) 4.0 μT (c) 0
73. (a) $B \approx \dfrac{\mu_0 I}{2w}$ (b) $B \approx \dfrac{\mu_0 I}{2\pi r}$
75. (a) $\pi R^2 J_0/3$ (b) $B = \dfrac{\mu_0 J_0 R^2}{6r}$
 (c) $B = \dfrac{\mu_0 J_0 r}{2}\left(1 - \dfrac{2r}{3R}\right)$
77. Since $\tau \propto 1/N$, more torque from a 1-turn loop.
81. $\mu_0 nIl/\sqrt{l^2 + 4a^2}$
83. 0.64 T
85. The hall potential is 10,000 times smaller than bioelectric potentials.
87. d
89. d

Chapter 27

13. 1.4×10^{-4} Wb
15. 160 T/s
17. 3.2 mH
19. 40 kV
21. 220 mH
23. 3.1 kJ
25. 57 mJ
29. 4.4 T
31. $-rb/2$
33. (a) -0.30 A (b) -0.20 A
35. 15 mT
37. (a) 3 s (b) clockwise
39. (a) 18 mA (b) 40 mA
41. -42 mA, clockwise
43. 130
45. (a) Upper bar (b) 0
47. (a) 25 mA (b) 1.3×10^{-3} N
 (c) 2.5 mW (d) 2.5 mW
49. 58 T/ms
53. 0.76 s
55. 20 s
57. (a) 5 Ω (b) 500 J
59. (a) 1.0 A (b) 0.43 A (c) -1.7 A
61. 190 mΩ
63. 3.4×10^{21} J/m^3
65. $\dfrac{\mu_0 I^2}{16\pi}$
67. 3×10^8 m/s (speed of light)
69. (a) $-br/(2\rho)$ (b) $= \dfrac{\pi b^2 h a^4}{8\rho}$
71. $v(t) = \dfrac{FR}{B^2l^2}\left[1 - \exp\left(-\dfrac{B^2l^2}{Rm}t\right)\right]$
73. $\dfrac{\mu_0}{2\pi}\ln(b/a)$
77. c
79. a

Chapter 28

15. (a) 294 V (b) 2.51×10^3 s^{-1}
17. (a) $V(0) \approx V_p/\sqrt{2}$, 45°
 (b) $V(0) = 0$, $\phi_b = 0$
 (c) $V(0) = V_p$, $\phi_c = 90°$
 (d) $V(0) = 0$, $\phi_d = \pm\pi$
 (e) $V(0) = -V_p$, $\phi_e = -90°$
19. $I_{R,\text{rms}} = 13$ mA, $I_{C,\text{rms}} = 24$ mA, $I_{L,\text{rms}} = 22$ mA
21. (a) 250 V (b) 15 V
23. 16 kHz
25. 8.1 H
27. (a) 32 mH (b) 1.0 V
29. (a) 79 nF (b) 300 Ω
31. (a) 95 Hz (b) 18 kΩ
33. 500 W
35. 4.8 kW, 4.5 kW
39. (a) 80 Hz (b) 2.0 H (c) $X_L = 4X_C$
41. $(45\ \mu\text{V})\sin[(63\ \text{s}^{-1})t]$
43. 3.2 kHz
45. (a) 52 nF (b) 350 Hz
47. 0.199 μH
49. (a) $1/\sqrt{2}$ (b) 1/2 (c) $-1/\sqrt{2}$ (d) 1/2
51. 50
53. 6.2 Ω
55. (a) Above resonance
57. (a) 0.333 (b) 4.00 W
59. (a) 5.5% (b) 9.1%
61. 3.7 mF
63. 2.7 V
65. 1620 Hz
67. 12 and 36 nF
71. $R = 400$ W, $L = 68$ mH, $C = 94$ nF
75. $\omega = \sqrt{\dfrac{1}{LC} - \dfrac{R^2}{4L^2}}$
77. Yes
79. a
81. d

Chapter 29

13. 1.3 nA
15. $-\hat{k}$
19. 7.5 km
21. 2.57 s
23. 5.00×10^6 m
25. $\hat{k}$
27. 12%
29. 1×10^{10} W/m^2
31. The radio has a minimum intensity of 0.27 nW/m^2, so it will work at the cabin.
33. 20 kW
35. (a) 7.2×10^{11} V/(m·s) (b) increasing
37. 0.94−1.0 PHz
39. 1.07 pT
41. 91%
43. 19%
45. 0.00004%
47. 4×10^{10}
49. (a) 4.6 kW (b) 53 mV/m
51. (a) $1/r$ (b) $1/r^2$
53. (a) 8.9×10^8 W/m^2 (b) 58×10^3 V/m
55. 6.2×10^3 y
57. 2.52 kPa
59. (a) 1.0 MV/m (b) 4.3 mm (c) 64 mJ
 (d) 2.1×10^{-10} kg·m/s (e) 64 W
61. 6
65. 2.75 m
67. 2.2 km
69. b
71. d

Chapter 30

11. 15°
13. 0.5°
15. Ice
17. 77.7°
19. 14.2°
21. 1.9
23. 79.1°

25. 1.66
27. 6.41°
29. (a) 18° (b) 390 nm
31. Ethyl alcohol
33. 1.83
35. 5.1 m
37. 139 nm
39. 35°
41. Diagonal face, 23°
43. 1.07
47. 53.5°
49. 36.9°
51. 2.7 m
53. 1.9 m
57. (c) 50.9°
61. $\frac{d}{c}\left(\frac{2}{3}n_1 + \frac{1}{3}n_2\right)$
63. a
65. c

Chapter 31

17. 35°
19. (a) $-1/4$ (b) real, inverted
21. (a) $2f$ (or R) (b) real
23. -2
25. 21 cm
27. 27 cm
29. 40 cm
31. 0.86 mm
33. 2.2 diopters
35. -1.3 diopters
37. -200
39. (a) -24 cm (b) 29 mm
(c) virtual, upright, enlarged
41. 18 cm
43. 7.4 cm
45. 7.59 cm
47. 12 cm
51. 29 or 41 cm
53. 11 cm
55. $s' = 1.1$ m, inverted image, real image
57. -68 cm
59. 2.0
61. 2
63. Choose plastic, because it meets requirements and is cheaper.
65. -2.8 cm
67. 3.3 diopters
69. 0.3°
71. 72 cm
81. 0.025 cm
83. b
85. b

Chapter 32

11. 1.7 cm
13. 420 nm
15. 4
17. (a) 4.8°, 9.7° (b) 2.9°, 6.8°
19. (a) 2 (b) 1
21. 103 nm
23. 594 nm, 424 nm
25. The top 1.5-cm of the film
27. 29°
29. 1.62%
31. 36 cm
33. 3×10^{-4} rad
35. 94°
37. (a) 38 (b) 3
39. 44 μm
43. 2
45. Not feasible because a 2-km-wide telescope is needed
47. 3.3 Å
49. 5
51. 236
53. 128.8 μm
55. $1 + 2.93\times10^{-4}$
57. 34 m
59. 2.0 μm
61. 6.9 km
63. Yes, since the resolution size scales with the wavelength.
65. $n_{gas} = 1 + \frac{m\lambda}{2L}$
67. $\Delta y = D\lambda/2d$
69. 92
71. c
73. a

Chapter 33

13. (a) 4.50 h (b) 4.56 h (c) 4.62 h
15. 33 ly
17. 40 m
19. 0.14c
21. (a) 2.0 (b) 2.5
23. 0.14c
25. (a) 2.1 MeV (b) 1.6 MeV
29. (a) 0.86c (b) 9.7 min
31. $c/\sqrt{2}$
33. Twin A = 83.2 years old, twin B = 39.7 years old
35. 0.96c
39. Civilization B, 3.5×10^5 y
41. $v = 0.50c$
45. 0.94c
47. (a) 10 ly, 13 y (b) 0 ly, 7.5 y
51. (a) 4.2 ly (b) -2.4 ly
53. (a) 0.758c (b) 1.09 GeV/c
55. 25 h
57. (a) 0.26 eV (b) 1.3 keV
(c) 3.1 MeV
63. 0.866c
65. 0.95c
71. 0.31 c; 27 kV
73. a
75. a

Chapter 34

15. 16
17. $\lambda_{peak} = 10.1\ \mu$m, $\lambda_{median} = 14.3\ \mu$m
19. (a) 500.0 nm (b) 708.6 nm
21. 2.8×10^{-19} J to 5.0×10^{-19} J
23. 1.44
25. 122 nm, 103 nm, 97.2 nm
27. 91.2 nm
29. (a) 3.7×10^{-63} m (b) 73 nm
31. The electron moves 1836 times faster than the proton.
33. 6×10^7 m/s
35. 130 nm
37. 23 keV
39. 5.4×10^{-2}
41. (a) 4.39×10^3 K (b) 0.474
43. (a) $1.7\times10^{28}\ s^{-1}$ (b) $3.2\times10^{15}\ s^{-1}$
(c) $1.3\times10^{18}\ s^{-1}$
45. (a) 1.12×10^{15} Hz (b) 2.79 eV
47. (a) 2.9 eV and 1.9 eV
(b) Plants absorb blue and red, reflect green.
49. 440 nm
51. (a) 154 pm (b) 222 eV
53. No
55. (a) 26.4 cm (b) 4.70 μeV
57. 229
59. 3.40 eV
61. (a) 0.0265 nm (b) 40.8 eV
63. 1.62 km/s
65. 2.9 km/s
67. 1 ps
71. $E_0 = \frac{1}{2}m_ec^2[(\gamma - 1) + \sqrt{(\gamma - 1)(\gamma + 3)}]$
79. c
81. d

Chapter 35

11. (a) 0 (b) $\pm a\sqrt{\ln 2/2}$
13. 5
15. 3.8 meV
17. (a) 0.38 eV (b) 1.5 eV
19. Electron
21. 0.2 MeV
23. 33 eV
25. 8.0 eV
27. $E \to E/4$
29. 930 pm
33. (a) 2.2 eV (b) 570 nm
35. 21 μm
37. 2.2 nm, 4.7 nm, 6.6 nm, 2.8 nm, 4.1 nm, 11 nm
39. (a) $\psi_{n-odd}(x') = \sqrt{\frac{2}{L}}\cos\left(\frac{n\pi x'}{L}\right)$,
$\psi_{n-even}(x') = \sqrt{\frac{2}{L}}\sin\left(\frac{n\pi x'}{L}\right)$
(b) $E_n = n^2h^2/(8mL^2)$
41. 0.759 nm
43. 2.5×10^{-17} eV; quantization is insignificant
45. (a) 0.30 (b) 0.15
51. 4
53. (c) $A_0 = (\alpha^2/\pi)^{1/4}$
55. b
57. a

Chapter 36

15. 3
17. d
19. 5
21. 3d

23. 2.58×10^{-34} J·s
25. 3/2, 5/2
27. $11.5\hbar\omega$
29. $1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^1$
33. 1.137 meV
35. $n = 4, l = 3$
37. 2.67×10^{68}
39. 90°, 65.9°, 114°, 35.3°, 145°
41. 0, ±1, ±2, ±3
45. (a) $E_1/16$ (b) $\sqrt{12}\,\hbar$ (c) $\frac{1}{2}\sqrt{35}\,\hbar$
47. (a) $16\,\hbar\omega$ (b) $4\,\hbar\omega$
49. $1s^2 2s^2 2p^6 3s^2 3p^6 4s^1 3d^{10}$
51. 3.0×10^{17}
53. 0.1
55. $3\hbar$, $6h$
57. (a) 0.966 (b) 0.0595
59. $P(r)dr = 4\pi r^2 \psi_{2s}^2\, dr$, $3 + \sqrt{5}$
61. (b) 54.4 eV, 870 eV, 91.4 keV, 115 keV
63. (b) 141 eV, 65.8 eV, 47.0 eV, 28.2 eV
65. $3a_0/2$
69. b
71. d

Chapter 37

17. 3.48 mm
19. 9.41×10^{-46} kg·m^2
21. 7.08×10^{13} Hz
23. 181 kcal/mol
25. 549 nm
27. 3.54 μm
29. 1.25 meV
31. $l\hbar^2/I$
33. 0.121 nm
35. (a) 0.179 eV (b) 0.358 eV
37. 14.95 μm
39. 15 meV
41. 35.8 μm
43. 10.2
45. −8.40 eV
49. 4.68 eV
51. 6.36×10^4 K, ~200 times room temperature
53. 709 nm, no
55. 1.8 kA
57. 508 nm
59. $I = m_1 m_2 R^2/(m_1 + m_2)$; 0.128 nm
63. (a) $(2^{9/2}\pi m^{3/2} L^3/3h^3)E^{3/2}$
65. 64 kA
67. a
69. a

Chapter 38

13. $^{211}_{86}$Ra, $^{220}_{86}$Ra, and $^{222}_{86}$Ra.
15. (a) $A = 35$ for both (b) $Z_K = Z_{Cl} + 2$
17. 5.9 fm
19. $^{64}_{29}\text{Cu} \to {}^{64}_{30}\text{Zn} + \beta^- + \bar{\nu}$; $^{64}_{29}\text{Cu} \to {}^{64}_{28}\text{Ni} + \beta^+ + \nu$; $^{64}_{29}\text{Cu} + e^- \to {}^{64}_{28}\text{Ni} + \nu$
21. 17 Bq/L
23. (a) 19×10^1 y (b) 29×10^1 y
25. 59.930 u
27. 5.612 MeV
29. 2
31. 1.0×10^{20} s^{-1}
33. 2×10^{20} m^{-3}
35. 10^3 s
37. 5.3×10^{-12} eV
39. 8.80 MeV
41. 0 atoms; 5×10^5 atoms; 8×10^4 atoms; U-238 and K-40 are suitable
43. 9.6 d
45. (a) $^{228}_{90}$Th
47. 8.9×10^3 y
49. Poland: 8.04 d; Austria: 16.2 d, Germany: 10.0 d
51. 3.0×10^9 y
53. 3.31%
55. 3×10^{-13}
57. 1.3×10^3 kg
59. 88.9%
61. 580 kg
63. 0.461 s
65. (a) 4×10^{38} s^{-1} (b) 7×10^9 y
67. 8×10^{17}s, which is about 20 billion years longer than the sun will shine
69. Borhium-262 ($^{262}_{107}$Bh)
71. (a) $^{65}_{29}$Cu (b) 4 h
73. (a) 1.1 GW (b) 2.8 s^{-1} (c) 0.45 metric tons
77. $R(t) = \lambda_A N_0 e^{-\lambda_A t} + \dfrac{\lambda_B \lambda_A N_0}{\lambda_B - \lambda_A}(e^{-\lambda_A t} - e^{-\lambda_B t})$
79. Yes
81. b
83. d

Chapter 39

19. 0.336 fs
21. $\pi^+ \to \mu^+ + \nu_\mu$
23. $\eta \to \pi^+ + \pi^- + \pi^0$
25. No, violates conservation of baryon number and angular momentum
27. sss
29. 4.54×10^7 L
31. 10^{28} K
33. 900 Mly
35. 12 Gyr
37. Reaction (a) is not possible because it violates conservation of baryon number and angular momentum.
39. (a) No (b) yes
41. $c\bar{c}$
43. (a) 0.16 µJ (b) 0.02 mm
45. 90 µs
47. 313 ly
49. (a) 256 fm (b) −2.81 keV
51. (a) 5.740×10^3km/s (b) 253 Mly
53. 2.6×10^{-25} s
55. 13.6 Gy, 17.6 Gy
57. 5.0 km·s^{-1} · Mly
59. c
61. b

Credits

Front Matter
Page i: Andrew Lambert Photography/Science Photo Library. Page iv: Robert Keren.

Chapter 1
Page 1: David Parker/Photo Researchers. Page 3: NASA. Page 4, Figure 1.3: iStockphoto. Page 5, Figure 1.4(a): NASA (b) Andrew Syred/Photo Researchers.

Chapter 2
Page 12: AP Wide World Photos. Page 12 B: Getty Images/ Aurora Creative. Page 13: Nik Wheeler/Corbis. Page 22, Figure 2.11: Richard Megna/Fundamental Photographs. Page 24: National Institute of Standards and Technology (NIST).

Chapter 3
Page 30: Martha Holmes/Nature Picture Library. Page 36, Figure 3.11: Richard Megna/Fundamental Photographs. Page 39: Preston Wilson/Getty Images.

Chapter 4
Page 48: Superstock/AGEfotostock. Page 53, Figure 4.6: Solstice Photography/Getty Images. Page 54, Figure 4.8: NASA. Page 58: NASA.

Chapter 5
Page 66: Peter/Georgina Bowater/Mira. Page 71, Figure 5.10: Nic Bothma/AP Wide World. Page 84, Figure 5.37: AP Photo/Lionel Cironneau.

Chapter 6
Page 85: Paolo Cocco/AFP/Getty Images.

Chapter 7
Page 101: Kennan Ward/Corbis. Page 105: Courtesy Northfield Mountain Pumped Storage Project. Page 109 L, R: Erik Borg/Addison Wesley Longman.

Chapter 8
Page 118 L: Pete Stopher/iStockphoto. Page 118 R: Courtesy Northrup Grumman. Page 119, Figure 8.2: Lowell Observatory. Page 123, Figure 8.6: David Hardy/Photo Researchers.

Chapter 9
Page 133: Patrick Reeves. Page 137, Figure 9.7: Wally MacNamee/Corbis. Page 142: NASA Headquarters. Page 154, Figure 9.28: Richard Megna/Fundamental Photographs.

Chapter 10
Page 155: Wind Prospect Ltd. Page 166: Courtesy Mark Flynn/University of Texas-Austin.

Chapter 11
Page 174: NASA/Goddard Space Flight Center. Page 181: REUTERS/Stefano Rellandini.

Chapter 12
Page 186: Robert Harding Picture Library Ltd./Alamy.

Chapter 13
Page 203: Shutterstock. Page 204: Courtesy Project Bandaloop. 209 L: Peter Tsai Photography/Alamy. Page 209 R: Shutterstock. Page 217, Figure 13.26: AP Wide World. Page 223, Figure 13.38: NASA.

Chapter 14
Page 224: Pixtal/AGEfotostock. Page 234, Figure 14.21: Uri Haber-Schaim/Kendall/Hunt Publishing. Page 236, Figure 14.26: Richard Megna/Fundamental Photographs. Page 238: Figure 14.29: Michael Freeman. Page 245, Figure 14.38: David Rydevik via Wikipedia.

Chapter 15
Page 246: Corbis/Bettmann. Page 252: Courtesy Pagani Automobili s.p.a. Page 256, Figure 15.16: Courtesy Columbia University Physics Dept. Page 257: MACIEJ NOSKOWSKI/ iStockphoto. Page 257, Figure 15.21: Anupam Pal.

Chapter 16
Page 264: Mark Antman/The Imageworks, Page 264 (inset): Steve Allen/Getty Images. Page 265: Ted Kinsman/Photo Researchers. Page 271, Figure 16.9: Hal Lott/The Stock Connection. Page 272, Figure 16.10: Manuel G. Velarde. Page 281, Figure 16.18 (inset): Branko Miokovic/iStockphoto.

Chapter 17
Page 282: Digital Vision/Getty Images. Page 289, Figure 17.7: AP/Wide World. Page 290: Brian J. Skerry/National Geographic Image Collection. Page 294, Figure 17.12: Shutterstock.

Chapter 18
Page 295: Photodisc/Getty Images. Page 302: Jon Arnold/AGEfotostock.

Chapter 19
Page 312: James Hardy/AGEfotostock. Page 318: Peter Bowater/Photo Researchers.

Chapter 20
Page 331: NASA Earth Observing System. Page 332: Photodisc Red/Getty Images. Page 333, Figure 20.1: Richard Megna/Fundamental Photographs. Page 345: Jason Brindel Commercial/Alamy. Page 345 (inset): Bohdan Senyuk, Kent State University.

Chapter 21
Page 351: Peter Menzel/Photo Researchers, Inc. Page 366: Fox Photos/Hulton/Getty Images. Page 370, Figure 21.36: Eric Schrader—Pearson Science.

Chapter 22
Page 372: AP/Wide World. Page 383: Courtesy of EPRI.

Chapter 23
Page 389: Imagemore/Getty Images. Page 392, Figure 23.5: Erik Borg/Pearson Education. Page 395: Cindy Charles/PhotoEdit. Page 401, Figure 23.18: National Ignition Facility.

Chapter 24
Page 403: Danial Sambraus/Photo Researchers. Page 414, Figure 24.15: Erik Borg/Addison Wesley Longman.

Chapter 25
Page 419: Tom Pantages. Page 422, Figure 25.6: Erik Borg/Addison Wesley Longman.

Chapter 26
Page 439: M. Aschwanden et al. (LMSAL), TRACE, NASA. Page 440, Figure 26.1: Cordelia Molloy/Photo Researchers, Inc. Page 443: Ivan Massar. Page 454, Figure 26.29: University Corporation for Atmospheric Research UCAR. Page 459, Figure 26.36: Richard Megna/Fundamental Photographs.

Chapter 27
Page 466: Charles E. Rotkin/Corbis. Page 481, Figure 27.28: NASA. Page 485, Figure 27.34: Milt and Joan Mann/Cameramann International, Ltd.

Chapter 28
Page 491: Lester Lefkowitz/Corbis. Page 499, Figure 28.13: Erik Borg/Addison Wesley. Page 503, Figure 28.21 TL: Shutterstock. Figure 28.21 TM: Shutterstock. Figure 28.21 TR: Larry Lawhead/iStockphoto. Figure 28.21 ML: Jonathan Nourak/PhotoEdit. Figure 28.21 BL: Robert Robinson/iStockphoto. Figure 28.21 BR: jlsohio/iStockphoto. Figure 28.21 MR: M. Eric Honeycutt/iStockphoto.

Chapter 29
Page 510 T: PhotoDisc/Getty Images. Page 510 BL: Brent Bossom/iStockphoto. Page 510 BR: Neustockimages/iStockphoto. Page 520, Figure 29.8: Richard Megna/Fundamental Photographs. Page 529, Figure 29.15: Japan Aerospace Exploration Agency.

Chapter 30
Page 531: Craig Tuttle/Corbis. Page 532: Courtesy of Schott Corporation. Page 536, Figure 30.9: Richard Megna/Fundamental Photographs. Page 537: Spencer Grant/Photo Researchers. Page 539, Figure 30.13: Courtesy of R. Giovanelli and H. R. Gillett/CSIRO National Measurement Laboratory, Australia. Page 541, Figure 30.15: Library of Congress.

Chapter 31
Page 544: Huntstock, Inc/Alamy. Page 545, Figure 31.2: Martin Bough/Fundamental Photographs. Page 546, Figure 31.4: Courtesy of the Space Telescope Institute/NASA. Page 547, Figure 31.7: Erik Borg/Addison Wesley Longman. Page 547, Figure 31.9: Scott Leigh/iStockphoto. Page 549, Figure 31.11: Perkin Elmer. Page 552, Figure 31.21: Erik Borg/Addison Wesley Longman. Page 556: © PHOTOTAKE Inc./Alamy. Page 559, Figure 31.34: Giant Magellan Telescope, Carnegie Observatories. Page 563, Figure 31.37: Shutterstock.

Chapter 32
Page 564: Chris Collins/Corbis. Page 566: Andrew Syred/Photo Researchers. Page 566, Figure 32.3: M. Cagnet et al. Atlas of Optical Phenomena Springer-Cerlag, 1962. Page 569, Figure 32.8: Chris Jones. Page 572, Figure 32.13: Courtesy of Jay M. Pasachoff. Page 573, Figure 32.14: Richard Megna/Fundamental Photographs. Page 573, 32.15: Erik Borg/Addison Wesley. Page 573, Figure 32.16: NASA/Ames Research Center/Ligo Project. Page 577, Figure 32.22: Courtesy of Cagnet Atlas. Page 577, Figure 32.23: Courtesy of Cagnet Atlas. Page 578, 32.26: Chris Jones. Page 579: Courtesy of the Blu-Ray Disc Association. Page 580: Chris Jones.

Chapter 33
Page 585: Delft University of Technology/Photo Researchers, Inc. Page 586: AP/Wide World. Page 589, Figure 33.4: Albert Einstein Archives Hebrew University of Jerusalem. Page 604, Figure 33.21: Space Telescope Science Institute/Photo Researchers.

Chapter 34
Page 609: Courtesy Wolfgang Ketterle, MIT. Page 620: Andrew Syred/Photo Researchers. Page 621, Figure 34.13: Millie LeBlanc/Courtesy of The Educational Development Center.

Chapter 35
Page 628: IBM Corporation. Page 629, Figure 35.2: Courtesy of Elisha Huggins, Dartmouth. Page 637: Courtesy of IBM. Page 643, Figure 35.19: Stefan Diaz.

Chapter 36
Page 644: Chip Clark. Page 652, Figure 36.14: Courtesy Wolfgang Ketterle, MIT. Page 661, Figure 36.20: Medical Body Scans/Photo Researchers, Inc.

Chapter 37
Page 662: Alfred Pasieka/Photo Researchers. Page 673, Figure 37.20: CB2/ZOB/WENN Photos/Newscom. Page 675: Koichi Kamoshida/Liaison/Getty Images.

Chapter 38
Page 680: WDCN/Univ. College London/Photo Researchers. Page 684: David Job/Getty Images. Page 700, Figure 38.20: Lawrence Livermore National Laboratory/Photo Researchers.

Chapter 39
Page 706: CERN/European Organization for Nuclear Research. Page 714, Figure 39.7: CERN/European Organization for Nuclear Research. Page 714, Figure 39.8: ICRR Institute for Cosmic Ray Research. Page 715: CERN. Page 716, Figure 39.11: NASA Headquarters. Page 717, Figure 39.13: NASA Headquarters. Page 719, Figure 39.16: Licia Verde.

Index

C

GEOPHYSICAL AND ASTROPHYSICAL DATA

EARTH

Mass	5.97×10^{24} kg
Mean radius	6.37×10^{6} m
Orbital period	3.16×10^{7} s (365.3 days)
Mean distance from Sun	1.50×10^{11} m
Mean density	5.5×10^{3} kg/m^3
Surface gravity	9.81 m/s^2
Surface pressure	1.013×10^{5} Pa
Magnetic moment	8.0×10^{22} A·m^2

MOON

Mass	7.35×10^{22} kg
Mean radius	1.74×10^{6} m
Orbital period	2.36×10^{6} s (27.3 days)
Mean distance from Earth	3.85×10^{8} m
Mean density	3.3×10^{3} kg/m^3
Surface gravity	1.62 m/s^2

SUN

Mass	1.99×10^{30} kg
Mean radius	6.96×10^{8} m
Orbital period (about galactic center)	6×10^{15} s (200 My)
Mean distance from galactic center	2.6×10^{20} m
Power output (luminosity)	3.85×10^{26} W
Mean density	1.4×10^{3} kg/m^3
Surface gravity	274 m/s^2
Surface temperature	5.8×10^{3} K

PERIODIC TABLE OF THE ELEMENTS

2	Atomic number
He	Symbol
4.003	Atomic mass (u)*

Metals

Semimetals

Nonmetals

1 H 1.008																	2 He 4.003
3 Li 6.941	4 Be 9.012											5 B 10.81	6 C 12.01	7 N 14.01	8 O 16.00	9 F 19.00	10 Ne 20.18
11 Na 22.99	12 Mg 24.31											13 Al 26.98	14 Si 28.09	15 P 30.97	16 S 32.07	17 Cl 35.45	18 Ar 39.95
19 K 39.10	20 Ca 40.08	21 Sc 44.96	22 Ti 47.88	23 V 50.94	24 Cr 52.00	25 Mn 54.94	26 Fe 55.85	27 Co 58.93	28 Ni 58.69	29 Cu 63.55	30 Zn 65.39	31 Ga 69.72	32 Ge 72.61	33 As 74.92	34 Se 78.96	35 Br 79.90	36 Kr 83.80
37 Rb 85.47	38 Sr 87.62	39 Y 88.91	40 Zr 91.22	41 Nb 92.91	42 Mo 95.94	43 Tc (98)	44 Ru 101.07	45 Rh 102.91	46 Pd 106.42	47 Ag 107.87	48 Cd 112.41	49 In 114.82	50 Sn 118.71	51 Sb 121.75	52 Te 127.60	53 I 126.90	54 Xe 131.29
55 Cs 132.91	56 Ba 137.33	57–71 Lanthanide series	72 Hf 178.49	73 Ta 180.95	74 W 183.85	75 Re 186.21	76 Os 190.2	77 Ir 192.22	78 Pt 195.08	79 Au 196.97	80 Hg 200.59	81 Tl 204.38	82 Pb 207.2	83 Bi 208.98	84 Po (209)	85 At (210)	86 Rn (222)
87 Fr (223)	88 Ra (226)	89–103 Actinide series	104 Rf (261)	105 Db (268)	106 Sg (266)	107 Bh (272)	108 Hs (277)	109 Mt (276)	110 Ds (281)	111 Rg (280)	112 Cn (285)	113 (284)	114 (289)	115 (288)	116 (292)	117 (294)	118 (294)

Lanthanide series	57 La 138.91	58 Ce 140.12	59 Pr 140.91	60 Nd 144.24	61 Pm (145)	62 Sm 150.36	63 Eu 151.97	64 Gd 157.25	65 Tb 158.93	66 Dy 162.50	67 Ho 164.93	68 Er 167.26	69 Tm 168.93	70 Yb 173.04	71 Lu 174.97
Actinide series	89 Ac (227)	90 Th 232.04	91 Pa (231)	92 U 238.03	93 Np (237)	94 Pu (244)	95 Am (243)	96 Cm (247)	97 Bk (247)	98 Cf (251)	99 Es (252)	100 Fm (257)	101 Md (258)	102 No (259)	103 Lr (260)

* Atomic mass is average over abundances of stable isotopes. For radioactive elements other than uranium and thorium, mass is in parentheses and is that of the most stable important (in availability, etc.) isotope.

A list of the elements is given in Appendix D.