



HOW THINGS WORK

THE PHYSICS OF EVERYDAY LIFE

SIXTH EDITION

LOUIS A. BLOOMFIELD

WILEY

6TH
EDITION

How Things Work

THE PHYSICS OF EVERYDAY LIFE

Louis A. Bloomfield

The University of Virginia

WILEY

*To Karen for being such a wonderful friend and companion,
to Aaron, Elana, and Rich for being everything a father could want,
to Max and Rosie for being so cheerful and attentive,
and to the students of the University of Virginia for making teaching, research, and writing fun.*

VP & DIRECTOR:	Petra Recter
EXECUTIVE EDITOR:	Jessica Fiorillo
DEVELOPMENT EDITOR:	Jennifer Yee
ASSISTANT DEVELOPMENT EDITOR:	Mallory Fryc
EXECUTIVE MARKETING MANAGER:	Christine Kushner
ASSOCIATE DIRECTOR, PRODUCT DELIVERY:	Kevin Holm
SENIOR PRODUCTION EDITOR:	Sandra Dumas
PRODUCT DESIGNER:	Geraldine Osnato
SENIOR PHOTO EDITOR:	Billy Ray
DESIGN DIRECTOR:	Harry Nolan
COVER AND TEXT DESIGNER:	Thomas Nery

This book was set in 10/12 Times Roman by Aptara Corporation. Book and cover were printed and bound by Quad Graphics/Versailles.

The book is printed on acid-free paper.

Founded in 1807, John Wiley & Sons, Inc. has been a valued source of knowledge and understanding for more than 200 years, helping people around the world meet their needs and fulfill their aspirations. Our company is built on a foundation of principles that include responsibility to the communities we serve and where we live and work. In 2008, we launched a Corporate Citizenship Initiative, a global effort to address the environmental, social, economic, and ethical challenges we face in our business. Among the issues we are addressing are carbon impact, paper specifications and procurement, ethical conduct within our business and among our vendors, and community and charitable support. For more information, please visit our website: www.wiley.com/go/citizenship.

Copyright © 2016, 2013, 2010, 2006, 1997. John Wiley & Sons, Inc. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, website www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, (201) 748-6011, fax (201) 748-6008, website www.wiley.com/go/-permissions.

Evaluation copies are provided to qualified academics and professionals for review purposes only, for use in their courses during the next academic year. These copies are licensed and may not be sold or transferred to a third party. Upon completion of the review period, please return the evaluation copy to Wiley. Return instructions and a free-of-charge return shipping label are available at www.wiley.com/go/returnlabel. If you have chosen to adopt this book for use in your course, please accept this book as your complimentary desk copy. Outside of the United States, please contact your local representative.

Library of Congress Cataloging-in-Publication Data

Names: Bloomfield, Louis, author.

Title: How things work : the physics of everyday life / Louis A. Bloomfield,
The University of Virginia.

Description: Sixth edition. | Hoboken, NJ : John Wiley & Sons, Inc., [2015] |
?2016 | Includes index.

Identifiers: LCCN 2015033708I ISBN 9781119013846 (loose-leaf : alk. paper) |
ISBN 1119013844 (loose-leaf: alk. paper)

Subjects: LCSH: Physics—Textbooks.

Classification: LCC QC21.3 .B56 2015 | DDC 530—dc23 LC record available at <http://lccn.loc.gov/2015033708>

ISBN 978-1119-01384-6 (Binder Version)

The inside back cover will contain printing identification and country of origin if omitted from this page. In addition, if the ISBN on the back cover differs from the ISBN on this page, the one on the back cover is correct.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Foreword

In today's world we are surrounded by science and by the technology that has grown out of that science. For most of us, this is making the world increasingly mysterious and somewhat ominous as technology becomes ever more powerful. For instance, we are confronted by many global environmental questions such as the dangers of greenhouse gases and the best choices of energy sources. These are questions that are fundamentally technical in nature and there is a bewildering variety of claims and counterclaims as to what is "the truth" on these and similar important scientific issues. For many people, the reaction is to throw up their hands in hopeless frustration and accept that the modern world is impossible to understand and one can only huddle in helpless ignorance at the mercy of its mysterious and inexplicable behavior.

In fact, much of the world around us and the technology of our everyday lives is governed by a few basic physics principles, and once these principles are understood, the world and the vast array of technology in our lives become understandable and predictable. How does your microwave oven heat up food? Why is your radio reception bad in some places and not others? And why can birds happily land on a high-voltage electrical wire? The answers to questions like these are obvious once you know the relevant physics. Unfortunately, you are not likely to learn that from a standard physics course or physics textbook. There is a large body of research showing that, instead of providing this improved understanding of everyday life, most introductory physics courses are doing quite the opposite. In spite of the best intentions of the teachers, most students are "learning" that physics is abstract, uninteresting, and unrelated to the world around them.

How Things Work is a dramatic step toward changing that by presenting physics in a new way. Instead of starting out with abstract principles that leave the reader with the idea that physics is about artificial and uninteresting ideas, Lou Bloomfield starts out talking about real objects and devices that we encounter in our everyday lives. He then shows how these seemingly magical devices can be understood in terms of the basic physics principles that govern their behavior. This is much the way that most physics was discovered in the first place: people asked why the world around them behaved as it did and as a result discovered the principles that explained and predicted what they observed.

I have been using this book in my classes for several years, and I continue to be impressed with how Lou can take seemingly highly complex devices and strip away the complexity to show how at their heart are simple physics ideas. Once these ideas are understood, they can be used to understand the behavior of many devices we encounter in our daily lives, and often even fix things that before had seemed impossibly complex. In the process of teaching from this book, I have increased my own understanding of the physics behind much of the world around me. In fact, after consulting *How Things Work*, I have had the confidence to confront both plumbers and airconditioner repairmen to tell them (correctly as it turned out) that their diagnosis did not make sense and they needed to do something different to solve my plumbing and AC problems. Now I am regularly amused at the misconceptions some trained physicists have about some of the physics they encounter in their daily lives, such as how a microwave oven works and why it can be made out of metal walls, but putting aluminum foil in it is bad. It has convinced me that we need to take the approach used in this book in far more of our science texts.

Of course, the most important impact is on the students in my classes that use this book. These are typically nonscience students majoring in fields such as film studies, classics, English, business, etc. They often come to physics with considerable trepidation. It is inspiring to see many of them discover to their surprise that physics is very different from what they thought—that physics can actually be interesting and useful and makes the world a much less mysterious and more understandable place. I remember many examples of seeing this in action: the student who, after learning how both speakers and TVs work, was suddenly able to understand that it was not magic that putting his large speaker next to the TV distorted the picture but in fact it was just physics, and now he knew just how to fix it; the young woman scuba diver who, after learning about light and color, suddenly interrupted class to announce that now she understood why it was that you could tell how deep you were by seeing what color lobsters appeared; or the students who announced that suddenly it made sense that the showers on the first floor of the dorm worked better than those on the second floor. In addition, of course everyone is excited to learn how a microwave oven works and why there are these strange rules as to what you can and cannot put in it.

These examples are particularly inspiring to a teacher, because they tell you that the students are not just learning the material presented in class but they are then able to apply that understanding to new situations in a useful way, something that happens far too seldom in science courses.

Whether a curious layperson, a trained physicist, or a beginning physics student, most everyone will find this

book an interesting and enlightening read and will go away comforted in that the world is not so strange and inexplicable after all.

Carl Wieman

Nobel Laureate in Physics 2001
CASE/Carnegie US University Professor

Contents

CHAPTER 1 ↓ THE LAWS OF MOTION, PART 1

1

Active Learning Experiment: Removing a Tablecloth from a Table 1

Chapter Itinerary 2

1.1 Skating 2

(inertia, coasting, vector quantities, position, velocity, force, acceleration, mass, net force, Newton's first and second laws, inertial frames of reference, units)

1.2 Falling Balls 12

(gravity, weight, constant acceleration, projectile motion, vector components)

1.3 Ramps 21

(support forces, Newton's third law, energy, work, conservation of energy, kinetic and potential energies, gravitational potential energy, ramps, mechanical advantage)

Epilogue for Chapter 1 31 / Explanation: Removing a Tablecloth from a Table 31 / Chapter Summary and Important Laws and Equations 31

CHAPTER 2 ↓ THE LAWS OF MOTION, PART 2

33

Active Learning Experiment: A Spinning Pie Dish 33

Chapter Itinerary 34

2.1 Seesaws 34

(rotational inertia; angular velocity; torque; angular acceleration; rotational mass; net torque; Newton's first, second, and third laws of rotation; centers of mass and gravity; levers; balance)

2.2 Wheels 48

(friction, traction, ordered and thermal energies, wheels, bearings, kinetic energy, power, rotational work)

2.3 Bumper Cars 59

(momentum, impulse, conservation of momentum, angular momentum, angular impulse, conservation of angular momentum, gradients, potential energy, acceleration, and forces)

Epilogue for Chapter 2 70 / Explanation: A Spinning Pie Dish 70 / Chapter Summary and Important Laws and Equations 70

CHAPTER 3 ↓ **MECHANICAL OBJECTS PART 1****72****Active Learning Experiment: Swinging Water Overhead 72****Chapter Itinerary 73****3.1 Spring Scales 73**

(equilibrium, stable equilibrium, restoring force, Hooke's law, elastic potential energy, oscillation, calibration)

3.2 Ball Sports: Bouncing 79

(collisions, energy transfers, elastic and inelastic collisions, vibration)

3.3 Carousels and Roller Coasters 86

(uniform circular motion, feeling of acceleration, apparent weight, centripetal acceleration)

Epilogue for Chapter 3 94 / Explanation: Swinging Water Overhead 94 / Chapter Summary and Important Laws and Equations 95**CHAPTER 4** ↓ **MECHANICAL OBJECTS PART 2****96****Active Learning Experiment: High-Flying Balls 96****Chapter Itinerary 97****4.1 Bicycles 97**

(stable, neutral, and unstable equilibria; static and dynamic stability; precession)

4.2 Rockets and Space Travel 104

(reaction forces, law of universal gravitation, elliptical orbits, escape velocity, Kepler's laws, speed of light, special and general relativity, equivalence principle)

Epilogue for Chapter 4 117 / Explanation: High-Flying Balls 117 / Chapter Summary and Important Laws and Equations 117**CHAPTER 5** ↓ **FLUIDS****119****Active Learning Experiment: A Cartesian Diver 119****Chapter Itinerary 120****5.1 Balloons 120**

(pressure, density, temperature, thermal motion, absolute zero, Archimedes' principle, buoyant force, ideal gas law)

5.2 Water Distribution 131

(hydrostatics, Pascal's principle, hydraulics, hydrodynamics, steady state flow, streamlines, pressure potential energy, Bernoulli's equation)

Epilogue for Chapter 5 140 / Explanation: A Cartesian Diver 140 / Chapter Summary and Important Laws and Equations 141

CHAPTER 6 ↓ FLUIDS AND MOTION**142****Active Learning Experiment: A Vortex Cannon 142****Chapter Itinerary 143****6.1 Garden Watering 143**

(viscous forces, Poiseuille's law, laminar and turbulent flows, speed and pressure in a fluid, Reynolds number, chaos, momentum in a fluid)

6.2 Ball Sports: Air 153

(aerodynamics, aerodynamic lift and drag, viscous drag, pressure drag, boundary layers, stalls, Magnus and wake deflection forces)

6.3 Airplanes 161

(airfoils, streamlining, lifting wings, angle of attack, induced drag, stalled wings, thrust)

Epilogue for Chapter 6 171 / Explanation: A Vortex Cannon 171 / Chapter Summary and Important Laws and Equations 171**CHAPTER 7 ↓ HEAT AND PHASE TRANSITIONS****173****Active Learning Experiment: A Ruler Thermometer 173****Chapter Itinerary 174****7.1 Woodstoves 174**

(thermal energy, heat, temperature, thermal equilibrium, chemical bonds and reactions, conduction, thermal conductivity, convection, radiation, heat capacity)

7.2 Water, Steam, and Ice 184

(phases of matter, phase transitions, melting, freezing, condensation, evaporation, relative humidity, latent heats of melting and evaporation, sublimation, deposition, boiling, nucleation, superheating)

7.3 Clothing, Insulation, and Climate 192

(thermal conductivity, electromagnetic spectrum, light, blackbody spectrum, emissivity, Stefan-Boltzmann law, thermal expansion, greenhouse effect)

Epilogue for Chapter 7 205 / Explanation: A Ruler Thermometer 206 / Chapter Summary and Important Laws and Equations 206**CHAPTER 8 ↓ THERMODYNAMICS****208****Active Learning Experiment: Making Fog in a Bottle 208****Chapter Itinerary 209****8.1 Air Conditioners 209**

(laws of thermodynamics, temperature, heat, entropy, heat pumps and thermodynamic efficiency)

8.2 Automobiles 219

(heat engines and thermodynamic efficiency)

Epilogue for Chapter 8 228 / Explanation: Making Fog in a Bottle 228 / Chapter Summary and Important Laws and Equations 228**CHAPTER 9****RESONANCE AND MECHANICAL WAVES****230****Active Learning Experiment: A Singing Wineglass 230****Chapter Itinerary 231****9.1 Clocks 231**

(time and space, natural resonance, harmonic oscillators, simple harmonic motion, frequency, period, amplitude)

9.2 Musical Instruments 241

(sound; music; vibrations in strings, air, and surfaces; fundamental and higher-order modes; harmonic and nonharmonic overtones; sympathetic vibration; standing and traveling waves; transverse and longitudinal waves; velocity and wavelength in mechanical waves; superposition; Doppler effect)

9.3 The Sea 254

(tidal forces; surface waves; dispersion, refraction, reflection, and interference in mechanical waves)

Epilogue for Chapter 9 263 / Explanation: A Singing Wineglass 263 /**Chapter Summary and Important Laws and Equations 264****CHAPTER 10****ELECTRICITY****266****Active Learning Experiment: Moving Water without Touching It 266****Chapter Itinerary 267****10.1 Static Electricity 267**

(electric charge, electrostatic forces, Coulomb's law, electrostatic potential energy, voltage, charging by contact, electric polarization, electrical conductors and insulators)

10.2 Xerographic Copiers 276

(electric fields and voltage gradients, electric fields inside and outside conductors, discharges, charging by induction, capacitors)

10.3 Flashlights 287

(electric current; electric circuits; direction of current flow; electrical resistance; voltage drops; voltage rises; relationship among current, voltage, and power; Ohm's law; resistors; series and parallel circuits)

Epilogue for Chapter 10 299 / Explanation: Moving Water without Touching It 300 / Chapter Summary and Important Laws and Equations 301

CHAPTER 11**MAGNETISM AND ELECTRODYNAMICS****302****Active Learning Experiment: A Nail and Wire Electromagnet 302****Chapter Itinerary 303****11.1 Household Magnets 303**

(magnetic pole, magnetostatic forces, Coulomb's law for magnetism, ferromagnetism, magnetic polarization, magnetic domains, magnetic materials, magnetic fields, magnetic flux lines, relationship between currents and magnetic fields)

11.2 Electric Power Distribution 313

(direct and alternating currents, superconductivity, transformers, induction, magnetic field energy, relationship between changing magnetic fields and electric fields, Lenz's law, inductors, induced emf, electrical safety, generators, motors)

Epilogue for Chapter 11 329 / Explanation: A Nail and Wire Electromagnet 330 / Chapter Summary and Important Laws and Equations 330**CHAPTER 12****ELECTROMAGNETIC WAVES****332****Active Learning Experiment: A Disc in the Microwave Oven 332****Chapter Itinerary 333****12.1 Radio 333**

(relationship between changing electric fields and magnetic fields, electric field energy, tank circuits, antennas, electromagnetic waves, speed of light, wave polarization, amplitude modulation, frequency modulation, bandwidth)

12.2 Microwave Ovens 343

(speed, frequency, and wavelength in electromagnetic waves; polar and nonpolar molecules; Lorentz force; cyclotron motion)

Epilogue for Chapter 12 351 / Explanation: A Disc in the Microwave Oven 351 / Chapter Summary and Important Laws and Equations 351**CHAPTER 13****LIGHT****353****Active Learning Experiment: Splitting the Colors of Sunlight 353****Chapter Itinerary 354****13.1 Sunlight 354**

(light, Rayleigh scattering, index of refraction, impedance, refraction, reflection, dispersion, and interference in electromagnetic waves, polarized reflection)

13.2 Discharge Lamps 363

(color vision, primary colors of light and pigment, illumination, gas discharges, quantum physics, wave-particle duality, atomic orbitals, Pauli exclusion principle,

atomic structure, periodic chart, radiative transitions, Planck's constant, atomic fluorescence, radiation trapping)

13.3 LEDs and Lasers 377

(levels in solids; band structure; Fermi level; metals, insulators, and semiconductors; photoconductors; p-n junction; diodes; light-emitting diodes; incoherent and coherent light; spontaneous and stimulated emission; population inversion; laser amplification and oscillation; laser safety)

Epilogue for Chapter 13 390 / Explanation: Splitting the Colors of Sunlight 390 / Chapter Summary and Important Laws and Equations 391

CHAPTER 14 OPTICS AND ELECTRONICS

392

Active Learning Experiment: Magnifying Glass Camera 392

Chapter Itinerary 393

14.1 Cameras 393

(refracting optics, converging lenses, real images, focus, focal lengths, f-numbers, the lens equation, diverging lenses, virtual images, light sensors, vision and vision correction)

14.2 Optical Recording and Communication 403

(analog vs. digital representations, decimal and binary representations, diffraction, diffraction limit, plane and circular polarization, total internal reflection)

14.3 Audio Players 413

(transistors, MOSFETs, bits and bytes, logic elements, amplifiers, feedback)

Epilogue for Chapter 14 422 / Explanation: Magnifying Glass Camera 422 / Chapter Summary and Important Laws and Equations 423

CHAPTER 15 MODERN PHYSICS

425

Active Learning Experiment: Radiation-Damaged Paper 425

Chapter Itinerary 425

15.1 Nuclear Weapons 426

(nuclear structure, Heisenberg uncertainty principle, quantum tunneling, radioactivity, half-life, fission, chain reaction, isotopes, alpha decay, fusion, transmutation of elements, radioactive fallout)

15.2 Nuclear Reactors 438

(controlling nuclear fission, delayed neutrons, thermal fission reactors, moderators, boiling water and pressurized water reactors, fast fission reactors, nuclear reactor safety and accidents, inertial confinement and magnetic confinement fusion)

15.3 Medical Imaging and Radiation 448

(X-rays, X-ray fluorescence, Bremsstrahlung, photoelectric effect, Compton scattering, antimatter, gamma rays, beta decay, fundamental forces, particle accelerators, magnetic resonance)

Epilogue for Chapter 15 458 / Explanation: Radiation-Damaged Paper 458 / Chapter Summary and Important Laws and Equations 459

APPENDICES ↓

460

A Vectors 460

B Units, Conversion of Units 462

Glossary 465

Index 481

Physics is a remarkably practical science. Not only does it explain how things work or why they don't, it also offers great insight into how to create, improve, and repair those things. Because of that fundamental relationship between physics and real objects, introductory physics books are essentially users' manuals for the world in which we live.

Like users' manuals, however, introductory physics books are most accessible when they're based on real-world examples. Both users' manuals and physics texts tend to go unread when they're written like reference works, organized by abstract technical issues, indifferent to relevance, and lacking in useful examples. For practical guidance, most readers turn to "how to" books and tutorials instead; they prefer the "case-study" approach.

How Things Work is an introduction to physics and science that starts with whole objects and looks inside them to see what makes them work. It follows the case-study method, exploring physics concepts on a need-to-know basis in the context of everyday objects. More than just an academic volume, this book is intended to be interesting, relevant, and useful to non-science students.

Most physics texts develop the principles of physics first and present real-life examples of these principles reluctantly if at all. That approach is abstract and inaccessible, providing few conceptual footholds for students as they struggle to understand unfamiliar principles. After all, the comforts of experience and intuition lie in the examples, not in the principles. While a methodical and logical development of scientific principles can be satisfying to the seasoned scientist, it's alien to someone who doesn't even recognize the language being used.

In contrast, *How Things Work* brings science to the reader rather than the reverse. It conveys an understanding and appreciation for physics by finding physics concepts and principles within the familiar objects of everyday experience. Because its structure is defined by real-life examples, this book necessarily discusses concepts as they're needed and then revisits them whenever they reappear in other objects. What better way is there to show the universality of the natural laws?

I wrote this book to be read, not merely referred to. It has always been for nonscientists and I designed it with them in mind. In the seventeen years I have been teaching

How Things Work, many of my thousands of students have been surprised at their own interest in the physics of everyday life, have asked insightful questions, have experimented on their own, and have found themselves explaining to friends and family how things in their world work.

Changes in the Sixth Edition

Content Changes

- **Video figures.** If a picture is worth a thousand words, a video is worth a thousand pictures. That's particularly true for this book because so much of physics is about how things evolve with time. Most students consider themselves visual learners—they need to see what happens in order to understand it. Given that requirement, still images are so 20th century.

In this edition, I have replaced many of the static figures with video figures, using the tools of modern 3D animation and video editing. In print, those video figures are distilled into motionless images but online, they move and talk. Whenever possible and practical, the video figures are quantitatively accurate in both time and space. They're not just cartoons; they're careful models of the real world.

- **Rewriting and editing.** Despite teaching *How Things Work* for almost 25 years, I am still learning how to explain the physics of everyday life. I continue to discover clearer approaches, better analogies, and more effective techniques for conveying understanding and avoiding misconceptions. For this edition, I have examined every word of the book, editing and rewriting it to make sure that it is doing the best job possible.

- **Improved discussions of many physics concepts.** No one book can or should cover all of physics, but whatever physics is included should be presented carefully enough to be worthwhile. In this edition, I have refined the discussions of many physics issues and added some new ones. Look for improved coverage of concepts such as orbits, magnetic induction, and antennas, to name just a few.

The Goals of This Book

As they read this book, students should:

1. Begin to see science in everyday life. Science is everywhere; we need only open our eyes to see it. We're surrounded by things that can be understood in terms of science, much of which is within a student's reach. Seeing science doesn't mean that when viewing an oil painting they should note only the selective reflection of incident light waves by organic and inorganic molecules. Rather, they should realize that there's a beauty to science that complements aesthetic beauty. They can learn to look at a glorious red sunset and appreciate both its appearance and why it exists.

2. Learn that science isn't frightening. The increasing technological complexity of our world has instilled within most people a significant fear of science. As the gulf widens between those who create technology and those who use it, their ability to understand one another and communicate diminishes. The average person no longer tinkers with anything and many modern devices are simply disposable, being too complicated to modify or repair. To combat the anxiety that accompanies unfamiliarity, this book shows students that most objects can be examined and understood, and that the science behind them isn't scary after all. The more we understand how others think, the better off we'll all be.

3. Learn to think logically in order to solve problems. Because the universe obeys a system of well-defined rules, it permits a logical understanding of its behaviors. Like mathematics and computer science, physics is a field of study where logic reigns supreme. Having learned a handful of simple rules, students can combine them logically to obtain more complicated rules and be certain that those new rules are true. So the study of physical systems is a good place to practice logical thinking.

4. Develop and expand their physical intuition. When you're exiting from a highway, you don't have to consider velocity, acceleration, and inertia to know that you should brake gradually—you already have physical intuition that tells you the consequences of doing otherwise. Such physical intuition is essential in everyday life, but it ordinarily takes time and experience to acquire. This book aims to broaden a student's physical intuition to situations they normally avoid or have yet to encounter. That is, after all, one of the purposes of reading and scholarship: to learn from other people's experiences.

5. Learn how things work. As this book explores the objects of everyday life, it gradually uncovers most of the physical laws that govern the universe. It reveals those laws as they were originally discovered: while trying to

understand real objects. As they read this book and learn these laws, students should begin to see the similarities between objects, shared mechanisms, and recurring themes that are reused by nature or by people. This book reminds students of these connections and is ordered so that later objects build on their understanding of concepts encountered earlier.

6. Begin to understand that the universe is predictable rather than magical. One of the foundations of science is that effects have causes and don't simply occur willy-nilly. Whatever happens, we can look backward in time to find what caused it. We can also predict the future to some extent, based on insight acquired from the past and on knowledge of the present. And where predictability is limited, we can understand those limitations. What distinguishes the physical sciences and mathematics from other fields is that there are often absolute answers, free from inconsistency, contraindication, or paradox. Once students understand how the physical laws govern the universe, they can start to appreciate that perhaps the most magical aspect of our universe is that it is not magic; that it is orderly, structured, and understandable.

7. Obtain a perspective on the history of science and technology. None of the objects that this book examines appeared suddenly and spontaneously in the workshop of a single individual who was oblivious to what had been done before. These objects were developed in the context of history by people who were generally aware of what they were doing and usually familiar with any similar objects that already existed. Nearly everything is discovered or developed when related activities make their discoveries or developments inevitable and timely. To establish that historical context, this book describes some of the history behind the objects it discusses.

Visual Media

Because this book is about real things, its videos, illustrations, and photographs are about real things, too. Whenever possible, these visual materials are built around familiar objects so that the concepts they are meant to convey become associated with objects students already know. Many students are visual learners—if they see it, they can learn it. By superimposing the abstract concepts of physics onto simple realistic visuals, this book attempts to connect physics with everyday life. That idea is particularly evident at the opening of each section, where the object examined in that section appears in a carefully rendered drawing. This drawing provides students with something concrete to keep in mind as they encounter the more abstract physical concepts that appear in that section. By lowering the boundaries between what the

students see in the book and what they see in their environment, the rich visual media associated with this book makes science a part of their world.

Features

This printed book contains 40 sections, each of which discusses how something works. The sections are grouped together in 15 chapters according to the major physical themes developed. In addition to the discussion itself, the sections and chapters include a number of features intended to strengthen the educational value of this book. Among these features are:

- **Chapter introductions, experiments, and itineraries.** Each of the 15 chapters begins with a brief introduction to the principal theme underlying that chapter. It then presents an experiment that students can do with household items to observe firsthand some of the issues associated with that physical theme. Lastly, it presents a general itinerary for the chapter, identifying some of the physical issues that will come up as the objects in the chapter are discussed.
- **Section introductions, questions, and experiments.** Each of the 40 sections explains how something works. Often that something is a specific object or group of objects, but it is sometimes more general. A section begins by introducing the object and then asks a number of questions about it, questions that might occur to students as they think about the object and that are answered by the section. Lastly, it suggests some simple experiments that students can do to observe some of the physical concepts that are involved in the object.
- **Check your understanding and check your figures.** Sections are divided into a number of brief segments, each of which ends with a “Check Your Understanding” question. These questions apply the physics of the preceding segment to new situations and are followed by answers and explanations. Segments that introduce important equations also end with a “Check Your Figures” question. These questions show how the equations can be applied and are also followed by answers and explanations.
- **Chapter epilogue and explanation of experiment.** Each chapter ends with an epilogue that reminds the students of how the objects they studied in that chapter fit within the chapter’s physical theme. Following the epilogue is a brief explanation of the experiment suggested at the beginning of the chapter, using physical concepts that were developed in the intervening sections.

• **Chapter summary and laws and equations.** The sections covered in each chapter are summarized briefly at the end of the chapter, with an emphasis on how the objects work. These summaries are followed by a restatement of the important physical laws and equations encountered within the chapter.

• **Chapter exercises and problems.** Following the chapter summary material is a collection of questions dealing with the physics concepts in that chapter. Exercises ask the students to apply those concepts to new situations. Problems ask the students to apply the equations in that chapter and to obtain quantitative results.

• **Three-way approach to the equation of physics.** The laws and equations of physics are the groundwork on which everything else is built. But because each student responds differently to the equations, this book presents them carefully and in context. Rather than making one size fit all, these equations are presented in three different forms. The first is a word equation, identifying each physical quantity by name to avoid any ambiguities. The second is a symbolic equation, using the standard format and notation. The third is a sentence that conveys the meaning of the equation in simple terms and often by example. Each student is likely to find one of these three forms more comfortable, meaningful, and memorable than the others.

• **Glossary.** The key physics terms are assembled into a glossary at the end of the book. Each glossary term is also marked in bold in the text when it first appears together with its contextual definition.

• **Historical, technical, and biographical asides.** To show how issues discussed in this book fit into the real world of people, places, and things, a number of brief asides have been placed in the margins of the text. An appropriate time at which to read a particular aside is marked in the text by a color-coded mark such as ■.

Organization

The 40 sections that make up this book are ordered so that they follow a familiar path through physics: mechanics, heat and thermodynamics, resonance and mechanical waves, electricity and magnetism, light, optics, and electronics, and modern physics. Because there are too many topics here to cover in a single semester, the book is designed to allow shortcuts through the material. In general, the final sections in each chapter and the final chapters in each of the main groups mentioned above can be omitted

without serious impact on the material that follows. The only exceptions to that rule are the first two chapters, which should be covered in their entirety as the introduction to any course taught from this book. The book also divides neatly in half so that it can be used for two independent one-semester courses—the first covering Chapters 1–9 and the second covering Chapters 1, 2, and 10–15. That two-course approach is the one I use myself. A detailed guide to shortcuts appears on the instructor’s website.

WileyPLUS Learning Space With Orion

Within WileyPLUS Learning Space, instructors can organize learning activities, manage student collaboration, and customize their course. Students can collaborate and have meaningful discussions on concepts they are learning. ORION provides students with a personal, adaptive learning experience so they can build their proficiency on concepts and use their study time most effectively. ORION helps students learn by learning about them and providing them with a personalized experience that helps them to pace themselves through the course based on their ongoing performance and level of understanding.

The WileyPLUS Learning Space course includes the following:

- **Online book with extensive video figures and annotation.** Although this book aims to be complete and self-contained, its pages can certainly benefit from additional explanations, answers to open questions, discussions of figures and equations, and real-life demonstrations of objects, ideas, and concepts. Using the web, I can provide all of those features. The online version of this book is annotated with hundreds, even thousands of short videos that bring it to life and enhance its ability to teach.
- **Computer simulations of the book’s objects.** One of the best ways to learn how a violin or nuclear reactor works is to experiment with it, but that’s not

always practical or safe. Computer simulations are the next best thing and the student website includes many simulations of the book’s objects. Associated with each simulation is a sequence of interactive questions that turn it into a virtual laboratory experiment. In keeping with the *How Things Work* concept, the student is then able to explore the concepts of physics in the context of everyday objects themselves.

- **Interactive exercises and problems.** Homework is most valuable when it’s accompanied by feedback and guidance. By providing that assistance immediately, along with links to videos, simulations, the online book, and even additional questions, the website transforms homework from mere assessment into a tutorial learning experience.

For more information, go to: www.wileypluslearningspace.com.

Instructor Companion Website

A broad spectrum of ancillaries are available to support instructors:

- **Test Bank** (Word) – Includes over 800 multiple choice, short answer, and fill-in-the-blank questions.
- **Lecture PowerPoints** highlight topics to help reinforce students’ grasp of essential concepts.
- **Image PowerPoints** contain text images and figures, allowing instructors to customize their presentations and providing additional support for quizzes and exams.
- **Solutions to Selected Exercises and Problems**
- **Additional Web Chapters - Materials Science and Chemical Physics**

To see a complete listing of these ancillaries, or to gain access to them upon adoption and purchase, please visit:

www.wiley.com/college/sc/bloomfield

Acknowledgments

Many people have contributed to this book in one way or another and I am enormously grateful for their help. First among them are my editors, Jessica Fiorillo and Stuart Johnson, who together have guided this project and supported me for twenty years. Jennifer Yee has been amazingly generous with her time and attention in helping me develop this sixth edition, and Sandra Dumas and Jackie Henry have done a fantastic job of shepherding it through production. I’m delighted to have had Thomas Nery working on the graphic design, and Billy Ray working on the photographs. The online component that accompanies the print book would not have been possible or even conceivable without the help of John Duval and Mallory Fryc. And none of this could have happened without the support,

guidance, and encouragement of Christine Kushner, Geraldine Osnato, and Petra Recter. To my numerous other friends and collaborators at John Wiley, many thanks.

I continue to enjoy tremendous assistance from colleagues here and elsewhere who have supported the How Things Work concept, discussed it with me, and often taught the course themselves. They are now too many to list, but I appreciate them all. I am particularly grateful, however, to my colleagues in AMO physics at Virginia, Tom Gallagher, Bob Jones, Olivier Pfister, and Cass Sackett, for more than making up for my reduced scientific accomplishments while working on this project, and to Carl Weiman for his vision of physics education as outlined in the foreword to this book. I must also thank our talented lecture-demonstration group, Al Tobias, Max Bychkov, Mike Timmins, Nikolay Sandev, and Roger Staton, for help to bring physics to life in my class and in the videos for this book.

Of course, the best way to discover how students learn science is to teach it. I am ever so grateful to the students of the University of Virginia for being such eager, enthusiastic, and interactive participants in this long educational experiment. It has been a delight and a privilege to get to know so many of them as individuals and their influence on this enterprise has been immeasurable.

Lastly, this book has benefited more than most from the constructive criticism of a number of talented reviewers. Their candid, insightful comments were sometimes painful to read or hear, but they invariably improved the book. Not only did their reviews help me to present the material more effectively, but they taught me some interesting physics as well. My deepest thanks to all of these fine people:

Brian DeMarco, University of Illinois, Urbana Champaign	Tim Kidd, University of Northern Iowa
Dennis Duke, Florida State University	Judah Levine, University of Colorado, Boulder
Donald R. Franceschetti, University of Memphis	Darryl J. Ozimek, Duquesne University
Alejandro Garcia, San Jose State University	Michael Roth, University of Northern Iowa
Richard Gelderman, Western Kentucky University	Anna Solomey, Wichita State University
Robert B. Hallock, University of Massachusetts, Amherst	Bonnie Wylo, Eastern Michigan University
Mark James, Northern Arizona University	

The real test of this book, and of any course taught from it, is its impact on students' lives long after their classroom days are done. Theirs' is a time both exciting and perilous; one in which physics will play an increasingly important and multifaceted role. It is my sincere hope that their encounter with this book will leave those students better prepared for what lies ahead and will help them make the world a better place in the years to come.

Louis A. Bloomfield
Charlottesville, Virginia
bloomfield@virginia.edu

The Laws of Motion

PART 1

The aim of this book is to broaden your perspectives on familiar objects and situations by helping you understand the science that makes them work. Instead of ignoring that science or taking it for granted, we'll seek it out in the world around us, in the objects we encounter every day. As we do so, we'll discover that seemingly "magical" objects and effects are quite understandable once we know a few of the physical concepts that make them possible. In short, we'll learn about *physics*—the study of the material world and the rules that govern its behavior.

To help us get started, this first pair of chapters will do two main things: introduce the language of physics, which we'll be using throughout the book, and present the basic laws of motion on which everything else will rest. In later chapters, we'll examine objects that are interesting and important, both in their own right and because of the scientific issues they raise. Most of these objects, as we'll see, involve many different aspects of physics and thus bring variety to each section and chapter. These first two chapters are special, though, because they must provide an orderly introduction to the discipline of physics itself.

ACTIVE LEARNING EXPERIMENTS

Removing a Tablecloth from a Table

One famous "magical" effect allows a tablecloth to be removed from a set table without breaking the dishes. The person performing this stunt pulls the tablecloth sideways in one lightning-fast motion. The smooth, slippery tablecloth slides out from under the dishes, leaving them behind and nearly unaffected.

With some practice, you too can do this stunt. Choose a slick, unhemmed tablecloth, one with no flaws that might catch on the dishes. A supple fabric such as silk

helps because you can then pull the cloth slightly downward at the edge of the table. When you get up the nerve to try—with unbreakable dishes, of course—make sure that you pull suddenly and swiftly, so as to minimize the time it takes for the cloth to slide out from under the dishes. Leaving a little slack in the cloth at first helps you get your hands up to speed before the cloth snaps taut and begins to slide off the table. Don't make the mistake of starting slowly or you'll decorate the floor.



Courtesy Lou Bloomfield

Give the tablecloth a yank and watch what happens. With luck, the table will remain set. If it doesn't, try again, but this time go faster or change the types of dishes or the way you pull the cloth.

If you don't have a suitable tablecloth, or any dishes you care to risk, there are many similar experiments you can try. Put several coins on a sheet of paper and whisk that sheet out from under them. Or stack several books on a table and use a stiff ruler to knock out the bottom one. Especially impressive is balancing a

short eraserless pencil on top of a wooden embroidery hoop that is itself balanced on the open mouth of a glass bottle. If you yank the ring away quickly enough, the pencil will be left behind and will drop right into the bottle.

The purpose of this experiment is addressed in a simple question: Why do the dishes stay put as you remove the tablecloth? We'll return to that question at the end of this chapter. In the meantime, we'll explore some of the physics concepts that allow us to answer it.

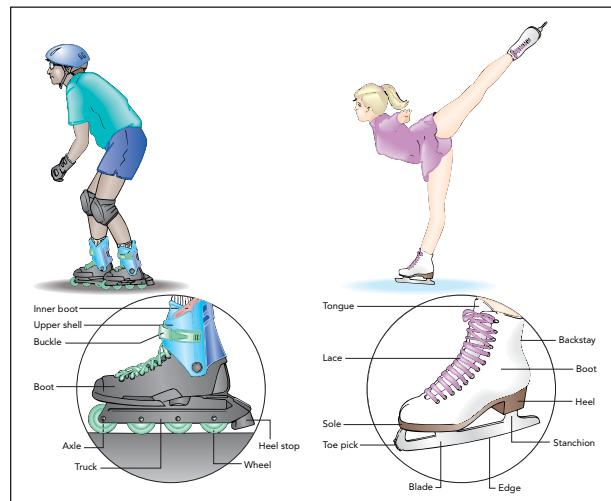
Chapter Itinerary

To examine these concepts, we'll look carefully at three kinds of everyday activities and objects: (1) *skating*, (2) *falling balls*, and (3) *ramps*. In Skating, we see how objects move when nothing pushes on them. In Falling Balls, we find out how that movement can be influenced by gravity. In Ramps, we explore mechanical advantage and how gradual inclines make it possible to lift heavy objects without pushing very hard. For a more complete preview of what we examine in this chapter, flip ahead to the Chapter Summary and Important Laws and Equations at the end of the chapter.

These activities may seem mundane, but understanding them in terms of basic physical laws requires considerable thought. These two introductory chapters will be like climbing up the edge of a high plateau: the ascent won't be easy, and our destination will be hidden from view. However, once we arrive at the top, with the language and basic concepts of physics in place, we'll be able to explain a broad variety of objects with only a small amount of additional effort. And so we begin the ascent.

SECTION 1.1

Skating



Like many sports, skating is trickier than it appears. If you're a first-time skater, you're likely to find yourself getting up repeatedly from the ground or ice, and it will take some practice before you can glide smoothly forward or come gracefully to a stop. But whether you're wearing ice skates or roller skates, the physics of your motion is surprisingly simple. When you're on a level surface with your skates pointing forward, you coast!

Coasting is one of the most basic concepts in physics and our starting point in this book. Joining it in this section will be starting, stopping, and turning, which together will help us

understand the first few laws of motion. We'll leave sloping surfaces for the section on ramps and won't have time to teach you how to do spins or win a race. Nonetheless, our exploration of skating will get us well on the way to an understanding of the fundamental principles that govern all movement and thereby prepare us for many of the objects we'll examine in the rest of this book.

Questions to Think About: What do we mean by "movement"? What makes skaters move, and once they're moving, what keeps them in motion? What does it take to stop a moving skater or turn that skater in another direction?

Experiments to Do: A visit to the ice or roller rink would be ideal, but even a skateboard or a chair with wheels will suffice. Get yourself moving forward on a level surface and then let yourself coast. Why do you keep moving? Is anything pushing you forward? Does your direction ever reverse as you coast? How could you describe the details of your motion to someone on your cell phone? How would you measure your speed?

Before you run into a wall or tree, slow yourself to a stop. What was it that slowed you down? Were you still coasting as you stopped? Did anything push on you as you slowed yourself?

Get yourself moving again. What caused you to speed up? How quickly can you pick up speed, and what do you do differently to speed up quickly? Now turn to one side or the other. Did anything push on you as you turned? What happened to your speed? What happened to your direction of travel?

Gliding Forward: Inertia and Coasting

While you're putting on your skates, let's take a moment to think about what happens to a person who has nothing pushing on her at all. When she's completely free of outside influences (Fig. 1.1.1), free of pushes and pulls, does she stand still? Does she move? Does she speed up? Does she slow down? In short, what does she do?

The correct answer to that apparently simple question eluded people for thousands of years; even Aristotle, perhaps the most learned philosopher of the classical world, was mistaken about it (see 1). What makes this question so tricky is that objects on Earth are never truly free of outside influences; instead, they all push on, rub against, or interact with one another in some way.

As a result, it took the remarkable Italian astronomer, mathematician, and physicist Galileo Galilei many years of careful observation and logical analysis to answer that question 2. The solution he came up with, like the question itself, is simple: if the person is stationary, she will remain stationary; if she is moving in some particular direction, she will continue moving in that direction at a steady pace, following a straight-line path. This property of steady motion in the absence of any outside influence is called **inertia** (Fig. 1.1.2).

INERTIA

A body in motion tends to remain in motion; a body at rest tends to remain at rest.

The main reason that Aristotle failed to discover inertia, and why we often overlook inertia ourselves, is friction. When you slide across the floor in your shoes, friction quickly slows you to a stop and masks your inertia. To make inertia more obvious, we must get rid of friction. That's why you're wearing skates.

Skates almost completely eliminate friction, at least in one direction, so that you can glide effortlessly across the ice or roller rink and experience your own inertia. For simplicity, imagine that your skates are perfect and experience no friction at all as you glide. Also, for this and the next couple of sections, let's forget not only about friction but also about air resistance. Since the air is calm and you're not moving too fast, air resistance isn't all that important to skating anyway.

Now that you're ready to skate, we'll begin to examine five important physical quantities relating to motion and look at their relationships to one another. These quantities are position, velocity, mass, acceleration, and force.

Let's start by describing where you are. At any particular moment, you're located at a **position**—that is, at a specific point in space. Whenever we report your position, it's always as a **distance** and **direction** from some reference point, how many meters north of the refreshment stand or how many kilometers west of Cleveland. For our discussion of skating, we'll choose as our reference point the bench you used while putting on your skates.

Position is an example of a vector quantity. A **vector quantity** consists of both a magnitude and a direction; the **magnitude** tells you how much of the quantity there is, while the direction tells you which way the quantity is pointing. Vector quantities are common in nature. When you encounter one, pay attention to the direction part; if you're looking for buried treasure 30 paces from the old tree but forget that it's due east of that tree, you'll have a lot of digging ahead of you.

You're on your feet and beginning to skate. Once you're moving, your position is changing, which brings us to our second vector quantity—velocity. **Velocity** measures the

Courtesy of Lou Bloomfield



Fig. 1.1.1 This skater glides without any horizontal influences. If she's stationary, she'll tend to remain stationary; if she's moving, she'll tend to continue moving.

1 Aristotle (Greek philosopher, 384–322 BC) theorized that objects' velocities were proportional to the forces exerted on them. While this theory correctly predicted the behavior of a sliding object, it incorrectly predicted that heavier objects should fall faster than lighter objects. Nonetheless, Aristotle's theory was respected for a long time, in part because finding the simpler and more complete theory was hard and in part because the scientific method of relating theory and observation took time to develop.

2 While a professor in Pisa, Galileo Galilei (Italian scientist, 1564–1642) was obliged to teach the natural philosophy of Aristotle. Troubled with the conflict between Aristotle's theory and observations of the world around him, Galileo devised experiments that measured the speeds at which objects fall and determined that all falling objects fall at the same rate.

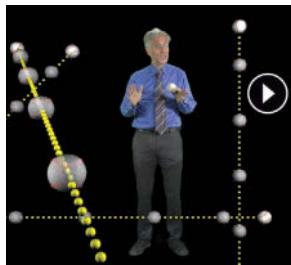


Fig. 1.1.2 These baseballs are in deep space and free from outside influences. Each ball moves according to inertia alone, following a straight-line path at a steady pace.

3 In 1664, while Sir Isaac Newton (English scientist and mathematician, 1642–1727) was a student at Cambridge University, the university was forced to close for 18 months because of the plague. Newton retreated to the country, where he discovered the laws of motion and gravitation and invented the mathematical basis of calculus. These discoveries, along with his observation that celestial objects such as the moon obey the same simple physical laws as terrestrial objects such as an apple (a new idea for the time), are recorded in his *Philosophiae Naturalis Principia Mathematica*, first published in 1687. This book is perhaps the most important and influential scientific and mathematical work of all time.

rate at which your position is changing with time. Its magnitude is your **speed**, the distance you travel in a certain amount of time,

$$\text{speed} = \frac{\text{distance}}{\text{time}}$$

and its direction is the direction in which you're heading.

For example, if you move 2 meters (6.6 feet) west in 1 second, then your velocity is 2 meters per second (6.6 feet per second) toward the west. If you maintain that velocity, your position moves 20 meters west in 10 seconds, 200 meters west in 100 seconds, and so on. Even when you're motionless, you still have a velocity—zero. A velocity of zero is special, however, because it has no direction.

When you're gliding freely, however, with nothing pushing you horizontally, your velocity is particularly easy to describe. Since you travel at a steady speed along a straight-line path, your velocity is constant—it never changes. In a word, you **coast**. And if you happen to be at rest with nothing pushing you horizontally, you remain at rest. Your velocity is constantly zero.

Thanks to your skates, we can now restate the previous description of inertia in terms of velocity: an object that is not subject to any outside influences moves at a constant velocity, covering equal distances in equal times along a straight-line path. This statement is frequently identified as **Newton's first law of motion**, after its discoverer, the English mathematician and physicist Sir Isaac Newton **3**. The outside influences referred to in this law are called **forces**, a technical term for pushes and pulls. An object that moves in accordance with Newton's first law is said to be **inertial**.

NEWTON'S FIRST LAW OF MOTION

An object that is not subject to any outside forces moves at a constant velocity, covering equal distances in equal times along a straight-line path.

INTUITION ALERT: Coasting

Intuition says that when nothing pushes on an object, that object slows to a stop; you must push it to keep it going.

Physics says that when nothing pushes on an object, that object coasts at constant velocity.

Resolution: Objects usually experience hidden forces, such as friction or air resistance, that tend to slow them down. Eliminating those hidden forces is difficult, so that you rarely see the full coasting behavior of force-free objects.

Check Your Understanding #1: A Puck on Ice

Why does a moving hockey puck continue to slide across an ice rink even though no one is pushing on it?

Answer: The puck coasts across the ice because it has inertia.

Why: A hockey puck resting on the surface of wet ice is almost completely free of horizontal influences. If someone pushes on the puck, so that it begins to travel with a horizontal velocity across the ice, inertia will ensure that the puck continues to slide at constant velocity.

The Alternative to Coasting: Acceleration

As you glide forward with nothing pushing you horizontally, what prevents your speed and direction from changing? The answer is your mass. **Mass** is the measure of your inertia, your resistance to changes in velocity. Almost everything in the universe has mass. Mass has no direction, so it's not a vector quantity—that is, a quantity that has only an amount.

Because you have mass, your velocity will change only if something pushes on you—that is, only if you experience a force. You'll keep moving steadily in a straight path until something exerts a force on you to stop you or send you in another direction. **Force** is our third vector quantity, having both a magnitude and a direction. After all, a push to the right is different from a push to the left.

When something pushes on you, your velocity changes; in other words, you accelerate. **Acceleration**, our fourth vector quantity, measures the rate at which your velocity is changing with time (Fig. 1.1.3). Any change in your velocity is acceleration, whether you're speeding up, slowing down, or turning. If either your speed or direction of travel is changing, you're accelerating!

Like any vector quantity, acceleration has a magnitude and a direction. To see how these two parts of acceleration work, imagine that you're at the starting line of a speed-skating race, waiting for it to begin. The starting buzzer sounds, and you're off! You dig your skates into the surface beneath you and begin to accelerate—your speed increases and you cover ground more and more quickly. The magnitude of your acceleration depends on how hard the skating surface pushes you forward. If it's a long race and you're not in a hurry, you take your time getting up to full speed. The surface pushes you forward gently and the magnitude of your acceleration is small. Your velocity changes slowly. However, if the race is a sprint and you need to reach top speed as quickly as possible, you spring forward hard and the surface exerts an enormous forward force on you. The magnitude of your acceleration is large, and your velocity changes rapidly. In this case, you can actually feel your inertia opposing your efforts to pick up speed.

Acceleration has more than just a magnitude, though. When you start the race, you also select a direction for your acceleration—the direction toward which your velocity is shifting with time. This acceleration is in the same direction as the force causing it. If you obtain a forward force from the surface, you'll accelerate forward—your velocity will shift more and more forward. If you obtain a sideways force from the surface, the other racers will have to jump out of your way as you careen into the wall. They'll laugh all the way to the finish line at your failure to recognize the importance of direction in the definitions of force and acceleration.

Once you're going fast enough, you can stop fighting inertia and begin to glide. You coast forward at a constant velocity. Now inertia is helping you; it keeps you moving

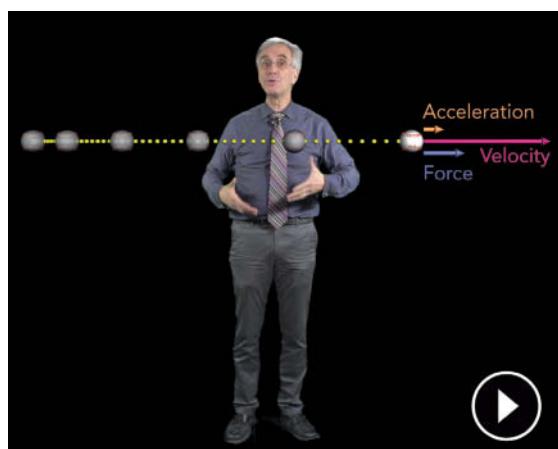


Fig. 1.1.3 A rightward force is causing this baseball to accelerate toward the right. Its velocity is increasing toward the right so that it travels farther with each passing second.

steadily along even though nothing is pushing you forward. (Recall that we're neglecting friction and air resistance. In reality, those effects push you backward and gradually slow you down as you glide.)

Even when you're not trying to speed up or slow down, you can still accelerate. As you steer your skates or go over a bump, you experience sideways or up-down forces that change your *direction of travel* and thus cause you to accelerate.

Finally the race is over, and you skid to a stop. You're accelerating again, but this time in the backward direction—opposite your forward velocity. Although we often call this process *deceleration*, it's just a special type of acceleration. Your forward velocity gradually diminishes until you come to rest.

To help you recognize acceleration, here are some accelerating objects:

1. A runner who's leaping forward at the start of a race—the runner's velocity is changing from zero to forward, so the runner is accelerating *forward*.
2. A bicycle that's stopping at a crosswalk—its velocity is changing from forward to zero, so it's accelerating *backward* (that is, it's decelerating).
3. An elevator that's just starting upward from the first floor to the fifth floor—its velocity is changing from zero to upward, so it's accelerating *upward*.
4. An elevator that's stopping at the fifth floor after coming from the first floor—its velocity is changing from upward to zero, so it's accelerating *downward*.
5. A car that's beginning to shift left to pass another car—its velocity is changing from forward to left-forward, so it's accelerating mostly *leftward*.
6. An airplane that's just beginning its descent—its velocity is changing from level-forward to descending-forward, so it's accelerating almost directly *downward*.
7. Children riding a carousel around in a circle—while their speeds are constant, their directions of travel are always changing. We'll discuss the directions in which they're accelerating in Section 3.3.

Here are some objects that are *not* accelerating:

1. A parked car—its velocity is always zero.
2. A car traveling straight forward on a level road at a steady speed—there is no change in its speed or direction of travel.
3. A bicycle that's climbing up a smooth, straight hill at a steady speed—there is no change in its speed or direction of travel.
4. An elevator that's moving straight upward at a steady pace, halfway between the first floor and the fifth floor—there is no change in its speed or direction of travel.

Seeing acceleration isn't as easy as seeing position or velocity. You can determine a skater's position in a single glance and her velocity by comparing her positions in two separate glances. To observe her acceleration, however, you need at least three glances because you are looking for how her velocity is changing with time. If her speed isn't steady or her path isn't straight, then she's accelerating.

Check Your Understanding #2: Changing Trains

Trains spend much of their time coasting along at constant velocity. When does a train accelerate forward? backward? leftward? downward?

Answer: The train accelerates forward when it starts out from a station, backward when it arrives at the next station, to the left when it turns left, and downward when it begins its descent out of the mountains.

Why: Whenever the train changes its speed or its direction of travel, it is accelerating. When it speeds up on leaving a station, it is accelerating forward (more forward-directed speed). When it slows down at the next station, it is accelerating backward (more backward-directed speed or, equivalently, less forward-directed speed). When it turns left, it is accelerating to the left (more leftward-directed speed). When it begins to descend, it is accelerating downward (more downward-directed speed).

How Forces Affect Skaters

Your friends skate over to congratulate you after the race, patting you on the back and giving you high-fives. They're exerting forces on you, so you accelerate—but how much do you accelerate and in which direction?

First, although each of your friends is exerting a separate force on you, you can't accelerate in response to each force individually. After all, you have only one acceleration. Instead, you accelerate in response to the **net force** you experience—the sum of all the individual forces being exerted on you. Drawing this distinction between individual forces and net force is important whenever an object is experiencing several forces at once. For simplicity now, however, let's wait until you have only one friend left on the ice. When that friend finally pats you on the back, you experience only that one force, so it is the net force on you and it causes you to accelerate.

Your acceleration depends on the strength of that net force: the stronger the net force, the more you accelerate. However, your acceleration also depends on your mass: the more massive you are, the less you accelerate. For example, it's easier to change your velocity before you eat Thanksgiving dinner than afterward.

There is a simple relationship among the net force exerted on you, your mass, and your acceleration. Your acceleration is equal to the net force exerted on you divided by your mass or, as a word equation,

$$\text{acceleration} = \frac{\text{net force}}{\text{mass}}. \quad (1.1.1)$$

Your acceleration, as we've seen, is in the same direction as the net force on you.

This relationship was deduced by Newton from his observations of motion and is referred to as **Newton's second law of motion**. Structuring the relationship this way sensibly distinguishes the causes (net force and mass) from their effect (acceleration). However, it has become customary to rearrange this equation to eliminate the division. The relationship then takes its traditional form, which can be written in a word equation:

$$\text{net force} = \text{mass} \cdot \text{acceleration} \quad (1.1.2)$$

in symbols:

$$\mathbf{F}_{\text{net}} = m \cdot \mathbf{a},$$

and in everyday language:

Throwing a baseball is much easier than throwing a bowling ball (Fig. 1.1.4).

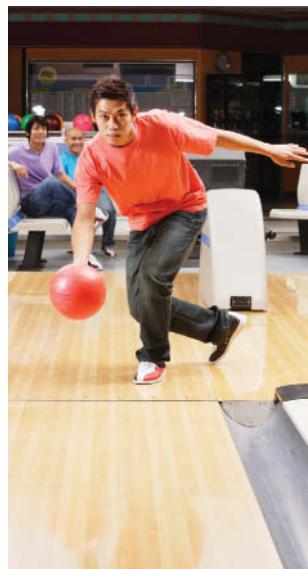
Remember that in Eq. 1.1.2 the direction of the acceleration is the same as the direction of the net force.

● NEWTON'S SECOND LAW OF MOTION

**The net force exerted on an object is equal to that object's mass times its acceleration.
The acceleration is in the same direction as the net force.**

Because it's an equation, the two sides of Eq. 1.1.1 are equal. Your acceleration equals the net force on you divided by your mass. Since your mass is constant unless you visit the snack bar, Eq. 1.1.1 indicates that an increase in the net force on you is accompanied by a

© Kent C. Horner/Getty Images, Inc.



© Asia Images Group/Getty Images, Inc.

Fig. 1.1.4 A baseball accelerates easily because of its small mass. A bowling ball has a large mass and is harder to accelerate.

corresponding increase in your acceleration. That way, as the right side of the equation increases, the left side increases to keep the two sides equal. Thus the harder your friend pushes on you, the more rapidly your velocity changes in the direction of that push.

We can also compare the effects of equal forces on two different masses, for example, you and the former sumo wrestler to your left. I'll assume, for the sake of argument, that you're the less massive of the two. Equation 1.1.1 indicates that an increase in mass must be accompanied by a corresponding decrease in acceleration. Sure enough, your velocity changes more rapidly than the velocity of the sumo wrestler when the two of you are subjected to identical forces (Fig. 1.1.5).

So far we've explored five principles:

1. Your position indicates exactly where you're located.
2. Your velocity measures the rate at which your position is changing with time.
3. Your acceleration measures the rate at which your velocity is changing with time.
4. To accelerate, you must experience a net force.
5. The greater your mass, the less acceleration you experience for a given net force.

We've also encountered five important physical quantities—mass, force, acceleration, velocity, and position—as well as some of the rules that relate them to one another. Much of the groundwork of physics rests on these five quantities and on their interrelationships.

Skating certainly depends on these quantities. We can now see that, in the absence of any horizontal forces, you either remain stationary or coast along at a constant velocity. To start, stop, or turn, something must push you horizontally and that something is the ice or pavement. We haven't talked about how you obtain horizontal forces from the ice or pavement, and we'll leave that problem for later sections. As you skate, however, you should be aware of these forces and notice how they change your speed, direction of travel, or both. Learn to watch yourself accelerate.



Check Your Understanding #3: Hard to Stop

It's much easier to stop a bicycle traveling toward you at 5 kilometers per hour (3 miles per hour) than an automobile traveling toward you at the same velocity. What accounts for this difference?

Answer: An automobile has a much greater mass than a bicycle.

Why: To stop a moving vehicle, you must exert a force on it in the direction opposite its velocity. The vehicle will then accelerate backward so that it eventually comes to rest. If the vehicle is heading toward you, you must push it away from you. The more mass the vehicle has, the less it will accelerate in response to a certain force and the longer you will have to push on it to stop it completely. Although it's easy to stop a bicycle by hand, stopping even a slowly moving automobile by hand requires a large force exerted for a substantial amount of time.

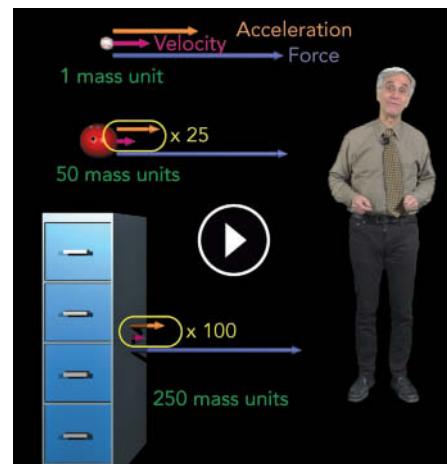


Fig. 1.1.5 A baseball, bowling ball, and file cabinet have different masses and accelerate quite differently in response to equal rightward forces. Arrows representing the accelerations and velocities of the bowling ball and file cabinet are magnified to make them visible.

Check Your Figures #1: At the Bowling Alley

Bowling balls come in various masses. Suppose that you try bowling with two different balls, one with twice the mass of the other. If you push on them with equal forces, which one will accelerate faster and how much faster?

Answer: The less massive ball will accelerate twice as rapidly.

Why: Equation 1.1.1 shows that an object's acceleration is inversely proportional to its mass:

$$\text{acceleration} = \frac{\text{force}}{\text{mass}}.$$

If you push on both bowling balls with equal forces, then their accelerations will depend only on their masses. Doubling the mass on the right side of this equation halves the acceleration on the left side. That means that the more massive ball will accelerate only half as quickly as the other ball.

Several Skaters: Frames of Reference

While skating alone is peaceful, it's usually more fun with other skaters around. That way, you have people to talk to and an audience for your athleticism and artistry.

However, with several skaters coasting on the ice at once, there's a question of perspective. As you glide steadily past a friend, the two of you see the world somewhat differently. From your perspective, you are motionless and your friend is moving. From your friend's perspective, though, your friend is motionless and you are moving. Who is right?

It turns out that you're both right and that physics has a way of accommodating this apparent paradox. Each of you is observing the world from a different **inertial frame of reference**, the viewpoint of an *inertial* object—an object that is not accelerating and that moves according to Newton's first law. One of the remarkable discoveries of Galileo and Newton is that the laws of physics work perfectly in any inertial frame of reference (Fig. 1.1.6). From an inertial frame, everything you see in the world around you obeys the laws of motion that we're in the process of exploring. Although you may find it odd to think of scenery as moving, your inertial frame of reference is as good as any and in your frame you are at rest amid the moving landscape.

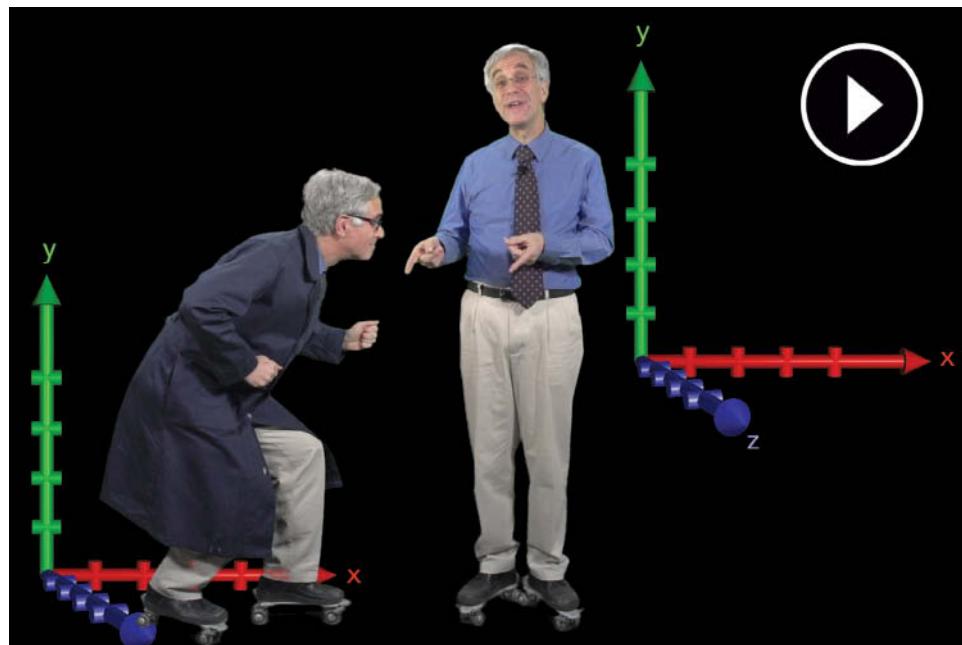


Fig. 1.1.6 Two skaters who are moving relative to one another, but not accelerating, have different inertial frames of reference. Although each skater will have a different coordinate system in which to measure physical quantities, the laws of physics will correctly describe what that skater observes from his own inertial frame of reference.

Since both you and your friend are coasting, each of you views the world from an inertial frame of reference and sees the surrounding objects moving in perfect accordance with the laws of motion. Some objects travel at constant velocity, while others accelerate in response to forces. However, because the two of you are observing those objects from different inertial frames, you will disagree on the particular values of some of the physical quantities you might measure.

In the present case, you see yourself as motionless because you view the world from your own inertial frame. In that frame, your friend is coasting westward at 2 meters per second (2 m/s, which is roughly 6.6 feet per second, or 6.6 ft/s). However, your friend sees things differently. In your friend's inertial frame, your friend is motionless and you are coasting eastward at 2 m/s. As long as the two of you don't try to compare the positions or velocities of the objects you observe, or certain physical quantities derived from those values, there will be no disagreements and no inconsistencies. However, if you forget to watch where you're going and crash into a wall, don't expect your friend to sympathize when you claim that you were motionless and that the moving wall ran into you. That's not how your friend saw it.

Each time we examine an object in this book, we'll pick a specific inertial frame of reference from which to view that object. We'll normally select an inertial frame that makes the object and its motions appear as simple as possible and then stick with that frame consistently. The best choice of inertial frame will usually be so obvious that we'll adopt it without even a moment's thought. On occasion, however, we'll have to pick the frame carefully and deliberately. Finally, although there are formal methods for working with two or more inertial frames at once, I'll leave that for another book.

Check Your Understanding #4: Two Views

You are standing on the sidewalk, watching a train coast eastward at constant velocity. Your friend is riding in that train. In her inertial frame of reference, the sweater in her lap is motionless. Describe the sweater's motion in *your* inertial frame of reference.

Answer: The sweater is coasting eastward at constant velocity.

Why: Although both of you agree that the sweater is not accelerating and that it is moving according to Newton's first law, you disagree on its specific velocity. She sees the sweater at rest, while you see it coasting eastward at constant velocity. Your viewpoints are equally valid.

Measure for Measure: The Importance of Units

If you went to the grocery store and asked for “6 of sugar,” the clerk wouldn’t know how much sugar to give you. The number 6 wouldn’t be enough information; you need to specify which units—cups, pounds, cubes, or tons—you have in mind. This need to specify units applies to almost all physical quantities—velocity, force, mass, and so on—and has led our society to develop units that everyone agrees on, also known as **standard units**.

For example, when you say that a skater’s speed is 20 miles per hour, you have chosen *miles per hour* as the standard unit of speed and you’re asserting that the skater is moving 20 times that fast. You can report the skater’s speed as a multiple of any standard unit of speed—feet per second, yards per day, or inches per century, to name only a few—and you can always find a simple relationship to convert from one unit of speed to another. For example, to convert the skater’s speed from miles per hour to kilometers per hour, you multiply it by 1.609 kilometers/mile:

$$20 \frac{\text{mi}}{\text{h}} \cdot \frac{1.609 \text{ km}}{1 \text{ mi}} = 32.2 \frac{\text{km}}{\text{h}}$$

Many of the common units in the United States come from the old **English system of units**, which most of the world has abandoned in favor of **SI units** (*Système Internationale*

d'Unités). The continued use of English units in the United States often makes life difficult. If you have to triple a cake recipe that calls for $\frac{3}{4}$ cup of milk, you must work hard to calculate that you need $2\frac{1}{4}$ cups. Then you go to buy $2\frac{1}{4}$ cups of milk, which is slightly more than half a quart, but end up buying 2 pints instead. You now have 14 ounces of milk more than you need—but is that 14 fluid ounces or 14 ounces of weight? And so it goes.

The SI system has two important characteristics that distinguish it from the English system and make it easier to use. In the SI system:

1. Different units for the same physical quantity are related by factors of 10.
2. Most of the units are constructed out of a few basic units: the meter, the kilogram, and the second.

Let's start with the first characteristic: different units for the same physical quantity are related by factors of 10. When measuring volume, 1000 milliliters is exactly 1 liter and 1000 liters is exactly 1 cubic meter (1 meter³). When measuring mass, 1000 grams is exactly 1 kilogram and 1000 kilograms is exactly 1 metric ton. Because of this consistent relationship, enlarging a recipe that's based on the SI system is as simple as multiplying a few numbers. You never have to think about converting pints into quarts, teaspoons into tablespoons, or ounces into pounds. Instead, if you want to triple a recipe that calls for 500 milliliters of sugar, you just multiply the recipe by 3 to obtain 1500 milliliters of sugar. Since 1000 milliliters is 1 liter, you'll need 1.5 liters of sugar. Converting milliliters to liters is as simple as multiplying by 0.001 liter/milliliter. (See Appendix B, online, for more conversion factors.)

SI units remain somewhat mysterious to many U.S. residents, even though some of the basic units are slowly appearing on our grocery shelves and highways. As a result, although the SI system really is more sensible than the old English system, developing a feel for some SI units is still difficult. How many of us know our heights in meters (the SI unit of length) or our masses in kilograms (the SI unit of mass)? If your car is traveling 200 kilometers per hour and you pass a police car, are you in trouble? Yes, because 200 kilometers per hour is about 125 miles per hour. Actually, the hour is not an SI unit—the SI unit of time is the second—but the hour remains customary for describing long periods of time. Thus the kilometer per hour is a unit that is half SI (the *kilometer* part) and half customary (the *hour* part).

The second characteristic of the SI system is its relatively small number of basic units. So far, we've noted the SI units of mass (the **kilogram**, abbreviated kg), length (the **meter**, abbreviated m), and time (the **second**, abbreviated s). One kilogram is about the mass of a liter of water; 1 meter is about the length of a long stride; 1 second is about the time it takes to say “one banana.” From these three basic units, we can create several others, such as the SI units of velocity (the **meter per second**, abbreviated m/s) and acceleration (the **meter per second squared**, abbreviated m/s²). One meter per second is a healthy walking speed; 1 meter per second² is about the acceleration of an elevator after the door closes and it begins to move upward. This conviction that many units are best constructed out of other, more basic units dramatically simplifies the SI system; the English system doesn't usually suffer from such sensibility.

The SI unit of force is also constructed out of the basic units of mass, length, and time. If we choose a 1-kilogram object and ask just how much force is needed to make that object accelerate at 1 meter per second², we define a specific amount of force. Since 1 kilogram is the SI unit of mass and 1 meter per second² is the SI unit of acceleration, it's only reasonable to let the force that causes this acceleration be the SI unit of force, the **kilogram-meter per second²**. Since this composite unit sounds unwieldy but is very important, it has been given its own name: the **newton** (abbreviated N)—after, of course, Sir Isaac, whose second law defines the relationship among mass, length, and time that the unit expresses. Conveniently, 1 newton is about the weight of a small apple; that is, if you hold that apple steady in your hand, you'll feel a downward force of about 1 newton.

Because a complete transition to the SI system will take generations, this book uses both unit systems whenever possible. Although I will emphasize the SI system, English and customary units may give you a better intuitive feel for a particular physical quantity. A bullet train traveling “67 meters per second” doesn’t mean much to most of us, whereas one moving “150 miles per hour” (150 mph) or “240 kilometers per hour” (240 km/h) should elicit our well-deserved respect.

Quantity	SI Unit	English Unit	SI → English	English → SI
Position	meter (m)	foot (ft)	1 m = 3.2808 ft	1 ft = 0.30480 m
Velocity	meter per second (m/s)	foot per second (ft/s)	1 m/s = 3.2808 ft/s	1 ft/s = 0.30480 m/s
Acceleration	meter per second ² (m/s ²)	foot per second ² (ft/s ²)	1 m/s ² = 3.2808 ft/s ²	1 ft/s ² = 0.30480 m/s ²
Force	newton (N)	pound-force (lbf) [†]	1 N = 0.22481 lbf	1 lbf = 4.4482 N
Mass	kilogram (kg)	pound-mass (lbm) [†]	1 kg = 2.2046 lbm	1 lbm = 0.45359 kg

[†] The English units of force and mass are both called the *pound*. To distinguish these two units, it has become standard practice to identify them explicitly as pound-mass and pound-force.

Check Your Understanding #5: Going for a Walk

If you’re walking at a pace of 1 m per second, how many miles will you travel in an hour?

Answer: About 2.24 miles.

Why: There are many different units in this example, so we must do some converting. First, an hour is 3600 s, so in an hour of walking at 1 m per second you will have walked 3600 m. Second, a mile is about 1609 m, so each time you travel 1609 m you have traveled 1 mile. By walking 3600 m, you have completed 2 miles and are about one-quarter of the way into your third mile.

SECTION 1.2

Falling Balls



We’ve all dropped balls from our hands or seen them arc gracefully through the air after being thrown. These motions are simplicity itself and, not surprisingly, they’re governed by only a few universal rules. We encountered several of those rules in the previous section, but we’re about to examine our first important type of force—gravity. Like Newton, who reportedly began his investigations after seeing an apple fall from a tree,

we’ll start simply by exploring gravity and its effects on motion in the context of falling objects.

Questions to Think About: What do we mean by “falling,” and why do balls fall? Which falls faster, a heavy ball or a light ball? Can a ball that’s heading upward still be falling? How does gravity affect a ball that’s thrown sideways?

Experiments to Do: A few seconds with a baseball will help you see some of the behaviors that we’ll be exploring. Toss the ball into the air to various heights, catching it in your hand as it returns. Have a friend time the flight of the ball. As you toss the ball higher, how much more time does it spend in the air? How does it feel coming back to your hands? Is there any difference in the impact it makes? Which takes the ball longer: rising from your hand to its peak height or returning from its peak height back to your hand?

Now drop two different balls—a baseball, say, and a golf ball. If you drop them simultaneously, without pushing either one up or down, does one ball strike the ground first or do they arrive together? Now throw one ball horizontally while dropping the second. If they both leave your hands at the same time and the first one’s initial motion is truly horizontal, which one reaches the ground first?

Weight and Gravity

Like everything else around us, a ball has a weight. For example, a golf ball weighs about 0.45 N (0.10 lbf)—but what is weight? Evidently it's a force, since both the newton (N) and the pound-force (lbf) are units of force. To understand what weight is, however—and, in particular, where it comes from—we need to look at gravity.

Gravity is a physical phenomenon that produces attractive forces between every pair of objects in the universe. In our daily lives, however, the only object massive enough and near enough to have obvious gravitational effects on us is our planet, Earth. Gravity weakens with distance; the moon and sun are so far away that we notice their gravities only through such subtle effects as the ocean tides.

Earth's gravity exerts a downward force on any object near its surface. That object is attracted directly toward the center of Earth with a force we call the object's **weight** (Fig. 1.2.1). Remarkably enough, this weight is exactly proportional to the object's mass—if one ball has twice the mass of another ball, it also has twice the weight. Such a relationship between weight and mass is astonishing because weight and mass are very different attributes: weight is how hard gravity pulls on a ball, and mass is how difficult that ball is to accelerate. Because of this proportionality, a ball that's heavy is also hard to shake back and forth!

An object's weight is also proportional to the local strength of gravity, which is measured by a downward vector called the **acceleration due to gravity**—an odd name that I'll explain shortly. At the surface of Earth, the acceleration due to gravity is about 9.8 N/kg (1.0 lbf/lbm). That value means that a mass of 1 kilogram has a weight of 9.8 newtons and that a mass of 1 pound-mass has a weight of 1 pound-force.

More generally, an object's weight is equal to its mass times the acceleration due to gravity, which can be written as a word equation:

$$\text{weight} = \text{mass} \cdot \text{acceleration due to gravity}, \quad (1.2.1)$$

in symbols:

$$\mathbf{w} = m \cdot \mathbf{g},$$

and in everyday language:

You can lose weight either by reducing your mass or by going someplace, like a small planet, where the gravity is weaker.

But why *acceleration* due to gravity? What acceleration do we mean? To answer that question, let's consider what happens to a ball when you drop it.

If the only force on the ball is its weight, the ball accelerates downward; in other words, it falls. Although a ball moving through Earth's atmosphere encounters additional forces due to the air, let's ignore those forces for the time being. Doing so costs us only a little in terms of accuracy—the air's forces on the ball are negligible as long as the ball is dense and its speed relatively small—and allows us to focus exclusively on the effects of gravity.

How much does the falling ball accelerate? According to Eq. 1.1.1, the ball's acceleration is equal to the net force exerted on it divided by its mass. Because the ball is *falling*, however, the only force on it is its own weight. That weight, according to Eq. 1.2.1, is equal to the ball's mass times the acceleration due to gravity. Using a little algebra, we get

$$\begin{aligned} \text{falling ball's acceleration} &= \frac{\text{ball's weight}}{\text{ball's mass}} \\ &= \frac{\text{ball's mass} \cdot \text{acceleration due to gravity}}{\text{ball's mass}} \\ &= \text{acceleration due to gravity}. \end{aligned}$$



Fig. 1.2.1 A dropped baseball experiences only its weight—the force due to gravity. It accelerates downward.

As you can see, the falling ball's acceleration is equal to the acceleration due to gravity. Thus, acceleration due to gravity really is an acceleration after all: it's the acceleration of a freely falling object. Moreover, the units of acceleration due to gravity can be transformed easily from those relating weight to mass, 9.8 N/kg (1.0 lbf/lbm), into those describing the acceleration of free fall, 9.8 m/s^2 (32 ft/s^2).

Thus a ball falling near Earth's surface experiences a downward acceleration of 9.8 m/s^2 (32 ft/s^2), regardless of its mass (Fig. 1.2.2). This downward acceleration is substantially more than that of an elevator starting its descent. When you drop a ball, it picks up speed very quickly in the downward direction.

Because all falling objects at Earth's surface accelerate downward at exactly the same rate, a billiard ball and a bowling ball dropped simultaneously from the same height will reach the ground together. (Remember that we're not considering forces due to the air yet.) Although the bowling ball weighs more than the billiard ball, it also has more mass; so while the bowling ball experiences a larger downward force, its larger mass ensures that its downward acceleration is equal to that of the lighter and less massive billiard ball.

Check Your Understanding #1: Weight and Mass

Out in deep space, far from any celestial object that exerts significant gravity, would an astronaut weigh anything? Would that astronaut have a mass?

Answer: The astronaut would have zero weight but would still have a normal mass.

Why: Weight is a measure of the force exerted on the astronaut by gravity. Far from Earth or any other large object, the astronaut would experience virtually no gravitational force and would have zero weight. But mass is a measure of inertia and doesn't depend at all on gravity. Even in deep space, it would be much harder to accelerate a school bus than to accelerate a baseball because the school bus has more mass than the baseball.

Check Your Figures #1: Weighing In on the Moon

You're in your spacecraft on the surface of the moon. Before getting into your suit, you weigh yourself and find that your moon weight is almost exactly one-sixth your Earth weight. What is the moon's acceleration due to gravity?

Answer: It is about 1.6 m/s^2 (5.3 ft/s^2).

Why: You can rearrange Eq. 1.2.1 to show that the acceleration due to gravity is proportional to an object's weight:

$$\text{acceleration due to gravity} = \frac{\text{weight}}{\text{mass}}.$$

Your mass doesn't change in going to the moon, so any change in your weight must be due to a change in the acceleration due to gravity. Since your moon weight is one-sixth of your Earth weight, the moon's acceleration due to gravity must be one-sixth that of Earth, or about 1.6 m/s^2 .



Fig. 1.2.2 A baseball (left) and a golf ball (right) both accelerate downward at the acceleration due to gravity. Their differences in weight and mass perfectly compensate for one another.

The Velocity of a Falling Ball

We're now ready to examine the motion of a falling ball near Earth's surface. A falling ball is one that has only its weight, the force due to gravity, acting on it and gravity, as we've seen, causes any falling object to accelerate downward at a constant rate. However, we're usually less interested in the falling object's acceleration than we are in its position and velocity. Where will the object be in 3 s, and what will its velocity be then? When you're trying to summon up the courage to jump off the high dive, you want to know how long it'll take you to reach the water and how fast you'll be going when you hit.

The first step in answering these questions is to look at how a ball's velocity is related to the time you've been watching it fall. To do that, you'll need to know the ball's *initial velocity*—that is, its speed and direction at the moment you start watching it. If you drop the ball from rest, its initial velocity is zero.

You can then describe the ball's present velocity in terms of its initial velocity, its acceleration, and the time that has passed since you started watching it. Because a constant acceleration causes the ball's velocity to change by the same amount each second, the ball's present velocity differs from its initial velocity by the acceleration times the time that you've been watching it. We can relate these quantities as a word equation:

$$\text{present velocity} = \text{initial velocity} + \text{acceleration} \cdot \text{time}, \quad (1.2.2)$$

in symbols:

$$v_f = v_i + \mathbf{a} \cdot t,$$

and in everyday language:

A stone dropped from rest descends faster with each passing second, but you can give it a boost by throwing it downward instead of just letting go.

For a ball falling from rest, the initial velocity is zero, the acceleration is downward at 9.8 m/s^2 (32 ft/s^2), and the time you've been watching it is simply the time since it started to drop (Fig. 1.2.3). After 1 s, the ball has a downward velocity of 9.8 m/s (32 ft/s). After 2 s, the ball has a downward velocity of 19.6 m/s (64 ft/s). After 3 s, its downward velocity is 29.4 m/s (96 ft/s), and so on.

Because a ball falls in a three-dimensional world, Eq. 1.2.2 is a relationship among *vector* quantities. When a ball is dropped from rest, however, its motion is strictly vertical and all of those vector quantities point either up or down. In that special case of motion along a vertical line, we can represent vector quantities that point up by ordinary positive values and vector quantities that point down by ordinary negative values. The acceleration due to gravity is then -9.8 m/s^2 , and a ball that has been falling from rest for 2 s has a velocity of -19.6 m/s . As we'll see later in this section, those simplified values are actually the *upward components* of the ball's acceleration and velocity.

Check Your Understanding #2: Half a Fall

You drop a marble from rest, and after 1 s, its velocity is 9.8 m/s (32 ft/s) in the downward direction. What was its velocity after only 0.5 s of falling?

Answer: 4.9 m/s (16 ft/s) in the downward direction.

Why: A freely falling object accelerates downward at a steady rate. Its velocity changes by 9.8 m/s (16 ft/s) in the downward direction each and every second. In half a second, the marble's velocity changes by only half that amount, or 4.9 m/s (16 ft/s).

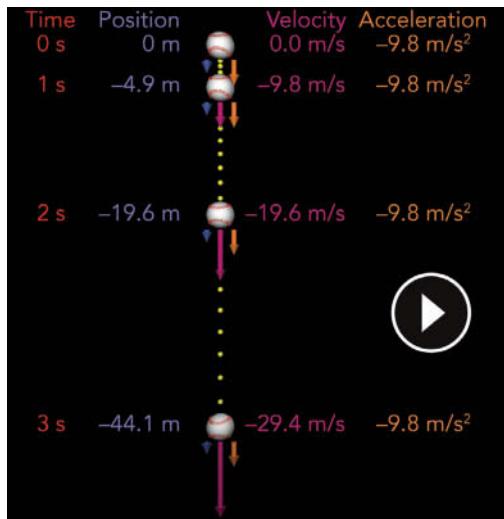


Fig. 1.2.3 The moment you let go of a ball that was resting in your hand, it begins to fall. Its weight causes it to accelerate downward. After 1 s, it has fallen 4.9 m and has a velocity of 9.8 m/s downward. After 2 s, it has fallen 19.6 m and has a velocity of 19.6 m/s downward, and so on. As the ball continues to accelerate downward, its velocity continues to increase downward. Negative values for the position and velocity are meant to indicate downward movement caused by a negative or downward acceleration.

Check Your Figures #2: The High Dive

If it takes you about 1.4 s to reach the water from the 10-m (32-ft) diving platform, how fast are you going just before you enter the water?

Answer: About 14 m/s (45 ft/s, 50 km/h, or 31 mph).

Why: The downward acceleration due to gravity is 9.8 m/s^2 (32 ft/s^2). You fall for 1.4 s, during which time your velocity increases steadily in the downward direction. Since you start with zero velocity, Eq. 1.2.2 gives a final velocity of

$$\text{final velocity} = 9.8 \text{ m/s}^2 \cdot 1.4 \text{ s} = 13.72 \text{ m/s.}$$

Since the time of the fall is given only to two digits of accuracy (1.4 s could really be 1.403 s or 1.385 s), we shouldn't claim that our calculated final velocity is accurate to four digits. We should round the value to 14 m/s (45 ft/s).

The Position of a Falling Ball

The ball's velocity continues to increase as it falls, but where exactly is the ball located? To answer that question, you need to know the ball's *initial position*—that is, where it was when you started to watch it fall. If you dropped the ball from rest, the initial position was your hand and you can define that spot as 0.

You can then describe the ball's present position in terms of its initial position, its initial velocity, its acceleration, and the time that has passed since you started watching it. However, because the ball's velocity is changing, you can't simply multiply its present velocity by the time that it's been falling to determine how much the ball's present position differs from its initial position. Instead, you must use the ball's average velocity during the whole period you've been watching it. Since the ball's velocity has been changing uniformly from its initial velocity to its present velocity, the ball's average velocity is exactly halfway in between the two individual velocities:

$$\text{average velocity} = \text{initial velocity} + \frac{1}{2} \cdot \text{acceleration} \cdot \text{time.}$$

The ball's present position differs from its initial position by this average velocity times the time that you've been watching it. We can relate these quantities as a word equation:

$$\text{present position} = \text{initial position} + \text{initial velocity} \cdot \text{time} + \frac{1}{2} \cdot \text{acceleration} \cdot \text{time}^2, \quad (1.2.3)$$

in symbols:

$$\mathbf{x}_f = \mathbf{x}_i + \mathbf{v}_i \cdot t + \frac{1}{2} \cdot \mathbf{a} \cdot t^2,$$

and in everyday language:

The longer a stone has been falling, the more its height diminishes with each passing second. However, it won't overtake a stone that was dropped next to it at an earlier time or dropped from beneath it at the same time.

For a ball falling from rest, the initial velocity is zero, the acceleration is downward at 9.8 m/s^2 (32 ft/s^2), and the time you've been watching it is simply the time since it started to drop (Fig. 1.2.3). After 1 s, the ball has fallen 4.9 m (16 ft). After 2 s, it has fallen a total of 19.6 m (64 ft). After 3 s, it has fallen a total of 44.1 m (145 ft), and so on.

Equations 1.2.2 and 1.2.3 depend on the definition of acceleration as the measure of how quickly *velocity* changes and the definition of velocity as the measure of how quickly *position* changes. Because the acceleration of a falling ball doesn't change with time, the two equations can be derived using algebra. In more complicated situations, however, where an object's acceleration changes with time, predicting position and velocity usually requires the use of calculus. *Calculus* is the mathematics of change, invented by Newton to address just these sorts of problems.

We've been discussing what happens to a falling ball, but we could have chosen another object instead. Everything falls the same way; heavy or light, large or small, all objects take the same amount of time to fall a given distance, as long as they're dense enough to be unaffected by the air. If there were no air, this statement would be exactly true for any object; a feather and a lead brick would plummet downward together if you dropped them simultaneously.

Now that we've explored acceleration due to gravity, we can see why a ball dropped from a tall ladder is more dangerous than the same ball dropped from a short stool. The farther the ball has to fall, the longer it takes to reach the ground and the more time it has to accelerate. During its long fall from the tall ladder, the ball acquires a large downward velocity and becomes very hard to stop. If you try to catch it, you'll have to exert a very large upward force on it to accelerate it upward and bring it to rest quickly. Exerting that large upward force may hurt your hand.

The same notion holds if you're the falling object. If you leap off a tall ladder, a substantial amount of time will pass before you reach the ground. By the time you arrive, you'll have acquired considerable downward velocity. The ground will then accelerate you upward and bring you to rest with a very large and unpleasant upward force. (For an interesting and less painful application of long falls, see 4.)

4 In 1782, William Watts, a plumber from Bristol, England, patented a technique for forming perfectly spherical, seamless lead shot for use in guns. His idea was to pour molten lead through a sieve suspended high above a pool of water. The lead droplets cool in the air as they fall, solidifying into perfect spheres before reaching the water. Shot towers based on this idea soon appeared throughout Europe and eventually in the United States.

Nowadays, iron shot has all but replaced environmentally dangerous lead shot. Iron shot is cast, rather than dropped, because the longer cooling time needed to solidify molten iron would require impractically tall shot towers.

Check Your Understanding #3: Half a Fall Again

You drop a marble from rest, and after 1 s, it has fallen downward a distance of 4.9 m (16 ft). How far had it fallen after only 0.5 s?

Answer: It had fallen about 1.2 m (4 ft).

Why: While a freely falling object's velocity changes steadily in the downward direction, its change in height is more complicated. When you drop the marble from rest, it starts its descent slowly but picks up speed and covers the downward distance faster and faster. In the first 0.5 s, it travels only a quarter of the distance it travels in the first 1 s, or about 1.2 m (4 ft).

Check Your Figures #3: Extreme Physics

You're planning to construct a bungee-jumping amusement at the local shopping center. If you want your customers to have a 5-s free-fall experience, how tall will you need to build the tower from which they'll jump? (Don't worry about the extra height needed to stop people after the bungee pulls taut.)

Answer: The tower should be about 122 m (402 ft or as high as a 40-story building).

Why: As they fall, the jumpers will travel downward at ever increasing speeds. Since the jumpers start from rest and fall downward for 5 s, we can use Eq. 1.2.3 to determine how far they will fall:

$$\begin{aligned}\text{final height} &= \text{initial height} - \frac{1}{2} \cdot 9.8 \text{ m/s}^2 \cdot (5 \text{ s})^2 \\ &= \text{initial height} - 122.5 \text{ m}.\end{aligned}$$

The downward acceleration is indicated here by the negative change in height. At the end of 5 s, the jumpers will have fallen more than 122 m (402 ft) and will be traveling downward at about 50 m/s (160 ft/s). The tower will need additional height to slow the jumpers down and begin bouncing them back upward. Clearly, a 5-s free fall is pretty unrealistic. Try for a 2- or 3-s free fall instead.

Tossing the Ball Upward

If the only force acting on an object is its weight, then the object is falling. So far, we've explored this principle only as it pertains to balls dropped from rest. However, a thrown ball is falling, too; once it leaves your hand, it's subject only to its weight and it accelerates downward at 9.8 m/s^2 (32 ft/s^2).

Equation 1.2.2 still describes how the ball's velocity depends on the fall time, but now the initial velocity isn't zero. If you toss the ball straight up in the air, it leaves your hand with a large upward velocity (Fig. 1.2.4). As soon as you let go of the ball, it begins to accelerate downward. If the ball's initial upward velocity is 29.4 m/s (96 ft/s), then after 1 s its upward velocity is 19.6 m/s (64 ft/s). After another second, its upward velocity is only 9.8 m/s (32 ft/s). After a third second, the ball momentarily comes to a complete stop with a velocity of zero. It then descends from this peak height, falling just as it did when you dropped it from rest.

The ball's flight before and after its peak is symmetrical. It travels upward quickly at first, since it has a large upward velocity. As its upward velocity diminishes, it travels more and more slowly until it comes to a stop. It then begins to descend, slowly at first and then faster and faster as it continues its constant downward acceleration. The time the ball takes to rise from its initial position in your hand to its peak height is exactly equal to the time it takes to descend back down from that peak to your hand. Equation 1.2.3 indicates how the position of the ball depends on the fall time, with the initial velocity being the upward velocity of the ball as it leaves your hand.

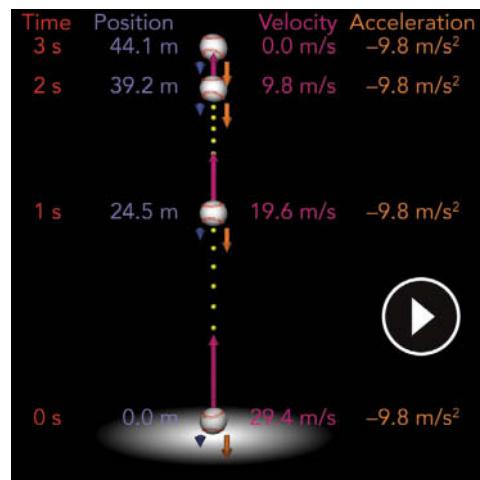


Fig. 1.2.4 The moment you let go of a ball thrown straight upward, it begins to accelerate downward at 9.8 m/s^2 . The ball rises but its upward velocity diminishes steadily until it momentarily comes to a stop. It then descends with its downward velocity increasing steadily. In this example, the ball rises for 3 s and comes to rest. It then descends for 3 s before returning to your hand in a very symmetrical flight.

The larger the initial upward velocity of the ball, the longer it rises and the higher it goes before its velocity is reduced to zero. It then descends for the same amount of time it spent rising. The higher the ball goes before it begins to descend, the longer it takes to return to the ground and the faster it's traveling when it arrives. That's why catching a high fly ball with your bare hands stings so much: the ball is traveling very, very fast when it hits your hands, and a large force is required to bring the ball to rest quickly.

Check Your Understanding #4: A Toss-Up

You toss a coin straight up, and it rises well above your head. At the moment the coin reaches its peak height, what is its velocity? Is that velocity constant or changing with time? Is the coin's acceleration constant or changing with time?

Answer: The coin's velocity is momentarily zero at its peak, but that velocity is changing with time. The coin's acceleration, however, is constant—the acceleration due to gravity.

Why: Once the coin leaves your hand, it's a falling object and constantly accelerates downward at the acceleration due to gravity. Because it begins its fall traveling upward, it rises at a gradually decreasing speed, is momentarily motionless at its peak, and then descends at a gradually increasing speed.

How a Thrown Ball Moves: Projectile Motion

What happens if you don't toss the ball exactly straight up? Suppose you throw the ball upward at some angle. The ball still rises to a peak height and then descends, but it also travels away from you horizontally so that it strikes the ground at some distance from your feet. How much does this horizontal travel complicate the motion of a falling ball?

The answer is not very much. One of the beautiful simplifications of physics is that you can often treat an object's vertical motion independently of its horizontal motion. This technique involves separating the vector quantities—acceleration, velocity, and position—into **components**, those portions of the quantities that lie along specific directions (Fig. 1.2.5). For example, the upward component of an object's position is that object's altitude.

A ball's altitude, however, is only part of its position; you still need to know where it is in relation to your right or left and to your front or back. In fact, you can specify its position (or any other vector quantity) in terms of three components along three directions that are perpendicular to one another. This means that you can completely specify the ball's position by its distance to your right, its distance in front of you, and its altitude above you. If it's to your left, behind you, or below you, the corresponding components have negative values. For example, the ball might be 3 m to your right, -3 m in front of you (which is 3 m behind you), and 2 m above you.

When you drop the ball from rest or toss it straight up, its motion is entirely vertical and only the upward components of its position, velocity, and acceleration are important. When the falling ball is also moving horizontally, however, you'll need to pay attention to the rightward and forward components of those vector quantities. Throwing the ball forward helps because it eliminates any rightward components of the motion. The ball then arcs forward with a position specified at each moment by its altitude (its upward component) and its distance *downfield* (its forward component).

Although Eqs. 1.2.2 and 1.2.3 were introduced to describe a ball's vertical motion, they also describe its horizontal motion. More generally, they relate three vector quantities—the ball's position, velocity, and constant acceleration—to one another and describe how the ball's position and velocity change with time when that ball undergoes constant acceleration in any direction. Of course, since we're concerned now with falling balls, the relevant constant acceleration is the downward acceleration due to gravity.

These equations also apply to the *components* of the ball's position, velocity, and constant acceleration. If you add the words *upward component of* in front of each vector quantity

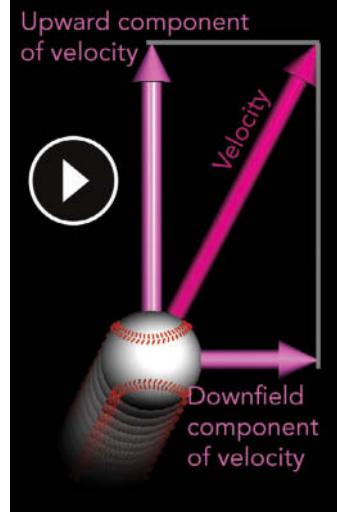


Fig. 1.2.5 Even if the ball has a velocity that is neither purely vertical nor purely horizontal, its velocity may nonetheless be viewed as having an upward component and a downfield component. Part of its total velocity acts to move this ball upward, and part of its total velocity acts to move this ball downfield.

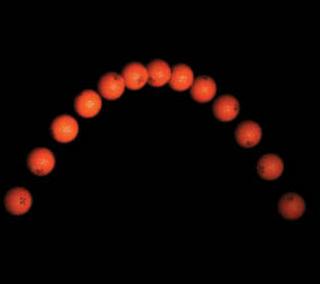


Fig. 1.2.6 This golf ball drifts steadily to the right after being thrown because gravity affects only the ball's upward component of velocity.

in Eqs. 1.2.2 and 1.2.3, the equations correctly describe the ball's vertical motion. Now it's time to add the words *forward component of* to those equations so that they describe the ball's horizontal motion.

Once you've thrown the ball forward and are no longer touching it, its motion can be broken into two parts: its upward motion and its forward motion (Fig. 1.2.5). Part of the ball's initial velocity is in the upward direction, and that upward component of velocity determines the object's ascent and descent. Part of the ball's initial velocity is in the forward direction, and that forward component of velocity determines the ball's progress downfield.

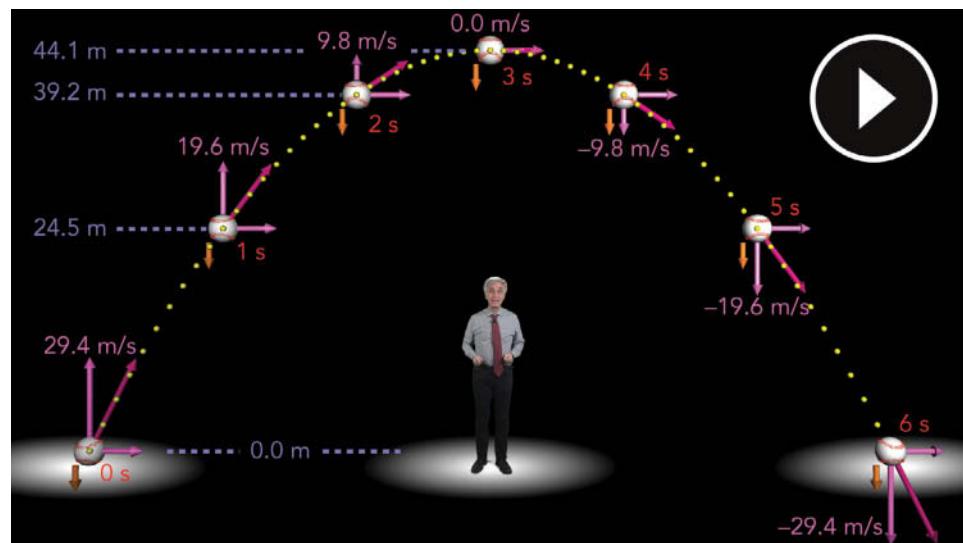
Because the ball is accelerating straight downward, the forward component of its acceleration is zero and the forward component of its velocity thus remains constant. The ball travels downfield at a steady rate throughout its flight (Fig. 1.2.6). Overall, the upward component of the ball's initial velocity determines how high the ball goes and how long it stays aloft before striking the ground, while the forward component of the initial velocity determines how quickly the ball travels downfield during its time aloft (Fig. 1.2.7).

Just before the ball hits the ground it still has its original forward component of velocity, but its upward component of velocity is now negative—the ball is moving downward. The total velocity of the ball is composed of these two components. The ball starts with its velocity directed up and forward, and finishes with its velocity directed down and forward.

If you want a ball or shot put to hit the ground as far from your feet as possible, you should keep it aloft for a long time *and* give it a sizable forward component of velocity; in other words, you must achieve a good balance between the upward and forward components of velocity (Fig. 1.2.8). These components of velocity together determine the ball's flight path, its **trajectory**. If you throw the ball straight up, it will stay aloft for a long time but will not travel downfield at all (and you will need to wear a helmet). If you throw the ball directly forward, it will travel downfield quickly but hit the ground almost immediately.

Neglecting air resistance and the altitude difference between your throwing arm and the ground that the ball will eventually hit, your best choice is to throw the ball forward at an angle of 45° above horizontal. At that angle, the initial upward component of velocity will be the same as the initial forward component of velocity. The ball will stay aloft for a reasonably long time and will make good use of that time to move downfield. Other angles won't make such good use of the initial speed to move the ball downfield. (A discussion of how to determine the upward and forward components of velocity appears in Appendix A, online.)

Fig. 1.2.7 If you throw a ball upward, at an angle, part of the initial velocity will be in the upward direction and part will be in the downfield direction. The vertical and horizontal motions will take place independently of one another. The ball will rise and fall just as it did in Figs. 1.2.3 and 1.2.4; at the same time, however, it will move downfield. Because there is no horizontal component of acceleration (gravity acts only in the vertical direction), the downfield component of velocity remains constant during the ball's 6-s trip.



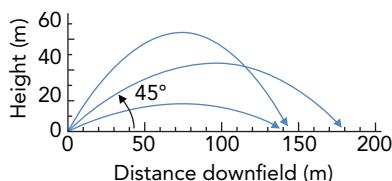
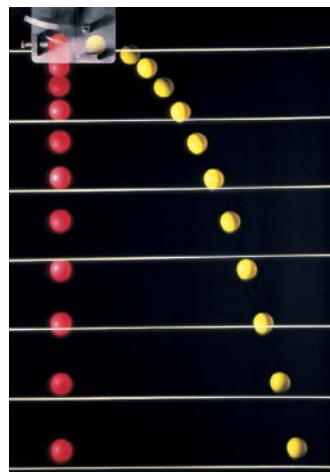


Fig. 1.2.8 If you want the ball to hit the ground as far from your feet as possible, given a certain initial speed, throw the ball at 45° above horizontal. Halfway between horizontal and vertical, such a throw gives the ball equal initial upward and downfield components of velocity. The ball then stays aloft for a relatively long time and makes good use of that flight time to travel downfield.

© Richard Megna/Fundamental Photographs



These same ideas apply to two baseballs, one dropped from a cliff and the other thrown directly forward from that same cliff. If both leave your hands at the same time, they will both hit the ground below at the same time (Fig. 1.2.9). The fact that the second ball has an initial forward velocity doesn't affect the time it takes to descend to the ground because the horizontal and vertical motions are independent. Of course, the ball thrown directly forward will strike the ground far from the base of the cliff, while the dropped ball will land directly below your hand.

Check Your Understanding #5: Aim High

Why must a sharpshooter or an archer aim somewhat above her target? Why can't she simply aim directly at the bull's-eye to hit it?

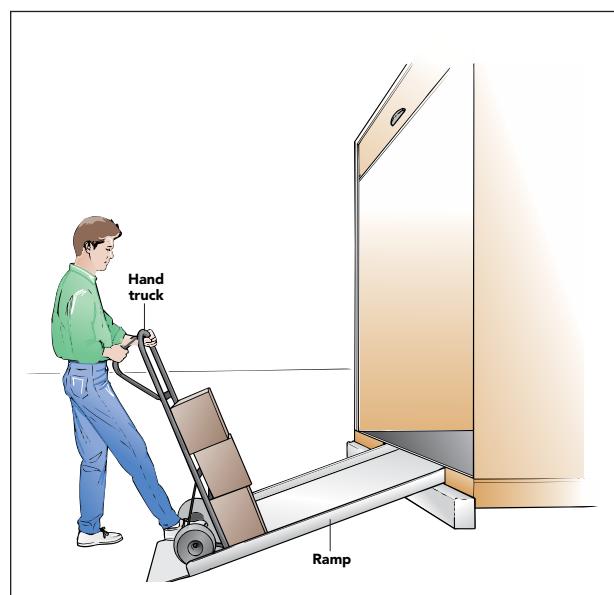
Answer: The bullet or arrow will fall in flight, so she must compensate for its loss of height.

Why: To hit the bull's-eye, the sharpshooter or archer must aim above the bull's-eye because the projectile will fall in flight. Even if the target is higher or lower than the sharpshooter, the projectile will fall below the point at which she is aiming. The longer the bullet or arrow is in flight, the more it will fall and the higher she must aim. As the distance to the target increases, the flight time increases and her aim must move upward.

Fig. 1.2.9 When these two balls were dropped, they accelerated downward at the constant rate of 9.8 m/s^2 and their velocities increased steadily in the downward direction. Even though one of them was initially moving to the right, they descended together.

SECTION 1.3

Ramps



In the previous section, we looked at what happens to an object experiencing only a single force, its weight. But what happens to objects that experience two or more forces at the same time? Imagine, for example, an object resting on the floor. That object experiences both its downward weight and an upward force from the floor. If the floor is level, the object doesn't accelerate; but if the floor is tilted, so that it forms a ramp, the object accelerates downhill. In this section, we'll examine the motion of objects traveling along ramps. In addition to making sports such as skateboarding and skiing more fun, ramps are common tools that help us lift and move heavy objects.

Questions to Think About: How does a ramp make it possible for one person to lift a very heavy object? Why is lowering a heavy object so much easier than raising it? What changes about that object as you raise it? Why is a steep hill so much scarier to ski or sled down than a more gradual slope? Why is the steeper hill so much harder to bicycle up?

Experiments to Do: Place an unbreakable water bottle on a level table or board. Hold the bottle steady for a second and then let go of it without pushing on it. Why does the bottle remain motionless?

Now, equip yourself with a pencil. Have a friend tilt the surface of the table or board slightly so that the bottle begins to roll downhill. Can you stop the bottle by pushing on it with the pencil? Place the bottle back on the table and have your friend tilt the table more sharply. Does the tilt of the table

affect your ability to stop the bottle from rolling downhill? Now try to push the bottle uphill with the pencil (1) when the table is slightly tilted and (2) when it is more sharply tilted. Which task requires more force? Why?

Evidently a gentle push is all that may be needed to raise a relatively heavy object if you use a ramp to help you. To understand why this feat is possible, we need to explore a handful of physical concepts and a few basic laws of motion.

A Piano on the Sidewalk

Imagine that you have a friend who's a talented but undiscovered pianist. She's renting a new apartment, and because she can't afford professional movers (Fig. 1.3.1), she's asked you to help her move her baby grand. Fortunately, her new apartment is only on the second floor. The two of you still face a difficult challenge, however: how do you get that heavy piano up there? More important, how do you keep it from falling on you during the move?

The problem is that you can't push upward hard enough to lift the entire piano at once. One solution to this problem, of course, would be to break the piano into pieces and carry them up one by one. This method has obvious drawbacks—your friend isn't expecting a firewood delivery. A better solution would be to find something else to help you push upward, and one of your best choices is the simple machine known as a ramp.

Throughout the ages, **ramps**, also known as inclined planes, have made tasks like piano-moving possible. Because ramps can exert the enormous upward forces needed to lift stone and steel, they've been essential building equipment since the days of the pyramids. To see how ramps provide these lifting forces, we'll continue to explore the example of the piano, looking first at the force that the piano experiences when it touches a surface. For the time being, we'll continue to ignore friction and forces due to the air; they will needlessly complicate our discussion. Besides, as long as the piano is on wheels, friction is negligible.

With the piano resting on the sidewalk outside the apartment, you make a startling discovery: the piano is *not* falling. Has gravity disappeared? The answer to that question would be painfully obvious if your foot were underneath one of the piano's wheels. No, the piano's weight is still all there. But something is happening at the surface of the sidewalk to keep the piano from falling. Let's take a careful look at the situation.

To begin with, the piano is clearly pushing down hard on the sidewalk. That's why you're keeping your toes out of the way. But the presence of a new downward force *on the sidewalk* doesn't explain why *the piano* isn't falling. Instead, we must look at the sidewalk's response to the piano's downward push: the sidewalk pushes upward on the piano! You can feel this response by leaning over and pushing down on the sidewalk with your hand—the sidewalk will push back. Those two forces, your downward push on the sidewalk and its upward push on your hand, are exactly equal in magnitude but opposite in direction.

This observation—that two things exert equal but opposite forces on one another, isn't unique to sidewalks, pianos, or hands; in fact, it's always true. If you push on any object, that object will push back on you with an equal amount of force in exactly the opposite direction. This rule—often expressed as “for every action, there is an equal but opposite reaction”—is known as **Newton's third law of motion**, the last of his three laws.

© SSPL via/Getty Images, Inc.



Fig. 1.3.1 A ramp would make this move much easier.

NEWTON'S THIRD LAW OF MOTION

For every force that one object exerts on a second object, there is an equal but oppositely directed force that the second object exerts on the first object.

The universality of this law is astounding. Whether an object is large or small, hard or soft, stationary or faster than a rocket, if you can push on it, it *will* push back on you with an equal but oppositely directed force.

In the present case, the sidewalk and piano push on one another with equal but oppositely directed forces. Of this pair of equal-but-opposite forces, only one force acts *on the piano*: the sidewalk's upward push. This upward push on the piano is what keeps the piano from falling. We've solved the mystery.

INTUITION ALERT: Action and Reaction

Intuition says that, when you push an object that's moving away from you, it pushes back more gently on you than you push on it and that, when you push an object that's moving toward you, it pushes back harder on you than you push on it.

Physics says that when you push on an object, it always pushes back on you exactly as hard as you push on it.

Resolution: It's difficult to push on an object that's moving away from you, so you naturally push on it more gently than you expect. The gentle force it exerts on you is simply an equal but oppositely directed response to your gentle force on it. In contrast, it's difficult not to push strongly on an object that's moving toward you, so you naturally push on it harder than you expect. The strong force it exerts on you is again an equal but oppositely direct response to your strong force on it.

SUMMARY OF NEWTON'S LAWS OF MOTION

1. An object that is not subject to any outside forces moves at a constant velocity, covering equal distances in equal times along a straight-line path.
2. The net force exerted on an object is equal to that object's mass times its acceleration. The acceleration is in the same direction as the force.
3. For every force that one object exerts on a second object, there is an equal but oppositely directed force that the second object exerts on the first object.

Check Your Understanding #1: Swinging

You are pushing a child on a playground swing. If you exert a 50-N (11-lbf) force on him as he is swinging away from you, how much force will he exert back on you?

Answer: He will exert 50 N (11 lbf) on you.

Why: Whenever you exert a 50-N force on an object, whether it's moving or stationary, it will exert a 50-N force back on you. There are no exceptions. If that object is a friend, it doesn't matter whether she is stationary or moving or wearing roller skates or even sound asleep; she will push back with 50 N of force. She has no choice in the matter. Similarly, if someone pushes on you, you will feel yourself pushing back. That's how Newton's third law works.

Looking for Support and Adding Up the Forces

Although we've figured out why the piano isn't falling, we still don't know what type of force the sidewalk is using to hold it up or why that upward force so perfectly balances the piano's downward weight.

Let's begin with the type of force. Since two objects can't occupy the same space at the same time, their surfaces push apart whenever they're in contact. They exert **support forces**

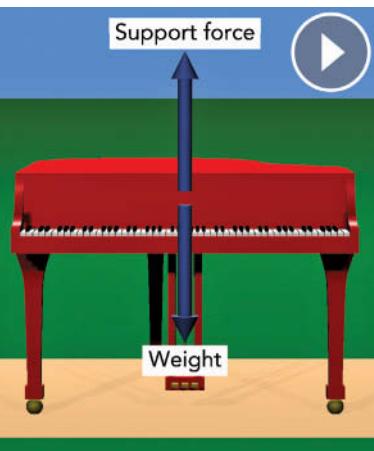


Fig. 1.3.2 A piano resting on the sidewalk. The sidewalk exerts an upward support force that exactly balances the piano's downward weight. The net force on the piano is zero, so the piano doesn't accelerate.

5 Forces that are directed exactly away from surfaces are called **normal forces**, since the term **normal** is used by mathematicians to describe something that points exactly away from a surface—at a right angle or perpendicular to that surface.

on one another, each pushing the other directly away from its surface—the direction *normal* or *perpendicular* to its surface (see 5). Since the sidewalk is horizontal, the support force it exerts on the piano is vertical—straight up (Fig. 1.3.2).

How large is that support force? To answer this question, suppose the sidewalk's support force were strong enough to make the piano accelerate upward. The piano would soon lift off the sidewalk, and as their surfaces stopped touching, the sidewalk's support force on the piano would grow weaker. Alternatively, suppose the sidewalk's support force was weak enough to let the piano accelerate downward. The piano would soon drop into the sidewalk, and as their surfaces overlapped more, the sidewalk's support force on the piano would grow stronger.

Because of these behaviors, the sidewalk's upward support force on the piano adjusts automatically until it exactly balances the piano's downward weight and the piano accelerates neither up nor down. When you sit on the piano during a break, the sidewalk's upward support force quickly readjusts to balance your weight as well.

Another way to state that the upward support force on the piano exactly balances the piano's downward weight is to say that the net force on the piano is zero, meaning that the sum of all the forces on the piano is zero. Objects often experience more than one force at a time, and it's the net force, together with the object's mass, that determines how it accelerates. When you and your friend push the piano in the same direction, your forces add up, assisting one another so that the piano accelerates in that direction (Fig. 1.3.3a). When the two of you push the piano in opposite directions, your forces oppose and at least partially cancel one another (Fig. 1.3.3b).

When the two of you push the piano at an angle with respect to one another, the net force points somewhere in between. For example, if you push the piano eastward while your friend pushes it northward, the net force will point to the northeast and the piano will accelerate in that direction (Fig. 1.3.3c). The precise angle of the net force and the piano's subsequent acceleration depend on exactly how hard each person pushes. For most of the following discussion, we'll need only a rough estimate of the net force's magnitude and direction, and we'll obtain that estimate using common sense.

Apart from direction, there's one crucial difference between the force that gravity exerts on the piano and the support force that the sidewalk exerts on it. While the piano's weight is dispersed throughout the piano, the sidewalk's upward support force acts only on the piano's wheels. Even when the net force on the piano is zero, having individual forces act on it at different locations can lead to considerable stress within the piano. If it weren't built so well, the piano might lose a leg or two during the move.

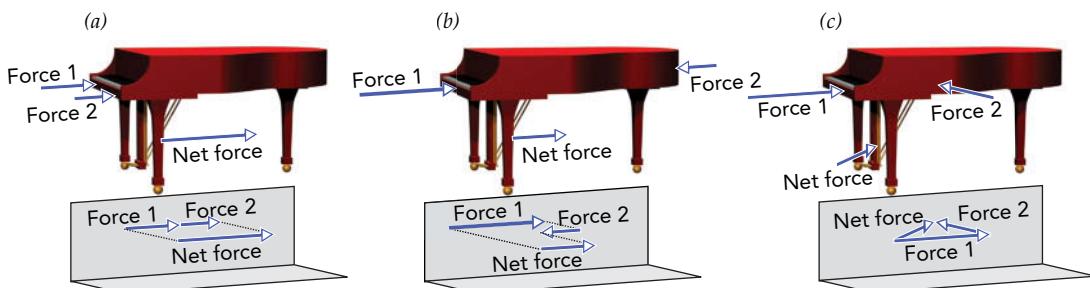


Fig. 1.3.3 When several forces act simultaneously on an object, the object responds to the sum of the forces. This sum is called the net force, and it has both a magnitude and a direction. Here, as elsewhere in this book, the length of each force arrow indicates its magnitude. You can sum the forces graphically by arranging the force arrows in sequence, head to tail. The net force arrow then points from the tail of the first arrow to the head of the last arrow. Some of the force arrows are shown displaced slightly for clarity, a shift that doesn't affect the summing process.

COMMON MISCONCEPTIONS: Newton's Third Law and Balanced Forces

Misconception: When you push on an object and it pushes back on you, those two equal-but-opposite forces somehow cancel one another perfectly and therefore have no effect on either you or the object.

Resolution: The two forces described by Newton's third law always act on *two different things*. Your push acts on the object, while the object's push acts on you. Since the object accelerates in response to the net force it experiences—the sum of all the individual forces acting *on it*—it is affected only by your force *on it*, not by its force *on you*. If you are the only thing pushing on it, it will accelerate. If the object is the only thing pushing on you, you'll accelerate, too!

Check Your Understanding #2: Riding the Elevator

As you ride upward in an elevator at a constant velocity, what two forces act on your body and what is the net force on you?

Answer: The two forces are your downward weight and an upward support force from the floor. They balance, so that the net force on you is zero.

Why: Whenever anything is moving with constant velocity, it's not accelerating and thus has zero net force on it. Although the elevator is moving upward, the fact that you are not accelerating means that the car must exert an upward support force on you that exactly balances your weight. You experience zero net force.

What the Piano Needs: Energy

As you approach the task of lifting your friend's piano into her apartment, you might begin to worry about safety. There is clearly a difference between the piano resting on the sidewalk and the piano suspended on a board just outside the second-floor apartment. After all, which one would you rather be sitting beneath? The elevated piano has something that the piano on the sidewalk doesn't have: the ability to do things—to break the board, to produce motion, and to squash whatever is beneath it. This capacity to make things happen is called *energy*, and the process of making them happen is called *work*.

Energy and work are both important physical *quantities*, meaning that both are measurable. For example, you can measure the amount of energy in the suspended piano and the amount of work the piano does when the board breaks and it falls to the sidewalk. As you may suspect, the physical definitions of *energy* and *work* are somewhat different from those of common English. Physical **energy** isn't the exuberance of a child at the amusement park or what's contained in a so-called "energy drink"; instead, it's defined as the capacity to do work. Similarly, physical **work** doesn't refer to activities at the office or in the yard; instead, it refers to the process of transferring energy.

Energy is what's transferred, and work does the transferring. The most important characteristic of energy is that it's conserved. In physics, a **conserved quantity** is one that can't be created or destroyed but that can be transferred between objects or, in the case of energy, be converted from one form to another. Conserved quantities are very special in physics; there are only a few of them. An object that has energy can't simply make that energy disappear; it can only get rid of energy by giving it to another object, and it makes this transfer by doing work on that object.

The relationship between energy and work is analogous to the relationship between money and spending: money is what is transferred, and spending does the transferring. Sensible, law-abiding citizens don't create or destroy money; instead, they transfer it among themselves through spending. Just as the most interesting aspect of money is spending it, so the most interesting aspect of energy is doing work with it. We can define money as the capacity to spend, just as we define energy as the capacity to do work.

So far we've been using a circular definition: work is the transfer of energy, and energy is the capacity to do work. But what is involved in doing work on an object? You do work on an object by exerting a force on it as it moves in the direction of that force. As you throw a ball, exerting a forward force on the ball as the ball moves forward, you do work on the ball; as you lift a rock, pushing the rock upward as it moves upward, you do work on the rock. In both cases, you transfer energy from yourself to an object by doing work on it.

This transferred energy is often apparent in the object. When you throw a ball, it picks up speed and undergoes an increase in **kinetic energy**, energy of motion that allows the ball to do work on whatever it hits. When you lift a rock, it shifts farther from Earth and undergoes an increase in **gravitational potential energy**, energy stored in the gravitational forces between the rock and Earth that allows the rock to do work on whatever it falls on. In general, **potential energy** is energy stored in the forces between or within objects.

Returning to the task at hand, it's now apparent that raising the piano to the second-floor apartment is going to increase the piano's gravitational potential energy by a substantial amount. Since energy is a conserved quantity, this additional energy must come from something else. Unfortunately, that something is you! To deliver the piano, you are going to have to provide it with the gravitational potential energy it needs by doing exactly that amount of work on it. As we'll see, you can do that work the hard way by carrying it up a ladder or the easy way by pushing it up a ramp.

Check Your Understanding #3: No Shortage of Energy

Do any of these objects have energy they can spare: a compressed spring, an inflated toy balloon, a stick of dynamite, and a falling ball?

Answer: Yes, they all do.

Why: Each of these four objects can easily do work on you and thereby give you some of its spare energy. It does this work by pushing on you as you move in the direction of that push.

Lifting the Piano: Doing Work

To do work on an object, you must push on it while it moves in the direction of your push. The work you do on it is the force you exert on it times the distance it travels along the direction of your force. We can express this relationship as a word equation:

$$\text{work} = \text{force} \cdot \text{distance}, \quad (1.3.1)$$

in symbols:

$$W = \mathbf{F} \cdot \mathbf{d},$$

and in everyday language:

If you're not pushing or it's not moving, then you're not working.

This simple relationship assumes that your force is constant while you're doing the work. If your force varies, the calculation of work will have to recognize that variation and may require the use of calculus.

Calculating the work you do is easy if the object moves exactly in the direction of your constant push; you simply multiply your force times the distance the object travels. However, if the object doesn't move in the direction of your push, you must multiply your

force times the *component* of the object's motion that lies along the direction of your force.

As long as the angle between your force and the object's motion is small, you can often ignore this complication. However, as the angle becomes larger, the work you do on the object decreases. When the object moves at right angles to your force, the work you do on it drops to zero—it's not moving along the direction of your force at all. And for angles larger than 90° , the object moves *opposite* your force and the work you do on it actually becomes negative!

Recalling that forces always come in equal but oppositely directed pairs, we can now explain why energy is conserved: whenever you do work on an object, that object simultaneously does an equal amount of negative work on you! After all, if you push an object and it moves along the direction of your force, then it pushes back on you and you move along the direction opposite its force. You do positive work on it, and it does negative work on you.

For example, when you lift the piano to judge its weight, you push it up as it moves up and thus do work on it. At the same time, the piano pushes your hand down but your hand moves up, so it does negative work on your hand. Overall, the piano's energy increases by exactly the same amount that your energy decreases—a perfect transfer! The energy that you're losing is mostly food energy, a form of chemical potential energy, and the energy the piano is gaining is mostly gravitational potential energy.

When you lower the piano after realizing that it's too heavy to carry up a ladder, the process is reversed and the piano transfers energy back to you. Now the piano does work on you, and you do an equal amount of negative work on the piano. The piano is losing mostly gravitational potential energy, and you are gaining mostly thermal energy—a disordered form of energy that we'll examine in Section 2.2. Unlike a rubber band, your body just isn't good at storing work done on it, so it simply gets hotter. Nonetheless, it's usually easier to have work done on you than to do work on something else. That's why it's easier to lower objects than to lift them.

Finally, when you hold the piano motionless above the pavement, while waiting for your friend to reinstall the wheel that fell off, you and the piano do no work on one another. You are simply converting chemical potential energy from your last meal into thermal energy in your muscles and getting overheated, probably in more ways than one.



CONSERVED QUANTITY: ENERGY

TRANSFERRED BY: WORK

Energy: The capacity to do work. Energy has no direction. It can be hidden as potential energy.

Kinetic energy: The form of energy contained in an object's motion.

Potential energy: The form of energy stored in the forces between or within objects.

Work: The mechanical means for transferring energy; $\text{work} = \text{force} \cdot \text{distance}$.



Check Your Understanding #4: Pitching

When you throw a baseball horizontally, you're not pushing against gravity. Are you doing any work on the baseball?

Answer: Yes.

Why: Any time you exert a force on an object and the object moves in the direction of that force, you are doing work on the object. Since gravity doesn't affect horizontal motion, the work you do on the baseball as you throw it ends up in the baseball as kinetic energy (energy of motion). As anyone who has been hit by a pitch can attest, a moving baseball has more energy than a stationary baseball.

 Check Your Figures #1: Light Work, Heavy Work

You are moving books to a new shelf, 1.20 m (3.94 ft) above the old shelf. The books weigh 10.0 N (2.25 lbf) each, and you have 10 of them to move. How much work must you do on them as you move them? Does it matter how many you move at once?

Answer: It takes $120 \text{ N} \cdot \text{m}$ (88.6 ft · lbf), no matter how many you lift at once.

Why: To keep each book from accelerating downward, you must support its weight with an upward force of 10.0 N. You must then move it upward 1.20 m. The work you do pushing upward on the book as it moves upward is given by Eq. 1.3.1:

$$\text{work} = \text{force} \cdot \text{distance} = 10.0 \text{ N} \cdot 1.20 \text{ m} = 12.0 \text{ N} \cdot \text{m}.$$

It takes $12.0 \text{ N} \cdot \text{m}$ of work to lift each book, whether you lift it together with other books or all by itself. The total work you must do on all 10 books is $120 \text{ N} \cdot \text{m}$.

Gravitational Potential Energy

How much work would you do on the piano while lifting it straight up a ladder to the apartment? Apart from a little extra shove to get the piano moving upward, lifting it would entail supporting the piano's weight while it coasted upward at constant velocity from the sidewalk to the second floor. Since you would be pushing upward on the piano with a force equal in amount to its weight, the work you would do on it would be its weight times the distance you lifted it.

As the piano rises in this scenario, its gravitational potential energy increases by an amount equal to the work you do on it. If we agree that the piano has zero gravitational potential energy when it rests on the sidewalk, then the suspended piano's gravitational potential energy is simply its weight times its height above the sidewalk. Since the piano's weight is equal to its mass times the acceleration due to gravity, its gravitational potential energy is its mass times the acceleration due to gravity times its height above the sidewalk.

These ideas aren't limited to pianos. You can determine the gravitational potential energy of any object by multiplying its mass times the acceleration due to gravity times its height above the level at which its gravitational potential energy is zero. This relationship can be expressed as a word equation:

$$\text{gravitational potential energy} = \text{mass} \cdot \text{acceleration due to gravity} \cdot \text{height}, \quad (1.3.2)$$

in symbols:

$$U = m \cdot g \cdot h,$$

and in common language:

The higher it was, the harder it hit.

Of course, if you know the object's weight, you can use it in place of the object's mass times the acceleration due to gravity.

So what is the piano's gravitational potential energy when it reaches the second floor? If it weighs 2000 N (450 lbf) and the second floor is 5 m (16 ft) above the sidewalk, you will have done $10,000 \text{ N} \cdot \text{m}$ (about 7200 ft · lbf) of work in lifting it up there, and the piano's gravitational potential energy will thus be $10,000 \text{ N} \cdot \text{m}$. The **newton-meter** is the SI unit of energy and work; it's so important that it has its own name, the **joule** (abbreviated J). At the second floor, the piano's gravitational potential energy is 10,000 J.

A few everyday examples should give you a feeling for how much energy a joule is. Lifting a liter bottle of water 10 centimeters (4 inches) upward requires about 1 J of work.

A 1500-watt hairdryer needs 1500 J every second to operate. Your body is able to extract about 2,000,000 J from a slice of cherry pie. When you're bicycling or rowing hard, your body can do about 1000 J of work each second. A typical flashlight battery has about 10,000 J of stored energy.

Quantity	SI Unit	English Unit	SI → English	English → SI
Energy	joule (J) = newton-meter (N · m)	foot-pound (ft · lbf)	$1 \text{ J} = 0.73757 \text{ ft} \cdot \text{lbf}$	$1 \text{ ft-lbf} = 1.3558 \text{ J}$

Check Your Understanding #5: Mountain Biking

Bicycling to the top of a mountain is much harder than rolling back down to the bottom. At which place do you have the most gravitational potential energy?

Answer: You have the most at the top of the mountain.

Why: Bicycling up the mountain is hard because you must do work against the force of gravity. This work is stored as an increasing gravitational potential energy on your way uphill. Gravity then does work on you as you roll downhill and your gravitational potential energy decreases.

Check Your Figures #2: Watch Out Below

If you carry a U.S. penny (0.0025 kg) to the top of the Empire State Building (443.2 m, or 1453.7 ft), how much gravitational potential energy will it have?

Answer: About 11 J.

Why: The penny's gravitational potential energy is given by Eq. 1.3.2:

$$\begin{aligned}\text{gravitational potential energy} &= 0.0025 \text{ kg} \cdot 9.8 \text{ N/kg} \cdot 443.2 \text{ m} \\ &= 11 \text{ N} \cdot \text{m} = 11 \text{ J}.\end{aligned}$$

This 11-J increase in energy would be quite evident if you were to drop the penny. In principle, the penny could accelerate to very high speed (up to 340 km/h or 210 mph) and do lots of damage when it hit the ground. Fortunately, a falling penny actually tumbles through the air, which slows it to about 40 km/h or 25 mph.

Lifting the Piano with a Ramp

Unfortunately, you probably can't carry a grand piano up a ladder by yourself. You need a ramp, both to help you support the piano and to make it easier for you to raise the piano to the second floor.

Like the sidewalk, a ramp exerts a support force on the piano to prevent the piano from passing through its surface. However, since the ramp isn't exactly horizontal, that support force isn't exactly vertical (Fig. 1.3.4). The piano's weight still points straight down, but since the ramp's support force doesn't point straight up, the two forces can't balance one another. There is a nonzero net force on the piano.

This net force can't point into or out of the ramp. If it did, the piano would accelerate into or out of the ramp and the two objects would soon either lose contact or travel through one another. Instead, the net force points exactly along the surface of the ramp—a direction *tangent* or *parallel* to the surface. More specifically, it points directly downhill, so the piano accelerates down the ramp!

However, because this net force is much smaller than the piano's weight, the piano's acceleration down the ramp is slower than if it were falling freely. This effect is familiar to anyone who has bicycled downhill or watched a cup slip slowly off a tilted table. Although gravity is still responsible, these objects accelerate more slowly than they would if falling and in the direction of the downward slope.



Fig. 1.3.4 When a piano is on a ramp, its weight and the ramp's support force don't cancel. Instead, they sum to a downhill ramp force. If no other forces act on the piano, the ramp force is its net force and it accelerates downhill.

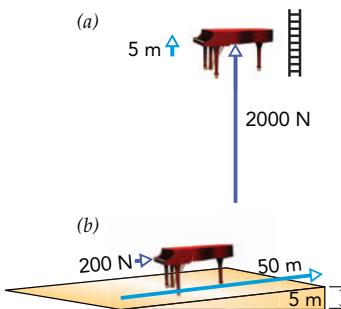


Fig. 1.3.5 To lift a piano weighing 2000 N, you can either (a) push it straight up or (b) push it along a ramp. To keep the piano moving at a constant velocity, you must make sure it experiences a net force of zero. If you lift it straight up the ladder in (a), you must exert an upward lifting force of 2000 N to balance the piano's downward weight. If you push it up the ramp shown in (b), you will only have to push the piano uphill with a force of 200 N to give the piano a net force of zero.

Therein lies the beauty of the ramp. By putting the piano on a ramp, you let the ramp support most of the piano's weight. The piano experiences only a small residual force, which I call a ramp force, pushing it downhill along the ramp. If you now push uphill on the piano with a force that exactly balances that downhill ramp force, the net force on the piano drops to zero and the piano stops accelerating. If you push uphill a little harder, the piano will accelerate up the ramp!

How does the ramp change the job of moving the piano? Suppose that you build a ramp 50 m (164 ft) long that extends from the sidewalk to the apartment's balcony, 5 m (16.4 ft) above the pavement. This ramp is sloped so that traveling 50 m uphill along its surface lifts the piano only 5 m upward (Fig. 1.3.5). Because of the ramp's 10 to 1 ratio between distance traveled along its surface and altitude change, you can push the 2000-N (450-lbf) piano up it at constant velocity with a force of only 200 N (45 lbf). Most people can push that hard, so the moving job is now realistic. To reach the apartment, you must push the piano 50 m along this ramp with a force of 200 N so that you will do a total of 10,000 J of work.

By pushing the piano up the ramp, you've used physical principles to help you perform a task that would otherwise have been nearly impossible. You didn't get something for nothing, however. The ramp is much longer than the ladder, and you have had to push the piano for a longer distance to raise it to the second floor. Of course, you have had to push with less force.

Remarkably, the amount of work you do in either case is 10,000 J. When you carry the piano up the ladder, the force you exert is large but the distance the piano travels in the direction of that force is small. When you push the piano up the ramp, the force is small but the distance is large. Either way, the final result is the same: the piano ends up on the second floor with an additional 10,000 J of gravitational potential energy and you have done 10,000 J of work. Expressed graphically in an equation, this relationship appears as:

$$\text{work} = \text{large force} \cdot \text{small distance} = \text{small force} \cdot \text{large distance}.$$

In the absence of friction, the amount of work you do on the piano to get it to the second floor doesn't depend on how you raise it. No matter how you move that piano up to the second floor, its gravitational potential energy will increase by 10,000 J, so you'll have to do 10,000 J of work on it. Even if you disassemble the piano into parts, carry them individually up the stairs, and reconstruct the piano in your friend's living room, you will have done 10,000 J of work lifting the piano.

Unless you're an experienced piano tuner, you'll probably be better off sticking with the ramp. It offers an easy method for one person to lift a baby grand piano. The ramp provides **mechanical advantage**, the process whereby a mechanical device redistributes the amounts of force and distance that go into performing a specific amount of mechanical work. In moving the piano with the help of the ramp, you've performed a task that would normally require a large force over a small distance by supplying a small force over a large distance. You might wonder whether the ramp itself does any work on the piano; it doesn't. Although the ramp exerts a support force on the piano and the piano moves along the ramp's surface, this force and the distance traveled are at right angles to one another. The ramp does no work on the piano.

Mechanical advantage occurs in many situations involving ramps. For example, it appears when you ride a bicycle up a hill. Climbing a gradual hill takes far less uphill force than climbing a steep hill of the same height. Since your pedaling ultimately provides the uphill force, it's much easier to climb the gradual hill than the steep one. Of course, you must travel a longer distance along the road as you climb the gradual hill than you do on the steep hill, so the work you do is the same in either case.

Ramps and inclined planes show up in many devices, where they reduce the forces needed to perform otherwise difficult tasks. They also change the character of certain activities. Skiing wouldn't be very much fun if the only slopes available were horizontal or vertical. By choosing ski slopes of various grades, you can select the ramp forces that set

you in motion. Gentle slopes leave only small ramp forces and small accelerations; steep slopes produce large ramp forces and large accelerations.

Finally, our observation about mechanical advantage is this: mechanical advantage allows you to do the same work, but you must make a trade-off—you must choose whether you want a large force or a large distance. The product of the two parts, force times distance, remains the same.

Check Your Understanding #6: Access Ramps

Ramps for handicap entrances to buildings are often quite long and may even involve several sharp turns. A shorter, straighter ramp would seem much more convenient. What consideration leads the engineers designing these ramps to make them so long?

Answer: The engineers must limit the amount of force needed to propel a wheelchair steadily up the ramp. The steeper the ramp, the more force is required.

Why: A person traveling in a wheelchair on a level surface experiences little horizontal force and can move at constant velocity with very little effort. However, climbing a ramp at constant velocity requires a substantial uphill force equal in magnitude to the downhill force from gravity. The steeper the ramp, the more uphill force is needed to maintain constant velocity. A 12 to 1 grade (12 meter of ramp surface for each meter of rise in height) is the accepted limit to how steep such a long ramp can be.

Epilogue for Chapter 1

In this chapter we have examined three everyday things and explored the basic physical laws that govern their behaviors. In Skating, we looked at the concept of inertia and observed that objects accelerate only in response to forces. In Falling Balls, we introduced an important type of force—weight—and saw how weight causes unsupported objects to fall with equal downward accelerations.

In Ramps, we encountered another type of force—support force. We also saw how the work done in changing an object's altitude doesn't depend on how you raise the object, since work is actually the mechanical means for transferring energy from one object to another. Energy, we noted, is one of the conserved physical quantities that govern the motion of objects in our universe. The particular type of energy we studied in Ramps was gravitational potential energy, the potential energy associated with the force of gravity.

Explanation: Removing a Tablecloth from a Table

The dishes remain in place because of their inertia. As we've seen, an object in motion tends to remain in motion, while an object at rest tends to remain at rest.

Before you pull on the tablecloth, the dishes sit motionless on its surface, and they tend to remain that way. By sliding the tablecloth off the table as quickly and smoothly as possible, you ensure that whatever force the tablecloth exerts on each dish occurs only for a very short time. As a result, the dishes undergo only the tiniest changes in velocity and remain essentially stationary on the tabletop.

Chapter Summary and Important Laws and Equations

How Skating Works: When you're gliding forward on frictionless skates, you experience no horizontal forces and move at a constant velocity. Inertia alone carries you forward. To change your velocity—to accelerate—something must exert a horizontal force on you. A forward force speeds you up, while a backward force slows you down. Since sideways forces change your direction of travel, they, too, make you accelerate.

How Falling Balls Work: Any ball that's subject only to its weight—the force due to gravity—is a falling ball. It accelerates downward at a steady rate. Gravity affects only the ball's vertical motion, causing the ball's vertical component of velocity to increase steadily in the downward direction. If the ball were initially moving horizontally, it would continue that horizontal motion and drift steadily downfield as it falls.

A falling ball that's initially rising soon stops rising and begins to descend. The larger its initial upward component of velocity, the longer the ball rises and the greater its peak height. When the ball then begins to descend, the peak height determines how long it takes for the ball to reach the ground.

When you throw a ball, the vertical component of initial velocity determines how long the ball remains aloft. The horizontal component of initial velocity determines how quickly the ball moves downfield. A thrower intuitively chooses an initial speed and direction for a ball so that it moves just the right distance downfield by the time it descends to the desired height.

How Ramps Work: An object at rest on level ground experiences two forces: its downward weight and an upward support force from the ground that exactly balances that weight. The net force on the object is thus zero. If the ground is replaced with a ramp, however, the support force is no longer directly upward and the net force on the object isn't zero. Instead, the net force points downhill along the ramp and is a ramp force that's equal to the weight of the object multiplied by the ratio of the ramp's rise to the ramp's length. If a 10-m-long ramp rises 1 m in height, then this ratio is 1 m divided by 10 m, or 0.10. The ramp force downhill along this ramp is thus only 10% of the object's weight.

To stop an object from accelerating down a ramp, you must balance the downhill ramp force by pushing equally hard uphill. In fact, if you exert more force up the ramp than it experiences down the ramp, the object will begin to accelerate up the ramp.

It takes less force to push an object up a ramp than to lift it directly upward, but you must push that object a longer distance along the ramp. Overall, the work you do in raising the object from one height to another is the same, whether or not you use the ramp. However, the ramp gives you a mechanical advantage, allowing you to do work that would require an unrealistically large force by instead exerting a smaller force for a longer distance.

1. Newton's first law of motion: An object that is free from all outside forces travels at a constant velocity, covering equal distances in equal times along a straight-line path.

2. Newton's second law of motion: An object's acceleration is equal to the net force exerted on that object divided by the object's mass, or

$$\text{net force} = \text{mass} \cdot \text{acceleration}. \quad (1.1.2)$$

3. Relationship between mass and weight: An object's weight is equal to its mass times the acceleration due to gravity, or

$$\text{weight} = \text{mass} \cdot \text{acceleration due to gravity}. \quad (1.2.1)$$

4. Velocity of an object experiencing constant acceleration: The object's present velocity differs from its initial velocity by its acceleration times the time since it was at that initial velocity, or

$$\begin{aligned} \text{present velocity} &= \text{initial velocity} \\ &\quad + \text{acceleration} \cdot \text{time}. \end{aligned} \quad (1.2.2)$$

5. Position of an object experiencing constant acceleration: The object's present position differs from its initial position by its

average velocity since it was at that initial position times the time since it was at that initial position, or

$$\begin{aligned} \text{present position} &= \text{initial position} + \text{initial velocity} \cdot \text{time} \\ &\quad + \frac{1}{2} \cdot \text{acceleration} \cdot \text{time}^2. \end{aligned} \quad (1.2.3)$$

6. Newton's third law of motion: For every force that one object exerts on a second object, there is an equal but oppositely directed force that the second object exerts on the first object.

7. Definition of work: The work done on an object is equal to the force exerted on that object times the distance that object travels in the direction of the force, or

$$\text{work} = \text{force} \cdot \text{distance}. \quad (1.3.1)$$

8. Gravitational potential energy: An object's gravitational potential energy is its mass times the acceleration due to gravity times its height above a zero level, or

$$\begin{aligned} \text{gravitational potential energy} &= \text{mass} \cdot \text{acceleration} \\ &\quad \text{due to gravity} \cdot \text{height}. \end{aligned} \quad (1.3.2)$$

The Laws of Motion

PART 2

In the previous chapter, we saw how things move from place to place and encountered energy, an important conserved quantity. But motion doesn't always involve a change of position, and energy isn't nature's only conserved quantity. In this chapter, we'll take a look at a second type of motion—rotation—and at two other conserved quantities—momentum and angular momentum. Spinning objects are quite common, and we'll do well to explore their laws of motion before proceeding much further. With those additional concepts under our belts, we'll be ready to explore the physics behind a broad assortment of mechanical objects.

ACTIVE LEARNING EXPERIMENTS

A Spinning Pie Dish

Spinning a dish on the top of a narrow post seems like a simple activity. But don't let its uncomplicated appearance deceive you: there are lots of physics involved in keeping the dish turning, in gradually slowing it down, and in preventing it from falling off the post.

An easy way to experiment with a spinning dish is to tape a pencil vertically to the edge of a table or chair so that its eraser projects several inches upward into the air. To avoid wobbling problems, the tape should hold the pencil rigidly in place, and the table or chair should be sturdy and stable.

Now prepare to balance a metal pie dish on the eraser before giving it a twist. If you don't have a pie dish, you can use a Frisbee, a deep plastic plate, or a shallow plastic bowl instead. Be creative. You probably have something that will work; just don't use Grandma's heirloom porcelain unless you're willing to face the possible consequences.

The first thing you'll need to do is to balance the dish on the eraser. Why is it easiest to balance the dish when

it's upside down? Now give the dish a gentle spin. What sort of influence do you have to exert on the dish to start it rotating? If the dish doesn't wobble and the pencil remains stationary, the dish should spin for a while before coming to a stop. Once you're no longer touching the dish, what keeps it turning? On the other hand, why doesn't it keep turning forever?

Now flip the pencil over so that its sharp point projects upward. What will happen when you place the dish on that point? Will the dish still balance? When you spin the dish, will it turn for a longer or shorter time than on the bare eraser? If the dish is soft and the point digs into it, protect the dish's bottom by taping a coin to it. How does this improved pivot affect the dish's rotation? Can you prolong the spin by taping weights around the outer edge of the dish? Is there a way to get the dish spinning just by blowing on it? Can you relate this motion to that of a spinning skater or a bicycle wheel?



Courtesy Lou Bloomfield



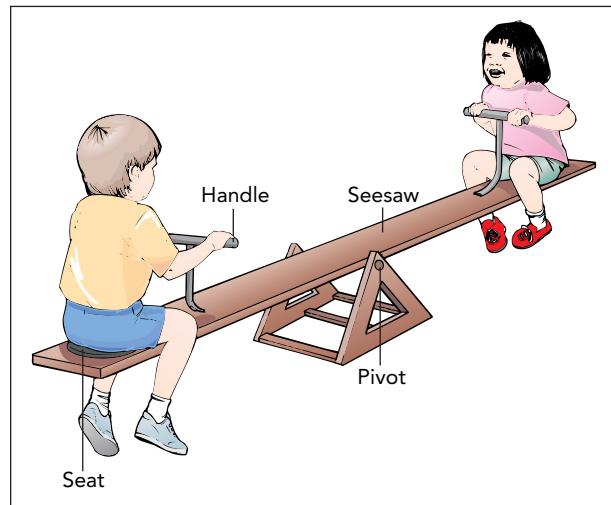
Chapter Itinerary

We're going to explore the laws of rotational motion and two new conserved quantities in the context of three everyday objects: (1) *seesaws*, (2) *wheels*, and (3) *bumper cars*. In *Seesaws*, we look at twists and turns, and see how two children manage to rock a seesaw back and forth. In *Wheels*, we examine how friction affects motion and learn how wheels make a

vehicle more mobile. In *Bumper Cars*, we learn the physics behind collisions and uncover some of the simple rules that govern what initially appear to be complicated motions. For a more complete preview of what we examine in this chapter, flip ahead to the Chapter Summary and Important Laws and Equations at the end of the chapter.

SECTION 2.1

Seesaws



The ramp that we examined in Section 1.3 is only one tool that provides mechanical advantage. In this section, we'll look at another such device: the type of lever known as a seesaw. Although seesaws move, their motion differs from that of the objects in Chapter 1. Seesaws don't go anywhere; they rotate. In this section, we'll revisit many of the laws of motion that we encountered previously, but this time we'll see these laws in a new context—rotational motion.

Questions to Think About: A playground seesaw balances only when the children riding it are properly situated. What do we

mean by a balanced seesaw? Why does it matter just where the children sit on the seesaw? What must the children do to start the seesaw turning? To keep it turning? Who is doing work on whom as they rock back and forth?

Experiments to Do: To get a feel for how rotating objects work, find a rigid ruler with a hole in its center—the kind that can be clipped into a three-ring binder. If you support the ruler by putting the tip of an upright pencil into the central hole, you'll find that the ruler exhibits rotational inertia; that is, it either remains stationary, at whatever orientation you choose, or rotates steadily about the central hole. (Eventually, the ruler comes to rest because of friction, a detail that we'll continue to ignore for now.) Now, push on one end of the ruler. What happens? Try pushing the ruler's end toward its central hole. What happens then? What is the most effective way to make the ruler spin?

Now lay the pencil on a table and place the ruler flat on top of it so that the pencil and the ruler are at right angles, or perpendicular, to each other. If you center the ruler on the pencil, the ruler will balance. How is this balanced ruler similar to the freely turning, inertial ruler of the previous paragraph? What role does gravity play in balancing the ruler? Load the two ends of the ruler with coins or other small weights, trying as you do so to keep the ruler balanced. Try placing the coins at different positions relative to the pencil. Is there any way you can balance a light weight on one end with a heavy weight on the other end?

The Seesaw

Any child who has played on a seesaw with friends of different sizes knows that the toy works best for two children of roughly the same weight (Fig. 2.1.1a). Evenly matched riders balance each other, and this balance allows them to rock back and forth easily. In contrast, when a light child tries to play seesaw with a heavy child, the heavy child's side of the seesaw drops rapidly and hits the ground with a thud (Fig. 2.1.1b). The light child is tossed into the air.

There are several solutions to the heavy child-light child problem. Of course, two light children could try to balance one heavy child. But most children eventually figure out that if the heavy child sits closer to the seesaw's pivot, the seesaw will balance (Fig. 2.1.1c). The children can then make the seesaw tip back and forth easily, just as it does when two evenly matched children ride at its ends. This is a pretty useful trick, and we'll explore it later in this section. First, though, we'll need to look carefully at the nature of rotational motion.

For simplicity, let's ignore the mass and weight of the seesaw itself. There are then only three forces acting on the occupied seesaw shown in Fig. 2.1.1: two downward forces (the weights of the two children) and one upward force (the support force of the central pivot). Seeing those three forces, we may immediately think about net forces and begin to look for some overall acceleration of this toy and its riders. But we know that the seesaw remains where it is in the playground and isn't likely to head off for Kalamazoo or the center of Earth anytime soon. Because the seesaw's fixed pivot always provides just enough upward and sideways force to keep the seesaw from accelerating as a whole, the seesaw always experiences zero net force and never leaves the playground. Overall movement of an object from one place to another is called **translational motion**. Although the seesaw never experiences translational motion, it can turn around the pivot, and thus it experiences a different kind of motion. Motion around a fixed point (which prevents translation) is called **rotational motion**.

Rotational motion is what makes a seesaw interesting—the whole point of a seesaw is that it can rotate so that one child rises and the other descends. (You may not think of going up and down as rotating, but if the ground weren't there, the seesaw would be able to rotate in a big circle.) But what causes the seesaw to rotate, and what observations can we make about the process of rotation?

To answer those questions, we'll need to examine several new physical quantities associated with rotation and explore the laws of rotational motion that relate them to one another. We'll do these things both by studying the workings of seesaws and other rotating objects and by looking for analogies between translational motion and rotational motion.

Imagine holding onto the seesaw in Fig. 2.1.1a to keep it level for a moment while the child on the left climbs off the seesaw. Now imagine letting go of the seesaw. As soon as you let go, the seesaw begins to rotate clockwise, and the child on the right descends toward the ground. The seesaw's motion is fairly slow at first, but it moves more and more quickly until that child strikes the ground with a teeth-rattling thump. We could describe the seesaw being twisted from rest in the following way:

“The seesaw starts out not rotating at all. When we release the seesaw, it begins to rotate clockwise. The seesaw's rate of rotation increases continuously in the clockwise direction until the seesaw strikes the ground.”

This description sounds a lot like the description of a falling ball released from rest:

“The ball starts out not moving at all. When we release the ball, it begins to move downward. The ball's rate of translation increases steadily in the downward direction until the ball strikes the ground.”

The statement about the seesaw involves rotational motion, while the statement about the ball involves translational motion. Their similarity isn't a coincidence; the concepts and laws of rotational motion have many analogies in the concepts and laws of translational motion. The familiarity that we've acquired with translational motion will help us examine rotational motion.

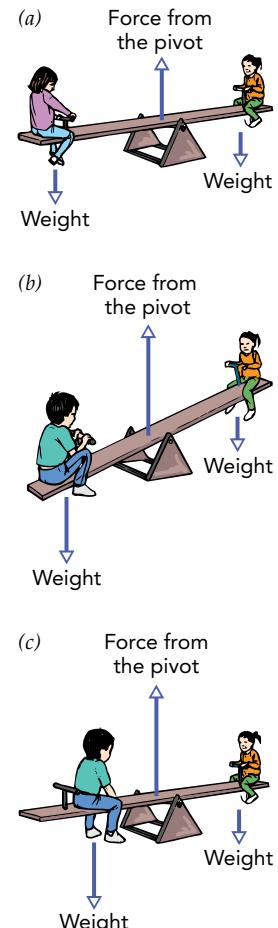


Fig. 2.1.1 (a) When two children of equal weight sit at opposite ends of a seesaw, it balances. (b) When their weights are not equal, the heavy child descends. (c) If the heavy child moves closer to the pivot, the seesaw can balance.

Check Your Understanding #1: Wheel of Fortune Cookies

The guests at a large table in a Chinese restaurant use a revolving tray, a lazy Susan, to share the food dishes. How does the motion of the lazy Susan differ from that of the passing dessert cart?

Answer: The lazy Susan undergoes rotational motion, while the dessert cart undergoes translational motion.

Why: The lazy Susan has a fixed pivot at its center. This pivot never goes anywhere, no matter how you rotate the lazy Susan. In contrast, the dessert cart moves about the room and has no fixed point. The server can rotate the dessert cart when necessary, but its principal motion is translational.

The Motion of a Dangling Seesaw

In the previous chapter we looked at the concept of translational inertia, which holds that a body in motion tends to stay in motion and a body at rest tends to stay at rest. This concept led us to Newton's first law of translational motion. Inserting the word *translational* here is a useful revision because we're about to encounter the corresponding concepts associated with rotational motion. We'll begin that encounter by observing a seesaw that's free of outside rotational influences. We'll then examine how the seesaw responds to outside influences such as its pivot or a handful of young riders. Because of the similarities between rotational and translational motions, this section closely parallels our earlier examinations of skating and falling balls.

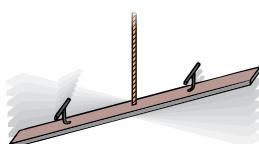


Fig. 2.1.2 A seesaw that's dangling from a rope at its middle. Since nothing twists it, the seesaw rotates steadily about a fixed line in space.

Let's suppose that your local playground is installing a new seesaw and that this seesaw is presently dangling from a rope (Fig. 2.1.2). The rope is attached to the middle of the seesaw, where it supports the seesaw's weight but exerts no other influences on the seesaw. Most important, let's suppose that the dangling seesaw can spin and pivot with complete freedom—nothing pushes on it or twists it—and that the rope doesn't get tangled or in the way. This dangling seesaw is free to turn in any direction, even completely upside down. You, the observer, are standing motionless near the seesaw. When you look over at the seesaw, what does it do?

If the seesaw is stationary, then it will remain stationary. However, if it's rotating, it will continue rotating at a steady pace about a fixed line in space. What keeps the seesaw rotating? Its **rotational inertia**. A body that's rotating tends to remain rotating; a body that's not rotating tends to remain not rotating. That's how our universe works.

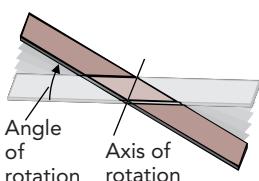


Fig. 2.1.3 You can specify this seesaw's angular position, relative to its horizontal reference orientation, as the axis about which it was rotated to reach its new orientation and the angle through which it was rotated.

To describe the seesaw's rotational inertia and rotational motion more accurately, we'll need to identify several physical quantities associated with rotational motion. The first is the seesaw's orientation. At any particular moment, the seesaw is oriented in a certain way—that is, it has an **angular position**. Angular position describes the seesaw's orientation relative to some reference orientation; it can be specified by determining how far the seesaw has rotated away from its reference orientation and the axis or line about which that rotation has occurred. The seesaw's angular position is a vector quantity, pointing along the rotation axis with a magnitude equal to the rotation angle (Fig. 2.1.3). Because *changes* in orientation are usually more interesting than orientation itself, angular position is a relatively little-used physical quantity.

The SI unit of angular position is the **radian**, the natural unit for angles. It's a natural unit because it follows directly from geometry, not from an arbitrary human choice or convention the way most units do. Geometry tells us that a circle of radius 1 has a circumference of 2π . By letting arc lengths around that circle's circumference specify angles, we are using radians. For example, there are 2π radians (or 360°) in a full circle and $\pi/2$ radians (or 90°) in a right angle. Since the radian is a natural unit, it is often omitted from calculations and derived units.

If the seesaw is rotating, then its angular position is changing; this brings us to our first important vector quantity of rotational motion—angular velocity. **Angular velocity** measures the rate at which the seesaw’s angular position is changing with time. Its magnitude is the seesaw’s **angular speed**, the angle through which the seesaw turns in a certain amount of time,

$$\text{angular speed} = \frac{\text{change in angle}}{\text{time}}$$

and its direction is the axis about which that rotation proceeds. The SI unit of angular velocity is the **radian per second** (abbreviated 1/s).

The seesaw’s **axis of rotation** is the line in space about which the seesaw is rotating. However, just knowing that line isn’t quite enough: Is the seesaw rotating clockwise or counterclockwise?

To resolve this ambiguity, we take advantage of the fact that any line has two directions to it. Once we have identified the line about which the seesaw is rotating, we can look down that line at the seesaw from both directions. From one, the seesaw appears to be rotating clockwise; from the other, counterclockwise. By convention, we choose to view the seesaw from the direction in which it appears to be rotating clockwise and say that the seesaw’s rotation axis points away from our eye toward the seesaw. This convention is called the **right-hand rule** because if the fingers of your right hand are curling around the axis in the way the seesaw is rotating, then your thumb is pointing along the seesaw’s rotation axis (Fig. 2.1.4). (Even if you’re left-handed, this rule still requires your right hand, Fig. 2.1.5).

Remembering this convention isn’t as important as understanding why you must specify the direction when reporting a rotating object’s angular velocity. Just as translational velocity consists of a translational speed and a direction in which the translational motion occurs, so angular velocity consists of a rotational speed and a direction about which the rotational motion occurs.

We’re now prepared to describe the rotational motion of the dangling seesaw. Because of its freedom from outside influences and its rotational inertia, its angular velocity is constant. If the seesaw is rotating, it keeps on rotating, always at the same angular speed, always about the same axis of rotation. If it is not rotating, however, its angular velocity is zero and remains zero.

As you might suspect, this observation isn’t unique to seesaws. It is **Newton’s first law of rotational motion**, which states that a rigid object that is not wobbling and is not subject to any outside influences rotates at a constant angular velocity, turning equal amounts in equal times about a fixed axis of rotation. The outside influences referred to in this law are called **torques**—a technical term for twists and spins. When you twist off the lid of a jar or spin a top with your fingers, you’re exerting a torque.

This law excludes objects that wobble or can change shape as they rotate because those objects have more complicated motions. They are covered instead by a more general principle—the conservation of angular momentum—that we’ll meet up with in Section 2.3.

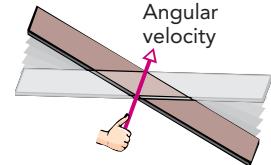


Fig. 2.1.4 This seesaw is spinning about the rotation axis shown. The direction of the seesaw’s angular velocity is defined by the right-hand rule.

Courtesy Lou Bloomfield



Fig. 2.1.5 Right-hand rules were created to help us remember several physics conventions, such as the agreed-on direction of angular velocity. These rules start with a convention that we all find extremely familiar, specifically which hand is your right hand, and use it to specify the conventions associated with physical laws. Even if you are left-handed, you still know which hand to call your right hand and therefore understand this convention.

NEWTON’S FIRST LAW OF ROTATIONAL MOTION

A rigid object that is not wobbling and is not subject to any outside torques rotates at a constant angular velocity, turning equal amounts in equal times about a fixed axis of rotation.



Check Your Understanding #2: Going for a Spin

A rubber basketball floats in a swimming pool. It experiences zero torque, no matter which end of it is up. If you spin the basketball and then let go, how will it move?

Answer: It will continue to spin at a steady pace about a fixed rotational axis (although friction with the water will gradually slow the ball's rotation).

Why: Because the basketball is free of torques, the outside influences that affect rotational motion, it has a constant angular velocity. If you spin the basketball, it will continue to spin about whatever axis you chose. If you don't spin the basketball, its angular velocity will be zero and it will remain stationary.

The Seesaw's Center of Mass

Even without visiting the playground, you can find many objects that are nearly free from torques: a baton thrown overhead by a baton twirler, for example, or a juggler's club whirling through the air between two clowns. These motions, however, are complicated because those freely moving objects rotate and translate at the same time. The spinning baton travels up and down, the turning club arcs through the air, and if the rope breaks, your seesaw will fall as it spins. How can we distinguish their translational motions from their rotational motions?

Once again, we can make use of a wonderful simplification of physics. There's a special point in or near a free object about which all of its mass is evenly distributed and about which it naturally spins—its **center of mass**. The axis of rotation passes right through this point so that, as the free object rotates, the center of mass doesn't move unless the object has an overall translational velocity. The center of mass of a typical ball is at its geometrical center, while the center of mass of a less symmetrical object depends on how the mass of that object is distributed. You can begin to find a small object's center of mass by spinning it on a smooth table and looking for the fixed point about which it spins (Fig. 2.1.6).

Center of mass allows us to separate an object's translational motion from its rotational motion. As a juggler's club arcs through space, its center of mass follows the simple path we discussed in Section 1.2 on falling balls (Fig. 2.1.7). At the same time, the club's rotational motion about its center of mass is that of an object that's free of outside torques: if it's not wobbling, it rotates with a constant angular velocity.

Many of the objects we'll examine in this book translate and rotate simultaneously, and it's worth remembering that we can often separate these two motions by paying attention to an object's center of mass. For example, your seesaw has been designed with its center of mass located exactly where the pivot will go. As a result, the pivot will prevent any translational motion of the installed seesaw while permitting nearly free rotational motion of the seesaw, at least about one axis.

When examining rotational motion, we need to choose the pivot point or **center of rotation**, the point around which all the physical quantities of rotation are defined. For a free object, the natural pivot point is its center of mass. For a constrained object, the best pivot point may be determined by the constraints—a door, for example, pivots about its hinges. Since your installed seesaw will continue to pivot about its center of mass, that point will remain the best choice for the center of rotation.

Once we've chosen the pivot, however, we must calculate all the physical quantities of rotation about that pivot. Because it's tedious to write *about its hinges* or *about its center of mass* every time I mention a physical quantity of rotation, I will often omit those phrases when the pivot point is obvious or already identified.

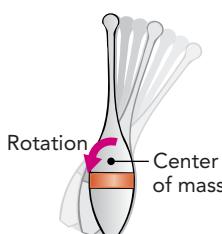


Fig. 2.1.6 This club spins about its center of mass, which remains stationary.

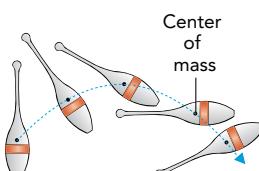


Fig. 2.1.7 A juggler's club that is traveling through space rotates about its center of mass as its center of mass travels in the simple arc associated with a falling object.

Check Your Understanding #3: Tracking the High Dive

When a diver does a rigid, open somersault off a high diving board, his motion appears quite complicated. Can this motion be described simply? How?

Answer: Yes. His center of mass falls smoothly, obeying the rules governing falling objects. As he falls, his body rotates at constant angular velocity about his center of mass.

Why: Like a thrown football or tossed baton, the diver is a rigid, rotating object. His motion can be separated into translational motion of his center of mass (it falls) and rotational motion about his center of mass (he rotates about it at constant angular velocity). While the diver may never think of his motion in these terms, he is aware intuitively of the need to handle both his rotational and translational motions carefully. Hitting the water with his chest because he mishandled his rotation isn't much more fun than hitting the board because he mishandled his translation.

How the Seesaw Responds to Torques

The workers are eating lunch, so the seesaw is still hanging from the rope. Why can't this dangling seesaw change its rotational speed or axis of rotation? Because it has rotational mass **1**. **Rotational mass** is the measure of an object's *rotational inertia*, its resistance to changes in its *angular velocity*. An object's rotational mass depends both on its ordinary mass and on how that mass is distributed within the object. The SI unit of rotational mass is the **kilogram-meter²** (abbreviated $\text{kg} \cdot \text{m}^2$). Because the seesaw has rotational mass, its angular velocity will change only if something twists it or spins it. In other words, it must experience a torque.

Torque—our second important vector quantity of rotational motion—has both a magnitude and a direction. The more torque you exert on the seesaw, the more rapidly its angular velocity changes. Depending on the direction of the torque, you can make the seesaw turn more rapidly or less rapidly or even make it rotate about a different axis. How do you determine the direction of a particular torque? One way is to imagine exerting this torque on a stationary ball floating in water (Fig. 2.1.8a,b). The ball will begin to rotate, acquiring a nonzero angular velocity (Fig. 2.1.8c). The direction of this angular velocity is that of the torque. The SI unit of torque is the **newton-meter** (abbreviated $\text{N} \cdot \text{m}$).

The larger an object's rotational mass, the more slowly its angular velocity changes in response to a specific torque (Fig. 2.1.9). You can easily spin a basketball with the tips of your fingers, but it's much harder to spin a bowling ball. The bowling ball's larger rotational mass comes about primarily because it has a greater ordinary mass than the basketball.

However, rotational mass also depends on an object's shape, particularly on how far each portion of its ordinary mass is from the axis of rotation. The farther a portion of mass is from that axis, the more rapidly it must accelerate as the entire object undergoes angular acceleration and the more leverage it has with which to oppose that acceleration. We'll examine levers shortly, but the consequence of these two effects of distance from the rotation axis is that each portion of mass contributes to the object's rotational mass in proportion to the square of its distance from that axis. That's why an object that has most of its mass located near the axis of rotation will have a much smaller rotational mass than an object of the same mass that has most of its mass located far from that axis. Thus a ball of pizza dough has a smaller rotational mass than the finished pizza. The bigger the pizza gets, the harder it is to start or stop spinning.

Because an object's rotational mass depends on how far its mass is from the axis of rotation, its rotational mass may change when its axis of rotation changes, even if it's rotating about its center of mass. For example, less torque is required to spin a tennis racket about its handle (Fig. 2.1.10a) than to flip the racket head-over-handle (Fig. 2.1.10b). When you spin the tennis racket about its handle, the axis of rotation runs right through the handle so that most of the racket's mass is fairly close to the axis and the

1 For clarity and simplicity, this book refers to the measure of an object's rotational inertia as *rotational mass*. However, this quantity is known more formally as **moment of inertia**.

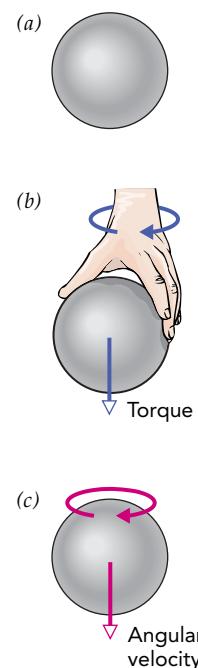


Fig. 2.1.8 If you start with a ball that's not spinning (a) and twist it with a torque (b), the ball will acquire an angular velocity (c) that's in the same direction as that torque.

Fig. 2.1.9 Spinning a merry-go-round is difficult because of its large rotational mass. Despite the large torque exerted by this boy, the merry-go-round's angular velocity increases slowly.



© Chris Harvey/Stone/Getty Images

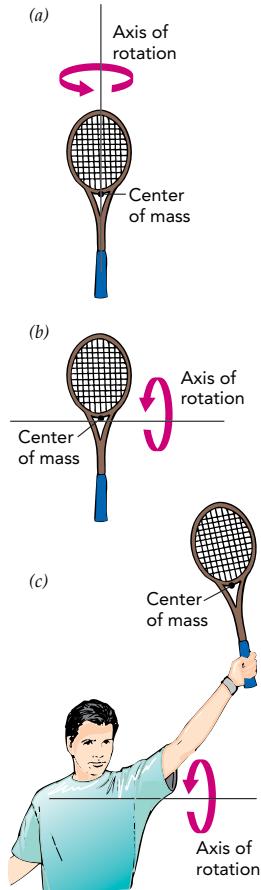


Fig. 2.1.10 A tennis racket's rotational mass depends on the axis about which it rotates. Its rotational mass is small (a) when it rotates about its handle and large (b) when it rotates head-over-handle. (c) If you make it rotate about your shoulder, its rotational mass becomes even larger.

rotational mass is small. When you flip the tennis racket head-over-handle, the axis of rotation runs across the handle so that both the head and the handle are far away from the axis and the rotational mass is large. The tennis racket's rotational mass becomes even larger when you hold it in your hand and make it rotate about your shoulder rather than about its center of mass (Fig. 2.1.10c).

When something exerts a torque on the dangling seesaw, its angular velocity changes; in other words, it undergoes angular acceleration, our third important vector quantity of rotational motion. **Angular acceleration** measures the rate at which the seesaw's *angular* velocity is changing with time. It's analogous to acceleration, which measures the rate at which an object's *translational* velocity is changing with time. Just as with acceleration, angular acceleration involves both a magnitude and a direction. An object undergoes angular acceleration when its angular speed increases or decreases or when its angular velocity changes directions. The SI unit of angular acceleration is the **radian per second²** (abbreviated $1/\text{s}^2$).

If the seesaw is experiencing several torques at once, it can't respond to them individually. Instead, it undergoes angular acceleration in response to the **net torque** it experiences, the sum of all the individual torques being exerted on it.

There is a simple relationship between the net torque exerted on the seesaw, its rotational mass, and its angular acceleration. The seesaw's angular acceleration is equal to the net torque exerted on it divided by its rotational mass or, as a word equation,

$$\text{angular acceleration} = \frac{\text{net torque}}{\text{rotational mass}}. \quad (2.1.1)$$

The seesaw's angular acceleration, as we've seen, is in the same direction as the net torque exerted on it.

This relationship is **Newton's second law of rotational motion**. Structuring the relationship this way distinguishes the causes (net torque and rotational mass) from their effect (angular acceleration). Nonetheless, it has become customary to rearrange the relationship to eliminate the division. In its traditional form, the relationship can be written in a word equation:

$$\text{net torque} = \text{rotational mass} \cdot \text{angular acceleration}, \quad (2.1.2)$$

in symbols:

$$\tau_{\text{net}} = I \cdot \alpha,$$

and in everyday language:

Spinning a marble is much easier than spinning a merry-go-round.

It's like Newton's second law of translational motion (net force = mass \cdot acceleration), except that net torque has replaced net force, rotational mass has replaced mass, and angular acceleration has replaced acceleration. This new law doesn't apply to nonrigid or wobbling objects, however, because nonrigid objects can change their rotational masses and wobbling ones are affected by more than one rotational mass simultaneously (see the earlier discussion of tennis rackets).

NEWTON'S SECOND LAW OF ROTATIONAL MOTION

The net torque exerted on a rigid object that is not wobbling is equal to that object's rotational mass times its angular acceleration. The angular acceleration points in the same direction as the net torque.

Because it's an equation, the two sides of Eq. 2.1.1 are equal. Any change in the net torque you exert on the seesaw must be accompanied by a proportional change in its angular acceleration. As a result, the harder you twist the seesaw, the more rapidly its angular velocity changes.

We can also compare the effects of a specific torque on two different rotational masses. Equation 2.1.1 indicates that a decrease in rotational mass must be accompanied by a corresponding increase in angular acceleration. If we replace the playground seesaw with one from a dollhouse, the rotational mass will decrease and the angular acceleration will increase. The angular velocity of the doll's seesaw thus changes more rapidly than the angular velocity of a playground seesaw when the two experience identical net torques.

In summary:

1. Your angular position indicates exactly how you're oriented.
2. Your angular velocity measures the rate at which your angular position is changing with time.
3. Your angular acceleration measures the rate at which your angular velocity is changing with time.
4. For you to undergo angular acceleration, you must experience a net torque.
5. The more rotational mass you have, the less angular acceleration you experience for a given net torque.

This summary of the physical quantities of rotational motion is analogous to the summary for translational motion in Section 1.1. Take a moment to compare the two.

Quantity	SI Unit	English Unit	SI → English	English → SI
Angular position	radian (1)	radian (1)		
Angular velocity	radian per second (1/s)	radian per second (1/s)		
Angular acceleration	radian per second ² (1/s ²)	radian per second ² (1/s ²)		
Torque	newton-meter (N · m)	foot-pound (ft · lbf)	$1 \text{ N} \cdot \text{m} = 0.73757 \text{ ft} \cdot \text{lbf}$	$1 \text{ ft} \cdot \text{lbf} = 1.3558 \text{ N} \cdot \text{m}$
Rotational mass	kilogram-meter ² (kg · m ²)	pound-foot ² (lbf · ft ²)	$1 \text{ kg} \cdot \text{m}^2 = 23.730 \text{ lbf} \cdot \text{ft}^2$	$1 \text{ lbf} \cdot \text{ft}^2 = 0.042140 \text{ kg} \cdot \text{m}^2$

Check Your Understanding #4: The Merry-Go-Round

The merry-go-round is a popular playground toy (see Fig. 2.1.9). Already challenging to spin empty, a merry-go-round is even harder to start or stop when there are lots of children on it. Why is it so difficult to change a full merry-go-round's angular velocity?

Answer: The full merry-go-round has a huge rotational mass.

Why: Starting or stopping a merry-go-round involves angular acceleration. As the pusher, you exert a torque on the merry-go-round, and it undergoes angular acceleration. This angular acceleration depends on the merry-go-round's rotational mass, which in turn depends on how much mass it has and how far that mass is from the axis of rotation. With many children adding to the merry-go-round's rotational mass, its angular acceleration tends to be small.

Check Your Figures #1: Hard to Turn

Automobile tires are normally hollow and filled with air. If they were made of solid rubber, their rotational masses would be about 10 times as large. With the wheel lifted off the ground, how much more torque would an automobile have to exert on a solid tire to make it undergo the same angular acceleration as a hollow tire?

Answer: It would need about 10 times as much torque.

Why: To keep the angular acceleration in Eq. 2.1.1 unchanged while increasing the rotational mass by a factor of 10, the torque must also increase by a factor of 10. Solid tires are extremely difficult to spin or to stop from spinning, which is why automobiles use hollow tires.

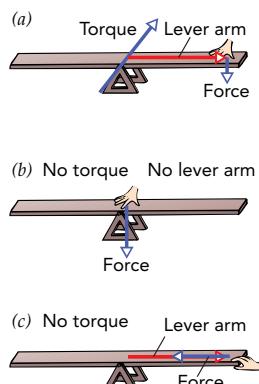


Fig. 2.1.11 (a) When you push on the seesaw, perpendicular to the lever arm, you produce a torque on the seesaw. But when you (b) push at the pivot or (c) push parallel to the lever arm, you produce no torque.

Forces and Torques

The workers have finally installed the seesaw. Because the pivot supporting the seesaw passes directly through the seesaw's center of mass, the seesaw rotates about its own natural pivot. Moreover, the pivot supports the seesaw's weight in a manner that leaves it free to obey Newton's first law of rotational motion. That is, when unoccupied and not influenced by anything else, the seesaw is either motionless or rotates with constant angular velocity about its pivot.

To change the seesaw's angular velocity, you have to exert a torque on it. But how do you actually exert a torque? To find out, you put your hand on one end of the seesaw and push that end down (Fig. 2.1.11a). If the seesaw was motionless, it starts turning. If it was already turning, its rate of rotation changes. You have indeed exerted a torque on the seesaw.

You started by exerting a *force* on the seesaw—you pushed on it—so forces and torques must be related somehow. Sure enough, a force can produce a torque and a torque can produce a force. To help us explore their relationship, let's think of all the ways *not* to produce a torque by pushing on the seesaw.

What happens if you push on the seesaw exactly where the pivot passes through it (Fig. 2.1.11b)? Nothing happens—there is no angular acceleration. If you move a little away from the pivot, you can get the seesaw rotating, but you have to push hard. You do much better if you push on the end of the seesaw, where even a small force can start the seesaw rotating. The shortest distance and direction from the pivot to the place where you push on the seesaw is a vector quantity called the **lever arm**; in general, the longer the lever arm, the less force it takes to cause a particular angular acceleration. Our first observation about producing a torque with a force is this: you obtain more torque by exerting that force farther from the pivot or axis of rotation. In other words, the torque is proportional to the length of the lever arm.

Another ineffective way to start the seesaw rotating is to push its end directly toward or away from the pivot (Fig. 2.1.11c). When your force is directed parallel to the lever arm, as it is in this case, it produces no torque on the seesaw. Our second observation about producing a torque with a force is that your force must have a component that is perpendicular to the lever arm and only that perpendicular component contributes to the torque.

We can summarize these two observations as follows: the torque produced by a force is equal to the lever arm times that force, where we include only the component of the force that is perpendicular to the lever arm. This relationship can be written as a word equation:

$$\text{torque} = \text{lever arm} \cdot \text{force perpendicular to lever arm}, \quad (2.1.3)$$

in symbols:

$$\tau = r \cdot F_{\perp},$$

and in everyday language:

When twisting an unyielding object, it helps to use a long wrench.

The directions of the force and lever arm also determine the direction of the torque. The three directions follow another right-hand rule (Fig. 2.1.12). If you point your right index finger in the direction of the lever arm and your bent middle finger in the direction of the force, then your thumb will point in the direction of the torque. Thus in Fig. 2.1.12a, the lever arm points to the right, the force points downward, and the resulting torque points into the page so that the seesaw undergoes angular acceleration in the clockwise direction. In Fig. 2.1.12b, the lever arm has reversed directions and so has the torque.

► Check Your Understanding #5: Cutting Up Cardboard

When you cut cardboard with a pair of scissors, it's best to move the cardboard as close as possible to the scissors' pivot. Explain.

Answer: The closer the cardboard is to the pivot, the more force it must exert on the scissors to produce enough torque to keep the scissors from rotating closed. When the cardboard is unable to produce enough torque, the scissors cut through it.

Why: When you place paper close to the pivot of a pair of scissors, you are requiring that paper to exert enormous forces on the scissors to keep them from rotating closed. Rotations are started and stopped by torques, and forces exerted close to the pivot exert relatively small torques.

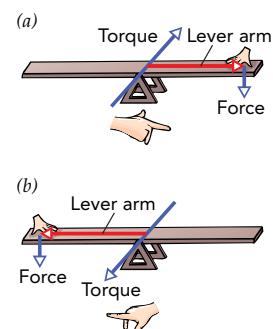


Fig. 2.1.12 The torque on a seesaw obeys a right-hand rule: if your index finger points along the lever arm and your middle finger points along the force, your thumb points along the torque.

Check Your Figures #2: A Few Loose Screws

You're trying to remove some rusty screws from your refrigerator, using an adjustable wrench with a 0.2-m (20-cm) handle. Although you push as hard as you can on the handle, you can't produce enough torque to loosen one of the screws. You have a 1.0-m (100-cm) pipe that you can slip over the handle of the wrench to make the wrench effectively 1 m long. How much more torque will you then be able to exert on the screw?

Answer: Five times as much torque as before.

Why: The pipe increases the wrench's lever arm by a factor of 5, from 0.2-m to 1.0 m. According to Eq. 2.1.3, the same force exerted five times as far from the pivot will produce five times as much torque about that pivot. Extending the handle of a lever-like tool is a common technique to increase the available torque, although it can be hazardous for both the tool and its user. Some tools that are designed for such extreme use come with removable handle extensions.

Balancing and Unbalancing the Seesaw

When a child sits on one end of your seesaw, the child's weight produces a torque on the seesaw about its pivot. Gravity is ultimately responsible for that torque, so we'll call it a *gravitational* torque. Because it depends on the lever arm (the vector from the pivot to the child), moving the child to the other end of the seesaw reverses the torque's direction.

What happens when two children are sitting on opposite ends of the seesaw? Each child's weight produces a gravitational torque on the seesaw about its pivot, but those two torques have opposite directions and thus at least partially cancel. If the two children have equal weights and sit at equal distances from the pivot, the two gravitational torques cancel exactly—they sum to zero. The seesaw then experiences no overall gravitational torque about its pivot and it balances.

A **balanced** seesaw experiences no overall gravitational torque about its pivot. If nothing twists it, the balanced seesaw is inertial—the net torque on it is zero. When the children aren't fidgeting, the seesaw is rigid and wobble-free, so it rotates at constant angular velocity, in accordance with Newton's first law of rotational motion. If it's motionless, it stays motionless. If it's turning, it continues turning at a steady pace.

COMMON MISCONCEPTIONS: A Balanced Object Is Motionless

Misconception: A balanced object is perfectly vertical or horizontal and doesn't move.

Resolution: A balanced object is experiencing zero gravitational torque about its pivot, but it may well be tilted or even rotating. In fact, if it is free of other torques, Newton's first law of rotational motion may apply to it. Only when an object's balance is easily disturbed (for example, a pencil standing upright on its point or a pendulum hanging directly below its pivot) are orientation and motionlessness important and then only because any small rotation will give rise to a gravitational torque about its pivot and consequently to the loss of balance.

Two children with different weights can also balance the seesaw, but they must sit at different distances from the pivot. According to Eq. 2.1.3, the gravitational torque a child exerts on the seesaw is proportional to that child's distance from the pivot times that child's weight. A heavier child must therefore sit closer to the pivot to balance a lighter child on the other end of the seesaw.

With two children riding it, the balanced seesaw has a considerable weight, yet that weight produces no torque on the seesaw about its pivot. That's because the seesaw's **center of gravity**, the effective location of the seesaw's overall weight, is located at the pivot. Although gravity gives the children and the board their own individual weights, there

is a unique point—the seesaw's center of gravity—about which all those individual weights are balanced and at which gravity effectively pulls down on the entire seesaw. With the seesaw's center of gravity located at the pivot, gravity has no lever arm with which to produce a torque on the seesaw about the pivot.

Any object or system of objects has a center of gravity—the effective location of its overall weight. Although center of gravity (a gravitational concept) is different from center of mass (an inertial concept), the two conveniently coincide; the seesaw's center of gravity and its center of mass are located at the same point. That coincidence stems from the proportionality between an object's weight and its mass near Earth's surface (see Section 1.2).

COMMON MISCONCEPTIONS: Center of Mass Is Center of Gravity

Misconception: Center of mass and center of gravity are the same idea and can be used interchangeably.

Resolution: Center of mass is an inertial concept: the effective location of an object's mass as it translates and the natural pivot as it rotates. Center of gravity is a gravitational concept: the effective location of the object's weight. Although these two centers coincide for objects near Earth's surface, interchanging the terms is as incorrect as confusing mass and weight.

As two children ride the balanced seesaw, they're inertial much of the time—they're either motionless or turning steadily in one direction or the other. But motionlessness is tedious, and a turning seesaw eventually touches the ground. Therefore, the children deliberately cause the seesaw to undergo angular acceleration so that it rocks back and forth.

They can cause this angular acceleration in two different ways. First, a boy can push on the ground with his feet so that the ground pushes back and thereby produces an additional torque on the seesaw. Although the seesaw remains balanced, that extra torque from the ground causes it to undergo angular acceleration. When the boy stops pushing on the ground, the ground's torque vanishes and the seesaw resumes a constant angular velocity.

Second, they can change their lever arms and thereby unbalance the seesaw. For example, if one girl leans closer to the pivot, that girl's lever arm decreases and so does her gravitational torque. Since the children's gravitational torques no longer sum to zero, the seesaw isn't balanced—it experiences an overall torque due to gravity—and it undergoes angular acceleration. When the girl stops leaning so that the seesaw balances, its angular velocity becomes constant again.

Check Your Understanding #6: Rocking the Boat

Loading a large container ship requires some care in balancing the cargo and fastening it down firmly. The effective pivot about which the ship can rotate in the water is located roughly along the centerline of the ship, from its bow to its stern. Why is improperly fastened-down cargo so dangerous on such a ship, possibly causing it to capsize during a storm?

Answer: A substantial shift in the cargo's position during a storm can unbalance the ship, giving rise to a gravitational torque that may cause the ship to rotate about the effective pivot. The ship may then capsize.

Why: Although most boats can compensate for some amount of cargo imbalance, shifting cargo can easily flip even a fairly stable boat. It happens frequently in real life, often with fatal consequences. Some boats, particularly canoes and racing shells, are notoriously sensitive to unbalanced loading and are easily flipped by careless or moving occupants.

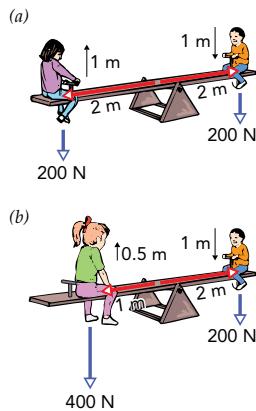


Fig. 2.1.13 When two children ride a balanced seesaw, the work the descending child does on the board is equal to the work the board does on the rising child. (a) When their weights are equal, those two works involve equal forces exerted over equal distance. (b) When one child is heavier than the other, those two works involve different forces exerted over different distances.

Levers and Mechanical Advantage

While discussing how to balance a seesaw, it was helpful to view the board and its riders as a single rotating seesaw, subject to various gravitational and other torques, and undergoing angular acceleration when those torques don't sum to zero. But we can also view the seesaw board individually as a lever that allows two children to do work on one another as it rotates. Most interestingly, that lever provides the mechanical advantage (see Section 1.3) necessary for a light child sitting far from the pivot to do the work of lifting a heavy child sitting near the pivot.

When two children sit on a motionless horizontal seesaw board (Fig. 2.1.13), each child pushes down on the board with a force equal to that child's weight. The force is directed perpendicular to the lever arm from the pivot to the force, so it produces a torque on the board about the pivot. If the children are properly situated on opposite sides of the board, their torques cancel and the seesaw board experiences zero net torque. The seesaw is balanced.

If the board is rotating steadily near horizontal, the forces and angles get a little complicated but the basic story doesn't change. Both children still push down on the board and their torques still sum to zero. Now, however, the descending child does work on the board by pushing it downward as it moves downward, and the board does work on the rising child by pushing that child upward as the child moves upward. Because the seesaw is balanced, the work done on the board by the descending child is equal to the work done by the board on the rising child. In effect, the board transfers energy perfectly from the descending child to the rising child.

To see this effect, suppose two 5-year-olds, each weighing 200-N (45 lbf), are sitting at opposite ends of the seesaw (Fig. 2.1.13a). By exerting a 200-N downward force on the seesaw board, 2 m (6.6 ft) from its pivot, each child produces a torque of $400\text{ N}\cdot\text{m}$ ($300\text{ ft}\cdot\text{lbf}$) on the board ($200\text{ N}\cdot 2\text{ m} = 400\text{ N}\cdot\text{m}$). Since those torques are in opposite directions, they sum to zero.

Now suppose the seesaw rotates clockwise, so that the girl rises 1 m and the boy descends 1 m. Since the boy exerted a 200-N downward force on the board as the board moved downward 1 m, he did $200\text{ N}\cdot 1\text{ m} = 200\text{ J}$ of work on the board. Simultaneously, the board exerted a 200-N upward force on the girl as the girl moved upward 1 m, so the board did 200 J of work on the girl. Overall, the board has transferred 200 J of energy from the boy to the girl.

Now let's replace the 5-year-old girl with a 400-N (90-lbf) teenager (Fig. 2.1.13b). To balance the seesaw, she must sit halfway to the pivot. She exerts a 400-N downward force on the board, 1 m (3.3 ft) from its pivot, leaving the torque unchanged ($400\text{ N}\cdot 1\text{ m} = 400\text{ N}\cdot\text{m}$). As before, the children's torques are in opposite directions and sum to zero. This effect explains how a small boy at the end of the seesaw can balance a large girl nearer the pivot.

Again, the seesaw rotates clockwise, but now the girl rises only 0.5 m when the boy descends 1 m. As before, the boy did 200 J of work on the board. The board exerted a 400-N upward force on the girl as the girl moved upward 0.5 m, so the board did $400\text{ N}\cdot 0.5\text{ m} = 200\text{ J}$ of work on the girl. Once again, the board has transferred 200 J of energy from the boy to the girl, even though the girl weighs more than the boy.

The seesaw's mechanical advantage allows even the tiniest child sitting at its end to lift the heaviest adult sitting near its pivot. Because the child travels much farther than the adult, the child's work on the seesaw equals the seesaw's work on the adult. This effect—a small force exerted for a long distance on one part of a rotating system producing a large force exerted for a short distance elsewhere in that system—is an example of the mechanical advantage associated with levers.

Check Your Understanding #7: Pulling Nails

Some hammers have a special claw designed to remove nails from wood. When you slide the claw under the nail's head and rotate the hammer by pulling on its handle, the claw pulls the nail out of the wood. The hammer's head contacts the wood to form a pivot that's about 10 times closer to the nail than to the handle. The torque you exert on the hammer twists it in one direction, while the torque that the nail exerts on the hammer twists it in the opposite direction. The hammer isn't undergoing any significant angular acceleration, so the torques must nearly balance. If you're exerting a force of 100 N (22 lbf) on the hammer's handle, how much force is the nail exerting on the hammer's claw?

Answer: The nail is exerting about 1000 N (220 lbf).

Why: Since the nail is 10 times closer to the pivot, the nail must exert 10 times the force on the hammer to create the same magnitude of torque as you do pulling on the handle. As the nail pulls on the hammer, the hammer pulls on the nail. Although the wood exerts frictional forces on the nail to keep it from moving, the extracting force overwhelms this friction and the nail slides slowly out of the wood.

Children Support Each Other

While we're thinking about the two children on a balanced seesaw as individual objects, using the seesaw board to exchange energy, we may notice something else about those children: they're using the seesaw board to support one another. In other words, each child is using the board to exert a torque on the other child about the pivot, and that torque cancels the other child's gravitational torque. For example, the boy in Fig. 2.1.13a experiences a clockwise torque due to gravity and a counterclockwise torque from the girl on the other end of the seesaw. Since those two torques sum to zero, the boy rotates at constant angular velocity about the pivot. Likewise, the boy exerts a torque on the girl so that she, too, rotates at constant angular velocity.

Since the girl is exerting a torque on the boy and the boy is exerting a torque on the girl, you might wonder if those two torques are related. It should come as no great surprise that they are equal in amount but oppositely directed. Just as there is a Newton's third law for translational motion, so there is a **Newton's third law of rotational motion**: if one object exerts a torque on a second object, then the second object will exert an equal but oppositely directed torque on the first object. The girl's torque on the boy and the boy's torque on the girl are a Newton's third law pair.

Newton's Third Law of Rotational Motion

For every torque that one object exerts on a second object, there is an equal but oppositely directed torque that the second object exerts on the first object.

It's worth noting that not all equal but oppositely directed torques are Newton's third law pairs. On a balanced seesaw, each child experiences two torques that *happen* to be equal but oppositely directed: a torque due to gravity and a torque due to the other child. Those two torques form a matched pair because the seesaw is balanced, not because of Newton's third law. In fact, if the seesaw weren't balanced, those two torques would not be equal but oppositely directed and the children would both be undergoing angular acceleration.

COMMON MISCONCEPTIONS: Newton's Third Law or Not?

Misconception: Every pair of equal but oppositely directed forces or torques is associated with Newton's third law.

Resolution: Newton's third law applies only to pairs of forces or torques that two objects exert on one another. In such cases only, the forces or torques must be equal but oppositely directed. Two forces or torques exerted on the same object are never a Newton's third law pair and can have any values, including equal but oppositely directed.

SUMMARY OF NEWTON'S LAWS OF ROTATIONAL MOTION

1. A rigid object that is not wobbling and is not subject to any outside torques rotates at a constant angular velocity, turning equal amounts in equal times about a fixed axis of rotation.
2. The net torque exerted on a rigid object that is not wobbling is equal to that object's rotational mass times its angular acceleration. The angular acceleration points in the same direction as the torque.
3. For every torque that one object exerts on a second object, there is an equal but oppositely directed torque that the second object exerts on the first object.

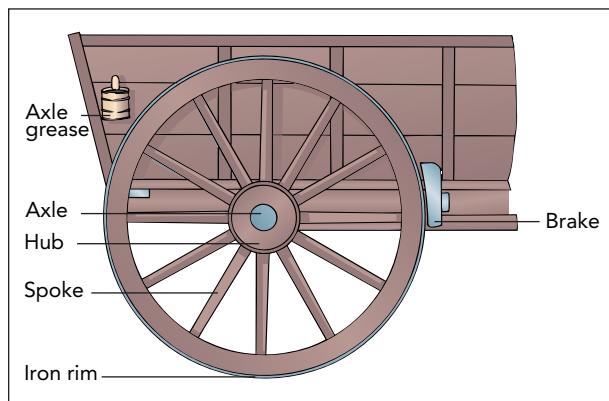
Note: These laws are the rotational analogs of the translational laws in Section 1.3.

Check Your Understanding #8: Turning on Ice

A light at your skating rink needs replacing, so you stand on the slippery ice and reach up overhead to unscrew the light. As you twist the light, you begin to rotate. What caused you to rotate?

Answer: The light exerted a torque on you.

Why: To unscrew the light, you must exert a torque on it. In accordance with Newton's third law of rotational motion, that light exerts an equal but opposite torque on you. The ice is too slippery to exert a torque on you, so you undergo angular acceleration and begin to rotate.

SECTION 2.2**Wheels**

Like ramps and levers, wheels are simple tools that make our lives easier. But a wheel's main purpose isn't mechanical advantage, it's overcoming friction. Up until now, we've ignored friction, looking at the laws of motion as they apply only in idealized situations. However, our real world does have friction, and most objects in motion tend to slow down and stop because of it. One of our first tasks in this section will therefore be to understand friction—though, for the time being, we'll continue to neglect air resistance.

Questions to Think About: If objects in motion tend to stay in motion, why is it so hard to drag a heavy box across the floor? If objects should accelerate downhill on a ramp, why won't a plate slide off a slightly tilted table? What makes the

wheels of a cart rotate as you pull the cart forward? How does spinning its wheels propel a car forward?

Experiments to Do: To observe the importance of wheels in eliminating friction, try sliding a book along a flat table. Give the book a push and see how quickly it slows down and stops. Which way is friction pushing on the book? Does the force that friction exerts on the book depend on how fast the book is

moving? Let the book come to a stop. Is friction still pushing on the book when it isn't moving? If you push gently on the stationary book, what force does friction exert on it?

Lay three or four round pencils on a table, parallel to one another and a few inches apart. Rest the book on top of the pencils and give the book a push in the direction that the pencils can roll. Describe how the book now moves. What do you think has caused the difference?

Moving a File Cabinet: Friction

When we imagined moving your friend's piano into a new apartment back in Section 1.3, we neglected a familiar force—friction. Luckily for us, your friend's piano had wheels on its legs, and wheels facilitate motion by reducing the effects of friction. We'll focus on wheels in this section. First, though, to help us understand the relationship between wheels and friction, we'll look at another item that needs to be moved—your friend's file cabinet.

The file cabinet is resting on a smooth and level hardwood floor; it's full of sheet music and weighs about 1000 N (225 lbf). Despite its large mass, you know that it should accelerate in response to a horizontal force, so you give it a gentle push toward the door. Nothing happens. Of course, the file cabinet accelerates in response to the net force it experiences, not to each individual force acting on it. Something else must be pushing on the file cabinet in just the right way to cancel your force and keep it from accelerating. Undaunted, you push harder and harder until finally, with a tremendous shove, you manage to get the file cabinet sliding across the floor. However, the cabinet moves slowly even though you continue to push on it. Something else is pushing on the file cabinet, trying to stop it from moving.

That something else is **friction**, a phenomenon that opposes the relative motion of two surfaces in contact with one another. Two surfaces that are in **relative motion** are traveling with different velocities so that a person standing still on one surface will observe the other surface as moving. In opposing relative motion, friction exerts forces on both surfaces in directions that tend to bring them to a single velocity.

For example, when the file cabinet slides by itself toward the left, the floor exerts a rightward frictional force on it (Fig. 2.2.1). The frictional force exerted on the file cabinet, *toward the right*, is in the direction opposite the file cabinet's velocity, *toward the left*. Since the file cabinet's acceleration is in the direction opposite its velocity, the file cabinet slows down and eventually comes to a stop.

According to Newton's third law of motion, an equal but oppositely directed force must be exerted by the file cabinet on the floor. Sure enough, the file cabinet does exert a leftward frictional force on the floor. However, the floor is rigidly attached to Earth, so it accelerates very little. The file cabinet does almost all the accelerating, and soon the two objects are traveling at the same velocity.

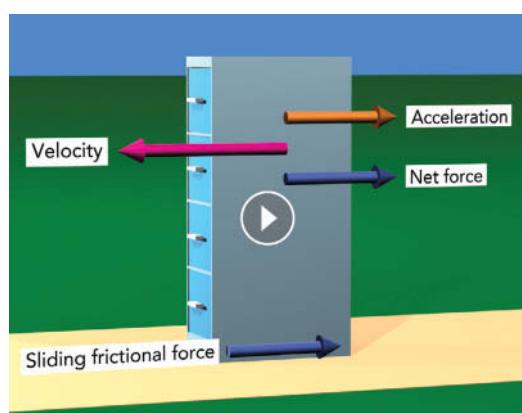


Fig. 2.2.1 A file cabinet sliding to the left across a sidewalk experiences a frictional force from the sidewalk to the right. Since that sliding frictional force is the net force on the file cabinet, the cabinet accelerates opposite its velocity and gradually slows to a stop.

Frictional forces always oppose relative motion, but they vary in strength according to (1) how tightly the two surfaces are pressed against one another, (2) how slippery the surfaces are, and (3) whether the surfaces are actually moving relative to one another. First, the harder you press two surfaces together, the larger the frictional forces they experience. For example, an empty file cabinet slides more easily than a full one. Second, roughening the surfaces generally increases friction, while smoothing or lubricating them generally reduces it. Riding a toboggan down the driveway is much more interesting when the driveway is covered with snow or ice than when the driveway is bare asphalt. We'll examine the third issue later on.

Check Your Understanding #1: The One That Got Away

Your table at the restaurant isn't level, and your water glass begins to slide slowly downhill toward the edge. Which way is friction exerting a force on it?

Answer: Friction is pushing the glass uphill.

Why: The glass is sliding downhill across the top of the stationary table. Since friction always opposes relative motion, it pushes the glass uphill, in the direction opposite its motion.

A Microscopic View of Friction

As the file cabinet slides by itself across the floor, it experiences a horizontal frictional force that gradually brings it to a stop. From where does this frictional force come? The obvious forces on the file cabinet are both vertical, not horizontal; the cabinet's weight is downward, and the support force from the floor is upward. How can the floor exert a horizontal force on the file cabinet?

The answer lies in the fact that neither the bottom of the file cabinet nor the top of the floor is perfectly smooth. They both have microscopic hills and valleys of various sizes. The file cabinet is actually supported by thousands of tiny contact points, where the file cabinet directly touches the floor (Fig. 2.2.2). As the file cabinet slides, the microscopic projections on the bottom of the file cabinet pass through similar projections on the top of the floor. Each time two projections collide, they experience horizontal forces. These tiny forces oppose the relative motion and give rise to the overall frictional forces experienced by the file cabinet and floor. Because even an apparently smooth surface still has some microscopic surface structure, all surfaces experience friction as they rub across one another.

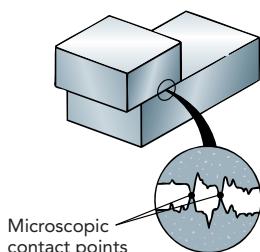


Fig. 2.2.2 Two surfaces that are pressed against one another actually touch only at specific contact points. When the surfaces slide across one another, these contact points collide, producing sliding friction and wear.

Increasing the size or number of these microscopic projections by roughening the surfaces generally leads to more friction. If you put sandpaper on the bottom of the file cabinet, it would experience larger frictional forces as it slides across the floor. On the other hand, a microscopically smoother “nonstick” surface, like that used in modern cookware, would let the file cabinet slide more easily.

Increasing the number of contact points by squeezing the two surfaces more tightly together also leads to more friction. The microscopic projections simply collide more often. That's why adding more sheet music to the file cabinet would make it harder to slide. Doubling the file cabinet's weight would roughly double the number of contact points and make it about twice as hard to move across the floor. A useful rule of thumb is that the frictional forces between two firm surfaces are proportional to the forces pressing those two surfaces together.

Friction also causes wear when the colliding contact points break one another off. With time, this wear can remove large amounts of material so that even seemingly indestructible stone steps are gradually worn away by foot traffic. The best way to reduce wear between two surfaces (other than to insert a lubricant between them) is to polish them so that they are extremely smooth. The smooth surfaces will still touch at contact points and experience friction as they slide across one another, but their contact points will be broad and round and will rarely break one another off during a collision.

Check Your Understanding #2: Weight and Friction

How much harder is it to slide a stack of two identical books across a table than it is to slide just one of those books?

Answer: About twice as hard.

Why: The frictional forces between the table and books are roughly proportional to the forces pressing them together. The book's weight is what pushes them together. When you effectively double the book's weight, by stacking a second book on top of it, you double the frictional forces between the table and the books.

Static Friction, Sliding Friction, and Traction

There are really two kinds of friction: sliding and static. When two surfaces are moving across one another, **sliding friction** acts to stop them from sliding. But even when those surfaces have the same velocity, **static friction** may act to keep them from starting to slide across one another in the first place.

You find it particularly hard to start the file cabinet sliding across the floor. Contact points between the cabinet and floor have settled into one another, so a small push does nothing. Static friction always exerts a frictional force that exactly balances your push. Since the net force on the file cabinet is zero, it doesn't accelerate.

However, the force that static friction can exert is limited. To get the file cabinet moving, you need to give it a mighty shove and thereby exert more horizontal force on it than static friction can exert in the other direction. The net force on the file cabinet is then no longer zero, and the file cabinet accelerates.

Once the file cabinet is moving, static friction is replaced by sliding friction. Because sliding friction acts to bring the file cabinet back to rest, you must push on the cabinet to keep it moving. With the file cabinet sliding across the floor, however, the contact points between the surfaces no longer have time to settle into one another, and they consequently experience weaker horizontal forces. That's why the force of sliding friction is generally weaker than that of static friction and why it's easier to keep the file cabinet moving than it is to get it started.

Both forms of friction are incorporated in the concept of **traction**, the largest amount of frictional force that the file cabinet can obtain from the floor at any given moment. When the cabinet is stationary, its traction is equal to the maximum amount of force that static friction can exert on it. Once it begins to slide across the floor, however, its traction reduces to the amount of force that sliding friction exerts.

While the file cabinet's traction is a nuisance that you must overcome, the traction of your shoes on the floor is crucial. Unless you can push against the wall, your shoes are going to need enough traction to provide the horizontal force required to move the file cabinet. Let's hope you're wearing your basketball shoes!

There is another difference between static and sliding friction—sliding friction wastes energy. It converts useful, **ordered energy**, work or energy that can easily do work, into relatively useless **thermal energy**, a disordered energy that's associated with temperature. Sliding friction makes things hotter by turning work into thermal energy.

We can observe work becoming thermal energy as you slide the file cabinet across the floor at constant velocity. Since the cabinet isn't accelerating, you and the floor are pushing the cabinet equally hard in opposite directions. Because you're pushing the cabinet forward as it moves forward, you're doing (positive) work on the cabinet. Sliding friction from the floor, however, is pushing the cabinet backward as it moves forward, so sliding friction is doing negative work on the cabinet. Overall, sliding friction is removing energy from the cabinet just as fast as you're adding it.

Where is your work going? The cabinet isn't transferring your work to the floor because the floor can't move and therefore can't have work done on it. Alas, sliding friction

is grinding your work up into tiny fragments by way of countless collisions between the contact points on the two sliding surfaces. The fragments of your energy are distributed randomly among particles that make up those surfaces and thereby become thermal energy. The cabinet and floor are both becoming hotter.

In contrast, static friction doesn't waste energy as thermal energy; it simply enables objects to do work on one another. For example, your efforts have made you thirsty, so you drink a glass of water. As you grip the sides of the glass with your fingers and lift it upward, each finger is exerting an upward static friction force on the glass. Since the glass moves upward as your fingers push it upward, your fingers are doing (positive) work on the glass. At the same time, the glass is exerting a downward static friction force on each of your fingers. Because each finger moves upward as the glass pushes it downward, the glass is doing negative work on your fingers. In this manner, static friction is transferring energy from you to the glass without wasting any of it as thermal energy.

Check Your Understanding #3: Skidding to a Stop

Antilock brakes keep an automobile's wheels from locking and skidding during a sudden stop. Apart from issues of steering, what is the advantage of preventing the wheels from skidding (sliding) on the pavement?

Answer: If the wheels continue to turn, they experience static friction with the pavement. If they lock and begin to skid, they experience sliding friction. Since the traction provided by static friction is greater than that provided by sliding friction, the car will decelerate faster if the wheels don't skid.

Why: For a rapid stop, the car needs the maximum possible force in the direction opposite its velocity. The most effective way to obtain that stopping force from the road is with static friction between the turning wheels and the pavement. Sliding friction, the result of skidding tires, is less effective at stopping the car, wears out the tires, and diminishes the driver's ability to steer the vehicle.

Friction and Energy

We've seen that sliding friction converts work into thermal energy, but what exactly is thermal energy? To answer that question, let's first reexamine energy in general.

Energy is the capacity to do work, and it is transferred between objects by doing that work. As noted in Section 1.3, energy is a conserved physical quantity—it cannot be created or destroyed; it can only be transferred between objects or converted from one form to another. To distinguish energy from two other conserved quantities we'll encounter in the next section, you can think of energy as the *conserved quantity of doing*. With practice, you'll be able to follow the flow of energy through a system just as an accountant follows the flow of money through a company.

Energy's two basic forms are kinetic (energy contained in the motions of objects) and potential (energy stored in the forces between or within those objects). Each force has its own specific form of potential energy, some of which appear in Table 2.2.1.

The individual forms of energy are *not* conserved quantities; only *total* energy—the sum of all the individual forms—is conserved. For example, when you drop a ball from

TABLE 2.2.1 Several Specific Forms and Examples of Potential Energy

Form of Potential Energy	Example
Gravitational potential energy	A bowling ball at the top of a hill
Elastic potential energy	A wound clock spring
Electrostatic potential energy	A cloud in a thunderstorm
Chemical potential energy	A firecracker
Nuclear potential energy	Uranium

rest, its gravitational potential energy decreases and its kinetic energy increases, but its total energy remains unchanged.

We measure energy in many common units: joules (J), calories, food Calories (also called kilocalories), and kilowatt-hours, to name only a few. All these units measure the same thing, and they differ from one another only by numerical conversion factors, some of which can be found in Appendix B. For example, 1 food Calorie is equal to 1000 calories or 4187 J. A jelly donut with about 250 food Calories thus contains about 1,000,000 J of energy. Since a joule is the same as a newton-meter, 1,000,000 J is the energy you'd use to lift your friend's file cabinet into the second-floor apartment 200 times (1000 N times 5 m upward equals 5000 J of work per trip). No wonder eating donuts is hard on your physique!

Let's return now to thermal energy. Thermal energy is actually a mixture of ordinary kinetic and potential energies. But unlike the kinetic energy in a moving ball or the potential energy in an elevated piano, the kinetic and potential energies in thermal energy are contained entirely within those objects. Any object has **internal energy**, energy held entirely within that object by its individual particles and forces between those particles; *thermal energy* is the portion of internal energy that's associated with temperature. Thermal energy makes every microscopic particle in the object jiggle randomly and independently; at any moment, each particle has its own tiny supply of potential and kinetic energies, and this dispersed and disordered energy is collectively referred to as thermal energy.

As you push the file cabinet across the floor, you do work on it, but it doesn't pick up speed. Instead, sliding friction converts your work into thermal energy, and the cabinet and floor become hotter as the energy you transfer to them disperses among their particles. But although sliding friction easily turns work into thermal energy, there's no easy way to turn thermal energy back into work. Disorder not only makes things harder to use, but it also is difficult to undo. When you drop your favorite coffee mug on the floor and it shatters into a thousand pieces, the cup is still all there; however, it's disordered and thus much less useful. Just as dropping the pieces on the floor a second time isn't likely to reassemble your cup, energy converted into thermal energy can't easily be reassembled into useful, ordered energy.

Check Your Understanding #4: Burning Rubber

If you push too hard on your car's accelerator pedal when the traffic light turns green, your wheels will slip and you'll leave a black trail of rubber behind. Such a "jackrabbit start" can cause as much wear on your tires as 50 km (31 miles) of normal driving. Why is skidding so much more damaging to the tires than normal driving?

Answer: Normal driving involves mostly static friction because the surfaces of the tires don't slide across the pavement. Skidding involves sliding friction as the tire surfaces move independently of the pavement. Because it involves sliding friction, skidding creates thermal energy and damages the tires.

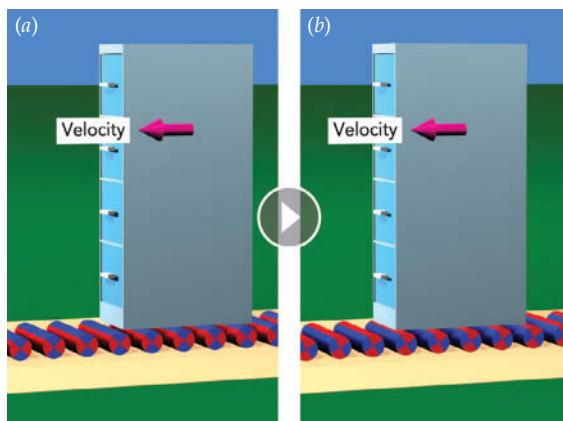
Why: The expression "burning rubber" is an appropriate name for skidding during a jackrabbit start. Substantial thermal energy is produced, and a trail of hot rubber is left on the pavement behind the car. At drag races, the frictional heating that results from skidding at the start can be so severe that the tires actually catch on fire.

Wheels

You've wrestled your friend's file cabinet out the door of the old apartment and are now dragging it along the sidewalk. You're doing work against sliding friction the whole way, producing large amounts of thermal energy in both the bottom of the cabinet and the surface of the sidewalk. You're also damaging both objects, since sliding friction is wearing out their surfaces. The four-drawer file cabinet may be down to three drawers by the time you arrive at the new apartment.

Fortunately, there are mechanical systems that can help you move one object across another without sliding. The classic example is a roller (Fig. 2.2.3). If you place the file cabinet on rollers, those rollers will rotate as the file cabinet moves so that their surfaces

Fig. 2.2.3 (a) A file cabinet that's supported on turning rollers experiences no sliding friction. (b) Since the top surface of a roller moves forward with the file cabinet, while its bottom surface stays behind with the sidewalk, the roller moves only half as fast as the file cabinet. As a result, the rollers are soon left behind.



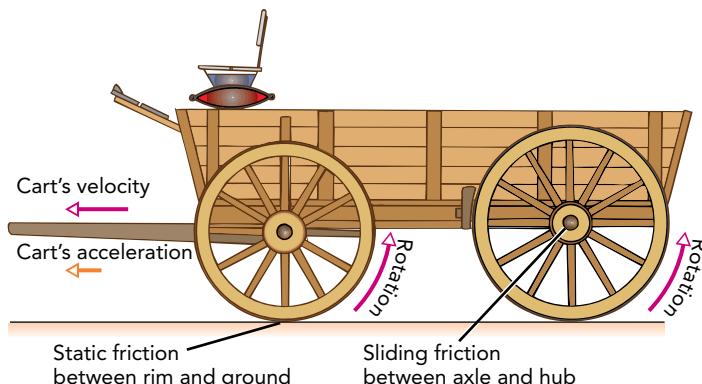
never slide across the bottom of the cabinet or the top of the sidewalk. To see how the rollers work, make a fist with one hand and roll it across the palm of your other hand. The skin of one hand doesn't slide across the skin of the other hand; since this silent motion doesn't convert work into thermal energy, your skin remains cool. Now slide your two open palms across one another; this time, sliding friction warms your skin.

Although the rollers don't experience sliding friction, they can experience static friction. The top of each roller is touching the bottom of the cabinet, and the two surfaces move along together even when the file cabinet is accelerating because of static friction; they grip one another tightly until the roller's rotation pulls them apart. A similar process takes place between the rollers and the top of the sidewalk; static friction can exert torques on the rollers and cause them to rotate in the first place. Again, you can illustrate this behavior with your hands. Try to drag your fist across your open palm. Just before your fist begins to slide, you'll feel a torque on it. Static friction between the skins of your two hands, acting to prevent sliding, causes your fist to begin rotating just like a roller.

Once you get the file cabinet moving on rollers, you can keep it rolling along the level sidewalk indefinitely. Without any sliding friction, the cabinet doesn't lose kinetic energy, so it continues at constant velocity without your having to push it. However, the rollers move out from under the file cabinet as it travels, and you frequently have to move a roller from the back of the cabinet to the front. In fact, you need at least three rollers to ensure that the file cabinet never falls to the ground when a roller pops out the back. Although the rollers have eliminated sliding friction, they've created another headache—one that makes the prospect of a long trip unappealing. Is there another device that can reduce sliding friction without requiring constant attention?

One alternative would be a four-wheeled cart. The simplest cart rests on fixed poles or axles that pass through central holes or hubs in the four wheels (Fig. 2.2.4). The ground exerts upward support forces on the wheels, the wheels exert upward support forces on the axles, and the axles support the cart and its contents. As the cart moves forward, its wheels

Fig. 2.2.4 As this cart accelerates toward the left, its wheels roll counterclockwise. Although the wheel rims experience only static friction with the ground, the wheel hubs slide around the axles and convert the cart's kinetic energy into thermal energy. To reduce this wasted energy, the cart has narrow axles that are lubricated with axle grease.



roll. Their bottom surfaces don't slide or skid across the ground; instead, each wheel lowers a portion of its surface onto the sidewalk, leaves it there briefly so that it may experience static friction, and then raises it back off the sidewalk, with a new portion of wheel surface taking its place. Because of the touch-and-release character of rolling, there is no sliding friction between the cart's wheels and the ground.

Unfortunately, as each wheel rotates, its hub slides across the stationary axle at its center. This sliding friction wastes energy and causes wear to both hub and axle. However, the narrow hub moves relatively slowly across the axle so that the work and wear done each second are small. Still, this sliding friction is undesirable and can be reduced significantly by lubricating the hub and axle with axle grease.

A better solution is to insert rollers between the hub and axle (Fig. 2.2.5). The result is a roller bearing, a mechanical device that eliminates sliding friction between a hub and an axle. A complete bearing consists of two rings separated by rollers that keep those rings from rubbing against one another. In this case, the bearing's inner ring is attached to the stationary axle, while its outer ring is attached to the spinning wheel hub. The nondriven wheels of an automobile are supported by roller bearings on essentially stationary axles. The non-driven wheels of lighter vehicles, such as bicycles and wagons, are similarly supported on stationary axles, but their bearings use balls instead of rollers—ball bearings. When a vehicle with free wheels starts forward, static friction from the ground pushes backward on the bottoms of those free wheels to keep them from skidding forward. Those frictional forces also produce torques that cause the wheels to begin turning and they roll forward (Fig. 2.2.6a).

A car's engine-powered or driven wheels are also supported by roller bearings, but these bearings act somewhat differently. Because the engine must be able to exert a torque on each driven wheel, those wheels are rigidly connected to their axles. As the engine spins these axles, the axles spin their wheels (Fig. 2.2.6b). A bearing prevents each spinning axle from rubbing against the car's frame. This bearing's outer ring is attached to the stationary car frame while its inner ring is attached to the spinning axle. When a vehicle begins to spin its driven wheels, static friction from the ground pushes forward on the bottoms of those driven wheels to keep them from skidding backward. Since those forward frictional forces are the only horizontal forces on the vehicle, the vehicle accelerates forward and the driven wheels roll forward.

Recognizing a good idea when you think of it, you load the file cabinet into the passenger seat of your red convertible sports car and start the engine. The car isn't quite as responsive as usual because of the added mass, but it's still able to accelerate respectably. In a few seconds, you're cruising down the road toward the new apartment and a very grateful friend.

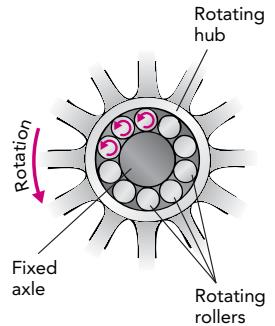


Fig. 2.2.5 In a roller bearing, the hub of the wheel doesn't touch the axle directly. Instead, the two are separated by a set of rollers that turn with the hub. The bottom few rollers bear most of the load since the hub pushes up on them and they push up on the axle. As the wheel turns, the rollers recirculate, traveling up to the right and over the top of the axle before returning down to the left to bear the load once again. The rollers, wheel, and axle can experience only static friction, not sliding friction. In a ball bearing, the cylindrical rollers are replaced by spherical balls.

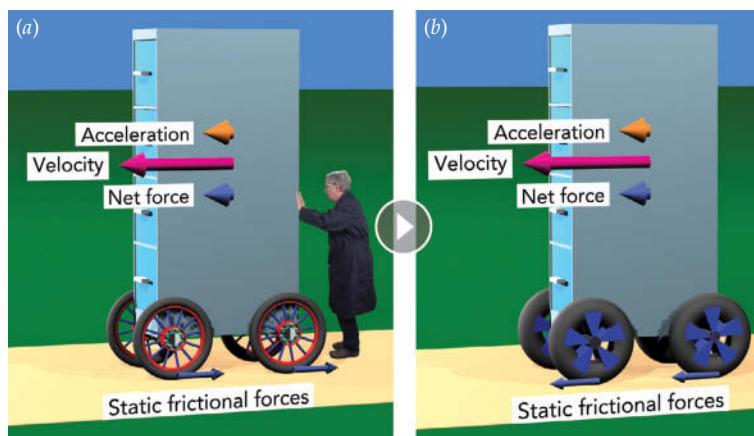


Fig. 2.2.6 (a) When my assistant pushes a cart forward and it accelerates forward, static friction from the ground pushes the wheel bottoms backward to prevent them from skidding. That same frictional force causes the wheels to roll forward. (b) When the engine of a go-cart twists its wheels forward, static friction from the ground pushes the wheel bottoms forward to prevent them from skidding. That same frictional force causes the vehicle to accelerate forward.


Check Your Understanding #5: Jewel Movements

Many antique mechanical watches and clocks proudly proclaim that they have “jewel movements.” Gears in these timepieces turn on axles that are pointed at either end and are supported at those ends by very hard, polished gemstones. What is the advantage of having needlelike ends on an axle and supporting those needles with smooth, hard jewels?

Answer: Because all the supporting forces are very close to the axis of rotation, the jewels exert almost zero torque on the axle. The axle turns remarkably freely.

Why: Mechanical timepieces need almost ideal motion to keep accurate time. One of the best ways to allow a rotating object free movement is to support it exactly on the axis of rotation, where the support can’t exert torque on the object.

Power and Rotational Work

It’s a beautiful afternoon, and you thrill to the power of your car as it drives swiftly up a steep hill. In this situation, power means more than just a sense of exhilaration and freedom; **power** is also the physical quantity that measures the rate at which your car’s engine does work—the amount of work it does in a certain amount of time, or

$$\text{power} = \frac{\text{work}}{\text{time}}.$$

The SI unit of power is the **joule per second**, also called the **watt** (abbreviated W). Other units of power include Calories per hour and horsepower; like the units for energy, these units differ only by numerical factors, which are listed in Appendix B online. For example, 1 horsepower is equal to 745.7 W, so the engine of your 200-horsepower roadster can supply about 150,000 W of power. With you and the file cabinet on board, the car weighs about 15,000 N (3400 lbf), so the car engine can lift it about 10 m upward each second. No wonder it’s climbing the hill so easily!

Up to now, we’ve considered work only in the context of *translation* motion, where work involves force and distance. For example, you do work on the file cabinet by exerting a force on it as it moves a distance in the direction of that force. The car’s engine, however, does work on each of its powered wheels in the context of *rotational* motion, where work involves torque and angle. Specifically, the engine does work on a wheel by exerting a torque on the wheel as the wheel rotates through an angle in the direction of the torque.

To do work on an object via rotational motion, you must twist it as it rotates through an angle in the direction of your twist. The work you do on it is equal to the torque you exert on its times the angle through which it rotates in the direction of your torque, where that angle is measured in radians, the natural unit of angles. We can express this relationship as a word equation:

$$\text{work} = \text{torque} \cdot \text{angle(in radians)} \quad (2.2.1)$$

in symbols:

$$W = \tau \cdot \theta,$$

and in everyday language:

If you’re not twisting or it’s not turning, then you’re not working.

This simple relationship assumes that your torque is constant while you’re doing the work. If your torque varies, the calculation of work will have to recognize that variation and may require the use of calculus.

As the engine does work on its wheels via rotational motion, those wheels do work on the car via translational motion. Because each wheel grips the pavement with static friction, its contact point with the ground can't move. Instead, the rotating wheel pushes the axle and car forward as they move forward. So energy that flows into the wheels via rotational motion returns to the car via translational motion and keeps the car climbing steadily up the hill.

Like work itself, power can be supplied by translational motion or rotational motion. When you're pushing the file cabinet across the floor, you're using translational motion to supply power to it. Since translational work is force times distance (Eq. 1.3.1) and power is work divided by time, translational power is force times distance divided by time, or

$$\text{power} = \text{force} \cdot \text{velocity}.$$

You can thus supply more power to the file cabinet either by pushing it harder or by having it move faster in the direction of your push.

As your car drives up the hill, the car engine is using rotational motion to supply power to its wheels. Since rotational work is torque times angle, rotational power is torque times angle divided by time, or

$$\text{power} = \text{torque} \cdot \text{angular velocity}.$$

The car can therefore supply more power to its wheels either by twisting them harder or by having them rotate faster in the direction of that twist.

Check Your Understanding #6: Going Nowhere One Foot at a Time

You're pedaling a stationary bicycle when your shoelace comes undone. While you retie the lace, you can pedal with only one foot and exert half as much torque on the pedals as before. The pedals are still turning just as quickly, so are you still exercising as hard?

Answer: No.

Why: By halving the torque you exert on the pedals, you halve the work that you do on those pedals each time they complete a rotation.

Check Your Figures #1: We All Knead Energy

When your kitchen mixer is kneading a loaf of sourdough bread, it exerts a torque of 20 newton-meters on its mixing blade, and that blade completes 500 rotations. How much energy does the mixer supply to its blade and the dough?

Answer: It supplies about 63,000 joules.

Why: According to Eq. 2.2.1, the energy that the mixer transfers to its blade is equal to the torque it exerts on the blade times the angle through which the blade turns. Since there are 2π radians per complete rotation, the blade turns approximately 3140 radians. The work that the mixer does is 20 N·m times 3140, or about 63,000 J.

Kinetic Energy

As you near your destination, you begin thinking about the car's brakes. They're designed to stop the car by turning its kinetic energy into thermal energy. They'll perform their task by rubbing stationary brake pads against spinning metal discs. Although you're confident that those brakes are up to the task, just how much kinetic energy are they going to have to convert into thermal energy?

One way to determine the car's kinetic energy is to calculate the work its engine did on it while bringing it from rest to its current speed. The result of that calculation is that the

moving car's kinetic energy is equal to one-half its mass times the square of its speed. This relationship can be written as a word equation:

$$\text{kinetic energy} = \frac{1}{2} \cdot \text{mass} \cdot \text{speed}^2, \quad (2.2.2)$$

in symbols:

$$K = \frac{1}{2} \cdot m \cdot v^2,$$

and in everyday language:

Racing around at twice the speed takes four times the energy.

With you and the file cabinet on board, the sports car has a mass of about 1500 kg (3300 lbm). At a speed of 100 km/h (62 mph), it has over 575,000 J of kinetic energy. That enormous energy is four times what it would be at 50 km/h (31 mph), so put down your cell phone and drive carefully. The dramatic increase in kinetic energy that results from a modest increase in speed explains why high-speed crashes are far deadlier than those at lower speeds and why that police officer is checking out your car with a radar gun. Red cars get all the attention.

You're traveling safely within the speed limit and exchange a polite wave with the officer. However, you soon pass another car that has been stopped for a ticket. The light on the nearby police car spins round and round, and rotating objects have kinetic energy, too. Like the kinetic energy of translational motion, the kinetic energy of rotational motion depends on inertia and speed; however, it's the rotational inertia and rotational speed that matter. The light's kinetic energy of rotational motion is equal to one-half its rotational mass times the square of its angular speed. This relationship can be written as a word equation:

$$\text{kinetic energy} = \frac{1}{2} \cdot \text{rotational mass} \cdot \text{angular speed}^2, \quad (2.2.3)$$

in symbols:

$$K = \frac{1}{2} \cdot I \cdot \omega^2,$$

and in everyday language:

It takes a very energetic person to spin his wheels twice as fast.

With the ticket complete, the police car pulls out into traffic with its light still spinning. The light's total kinetic energy is now the sum of two parts: translational kinetic energy and rotational kinetic energy. Its translational kinetic energy depends on the speed of the light's center of mass, which is equal to the police car's speed through traffic. Its rotational kinetic energy depends on the angular speed at which the light turns about its center of mass.

As the police car disappears in the distance, it occurs to you that the spinning wheels of your car also have rotational kinetic energy that adds to the car's substantial translational kinetic energy. Still, you trust your brakes. In a few minutes, you arrive at your destination and brake to a stop. Although you're aware of the added mass because the car decelerates less quickly than usual, the brakes successfully transform the car's kinetic energy into thermal energy. You've reached your goal safely and are now a hero.

Check Your Understanding #7: Throwing a Fastball

A typical grade school pitcher can throw a baseball at 80 km/h (50 mph), but only a few professional athletes have the extraordinary strength needed to throw a baseball at twice that speed. Why is it so much harder to throw the baseball only twice as fast?

Answer: Doubling the speed of the baseball requires quadrupling the energy transferred to it by the pitcher.

Why: To throw a 160-km/h (100-mph) fastball, a major league pitcher must put four times as much kinetic energy into both the ball and his arm as when pitching an 80-km/h slow ball. He also pitches the fastball in half the time needed to pitch the slow ball. Overall, he must do four times as much work throwing a fastball, and he must do that work in one-half the time. That means the pitcher produces eight times as much power while throwing a fastball as when throwing a slow ball. No wonder amateurs have trouble duplicating that feat.

Check Your Figures #2: Blowing in the Wind

The air in a hurricane travels at 200 km/h (124 mph). How much more kinetic energy does 1 kg of this air have than 1 kg of air moving at only 20 km/h?

Answer: It has 100 times as much kinetic energy.

Why: Because kinetic energy is proportional to speed squared, the kilogram of air in the hurricane moves 10 times as fast but has 100 times as much kinetic energy as the slower moving air. This enormous increase in energy is what makes a hurricane's wind dangerous. The air's terrific speed also brings large quantities of it to you quickly, so that the wind power arriving each second is overwhelming.

Check Your Figures #3: Playing Around at the Playground

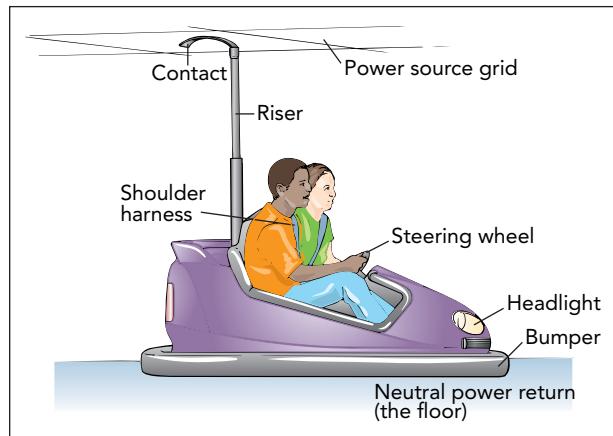
When children climb onto a playground merry-go-round, they increase its rotational inertia. If the children triple the merry-go-round's rotational mass, how will they alter the kinetic energy it has when it spins at a certain angular speed?

Answer: The children will triple the merry-go-round's kinetic energy.

Why: The kinetic energy of a spinning object is proportional to its rotational mass. Since the children triple the merry-go-round's rotational mass, they triple its kinetic energy.

SECTION 2.3

Bumper Cars



While car crashes normally aren't much fun outside of movies or television, there is one delightful exception—bumper cars.

For a few minutes, drivers in this amusement park ride race madly about an oval track, deliberately crashing their vehicles into one another and laughing hysterically at the violent impacts. Jolts, jerks, and spins are half the fun, and it's a wonder that no one gets whiplash. Hidden in the excitement, though, are several important physics concepts that influence everything from tennis to billiards.

Questions to Think About: Why does a stationary car begin rolling forward after being struck by a moving car? What aspects of motion are passed between cars as they collide? Why does your car jolt more when the car that hits it contains two big adults rather than one small child? What would happen if the bumper cars had hard steel bumpers rather than soft rubber ones? Why is your car often set spinning by collisions, and what keeps it spinning?

Experiments to Do: Place a coin on a smooth table and flick a second identical coin so that it slides along the table and

strikes the stationary coin squarely. What happens? Try this experiment again, but now use two coins with different masses. How is the collision different? Does it matter which coin you crash into the other?

Now line up several identical coins so that they touch, and slide another coin into one end of this line. How does the

collision affect the coin that was originally moving? How does it affect the line of coins? What was transferred among the coins by the collision?

Now stand a coin on its edge and flick it so that it spins rapidly. Did you give it something that keeps it spinning? Why does the coin eventually stop spinning?

Coasting Forward: Linear Momentum

Bumper cars are small, electrically powered vehicles that can turn on a dime and are protected on all sides by rubber bumpers. Each car has only two controls: a pedal that activates its motor and a steering wheel that controls the direction in which the motor pushes the car. Since the car itself is so small, its occupants account for much of the car's total mass and rotational mass.

Imagine that you have just sat down in one of these cars and put on your safety strap. The other people also climb into their cars, usually one person per car, and the ride begins.

With your car free to move or turn, you quickly become aware of its translational and rotational inertias. The car's translational inertia makes it hard to start or stop, and its rotational inertia makes it difficult to spin or stop from spinning. While we've seen these two types of inertia before, let's take another look at them and at how they affect your bumper car. This time, we'll see that they're associated with two new conserved quantities: linear momentum and angular momentum. As promised, energy isn't the only conserved quantity in nature!

When fast-moving bumper cars crash into one another, they exchange more than just energy. Energy is a scalar rather than vector quantity, so it's directionless and can't account for the fact that these cars seem to be exchanging some physical quantity that incorporates both speed and direction of travel. For example, if your car is hit squarely by a car speeding toward the right, then your car's motion shifts rightward in response. What your car is receiving from the other car is a rightward-directed dose of a conserved vector quantity known as linear momentum.

Linear momentum, usually just called **momentum**, is the measure of an object's translational impetus or tenacity—its determination to keep moving the way it's currently moving. Roughly speaking, your car's momentum indicates the effort it took to get the car moving with its present speed and direction of motion. To distinguish it from energy and angular momentum, you can think of momentum as the *conserved quantity of moving*.

The car's momentum is equal to its mass times its velocity and can be written as a word equation:

$$\text{momentum} = \text{mass} \cdot \text{velocity}, \quad (2.3.1)$$

in symbols:

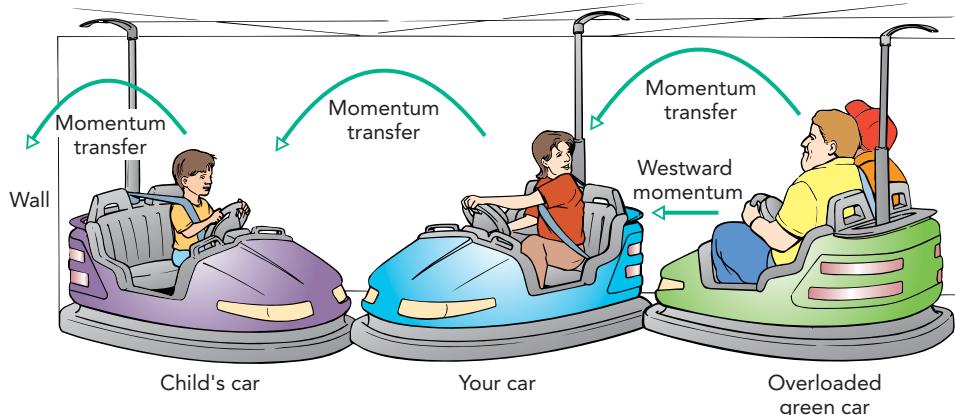
$$\mathbf{p} = m \cdot \mathbf{v},$$

and in everyday language:

It's hard to stop a fast-moving truck.

Note that momentum is a vector quantity and that it has the same direction as the velocity. As we might expect, the faster your car is moving or the more mass it has, the more momentum it has in the direction of its velocity. The SI unit of momentum is the **kilogram-meter per second** (abbreviated $\text{kg} \cdot \text{m/s}$).

To physicists, conserved quantities are rare treasures that make it easier to understand otherwise complicated motions. Like all conserved quantities, momentum can't be created or destroyed. It can only be transferred between objects. Momentum plays a very basic role in bumper cars; the whole point of crashing them into one another is to enjoy the momentum



transfers. During each collision, momentum shifts from one car to the other so that they abruptly change their speeds or directions or both. As long as these momentum transfers aren't too jarring, everyone has a good time.

You've stopped your car, so it has zero velocity and zero momentum. To begin moving again, something must transfer momentum to your car. While you could press the pedal and let the motor gradually transfer momentum from the ground to your car, that's not much fun. Instead, you let two grinning couch potatoes in an overloaded green car slam into you at breakneck speed (Fig. 2.3.1).

The green car was heading westward, and in a few moments your car is moving westward, too, while the green car has slowed significantly. Before you recover from the jolt, your car pounds a child's car westward and your car slows down abruptly. Finally, its impact with a wall stops the child's car. Despite disapproving looks from the child's parents, there's no harm done. Overall, westward momentum has flowed from the spudmobile to your car, to the child's car, and into the wall. No momentum has been created or destroyed; you've all simply enjoyed passing it along from car to car.

Fig. 2.3.1 Your car is hit by a fast-moving, massive car with westward momentum. Much of that westward momentum is transferred to your car. You crash into a child's car, transferring the westward momentum to it. It then crashes into the wall, transferring the westward momentum to the wall. Although that wall doesn't appear to move, it and the Earth, to which it's attached, actually accelerate westward a tiny amount as they receive the westward momentum.

Check Your Understanding #1: Stuck on the Ice

Suppose you're stuck in the middle of a frozen lake, with a surface so slippery that you can't get any traction. You take off a shoe and throw it toward the southern shore. You find yourself coasting toward the northern shore and soon escape from the lake. Why did this scheme work?

Answer: By transferring southward momentum to the shoe, you obtained northward momentum.

Why: Initially, both you and your shoe have zero momentum. But when you throw the shoe southward, you give it southward momentum. Since the only source of that southward momentum is you, you must have lost southward momentum. A negative amount of southward momentum is actually northward momentum, and thus you coast northward. Interestingly enough, the total momentum of you and the shoe hasn't changed. It's still zero, as it must be because momentum is conserved. It has simply been redistributed.

Check Your Figures #1: Follow That Train!

The bad guys are getting away in a four-car train, and you're trying to catch them. The train has a mass of 20,000 kg and it's rolling forward at 22 m/s (80 km/h or 50 mph). What is the train's momentum?

Answer: The train's momentum is 4400,000 kg · m/s, in the forward direction.

Why: You can use Eq. 2.3.1 to calculate the train's momentum from its mass and velocity:

$$\begin{aligned}\text{linear momentum} &= 20,000 \text{ kg} \cdot 22 \text{ m/s} \\ &= 440,000 \text{ kg} \cdot \text{m/s.}\end{aligned}$$

That momentum is in the same direction the train is moving, the forward direction.

Exchanging Momentum in a Collision: Impulses

Momentum is transferred to a car by giving it an **impulse**, a force exerted on it for a certain amount of time. When the motor and floor push your bumper car forward for a few seconds, they give your car an impulse and transfer momentum to it. This impulse is the change in your car's momentum and is equal to the force exerted on the car times the duration of that force. This relationship can be written as a word equation:

$$\text{impulse} = \text{force} \cdot \text{time}, \quad (2.3.2)$$

in symbols:

$$\Delta\mathbf{p} = \mathbf{F} \cdot t,$$

and in everyday language:

The harder and longer you push a bobsled forward at the start of a race, the more momentum it will have when it starts down the hill.

The more force or the longer that force is exerted, the larger the impulse and the more your bumper car's momentum changes. Remember that an impulse, like momentum itself, is a vector quantity; it points in the same direction as the force. If your aim is off and the misdirected impulse you obtain from the floor sends you crashing into the wall, don't say you hadn't been warned!

Different forces exerted for different amounts of time can transfer the same momentum to a car:

$$\begin{aligned} \text{impulse} &= \text{large force} \cdot \text{short time} \\ &= \text{small force} \cdot \text{long time}. \end{aligned} \quad (2.3.3)$$

Thus you can get your car moving with a certain forward momentum either by letting the motor and floor push it with a small forward force of long duration or by letting the colliding green car push on it with a large forward force of short duration.

Forces that colliding objects exert on one another as they exchange momentum are often called *impact forces*, and they explain why bumper cars have soft rubber bumpers. If the bumpers were hard steel, the collision between the green car and your car would last only an instant and would involve enormous impact forces. You'd be in need of a neck brace and the services of a personal injury lawyer. However, amusement parks don't like lawsuits and sensibly limit the impact forces. To do this, they use rubber bumpers and rather slow-moving cars.

Nonetheless, you can get a pretty good jolt when you collide head-on with another car. Your two cars then start with oppositely directed momenta, and the collision roughly exchanges those momenta between cars. In almost no time, you go from heading forward to heading backward. The impulse that causes this reversal of motion is especially large because it not only stops your forward motion but also causes you to begin heading backward.

During the collision, you gave the other car more forward momentum than you had and ended up with less than zero forward momentum. A vector of negative magnitude is a vector of positive magnitude pointing in the opposite direction. In this case, your deficit of forward momentum after the collision means that you have backward momentum.

Momentum is conserved because of Newton's third law of motion. When the green car exerts a force on your car for a certain amount of time, your car exerts an equal but oppositely directed force on the green car for exactly the same time. Because of the equal but oppositely directed nature of the two forces, the cars receive impulses that are equal in amount but opposite in direction. Since the momentum gained by one car is exactly equal to the momentum lost by the other car, we say that momentum is transferred from one car to the other.

The more mass a car has, the less its velocity changes as a consequence of a momentum transfer. That's why the green car doesn't stop completely when it crashes into your car, whereas your car speeds up dramatically. Just a fraction of the green car's forward momentum causes a large change in your car's velocity. Like a bug being hit by an automobile windshield, your bumper car does most of the accelerating.

CONSERVED QUANTITY: MOMENTUM

TRANSFERRED BY: IMPULSE

Momentum: The measure of an object's translational motion, its tendency to continue moving in a particular direction. Momentum is a vector quantity, meaning that it has a direction. It has no potential form and therefore cannot be hidden; momentum = mass · velocity.

Impulse: The mechanical means for transferring momentum; impulse = force · time.

COMMON MISCONCEPTIONS: Momentum and Force

Misconception: A massive moving object carries a force with it—the “force of its momentum.”

Resolution: Although the impulses that transfer momentum involve forces, momentum itself does not. A moving object carries only momentum, not force. Most important, a coasting object is free of any net force. However, when that object hits an obstacle, the two exchange momentum via impulses, and those impulses do involve forces.

Check Your Understanding #2: Bowling Them Over

When a beanbag hits the wall, it transfers all its forward momentum to the wall and comes to a stop. When a rubber ball hits the wall, it transfers all its forward momentum, comes to a stop, and then rebounds. During the rebound, it transfers still more forward momentum to the wall. If you wanted to knock over a weighted bowling pin at the county fair, which would be the more effective projectile: the rubber ball or the beanbag, assuming they have identical masses and you throw them with identical velocities?

Answer: The bouncy rubber ball would be more effective.

Why: Either projectile will transfer all its original momentum to the bowling pin while coming to a stop. However, the bouncy rubber ball will bounce back and continue to exert a force on the bowling pin. The impulse (force · time) delivered by the rubber ball will be greater than that delivered by the beanbag because the ball will exert its forward force for a longer time (during stopping *and* rebounding). The ball will rebound with its momentum reversed, having transferred roughly twice its original momentum to the pin.

Check Your Figures #2: Stop That Train!

The engine of the train you're chasing (see Check Your Figures #1) has broken down, but it's still rolling forward. To stop it, you grab onto the last car and begin to drag your boot heels on the ground. The backward force on the train is 200 N. How long will it take you to stop the train?

Answer: It will take 2200 s, so kiss your boot heels goodbye!

Why: To stop the train, you must give it a backward impulse that completely cancels its forward momentum. Since its forward momentum is 440,000 kg · m/s, the backward impulse must be 440,000 kg · m/s. Since 200 N can also be written as 200 kg · m/s², we can use Eq. 2.3.2 to find the time:

$$\text{time} = \frac{440,000 \text{ kg} \cdot \text{m/s}}{200 \text{ kg} \cdot \text{m/s}^2}$$

$$= 2200 \text{ s.}$$

Spinning in Circles: Angular Momentum

When bumper cars are sent spinning during crashes, they are exchanging yet another conserved quantity. Like momentum, it's a conserved *vector* quantity, but now it's associated with the angular speed and direction of rotation around a specific pivot. For example, when your car receives a glancing blow from a car that is circling it rapidly in a clockwise direction, your car acquires more clockwise rotation about its center of mass. What your car is receiving from the other car is a clockwise-directed dose of a conserved vector quantity known as angular momentum.

Angular momentum is the measure of an object's rotational impetus or tenacity, its determination to keep rotating the way it's currently rotating. Simply put, your car's angular momentum indicates the effort it took to get the car rotating with its present angular speed and axis of rotation. To distinguish it from energy and momentum, you can think of angular momentum as the *conserved quantity of turning*.

The car's angular momentum is equal to its rotational mass times its angular velocity and can be written as a word equation:

$$\text{angular momentum} = \text{rotational mass} \cdot \text{angular velocity}, \quad (2.3.4)$$

in symbols:

$$\mathbf{L} = I \cdot \boldsymbol{\omega},$$

and in everyday language:

It's hard to stop a spinning carousel.

Note that angular momentum is a vector quantity and that it has the same direction as the angular velocity. The faster your car is spinning or the larger its rotational mass, the more angular momentum it has in the direction of its angular velocity. The SI unit of angular momentum is the **kilogram-meter² per second** (abbreviated kg · m²/s).

Angular momentum is another conserved quantity; it can't be created or destroyed, only transferred between objects. For your car to begin spinning, something must transfer angular momentum to it, and your car will then continue to spin until it transfers this angular momentum elsewhere. To study angular momentum properly, however, we must pick the pivot about which all the spinning will occur. In the present situation, a good choice for this pivot is your car's initial center of mass.

Your bumper car is stationary again, so it has zero angular velocity and zero angular momentum. Suddenly, a purple car sweeps by and strikes your car a glancing blow (Fig. 2.3.2). Because the purple car was circling your car counterclockwise, it had counterclockwise angular momentum about the pivot. Its impact transfers some of this angular momentum to your car, which begins spinning counterclockwise itself. Since it has given up some of its angular momentum, the purple car circles your car less rapidly. Your car gradually stops spinning as its wheels and friction transfer the angular momentum to the ground and Earth. Overall, no angular momentum was created or destroyed during the collision. Instead, it was transferred from the purple car to your car to Earth.



Check Your Understanding #3: Many Happy Returns

Satellites are often set spinning during launch to give them added stability. When astronauts visit these satellites years later, they find them still spinning. Why don't the satellites stop spinning?

Answer: The satellites are unable to get rid of their angular momentum.

Why: Because of their extreme isolation, orbiting satellites have nothing with which to exchange angular momentum. The angular momentum given to them at launch stays with them indefinitely, so they continue to spin for decades.

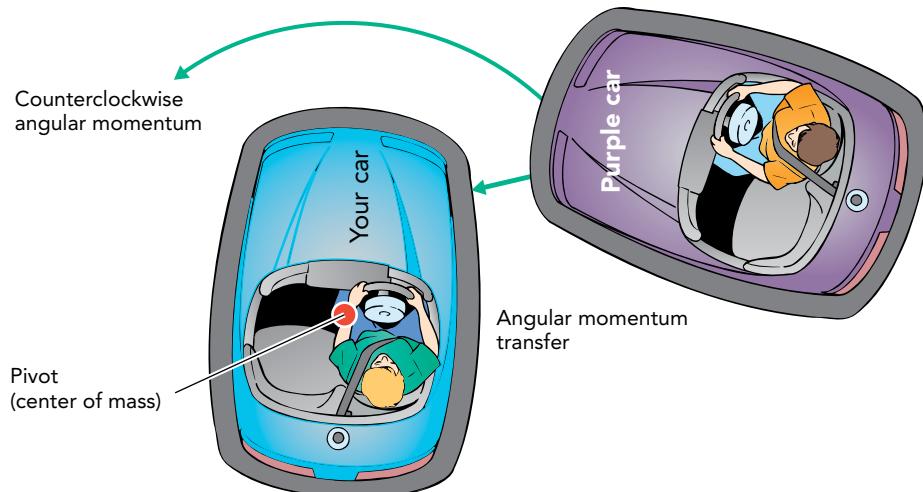


Fig. 2.3.2 Since the purple car is circling your car counterclockwise, it has counterclockwise angular momentum. When it hits your car, it transfers some of that angular momentum to your car. Because of this transfer, the purple car stops circling quickly as your car begins to spin counterclockwise.

Check Your Figures #3: Want to Go for a Spin?

Spinning satellites are particularly stable. Suppose that the astronauts launching a particular satellite decide to increase its angular velocity by a factor of 5. How will that change affect the satellite's angular momentum?

Answer: The angular momentum will increase by a factor of 5.

Why: Because the satellite's angular momentum is proportional to its angular velocity, spinning it five times faster will increase its angular momentum by that same factor.

Glancing Blows: Angular Impulses

Angular momentum is transferred to a bumper car by giving it an **angular impulse**, that is, a torque exerted on it for a certain amount of time. When the purple car hits your car and exerts a torque on it briefly, it gives your car an angular impulse and transfers angular momentum to it. This angular impulse is the change in your car's angular momentum and is equal to the torque exerted on your car times the duration of that torque. This relationship can be written as a word equation:

$$\text{angular impulse} = \text{torque} \cdot \text{time}, \quad (2.3.5)$$

in symbols:

$$\Delta \mathbf{L} = \boldsymbol{\tau} \cdot t,$$

and in everyday language:

To get a merry-go-round spinning rapidly, you must twist it hard and for a long time.

The more torque is exerted or the longer that torque is exerted, the larger the angular impulse and the more your car's angular momentum changes. Once again, an angular impulse is a vector quantity and points in the same direction as the torque. Had the purple car been circling your car clockwise when it struck the glancing blow, its angular impulse would have been in the opposite direction and you'd be spinning the other way.

Different torques exerted for different amounts of time can transfer the same angular momentum to a car:

$$\begin{aligned}\text{angular impulse} &= \text{large torque} \cdot \text{short time} \\ &= \text{small torque} \cdot \text{long time}.\end{aligned}\quad (2.3.6)$$

Thus you can get your car spinning with a certain angular momentum either by letting the motor and floor twist it with a small torque of long duration or by letting the colliding purple car twist it with a large torque of short duration. As with linear momentum, sudden transfers of angular momentum can break things, so the cars are designed to limit their *impact torques* to reasonable levels. Even so, you may find yourself reaching for the motion sickness bag after a few spinning collisions.

Angular momentum is conserved because of Newton's third law of rotational motion. When the purple car exerts a torque on your car for a certain amount of time, your car exerts an equal but oppositely directed torque on the purple car for exactly the same amount of time. Because of the equal but oppositely directed nature of the two torques, the cars receive angular impulses that are equal in amount but opposite in direction. Since the angular momentum gained by one car is exactly equal to the angular momentum lost by the other car, we say that angular momentum is transferred from one car to the other.

Because a car's angular momentum depends on its rotational mass, two different cars may end up rotating at different angular velocities even though they have identical angular momenta. For example, when the purple car hits the overloaded green car and transfers angular momentum to it, the green car's enormous rotational mass makes it spin relatively slowly. The same sort of behavior occurs with linear momentum, where a car's mass affects how fast it travels when it's given a certain amount of linear momentum. But while a bumper car can't change its mass, it can change its rotational mass. If it does so while it's spinning, its angular *momentum* won't change, but its angular *velocity* will!

To see this change in angular velocity, consider the overloaded green car. Its two large occupants are disappointed with the ride because their huge mass and rotational mass prevent them from experiencing the intense jolts and spins that you've been enjoying. Suddenly they get a wonderful idea. As their car slowly spins, one of them climbs into the other's lap and the two sit very close to the car's center of mass. By rearranging the car's mass this way, they have reduced the car's overall rotational mass and the car actually begins to spin faster than before.

As the green car's mass is being redistributed, it's not a freely turning rigid object covered by Newton's first law of rotational motion. However, it is freely turning and thus covered by a more general and equally useful rule: an object that is not subject to any outside torques has constant angular momentum. As the car's rotational mass becomes smaller, its angular velocity must increase to keep its angular momentum constant. That's just what happens. This effect of changing one's rotational mass explains how an ice skater can achieve an enormous angular velocity by pulling herself into a thin, spinning object on ice (Fig. 2.3.3).

© Brian Kersey/NewsCom



Fig. 2.3.3 When skater Sarah Hecken pulls in her arms, she reduces her rotational mass. Since she is experiencing zero net torque, her angular momentum must remain constant and she begins to spin more rapidly.



CONSERVED QUANTITY: ANGULAR MOMENTUM

TRANSFERRED BY: ANGULAR IMPULSE

Angular momentum: The measure of an object's rotational motion, its tendency to continue spinning about a particular axis. Angular momentum is a vector quantity, meaning that it has a direction. It has no potential form and therefore cannot be hidden; angular momentum = rotational mass · angular velocity.

Angular impulse: The mechanical means for transferring angular momentum; angular impulse = torque · time.

Check Your Understanding #4: Spinning the Merry-Go-Round

A person who is initially motionless starts a merry-go-round spinning and then returns to being motionless. If angular momentum is truly conserved, what is the source of the angular momentum that the spinning merry-go-round now has?

Answer: It came from the entire Earth.

Why: Because the person stood on Earth as he started the merry-go-round spinning, he transferred angular momentum from Earth to the merry-go-round. The merry-go-round spins in one direction, and Earth's rotation changes ever so slightly in the other direction. Because Earth is so huge and has such an enormous rotational mass, its slight change in rotation is undetectable.

Check Your Figures #4: Spin Away!

How much longer will it take the astronauts launching the satellite in Check Your Figures #3 to bring it to the faster angular velocity, if they use the initially planned torque?

Answer: It will take them five times as long.

Why: To reach the new, faster angular velocity, the astronauts will need an angular impulse that's five times as large as originally planned. Since they will be using the same torque, they will have to exert that torque for five times as long.

The Three Conserved Quantities

As you drive your bumper car around the oval track, its motion is governed in large part by three conserved quantities: energy, linear momentum, and angular momentum (Table 2.3.1). While you can exchange those quantities with Earth and the power company by steering your car or switching on its motor, most of the interesting exchanges involve collisions.

Each time your car shoves another car forward, your car does work on that other car and transfers energy to it. Each time your car pushes another car northward, your car gives a northward impulse to that other car and transfers northward momentum to it. And each time your car twists another car clockwise about its center of mass, your car gives a clockwise angular impulse to the other car and transfers clockwise angular momentum to it. These exchanges of energy, momentum, and angular momentum are fast and furious and make for an exciting ride.

Quantity	SI Unit	English Unit	SI → English	English → SI
Momentum	kilogram-meter per second ($\text{kg} \cdot \text{m/s}$)	pound-foot per second ($\text{lbm} \cdot \text{ft/s}$)	$1 \text{ kg} \cdot \text{m/s} =$ 7.2329 $\text{lbm} \cdot \text{ft/s}$	$1 \text{ lbm} \cdot \text{ft/s} =$ 0.13826 $\text{kg} \cdot \text{m/s}$
Angular momentum	kilogram-meter ² per second ($\text{kg} \cdot \text{m}^2/\text{s}$)	pound-foot ² per second ($\text{lbm} \cdot \text{ft}^2/\text{s}$)	$1 \text{ kg} \cdot \text{m}^2/\text{s} =$ 23.730 $\text{lbm} \cdot \text{ft}^2/\text{s}$	$1 \text{ lbm} \cdot \text{ft}^2/\text{s} =$ 0.042140 $\text{kg} \cdot \text{m}^2/\text{s}$

TABLE 2.3.1 The Three Conserved Quantities of Motion and Their Transfer Mechanisms

Context	Conserved Quantity	Transfer Mechanism
Doing	Energy	Work
Moving	Linear momentum	Impulse
Turning	Angular momentum	Angular impulse

Check Your Understanding #5: Hitting the Wall

You're backing out of a parking space and accidentally hit a concrete wall. The wall doesn't move, and your car sustains some damage. Did your car transfer any energy or momentum to the wall?

Answer: Your car transferred momentum but no energy.

Why: To transfer momentum to the wall, your car must give it an impulse: it must push on the wall for an amount of time. It did so and thus transferred all its backward momentum to the wall. But to transfer energy to the wall, your car must do work on it: it must push on the wall as the wall moves in the direction of that push. The wall is immobile, however, so your car couldn't do work on it. Instead, the car's energy stayed in the car, where it caused damage.

Potential Energy, Acceleration, and Force

Shortly before the bumper car ride stops, you notice that the floor is somewhat uneven and that cars accelerate down whatever slopes they encounter. These accelerations often deflect cars as they cross depressions or bumps, adding to the craziness of the ride. Depending on its initial velocity, a car that moves across a slope may change its speed, its direction, or both.

COMMON MISCONCEPTIONS: Acceleration and Velocity

Misconception: An object's acceleration and velocity always point along the same line.

Resolution: Although acceleration measures the rate at which velocity changes with time, those two quantities are independent at any given moment and may or may not point along the same line. When they are not aligned, the direction of the object's velocity will shift over time toward the direction of its acceleration and the object's path will bend.

We have seen this tendency to accelerate downhill before with ramps, but now let's look at it in terms of potential energy. When a bumper car accelerates downhill, it also accelerates in the direction that reduces its potential energy as quickly as possible. That's not a coincidence; there is a direct relationship between the car's acceleration and how the car's potential energy changes with position. That relationship, which we're about to explore, will prove to be quite useful throughout this book.

Before elaborating further, however, I need to introduce a type of vector quantity known as a gradient. A **gradient** characterizes how some physical quantity changes gradually with position. It points in the direction of fastest increase in its physical quantity, and its magnitude is the rate of that increase. To find that rate, take a small step in the direction of fastest increase and divide the resulting increase by the length of the step.

For example, this  has a redness gradient. That gradient points toward the right because that's the direction of fastest redness increase, and its magnitude is 100% per inch because the redness increases by 100% over a distance of 1 inch to the right. Similarly, this  has a blueness gradient that points to the left and has a magnitude of 200% per inch.

A ramp has an altitude gradient; that is, the ramp's altitude increases gradually along its surface from the bottom of the ramp to the top. That altitude gradient points uphill—in the direction of quickest altitude increase—and its magnitude is the increase in altitude caused by a small uphill step divided by the length of that step. The steeper the ramp, the larger its altitude gradient.

Last, a bumper car riding on a ramp has a potential energy gradient; that is, the car's potential energy increases gradually as the car moves along the surface of the ramp from the bottom of the ramp to the top. This potential energy gradient, or simply **potential gradient**, points uphill—in the direction of quickest potential energy increase—and its magnitude is the increase in the car's potential energy caused by a small uphill step divided by the length of that step. The steeper the ramp, the larger the car's potential gradient.

Interestingly, the car's potential gradient and its acceleration point in exactly opposite directions. The car's potential gradient points in the direction that *increases* its potential energy as quickly as possible, whereas the bumper car accelerates in the direction that *reduces* its potential energy as quickly as possible.

That's not a coincidence. Potential energies are energies stored in forces and pushing the bumper car in the direction of the potential gradient stores potential energy in a force that must be pushing the bumper car in the opposite direction. In fact, the negative of the car's uphill potential gradient *is* the downhill force on the car. We can write this relationship as a word equation:

$$\text{force} = -\text{potential gradient}, \quad (2.3.7)$$

in symbols:

$$\mathbf{F} = -\mathbf{Gradient} (U),$$

and in everyday language:

An object accelerates in the direction that reduces its potential energy as quickly as possible,

where the force points in the direction of quickest potential energy decrease.

A similar rule applies for rotational motion and torques. When there are multiple forms of potential energy, we must remember to sum them to obtain *total* potential energy, *total* potential gradient, and *net* force. These rules provide a useful way to determine how motion will proceed: which way a spring will leap, a chair will tip, or a bumper car will roll. We'll use them frequently in this book.

POTENTIAL ENERGY AND ACCELERATION

An object accelerates in the direction that reduces its total potential energy as quickly as possible.

Check Your Understanding #6: Heading Down

When you pull a child back on a playground swing and let go, which way does that child accelerate?

Answer: The child accelerates forward, in the direction that will reduce the child's potential energy as quickly as possible.

Why: The child has only one form of potential energy—gravitational potential energy. This gravitational potential energy is lowest when the child is directly below the swing's supporting bar. The child accelerates forward because that will put the child below the support as quickly as possible.

Check Your Figures #5: Pumpkin Chucking

You've constructed a catapult for a pumpkin-chucking contest. When you let go of the basket holding the pumpkin, the catapult will swing that basket forward and release 10,000 J of potential energy as it travels 1 m forward. What force must you exert on the basket to hold it in place prior to letting go?

Answer: You must exert 10,000 N on the basket in the backward direction.

Why: When you let go, the basket will accelerate forward—the direction that reduces its potential energy as quickly as possible. According to Eq. 2.3.7, since the basket acquires 10,000 J in traveling backward 1 m, the forward force acting on the basket is:

$$\frac{10,000 \text{ J}}{1 \text{ m}} = 10,000 \text{ N.}$$

To keep the basket stationary, you must exert a force of 10,000 N in the backward direction.

Epilogue for Chapter 2

In this chapter we have looked at rotating and colliding objects and have studied the physical laws that describe their motions. In Seesaws, we examined rotational inertia and saw how torques cause angular accelerations. We also noticed how useful it can be to separate an object's rotational motion from its translational motion. In Wheels, we discussed another important type of force, friction, as well as a new type of energy, thermal energy, the energy associated with heat and temperature. In Bumper Cars, we introduced two more conserved physical quantities: momentum and angular momentum. As we'll see, following the transfers of energy, momentum, and angular momentum between objects often helps our understanding of how those objects work.

Explanation: A Spinning Pie Dish

Because the balanced dish has rotational inertia, torques are required both to start it spinning and to stop it from doing so. When you twist the dish with your hand, the torque you exert gives the dish an angular impulse and sets it spinning with a certain amount of angular momentum. If the pivot were truly frictionless and there were no air resistance, the dish would spin indefinitely because it would be unable to get rid of its angular momentum. However, friction in the pivot exerts a small but significant torque that opposes the dish's motion and gradually slows it down. As this frictional torque transfers angular momentum out of the dish and into the pencil, chair, and Earth, the dish turns more and more slowly until it finally comes to a stop. The sharper the pivot and the smaller the contact area between point and dish, the less frictional torque the dish experiences and the longer it spins.

Balancing the dish is easy as long as it's upside-down. With its edge drooping downward, the dish has relatively little gravitational potential energy and is surprisingly stable. If it begins to tip to one side, the dish's average height rises and so does its gravitational potential energy. Since objects naturally accelerate in whatever direction lowers their potential energy as quickly as possible, the upside-down dish quickly tips back toward level after being disturbed. In contrast, an upright dish is virtually impossible to balance on a point because any tip will lower its gravitational potential energy and lead quickly to catastrophe. We'll look at these stabilizing/destabilizing effects more carefully later on in this book.

Chapter Summary and Important Laws and Equations

How Seesaws Work: A seesaw is a rotating toy that works best when it's almost perfectly balanced, meaning that it experiences no overall torque about its pivot due to gravity. The seesaw's pivot usually passes through its center of gravity so that the seesaw balances when it's not occupied. The riders arrange themselves so that their gravitational torques cancel one another and the occupied seesaw continues to balance. The seesaw then experiences zero net torque and zero angular acceleration, and it rotates with constant angular velocity. It either remains motionless or turns steadily in one direction or the other.

To make the seesaw tip back and forth, the riders briefly upset its inertial motion. Often they do this by pushing against the ground with their feet to produce additional torques on the seesaw. The seesaw then experiences both a net torque and an angular acceleration. By rhythmically changing the net torque on the seesaw, the riders cause it to rotate back and forth.

How Wheels Work: Wheels facilitate motion by eliminating or reducing sliding friction between an object and a surface. The wheels convey the support forces needed to hold the object up but allow the object to move without sliding. As a cart with freely turning wheels moves along the surface, static friction between each wheel and the surface exerts a torque on that wheel and causes it to turn. However, rubbing may occur between the wheel's hub

and the axle, where sliding friction can waste energy and cause wear. To eliminate this sliding friction, roller or ball bearings are often used.

The torque that causes a powered wheel on a vehicle to turn comes from an engine by way of an axle. In this case, static friction between the outside of the wheel and the ground exerts a torque on the wheel that opposes the torque from the engine. This static frictional force also contributes to the net force on the vehicle and propels it.

Once supported on wheels and bearings, objects can move freely and can retain linear momentum, angular momentum, and energy for long periods of time. By eliminating sliding friction, wheels can also keep objects from converting ordered energy into thermal energy. Wheels allow vehicles to hold onto these conserved quantities for extended periods and make transportation far more practical.

How Bumper Cars Work: Since they start from rest, bumper cars must obtain their initial momenta and angular momenta from the ground and their initial kinetic energies from the power company. They do this with the help of motors and wheels, which gradually transfer energy, momentum, and angular momentum into the cars.

Once the cars are moving, they can begin to exchange those conserved quantities by way of collisions. Each impact usually changes the cars' speeds and directions of travel in a manner that may seem rather complicated. However, following the exchanges of momentum, angular momentum, and energy often makes it easier to understand these collisions.

Cars containing massive riders respond weakly when collisions transfer momentum and angular momentum to them. That's because their large masses and rotational masses minimize their changes in velocity and angular velocity. Because of their small masses, children experience the wildest rides.

1. Newton's first law of rotational motion: A rigid object that is not wobbling and is not subject to any outside torques rotates at a constant angular velocity, turning equal amounts in equal times about a fixed axis of rotation.

2. Newton's second law of rotational motion: The net torque exerted on an object is equal to that object's rotational mass times its angular acceleration, or

$$\text{net torque} = \text{rotational mass} \cdot \text{angular acceleration}. \quad (2.1.2)$$

The angular acceleration points in the same direction as the torque. This law doesn't apply to objects that are nonrigid or wobbling.

3. Relationship between force and torque: The torque produced by a force is equal to the lever arm times the component of that force perpendicular to the lever arm, or

$$\text{torque} = \text{lever arm} \cdot \text{force perpendicular to lever arm}. \quad (2.1.3)$$

4. Rotational definition of work: The work done on an object is equal to the torque exerted on that object times the angle (in radians) through which that object rotates in the direction of the torque, or

$$\text{work} = \text{torque} \cdot \text{angle (in radians)}. \quad (2.2.1)$$

5. Kinetic energy: An object's translational kinetic energy is one-half its mass times the square of its speed, or

$$\text{kinetic energy} = \frac{1}{2} \cdot \text{mass} \cdot \text{speed}^2. \quad (2.2.2)$$

An object's rotational kinetic energy is one-half its rotational mass times the square of its angular speed, or

$$\text{kinetic energy} = \frac{1}{2} \cdot \text{rotational mass} \cdot \text{angular speed}^2. \quad (2.2.3)$$

6. Linear momentum: An object's linear momentum is its mass times its velocity, or

$$\text{linear momentum} = \text{mass} \cdot \text{velocity}. \quad (2.3.1)$$

7. Definition of impulse: The impulse given to an object is equal to the force exerted on that object times the length of time that force is exerted, or

$$\text{impulse} = \text{force} \cdot \text{time}. \quad (2.3.2)$$

8. Angular momentum: An object's angular momentum is its rotational mass times its angular velocity, or

$$\text{angular momentum} = \text{rotational mass} \cdot \text{angular velocity}. \quad (2.3.4)$$

9. Definition of angular impulse: The angular impulse given to an object is equal to the torque exerted on that object times the length of time that torque is exerted, or

$$\text{angular impulse} = \text{torque} \cdot \text{time}. \quad (2.3.5)$$

10. Newton's third law of rotational motion: For every torque that one object exerts on a second object, there is an equal but oppositely directed torque that the second object exerts on the first object.

11. Potential energy, acceleration, and force: An object experiences a force and accelerates in the direction that reduces its total potential energy as quickly as possible. That force is equal but opposite to the potential gradient, or

$$\text{force} = -\text{potential gradient}. \quad (2.3.7)$$

Mechanical Objects

PART 1

Now that we've surveyed the laws of motion, we can begin using those laws to explain the behaviors of everyday objects. But while we can already address some of the central features at work in a toy wagon, a weight machine, or a ski lift, we're still missing a number of mechanical concepts that are important in the world around us. In this chapter, we look at some of those additional concepts.

One of the most important new concepts will be the feeling of acceleration. If we treat acceleration passively, it can be fairly uninteresting: we push on the cart, and the cart accelerates. If we think of it more actively, though—for example, if we envision ourselves on a roller coaster as it plummets down that first big hill—then the experience of acceleration becomes much more intriguing. In fact, we might even need to hold on to our hats.

ACTIVE LEARNING EXPERIMENTS

Swinging Water Overhead

To examine some of the novel effects of acceleration, try experimenting with a bucket of water. If you're careful, you can swing the bucket over your head and upside down without spilling a drop. In the process, you'll be demonstrating a number of important physical concepts.

To do this experiment, you'll need a bucket with a handle. (You might substitute some equivalent container; even a plastic cup will do in a pinch.) Fill the bucket

partway full of water and then hold it by the handle so that it hangs down by your side.

Now swing the bucket backward about an eighth of a turn and bring it forward rapidly. In one smooth, fluid motion, swing it forward, up, and over your head. Continue this motion all the way around behind you and then bring the bucket forward again. You'll need to swing the bucket quickly to avoid getting wet. As you swing the



Courtesy Lou Bloomfield

bucket around and around, you'll notice that the water stays in it even when it travels over your head and is upside down. Why doesn't the water fall out?

You can carry this experiment a step further by swinging the bucket at various speeds—that is, if you don't mind getting wet. What will happen if you swing the bucket less rapidly or more rapidly? Is the pull stronger or weaker when you swing less rapidly? more rapidly? Is there any relationship between the upward pull you feel from the inverted bucket and the water's tendency to remain inside?

You might vary this experiment in several ways. Try swinging a plastic cup held in your fingers, or try placing

a full wine glass in the bucket and swinging the two objects together. In the latter case, you'll find that the wine will stay in the glass and the glass will stay at the bottom of the bucket, even when the bucket is upside down.

By the way, the hardest part of all these tricks is stopping. To avoid a catastrophe, you'll need to do the same thing you did to get started, only in reverse. Come to a smooth, gradual stop about an eighth of a turn in front of you, and then let the bucket loosely drop back to your side. If you stop moving the bucket too abruptly, the water, wine, or glass will spill or smash. Why do you suppose that happens?

Chapter Itinerary

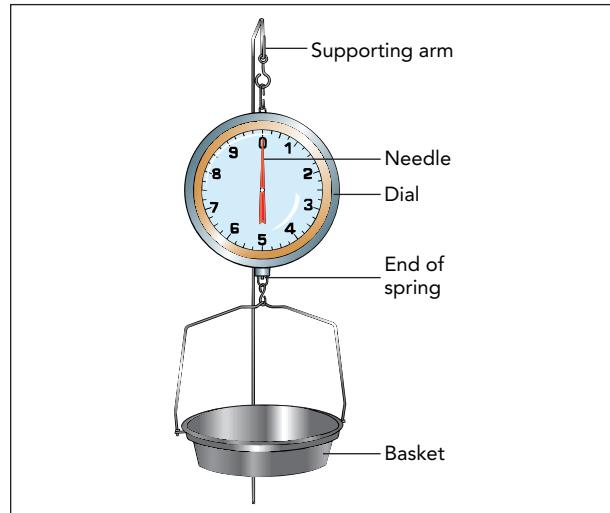
In this chapter, we examine three types of everyday objects: (1) *spring scales*, (2) *ball sports: bouncing*, and (3) *carousels and roller coasters*. In Spring Scales, we review the relationship between mass and weight and explore how the distortion of a spring can be used to measure an object's weight. In Ball Sports: Bouncing, we study how balls store and return energy and how their bouncing depends both on their own characteristics and on those of the objects they hit. And in Carousels and Roller Coasters, we look at how acceleration gives rise to gravity-like apparent forces that can make us scream with delight at the amusement park. For a more complete preview of what we examine in this

chapter, skip ahead to the Chapter Summary and Important Laws and Equations at the end of the chapter.

The concepts these objects illustrate can explain other phenomena as well. Almost any solid object, from a mattress to a diving board to a tire, behaves like a spring scale's spring when you push on it. Bouncing balls offer a view of collisions that will help you comprehend what happens when two cars crash or when a hammer hits a nail. And the sensations associated with acceleration that you experience on a roller coaster are also present when you ride in airplanes, on subways, or on swing sets. When it comes to the physics of everyday objects, there really is nothing new under the sun.

SECTION 3.1

Spring Scales



How much of you is there? From day to day, depending on how much you eat, the amount of you stays approximately the same. But how can you tell how much that is? The best measure of quantity is mass: kilograms of gold, grain, or you. Mass is the

measure of an object's inertia and, as we saw in Section 1.1, it doesn't depend on the object's environment or on gravity. A kilogram box of cookies always has a mass of 1 kilogram, no matter where in the universe you take it.

But mass is difficult to measure directly. Moreover, the very concept of mass is only about 300 years old. Consequently, people began quantifying the material in an object by measuring its weight. Spring scales eventually became one of the simplest and most practical tools for accomplishing this task, and they are still found in bathrooms and grocery stores today. They really do contain springs, although these are normally hidden from view.

Questions to Think About: How is your weight related to your mass? If Earth's gravity became twice as strong, how would your mass be affected? What about your weight? Does jumping up and down change either your mass or your weight? If you stand on a strong spring, how does your weight affect the shape of the spring? Why should there be a relationship between your weight and how much the spring bends?

Experiments to Do: Find a hanging spring scale of the type used in the produce section of a grocery store and watch the

scale's basket and weight indicator as you put objects in the basket. What happens to the basket as you fill it up? Can you find a relationship between the basket's height and the weight reported by the scale? If you drop something into the basket, instead of lowering it gently, how does the scale respond? Why does the weight indicator bounce back and forth rhythmically? What happens to the gravitational potential energy of the dropped item?

Now find a spring bathroom scale—the short, flat kind with a rotating dial. Stand on it. Why does it read your correct

weight only when you are standing still? If you jump upward, how does the scale's reading change? What about if you let yourself drop? Bounce up and down gently. How does the scale's average reading compare with your normal weight? Does bouncing really change your weight? You can also change the scale's reading by pushing on the floor, wall, or other nearby objects. Which way must you push to increase the scale's reading? to decrease it? When you change the reading in these ways, are you actually changing your weight?

Why You Must Stand Still on a Scale

When you stand on a bathroom scale, the scale doesn't directly measure your weight. That's because your weight is a force that Earth's gravity exerts on *you*, not on *the scale*, so it affects the scale only indirectly. What the scale actually reports is how much upward force it exerts on you to keep you from falling. When you're not falling or otherwise accelerating, the scale pushes you upward just as hard as your weight pulls you downward, so knowing the scale's force on you is equivalent to knowing your weight.

That subtle difference between your weight and what the scale reports is important, however, because it makes the weighing process sensitive to acceleration. If you're accelerating, the scale won't report your true weight. For example, when you jump up and down on the scale, the scale's reading varies wildly. You accelerate as you jump, so the two forces acting on you—your weight downward and the scale's push upward—don't sum to zero. If you want an accurate measurement of your weight, therefore, you have to stand still.

Even when you stand still, weighing is not a perfect way to quantify the amount of material in your body. That's because your weight depends on your environment. If you always weigh yourself in the same place, the readings will be pretty consistent, as long as you don't routinely eat a dozen jelly doughnuts for lunch. But if you moved to the moon, where gravity is weaker, you'd weigh only about one-sixth as much as on Earth. Even a move elsewhere on Earth will affect your weight: Earth bulges outward slightly at the equator, and gravity there is about 0.5% weaker than at the poles. That change, together with a small acceleration effect due to Earth's rotation, means that a scale will read 1.0% less when you move from the north pole to the equator. Obviously, moving south is not a useful weight-loss plan.



Check Your Understanding #1: Space Merchants

You're opening a company that will export gourmet food from Earth to the moon. You want the package labels to be accurate at either location. How should you label the amount of food in each package—by mass or by weight?

Answer: You should sell by mass—for example, by kilogram or pound-mass.

Why: If you label your product by weight, you are specifying the force that gravity exerts on it near Earth's surface. When it's exported to the moon, such a product will weigh just $\frac{1}{6}$ as much, and your company may be fined for selling underweight groceries. If you label the packages according to their masses, that labeling will remain correct no matter where you ship the packages. Mass is the measure of inertia and depends only on the object, not on its environment.

Stretching a Spring

At the grocery store, you weigh a melon by placing it in the basket of a spring scale. Like a bathroom scale, the grocery store scale can't measure the melon's weight directly;

instead, it reports how much upward force it exerts on the melon. As long as the melon isn't accelerating, the scale's upward force on the melon is equal but opposite to the melon's weight, so measuring one force is equivalent to measuring the other. But how does the spring scale determine how hard it's pushing upward on the melon?

In keeping with its name, the spring scale uses a spring to push upward on the melon. The beauty of that arrangement is that a simple relationship exists between a spring's length and the forces it's exerting on its ends. The spring scale determines how much upward force it's exerting on the melon by measuring the length of its spring.

The spring shown in Fig. 3.1.1 consists of a wire coil that pulls inward on its ends when it's stretched and pushes outward on them when it's compressed. When that spring is neither stretched nor compressed (Fig. 3.1.1a), it exerts no forces on its ends. If no other forces act on this relaxed spring, each end is in **equilibrium**—it experiences zero net force. As the phrase *zero net force* suggests, equilibrium occurs whenever the forces acting on an object sum to zero so that the object doesn't accelerate. When you sit motionless in a chair, for example, you are in equilibrium. The relaxed spring is also at its **equilibrium length**, its natural length when you leave it alone. No matter how you distort this spring, it tries to return to this equilibrium length.

Let's attach the left end of our spring to a wall (Fig. 3.1.1b) and its right end to a moveable bar. When the bar isn't pulling or pushing on the spring's right end, that end is in equilibrium at a particular location—its **equilibrium position**. Since the spring's right end naturally returns to this equilibrium position if the bar disturbs it and then lets go, the end is in a **stable equilibrium**.

What happens if the bar moves the spring's right end to the right (Fig. 3.1.1c)? The stretched spring now exerts a leftward force on that end, trying to restore it to its original equilibrium position. Because the force that the spring exerts on its end acts to restore that end to equilibrium, we call it a **restoring force**. The spring actually exerts restoring forces on both of its ends, pulling them equally hard in opposite directions. Because the left end of the spring is held in place by the wall, however, we'll focus our attention on its moveable right end and the restoring force on that end.

The farther the bar stretches the spring to the right, the stronger its restoring force on the right end becomes (Fig. 3.1.1d). Remarkably, that restoring force is exactly proportional to how far the bar has stretched the spring.

If the bar changes directions and moves the spring's right end to the left (Fig. 3.1.1e), the compressed spring exerts a rightward force on its right end. Nonetheless, that force is still a restoring force and it is still proportional to the spring's distortion.

In general, whenever a spring is distorted away from its equilibrium length, it exerts a restoring force on its moveable end that is proportional to the distortion and points in the opposite direction. This observation is expressed in **Hooke's law**, named after the Englishman Robert Hooke, who discovered it in the late seventeenth century. This law can be written in a word equation:

$$\text{restoring force} = -\text{spring constant} \cdot \text{distortion}, \quad (3.1.1)$$

in symbols:

$$\mathbf{F} = -k \cdot \mathbf{x},$$

and in everyday language:

The stiffer the spring and the farther you stretch it, the harder it pulls back.

Here the **spring constant**, k , is a measure of the spring's stiffness. The larger the spring constant—that is, the firmer the spring—the larger the restoring force the spring exerts for

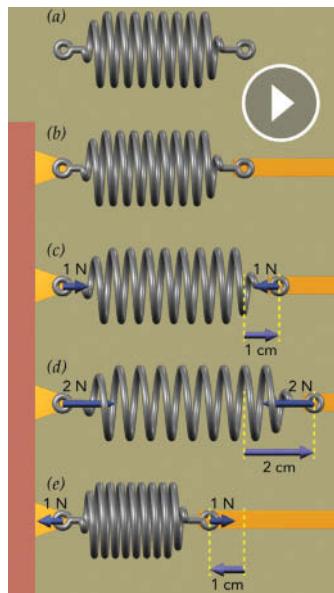


Fig. 3.1.1 Five identical springs. The ends of spring (a) are free so that it can adopt its equilibrium length. The left ends of the other four springs are fixed so that only their right ends can move. The right end of spring (b) remains at its equilibrium position, but springs (c, d, and e) are distorted and exert restoring forces on their ends that are proportional to how far their right ends have been distorted from their equilibrium positions.

a given distortion. The negative signs in these equations indicate that a restoring force always points in the direction opposite the distortion.

HOOKE'S LAW

The restoring force exerted by an elastic object is proportional to how far it has been distorted from its equilibrium shape.

Springs are distinguished by their stiffness, as measured by their spring constants. Some springs are **soft** and have small spring constants—for example, you need just one finger to compress the spring in your retractable ballpoint pen. Other springs, like the large ones that suspend an automobile chassis above the wheels, are **firm** and have large spring constants. But no matter the stiffness, nearly all simple springs obey Hooke's law.

Hooke's law is remarkably general and isn't limited to the behavior of coil springs. Many objects respond to distortion with restoring forces that are proportional to how far you've distorted them away from their equilibrium lengths—or, in the case of some complicated objects, their equilibrium shapes. *Equilibrium shape* is the shape an object adopts when it's not subject to any outside forces. If you bend a tree branch, it will push back with a force proportional to how far it has been bent. If you pull on a rubber band, it will pull back with a force proportional to how far it has been stretched, up to a point. If you squeeze a mattress, it will push outward with a force proportional to how far it has been compressed. If a heavy truck bends a bridge downward, the bridge will push upward with a force proportional to how far it has been bent (Fig. 3.1.2).

There is a limit to Hooke's law, however. If you distort an object too far, it will usually begin to exert less force than Hooke's law demands. This is because you will have exceeded the object's **elastic limit** and will have deformed it permanently in the process. If you stretch a spring too far, it won't return to its original equilibrium length when you release it; if you bend a tree branch too far, it will break. As long as you stay within the elastic limit, however, a great many things obey Hooke's law—a rope, a ruler, a tabletop, and a trampoline.

Distorting a spring requires work. When you stretch a spring with your hand, pulling its ends apart as they move apart, you transfer some of your energy to the spring. The spring stores this energy as **elastic potential energy**. If you reverse the motion, the spring returns most of this energy to your hand; the remainder is converted to thermal energy by frictional effects inside the spring itself. Work is also required to compress, bend, or twist a spring. In short, a spring that is distorted away from its equilibrium shape always contains elastic potential energy.

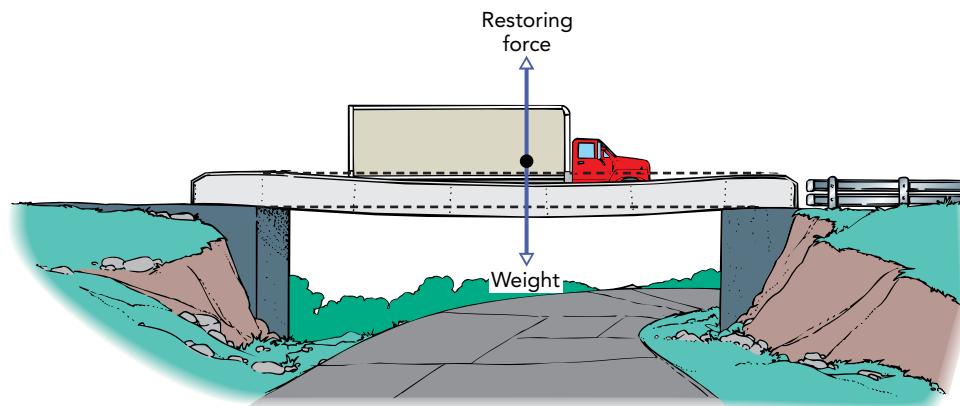


Fig. 3.1.2 A steel bridge sags under the weight of a truck. The bridge bends downward until the upward restoring force it exerts on the truck exactly balances the truck's weight.

Check Your Understanding #2: Going Down Anyone?

As you watch people walk off the diving board at a pool, you notice that it bends downward by an amount proportional to each diver's weight. Explain.

Answer: The diving board is behaving as a spring, bending downward in proportion to the weight of each diver.

Why: The heavier the diver, the more the board bends downward before exerting enough upward force on the diver to balance the diver's weight.

Check Your Figures #1: A Sinking Sensation

You're hosting a party in your third-floor apartment. When the first 10 guests begin standing in your living room, you notice that the floor has sagged 1 centimeter in the middle. How far will the floor sag when 20 guests are standing on it? when 100 guests are standing on it?

Answer: It will sag 2 centimeters and 10 centimeters (assuming that the floor doesn't break).

Why: A floor, like most suspended surfaces, behaves like a spring. Your floor distorts 1 cm before it exerts an upward restoring force equal to the weight of 10 guests. It will thus distort 2 cm before supporting 20 guests and 10 cm before supporting 100 guests. While this distortion should be within the elastic limit of the floor beams, it may cause the plaster and paint to crack. If the beams break, the floor will collapse.

How a Hanging Grocery Scale Measures Weight

Before you weigh another melon, let's examine the grocery store scale. It contains a coil spring that's responsible for the weighing process. One end of that spring is suspended from the ceiling, and the other end supports the weighing basket (Fig. 3.1.3). For the sake of simplicity, let's suppose the spring and the basket have negligible weights and that the basket is effectively the bottom end of the spring. With the basket empty, the spring adopts its equilibrium length and the basket is in a position of stable equilibrium. If you shift the basket up or down and then let go of it, the spring will push it back to this position. When you place a melon in the basket, the melon's weight pushes the basket downward and spoils its equilibrium. The basket starts descending and as it does, the spring stretches and begins to exert an upward force on the basket. The more the spring stretches, the greater this upward force so that eventually the spring is stretched just enough so that its upward restoring force exactly supports the melon's weight. The basket is now in a new stable equilibrium position—again experiencing zero net force.

The scale, then, uses Hooke's law to determine the weight of the melon. When the basket has settled at its new equilibrium position, where the melon's downward weight is exactly balanced by the spring's upward restoring force, the amount the spring has stretched is an accurate measure of the melon's weight.

The scales in Fig. 3.1.3 differ only in the way they measure how far the spring has stretched beyond its equilibrium length. In Fig. 3.1.3a, the scale uses a pointer attached to the end of the spring, while in Fig. 3.1.3b, the scale uses a rack and pinion gear system that converts the small linear motion of the stretching spring into a much more visible rotary motion of the dial needle. The rack is the series of evenly spaced teeth attached to the lower end of the spring; the pinion is the toothed wheel attached to the dial needle. As the spring stretches, the rack moves downward, and its teeth cause the pinion to rotate. The farther the rack moves, the more the pinion turns and the higher the weight reported by the needle.

Each of these spring scales reports a number for the weight of the melon you put in the basket. For that number to mean something, the scale has to be calibrated. **Calibration** is the process of comparing a local device or reference to a generally accepted standard to ensure accuracy. To calibrate a spring scale, the device or its reference components must be

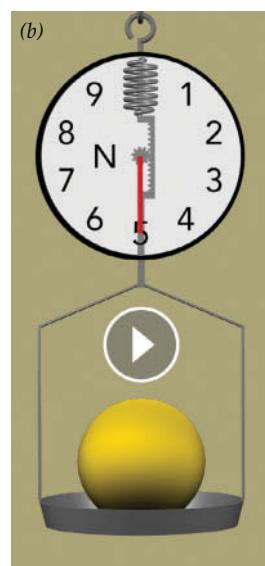
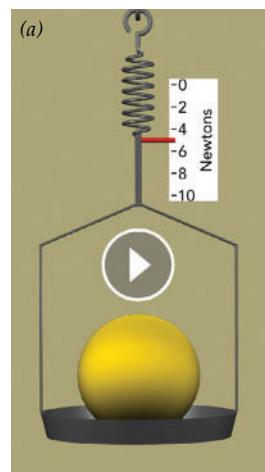


Fig. 3.1.3 Two spring scales weighing melons. Each scale balances the melon's downward weight with the upward force of a spring. The heavier the melon, the more the spring will stretch before it exerts enough upward force to balance the melon's weight. The top scale has a pointer to indicate how far the spring has stretched and thus how much the melon weighs. The bottom scale has a rack and pinion gear that turns a needle on a dial. As the comblike rack moves up and down, it turns the toothed pinion gear.

compared against standard weights. Someone must put a standard weight in the basket and measure just how far the spring stretches. Each spring is different, although spring manufacturers try to make all their springs as identical as possible.

Courtesy Lou Bloomfield



Fig. 3.1.4 When you step on this bathroom scale, its surface moves downward slightly and compresses a stiff spring. The extent of this compression is proportional to your weight, which is reported by the dial. Levers inside the scale make it insensitive to exactly where you stand.

Check Your Understanding #3: Scaling Down

If you pull the basket of a hanging grocery store scale downward 1 cm, it reports a weight of 5 N (about 1.1 lbf) for the contents of its basket. If you pull the basket downward 3 cm, what weight will it report?

Answer: The scale will read 15 N (about 3.3 lbf).

Why: The scale's dial is simply reporting the position of its basket. The dial is calibrated so that a 1-cm drop in the basket indicates that the spring is pulling up on it with a force of 5 N. Since the spring's restoring force is described by Hooke's law, a 3-cm drop in the basket means that the spring is exerting an upward force of 15 N on the basket.

Bouncing Bathroom Scales

As we noted earlier, the most common type of bathroom scale is also a spring scale (Fig. 3.1.4). When you step on this kind of scale, you depress its surface and levers inside it pull on a hidden spring. That spring stretches until it exerts, through the levers, an upward force on you that is equal to your weight. At the same time, a rack and pinion mechanism (see Fig. 3.1.3b) inside the scale turns a wheel with numbers printed on it. When the wheel stops moving, you can read one of these numbers through a window in the scale. That number depends on how far the spring has stretched and has been calibrated so that it accurately indicates your weight.

However, the wheel rocks briefly back and forth around your actual weight before it settles down. The wheel moves because you're bouncing up and down as the scale gradually gets rid of excess energy. When you first step on the scale's surface, its spring is not stretched and it isn't pushing up on you at all. You begin to fall. As you descend, the spring stretches and the scale begins to push up on your feet. By the time you reach the equilibrium height, where the scale is exactly supporting your weight, you are traveling downward quickly and coast right past that equilibrium. The scale begins to read more than your weight.

The scale now accelerates you upward. Your descent slows, and you soon begin to rise back toward equilibrium. Again you coast past equilibrium, but now the scale begins to read less than your weight. You are bouncing up and down because you have excess energy that is shifting back and forth between potential energy and kinetic energy. This bouncing continues until sliding friction in the scale has converted it all into thermal energy. Only then does the bouncing stop and the scale read your correct weight.

The bouncing that you experience about this stable equilibrium is a remarkable motion (Fig. 3.1.5), one that we'll study in detail when we examine clocks in Chapter 9. You are effectively a mass supported by a spring, and your rhythmic rise and fall is that of a *harmonic oscillator*. Harmonic oscillators are so common and important in nature that Chapter 9 is devoted to them. The details can wait, but there are two features of your present situation that I'll point out now.

First, your total potential energy is at its minimum when you're at the stable equilibrium. Even though both gravitational and elastic potentials are involved, their sum increases as you shift away from the equilibrium. Since an object always accelerates so as to reduce its total potential energy as quickly as possible, you always accelerate toward the stable equilibrium. Because the restoring force you experience increases as you move farther from the equilibrium, you reach your peak acceleration and potential energy at the moment when you are farthest from equilibrium and are turning around to head back toward that equilibrium.

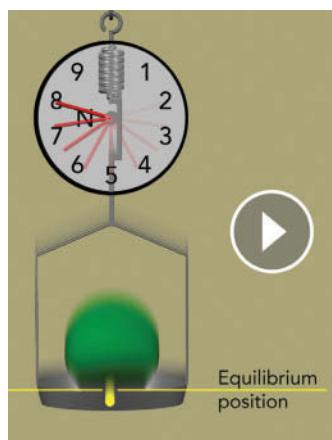


Fig. 3.1.5 When you place a melon in the basket of a spring scale, the basket and melon bounce briefly about their equilibrium position as the scale gradually removes their excess energy. Their rhythmic motion about the equilibrium position is that of a mass on a spring and is an example of a harmonic oscillator.

Second, your kinetic energy reaches its peak as you pass through the stable equilibrium. Having accelerated toward that equilibrium until the moment of arrival, you're moving fast and coast right through it. But once you leave equilibrium, you begin to accelerate toward it again. That acceleration is backward, opposite your velocity, so you are decelerating. Therefore, you reach your peak speed and kinetic energy at the moment you pass through equilibrium. As you bounce up and down, waiting for the scale to waste your excess energy, that energy transforms back and forth rhythmically between kinetic and potential forms.

COMMON MISCONCEPTIONS: Equilibrium and Motionlessness

Misconception: An object at equilibrium is always motionless.

Resolution: An object at equilibrium is not accelerating, but its velocity may not be zero. If it was moving when it reached equilibrium, it will coast through that equilibrium at constant velocity.

Check Your Understanding #4: Weighed Down

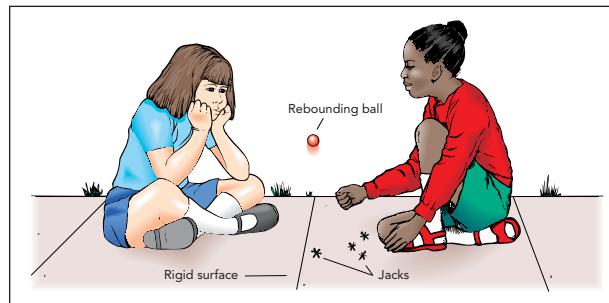
When you step on the surface of a spring bathroom scale, you can feel it move downward slightly. How is the distance that the scale's surface moves downward when you step on it related to the weight it reports?

Answer: The distance the scale's surface moves downward is proportional to the weight it reports.

Why: The scale's spring is connected to its surface by levers so that as the surface moves downward, the spring distorts by a proportional amount. The spring's distortion is reported on the dial. Thus the dial's reading is proportional to the surface's downward movement.

SECTION 3.2

Ball Sports: Bouncing



If you visit a toy or sporting goods store, you'll find many different balls—almost a unique ball for every sport or ball game. These balls differ in more than just size and weight. Some are very hard, others very soft; some are smooth, others rough or ridged.

In this section we'll focus primarily on another difference—their ability to bounce. A bouncy ball, for example, bounces extraordinarily well, while a foam rubber ball hardly bounces at all. Even balls that appear identical can be very different; a new tennis ball bounces much better than an old one. We'll begin this section by exploring these differences.

Questions to Think About: Is it possible for a ball to bounce higher than the height from which it was dropped? Where does a ball's kinetic energy go as it bounces, and what happens to the energy that doesn't reappear after the bounce? What happens when a ball bounces off a moving object, such as a baseball bat? What role does the baseball bat's structure have in the bouncing process? Does it matter which part of a baseball bat hits the ball?

Experiments to Do: Drop a ball on a hard surface and watch it bounce. What happens to the ball's shape during the bounce? Hold the ball in your hands and push its surface inward with your fingers. What is the relationship between the force it exerts on your fingers and how far inward you dent it? Denting the ball takes work. Why? How does the ball's energy change as it dents? What happens to the ball's energy as its shape returns to normal (to equilibrium)?

Drop the ball from various heights and see whether you can find a simple relationship between the ball's initial height and the height to which it rebounds. Now drop the ball on a soft surface, such as a pillow, or on a lively surface, such as an inflated balloon. Why does the surface it hits change the way the ball bounces?

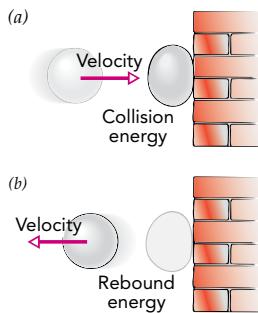


Fig. 3.2.1 A bounce from a wall has two halves: (a) the collision and (b) the rebound. During the collision between the ball and the wall, some of their kinetic energy is transformed into other forms—an amount called the collision energy. During the rebound, some stored energy is released as kinetic energy—an amount called the rebound energy. The rebound energy is always less than the collision energy because some energy is lost as thermal energy. However, a lively ball wastes less energy than a dead one.

The Way the Ball Bounces

What would sports be like without balls? Basketball, baseball, tennis, golf, soccer, volleyball ... the list of sports that depend on them is almost endless—and nearly all those ball sports involve bouncing. While most bounces are obvious, others are more subtle. When a baseball hits a bat or a player kicks a soccer ball, those balls are actually bouncing, albeit from moving objects. Why does a ball bounce, and how do its characteristics and circumstances affect its bounce?

We'll start by looking at a ball's shape. When nothing pushes on the ball, it adopts its equilibrium shape, typically a sphere. The term *equilibrium shape*, of course, is one we've seen already; the previous section used it to describe springs. That's no coincidence, for balls and springs have some important similarities. Chief among those similarities is their ability to store and return energy. Both balls and springs store energy as you distort them away from their equilibrium shapes and release that stored energy when you let them return to their equilibrium shapes.

This energy storage and return is evident when a ball bounces off a hard surface. As those two objects collide (Fig. 3.2.1a), the ball's surface distorts and its elastic potential energy increases. The energy responsible for that distortion is extracted from the kinetic energies of the colliding objects. As the two objects subsequently rebound (Fig. 3.2.1b), the ball's surface returns toward its equilibrium and its elastic potential energy decreases. That return to equilibrium releases energy that is added to the kinetic energies of the rebounding objects.

We'll call the kinetic energy absorbed during the collision the **collision energy** and the kinetic energy released during the rebound the **rebound energy**. How do those two energies compare? Since energy is conserved and the two objects appear unchanged by the bounce, shouldn't the rebound energy equal the collision energy?

In fact, the two energies would be equal except for one important detail: the two objects *have* changed. Their thermal energies have increased. As it distorts and then returns to equilibrium, the ball wastes some of the collision energy as thermal energy, so the rebound energy is always less than the collision energy (Fig. 3.2.2). While total energy is conserved during the bounce, individual forms of energy are not. No physical law prevents kinetic energy from becoming thermal energy, and it happens every bounce.

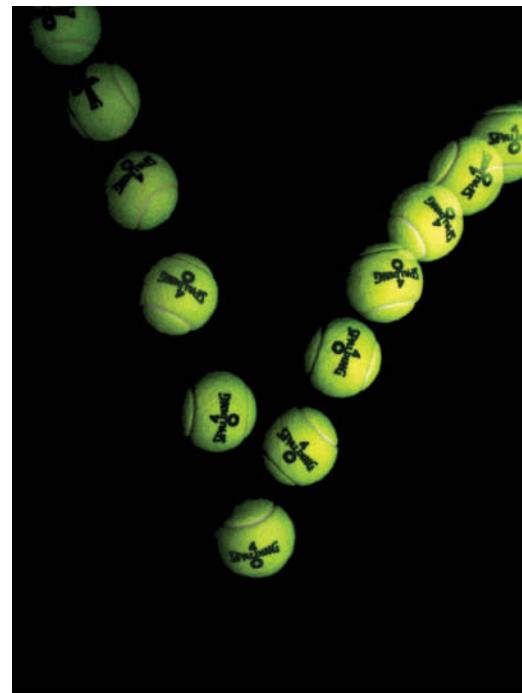


Fig. 3.2.2 When a tennis ball hits the floor, it dents inward to store energy and then rebounds somewhat more slowly than it arrived. These images show the ball's position at 12 equally spaced times. Is the ball bouncing to the left or the right? How can you tell?

Courtesy Lou Bloomfield

TABLE 3.2.1 Approximate Energy Ratios and Speed Ratios for a Variety of Balls

Type of Ball	Rebound Energy Collision Energy	Rebound Speed† Collision Speed
Superball	0.81	0.90
Racquet ball	0.72	0.85
Golf ball	0.67	0.82
Basketball	0.64	0.80
Tennis ball	0.56	0.75
Steel ball bearing	0.42	0.65
Baseball	0.30	0.55
Foam rubber ball	0.09	0.30
“Unhappy” ball	0.01	0.10
Beanbag	0.002	0.04

†Known as the *coefficient of restitution*, this speed ratio can be squared to obtain the energy ratio.

Some balls bounce better than others. A bouncy ball is often called “lively,” and a ball that doesn’t bounce well is said to be “dead.” A lively ball is extremely elastic—it converts most of the collision energy into elastic potential energy during the collision and converts most of that elastic potential energy into rebound energy during the rebound. A dead ball, on the other hand, is barely elastic at all—it converts most of the collision energy into thermal energy during the collision and converts most of what’s left into thermal energy during the rebound.

The ratio of rebound energy to collision energy (Table 3.2.1) determines how high a ball will bounce when you drop it from rest onto a hard floor (Fig. 3.2.3). Before you drop it, the ball has only gravitational potential energy, which is proportional to its height above the floor (Eq. 1.3.2). That potential energy becomes kinetic energy as the ball falls and then collision energy when it collides with the floor. The ball’s rebound energy becomes kinetic energy as it rebounds from the floor and then gravitational potential energy as it rises. At its peak, the ball has only gravitational potential energy, which is again proportional to its height. The ratio of the ball’s rebound energy to its collision energy is therefore equal to the ratio of its final height to its initial height.

For an ideally elastic ball, these ratios are 1.00; all the collision energy is returned as rebound energy and the ball rebounds to its initial height. Known as **elastic collisions**, such perfect bounces are common among the tiny atoms in a gas but are unattainable for ordinary objects; there are just too many ways for any large item to divert or dissipate energy, including as thermal energy, sound, vibration, and light.

A real ball experiences **inelastic collisions**—it wastes some of the collision energy as thermal energy and rebounds to a lesser height. For example, a baseball’s rebound energy is only 0.30 times its collision energy, so it will rebound to only 30% of its original height (Fig. 3.2.3a). Without that dramatic energy loss during a bounce, baseball pitchers would need to wear armor.

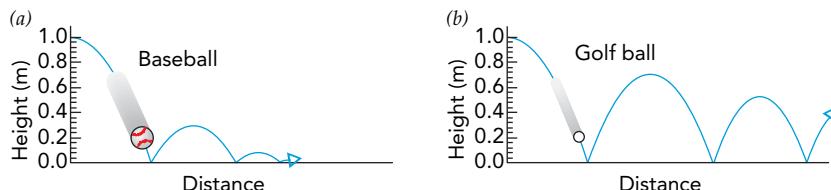


Fig. 3.2.3 (a) A baseball wastes 70% of the collision energy as thermal energy and bounces weakly. (b) In contrast, a golf ball wastes only 30% of the collision energy and bounces well.

While that *energy* ratio is useful, a ball is traditionally characterized by a *speed* ratio: its rebound speed divided by its collision speed. Here, the collision speed is how fast the two objects approach one another before the bounce, and the rebound speed is how fast those objects separate after the bounce. The speed ratio is known as the ball's **coefficient of restitution**:

$$\text{coefficient of restitution} = \frac{\text{rebound speed of ball}}{\text{collision speed of ball}} \quad (3.2.1)$$

(see Table 3.2.1). The larger a ball's coefficient of restitution, the faster it rebounds from a surface for a given collision speed. For example, a basketball's coefficient of restitution is 0.80, so when it bounces off the backboard during a foul shoot, its rebound speed is only 0.80 times its collision speed. That decrease in speed makes it easier for the ball to drop into the basket after it bounces than before it bounces.

Since a ball's kinetic energy is proportional to the square of its speed, its energy ratio is simply the square of its speed ratio. Measuring those ratios is more than an academic exercise because most of the organizations that govern ball sports specify their balls' coefficients of restitution. It's a matter of fairness and safety. For example, baseballs could easily be made so lively that every game would be a home-run derby. Just think of all the asterisks.

Balls bounce best when they store energy through compression rather than through surface bending. That's because most ball materials, such as leather or leather-like plastics, experience lots of wasteful internal friction during bending. Since solid balls involve compression, they usually bounce well, whether they're made of rubber, wood, plastic, or metal. Air-filled balls, however, bounce well only when properly inflated. A normal basketball, which stores most of its energy in its compressed air, has a high coefficient of restitution. In contrast, an underinflated basketball, which experiences lots of surface bending during a collision, barely bounces at all. Similarly, a tennis ball bounces best when new; after a while, the compressed air inside leaks out and the ball's coefficient of restitution drops.

Check Your Understanding #1: A Game of Marbles

As you head into the park to play a game of marbles with your friends, several of the glass marbles fall through a hole in your marble bag and bounce nicely on the granite walkway. How can a marble bounce?

Answer: When the marble collides with the hard granite surface, it dents inward and converts most of its kinetic energy into elastic potential energy. It then rebounds, converting this stored energy back into kinetic energy as it bounces from the surface.

Why: An elastic marble has a very high coefficient of restitution and bounces well from a hard surface.

How the Surface Affects the Bounce

Since the surface on which a ball bounces isn't perfectly hard, that surface also contributes to the bouncing process. It distorts and stores energy when the ball hits it and returns some of this stored energy to the rebounding ball. Overall, the collision energy is shared between the ball and the surface, and each provides part of the rebound energy.

Just how the collision energy is distributed between the surface and ball depends on how stiff each one is. During the bounce, they push on one another with equal but oppositely directed forces. Since the forces denting them inward are equal, the work done in distorting each object is proportional to how far inward it dents. The object that dents farthest receives most of the collision energy.

Courtesy Lou Bloomfield



Since the ball usually distorts more than the surface it hits, most of the collision energy normally goes into the ball. You might expect the ball to provide most of the rebound energy, too. However, that's not always true. Some lively elastic surfaces store collision energy very efficiently and return almost all of it as rebound energy. Since a relatively dead ball wastes most of the collision energy it receives, a lively surface's contribution to the rebound energy can be very important to the bounce.

For example, a lively racket is critical to the game of tennis because the racket's strings provide much of the rebound energy as the ball bounces off the racket (Fig. 3.2.4). The strings are livelier than the ball, so increasing the fraction of collision energy that is invested in the strings actually speeds up the rebounding ball. That's why a player who is willing to sacrifice aiming accuracy for greater ball speed will string his racket relatively loosely. The looser strings will receive more of the collision energy and return almost all of it as rebound energy.

Trampolines and springboards are even more extreme examples, with surfaces so lively that they can make people bounce. People, like beanbags, have coefficients of restitution near zero; when you land on a trampoline, it receives and stores most of the collision energy and then provides most of the rebound energy.

The stiffnesses of the ball and surface also determine how much force each object exerts on the other and thus how quickly the collision proceeds. When both objects are very hard, the forces involved are large and the acceleration is rapid. Thus a steel ball rebounds very quickly from a concrete floor because the two exert enormous forces on one another. If the ball or surface is relatively soft, the forces are weaker and the acceleration is slower.

What if the surface that a ball hits isn't very massive? In that case, the surface may do part or all of the "bouncing." During the bounce, the ball and the surface accelerate in opposite directions and share the rebound energy. Massive surfaces, such as floors and walls, accelerate little and receive almost none of the rebound energy. But when the surface a ball hits is not very massive, you may see it accelerate. For example, when a ball hits a lamp on the coffee table, the ball will do most of the accelerating, but the lamp is likely to fall over, too.

Similarly, when a baseball strikes a baseball bat, the ball and bat accelerate in opposite directions. The more massive the bat, the less it accelerates. To ensure that most of the rebound energy went to the baseball, the legendary hitters of the early twentieth century used massive bats. Such bats are no longer in vogue because they're too difficult to swing. But in the early days of baseball, when pitchers were less skillful, massive bats drove many long home runs.

Check Your Understanding #2: The Game Begins

You are playing the game of marbles on a soft dirt field. The goal is to knock glass marbles out of a circle by hitting them with other marbles. You initially drop several marbles onto the ground inside the circle, and they hardly bounce at all. What prevents them from bouncing well here?

Answer: The soft dirt distorts more than the hard marble and receives most of the collision energy. It converts most of that energy into thermal energy so the marble rebounds weakly.

Why: While a marble bounces nicely on a hard surface, the dirt is soft and receives virtually all the collision energy when the hard marble hits it. The dirt distorts during the impact but stores little energy because it's not very elastic. The marble doesn't rebound much.

How a Moving Surface Affects the Bounce

When a moving ball hits a stationary bat, they bounce. When a moving bat hits a stationary ball, they still bounce. In fact, any time the bat and ball approach one another and collide, it's a bounce and everything we've discussed about bouncing applies. The ball's coefficient of restitution, for example, is still the ratio of the ball's rebound speed to its collision speed.

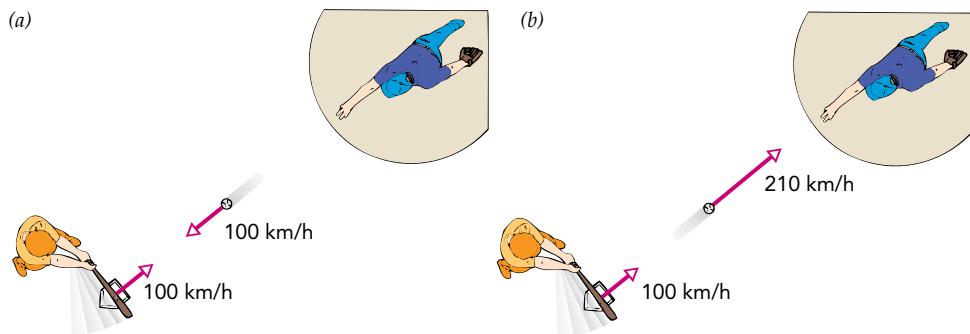


Fig. 3.2.5 (a) Before they collide, the bat and ball are approaching one another at an overall speed of 200 km/h. (b) After the collision, the two are separating from one another at a speed of 110 km/h. However, because the bat is moving toward the pitcher at 100 km/h, the outgoing ball is traveling at 210 km/h in that same direction.

But when the bat is moving before or after the bounce, we clearly have to be careful what we mean by collision speed and rebound speed.

Collision speed is neither the speed of the ball nor the speed of the bat; it's the speed at which those two objects approach one another before the bounce. Similarly, rebound speed is the speed at which the objects separate from one another after the bounce. Technically, both speeds stem from the difference between the bat's velocity and the ball's velocity. However, as long as each object is heading directly toward or away from the other object, collision and rebound speeds are relatively simple to calculate. When the bat and ball are heading in opposite directions, you add their speeds and when they're heading in the same direction, you subtract their speeds.

That simple approach should be familiar from watching cars on the highway. When two cars are moving 60 mph in opposite directions on a narrow road, they are approaching one another at a hair-raising 120 mph. Let's hope they don't collide. When a car traveling 60 mph comes up behind a second car traveling 55 mph in the same direction, however, they are approaching one another at a leisurely 5 mph. If they collide, it'll be only a fender-bender.

We can use these observations about speeds and bouncing to explain why the baseball you just hit is now traveling over the center fielder's head. Let's suppose that, right before the collision (Fig. 3.2.5a), the pitched baseball was approaching home plate at 100 km/h (62 mph) and that, as you swung to meet the ball, your bat was moving toward the pitcher at 100 km/h. Since each object is moving toward the other, their collision speed is the sum of their individual speeds, or 200 km/h (124 mph).

The baseball's coefficient of restitution is 0.55, so after the collision (Fig. 3.2.5b) their rebound speed was only 0.55 times their collision speed, or 110 km/h. The outgoing ball and the swinging bat separate from one another at 110 km/h. Since the bat is still moving toward the pitcher at 100 km/h, the ball must be traveling toward the pitcher even faster: at 100 km/h plus 110 km/h or a total speed of 210 km/h (130 mph)! That's why the baseball flies past everyone in the outfield and into the stands.



Check Your Understanding #3: Marble Frames of Reference

Two of you flick your marbles into the circle simultaneously from opposite sides of the circle, and they collide head-on. Each marble was traveling forward at 1 m/s (3.3 ft/s). How quickly were the two marbles approaching one another just before they hit?

Answer: They were approaching at 2 m/s (6.6 ft/s).

Why: The velocities reported in the question are those observed by people sitting still with respect to the circle. Since the marbles are heading toward one another, you must add their speeds to obtain the collision speed.

Surfaces Also Bounce ... and Twist and Bend

We've seen how a surface affects a ball when they bounce. Now let's look at how the ball affects the surface. When you swing your bat into a pitched ball, the bat doesn't continue on exactly as before. The ball pushes on the bat during the collision, and the bat responds in a number of interesting ways.

First, the ball's impact force causes the bat to accelerate backward during the collision, so its speed toward the pitcher decreases slightly. Since the ball finishes its rebound from the slower-moving bat, the ball doesn't bounce back toward the pitcher quite as fast it would if the bat had maintained its speed. Increasing the bat's mass reduces this slowing effect, but it also makes the bat harder to swing.

Second, the ball's impact force can also produce a torque on the bat about its center of mass, so that the bat undergoes angular acceleration (Fig. 3.2.6). Together, the bat's acceleration and angular acceleration produce interesting effects on the bat's handle. The bat's acceleration tends to jerk its handle away from the pitcher, while its angular acceleration tends to jerk its handle toward the pitcher. The extent of these two opposing motions depends on just where the ball hits the bat. If the ball hits the bat's **center of percussion**, the two motions cancel and the bat handle moves forward steadily (Fig. 3.2.6c). The smooth feel of such a collision explains why the center of percussion, located about 7 inches from the end of the bat, is known as a "sweet spot."

Third, the ball's impact often causes the bat to vibrate. Like a xylophone bar struck by a mallet (Fig. 3.2.7a), the bat bends back and forth rapidly with its ends and center moving in opposite directions (Fig. 3.2.7b). These vibrations sting your hands and can even break a wooden bat. However, near each end of the bat, there's a point that doesn't move when the bat vibrates—a **vibrational node**. When the ball hits the bat at its node, no vibration occurs. Instead, the bat emits a crisp, clear "crack" and the ball travels farther. Fortunately, the bat's vibrational node and its center of percussion almost coincide, so you can hit the ball with both sweet spots at once.

As these handle motions and bending vibrations illustrate, the science and engineering of bats are surprisingly complicated. That's why bat manufacturers are forever developing better, more potent ones. Multiwalled aluminum, titanium, and composite bats are the latest examples. Each of these hollow bats has a thin outer wall that dents considerably on impact with the ball and therefore receives much of the collision energy. Because that wall is also highly elastic, nearly all this energy emerges as rebound energy as the wall returns to its equilibrium shape. As a bonus, the bat's light outer wall and massive inner wall ensure that it retains very little energy in its vibrations following the bounce.

Unlike the hard surface of a wooden bat, which barely dents and thus barely participates in the energy storage process, one of these hi-tech bats acts as a trampoline—it stores and returns so much of the collision energy that it substantially increases the outgoing speed of the batted ball. You can find equally hi-tech equipment at the golf shop or tennis store. We're in an era in which scientific analysis and design are radically altering sports.

Check Your Understanding #4: Mass and Marbles

The marbles you're playing with are not all the same size and mass. You notice that larger marbles are particularly effective at knocking other marbles out of the circle. You decide to use a 10-cm-diameter glass ball as a marble, expecting to clean out the entire circle. But when you flick it with your thumb, your thumb merely bounces off. Why doesn't the glass ball move forward quickly?

Answer: The glass ball is so much more massive than your thumb that your thumb receives almost all the rebound energy. Your thumb bounces, not the glass ball.

Why: In any collision, it's the least massive object that experiences the greatest acceleration and that receives the largest share of the rebound energy. The effect is similar to what would happen if you swung a light aluminum baseball bat at a pitched bowling ball. The bat would rebound wildly but the bowling ball would continue to travel toward the catcher. This same effect is true in automobile collisions, where a massive sedan is much less disturbed than the tiny subcompact with which it collides.

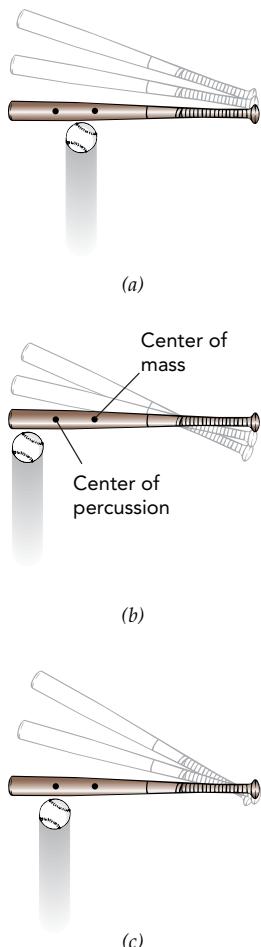


Fig. 3.2.6 When a ball hits a bat, the bat experiences both acceleration and angular acceleration. (a) If the ball hits near the bat's center, the angular acceleration is small and the bat's handle accelerates backward. (b) If the ball hits near the bat's end, the angular acceleration is large and the bat's handle accelerates forward. (c) But if the ball hits the bat's center of percussion, the angular acceleration is just right to keep the handle from accelerating.

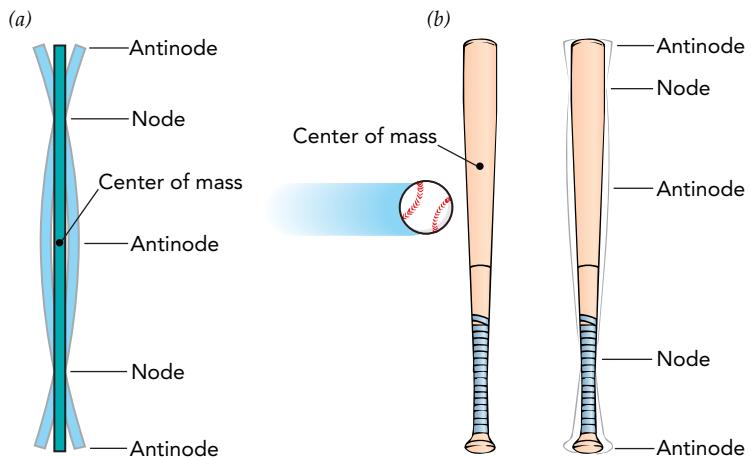
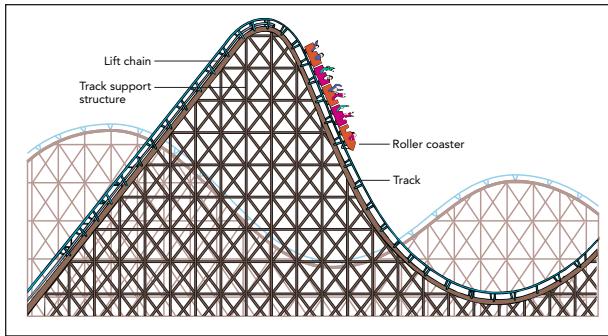


Fig. 3.2.7 (a) When struck by a mallet, a xylophone bar vibrates with its middle and ends moving back and forth in opposite directions. The parts that move farthest are antinodes, and the points that don't move at all are nodes. (b) When struck by a ball, a baseball bat vibrates in a similar fashion. However, an impact at one of the bat's nodes causes no vibration.

SECTION 3.3

Carousels and Roller Coasters



As your sports car leaps forward at a green light, you're pressed firmly back against your seat. It feels as though gravity were somehow pulling you down and backward at the same time. It's not gravity pulling backward on you, though; it's your own inertia trying to prevent you from accelerating forward with the car.

When this happens, you're experiencing the feeling of acceleration. We encounter this feeling many times each day, whether through turning in an automobile or riding up several floors in a fast elevator. Nowhere are the feelings of acceleration more acute than at the amusement park. We accelerate up, down, and around on the carousel, back and forth in the bumper cars, and left and right in the scrambler. The ultimate ride, of course, is the roller coaster, which is one big, wild experience of acceleration. When you close your eyes on a straight stretch

of highway, you can hardly tell the automobile is moving. When you close your eyes on a roller coaster, however, you have no trouble feeling every last turn in the track. It's not the speed you feel, but the acceleration. What is often called motion sickness should really be called acceleration sickness.

Questions to Think About: How does your body feel its own weight? When you swing a bucket full of water around in a circle, as I asked you to do in the chapter opening, why does the bucket pull outward on you? Why can you swing that bucket completely over your head without spilling the water inside it? What keeps you from falling out of a roller coaster as it goes over the top of a loop-the-loop? Which car of a roller coaster should you sit in to experience the best ride?

Experiments to Do: To begin associating the familiar sensations of motion with the physics of acceleration, travel as a passenger in a vehicle that makes lots of turns and stops. Close your eyes and find whether you can tell which way the vehicle is turning and when it's starting or stopping. Which way do you feel pulled when the vehicle turns left? turns right? starts? stops? How is this sensation related to the direction of the vehicle's acceleration? Now during a time when the vehicle is traveling at constant velocity on a level path, find if you feel any sensations that tell you which way it's heading. Try to convince yourself that it's heading backward or sideways rather than forward. Which is easier to feel: your acceleration or your velocity?

Carousels and Acceleration

While roller coasters offer interesting visual effects, such as narrowly missing obstacles, and strange orientations, such as running upside down, the real thrill of roller coasters

comes from their accelerations. Plenty of other amusement park rides suspend you sideways or upside down so that you feel ordinary gravity pulling at you from unusual angles. But why pay for those when you can stand on your head for free? For a *real* thrill, you need acceleration to give you the weightless feeling you experience as a roller coaster dives over its first big hill or the intense heaviness you feel as it whips you around a sharp corner. Why do accelerations give rise to these wonderful sensations, and why are they most extreme when you're sitting in the last car of the roller coaster?

To answer those questions, we'll need to take another look at acceleration. Although you may start your day at an amusement park by riding its biggest roller coaster, some of us might benefit from a warm-up. Let's start with a simpler ride, a carousel.

When you ride on a carousel, you travel in a *circular path* around a central pivot. That's not the motion of an object that's simply coasting forward and exhibiting inertia. If you were experiencing zero net force, you would travel in a straight line at a steady pace in accordance with Newton's first law. However, since your path is circular instead of straight, your direction of travel is changing; you are accelerating and must be experiencing a non-zero net force.

Which way are you accelerating? Remarkably, you are always accelerating toward the center of the circle. To see why that's so, let's look down on a simple carousel that's turning counterclockwise at a steady pace (Fig. 3.3.1). At first, the boy riding the carousel is directly east of its central pivot and is moving northward (Fig. 3.3.1a). If nothing were pulling on the boy, he would continue northward and fly off the carousel. Instead, he follows a circular path by accelerating toward the pivot—that is, toward the west. As a result, his velocity turns toward the northwest and he heads in that direction. To keep from flying off the carousel, he must continue to accelerate toward the pivot, which is now southwest of him (Fig. 3.3.1b). His velocity turns toward the west, and he follows the circle in that direction. And so it goes (Fig. 3.3.1c).

The boy's body is always trying to go in a straight line, but the carousel keeps pulling him inward so that he accelerates toward the central pivot. The boy is experiencing **uniform circular motion**. *Uniform* means that the boy is always moving at the same speed, although his direction keeps changing. *Circular* describes the path the boy follows as he moves, his trajectory.

Like any object undergoing uniform circular motion, the boy is always accelerating toward the center of the circle. An acceleration of this type, toward the center of a circle, is called a **centripetal acceleration** and is caused by a centrally directed force, a **centripetal force**. A centripetal force is not a new, independent type of force like gravity but the net result of whatever forces act on the object. *Centripetal* means “center-seeking,” and a centripetal force pushes the object toward that center. The carousel uses support forces and friction to exert a centripetal force on the boy, and he experiences a centripetal acceleration. Amusement park rides often involve centripetal acceleration (Fig. 3.3.2).

The amount of acceleration the boy experiences depends on his speed and the radius of the carousel. The faster the boy is moving and the smaller the radius of his circular trajectory, the more he accelerates. His acceleration is equal to the square of his speed divided by the radius of his path.

We can also determine the boy's acceleration from the carousel's angular speed and its radius. The faster the carousel turns and the larger the radius of the boy's circular trajectory, the more he accelerates. His acceleration is equal to the square of the carousel's angular speed times the radius of his path. We can express these two relationships as word equations:

$$\text{acceleration} = \frac{\text{speed}^2}{\text{radius}} = \text{angular speed}^2 \cdot \text{radius}, \quad (3.3.1)$$

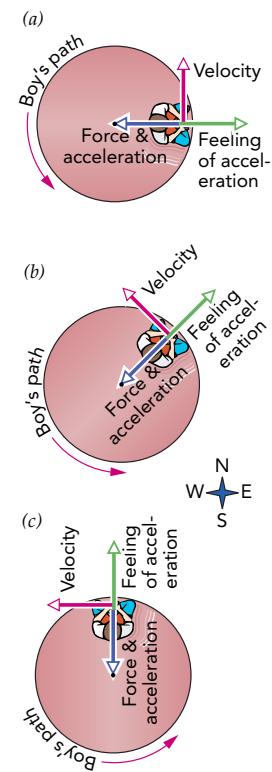


Fig. 3.3.1 A boy riding on a turning carousel is always accelerating toward the central pivot. His velocity vector shows that he is moving in a circle, but his acceleration vector points toward the pivot. When he is heading north (a), he is accelerating toward the west. His velocity gradually changes direction until he is heading northwest (b), at which time he is accelerating toward the southwest. He turns further until he is heading west (c) and is then accelerating toward the south. (North is upward.)



©Andrew Scott, Brisbane, Australia/Getty Images

Fig. 3.3.2 The people on this ride travel in a circle, pushed inward by the wall behind them so that they accelerate toward the center of the circle.

in symbols:

$$a = \frac{v^2}{r} = \omega^2 \cdot r,$$

and in everyday language:

Making a tight, high-speed turn involves lots of acceleration.

COMMON MISCONCEPTIONS: Centrifugal Force

Misconception: As you swing an object in a circle, that object is pulled outward by “centrifugal force.”

Resolution: If the object were free of forces, inertia would cause it to travel in a straight line. But because you are pulling the object inward, it accelerates inward and its path bends into a circle. Since you are exerting an inward force on the object, it is exerting an outward force on you, in accordance with Newton’s third law. However, that outward force acts on you, not on the object. There is no outward “centrifugal force” acting on the object itself.



Check Your Understanding #1: Banking on a Curve

Your racing car comes to a banked, left-hand turn, which it completes easily. Why is it essential that a racetrack turn be banked so that it slopes downhill toward the center of the turn?

Answer: A banked turn is needed so that the support force exerted by the racetrack on the car’s wheels can provide at least some of the centripetal force needed to accelerate the car around the turn.

Why: As a car and driver travel around a circular turn, they are accelerating toward the center of the track and require a huge centripetal force inward. On a level track, the only horizontal force available is static friction between the ground and the car’s tires. If static friction is unable to provide enough inward force, the car will skid off the track, following a straight-line path. This type of accident is typical of a highway curve on an icy day and is why designers bank the curves. The banks are ramps sloping down toward the center so that the horizontal component of the support force exerted by the track on the car’s wheels provides an additional, inward, centripetal force to help that car accelerate around the curve.

Check Your Figures #1: Going for a Spin

Some children are riding on a playground carousel with a radius of 1.5 m (4.9 ft). The carousel turns once every 2 s. How quickly are the children accelerating?

Answer: They are accelerating about 15 m/s^2 (50 ft/s^2).

Why: Because the children are in uniform circular motion, their acceleration is given by Eq. 3.3.1. The carousel turns once every 2 s, so its angular velocity is 2π radians divided by 2 s. Omitting *radians* because they're the natural unit of angle, the carousel's angular velocity is π 1/s. Since its radius is 1.5 m, the children's acceleration is

$$\text{children's acceleration} = (\pi \text{ 1/s})^2 \cdot 1.5 \text{ m} = 14.8 \text{ m/s}^2.$$

Since our measurements of the carousel's radius and its turning time are accurate only to about 10%, our calculation of the children's acceleration is accurate only to about 10%. We report it as 15 m/s^2 (50 ft/s^2).

The Experience of Acceleration

Nothing is more central to the laws of motion than the relationship between force and acceleration. Until now, we've looked at forces and noticed that they can cause accelerations; now let's take the opposite perspective, looking at accelerations and noticing that they require forces. For you to accelerate, something must push or pull on you. Just where and how that force is exerted on you determines what you feel when you accelerate.

The backward sensation you feel as your car accelerates forward is caused by your body's inertia, its resistance to acceleration (Fig. 3.3.3). For you to accelerate forward with the car, something must push you forward; otherwise, inertia will keep you in steady motion as the accelerating car drives out from under you. Fortunately, your seat provides you with that forward push. But the seat can't exert a force uniformly throughout your body. Instead, it pushes only on your back, and your back then pushes on your bones, tissues, and internal organs to make them accelerate forward. Each piece of tissue or bone is responsible for the forward force needed to accelerate forward the tissue in front of it. A whole chain of forces, starting from your back and working forward toward your front, causes your entire body to accelerate forward.

Let's compare this situation with what happens when you're standing motionless on the floor. Since gravity exerts a downward force on you that's distributed uniformly throughout your body, each part of your body has its own independent weight; these individual weights, taken together, add up to your total weight. The floor, for its part, is exerting an upward support force on you that keeps you from accelerating downward through its surface. But the floor can't exert a force uniformly throughout your body. Instead, it pushes only on your feet, and your feet then push on your bones, tissues, and internal organs to keep them from accelerating downward. Each piece of tissue or bone is responsible for the upward force needed to keep the tissue above it from accelerating downward. A whole chain of forces, starting from your feet and working upward toward your head, keeps your entire body from accelerating downward.

As you probably noticed, the two previous paragraphs are very similar—and so are the sensations of gravity and acceleration. When the ground is preventing you from falling, you feel "heavy"; your body senses all the internal forces needed to support its pieces so that they don't accelerate, and you interpret these sensations as weight. When the car seat is causing you to accelerate forward, you also feel "heavy"; your body senses all the internal forces needed to accelerate its pieces forward, and you interpret these sensations as weight. This time you experience the weight-like sensation toward the back of the car.

Try as you may, you can't distinguish the weight-like sensation that you experience as you accelerate from the weight sensation you experience due to gravity. And you're not the only one fooled by acceleration. Even the most sophisticated laboratory instruments can't

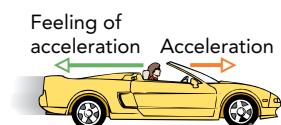


Fig. 3.3.3 As you accelerate forward in a car, you sense a gravity-like feeling of acceleration in the direction opposite to the acceleration. This feeling of acceleration is really the mass of your body resisting acceleration.

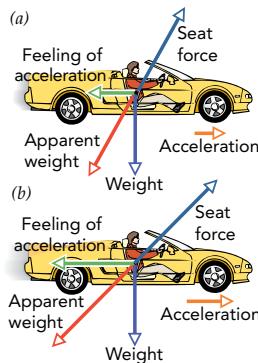


Fig. 3.3.4 (a) When you accelerate forward gently, the backward feeling of acceleration is small and your apparent weight is mostly downward. Your apparent weight is equal but opposite to the force your seat exerts on you. (b) When you accelerate forward quickly, you experience a strong backward feeling of acceleration and your apparent weight is backward and down.

determine directly whether they are experiencing gravity or are accelerating. However, despite the convincing sensations, the backward heavy feeling in your gut as you accelerate forward is the result of inertia and is not due to a real backward force.

We'll call this experience a **feeling of acceleration**. It always points in the direction opposite the acceleration that causes it, and its strength is proportional to that acceleration. When you turn your car to the left, you're accelerating leftward and experience a rightward feeling of acceleration. The tighter your turn, the stronger that feeling of acceleration becomes. When you're not driving and can safely close your eyes, you should be able to sense both the direction and amount of each acceleration you make.

So what does the boy on the spinning carousel feel? Since he is accelerating inward, toward the center of the circle, he experiences a feeling of acceleration outward, away from the center of the circle. Although the boy's actual weight can't change, his feeling of acceleration can have any strength. If the carousel is spinning fast enough, he can experience a feeling of acceleration that is much stronger than his weight. That's why he's holding on tight!

Since you often compare your feelings of acceleration to your feeling of weight, it is customary to measure feelings of acceleration in g's, where 1 g (1 gravity) is equal to the feeling of your actual weight. In this context, it is also customary to measure acceleration in g's, where 1 g is 9.8 m/s^2 (32 ft/s^2), the acceleration due to gravity. These two definitions lead to a simple observation: a 1-g acceleration produces a 1-g feeling of acceleration in the opposite direction. For example, a 1-g inward acceleration on a playground carousel produces a 1-g outward feeling of acceleration. If you accelerate five times that quickly, on the scrambler or on an airplane maneuvering sharply, your 5-g acceleration will produce a 5-g feeling of acceleration in the opposite direction.

As we've seen, you experience a backward feeling of acceleration when you accelerate forward in your car. However, you don't experience this feeling of acceleration all by itself; you also experience the downward feeling of your weight. Together these two effects feel like an especially strong weight at an angle somewhere between straight down and the back of the car (Fig. 3.3.4a). We'll call the combined feelings of weight and acceleration your **apparent weight**. The faster you accelerate forward, the stronger your apparent weight becomes and the more it points toward the back of the car (Fig. 3.3.4b).

You feel apparent weight because the seat is pushing on you, both to support your actual weight and to make you accelerate. In fact, your apparent weight is equal but opposite to the real force that the seat is exerting on you. As you accelerate forward in your car, the seat pushes you upward and forward, and your apparent weight is downward and backward (Fig. 3.3.4).

Like feeling of acceleration, apparent weight is customarily measured in g's. When you slam the car's accelerator pedal to the floor and accelerate forward rapidly, your apparent weight may reach 2 g. The seat is then pushing you forward and upward with a force that's twice as large as your actual weight, and you're pushing back just as hard. No wonder you're pressed tightly into the seat!

Check Your Understanding #2: The Feel of a Tight Turn

You're sitting in the passenger seat of a racing car that is moving rapidly along a level track. The track takes a sharp turn to the left, and you find yourself thrown against the door to your right. What horizontal forces are acting on you, and what feelings of acceleration do you experience?

Answer: The car seat and door are exerting a leftward force on you, causing you to accelerate leftward with the car. You also experience a rightward feeling of acceleration as your inertia acts to keep you from accelerating leftward.

Why: As the car turns left, it accelerates toward the left. The car seat and right door together exert a leftward force on you to prevent the car from driving out from under you. Because it has mass, your body resists this leftward acceleration. You feel your body trying to go in a straight line, which would carry it out the right door of the car as the car accelerates toward the left.

© Harry DiOrio/The Image Works

Roller Coaster Acceleration

We're now prepared to look at a roller coaster and understand what you feel when you go over hills (Fig. 3.3.5) and loop-the-loops (Fig. 3.3.6). Every time the roller coaster accelerates, you experience a feeling of acceleration in the direction opposite your acceleration. That feeling of acceleration gives you an apparent weight that's different from your actual weight.

As we saw with a car, rapid horizontal accelerations tip your apparent weight in the direction opposite your acceleration. A roller coaster, however, can do something a car normally can't—it can accelerate downward! In that case, the feeling of acceleration you experience is upward and opposes your downward feeling of weight. The two feelings at least partially cancel, so that your apparent weight points weakly downward or perhaps, if the downward acceleration is fast enough, points upward.

There is one more possibility: if you accelerate downward at just the right rate, your upward feeling of acceleration will exactly cancel your downward feeling of weight. Your apparent weight will then be zero, and you'll feel perfectly weightless. Since your feeling of weight is 1 g downward, perfect cancellation occurs when your feeling of acceleration is 1 g upward. You must be accelerating downward at 1 g; you must be in *free fall!*

When you're falling, you feel perfectly weightless. Your apparent weight is zero, and nothing is supporting you. Because your hat and sunglasses are falling with you, nothing is supporting them, either. If they come off your head, they will hover around you as you all fall together.

Similarly, your internal organs don't need to support one another as they fall, and the absence of internal support forces gives rise to the exhilarating sensation of free fall. Our bodies are very sensitive to even partial weightlessness, and this falling sensation is half the fun of a roller coaster. An astronaut falling freely in space has this disquieting weightless feeling for days on end. No wonder astronauts have such frequent troubles with motion (or, rather, acceleration) sickness.

Because a roller coaster is attached to a track, however, its rate of downward acceleration can actually exceed that of a freely falling object. In those special situations, the track will be assisting gravity in pushing the roller coaster downward. As a rider, you'll be pushed downward, too, and your apparent weight will point upward, as though the world had turned upside down!



Fig. 3.3.5 As this roller coaster plunges down the first hill, its last car is pulled over the edge by the cars in front of it and its riders feel almost weightless.

© Charles V. Angelo/Photo Researchers/Getty Images, Inc.



Fig. 3.3.6 As it goes over the loop-the-loop, this roller coaster is accelerating rapidly toward the center of the circle. When it reaches the top of the loop, the track pushes the roller coaster downward and the riders feel pressed into their seats. If they close their eyes, they won't even be able to tell that they're upside down.

Check Your Understanding #3: Drop Tower Panic

A drop tower is a terrifying amusement park ride in which you are strapped into a seat, lifted high in the air, and then dropped. While you're in free fall, what is your apparent weight?

Answer: Your apparent weight is zero.

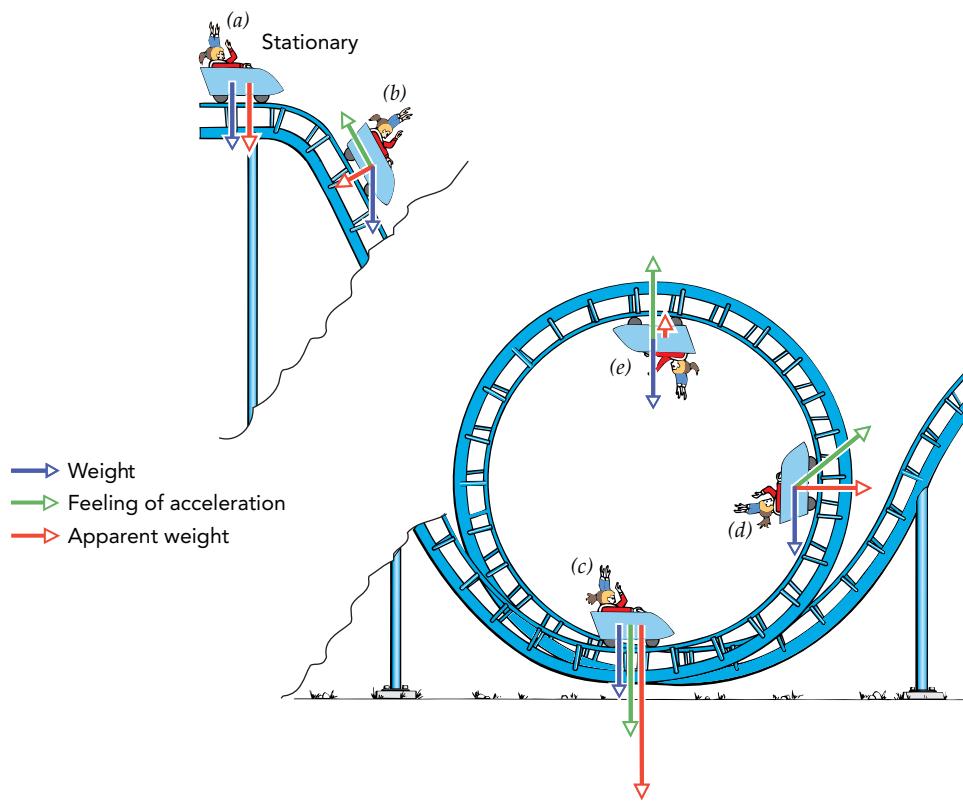
Why: As you fall freely, you are accelerating downward at 1 g, the full acceleration due to gravity. The 1-g upward feeling of acceleration you experience exactly matches your downward weight, so that your apparent weight is zero. You feel perfectly weightless.

Roller Coaster Loop-the-Loops

Figure 3.3.7 shows a single-car roller coaster at various points along a simple track with one hill and one loop-the-loop. You are riding in the car, and your weight, feeling of acceleration, and apparent weight are indicated with arrows of varying lengths that show each vector quantity's direction and magnitude. The longer the arrow, the greater the magnitude of the quantity that it represents.

At the top of the first hill (Fig. 3.3.7a), the single-car roller coaster is almost stationary. You feel only your weight, straight down—nothing exciting yet. However, as soon as the car begins its descent, accelerating down the track, a feeling of acceleration appears

Fig. 3.3.7 A single-car roller coaster going over the first hill and a loop-the-loop. At each point along the track, the car experiences its weight, a feeling of acceleration due to its current acceleration, and an apparent weight that is the sum of those two. The apparent weight always points toward the track, and the car doesn't fall off it.



pointing up the track (Fig. 3.3.7b). Your feelings of weight and acceleration together give you an apparent weight that is much less than 1 g and points down and into the track. Most people find this sudden reduction in apparent weight to be terrifying, so you're welcome to scream!

On Earth, weightless feelings can't last. They occur only during downward acceleration and disappear as your car levels off near the bottom of the hill. By the time the car begins its rise into the loop-the-loop, it is traveling at maximum speed and has begun to accelerate upward (Fig. 3.3.7c). This upward acceleration of about 2 g creates a downward feeling of acceleration of about 2 g. Your apparent weight is about 3 g downward, and you're pressed tightly into your seat.

As the car climbs the right side of the loop-the-loop against gravity, its speed decreases somewhat and its path bends inward, toward the center of the loop. Halfway up the right side, you're accelerating inward and downward, so your feeling of acceleration is outward and upward (Fig. 3.3.7d). Your apparent weight is greater than 1 g outward, away from the center of the loop, and you're still pressed into your seat (Fig. 3.3.8).

At the top of the loop-the-loop (Fig. 3.3.7e), the car's path is still bending inward, toward the center of the loop. You're accelerating downward, and your feeling of acceleration points in the opposite direction—upward toward the sky. That upward feeling of acceleration is greater than your downward feeling of weight, so your apparent weight is upward (Fig. 3.3.6)!

Not only does the inverted car stay on its track, but you also remain pressed into your seat. Actually, the car is pushing downward on you to help gravity accelerate you around the loop. If your hat were to come off at the top of the loop, it would land in your seat, even though that seems to involve some sort of upward movement. In fact, the car is accelerating downward at more than 1 g, while your hat has only the 1-g downward acceleration of a freely falling object. Gravity and the track together push the car downward so fast that it overtakes the hat—your hat is falling, but the car is plummeting even faster.



© Digital Vision/Getty Images, Inc.

Fig. 3.3.8 The roller coaster pushes these riders inward as it travels around the loop-the-loop. The riders experience strong feelings of acceleration outward and can hardly tell that they're upside down.

In truth, a typical loop-the-loop isn't perfectly circular; it's more sharply curved on top than on its sides or bottom. This varying radius curve, known as a clothoid, is chosen for safety and comfort. By sharpening the curve only on top, the clothoid track maximizes the downward acceleration there while reducing the accelerations elsewhere on the track. High acceleration is important only when the roller coaster is upside down. Everywhere else, it simply makes the riders feel heavy and uncomfortable, particularly at the bottom of the loop where the coaster is traveling fastest and accelerating upward rapidly.

Most roller coaster tracks are designed so that their riders are always pressed into their seats, even when the cars go upside down. On loop-the-loops, for example, these tracks ensure that the riders are accelerating inward rapidly enough to remain seated securely. In principle, roller coasters that travel on such tracks don't need seat belts to prevent their riders from falling out (although seat belts are comforting to the passengers and insurance companies). Their riders' apparent weights are always directed toward the track, so the riders and their seats push against one another throughout the ride.

However, some roller coaster tracks use special cars and restraints that allow them to direct their riders' apparent weights away from the track. At such moments, the cars must pull on their riders to keep them from popping out of their seats. These roller coasters can and do go upside down without the strong downward accelerations needed to press their riders into their seats. The riders are then hanging from their inverted cars, and if one of them loses a hat, it falls to the ground rather than into the car.

What about a roller coaster with more than one car? For the most part, the same rules apply. However, new forces now act on each car—forces exerted by the other cars in the train. The effects of these cars are most pronounced at the top of the first and biggest hill. As the train disconnects from the lift chain and approaches the descent, it is rolling forward slowly and the first cars are well over the crest of the hill before they pick up much speed (Fig. 3.3.9a). They're pulling hard on the cars behind them, and those cars are pulling back, slowing their descent. By the time the train is moving fast, the first car is well down the hill and the track is beginning to turn upward. The first car's riders experience mostly upward acceleration and downward feeling of acceleration. That's why riders in the first few cars of a roller coaster don't feel much weightlessness.

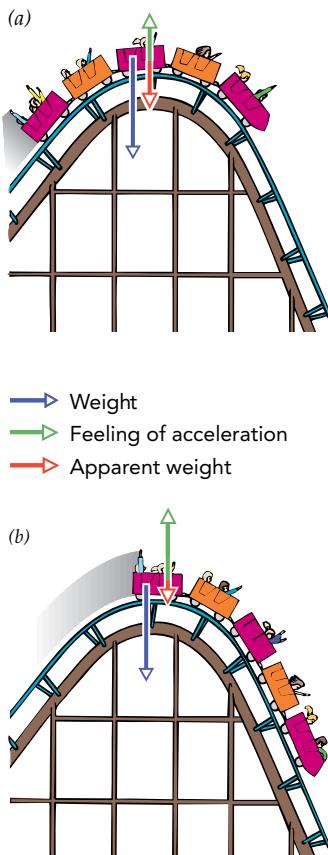


Fig. 3.3.9 When a multicar roller coaster descends the first hill, the ride experienced in the first cars is different from that in the last cars. (a) The first cars travel over the crest of the hill slowly and reach high speed only well down the hill. The cars behind them slow their descent. (b) The last cars are whipped over the top and are traveling very rapidly early on. The last cars accelerate downward dramatically as they go over the first hill and their riders experience a strong feeling of weightlessness.

In contrast, the last car is moving at high speed early in its descent. It undergoes a dramatic downward acceleration as it's yanked over the crest of the first hill by the cars in front of it (Fig. 3.3.9b). As a result, its riders experience large upward feelings of acceleration and quite extreme weightlessness. In fact, the designers of the track must be careful not to make the downward acceleration too rapid or the roller coaster will flick the riders in the last car right out of their seats. Each time the last car crests a hill, it accelerates downhill in a way that no other car on the coaster can match.

Obviously, it does matter where you sit on a roller coaster. The first seat offers the most exciting view, but it provides less than spectacular weightless feelings. The last car almost always offers the best weightless feelings. Probably the dullest seat in the roller coaster is the second; it offers a relatively tame ride and an unchanging view of the people in the front seat.

Check Your Understanding #4: Taking to the Air

Your racing car travels over a bump in the track and suddenly becomes airborne. What keeps you in the air? Has gravity disappeared?

Answer: Gravity is still present, but your inertia prevents you from following a rapid downturn in the surface you are traveling along.

Why: If the road you are traveling along suddenly turns downward, you must accelerate downward to stay in contact with its surface. The steeper and more abrupt the descent, the more downward acceleration you need. The only downward force you experience is your weight, which can cause a downward acceleration of no more than 9.8 m/s^2 (32 ft/s^2). If the surface drops out from under you faster than that, you will become airborne. You will then be falling freely and accelerating downward as fast as gravity will permit. Eventually, you will fall to the surface. In many sports, including skiing, motorcycle racing, and skateboarding, a person traveling along an uneven surface becomes airborne after passing over a bump.

Epilogue for Chapter 3

In this chapter we have looked at the physical concepts involved in four types of simple machines. In Spring Scales, we explored the relationship between the force acting on a spring and its distortion, and we examined how this distortion can be used to measure an object's weight. In Ball Sports: Bouncing, we examined the process of storing and releasing kinetic energy during a collision. As we saw, both the ball and the object it hits contribute to the bounce.

In Carousels and Roller Coasters, we explored both the sensations we feel as we accelerate and why those sensations occur. We saw that circular motion involves a centrally directed acceleration that gives rise to an outwardly directed feeling of acceleration.

Explanation: Swinging Water Overhead

As you swing the bucket over your head, you are pulling downward on it and causing it to accelerate downward very rapidly. The water remains in the inverted bucket because the bucket is accelerating downward faster than gravity alone can accelerate the water. As the water tries to fall, the bucket overtakes the falling water. As a result, the water is pressed into the bottom of the bucket. The same effect occurs when you throw a book toward the floor by pushing it downward rapidly with your open palm. As the book tries to fall, your palm overtakes it, and it remains pressed into your palm because your hand is accelerating it downward faster than gravity. Finally, if you stop swinging the bucket too abruptly, the bucket's contents will not decelerate with it. Instead, they will spill or smash.

Chapter Summary and Important Laws and Equations

How Spring Scales Work: A spring scale measures an object's weight by supporting that object with a spring. When the object is at rest, the spring's upward force exactly balances the object's downward weight; the scale then reports the upward restoring force its spring is exerting on the object. As Hooke's law describes, that restoring force is proportional to the spring's distortion, so the scale can determine it by measuring how far the spring has bent. This measurement is often done mechanically and is reported using a needle or dial.

How Bouncing Balls Work: A ball behaves like a spherical or oblong spring. It stores elastic potential energy as it distorts away from its equilibrium shape and releases some of that energy as it returns to normal. When a ball strikes a surface, some kinetic energy is removed from the ball and the surface, and is either stored within those objects as elastic potential energy or lost as thermal energy. As the objects rebound, some of the stored energy becomes kinetic energy again. The kinetic energy returned—the rebound energy—is always less than the kinetic energy initially removed from the objects—the collision energy. The missing energy has been converted into thermal energy.

How Carousels and Roller Coasters Work: A carousel uses centripetal acceleration to give each rider an outward feeling of acceleration. Combined with weight, this gravity-like sensation gives the rider an apparent weight that points downward and outward.

A roller coaster also uses rapid acceleration to create unusual apparent weights for its riders. Each time the coaster accelerates on a hill or turn, the rider experiences a feeling of acceleration in the direction opposite the acceleration. This feeling of acceleration, combined with the rider's weight, creates an apparent weight that varies dramatically in amount and direction throughout the ride. It's this fluctuating apparent weight, particularly its near approaches to zero, that make riding a roller coaster so exciting.

1. Hooke's law: The restoring force exerted by an elastic object is proportional to how far it is from its equilibrium shape, or

$$\text{restoring force} = -\text{spring constant} \cdot \text{distortion}. \quad (3.1.1)$$

2. Acceleration of an object in uniform circular motion: An object in uniform circular motion has a centripetal acceleration

equal to the square of its speed divided by the radius of its circular trajectory, which is equal to the square of its angular speed times the radius of its circular trajectory, or

$$\text{acceleration} = \frac{\text{speed}^2}{\text{radius}} = \text{angular speed}^2 \cdot \text{radius}. \quad (3.3.1)$$

Mechanical Objects

PART 2

No matter how sophisticated the machines around us appear, most are based in large part on the simple principles that we have already encountered. In this chapter, we take a look at two more fascinating machines and see what makes them tick. As we do, we'll find ourselves revisiting familiar issues and exploring a few new ones all the way to the frontiers of science and the cosmos.

ACTIVE LEARNING EXPERIMENTS

High-Flying Balls

Among the physical issues that we discuss in this chapter are the reaction effects that push rockets forward. These reaction effects appear in an interesting way in a simple experiment that involves two different-size balls: a basketball and a tennis ball.

If you can't find a basketball and a tennis ball, any two lively balls of very different masses will do. Drop the balls separately and see how high they bounce. You'll immediately notice that an individual ball can't bounce higher than the point from which you dropped it. Such a rebound would give it more energy than it had originally. In fact, some of the ball's energy will be lost to thermal energy, so it will not even reach its original height.

But what will happen when you stack the smaller ball on top of the larger ball and drop the two balls together? Think about the sequence of events that will occur, and then give it a try. Make sure that the small ball remains directly above the large ball as the two balls fall to the floor.

Why does the smaller ball rebound as it does? How does your choice of balls influence this effect? If you put the smaller ball on the bottom, will that change the outcome? What about dropping the pair of balls from a different height? Try the same experiment with balls of various sizes to gather more insight into what is happening.



Courtesy Lou Bloomfield



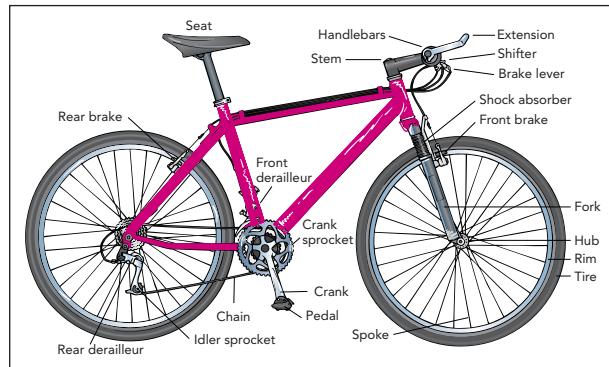
Chapter Itinerary

The objects in this chapter combine many of the concepts we have discussed and add a few new ones of their own. These objects are (1) *bicycles* and (2) *rockets*. In Bicycles, we see how motion and clever design can make an apparently unstable vehicle stable enough to ride easily, even without your hands on the handlebars. In Rockets and Space Travel, we look at the principle of action and reaction, and see how this basic idea makes it possible for spacecraft to leave Earth's surface and even travel toward the stars. For a more complete preview of this chapter, turn ahead to the Chapter Summary and Important Laws and Equations at the end of the chapter.

This chapter revisits many ideas from the previous chapters but it also introduces a variety of new concepts we can use. In Bicycles, we learn about dynamic stability and explore issues that waiters, water-skiers, and warehouse clerks must deal with all the time. We also encounter gyroscopes and their peculiar responses to torques. In Rockets and Space Travel, we see how gravity causes objects to orbit one another, giving rise to the intricate motions of the planets, moons, and comets. We also look at new physics that appears near the extremes of speed and gravity.

SECTION 4.1

Bicycles



A bicycle is a wonderfully energy-efficient, human-powered vehicle. Its wheels allow its rider to coast forward easily on level surfaces and accelerate effortlessly down hills. Compare the easy motion of a bicycle to that of walking, which requires effort every step of the way. Bicycles are very simple machines, and most of their moving parts are quite visible: the pedals, sprockets, brakes, and steering mechanisms, to name a few. Their simplicity and visibility make bicycles relatively easy to fix, even for a novice.

Questions to Think About: Why is a two-wheeled bicycle preferable to the apparently more stable three-wheeled tricycle? Why do you lean a bicycle in the direction that you are turning? How is it possible to ride a bicycle without hands on the handlebar? How does pedaling a bicycle make it move forward? Why does a bicycle have such a complicated drive system between its pedals and its rear wheel?

Experiments to Do: If you know how to ride a bicycle, pay attention to its stability as you ride it. Notice how you lean the bicycle in the direction that you are turning and how the bicycle naturally steers into that turn as you begin to lean. This automatic steering is part of the bicycle's self-stabilizing behavior. As you ride, observe how fast the pedals are turning and how hard you must push on the pedals to keep them turning at that rate. Go up and down some hills in various gears, and see how each gear choice affects the pedaling rate and the forces you must exert on the pedals. Why do you choose a certain gear for a certain situation? What pedaling rate do you find most comfortable? What pedaling force feels best?

Tricycles and Static Stability

Bicycles have a stability problem. With only two wheels to support them, stationary bicycles tip over easily. Why then do we use two-wheeled bicycles for transportation?

We can begin to answer that question by looking at **static stability**, an object's stability at rest. To have static stability, an object needs a stable equilibrium—a concept we first encountered in Section 3.1. Equilibrium means zero net force or torque, but near a *stable* equilibrium an object experiences restoring influences—forces and/or torques—that push it back toward that equilibrium. Aided by such influences, a statically stable object returns to its stable equilibrium after being displaced slightly.

A marble in a bowl exhibits static *translational* stability—when displaced from its equilibrium position at the bottom of the bowl, it experiences restoring forces that act to return it to the bottom of the bowl. In contrast, a stool exhibits static *rotational* stability—when displaced from its upright equilibrium orientation it experiences restoring torques



Fig. 4.1.1 A tricycle is very stable when it's standing still. However, it tips over easily during a high-speed turn because the rider can't lean in the direction that the tricycle is turning. The rider must also pedal furiously to move at a reasonable speed.

that act to return it to upright. The fact that bicycles don't exhibit static *translation* stability is a good thing; a bicyclist doesn't want to stay near one place. But bicycles don't exhibit static *rotational* stability either. If you want a pedal-powered vehicle that has statically rotational stability, try a tricycle (Fig. 4.1.1).

An upright tricycle has static rotational stability because it experiences restoring torques following a tip (Fig. 4.1.2a). Rather than look for those torques directly, however, let's take a much more general approach. Physicists are trained to look for the big-picture concepts hiding behind seemingly unrelated examples, and sometimes I just have to do it.

Recall that an object accelerates in whichever direction reduces its total potential energy as quickly as possible. In this case, the tricycle undergoes angular acceleration in whichever rotational direction reduces its total potential energy as quickly as possible. Since small tips always raise the tricycle's center of gravity and thus increase its gravitational potential energy, the tipped tricycle can reduce its total potential energy by rotating back to upright. The upright tricycle is thus in a stable **rotational equilibrium**—it will return spontaneously to that upright equilibrium after being tipped slightly. No wonder children love tricycles!

This relationship between static stability and total potential energy is universal. There is no need to look directly for the restoring influences, which may be complicated anyway. If an object's total potential energy rises whenever it's displaced, it's in a stable equilibrium and will tend to return there following the displacement. That useful rule applies not only to static rotational stability and tricycles but also to static translational stability and everything from canoes to bridge spans to decorative mobiles. If you want an object to be stable at rest, ensure that any small displacement increases its total potential energy.

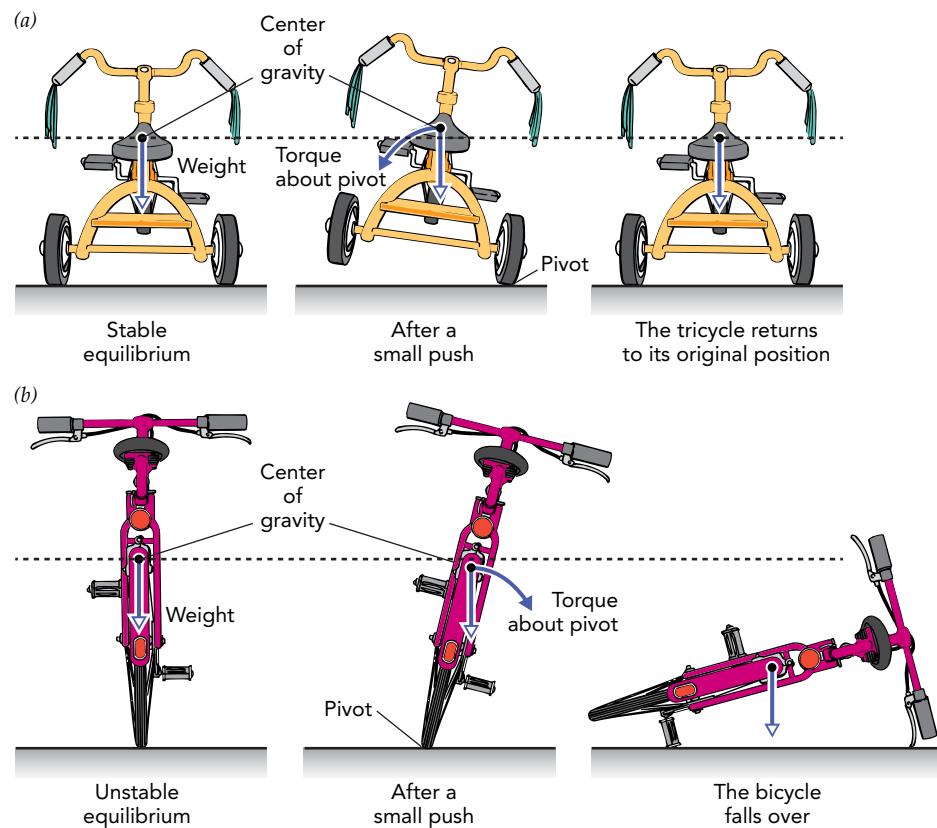


Fig. 4.1.2 (a) A tricycle is in a stable rotational equilibrium. When it's tipped, its center of gravity and gravitational potential energy rise and it experiences restoring forces that return it to upright. (b) A bicycle is in an unstable rotational equilibrium. Any tip causes it to fall.

STABLE EQUILIBRIUM AND POTENTIAL ENERGY

An object is in a stable equilibrium when any small displacement increases its total potential energy.

That general observation underlies a simple rule of thumb: an object resting on a surface is in a stable rotational equilibrium if a vertical line passing through its center of gravity also passes through the *interior* of its **base of support**—the polygon defined by its contact points with the ground. This rule follows from geometry: if a vertical line through the object’s center of gravity passes within its base of support, then tipping it always raises its center of gravity and increases its gravitational potential energy. Assuming that no other potential energies are significant, the tipped object will undergo angular acceleration back toward its original orientation, making that orientation a stable rotational equilibrium. For example, a tricycle is in a stable rotational equilibrium when its center of gravity is above the interior of its triangular base of support (Fig. 4.1.3).

Geometry also sets the limits of the object’s static rotational stability. If the vertical line passing through its center of gravity also passes through the *edge* of its base of support, there are three possibilities. If its center of gravity is below the edge, the object is in a stable rotational equilibrium. If its center of gravity is exactly on the edge, the object is in a **neutral equilibrium**—it remains in rotational equilibrium even when you tip it about that edge.

If the object’s center of gravity is above the edge, however, it is in an **unstable equilibrium**—that is, when displaced slightly in certain directions, the object will experience anti-restoring torques that cause angular acceleration away from equilibrium. Tipping the object in those directions lowers its center of gravity and decreases its gravitational potential energy, so it undergoes angular acceleration away from equilibrium and tips over. If the rider of a tricycle leans out so far that their combined center of gravity is located above the edge of the triangular base of support, they’ll be in an unstable equilibrium and in danger of tipping over.

Unlike a tricycle, a bicycle has no stable equilibrium (Fig. 4.1.2b). Its base of support is a line defined by its two wheels and a line has only an edge, no interior. Although the rider of a bicycle can achieve rotational equilibrium by locating their combined center of gravity above that edge, that equilibrium is unstable and the slightest displacement to the side will tip them over.

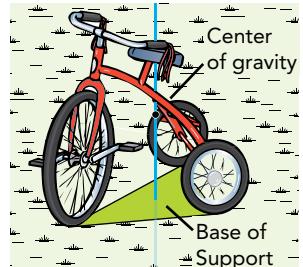


Fig. 4.1.3 An upright tricycle is in a stable equilibrium because a vertical line through its center of gravity passes through the interior of the triangular base of support formed by its three contact points with the ground.

UNSTABLE EQUILIBRIUM AND POTENTIAL ENERGY

An object’s equilibrium is unstable when a small displacement can decrease its total potential energy.

Check Your Understanding #1: Keeping Coffee in Its Place

Some travel mugs taper outward at the bottom so that they have very wide bases. Why does this shape make such a mug particularly stable and keep it from flipping over during the morning commute?

Answer: The mug’s wide base ensures that its total potential energy will continue to increase as it tips, even if that tip becomes quite severe, so it will naturally return to its stable equilibrium.

Why: A mug with a narrow base is stable, but only if it never tips very far. As soon as its center of gravity moves beyond its original base of support, over it will go. However, a mug with a wide base has a broad base of support and will recover even from severe tips.

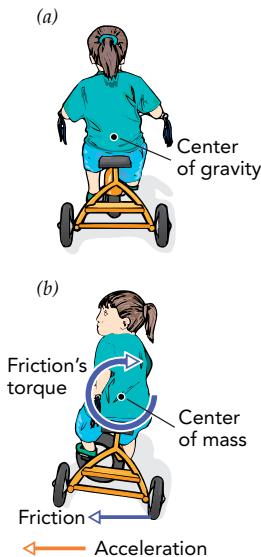


Fig. 4.1.4 (a) A tricycle that is heading straight is stable because any tip causes its center of gravity to move upward. (b) During a fast left turn, however, the tricycle accelerates left and friction exerts a large leftward force on the wheels. This frictional force produces a torque about the tricycle and rider's combined center of mass and can cause the tricycle to tip over.

Bicycles and Dynamic Stability

Static rotational stability matters most to people who have difficulty balancing. That's why children learn to ride tricycles first (Fig. 4.1.4a). But when a tricycle is moving, static rotational stability doesn't guarantee safety. If a child rolls down a steep hill and then makes a sudden sharp turn, he or she will probably flip over. What has gone wrong?

A moving tricycle stays upright only if the girl riding it avoids sudden accelerations, such as a left turn at high speed. To make such a turn, she steers the tricycle's wheels so that friction with the pavement pushes the tricycle to the left (Fig. 4.1.4b). Frictional forces on the wheel accelerate the tricycle leftward, redirecting its speed so that it turns. Of course, the girl needs to turn, too, so the tricycle pushes her along with it. As long as the turn is slow, a gentle push is all that's required and the tricycle and girl turn together safely.

If the turn is too abrupt, however, the girl doesn't complete the turn along with the tricycle. Instead, her body goes straight as the tricycle drives out from under her. Crash. As you can see, a tricycle has good static stability but poor **dynamic stability**, stability in motion.

The tricycle flips because it can't handle the enormous torque that friction exerts on it during a sharp turn. Because the horizontal frictional force that turns the tricycle is exerted well below the tricycle and rider's combined center of mass (Fig. 4.1.4b), it produces a torque about that center of mass. If this torque is small, the tricycle's static rotational stability will provide enough restoring torque in the opposite direction to prevent any angular acceleration. If the turn is too sharp, the huge frictional torque will overwhelm the limited restoring torque and the tricycle and rider will flip over. During high-speed turns, the tricycle is dynamically *unstable*.

Since the goal of a wheeled vehicle is to go somewhere, dynamic rotational stability is ultimately more important than static rotational stability. And while a bicycle lacks static stability, a moving bicycle is remarkably stable. Its dynamic stability is so good that it's almost hard to tip over and can even be ridden without any hands on the handlebars. This feat is a popular daredevil stunt among children who haven't yet realized how easy it is.

As British physicist David Jones discovered, the bicycle's incredible dynamic stability results from its tendency to steer automatically in whatever direction it's leaning. For example, if the bicycle begins leaning to the left, the front wheel will automatically steer toward the left so as to return the bicycle to an upright position. Although a stationary bicycle falls over when it's displaced from its unstable equilibrium, a forward-moving bicycle naturally drives under the combined center of mass and returns to that unstable equilibrium.

There are two physical mechanisms acting together to produce this automatic steering effect: one involving rotation and one involving potential energy. The first is based on the wheels alone: they behave as *gyroscopes*. Because each wheel is spinning, it has angular momentum and tends to continue spinning at a constant angular speed about a fixed axis in space. Since a wheel's angular momentum can be changed only by a torque, it naturally tends to keep its upright orientation.

But angular momentum alone doesn't prevent the bicycle from tipping over, any more than it prevents a tricycle from doing so. Instead, it prompts the bicycle to steer automatically by way of gyroscopic *precession*, the pivoting of a gyroscope's rotational axis caused by a torque exerted perpendicular to its angular momentum. When a bicycle is upright, the pavement's upward support force points toward the front wheel's center of mass and produces no torque on that wheel. But when the bicycle leans to the left, the pavement's upward support force no longer points at the wheel's center of mass and it therefore produces a torque on the wheel. That torque is perpendicular to the wheel's angular momentum, so the wheel precesses; its axis of rotation pivots toward the left and thereby steers the bicycle to safety!

Assisting gyroscopic precession in this automatic steering process is a second effect due to potential energy. Because of the shape and angle of the fork supporting its front wheel, a leaning bicycle can lower its center of gravity and reduce its total potential energy by steering its front wheel in the direction that the bicycle is leaning. When the bicycle

leans to the left, its front wheel steers toward the left to reduce the bicycle's total potential energy as quickly as possible. Once again the bicycle automatically steers in the direction that it's leaning and avoids falling over. These self-correcting effects explain why a riderless bicycle stays up so long when you roll it forward or down a hill.

There is little a bicycle designer can do to change the gyroscopic effect, but the potential energy effect depends on fork shape and angle. To be stable, the front wheel must touch the ground behind the steering axis (Fig. 4.1.5). If the fork is flawed, so that the wheel touches the ground ahead of the steering axis, the bicycle will steer the wrong way when it leans and be virtually unridable.

The front fork of a typical adult's bicycle arcs forward so that the wheel touches the ground just behind the steering axis. This situation leaves the bicycle dynamically stable enough to ride yet highly maneuverable. In contrast, the front fork of a typical child's bicycle is relatively straight so that the wheel touches the ground far behind the steering axis. The child's bicycle is therefore more dynamically stable than the adult's bicycle but also less easy to turn. That trade-off between dynamic stability and maneuverability is universal, appearing not only in pedal-powered vehicles but in cars, boats, and aircraft as well.

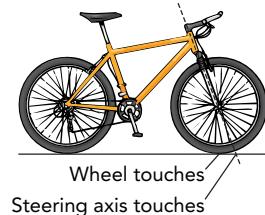


Fig. 4.1.5 A bicycle is stable when moving forward, in part because its front wheel touches the ground behind the steering axis. As a result, the front wheel naturally steers in the direction that the bicycle is leaning and returns the bicycle to an upright position.

Check Your Understanding #2: Like a Rolling Coin

Standing a coin on edge is difficult, but a rolling coin stays upright easily. What effect is stabilizing the moving coin?

Answer: The effect is gyroscopic precession.

Why: On its edge, the coin has little or no static stability and tips over easily. But when it's rolling forward quickly, the same gyroscopic precession effect that causes a bicycle to steer under its center of gravity also dynamically stabilizes the coin's otherwise unstable equilibrium.

Leaning While Turning

Why does a bicycle rider lean during a turn? The answer is that leaning can balance out the torque that friction exerts on him during the turn, the same frictional torque that flipped our unfortunate tricycle rider. With the proper lean, the bicyclist can safely complete even the sharpest turn.

As he rides along, the bicyclist tries to keep himself and the bicycle in rotational equilibrium. Since they experience a frictional torque about their combined center of mass each time they turn, the bicyclist balances that torque by leaning them toward the inside of the turn. The pavement's upward support force then produces a torque on them about their center of mass that opposes the frictional torque. When those two torques sum to zero, the bicyclist and bicycle are safely in rotational equilibrium.

These two opposing torques are both produced by pavement forces on the wheels, and their sum is the overall torque that the overall pavement force on the wheels produces. That overall torque drops to zero when the overall pavement force points directly toward the combined center of mass. As we saw in Section 2.1, a force exerted directly toward a pivot produces zero torque about that pivot. To stay in rotational equilibrium as he rides, the bicyclist always places the combined center of mass directly in line with the pavement's overall force on the wheels.

For example, when the bicycle is heading straight, the rider can obtain zero net torque by keeping the bicycle upright. The pavement pushes the wheels straight upward, directly toward the combined center of mass (Fig. 4.1.6a). But when the bicycle is turning left, the rider must lean to the left to stay in rotational equilibrium. During the leftward turn, the pavement pushes the wheels upward and leftward (Fig. 4.1.6b). Leaning leftward the proper amount ensures that the pavement force points toward the combined center of mass and produces zero torque (Fig. 4.1.7).

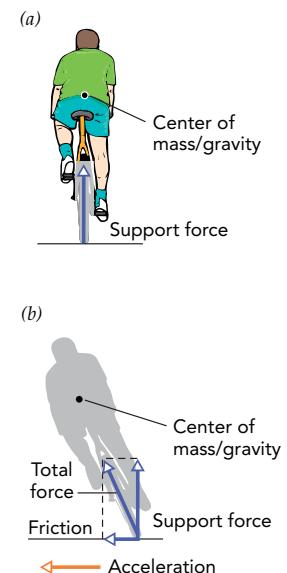


Fig. 4.1.6 (a) A bicycle that is heading straight is in rotational equilibrium when it's perfectly upright. The support force from the road produces no torque about its center of mass. (b) A bicycle that is turning left is in rotational equilibrium when it's tilted to the left. Together, the support and frictional forces from the road produce no torque about its center of mass.

© Bryn Lennon/Getty Images, Inc.



Fig. 4.1.7 As they round a turn during a race, these cyclists lean toward the inside of the turn. This leaning prevents the pavement from producing torques on them and tipping them over.

After a while, leaning the bike while turning becomes so automatic, so habitual, that you don't even think about it. You simply can't ride a bicycle or motorcycle without leaning as you turn. Even if you turn one of these vehicles so sharply that it skids, leaning can still keep it safely in rotational equilibrium.

All this discussion of leaning during turns begs the question: How do you make the bicycle lean prior to a turn? Actually, you cause that lean unconsciously by steering the bicycle briefly in the wrong direction—the direction opposite the turn itself! The bicycle then drives out from under you, and the two of you begin to lean in the desired manner. You then steer in the correct direction and remain in rotational equilibrium throughout the turn. When you're ready to stop turning, you steer extra hard briefly in the direction of the turn. The bicycle then drives under you, and the two of you return to upright. The turn is over.

Since only a statically unstable vehicle can lean, a statically stable one must rely on restoring torques to keep it near its rotational equilibrium. As we saw with tricycles, restoring torques are limited, and a vehicle can tip out of its rotational equilibrium during rapid accelerations. A car, truck, or SUV will flip if it turns too sharply, and some are particularly prone to such catastrophes. The higher a vehicle's center of mass and the narrower its base of support, the more limited its restoring torques and the more easily it flips during turns. SUVs are surprisingly vulnerable to such rollover accidents, and small trucks that have been boosted up to resemble tractor-trailer cabs are truly hazardous. Even some ordinary cars have been found unsafe in this regard.

Check Your Understanding #3: Cutting Corners

Why does a skier lean in the direction of a turn during a downhill run?

Answer: She leans into the turn to make sure that the force on her skis is directed toward her center of mass.

Why: In many sports and situations a person must lean in the direction of a turn. Since a turn always involves horizontal acceleration and horizontal force, the person must shift her center of mass toward the inside of the turn. When she leans just the right amount, the net force on her feet points directly toward her center of mass and exerts no torque on her about her center of mass. She remains in rotational equilibrium and doesn't tip over.

Pedaling Bicycles

So far, we have only discussed stability. Once we've settled on a bicycle as the most likely configuration for a useful person-powered vehicle, we need to figure out how to power it. The rider could push his feet on the ground, but that would be pretty inconvenient and even dangerous at high speeds. Instead, we use foot pedals to produce a torque on one of the wheels. But which wheel, and how do we produce that torque?

The original answer was to power the front wheel, using cranks attached directly to its axle. A crank is simply a lever that projects from the axle and produces a torque on that axle as you push its free end around in a circle. Each bicycle crank has a pedal installed on its free end so that you can use your foot to push on it. With its axle suspended in bearings, the front wheel turns as you pedal it. Friction between the turning wheel and the ground then pushes the bicycle forward.

This pedaling method is still used in children's tricycles, but it has three drawbacks. First, pedaling the front wheel interferes with its other responsibility: steering. Second, you can't take a break from pedaling; if the vehicle is moving, so are the pedals. Third, you can produce more than enough torque on the front wheel, but you often have trouble moving your legs quickly enough to keep up with the pedals. When you ride fast on level ground, you find yourself pedaling furiously and feeling very little resistance from the pedals.

The frantic pedaling problem goes to the heart of all pedal-powered vehicles, vehicles that draw power from you, the rider. You provide that power by doing a certain amount of

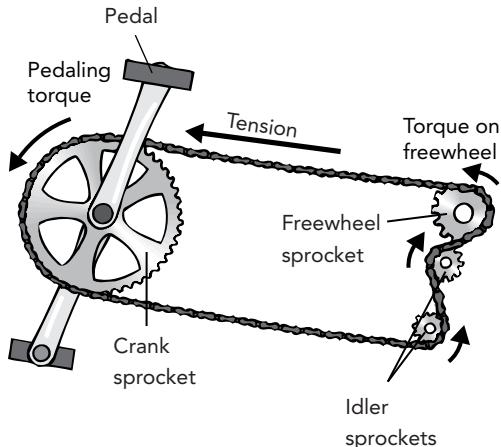


Fig. 4.1.9 The drive system for a modern bicycle. Pedaling produces torque on the crank sprocket, which in turn produces tension in the top segment of the chain. That segment of chain produces torque on the freewheel sprocket, which conveys that torque to the bicycle's rear wheel (not shown). The idler sprockets handle the extra chain.

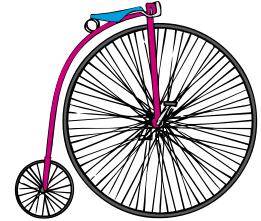


Fig. 4.1.8 The pennyfarthing used a large, directly driven front wheel to permit the rider to travel at a reasonable speed without having to pedal very rapidly. Its name came from its resemblance to two coins, the large English penny and the smaller farthing.

work on the pedals each second. Since work is equal to force times distance, you can do the same work on the pedals each second—that is, you can provide the same power to the vehicle—by pushing the pedals forward hard as they turn slowly or by pushing the pedals forward gently as they turn quickly. However, for reasons having to do with physiology more than physics, your legs are best at providing power when you are pushing the pedals forward medium hard as they turn medium fast.

Unfortunately, the pedals of an ordinary tricycle turn too quickly and too easily to make good use of your pedaling power. When the tricycle is moving fast on level ground, you can barely keep up with its pedals, let alone do work on them. The only time a tricycle makes good use of your capacity to supply power is when it's going uphill at moderate speed; only then do the pedals turn medium fast and require that you push them forward medium hard.

An early solution to the frantic pedaling problem was to use a gigantic front wheel. In such a configuration, one turn of the wheel would take you a considerable distance, so you no longer had to pedal furiously to go fast on a level road. At last you could reach your peak performance on level ground, pushing the pedals forward medium hard as they turned medium fast. The pennyfarthing of the mid-nineteenth century was this sort of bicycle (Fig. 4.1.8). But pedaling still interfered with steering and you couldn't stop pedaling while the bicycle was moving. Furthermore, this bicycle had a new problem—you couldn't push the pedals forward hard enough to keep its front wheel turning steadily on uphill stretches.

These problems were solved by removing the cranks from the front wheel's axle and using an indirect drive scheme to convey power to the rear wheel (Fig. 4.1.9). Employing toothed sprockets and a chain loop, that indirect drive allowed the pedals and the wheels to turn at different rates. This change lets you use mechanical advantage to choose how you supply power to the bicycle: whether you exert large forces on slowly moving pedals or small forces on rapidly moving pedals or, ideally, medium forces on medium-fast-moving pedals. Whether you're zooming along a level road or grinding slowly up a steep hill, you can always find a drive setting, or gear, that lets you comfortably supply your maximum power.

Finally, the nonstop pedaling problem was solved by incorporating a one-way drive or freewheel in the hub of the rear wheel. This freewheel (Fig. 4.1.10) allows the rear wheel to turn freely in one direction so that you can stop pedaling as you coast forward. A modern bicycle with these improvements appears in Fig. 4.1.11.

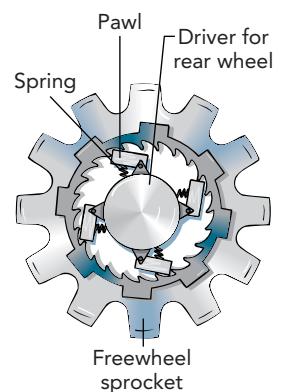


Fig. 4.1.10 The ratchet in a bicycle freewheel. If the relative rotation of the inner and outer parts is in the correct direction, the pawls transmit torque from the outer part to the inner part. If the relative rotation direction is reversed, the pawls compress the springs and skip along the teeth on the inside of the outer part. No torque is transmitted.



Fig. 4.1.11 A modern bicycle. The rear wheel is driven by a chain that allows the rider to vary the mechanical advantage between the pedals and the rear wheel. A freewheel in the hub of the rear wheel lets that wheel turn freely in one direction. This free motion allows the bicycle to coast forward, even when the pedals are stationary.



Check Your Understanding #4: Bigger Isn't Always Better

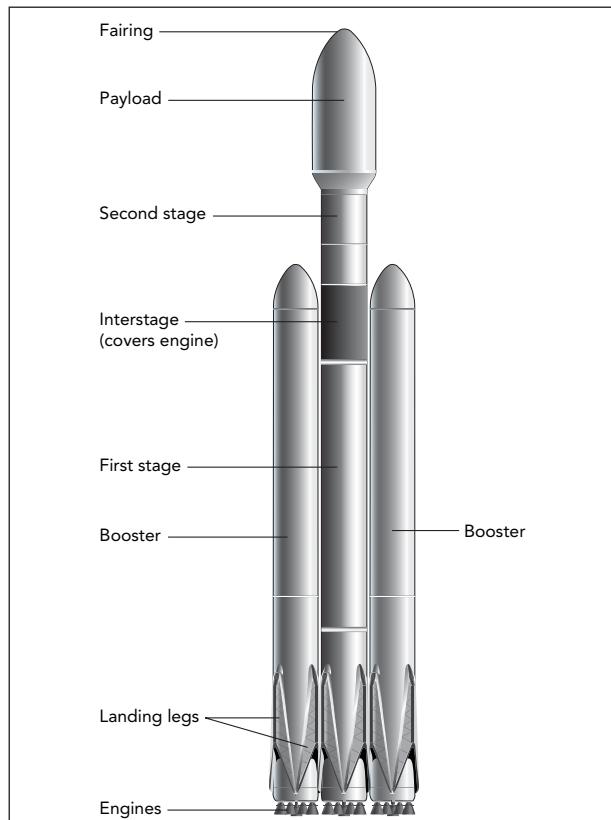
Why is it so difficult to climb a hill on a pennyfarthing?

Answer: Each turn of the huge wheel takes you a long distance up the hill and requires lots of work. To do this much work in a single turn of the pedals, you must push on the pedals very hard.

Why: As you pedal a bicycle, you are doing work on the pedals; you push the pedals as they move in the direction of that push. If too much work is required, as is the case when you're going uphill on a pennyfarthing, the force required becomes unbearable.

SECTION 4.2

Rockets and Space Travel



Despite the complexity of modern spacecraft, the rocket is one of the simplest of all machines. It's based on the principle that every action has a reaction. A rocket is propelled forward by pushing material out of its tail. As simple as rockets are, however, people have been developing better rockets for over 700 years. They're used for such pursuits as space exploration, weaponry, rescue operations, and amusement.

Questions to Think About: What pushes a rocket forward? How can a rocket work in space, where there seems to be nothing to push against? Why do modern rockets have such fancy exhaust nozzles? What is the fastest speed that a rocket can reach? Why do satellites travel endlessly around Earth? What sort of path does a rocket take as it travels from planet to planet?

Experiments to Do: While an afternoon spent launching model rockets (available at hobby and toy stores) would be the best introduction to this section, a toy balloon will do just fine. Blow up the toy balloon and let it go. It will sail around the room. What pushes the balloon forward? Are the room's air and walls involved in the propulsion, or is the balloon propelled by the very act of ejecting gas through its opening? How could you verify your answer to the previous question?

Rocket Propulsion

Among a rocket's most impressive features are its ability to propel itself forward even in the complete isolation of space and its capacity to reach astonishing speeds using that propulsion. It somehow manages to push itself forward without any outside help and to use that forward push to accelerate seemingly without limits.

Of course, a rocket can't really push itself forward, any more than you can lift yourself up by your boots, and it can't accelerate forever. In reality, it obtains a forward force, a **thrust** force, by pushing against its own limited store of fuel, and when that fuel runs out, it stops accelerating. To understand how a rocket obtains thrust from its fuel supply, let's look at how Newton's third law, the one describing action and reaction, applies to rockets.

Imagine that you're standing in the middle of a frozen pond with zero velocity and no momentum. It's a warm day and the wet ice is remarkably slippery. Try as you may, you can't seem to get moving at all. How can you get off the ice?

Because of your inertia, the only way you can start moving is if something pushes on you. Sure, you could order a pizza and then push against the delivery truck when it arrives. Instead you remove a shoe and throw it as hard as you can toward the east side of the pond (Fig. 4.2.1). As you throw the shoe, you exert a force on it with your hand. The shoe accelerates and heads off across the ice.

What happens to you? You head off toward the west side of the pond! You're moving because when you pushed the shoe toward the east side of the pond, it pushed you equally hard toward the west side of the pond. In the process, you transferred momentum to the shoe and it transferred momentum in the opposite direction to you. Momentum isn't being created or destroyed, it's only being redistributed. Even after you let go of the shoe, your combined momentum remains at zero. The shoe has as much momentum in one direction as you have in the other.

Of course, you are much more massive than the shoe, so you end up traveling slower than it does. Momentum is equal to mass times velocity, so the more massive the object, the less velocity it needs for the same amount of momentum. Still, you've achieved what you set out to do—you're sliding slowly toward the west side of the pond.

Your final speed is limited because you managed to transfer only a small amount of momentum to the shoe and thus received only a small amount of opposite momentum in return. If you'd been able to throw the shoe faster or if you'd thrown a whole display case of shoes, you'd have transferred more momentum and would be going faster.

Instead of throwing shoes, you'd have done better to throw very fast-moving gas molecules. Even at room temperature, the molecules in air are traveling about 1800 km/h (1100 mph). When gas molecules are heated to roughly 2800 °C (5000 °F), as they are in a liquid-fuel rocket engine, they move about three times that fast. If you throw something in one direction at that kind of speed, you receive quite a lot of momentum in the other direction.

That's what a conventional rocket engine does. It uses a chemical reaction to create very hot exhaust gas from fuels contained entirely within the rocket itself. What started as chemical potential energy in those fuels becomes thermal energy during the reaction and then mostly directed kinetic energy as the hot gas rushes out of the rocket engine in a plume of incandescent exhaust.

The rocket engine exerts a tremendous backward force on the gas, accelerating that gas to exhaust velocities measured in kilometers per second or miles per second. The gas pushes forward equally hard on the rocket engine. Propelled by such rocket engines, spacecraft can reach astonishing heights and speeds (Fig. 4.2.2).

Achieving those amazing exhaust velocities requires carefully designed nozzles. If you've ever watched the launch of a large rocket, you've probably noticed the bell-shaped nozzles through which the exhaust flows (Fig. 4.2.3). Each nozzle allows the rocket to

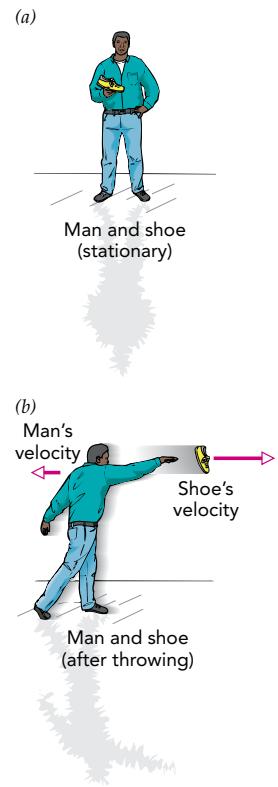


Fig. 4.2.1 (a) A man who is holding a shoe while standing still on ice has zero momentum. (b) Once he has thrown the shoe to the right, the shoe has a momentum to the right and the man has a momentum to the left. The total momentum of the man and shoe is still zero. Because the man is much more massive than the shoe, the shoe moves much faster than the man.

Fig. 4.2.2 This spacecraft's rocket engine pushes its fuel leftward in a plume of high-velocity exhaust and that fuel pushes the spacecraft rightward, action and reaction. As the spacecraft accelerates toward the right, its rightward velocity increases and the leftward velocity of its ejected fuel decreases. Throughout this propulsion process, the rocket's total momentum remains unchanged; it was zero prior to launch, so it remains zero as the spacecraft and fuel push on each other.

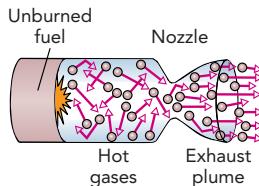
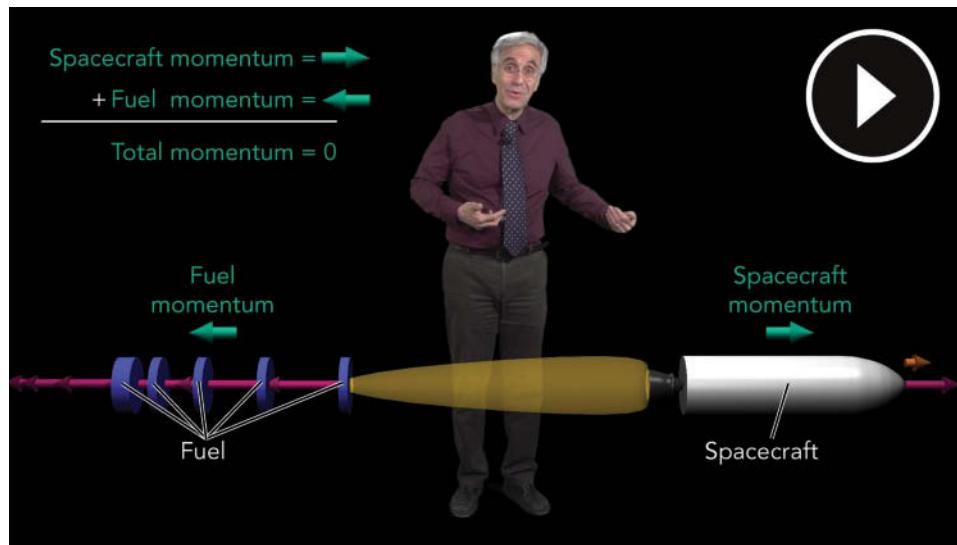


Fig. 4.2.3 A molecular picture of what happens in a chemical rocket engine. The engine burns its fuel in a confined chamber, and the exhaust gas flows out of a nozzle. The nozzle converts the random thermal motions of the exhaust-gas molecules into directed motion away from the rocket engine.

1 Swedish inventor and engineer Carl Gustaf de Laval's (1845–1913) invention of the converging-diverging nozzle predates the modern development of rockets by several decades. He invented this nozzle as a way to make steam turbines more efficient and is credited with laying the foundation for all future turbine technology. De Laval is also known for his invention of the cream separator for milk.

obtain as much forward momentum as possible from its exhaust by directing that exhaust backward and accelerating it to the greatest possible speed. As we'll see in Chapter 6, nozzles allow gases to convert their various internal energies into kinetic energy and are ideally suited for directing and accelerating gases. In the case of rocket exhaust, the most effective nozzle shape is a converging-diverging one, called a de Laval nozzle after its Swedish inventor, Carl Gustaf de Laval. **1**

To understand fully why this complicated nozzle structure works so well for rocket engines, we'd need to examine the physics of flowing gases up to and beyond the speed of sound. We'll encounter some of these issues later in this book, but for now a brief summary will have to suffice.

Inside the rocket and before the de Laval nozzle, the hot exhaust gas is tightly packed and its pressure is enormous. Like the gas in a spray bottle, this exhaust gas accelerates rapidly through the nozzle toward the lower pressure environment outside. The narrowing throat of that nozzle aids its acceleration, up to a point. When the gas reaches the narrowest part of the nozzle, it's traveling at the speed of sound and its characteristics begin to change dramatically. To coax the supersonic exhaust gas to accelerate still further, the nozzle stops narrowing and begins to widen. The tightly packed exhaust gas expands in volume as it flows through that widening bell and thereby prepares to enter the more open environment outside the nozzle.

Just how wide the diverging end of the de Laval nozzle must be to obtain the maximum thrust from its exhaust gas depends on the nozzle's surroundings. Near sea level, the exhaust gas flows into ordinary air outside the nozzle and a relatively narrow de Laval nozzle works best. At high altitude or in space, the exhaust gas enters thinner air or nothing at all, so a wider de Laval nozzle is more ideal. Rockets typically make a compromise in their nozzle shapes so as to operate reasonably well in all environments.

By the time the gas reaches the end of the de Laval nozzle, it has converted most of its original energy into kinetic energy, with its velocity directed away from the nozzle. In fact, because the gas actually continues to burn even as it flows through the nozzle, its kinetic energy and speed keep rising until they reach fantastic levels. With the help of the de Laval nozzle, exhaust gas leaves the rocket's engine at an **exhaust velocity**, or backward-directed flow speed, of between 10,000 and 16,000 km/h (6000 and 10,000 mph).

As it creates this plume of exhaust, the rocket pushes the gas backward and gives it backward momentum. The gas completes the momentum transfer by pushing the rocket forward. The very act of ejecting the exhaust is all that's required to obtain forward thrust;

the rocket doesn't need anything external to push and will operate perfectly well in empty space (see 2). When it pushes hard enough on its exhaust, the rocket not only can support its own weight, it can even accelerate upward.

COMMON MISCONCEPTIONS: Action and Reaction in Rockets

Misconception: A rocket needs some external object to react against to push itself forward.

Resolution: While rocket propulsion does involve a pair of equal but opposite forces, action and reaction, the rocket is pushing its exhaust backward (action) and the exhaust is pushing the rocket forward (reaction). What this exhaust plume hits, if anything, makes no difference to the propulsion effect.

2 On January 13, 1920, *The New York Times* ran an editorial attacking Robert Goddard for proposing that rockets could be used for travel in space. With modest financial support from the Smithsonian Institution, Goddard was pioneering the development of liquid-fuel rockets. The editorial began: "That Professor Goddard, with his 'chair' in Clark College and the countenancing of the Smithsonian Institution, does not know the relation of action to reaction, and of the need to have something better than a vacuum against which to react—to say that would be absurd. Of course he only seems to lack the knowledge ladled out daily in high schools."

Check Your Understanding #1: A Rocket with a Head Start

When a missile is launched from beneath the wing of a fighter aircraft, what does it push against to accelerate forward?

Answer: It pushes against its own exhaust, as do all rockets.

Why: While it may appear that the plume of exhaust beneath a ground-launched rocket is what lifts that rocket upward, the ground itself doesn't contribute to the thrust. The very act of pushing the gas out the nozzle propels the rocket forward.

The Ultimate Speed of a Spacecraft

At rest on the launchpad, a rocket consists principally of a spacecraft and a supply of fuel. Once the rocket's engine begins to fire, exhaust from the burned fuel accelerates backward and the spacecraft accelerates forward. The fuel is gradually consumed until eventually it runs out and the spacecraft coasts along on its own. Although weight and air resistance influence this story, let's neglect both for now to see what determines the spacecraft's eventual speed.

Remarkably, the ultimate speed of the spacecraft is not limited to the rocket's exhaust speed. As long as the rocket keeps pushing exhaust backward, it will continue to accelerate forward. For the spacecraft to reach speeds in excess of its exhaust speed, however, the rocket must push the majority of its initial mass backward as exhaust. After all, the momentum of the exhaust's larger mass heading backward at the exhaust speed cancels the momentum of the spacecraft's smaller mass heading forward at more than the exhaust speed. For a rocket that is 90% fuel and 10% spacecraft at launch, we might reasonably expect the spacecraft to end up traveling forward at nine times the exhaust speed.

Unfortunately, that simple analysis overestimates the spacecraft's speed. Because the rocket accelerates forward while its engine is firing, only the first portion of its exhaust travels backward at the full exhaust speed (Fig. 4.2.2.). As the rocket picks up speed in the forward direction, its exhaust moves less rapidly in the backward direction. When the rocket's forward speed exceeds its exhaust speed, the exhaust actually ends up moving forward!

Despite this problem, a spacecraft can still travel forward faster than the exhaust speed; it just needs more fuel. If we neglect air resistance and weight, the spacecraft's final speed is given by the rocket equation:

$$\text{spacecraft speed} = \text{exhaust speed} \cdot \log_e \left(\frac{\text{mass}_{\text{spacecraft}} + \text{mass}_{\text{fuel}}}{\text{mass}_{\text{spacecraft}}} \right). \quad (4.2.1)$$

For a rocket that is 90% fuel and 10% spacecraft at launch, its spacecraft can reach 2.3 times the speed of its exhaust gas. If it is more than 90% fuel, it can go even faster.

There's a problem, however, with trying to burn up and eject a huge fraction of the rocket's original mass as exhaust. It's difficult to construct a rocket that is 99.99% fuel and 0.01% spacecraft. Instead, a space-bound rocket is typically a stack of several separate rocket stages, each stage much smaller than the previous stage. Once the first stage has used up all its fuel, the whole stage is discarded and the next rocket stage begins to operate. In this manner, the rocket behaves as though it's ejecting almost all its mass as rocket exhaust. With the help of several stages and lots of fuel, rockets can travel substantially faster than their exhaust velocities and reach Earth orbit or the solar system beyond.



Check Your Understanding #2: Not Everything Is Disposable

The space shuttle didn't have the staged look of expendable rockets. How did it manage to eject most of its launch mass as exhaust?

Answer: It was actually staged subtly. Its two solid-fuel boosters were effectively the first stage, the external liquid-fuel tank was the second stage, and the orbiter itself was the third stage.

Why: Although the space shuttle was not stacked one stage on the next like a Saturn or Delta rocket, it didn't travel from ground to space as a single object. It discarded empty fuel containers as it accelerated upward. First to go were the two boosters, followed by the external fuel tank. The final mass of the orbiter itself was much less than what left the launchpad.

Orbiting Earth

If the spacecraft were heading straight up when it ran out of fuel, it would either fall back to the ground or leave Earth forever (more on that later). But if it were heading primarily horizontally when its engine turned off, it might find itself circling Earth endlessly. With no atmosphere to affect it, the spacecraft follows a path determined only by inertia and gravity, and since the spacecraft's weight causes it to accelerate toward the center of Earth, its trajectory can bend into a huge elliptical loop around Earth.

The spacecraft is following an orbit around Earth. An **orbit** is the path an object takes as it falls freely around a celestial object. Although the spacecraft accelerates directly toward Earth's center at every moment, its huge horizontal speed prevents it from actually hitting Earth's surface. Instead, Earth's curved surface bends away from it so rapidly that the fast-moving spacecraft falls around Earth rather than into it (Fig. 4.2.4). To orbit Earth just above the atmosphere, a spacecraft must travel at the enormous horizontal speed of 7.9 km/s (about 17,800 mph) and will circle Earth once every 84 minutes.

The farther the spacecraft's orbit is from Earth's surface, however, the longer its **orbital period**, the time it takes to complete one orbit. That longer period is due in part to the increased distance the spacecraft must travel when completing a larger orbit, but it's also due to an important characteristic of gravity: Earth's gravity becomes weaker as distance from Earth's center of mass increases.

Section 1.2 observed that an object's weight is equal to its mass times 9.8 N/kg, the acceleration due to gravity. That relationship (Eq. 1.2.1), however, is only an approximation, valid for an object near Earth's surface. As the object's altitude increases, both the acceleration due to gravity and the object's weight decrease—you truly weigh less on a mountaintop than you do in the valley nearby. Moreover, Eq. 1.2.1 completely ignores the fact that gravity attracts every object in the universe toward every other object in the universe (see [3]).

To determine orbits around Earth and other celestial bodies, scientists and engineers use a more general formula that relates the gravitational forces between two objects to their masses and the distance separating them. These forces are equal to the gravitational constant times the product of the two masses, divided by the square of the distance separating

[3] English physicist Henry Cavendish (1731–1810) proved that terrestrial objects do exert gravitational forces on one another. His experiment, performed in 1798, measured the tiny forces that two metal spheres exert on one another, using a very sensitive torsion balance. Comparing the forces between the two spheres with those between Earth and those same spheres (their weights), Cavendish was able to deduce the mass of Earth. In effect, Cavendish's tabletop experiment allowed him to weigh the Earth.

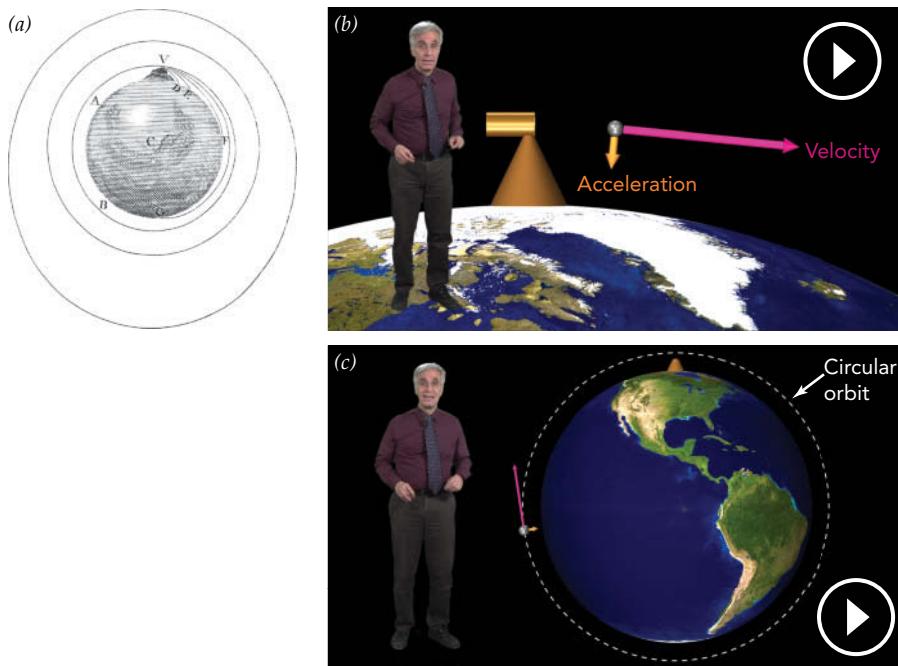


Fig. 4.2.4 (a) Newton's drawing of a cannonball fired horizontally from the top of a tall mountain. As the cannonball's speed increases, it travels farther from the mountain before hitting Earth. (b) When the cannonball moves fast enough, the curved Earth drops away beneath it and it never hits Earth at all. (c) The cannonball then orbits Earth.

them. This relationship, discovered by Newton and called the **law of universal gravitation**, can be written as a word equation:

$$\text{force} = \frac{\text{gravitational constant} \cdot \text{mass}_1 \cdot \text{mass}_2}{(\text{distance between masses})^2}, \quad (4.2.2)$$

in symbols:

$$F = \frac{G \cdot m_1 \cdot m_2}{r^2},$$

and in common language:

The pull of gravity is strongest between massive objects but diminishes rapidly as distance increases.

Note that the force on mass₁ is directed toward mass₂ and the force on mass₂ is directed toward mass₁. Those two forces are equal in magnitude but oppositely directed. The **gravitational constant** is a fundamental constant of nature, with a measured value of $6.6720 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2$.

THE LAW OF UNIVERSAL GRAVITATION

Every object in the universe attracts every other object in the universe with a force equal to the gravitational constant times the product of the two masses, divided by the square of the distance separating the two objects.

This relationship describes any gravitational attraction, whether it's between two planets or between Earth and you. The effective location of an object's mass is its center of mass, so the distance used in Eq. 4.2.2 is the distance separating the two centers of mass. For a spacecraft orbiting Earth just above its atmosphere, that distance is roughly Earth's radius of 6378 km (3964 miles). For a spacecraft far above the atmosphere, however, the distance is larger and the force of gravity is weaker; that spacecraft experiences a smaller acceleration due to gravity. To give it the additional time it needs for its path to bend around in a circle, the high-altitude spacecraft must travel more slowly than the low-altitude spacecraft. This reduced speed explains the long orbital periods of high-altitude spacecraft.

At 35,900 km (22,300 miles) above Earth's surface, the orbital period reaches 24 hours. A satellite traveling eastward in such an orbit turns with the Earth and is said to be *geosynchronous*. If a geosynchronous satellite orbits Earth around the equator, it's also *geostationary*—it always remains over the same spot on Earth's equator. Such a fixed orientation is useful for communications and weather satellites.

Not all orbits are circular (Fig. 4.2.5). The orbits of some spacecraft are elliptical, so their altitudes vary up and down once per orbit. At *apogee*, its greatest distance from Earth's center, a spacecraft travels relatively slowly because it has converted some of its kinetic energy into gravitational potential energy. At *perigee*, its smallest distance from Earth's center, the spacecraft travels relatively rapidly because it has converted some of its gravitational potential energy into kinetic energy. Of course, the perigee should not bring the spacecraft into Earth's atmosphere or it will crash.

The greater a spacecraft's total energy, the farther its orbit extends from Earth. If the spacecraft has too much total energy, Earth's gravity will be unable to bend its path into a closed loop and the spacecraft will coast off into interplanetary space. The spacecraft's path near Earth is then a parabola or hyperbola. The spacecraft follows this parabolic or hyperbolic path only once and then drifts away from Earth forever.

A spacecraft usually enters parabolic or hyperbolic orbit by firing its rocket engine. It starts in an elliptical orbit around Earth and uses its rocket engine to increase its kinetic energy. The spacecraft then arcs away from Earth, and its kinetic energy gradually transforms into gravitational potential energy. But Earth's gravity becomes weaker with distance, and the spacecraft's gravitational potential energy slowly approaches a maximum value even as its distance from Earth becomes infinite. If the spacecraft has enough kinetic energy to reach this maximum gravitational potential energy, it will be able to escape completely

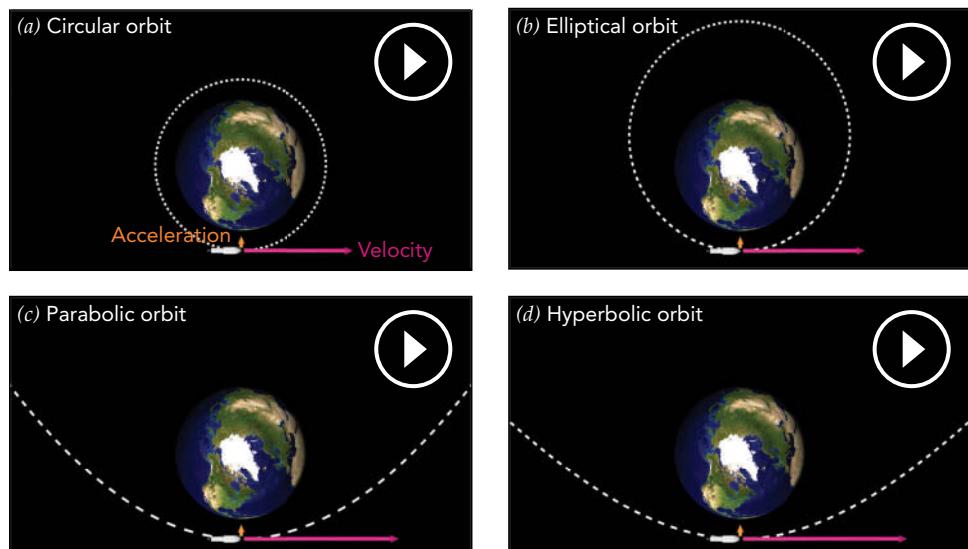


Fig. 4.2.5 As a spacecraft's total energy increases, its orbit evolves from circular (a) to elliptical (b) to parabolic (c), to hyperbolic (d). Circular and elliptical orbits are closed loops, so a spacecraft following one of those orbits does so repeatedly. Parabolic and hyperbolic orbits are open curves, so once a spacecraft following one of those orbits begins heading away from Earth, it never returns.

from Earth's gravity. If its kinetic energy approaches zero as it escapes, it's in a parabolic orbit. If it retains excess kinetic energy after the escape, it's in a hyperbolic orbit.

The speed that a spacecraft needs to escape from Earth's gravity is called the **escape velocity**. This escape velocity depends on the spacecraft's altitude and is about 11.2 km/s (25,000 mph) near Earth's surface. A spacecraft traveling at more than the escape velocity follows a hyperbolic orbital path and heads off toward the other planets or beyond.

COMMON MISCONCEPTIONS: Astronauts and "Weightlessness"

Misconception: An astronaut orbiting Earth is too far from Earth to experience gravity and is truly weightless.

Resolution: The astronaut is still so near Earth's surface that he experiences almost his full Earth weight. He feels weightless only because he is in free fall.

Check Your Understanding #3: Speeding Up the Lunar Month

The moon orbits Earth every 27.3 days, at a distance of 384,400 km from Earth's center of mass. For the moon to orbit in less time, how would its distance from Earth have to change?

Answer: The distance between Earth and moon would have to decrease.

Why: The moon behaves just like a spacecraft orbiting Earth at a distance of 384,400 km. Such a spacecraft would also have an orbital period of 27.3 days. To reduce this orbital period, the moon would have to move closer to Earth so that Earth's gravity could bend its path more rapidly.

Check Your Figures #1: Attractive Cars

How much gravitational force does a 1000-kg automobile exert on an identical car located 10 m away?

Answer: It exerts about 6.73×10^{-7} N.

Why: We use Eq. 4.2.2 to obtain the force:

$$\text{force} = \frac{6.6720 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2 \cdot 1000 \text{ kg} \cdot 1000 \text{ kg}}{(10 \text{ m})^2}$$

$$= 6.6720 \times 10^{-7} \text{ N.}$$

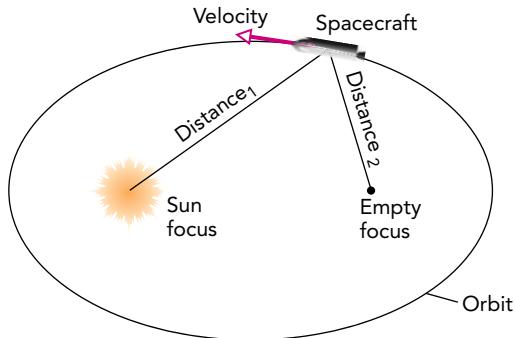
This force is roughly equal to the weight of a grain of sand. No wonder it's hard to feel gravity from anything but the entire Earth.

Orbiting the Sun: Kepler's Laws

Once it escapes from Earth's gravity and again turns off its rocket engine, the spacecraft behaves like a tiny planet and orbits the sun. If you watch it patiently as it travels and compare its orbital motion with the motions of the planets themselves, you may begin to notice three universal features of all these solar orbits. First recognized by German astronomer Johannes Kepler (1571–1630) through his careful analysis of the extensive observational data collected by Danish astronomer Tycho Brahe, those three orbital behaviors are known as Kepler's laws.

Kepler's first law is already rather familiar to us from our examination of Earth orbits. This law describes the shape of the spacecraft's looping orbit around the sun; it's an ellipse, with the sun at one focus of that ellipse (Fig. 4.2.6). An ellipse isn't an arbitrary oval; it's a planar curve with two foci and a rule stating that the sums of the distances between each point on the curve and the two foci are the same. In this case, one focus is occupied by the sun, and the other focus is usually empty. If you add the distance from the spacecraft to the

Fig. 4.2.6 A spacecraft's orbit around the sun is an ellipse, with one focus occupied by the sun and the other focus empty. The sum of distance₁ and distance₂ is the same for all points on this ellipse.



sun and the distance from the spacecraft to the empty focus, that sum will remain constant as the spacecraft orbits the sun. A circular orbit around the sun is a particularly simple elliptical one; its two foci coincide and the sun occupies them both.

Kepler recognized that every object orbiting the sun follows such an elliptical path. The planets move along nearly circular ellipses, while the comets travel in highly elongated ones. Our spacecraft's orbit may be circular or elongated, depending on its position and velocity at the time its engine stopped firing. To reach another planet, the spacecraft's solar orbit and that of its destination planet must overlap and the two objects must reach that overlapping point at the same time. Traveling from planet to planet is clearly a tricky business.

Newton later recognized that these elliptical orbits are a direct consequence of the law of universal gravitation (Eq. 4.2.2) and its inverse square relationship between force and distance ($\text{force} \propto 1/\text{distance}^2$). Any other relationship between force and distance would yield curving paths that don't close on themselves at all, let alone form ellipses. As you follow the spacecraft's elliptical orbit around the sun, you are witnessing an elegant exhibition of the law of universal gravitation.

KEPLER'S FIRST LAW: ORBITS

All planets move in elliptical orbits, with the sun at one focus of the ellipse.

Kepler's second law describes the area swept out by a line stretching from the sun to the spacecraft: that line sweeps out equal areas in equal times (Fig. 4.2.7). Regardless of how circular or elongated the spacecraft's orbit is, or where the spacecraft is along that orbit, the area marked off each second by that moving line is always the same.

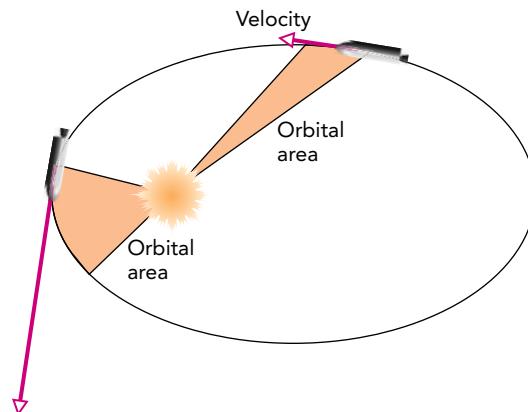


Fig. 4.2.7 The orbiting spacecraft sweeps out the same orbital area each second, despite variations in its distance from the sun. This steady sweep is a result of the spacecraft's constant angular momentum about the sun.

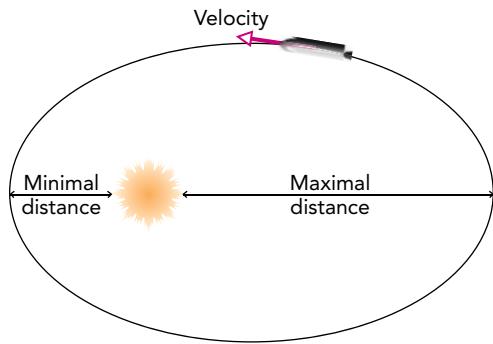


Fig. 4.2.8 The square of the spacecraft's orbital period is proportional to the cube of its mean distance from the sun (the average of its maximal and minimal distances from the sun).

This observation demonstrates another physical law—conservation of angular momentum. Since the sun's gravity pulls the spacecraft directly toward the sun, it exerts no torque on the spacecraft and the spacecraft's angular momentum about the sun is constant. Remarkably, the rate at which this line sweeps out area is proportional to the spacecraft's angular momentum, so the steadiness of that sweep demonstrates the constancy of the spacecraft's angular momentum about the sun.

KEPLER'S SECOND LAW: AREAS

A line stretching from the sun to a planet sweeps out equal areas in equal times.

Kepler's third law describes the spacecraft's orbital period around the sun: the square of its orbital period is proportional to the cube of its mean distance from the sun, that is, the average of its maximal and minimal distances from the sun (Fig. 4.2.8). This relationship can be derived from the law of universal gravitation (Eq. 4.2.2), the equations describing centripetal acceleration (for example, Eq. 3.3.1), and Newton's second law (Eq. 1.1.2).

KEPLER'S THIRD LAW: PERIODS

The square of a planet's orbital period is proportional to the cube of that planet's mean distance from the sun.

Check Your Understanding #4: Out of Round

Your spacecraft has just left Earth behind and is now orbiting the sun independently. Its elongated orbit has the same mean distance from the sun as Earth's orbit, but its maximal distance from the sun is twice its minimal distance. Compare its speeds at the maximal and minimal distances from the sun. Will it ever meet up with Earth again?

Answer: It is moving half as fast at its maximal distance from the sun as at its minimal distance, and it will meet Earth again in exactly 1 year.

Why: In accordance with Kepler's first law, the spacecraft is traveling in an elliptical orbit. Since Kepler's second law requires it to sweep out area at a constant rate, the spacecraft must move half as fast when it is twice as far from the sun. Finally, because the spacecraft's mean distance from the sun is the same as that of Earth, the two have the same orbital periods. They'll both complete their orbits in one year and meet up at that time.

Travel to the Stars: Special Relativity

4 In 1905, while Albert Einstein (German-born Swiss then American physicist, 1879–1955) was working as a patent examiner in Bern, he published four revolutionary papers in three different areas of physics. For the 26-year-old doctoral student from Germany, it was quite a banner year. Though Einstein is often portrayed as an elderly, wild-haired gentleman, his most important contributions to science were made when he was a vibrant young man who had married his first wife only just two years earlier.

Despite formidable challenges, it may one day be possible for human-crewed spacecraft to venture away from the solar system and travel to the stars. The distances involved are so vast that the only way to cover them in an astronaut's lifetime would be to move at fantastic speeds, speeds comparable to that of light itself.

Should spacecraft one day be able to attain such enormous speeds, they'll find that the basic laws of motion, the Galilean and Newtonian laws that we have been learning up to this point, are incomplete. Although extremely accurate at ordinary speeds, those laws falter near the **speed of light** (exactly 299,792,458 m/s). They turn out to be low-speed approximations for the more accurate laws of motion developed by Einstein in 1905 (**4**). Built on the observation that light always travels at the same speed, regardless of an observer's inertial frame of reference, these **relativistic laws of motion** are accurate at any speed. They are part of Einstein's **special theory of relativity**, the conceptual framework that describes space, time, and motion in the absence of gravity.

We saw in Section 1.1 that observers in different inertial frames of reference can disagree on an object's position and velocity. Special relativity recognizes that those observers can also disagree on the distance and time separating two events. More broadly, two inertial observers who are in relative motion perceive space and time somewhat differently. If they're moving at ordinary speeds, that difference in perceptions is negligible and the Newtonian laws of motion are nearly perfect. But if they're moving relative to one another at a substantial fraction of the speed of light, then they perceive space and time quite differently. In that case, the Newtonian approximations fail and the full laws of relativity are required.

Special relativity has many consequences for high-speed space travel, but we'll concentrate on how relativity alters two familiar conserved quantities: momentum and energy. At low speeds, our spacecraft's momentum takes its usual Newtonian value: mass times velocity (Eq. 2.3.1). But with increasing speed, a new relativistic factor enters the picture: $(1 - \text{speed}^2/\text{light speed}^2)^{-1/2}$. **Relativistic momentum** is equal to the object's mass times its velocity times that factor. This relationship can be written as a word equation:

$$\text{relativistic momentum} = \frac{\text{mass} \cdot \text{velocity}}{\sqrt{1 - \text{speed}^2/\text{light speed}^2}}, \quad (4.2.3)$$

in symbols:

$$\mathbf{p} = \frac{m \cdot \mathbf{v}}{\sqrt{1 - v^2/c^2}},$$

and in common language:

For a spacecraft to reach the speed of light, its momentum would have to be infinite.

At ordinary speeds, the relativistic factor is so nearly equal to 1 that this relativistic relationship is beautifully approximated by the Newtonian one. However, as the spacecraft's speed begins to approach that of light itself, the relativistic factor spoils the simple proportionality between momentum and velocity. Momentum then increases more rapidly than velocity. One result of this change is that it becomes impossible to reach the speed of light, let alone exceed it. Even if the spacecraft's thrust increases its forward momentum at a steady rate, its speed will increase less and less quickly. It will approach—but never reach—the speed of light.

A similar change happens to the spacecraft's energy as it approaches the speed of light. At low speeds, our isolated spacecraft's kinetic energy takes its usual Newtonian value:

half its mass times the square of its speed (Eq. 2.2.1). At high speeds, however, we must begin using **relativistic energy**. Relativistic energy is equal to the object's mass times the square of the speed of light times the relativistic factor. This relationship can be written as a word equation:

$$\text{relativistic energy} = \frac{\text{mass} \cdot \text{light speed}^2}{\sqrt{1 - \text{velocity}^2/\text{light speed}^2}}, \quad (4.2.4)$$

in symbols:

$$E = \frac{m \cdot c^2}{\sqrt{1 - v^2/c^2}},$$

and in common language:

A spacecraft starts with a rest energy and its energy grows toward infinity as its speed approaches the speed of light.

At ordinary speeds, the spacecraft's relativistic energy can be approximated as:

$$\text{relativistic energy} \approx \text{mass} \cdot \text{light speed}^2 + \frac{1}{2} \cdot \text{mass} \cdot \text{speed}^2.$$

The usual Newtonian kinetic energy appears at the right in this approximation, but to its left is a new energy that we've never seen before. Called the *rest energy*, it's present even when the spacecraft is motionless. Because the rest energy is constant, it doesn't affect low-speed motion and was overlooked in the Newtonian laws. However, this energy associated with mass itself (stated symbolically as $E = mc^2$) does have consequences and is surely the most famous feature of the special theory of relativity.

The relativistic version of energy has two implications for our spacecraft. First, the spacecraft's energy increases so quickly as it nears the speed of light that it can never reach that speed. Second, the spacecraft's initial store of energy before launch is associated with its initial mass. That mass and energy are so closely related is something to which we'll return in Chapter 15.

► Check Your Understanding #5: Stellar Tugboats

A stellar tugboat is pulling forward on a spacecraft that is already traveling close to the speed of light. Although the tugboat is steadily increasing the spacecraft's forward momentum, the spacecraft is speeding up less and less quickly. Why?

Answer: Although the spacecraft's forward momentum increases steadily, in accordance with the relativistic relationship between momentum and velocity, the spacecraft's velocity increases ever more slowly.

Why: At ordinary speeds, a steady transfer of momentum to an object will yield a steady increase in its speed. Near the speed of light, however, the object's increase in speed no longer keeps pace with its increase in momentum.

► Check Your Figures #2: The Real Speed Limit

A spacecraft passes your planetary base going half the speed of light. Using special equipment, you transfer exactly enough momentum to that spacecraft to double its forward momentum. How fast is it going now?

Answer: It is going $\sqrt{4/7}$ times the speed of light.

Why: When the spacecraft is traveling at $\frac{1}{2}$ the speed of light, Eq. 4.2.3 gives its forward momentum as $\sqrt{1/3}$ times its mass times the speed of light. Doubling that momentum and solving Eq. 4.2.3 for the velocity yields $\sqrt{4/7}$ times the speed of light.

 Check Your Figures #3: It Packs a Wallop

The speed of a 1-kg minispacecraft is $\sqrt{3}/4$ times the speed of light. What is its relativistic energy and what fraction of that energy is rest energy?

Answer: Its total energy is 1.8×10^{17} J, of which half is rest energy.

Why: For an object traveling at $\sqrt{3}/4$ times the speed of light, Eq. 4.2.4 gives a total energy of 2 times its mass times the square of the speed of light. For a 1-kg object, that energy is 1.8×10^{17} J. Its rest energy is simply its mass times the square of the speed of light, or 9.0×10^{16} J. Both values of energy are astonishingly large.

Visiting the Stars: General Relativity

In its travels near stars and other massive celestial objects, our spacecraft is likely to encounter another surprise: the Newtonian view of gravity is also an approximation! Near extremely dense, massive objects, gravity is no longer accurately described by Newton's law of universal gravitation. Instead, understanding gravity requires a new conceptual framework that Einstein first presented in 1916, the **general theory of relativity**.

This new framework is based on the observation that you can't distinguish between downward gravity and upward acceleration. As we found in Section 3.3, they feel exactly the same. For example, if you feel heavy as you stand inside a closed spacecraft, you can't be sure whether you're experiencing a weight due to downward gravity or a feeling of acceleration due to upward acceleration. In fact, the spacecraft's scientific instruments can't help you because they can't distinguish the effects of gravity from those of acceleration either. No matter how hard you try, you can't tell the difference.

At the heart of this problem is the concept of mass. Up to this point, we have seen mass play two apparently different roles that we can refer to as gravitational mass and inertial mass. When you're experiencing a weight, your **gravitational mass** is acting together with gravity to make you feel heavy. When you're experiencing a feeling of acceleration, your **inertial mass** is acting together with acceleration to make you feel heavy. However, in spite of their different roles, these two masses seem to be related. Without exception, an object that has a large inertial mass and is therefore difficult to shake back and forth also has a large gravitational mass and is therefore hard to support against gravity. In fact, the two masses seem to be the same. That observation led Einstein to propose the **principle of equivalence**, that these two masses, gravitational and inertial, are truly identical and therefore that no experiment you perform inside your spacecraft can distinguish between free fall and the absence of gravity. The general theory of relativity is based on this principle of equivalence.

As long as your spacecraft stays in regions of weak gravity, Newton's law of universal gravitation will adequately describe its motion. At the extremes of gravity, however, the general theory of relativity is necessary. That theory describes a universe in which massive objects distort the structure of nearby space and time, and in which extreme masses produce extreme distortions. One of the most startling predictions of this theory is the existence of objects so radical in their gravitational warping of nearby space and time that they are **black holes**, spherical or nearly spherical surfaces from which not even light can escape. A number of black holes have been discovered, including an enormous one at the center of our galaxy. You might want to avoid them.

 Check Your Understanding #6: Rip-van-Twinkle

You awaken aboard your small spacecraft only to discover that you have been asleep for almost 20 years and the crew has vanished. The windows are closed, and you have no idea where your spacecraft is located or what it is doing. You realize that you feel pulled toward the floor of the spacecraft. Are you experiencing weight or a feeling of acceleration?

Answer: Trick question! You can't possibly determine the answer!

Why: The central principle of general relativity is that no experiment you make inside your spacecraft can distinguish between the consequences of gravity and the consequences of acceleration. You're simply going to have to look out the window to figure it out.

Epilogue for Chapter 4

This chapter has discussed the physical concepts behind two types of machines. In Bicycles, we investigated the concept of dynamic stability, where an object that falls over when stationary becomes remarkably stable in motion. We also saw how the need for mechanical advantage between the pedals and the wheels led to the development of the modern multispeed bicycle.

In Rockets and Space Travel, we studied the ways in which reaction forces propel rockets forward. We found that a rocket's ultimate speed is not limited by its exhaust velocity, allowing it to lift itself into orbit around Earth. We also learned that gravity weakens as the distance between two gravitating objects increases, making it possible for a spaceship to break free from Earth's gravity and begin orbiting the sun. Finally, we took a brief look at the exotic physics a spacecraft would encounter if it approached the speed of light or passed through the extreme gravity near some celestial objects.

Explanation: High-Flying Balls

The top ball doesn't bounce off the ground; it bounces off the bottom ball. The top ball actually completes its bounce as the bottom ball is heading upward. After colliding, the two balls push off one another in just the same way that a rocket pushes off its exhaust. The balls exchange momentum and energy as they push against one another, and the small top ball—which has less mass and accelerates much more easily than the bottom ball—ends up with a large upward momentum and more than its fair share of energy. It flies upward as though it were hit by a massive upward-moving bat. In fact, it has been hit by a massive upward-moving ball, and it rebounds at high speed.

If the small ball's mass were really negligible compared to that of the large ball and the balls bounced without wasting any energy, then the small ball would rebound from the big ball traveling three times as fast as when it arrived. Its kinetic energy would be nine times as great as before, and it would rebound to nine times its original height. Of course, real balls aren't perfect, so the tennis ball won't bounce quite so high. Still, the effect is pretty impressive.

Chapter Summary and Important Laws and Equations

How Bicycles Work: While a stationary bicycle tips over easily, a moving bicycle is remarkably stable. It remains upright with the help of two stabilizing effects of motion: one due to gyroscopic precession and the other due to the shape and angle of the front fork. These effects work together to steer the bicycle in the direction that it's leaning. Whenever it tips, it automatically drives under its center of gravity and returns to upright. Leaning is also an essential part of turns. By leaning his body and/or the bicycle properly during a turn, the bicyclist ensures that there is no overall torque on the bicycle and that it doesn't tip over.

The rider powers the bicycle by pushing its pedals around in a circle. Since the rider can provide the most pedaling power when the pedals are turning at a moderate rate and he is pushing on them with moderate forces, a modern multispeed bicycle allows him to adjust the relative rotation rates of the pedals and the wheels. By choosing the right gear, the bicycle allows him to provide his peak power comfortably.

How Rockets Work: A rocket obtains its thrust by ejecting gas from an engine in its tail. The rocket pushes on the gas, and the gas pushes back. This gas usually comes from burning chemical fuels contained entirely inside the rocket itself and is released from the rocket's engine through a nozzle. A carefully designed nozzle permits the rocket to make efficient use of the energy stored in the fuel by ensuring that gas leaves the rocket at the maximum possible speed. The rocket's thrust is used to lift the rocket against the force of gravity and to accelerate it upward. Once in space, with its engine inactive, the rocket will orbit Earth or the sun, or it may travel off into interstellar space.

- 1.** Law of universal gravitation: Every object in the universe attracts every other object in the universe with a force equal to the gravitational constant times the product of the two masses, divided by the square of the distance separating the two objects, or

$$\text{force} = \frac{\text{gravitational constant} \cdot \text{mass}_1 \cdot \text{mass}_2}{(\text{distance between masses})^2}. \quad (4.2.2)$$

- 2.** Kepler's first law: All planets move in elliptical orbits, with the sun at one focus of the ellipse.
- 3.** Kepler's second law: A line stretching from the sun to a planet sweeps out equal areas in equal times.
- 4.** Kepler's third law: The square of a planet's orbital period is proportional to the cube of that planet's mean distance from the sun.

- 5.** Relativistic momentum: An object's relativistic-momentum is equal to its mass times its velocity times the relativistic factor, or

$$\text{relativistic momentum} = \frac{\text{mass} \cdot \text{velocity}}{\sqrt{1 - \text{speed}^2/\text{light speed}^2}}. \quad (4.2.3)$$

- 6.** Relativistic energy: An object's relativistic energy is equal to its mass times the square of the speed of light times the relativistic factor, or

$$\text{relativistic energy} = \frac{\text{mass} \cdot \text{light speed}^2}{\sqrt{1 - \text{velocity}^2/\text{light speed}^2}}. \quad (4.2.4)$$

So far all the everyday objects we've examined have been solids. However, since gases and liquids are also important parts of the world around us—as the air we breathe, the water we swim in, and even the blood we pump through our veins—we now turn to objects that, unlike solids, don't have well-defined shapes. These objects are called fluids, and the study of fluid behavior and motion is a broad field, extending across the sciences and engineering. Fluid dynamics, often called hydrodynamics, is as important to an oil-well engineer as to an animal physiologist or a stellar astrophysicist. The tools used to analyze fluids are somewhat more complicated than for solids because fluids themselves are more complicated—it's hard to exert a force directly on them, and even if we could, they usually don't move as a single rigid object. In this chapter, we'll look at some of the concepts and tools needed to understand their complex behaviors.

ACTIVE LEARNING EXPERIMENTS**A Cartesian Diver**

One of these concepts is buoyancy: an object immersed in a fluid experiences an upward force from that fluid. This buoyant force is what lifts a helium balloon into the sky and suspends a boat on the surface of water. How the object responds to buoyancy depends on the relative densities of the object and the fluid surrounding it, where density is the ratio of mass to volume. As we'll see in this chapter, an

object that's more dense than the surrounding fluid sinks, while one that's less dense than that fluid floats.

To see the importance of density in determining whether an object sinks or floats, you can construct a simple toy called a Cartesian diver. This once-popular parlor gadget consists of a small air-filled vial floating in a sealed container of water. Normally, the air bubble inside



Courtesy Lou Bloomfield



the vial keeps it floating at the surface of the water, but whenever you squeeze the container, the vial sinks.

To make a Cartesian diver, you'll need only a plastic soda bottle and a small vial that's open at one end. The vial can be made of almost anything—plastic, metal, or glass—as long as it's dense enough to sink in water. Fill the plastic soda bottle with water and float the vial in it upside down; air trapped inside the vial should keep the vial afloat. Now slowly reduce the size of the air bubble inside the vial until the vial barely floats. You can make this adjustment by tipping the vial to let some of its air escape or by removing it from the bottle and pouring water into it. Once you have the vial floating only a few millimeters out of the water at the top of the soda bottle, cap the bottle and prepare to test your diver.

Before you squeeze the soda bottle, think about how squeezing it will affect the air bubble. Now squeeze the bottle gently and watch the air bubble. How does the bubble's size depend on how hard you squeeze the bottle? Why should the two be related? Squeeze the bottle hard enough that the vial sinks. Why is there a relationship between the size of the air bubble and the diver's height in the water?

As you release the pressure on the bottle, the diver will float back up to the surface. Why does the sunken diver suddenly become buoyant again? By carefully squeezing the bottle, you can even make the diver hover in the middle of the bottle. Try making the diver hover while your eyes are closed. Why is hovering so difficult to sustain? Why must you watch the diver to make it hover?

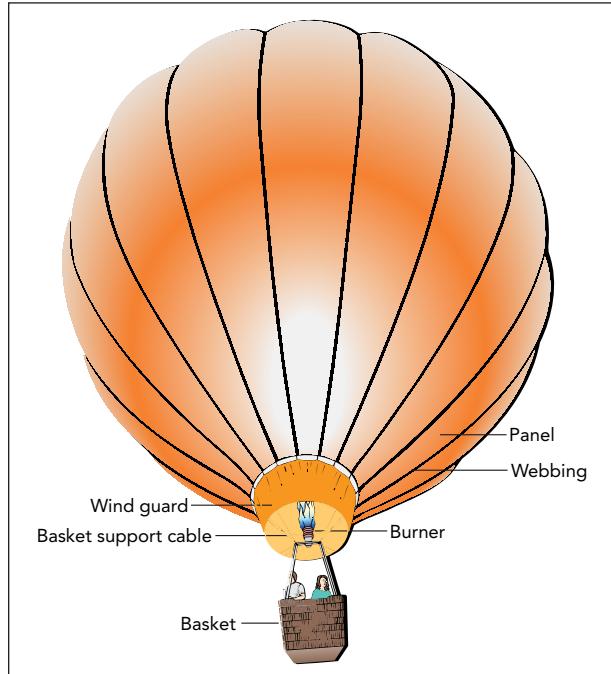
Chapter Itinerary

We'll return to the diver at the end of the chapter. First, we examine two things from the world around us: (1) *balloons* and (2) *water distribution*. In Balloons, we explore how the concepts of pressure and buoyancy help explain how Earth's atmosphere keeps hot-air and helium balloons from falling to the ground. In Water Distribution, we see both how pressure propels water through plumbing and the ways in which water can contain energy. For a more complete preview of this chapter, jump ahead to the Chapter Summary and Important Laws and Equations at the end of the chapter.

The issues we'll be looking at crop up frequently in our everyday experiences. Pressure plays an important role in aerosol cans, steam engines, firecrackers, and even the weather; buoyancy supports ships on water and keeps oil above vinegar in a bottle of salad dressing. Just as important, these concepts lay the groundwork for Chapter 6, where we'll examine objects in which motion affects the behavior of fluids.

SECTION 5.1

Balloons



Because gravity gives every object near Earth's surface a weight proportional to its mass, objects fall when you drop them. Why then does a helium-filled balloon—which, after all, is just another object with a mass and a weight—sail upward into the sky when you let go of it? Does the balloon have a negative mass and a negative weight, or are we forgetting something?

We're forgetting air—specifically, the layer of air that sits atop Earth's surface and is held in place by gravity. Since this air is difficult to see and moves out of our way so easily, we often forget that it's there. But air sometimes makes itself noticeable. When you ride a bicycle, you feel its forces; when you blow up a beach ball, you see that it takes up space. And when you release a helium balloon, air lifts the balloon upward.

Questions to Think About: Since most objects fall to the ground through the atmosphere, why doesn't the atmosphere itself fall downward? Why is the air "thinner" in the mountains than at sea level? If you suck all the air out of a plastic bag, what squeezes the bag into a thin sheet? Why does blowing air into the bag make it inflate? What happens to the bag's total mass when you fill it with air? with hot air? with helium? If you took a sturdy helium balloon to the moon, where there is no air, and then released it, which way would it move?

Experiments to Do: Pull on the string of a helium balloon to get a feel for how it behaves. If it pulls upward on your fingers, does that mean its weight (and mass) is negative? How would an object with a negative mass respond to a force? Shake the balloon and convince yourself that the balloon's mass is positive. Do you think there are any objects with negative masses?

Since the balloon's mass and weight are both positive, gravity must be pulling the balloon downward. But how can the stationary balloon pull upward on your finger? What other

forces might be pushing the balloon upward? You can enhance these upward forces by partially submerging the balloon in a container of water. Where else do similar upward forces appear in everyday life?

Take the helium balloon for a ride in a car. Which way does the balloon move when you start suddenly? when you stop suddenly? Again, it seems as though the balloon's mass is negative. What is pushing on the balloon to make it move in this counterintuitive way?

Air and Air Pressure

Hot-air and helium balloons don't really defy gravity. In fact, they're not even close to weightless. But even though these balloons may weigh hundreds of pounds, something supports their weight and holds them aloft—the surrounding air. To understand balloons, we must start by understanding air.

Like the objects we've already studied, air has weight and mass. If you don't believe me, compare two identical scuba tanks, one freshly filled with air and the other empty. If you weigh each tank carefully, you'll find that the full tank is heavier than the empty tank—an indication that air has weight. If you shake each tank, you'll find that the full tank is harder to accelerate than the empty tank—an indication that air has mass. So air really is like any other object ... or almost.

What makes air different from our previous objects is that air has no fixed shape or size. You can mold 1 kg of air into any form you like, and it can occupy a wide range of **volumes**. Air is **compressible**, that is, you can squeeze a certain mass of it into almost any space. For example, 1 kg of air could fill a single scuba tank or a whole basketball arena.

This flexibility of size and shape originates in the microscopic nature of air. Air is a **gas**, a substance consisting of tiny, individual particles that travel around independently. These individual particles are atoms and molecules. An **atom** is the smallest portion of an element that retains all the chemical characteristics of that element; a **molecule**, an assembly of two or more atoms, is the smallest portion of a chemical compound that retains all the characteristics of that compound. A molecule's atoms are held together by **chemical bonds**, linkages formed by electromagnetic forces between the atoms.

Air particles are extremely small, less than 1 millionth of a millimeter in diameter. Most are nitrogen and oxygen molecules, but others include carbon dioxide, water, methane, and hydrogen molecules, and also argon, neon, helium, krypton, and xenon atoms. Those particular atoms, which don't make strong chemical bonds and rarely form molecules, are called **inert gases** because of their chemical inactivity.

Like tiny marbles, these air particles have sizes, masses, and weights. But, while marbles quickly settle to the ground when you spill them from a bag, air particles don't seem to fall at all. Why don't they pile up on Earth's surface?

The answer has to do with air's thermal energy, the portion of air's internal energy that's associated with temperature. Air's individual particles have such minuscule masses that, even at room temperature, they exhibit frenetic **thermal motion** (Fig. 5.1.1a); thermal energy keeps them moving, spinning, and ricocheting off one another at bullet-like speeds of roughly 500 m/s (1100 mph). Their frequent collisions prevent them from making much progress in any particular direction and, between collisions, they travel in nearly straight-line paths because gravity doesn't have time to make them fall very far. This vigorous thermal motion spreads the air particles apart, so they don't accumulate on the ground. Real marbles, however, are too massive to exhibit noticeable thermal motion and fall to the ground in heaps.

Let's ignore gravity for the moment and consider what happens in a truck tire containing 1 kg of room-temperature air. The air particles whiz around inside the tire, and each

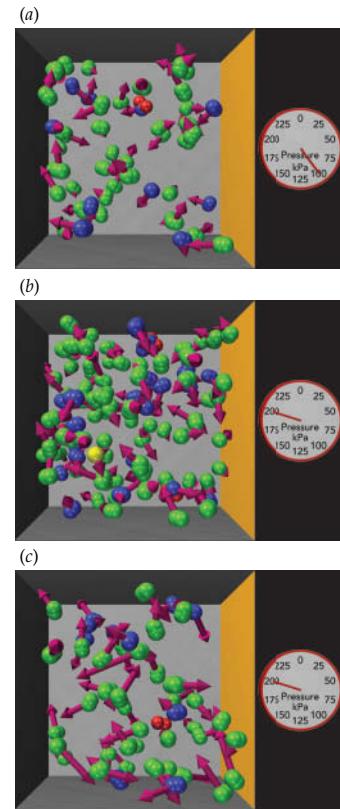


Fig. 5.1.1 (a) As air particles bounce off surfaces in this box, they exert pressure on those surfaces. (b) Packing the air particles more densely increases the number of particles that hit the surfaces each second and thus increases the air's pressure. (c) Increasing the air's temperature increases the speed of the air particles (show by the arrows) so that they hit the box's surfaces harder and more frequently, and thus increases the air's pressure.

time a particle bounces off a wall of the tire, it exerts a force on that wall. Although the individual forces are tiny, the number of particles is not, and together they produce a large average force. The size of this total force depends on the wall's **surface area**; the larger its surface area, the more average force it experiences. To characterize the air, however, we don't really need to know the wall's surface area; instead, we can refer to the average force the air exerts on each unit of surface area, a quantity called **pressure**:

$$\text{pressure} = \frac{\text{force}}{\text{surface area}}.$$

Pressure is measured in units of force per area. Since the SI unit of surface area is the **meter**² (abbreviated m² and often referred to as the *square meter*), the SI unit of pressure is the **newton per meter**². This unit is also called the **pascal** (abbreviated Pa), after French mathematician and physicist Blaise Pascal. One pascal is a small pressure; the air around you has a pressure of about 100,000 Pa (2100 lbf/ft² or 15 lbf/inch²), so it exerts a force of about 100,000 N on a 1-m² surface. Since 100,000 N (22,500 lbf) is about the weight of a city bus, air pressure can exert enormous forces on large surfaces.

Besides pushing on the walls of the tire, air also pushes on any object immersed in it. Its particles bounce off the object's surfaces, pushing them inward. As long as the object can withstand these compressive forces, the air won't greatly affect it since the uniform air pressure ensures that the forces on all sides of the object cancel one another perfectly. A sheet of paper, for example, will experience zero net force because the forces exerted on its two sides will sum to zero.

Air particles also bounce off one another, so that air pressure exerts forces on air, too. A cube of air inserted into the tire experiences all of the inward forces that a cube of metal would experience. The air around the cube pushes inward on it, and the cube pushes outward on the air around it. Since the net force on the cube of air is zero, the cube doesn't accelerate.

Check Your Understanding #1: Getting a Grip on Suction

After you push a suction cup against a smooth wall, the elastic cup bends back and a small, empty space is created between the cup and the wall. What keeps the suction cup against the wall?

Answer: Air pressure keeps it against the wall.

Why: Because the space between the suction cup and the wall is empty, the pressure there is zero. Pressure of the surrounding air exerts large inward forces on the outsides of both the cup and the wall, squeezing them together. As long as there is no air between them to push outward, the cup and wall remain tightly attached. Once air leaks into the suction cup, it's easily detached from the wall.

Pressure, Density, and Temperature

Since air pressure is produced by bouncing air particles, it depends on how often, and how hard, those particles hit a particular region of surface. The more frequent or harder the impacts, the greater is the air pressure.

To increase the rate at which air particles hit a surface, we can pack them more tightly. Suppose we use an air pump to add another 1 kg of room-temperature air to the tire. Since that truck tire is sturdy and stiff, its volume doesn't change significantly as the number of air molecules inside it doubles. However, doubling the number of air particles in the same volume doubles the rate at which they hit each surface and therefore doubles the pressure (Fig. 5.1.1b). Air's pressure is thus proportional to its **density**, that is, its mass per unit of volume:

$$\text{density} = \frac{\text{mass}}{\text{volume}}.$$

Since the SI unit of volume is the **meter³** (abbreviated m³ and often referred to as the *cubic meter*), the SI unit of density is the **kilogram per meter³** (abbreviated kg/m³). The air around you has a density of about 1.25 kg/m³ (0.078 lbm/ft³). Water, in contrast, has a much greater density of about 1000 kg/m³ (62.4 lbm/ft³).

We can also increase the rate at which air particles hit a surface by speeding them up (Fig. 5.1.1c). The hotter the air, the more thermal energy it contains and the faster its particles move. Thermal energy is a combination of **internal kinetic energy** in the particles' random thermal motion and **internal potential energy** stored as part of that random thermal motion. Since air particles are essentially independent, except during collisions, nearly all of air's thermal energy is internal kinetic energy.

If we double the internal kinetic energy of the calm air in the tire, we double the average kinetic energy of each particle. Because a particle's kinetic energy is proportional to the square of its speed, doubling its kinetic energy increases its speed by a factor of $\sqrt{2}$. As a result, each particle hits the surface $\sqrt{2}$ times as often and exerts $\sqrt{2}$ times as much average force when it hits. With each particle exerting $\sqrt{2} \cdot \sqrt{2}$, or two times as much average force, the pressure doubles. Calm air's pressure is thus proportional to the average kinetic energy of its particles. Even when air is moving as wind and has noninternal kinetic energy in its overall motion, its pressure is proportional to the average *internal* kinetic energy of its particles.

Temperature measures this average internal kinetic energy per particle; the hotter the air, the larger is the average internal kinetic energy per particle and the greater the air's pressure. That's why the tire's air pressure increases on hot days and when it heats up during highway driving.

The most convenient scales for relating air's temperature to air's pressure aren't the common **Celsius** (°C) and **Fahrenheit** (°F) scales; instead, they are special **absolute temperature scales**, in which the zero of the temperature scale is **absolute zero**, the temperature at which an object contains zero thermal energy. At absolute zero (−273.15 °C or −459.67 °F), air contains no internal kinetic energy at all and has no pressure. When you use an absolute temperature scale, air's pressure is proportional to its temperature.

The SI scale of absolute temperature is the **Kelvin** scale (K). The Kelvin scale is identical to the Celsius scale, except that it's shifted so that 0 K is equal to −273.15 °C. In addition to associating the zero of temperature with the zero of internal kinetic energy, the Kelvin scale avoids the need for negative temperatures. Room temperature is about 293 K.

Since air pressure is proportional to both the air's density and its absolute temperature, we can express the relationship among these quantities in the following way:

$$\text{pressure} \propto \text{density} \cdot \text{absolute temperature}. \quad (5.1.1)$$

This relationship is useful because it allows us to predict what will happen if we change the temperature or density of a specific gas, such as air. The relationship does have its limitations; in particular, it doesn't work if we compare the pressures of two different gases, such as air and helium, which differ in their chemical compositions. To make such a comparison, we'll need to improve on Eq. 5.1.1. We'll do that later when we examine helium balloons.

Even in describing a specific gas, Eq. 5.1.1 has other shortcomings. The main problem is that real gas particles aren't completely independent of one another. If the temperature drops too low, the particles begin to stick together to form a liquid and Eq. 5.1.1 becomes invalid. Despite its limitations, however, this simple relationship between pressure, density, and temperature will prove useful in understanding how hot-air balloons float. It will help us understand the basic structure of Earth's atmosphere, the origins of the upward force that keeps a hot-air balloon aloft, and the reason why hot air rises.

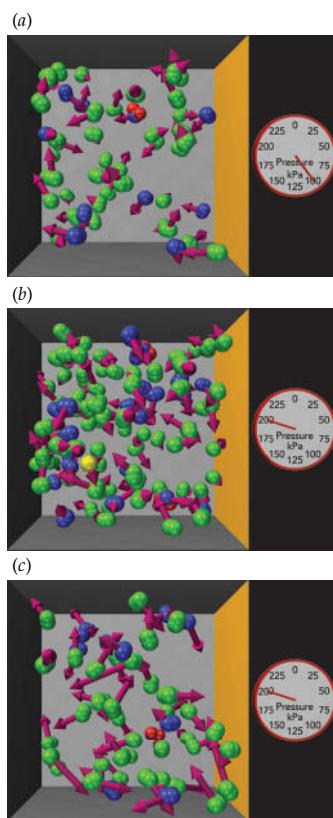


Fig. 5.1.1 (repeated)
 (a) As air particles bounce off surfaces in this box, they exert pressure on those surfaces. (b) Packing the air particles more densely increases the number of particles that hit the surfaces each second and thus increases the air's pressure. (c) Increasing the air's temperature increases the speed of the air particles (show by the arrows) so that they hit the box's surfaces harder and more frequently, and thus increases the air's pressure.

Quantity	SI Unit	English Unit	SI → English	English → SI
Area	meter ² (m ²)	foot ² (ft ²)	1 m ² = 10.764 ft ²	1 ft ² = 0.092903 m ²
Volume	meter ³ (m ³)	foot ³ (ft ³)	1 m ³ = 35.315 ft ³	1 ft ³ = 0.028317 m ³
Pressure	pascal (Pa)	pound-force per foot ² (lbf/ft ²)	1 Pa = 0.020885 lbf/ft ²	1 lbf/ft ² = 47.880 Pa
Density	kilogram per meter ³ (kg/m ³)	pound-mass per foot ³ (lbm/ft ³) [†]	1 kg/m ³ = 0.062428 lbm/ft ³	1 lbm/ft ³ = 16.018 kg/m ³



Check Your Understanding #2: Snacks That Go Pop in the Night

If you remove a partially filled container of food from the refrigerator and allow it to warm to room temperature, the lid will often bow outward and may even pop off. What has happened?

Answer: The pressure of the air trapped in the container increases as its temperature increases, causing the lid to bulge outward.

Why: Whenever a trapped quantity of gas changes temperature, it also changes volume or pressure or both. In this case, warming the air trapped in the container causes its pressure to increase. The unbalanced pressures inside and outside the container cause the lid to bow outward or even to pop off.

Earth's Atmosphere

Most of the mass of Earth's atmosphere is contained in a layer less than 6 km (4 miles) thick. Since Earth is 12,700 km (7,900 miles) in diameter, this layer is relatively thin—so thin that, if Earth were the size of a basketball, it would be no thicker than a sheet of paper.

The atmosphere stays on Earth's surface because of gravity. Every air particle, as we've seen, has a weight. Just as a marble thrown upward eventually falls back to the ground, so the particles of air keep returning toward Earth's surface. Although the particles are moving too fast for gravity to affect their motions significantly over the short term, gravity works slowly to keep them relatively near Earth's surface. An air particle, like a rapidly moving marble, may appear to travel in a straight line at first, but it will arc over and begin to fall downward eventually. Only the lightest and fastest moving particles in the atmosphere—hydrogen molecules and helium atoms—occasionally manage to escape from Earth's gravity and drift off into interplanetary space.

While gravity pulls the atmosphere downward, air pressure pushes the atmosphere upward (Fig. 5.1.2a). As the air particles try to fall to Earth's surface, their density increases and so does their pressure. It's this air pressure that supports the atmosphere and prevents it from collapsing into a thin pile on the ground.

To understand how gravity and air pressure structure the atmosphere, picture a 1-m² column of the atmosphere as though it were a tall stack of 1-kg air blocks (Fig. 5.1.2b). These blocks support one another with air pressure to form a stack of about 10,000 blocks. The bottom block must support the weight of all the blocks above it and is tightly compressed, with a height of about 0.8 m, a density of about 1.25 kg/m³, and a pressure of about 100,000 Pa. A block farther up in the stack has less weight to support and is less tightly compressed. The higher in the stack you look, the lower the density of the air and the less the air pressure.

The atmosphere has essentially the same structure as this stack of blocks. The air near the ground supports the weight of several kilometers of air above it, giving it a density of about 1.25 kg/m³ and a pressure of about 100,000 Pa. At higher altitudes, however, the air's density and pressure are reduced, since there is less atmosphere overhead and the air doesn't have to support as much weight. High-altitude air is thus “thinner” than low-altitude air.

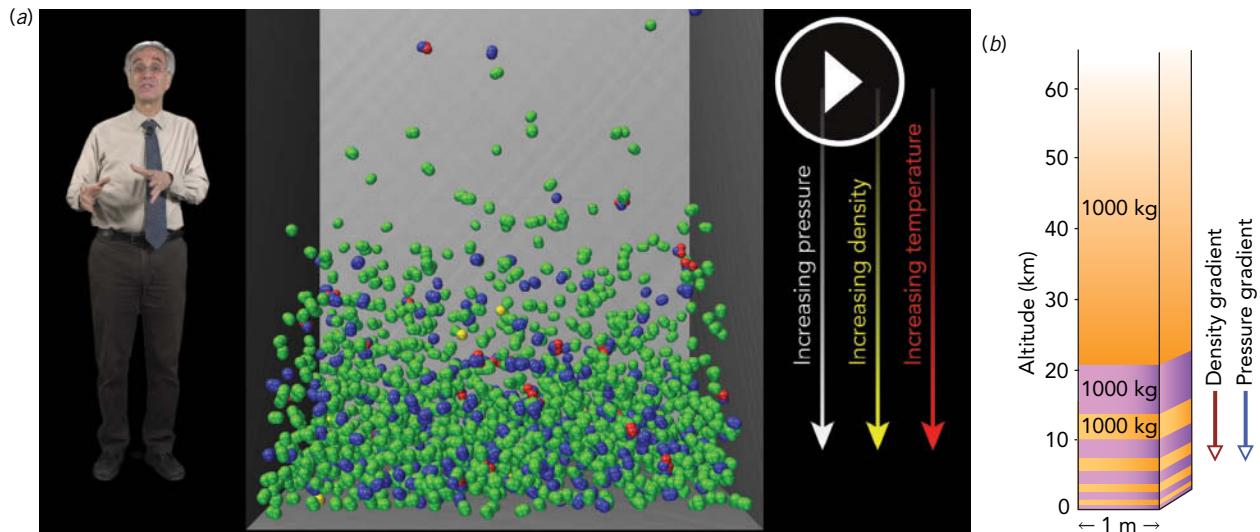


Fig. 5.1.2 (a) A model of Earth's atmosphere, showing that a competition between gravity and thermal energy gives rise to the structure of Earth's atmosphere, notably its increasing pressure, density, and temperature with decreasing altitude. (b) The air in a 1-m² column of atmosphere has a mass of about 10,000 kg. The bottom 1000 kg is the most tightly compressed because it supports the most weight above it. At higher altitudes, the air is less tightly compressed because it has less weight above it.

Since the atmosphere's density and pressure increase gradually with distance in the downward direction, the atmosphere has a downward **density gradient** and a downward **pressure gradient**. As we'll see shortly, that pressure gradient is what allows a balloon to float. Whatever the altitude, however, the pressure of the surrounding air is always referred to as **atmospheric pressure**. The atmosphere also has a downward temperature gradient that makes ballooning chilly at high altitudes; you might want to bring coffee or hot chocolate. We'll return to that temperature gradient in Section 7.3.

Check Your Understanding #3: Mountain Travel Is a Pain in the Ears

As you drive up and down in the mountains, you may feel a popping in your ears as air moves to equalize the pressures inside and outside your eardrum. What causes these pressure changes?

Answer: As you change altitude, the atmospheric pressure changes.

Why: The air inside your ears is trapped, so its temperature, density, and pressure are normally constant. As your altitude changes, the pressure outside your ear changes and your eardrum experiences a net force. It bows inward or outward, muting the sounds you hear and causing some discomfort. The pressure imbalance is relieved during swallowing, when air can flow into or out of your middle ear.

The Lifting Force on a Balloon: Buoyancy

So far we've examined air, air pressure, and the atmosphere. While it may seem that we've avoided dealing with balloons, these topics really are involved in keeping hot-air and helium balloons aloft. As we've seen, the air in Earth's atmosphere is a **fluid**, a shapeless substance with mass and weight. This air has a pressure and exerts forces on the surfaces it touches; that pressure is greatest near the ground and decreases with increasing altitude. Air pressure and its variation with altitude allow air to lift hot-air and helium balloons through an effect known as **buoyancy**.

Buoyancy was first described more than 2000 years ago by the Greek mathematician Archimedes (287–212 BC). Archimedes realized that an object partially or wholly immersed in a fluid is acted on by an upward **buoyant force** equal to the weight of the fluid it

displaces. **Archimedes' principle** is actually very general and applies to objects floating or submerged in any fluid, including air, water, or oil. The buoyant force originates in the forces that a fluid exerts on the surfaces of an object. We've seen that such forces can be quite large but tend to cancel one another. How then can pressure create a nonzero total force on an object, and why should that force be in the upward direction?

● ARCHIMEDES' PRINCIPLE

An object partially or wholly immersed in a fluid is acted on by an upward buoyant force equal to the weight of the fluid it displaces.

Without gravity, the forces would cancel each other perfectly because the pressure of a stationary fluid would be uniform throughout. But gravity causes a stationary fluid's pressure to increase with distance in the downward direction; the fluid has a downward pressure gradient. When nothing is moving, the air pressure beneath an object is always greater than the air pressure above it. Thus air pushes upward on the object's bottom more strongly than it pushes downward on the object's top, and the object consequently experiences an upward overall force from the air—a buoyant force.

How large is the buoyant force on this object? It's equal in magnitude to the weight of the fluid that the object displaces. To understand this remarkable result, imagine replacing the object with a similarly shaped portion of the fluid itself (Fig. 5.1.3a). Since the buoyant force is exerted by the surrounding fluid, not the object, it doesn't depend on the object's composition. A balloon filled with helium will experience the same buoyant force as a similar balloon filled with water or lead or even air. So replacing the object with a similarly shaped portion of fluid will leave the buoyant force on it unchanged.

However, a portion of fluid suspended in more of the same fluid doesn't accelerate anywhere; it just sits there, so the net force on it is clearly zero. It has a downward weight, but that weight must be canceled by some upward force that can come only from the surrounding fluid. This upward force is the buoyant force, and it's always equal in magnitude to the weight of the object-shaped portion of fluid displaced by the object.

This buoyant principle explains why some objects float while others sink. An object placed in a fluid experiences two forces: its downward weight and an upward buoyant force. If its weight is more than the buoyant force, it will accelerate downward (Fig. 5.1.3b); if its weight is less than the buoyant force, it will accelerate upward (Fig. 5.1.3c). If the two forces are equal, it won't accelerate at all and will maintain a constant velocity. If the balloon in this latter case starts motionless, it will remain motionless and will hover at a constant velocity of zero.

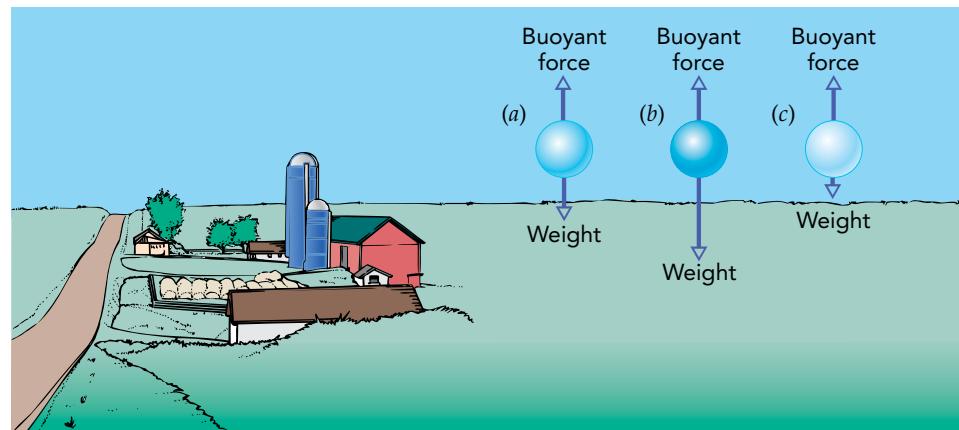


Fig. 5.1.3 (a) A portion of air immersed in that same air experiences an upward buoyant force equal to its weight and doesn't accelerate. (b) An object that is heavier than the air it displaces sinks, while (c) another object that is lighter than the air it displaces floats.

Whether an object will float in a fluid can also be viewed in terms of density. An object that has an average density greater than that of the surrounding fluid sinks, while one that has a lesser average density floats. A water-filled balloon, for example, will sink in air because water and rubber are denser than air. If you double the volume of the balloon, you double both its weight and the buoyant force on it, so it still sinks. The total volume of an object is less important than how its density compares to that of the surrounding fluid.

Check Your Understanding #4: Only Eyes above the Surface

Crocodiles sometimes swallow rocks, partly to aid their digestion but also to lower their heights in the water. To float with only its eyes above the water's surface, approximately what average density does a crocodile need?

Answer: It needs an average density just slightly less than the density of water.

Why: The crocodile should weigh a little less than the water it displaces when fully immersed. That way, it will experience a net upward force and will rise to the water's surface. Once its eyes are above the surface and not displacing water, the net force on the crocodile will drop to zero and it will hover. Since the crocodile's mass is slightly less than an equal volume of water, its average density is slightly less than that of water.

Check Your Figures #1: Why People Don't Float in Air

If a person displaces 0.08 m^3 (2.8 ft^3) of air, what is the buoyant force he experiences?

Answer: It is about 1 N (0.22 lbf).

Why: Air exerts a buoyant force on a person equal to the weight of the air he displaces. The density of air near sea level is about 1.25 kg/m^3 , so 0.08 m^3 of air has a mass of about 0.1 kg (1.25 kg/m^3 times 0.08 m^3) and a weight of about 1 N. So the upward buoyant force on him is about 1 N. This buoyant force due to the air is real and reduces the weight you read when you stand on a scale by about 0.125%.

Hot-Air Balloons

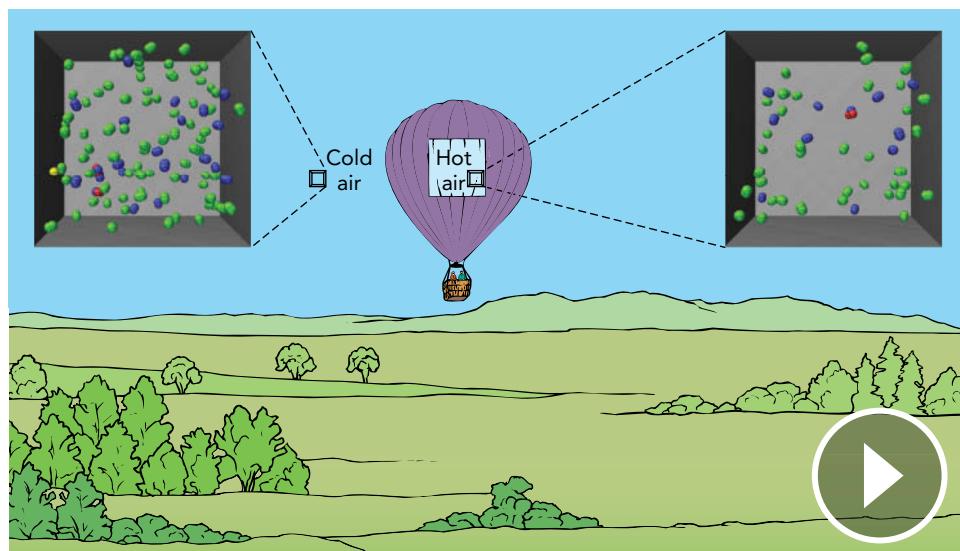
Since air is very light, with a density of only 1.25 kg/m^3 (0.078 lbm/ft^3), few objects float in it. One of these rare objects is a balloon with a vacuum inside it. Assuming that the balloon has a very thin outer shell or *envelope*, it will weigh almost nothing and have an average density near zero. Because its negligible weight is less than the upward buoyant force it experiences, the empty balloon will float upward nicely.

Unfortunately, this empty balloon won't last long. Because it's surrounded by atmospheric-pressure air, each square meter of its envelope will experience an inward force of 100,000 N. With nothing inside the balloon to support its envelope against this crushing force, it will smash flat. A thick, rigid envelope might be able to withstand the pressure of the surrounding air, but then the balloon's average density would be large and it would sink. So an empty balloon won't work.

What will work is a balloon filled with something that exerts an outward pressure on the envelope equal to the inward pressure of the surrounding air. Then each portion of the envelope will experience zero net force and the balloon will not be crushed. We could fill the balloon with outside air, but that would make its average density too high. Instead, we need a gas that has the same pressure as the surrounding air but a lower density.

One gas that has a lower density at atmospheric pressure is hot air. Filling our balloon with hot air takes fewer particles than filling it with cold air, since each hot-air particle is moving faster and contributes more to the overall pressure than does a cold-air particle. A hot-air balloon contains fewer particles, has less mass, and weighs less than it would if it contained cold air (Fig. 5.1.4). Now we have a practical balloon with an average density less than that of the surrounding air. The buoyant force it experiences is larger than its weight, and up it goes.

Fig. 5.1.4 A cube of hot air contains fewer air particles than an equivalent cube of cold air at the same pressure. Since it weighs less than the cold air it displaces, the hot air inside the balloon experiences an upward buoyant force that is greater than its downward weight.



© Courtesy Lou Bloomfield



Fig. 5.1.5 The bottom of a hot-air balloon is open so that heated air can flow in and cold air can flow out. The heated air displaces more than its weight in cold air and makes the balloon lighter.

Because the air pressure inside a hot-air balloon is the same as the air pressure outside the balloon, the air has no tendency to move in or out (an issue we will cover in the next section), and the balloon doesn't need to be sealed (Fig. 5.1.5). A large propane burner, located beneath the balloon's open end, heats the air that fills the envelope. The heated air expands to fill more volume at the same pressure and some of it flows out the balloon's open bottom.

The hotter the air in the envelope, the lower its density and the less the balloon weighs. The balloon's pilot controls the flame so that the balloon's weight is very nearly equal to the buoyant force on the balloon. If the pilot raises the air's temperature, particles leave the envelope, the balloon's weight decreases, and the balloon rises. If the pilot allows the air to cool, particles enter the envelope, the balloon's weight increases, and the balloon descends.

Even if the pilot heats the air until it is very hot, the balloon won't rise upward forever. As the balloon ascends, the air becomes thinner and the pressure decreases both inside and outside the envelope. Although the balloon's weight decreases as the air thins out, the buoyant force on it decreases even more rapidly and it becomes less effective at lifting its cargo. When the air becomes too thin to lift the balloon any higher, the balloon reaches a *flight ceiling* above which it can't rise. For each hot-air temperature, then, there is a cruising altitude at which the balloon will hover. When the balloon reaches that altitude, it's in a stable equilibrium. If the balloon shifts downward for some reason, the net force on it will be upward; if it shifts upward, the net force on it will be downward.

Because a balloon's fabric envelope ages quickly at high temperatures, the balloon's air mustn't be heated above about 120 °C (248 °F). Lower air temperatures prolong the life of the envelope, and a pilot usually tries to reduce the air temperature required by reducing the weight of the balloon's cargo. If you want to bring yet another friend on board, you might have to leave the champagne behind.

Check Your Understanding #5: Ballooning Weather

Can a hot-air balloon lift more on a hot day or a cold day?

Answer: It can lift more on a cold day.

Why: On a cold day, the outside air is relatively dense and the buoyant force on a balloon is larger than it would be on a hot day. The hot air in the balloon will cool off more quickly on a cold day, but the balloon will be able to carry a heavier load. Airplanes also fly better on cold days.

Helium Balloons

Although the particles in hot and cold air are similar, there are fewer of them in each cubic meter of hot air than in each cubic meter of cold air. We call the number of particles per unit of volume the **particle density**,

$$\text{particle density} = \frac{\text{particles}}{\text{volume}},$$

and hot air has a smaller particle density than cold air (Fig. 5.1.4). Because they contain similar particles, hot air also has a smaller density than cold air and is lifted upward by the buoyant force.

There's another way to make one gas float in another: use a gas consisting of very light particles. Helium atoms, for example, are much lighter than air particles. When they have equal pressures and temperatures, helium gas and air also have equal particle densities. Since each helium atom weighs 14% as much as the average air particle, 1 m³ of helium weighs only 14% as much as 1 m³ of air. Thus a helium-filled balloon has only a fraction of the weight of the air it displaces, and the buoyant force carries it upward easily.

Why should air and helium have the same particle densities whenever their pressures and temperatures are equal? Because a gas particle's contribution to the pressure doesn't depend on its mass (or weight). At a particular temperature, each particle in a gas has the same average internal kinetic energy in its translational motion, regardless of its mass. Although a helium atom is much less massive than a typical air particle, the average helium atom moves much faster and bounces more often. As a result, lighter but faster-moving helium atoms are just as effective at creating pressure as heavier but slower-moving air particles.

Thus, if you allow the helium atoms inside a balloon to spread out until the pressures and temperatures inside and outside the balloon are equal, the particle densities inside and outside the balloon will also be equal (Fig. 5.1.6). Since the helium atoms inside the balloon are lighter than the air particles outside it, the balloon weighs less than the air it displaces, and it will be lifted upward by the buoyant force.

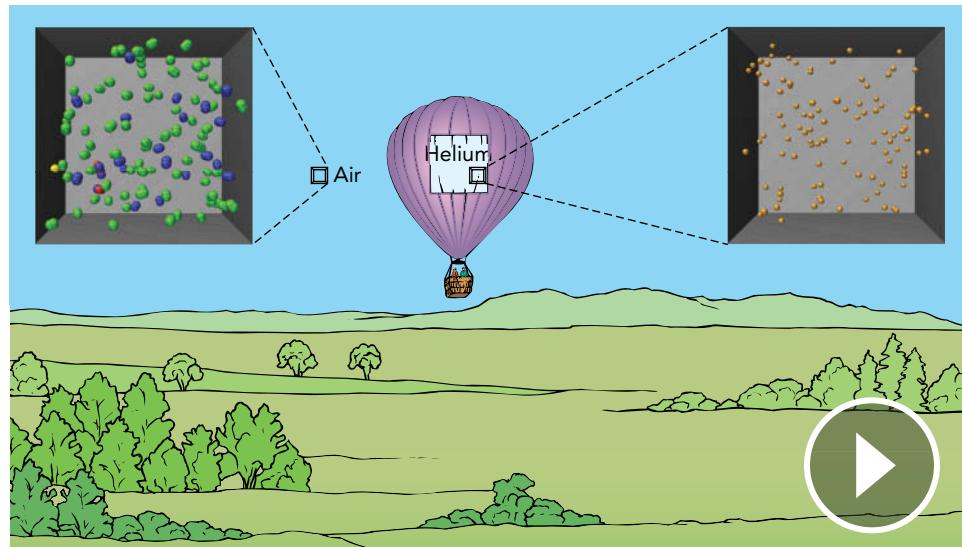


Fig. 5.1.6 A cube of helium contains the same number of particles as a cube of air at the same pressure and temperature. Since each helium atom weighs much less than the average air particle, the helium inside the balloon experiences an upward buoyant force that is greater than its downward weight.

The pressure of a gas is proportional to the product of its particle density and its absolute temperature, as the following formula indicates:

$$\text{pressure} \propto \text{particle density} \cdot \text{absolute temperature}. \quad (5.1.2)$$

This proportionality holds regardless of the gas's chemical composition. Our previous proportionality, Eq. 5.1.1, worked only as long as the gas's composition didn't change, so density and particle density remained proportional to one another. Now, however, we have a relationship with a wider applicability.

Equation 5.1.2, with an associated constant of proportionality, is called the **ideal gas law**. This law relates pressure, particle density, and absolute temperature for a gas in which the particles are perfectly independent. It's also fairly accurate for real gases in which the particles do interact somewhat. The constant of proportionality is the **Boltzmann constant**, with a measured value of $1.381 \times 10^{-23} \text{ Pa} \cdot \text{m}^3/(\text{particle} \cdot \text{K})$. Using the Boltzmann constant, the ideal gas law can be written as a word equation:

$$\text{pressure} = \text{Boltzmann constant} \cdot \text{particle density} \cdot \text{absolute temperature}, \quad (5.1.3)$$

in symbols:

$$p = k \cdot \rho_{\text{particle}} \cdot T,$$

and in everyday language:

Don't incinerate a spray can. A hot dense gas tends to burst its container.

THE IDEAL GAS LAW

The pressure of a gas is equal to the Boltzmann constant times the particle density times the absolute temperature.

1 Helium gas is obtained as a by-product of natural gas production from underground reservoirs in the United States, where it formed through the gradual radioactive decay of uranium and other unstable elements. While some of this gas is saved for industrial and commercial use, much is simply released into the atmosphere. The only other source of helium is the atmosphere, where helium is present at a level of 5 parts per million. Once the underground stores are consumed, helium will become a relatively rare and expensive gas.

2 Even helium-filled airships were easily destroyed by bad weather. The *Shenandoah*, one of two U.S. airships based on German designs, was destroyed by air turbulence on September 3, 1925, near Ava, Ohio. Crowds from a local fair immediately poured over the wreckage, collecting souvenirs.

Helium isn't the only "lighter-than-air" gas. Hydrogen gas, which is half as dense as helium, is also used to make balloons float. However, don't expect hydrogen to lift twice as much weight as helium. A balloon's lifting capacity is the difference between the upward buoyant force it experiences and its downward weight. Although the gas in a hydrogen balloon weighs half that in a similar helium balloon, the balloons experience the same buoyant force. Thus the hydrogen balloon's lifting capacity is only slightly more than that of the helium balloon. Hydrogen's main advantage is that it's cheap and plentiful, while helium is scarce (see **1**). Hydrogen is also dangerously flammable, so it's avoided in situations where safety is important. However, even helium-filled airships can have problems (see **2**).

Check Your Understanding #6: What Not to Put in a Balloon

A carbon dioxide molecule is heavier than an average air particle. If you pour carbon dioxide gas from a cup, which way does it flow in air, up or down?

Answer: It flows down.

Why: Carbon dioxide, found in carbonated beverages, dry ice, and fire extinguishers, is heavier than air because its molecules are heavier than air particles. The carbon dioxide you pour from the cup has the same pressure and temperature as the air around it and thus the same particle density. However, each carbon dioxide molecule weighs more, so the carbon dioxide is the denser gas (it has a higher mass density) and flows down in air. This tendency to flow along the floor makes carbon dioxide very good at extinguishing low-lying flames by depriving them of oxygen.

Check Your Figures #2: Popped Out of the Refrigerator

When you take an air-filled plastic container out of the refrigerator, it warms from 2 °C to 25 °C. How much does the pressure of the air inside it change?

Answer: It increases by 8.4%.

Why: To use Eq. 5.1.3 to determine the pressure change, we need temperatures measured on an absolute scale, such as the Kelvin scale. Since 0 °C is about 273 K, 2 °C is about 275 K and 25 °C is about 298 K. We can write Eq. 5.1.3 twice, once for each temperature:

$$\text{pressure}_{298 \text{ K}} = \text{Boltzmann constant} \cdot \text{particle density} \cdot 298 \text{ K},$$

$$\text{pressure}_{275 \text{ K}} = \text{Boltzmann constant} \cdot \text{particle density} \cdot 275 \text{ K}.$$

The particle density of the air in the container can't change as it warms up because its volume is fixed. Therefore we can divide the upper equation by the lower one and cancel out the Boltzmann constant and particle density on the right-hand side, giving:

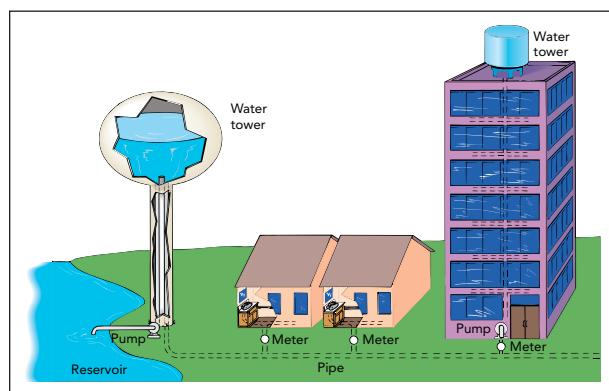
$$\frac{\text{pressure}_{298 \text{ K}}}{\text{pressure}_{275 \text{ K}}} = \frac{298 \text{ K}}{275 \text{ K}}$$

$$= 1.084.$$

The pressure in the container thus increases by a factor of almost 1.084, or about 8.4%. This elevated pressure will cause the container to emit a pop sound when you open it.

SECTION 5.2

Water Distribution



Now that we've explored the behavior of objects in fluids, let's turn to the behavior of fluids in objects. In this section, as we examine how plumbing distributes water, we'll see that pressure, density, and weight are just as important in plumbing as they are in ballooning. To keep things simple, we'll focus on the causes of water's motion through plumbing and leave the many complications of that motion itself for the next chapter. For the rest of this section, we'll ignore friction-like effects

and turbulence, and we'll let the water flow steadily most of the time.

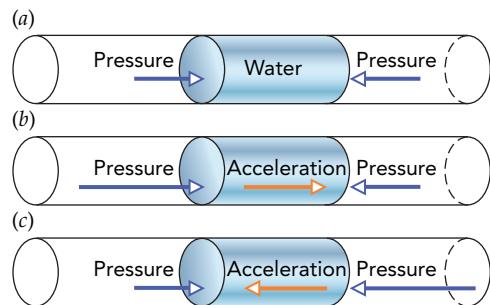
Questions to Think About: Why is the water pressure higher in the basement than it is in the attic? Why does a deep water well require a pump at the bottom? If a water tower is only a storage device, why is it so tall? Why do skyscrapers have complicated plumbing systems that include reservoirs at various levels in the building? What causes water to flow up through a drinking straw to your mouth?

Experiments to Do: To see the effects of pressure and weight on water, try these simple experiments with a drinking straw. First, suck water up the straw from a glass to your mouth. Are you exerting an attractive force on the water, or is some other force pushing it upward toward your mouth? With the straw full of water, seal the top with your finger; keeping the seal tight, remove the straw from the glass. What happens to the water inside? What happens to the water when you release the seal? Now blow into one end of a straw full of water while sealing the far end with your finger. When you release the seal on the far end, what happens to the water? What forces are responsible for this effect?

Water Pressure

Water distribution systems require two things: plumbing and water pressure. Plumbing is what delivers the water, and water pressure is what starts that water flowing. Water pressure is important because, like everything else, water has mass and accelerates only when pushed. If nothing pushed on the water when you opened a faucet, the water simply

Fig. 5.2.1 (a) If the water in a horizontal pipe is exposed to a uniform pressure, it will not accelerate. (b, c) If the pressure along the pipe is not uniform, however, the imbalance will create a net force on each portion of the water and the water will accelerate toward the side with the lower pressure.



wouldn't budge. Since the pushes that send water through pipes come principally from differences in water pressure, we need to look carefully at how this pressure is created and controlled.

We begin our study of water distribution by ignoring gravity. As we've seen with the atmosphere, gravity creates downward pressure gradients in stationary fluids—their pressures increase with depth and decrease with altitude. These downward pressure gradients complicate plumbing in hilly cities and skyscrapers. However, if all our plumbing is in a level region—for example, a single-story house in a very flat city—our job is much simpler. With no significant changes in height, the pressure gradients due to gravity produce negligible effects. In fact, we can neglect gravity altogether.

In this simplified situation, water accelerates only in response to unbalanced pressures. Just as unbalanced forces make a solid object accelerate, so too do unbalanced pressures make a fluid accelerate. If the water pressure inside a pipe is uniform, then each portion of water feels no net force and doesn't accelerate; it either remains stationary or coasts at constant velocity (Fig. 5.2.1a). If the pressure is out of balance, however, then each portion of water experiences a net force and accelerates toward the region of lowest pressure (Figs. 5.2.1b,c).

This acceleration doesn't mean that the water will instantly begin moving toward the lowest pressure. Because of its inertia, water changes velocity gradually; it speeds up, slows down, or turns to the side, depending on where the lowest pressure is located. A complicated arrangement of high and low pressures can steer water through an intricate maze of pipes, and that is exactly how water reaches your home from the city pumping station. Every change in its velocity during its trip through the plumbing is caused by a pressure imbalance.

You can create a pressure imbalance in water simply by squeezing parts of it. The pressure in a squeezed portion will rise and it will accelerate toward lower pressures elsewhere. Since this sort of pressure change isn't caused by the water's motion, it's a **static variation** in pressure. The water's motion can also affect its pressure and such **dynamic variations** in pressure can be complicated and fascinating. As we'll see in the next chapter, they contribute to such diverse effects as the spray of a garden hose nozzle, the lift on an airplane's wing, and the curve of a curve ball.



Check Your Understanding #1: Under Pressure in the Garden

With the water faucet open and the nozzle at the end of your garden hose shut tightly, the hose is full of high-pressure water. Why doesn't this water accelerate?

Answer: The water pressure (force on a unit of surface area) inside the hose is uniform throughout, so the water experiences no net force and doesn't accelerate.

Why: In the absence of gravity, fluids accelerate only when they experience pressure imbalances. Since water throughout the hose is at the same pressure, there is no pressure imbalance and no acceleration. When you open the nozzle, the pressure at that end of the hose drops and the water accelerates toward it.

Creating Water Pressure with Water Pumps

To start water flowing through the plumbing in a level house or city, you need a water pump, a device that uses mechanical work to deliver pressurized water through a pipe. At its most basic level, a water pump squeezes a portion of water to raise its local pressure so that it accelerates toward regions of lower pressure. The pump continues to squeeze that water as it flows out through the plumbing.

To understand how a pump works, picture a sealed plastic soda bottle full of water. When you don't squeeze the bottle, the pressure inside it is atmospheric and uniform (remember that we're neglecting gravity). But when you squeeze the sides of the bottle and push inward on the water, the water responds by pushing outward on you—Newton's third law—and it does this by increasing its pressure. The harder you push, the harder the water pushes back and the greater its pressure becomes.

Water, like all liquids, is **incompressible**; that is, it experiences almost no change in volume as its pressure increases. Although the bottle won't get smaller as you squeeze, the water pressure inside it can increase substantially. It doesn't take much force, exerted with your thumb on a small area of the bottle, to increase the pressure in the bottle from atmospheric to twice that value or more.

The pressure increases uniformly throughout the water bottle, an observation known as **Pascal's principle**: a change in the pressure of an enclosed incompressible fluid is conveyed undiminished to every part of the fluid and to the surfaces of its container. This uniform pressure rise leads to a large upward force on the bottle's cap. If the cap were wider and had more surface area, the upward force on it might be large enough to blow it off the bottle. That effect is the basis for hydraulic systems and lifts, where pressure produced in an incompressible fluid by a small force exerted on a small area of the fluid's container results in a large force exerted on a large area of the fluid's container (Fig. 5.2.2). It also explains why plastic drinking bottles usually have small caps and why wide-mouth plastic containers are better suited for candies, cookies, and nuts.

PASCAL'S PRINCIPLE

A change in the pressure of an enclosed incompressible fluid is conveyed undiminished to every part of the fluid and to the surfaces of its container.

It's time to remove the cap from the water bottle. Now when you squeeze the bottle and increase the local water pressure, the water can move. As its pressure increases inside the bottle, the water begins to accelerate toward the lower pressure outside the bottle's open top and the result is a fountain. You are pumping water!

You are also doing work; as the water flows out of the bottle, your hands move inward. Since you are pushing inward on the water and the water is moving inward, you are doing work on the water. Pumps do work when they deliver pressurized water, and pressurized water carries with it the energy associated with that work.

Although a water bottle can act as a pump briefly, it soon runs out of water. A more practical pump appears in Fig. 5.2.3. In this pump, a piston slides back and forth in the open end of a hollow cylinder, making a watertight seal. Pushing inward on that piston squeezes any water in the cylinder and increases the local water pressure. Water begins to flow.

In contrast to our simple bottle, this pump's cylinder is easy to refill. The cylinder actually has two openings, each with a valve that permits water to flow in only one direction. Water can leave the cylinder only through the top opening and can enter only through the bottom opening. As you push the pump's piston into the water-filled cylinder, the water pressure in the cylinder rises and water accelerates and flows out through the top valve. As you pull the pump's piston out of the water-filled cylinder, the water pressure inside the

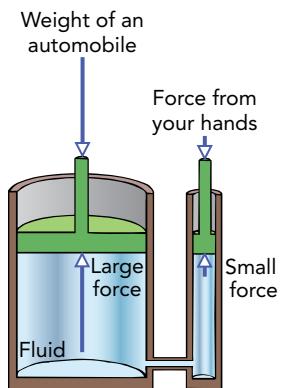


Fig. 5.2.2 The force a pressurized fluid exerts on a piston is proportional to the surface area of that piston. This fact underlies hydraulic systems, in which a small force exerted on a small piston pressurizes a trapped liquid so that it exerts a large force on a large piston. In this figure, your gentle downward push on the small piston is balanced by a gentle upward force exerted on that small piston by the pressurized fluid. At the same time, the strong downward push of an automobile on the large piston is balanced by a strong upward force exerted on that large piston by the pressurized fluid. If the small piston descends quickly at constant velocity, the large piston will rise slowly at constant velocity and you'll be doing the work of lifting the heavy car a short distance by pushing the small piston down gently for a large distance. The hydraulic system is providing you with mechanical advantage and allowing you to raise something that's too heavy for you to lift unaided.

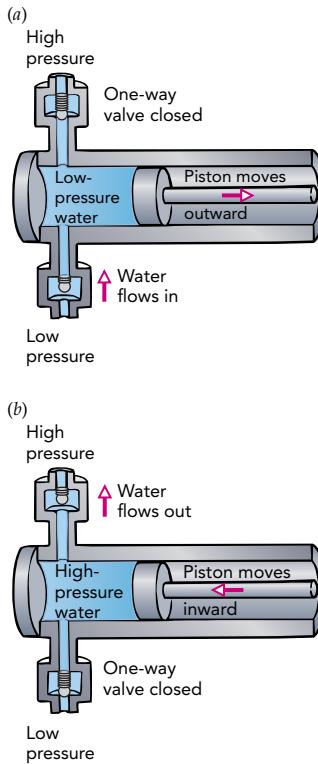


Fig. 5.2.3 Water is pumped from a region of low pressure to a region of high pressure by a reciprocating piston pump. (a) As the piston is drawn outward, water flows into the cylinder from the low-pressure region. (b) As the piston is pushed inward, the inlet one-way valve closes and water is driven out of the cylinder and into the high-pressure region.

cylinder drops and water accelerates and flows in through the bottom valve. In fact, as you withdraw the piston, the pressure inside the cylinder drops below atmospheric pressure, so even water in an open reservoir nearby will accelerate toward the partial vacuum in the cylinder and refill it.

Check Your Understanding #2: Working a Water Pump

Which normally requires more work: pulling the piston of a water pump out of the cylinder or pushing it back in?

Answer: Pushing it back in usually requires more work.

Why: When you pull the piston out of the cylinder, you are moving air out of the way and creating a partial vacuum inside the cylinder. This action requires a modest amount of work because the air doesn't push terribly hard on the back of the piston and pressure from the water flowing into the cylinder assists you. But as you push the piston back into the cylinder, you are pressurizing the water. Depending on the pressure of water in the outlet hose, the water in the cylinder may exert a very large force on the piston. In that case, you must do a great deal of work on the water as you push the piston inward and drive the water out of the cylinder.

Moving Water: Pressure and Energy

The pump of Fig. 5.2.3 can draw low-pressure water from a pond and fill a hose with high-pressure water. If the nozzle at the other end of the hose is open, the water will accelerate toward lower pressure outside the nozzle and spray toward the garden with considerable kinetic energy. From where does this kinetic energy come?

The energy comes from you and the pump. As you push the piston into the cylinder, pressurizing the water and squeezing it out through the top valve, you're doing work on the water—you're exerting an inward force on the water, and the water is moving a distance in the direction of your force. The inward force you exert on the water is equal to the water pressure times the surface area of the piston, or

$$\text{force} = \text{pressure} \cdot \text{surface area},$$

and the distance the water moves is equal to the volume of water pumped divided by the surface area of the piston, or

$$\text{distance} = \frac{\text{volume}}{\text{surface area}}.$$

Since work equals *force* times *distance*, the work you do pumping water is equal to the water pressure times the volume of water you pump, or

$$\text{work} = \text{pressure} \cdot \text{volume}.$$

As you pump high-pressure water through the hose, high-speed water sprays out of its nozzle. Since water is incompressible, water enters and exits the hose at equal rates and the work you do pumping 1 liter of water into the hose becomes kinetic energy in 1 liter of water leaving the nozzle. That energy travels directly from the pump to the nozzle, and the water doesn't actually store any of it along the way. Nonetheless, we can imagine that the water's kinetic energy after it reaches the nozzle was stored in that water's pressure before the nozzle. We create a useful fiction—**pressure potential energy**. Water that's under pressure has a pressure potential energy equal to the water's volume times its pressure.

Because pressure potential energy actually comes from the pump, it vanishes as soon as you break the link between the pressurized water and the pump that pressurizes it; you can't save a bottle of high-pressure water and expect it to retain this potential energy. The

concept of pressure potential energy is meaningful only if the water is flowing steadily, so that water leaving the plumbing is replaced exactly by water from the pump or something equivalent. The ideal situation in which to use the concept of pressure potential energy is also the simplest type of water flow—flow in which nothing ever changes. In keeping with its constancy, this situation is known as *steady-state flow*.

In **steady-state flow**, a fluid flows continuously and steadily through a stationary environment. When you watch steady-state flow, you can't detect the passage of time in either the fluid or its environment, and a video of steady-state flow looks exactly like a still photograph. With small allowances for the inevitable imperfections of real life, water traveling steadily through plumbing in your home, a soft wind blowing smoothly across your motionless cheek, and a gentle current flowing silently between the motionless banks of a lazy river are all cases of steady-state flow in fluids.

Although pumps and other devices with moving components aren't allowed in steady-state flow, they can be connected to it so long as they operate steadily. For example, if you pump water as steadily as you can through a motionless hose and nozzle, the water will exhibit steady-state flow from where it enters the hose to well after it sprays out of the nozzle.

The beauty of steady-state flow is in how it handles energy along streamlines. A **streamline** is the path taken by a tiny portion of water in the flow, a portion that I refer to here as a “drop.” In steady-state flow, an endless series of identical drops travels that same unchanging streamline, like the cars of an infinite train on a stationary track. In steady-state flow with no friction-like effects, a drop's ordered energy remains constant as it travels along its streamline. Moreover, since all the drops along that streamline are identical, the ordered energy per drop is constant everywhere along that streamline. As we'll soon see, those observations are surprisingly useful.

A drop's ordered energy can take three different forms: kinetic energy, pressure potential energy, and gravitational potential energy. Since we're still considering plumbing in a level region, we'll omit gravitational potential energy for now. We've already seen that the drop's pressure potential energy is equal to its pressure times its volume. The drop's kinetic energy is given by Eq. 2.2.2 as one-half its mass times the square of its speed. Since a drop's mass is its density times its volume, the drop's ordered energy is

$$\begin{aligned} \text{ordered energy} &= \text{pressure potential energy} + \text{kinetic energy} \\ &= \text{pressure} \cdot \text{volume} + \frac{1}{2} \cdot \text{density} \cdot \text{volume} \cdot \text{speed}^2 \\ &= \text{constant (along a streamline)}. \end{aligned} \quad (5.2.1)$$

If we divide both sides of this expression by the drop's volume, we can obtain another useful form of this relationship:

$$\begin{aligned} \frac{\text{ordered energy}}{\text{volume}} &= \frac{\text{pressure potential energy}}{\text{volume}} + \frac{\text{kinetic energy}}{\text{volume}} \\ &= \text{pressure} + \frac{1}{2} \cdot \text{density} \cdot \text{speed}^2 \\ &= \text{constant (along a streamline)}. \end{aligned} \quad (5.2.2)$$

Equation 5.2.2 is called **Bernoulli's equation**, after Swiss mathematician Daniel Bernoulli ³, whose work led to its development, although Swiss mathematician Leonhard Euler (1707–1783) actually completed it. It applies only to incompressible fluids such as water and assumes no ordered energy is wasted by friction-like effects.

According to Eq. 5.2.2, water that's in steady-state flow can exchange pressure for speed or speed for pressure as it flows along a streamline. As water accelerates out of a nozzle, for example, its pressure drops but its speed increases because it's converting pressure potential energy into kinetic energy. And as that moving water sprays against the car

³ As a professor in Basel, Daniel Bernoulli (Swiss mathematician, 1700–1782) taught not only physics but also botany, anatomy, and physiology. He correctly proposed that the pressure a gas exerts on the walls of its container results from the countless impacts of tiny particles that make up the gas. He also derived an important relationship among the pressure, motion, and height of a fluid—Bernoulli's equation.

you're washing, it slows down but its pressure increases because it's converting kinetic energy back into pressure potential energy. In both cases, the water's ordered energy is constant.

Check Your Understanding #3: How Does Your Garden Grow?

Water in your garden hose has considerable pressure and arcs several meters through the air as you water your plants. What is the water pressure in the falling water once it leaves the nozzle at the end of the hose?

Answer: It is atmospheric pressure.

Why: As the water accelerates out of the nozzle, its pressure drops. It is converting pressure potential energy into kinetic energy. The water pressure drops until it reaches the pressure of the surrounding air, which is atmospheric pressure.

Gravity and Water Pressure

Gravity creates a downward pressure gradient in water; the deeper the water, the more weight there is overhead and the greater the pressure. Since water is much denser than air, water pressure increases rapidly with depth. In a vertical pipe that's open on top, the water's surface is at atmospheric pressure (about 100,000 Pa). Only 10 m (33 feet) below the water's surface, however, the pressure has already doubled to 200,000 Pa. At that modest depth, the water overhead weighs as much as the air overhead, even though the atmosphere is many kilometers thick.

The shape of the pipe doesn't affect the relationship between pressure and depth. No matter how complicated the plumbing, the pressure of stationary water inside it increases with depth by 10,000 Pa per meter, or 10,000 Pa/m (Fig. 5.2.4). This uniform downward pressure gradient creates an upward buoyant force on anything immersed in the water. In fact, that buoyant force is what supports the water itself (Fig. 5.2.5).

The dependence of water pressure on depth has a number of important implications for water distribution. First, water pressure at the bottom of a tall pipe is substantially higher than at the top of that same pipe. Consequently, if only a single pipe supplies water to a skyscraper, then the water pressure on the ground floor will be dangerously high while the pressure in the penthouse will be barely enough for a decent shower. Tall buildings must therefore handle water pressure very carefully; they can't simply supply water to every floor directly from the same pipe.

Second, pressure in a city water main does more than just accelerate water out of a showerhead; it also supports water in the pipes of multistory buildings. Lifting water to the third floor against the downward force of gravity requires a large upward force, and that force is provided by water pressure. The higher you want to lift the water, the more water pressure you need at the bottom of the plumbing. Lifting the water also requires energy, which is often provided by a water pump.

Third, as water travels up and down the streets of a hilly city, its pressure varies with height. In the valleys the pressure can be very large, and at the tops of hills the pressure can be very small. Water mains in valleys must therefore be particularly strong to keep from bursting. The large pressure in a valley is quite useful because it helps push the water back uphill on the other side of the valley (Fig. 5.2.6). Nonetheless, a hilly city must have pumping stations and other water-pressure-control systems located throughout to provide reasonable water pressures to all the buildings, regardless of their altitudes.

Apart from those pressure-control systems, even a hilly city's plumbing often involves steady-state flow and can be explained using a version of Bernoulli's equation that includes gravity. Before we explore that topic, though, let's take a look at the non-steady-state flow that occurs when water in an isolated section of plumbing has free surfaces that can move up or down so that you see things change with time.

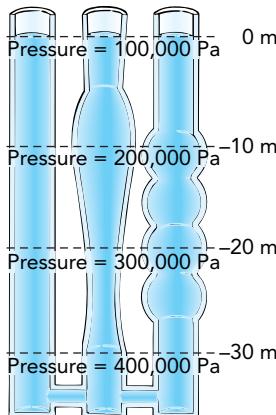


Fig. 5.2.4 The pressure of stationary water in pipes increases with depth by about 10,000 Pa per meter of depth. The shape of the pipes doesn't matter. For plumbing that's open on top and connected near the bottom, as shown here, water will tend to flow until its height is uniform throughout the plumbing.

When all of those free surfaces are at atmospheric pressure, there are no pressure imbalances and the water accelerates only in response to gravity. The water's only potential energy is gravitational and, like any object, it accelerates in the direction that lowers that gravitational potential as quickly as possible. If the water's free surface is higher in one place than another, it can reduce its average height and, therefore, its gravitational potential energy by letting its highest water fill in its lowest valley. After some sloshing about, the water settles down in stable equilibrium with all its free surfaces smooth and level at a single uniform height. No matter how complicated the plumbing, water that is open to the air on top always "seeks its level." The natural flow associated with this leveling effect is often used in water delivery (see ④).

Remarkably, the leveling effect works even when some of the water-filled plumbing is located above the water's free surfaces. For example, water can seek its level between two open containers through an elevated pipe known as a siphon (Fig. 5.2.7). In such cases, the pressure of water in those elevated segments of plumbing is less than atmospheric pressure.

If you seal off part of the isolated plumbing and reduce the pressure above one of the water's free surfaces, that surface will rise higher than all of the others. It will rise until the added pressure produced by the taller column of water replaces the missing pressure above the water's free surface. The less pressure there is above that surface, the higher the water must rise to make up for the missing pressure. This effect lifts water in a drinking straw and allows it to travel between two open containers in a siphon.

However, removing all the air pressure above the water's free surface inside a long straw or siphon will raise its height only about 10 m (33 feet) above the level of the water elsewhere in an open container. Even with no pressure above it, this 10-m column of elevated water completely replaces the absent air pressure and thus prevents water in the rest of the plumbing from lifting it any higher. It's therefore impossible to draw water from a deep well simply by lowering a pipe into that well and reducing the air pressure in the pipe; the water will rise no more than 10 m upward. Instead, a pump must be attached to the bottom of the pipe to pressurize the water and push it all the way to the top of the pipe.

Check Your Understanding #4: What's the Water Pressure?

If all the water to a 400-m-tall skyscraper were delivered from a single pipe, how much higher would the water pressure be on the ground floor than on the top floor?

Answer: It would be about 4,000,000 Pa (40 atmospheres) higher.

Why: The weight of water inside the pipe would create an enormous excess pressure near the bottom of the building. Water spraying from an open faucet on the first floor at this enormous pressure could accelerate to 319 km/h (200 mph), as it does in some high-pressure jet washing and cutting machines.

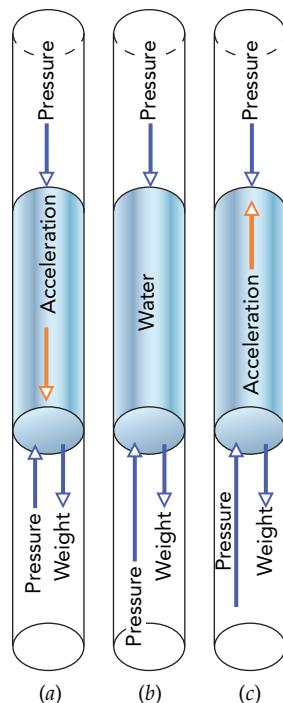


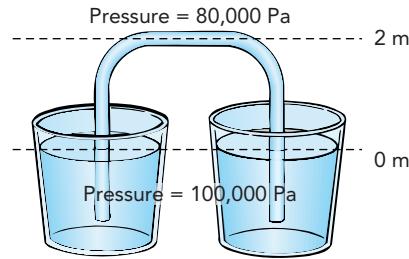
Fig. 5.2.5 When a pipe is oriented vertically, gravity affects the motion of water in the pipe. (a) If the water's pressure doesn't change with depth, the water will accelerate downward (fall) because of its weight. (b) If the water's pressure increases with depth by 10,000 Pa/m, the water won't accelerate. (c) If the water's pressure increases with depth by more than that amount, the water will accelerate upward.



Fig. 5.2.6 Los Angeles receives much of its water from Owens Valley, 300 km to the north. The water negotiates the mountains and valleys in between, driven by gravity alone. Giant pipes allow pressure to build during downhill stretches to push the water back uphill later on. Parts of the 1913 aqueduct support so much pressure that the steel pipe used in them has to be more than an inch thick.

④ The Romans used gravity to convey water to Rome from sources up to 90 km away. A very gradual slope in the aqueducts kept the water moving in spite of frictional effects that opposed the water's progress. Poisoning from the lead pipes used in some of the aqueducts is blamed in part for the decay of the Roman Empire.

Fig. 5.2.7 The two open containers of water are connected by a siphon. This U-shaped tube permits water to flow until its level is equal in both containers. The sturdy tube permits the water pressure in the siphon to drop below atmospheric pressure.



Moving Water Again: Gravity

As we've seen, it takes pressure and energy to lift water to the third floor of a building. We can now expand our description of fluids undergoing steady-state flow to include gravity and gravitational potential energy.

A drop's gravitational potential energy is given by Eq. 1.3.2 as its mass times the acceleration due to gravity times its height. The water's gravitational potential energy per volume is thus its density times the acceleration due to gravity times its height.

If we include gravitational potential energy in Eq. 5.2.2 and recognize that, for fluids in steady-state flow along a streamline, the energy per volume is constant, we obtain a relationship that can be written as a word equation:

$$\begin{aligned}
 \frac{\text{ordered energy}}{\text{volume}} &= \frac{\text{pressure potential energy}}{\text{volume}} + \frac{\text{kinetic energy}}{\text{volume}} \\
 &\quad + \frac{\text{gravitational potential energy}}{\text{volume}} \\
 &= \text{pressure} + \frac{1}{2} \cdot \text{density} \cdot \text{speed}^2 \\
 &\quad + \text{density} \cdot \text{acceleration due to gravity} \cdot \text{height} \\
 &= \text{constant (along a streamline)}, \tag{5.2.3}
 \end{aligned}$$

in symbols:

$$p + \frac{1}{2} \cdot \rho \cdot v^2 + \rho \cdot g \cdot h = \text{constant (along a streamline)},$$

and in everyday language:

When a stream of water speeds up in a nozzle or flows uphill in a pipe, its pressure drops.

This is a revised version of Bernoulli's equation, one that includes gravity. It correctly describes steady-state flow in streamlines that change height. As before, it applies only to incompressible fluids such as water and assumes no ordered energy is wasted by friction-like effects.

BERNOULLI'S EQUATION

For an incompressible, frictionless fluid in steady-state flow, the sum of its pressure potential energy, its kinetic energy, and its gravitational potential energy is constant along a streamline. Equation 5.2.3 expresses this law as a formula.



© Steve Lewis/The Image Bank/Getty Images

Fig. 5.2.8 Many buildings in New York City have water towers on their roofs. These towers maintain water pressure in the plumbing and help in firefighting.

Because energy is conserved, water that's in steady-state flow can exchange the energies associated with its speed, pressure, and height for one another. Thus as water flows downward, its speed or pressure or both increase; if it falls from an open faucet, its speed increases; and if it descends steadily inside a uniform pipe, its pressure increases. The reverse happens as water flows upward. Water rising from a fountain loses speed as it ascends, while water rising steadily in a uniform pipe loses pressure on its way up.

This interchangeability of height, pressure, and speed makes it possible to pressurize plumbing by connecting a tall column of water to the pipes. That's why cities, communities, and even individual buildings have water towers (Fig. 5.2.8). A water tower is built at a relatively high site within the region it serves. A pump fills the water tower with water, and then gravity maintains a constant high pressure throughout the plumbing that connects to it (Fig. 5.2.9). The water is at atmospheric pressure at the top of the water tower, but the pressure is much higher at the bottom; at the base of a 50-m-high water tower, for example, the pressure is about 600,000 Pa, or six times atmospheric pressure.

In addition to providing a fairly steady pressure in the water mains, a water tower stores energy efficiently and can deliver that energy quickly. When water is drawn out of the water tower, its gravitational potential energy at the top becomes pressure potential energy at the bottom. The water tower replaces a pump, supplying a steady flow of water at an almost constant high pressure. Unlike a pump, however, the water tower can supply this high-pressure water at an enormous rate. As long as the water level doesn't drop too far, high-pressure water keeps flowing.

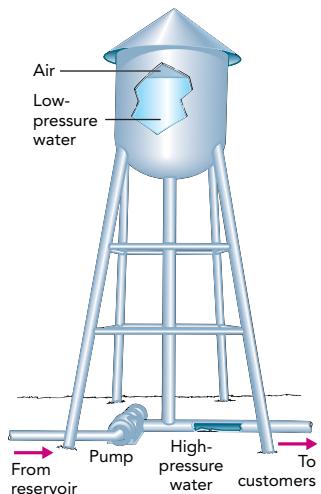


Fig. 5.2.9 A water tower uses the weight of water to create a large water pressure near the ground. The higher the tower, the greater the pressure near the ground. The water tower is able to maintain the water pressure passively and doesn't require constant pumping. Even during periods of peak water consumption, it maintains a fairly steady pressure. When the water level in the water tower drops below a certain set point, a pump refills the tower.

Check Your Understanding #5: Water Power

A hydroelectric power plant extracts energy from water that has descended from an elevated reservoir in a pipe. In the reservoir, this energy takes the form of gravitational potential energy. Just before the power plant, what form does the energy take?

Answer: It takes the form of pressure potential energy (and some kinetic energy).

Why: As the water descends inside the pipe, its gravitational potential energy is converted into pressure potential energy. The water reaching the power plant is under enormous pressure, and it's this pressure that exerts the forces needed to turn the turbines that run the generators. Work is required to turn the turbines, so the water gives up much of its energy in the power plant. This energy leaves the power plant via the electric power lines.

Check Your Figures #1: Up or Out

If the water pressure at the entrance to a building is 1,000,000 Pa, how high can the water rise inside the building and how fast will it flow out of a faucet right at the entrance? (A liter of water inside a pipe has a mass of 1 kg.)

Answer: It can rise about 100 m or emerge from the faucet at about 45 m/s (101 mph).

Why: As the water flows through the pipe (a streamline), its pressure potential energy can become gravitational potential energy or kinetic energy. At the start, the water's energy is all pressure potential energy so, from Eq. 5.2.3, the water's energy per volume is 1,000,000 Pa. If the water flows up the pipe, that energy will become gravitational potential energy. We can rearrange Eq. 5.2.3 to find the height it can reach:

$$\begin{aligned}\text{height} &= \frac{\text{energy}}{\text{volume}} \cdot \frac{1}{\text{density} \cdot \text{acceleration due to gravity}} \\ &= 1,000,000 \text{ Pa} \cdot \frac{1}{1000 \text{ kg/m}^3 \cdot 9.8 \text{ m/s}^2} = 102 \text{ m.}\end{aligned}$$

If the water flows out of the faucet, that energy will become kinetic energy. We can also rearrange Eq. 5.2.3 to find the speed it will attain:

$$\begin{aligned}\text{speed} &= \sqrt{\frac{\text{energy}}{\text{volume}} \cdot \frac{2}{\text{density}}} \\ &= \sqrt{\frac{2,000,000 \text{ Pa}}{1000 \text{ kg/m}^3}} = 45 \text{ m/s.}\end{aligned}$$

Epilogue for Chapter 5

In this chapter we have investigated some of the basic concepts associated with fluids. In Balloons, we explored the concept of pressure and the way in which air pressure structures and supports Earth's atmosphere. We saw that increased air pressure beneath an object produces an upward buoyant force on that object and that this buoyant force can float objects, such as hot-air and helium balloons, that are less dense than the surrounding air.

In Water Distribution, we examined how water pressure causes water to accelerate through plumbing, from high pressure to low pressure. We then focused on ways to produce water pressure, either using pumps or using gravity. By studying the forms of energy in water, we were led to Bernoulli's equation, which describes the interconversion of a fluid's ordered energy among pressure potential energy, kinetic energy, and gravitational potential energy. Although the most dramatic applications of Bernoulli's equation are ahead of us, we've already used it to understand the changes in pressure and speed that accompany water's movement up and down in pipes and in fountains.

Explanation: A Cartesian Diver

The diver floats because its average density—that is, the mass of the vial and its contents divided by the volume of space those two components occupy—is less than the density of water. Since the upward buoyant force on the diver exceeds its downward weight, the diver floats upward toward the top of the bottle. When the diver begins to stick out of the water, it displaces less water and more air and the buoyant force it experiences decreases. Eventually it experiences zero net force and floats without accelerating at the water's surface.

When you squeeze the soda bottle, you increase the pressure inside it. Because water is incompressible, its density doesn't change as the pressure goes up. However, the air

bubble inside the vial is compressed and takes up less space inside the vial. Water flows into the vial and increases the average density of the vial and its contents. When the average density of the diver finally exceeds the density of water, the diver sinks.

To keep the diver hovering in the water, you must adjust the water pressure until the diver's average density is exactly that of water. This adjustment is impossible to make without looking at the diver. Even the slightest overpressure or underpressure will cause the diver to drift slowly down or up.

Chapter Summary and Important Laws and Equations

How Balloons Work: A hot-air balloon floats because its total weight (basket, envelope, and hot air) is less than that of the cooler air it displaces. By heating the air in the envelope with a flame, the pilot reduces its density. As the air warms up, fewer particles are needed to fill the envelope, the extra particles flow out of the envelope through the opening at its bottom, and the balloon becomes lighter.

A helium balloon also weighs less than the air it displaces. However, its lower weight is due to the lightness of the individual helium atoms, each of which weighs much less than the air particle it replaces. By filling a balloon with helium, the balloon's weight is dramatically reduced. Since the buoyant force exerted on the balloon by the surrounding air exceeds the balloon's weight, the balloon accelerates upward.

How Water Distribution Works: Water distribution begins when a pump transfers low-pressure water from a reservoir to high-pressure plumbing. Along a level path, water accelerates toward regions of lower pressure, such as open hoses or showerheads. Pressure imbalances allow the water to negotiate bends in the pipes en route to its destination. During its travels, the water may rise or fall in height; as it does, its pressure changes, decreasing as the water moves upward and increasing as the water moves downward. In low-lying regions, where the water pressure may be too high to use directly, a pressure regulator may have to be added to the plumbing. In high-lying regions, where the water pressure may be too low to be practical, an additional pump may have to be employed to boost its pressure.

1. Archimedes' principle: An object partially or wholly immersed in a fluid is acted on by an upward buoyant force equal to the weight of the fluid it displaces.

2. Ideal gas law: The pressure of a gas is equal to the Boltzmann constant times the particle density times the absolute temperature, or

$$\text{pressure} = \text{Boltzmann constant} \cdot \text{particle density} \cdot \text{absolute temperature} \quad (5.1.3)$$

3. Pascal's principle: A change in the pressure of an enclosed incompressible fluid is conveyed undiminished to every part of the fluid and to the surfaces of its container.

4. Bernoulli's equation: For an incompressible fluid in steady-state flow, the sum of its pressure potential energy, its kinetic

energy, and its gravitational potential energy is constant along a streamline, or

$$\begin{aligned} \text{ordered energy} &= \frac{\text{pressure potential energy}}{\text{volume}} + \frac{\text{kinetic energy}}{\text{volume}} \\ &\quad + \frac{\text{gravitational potential energy}}{\text{volume}} \\ &= \text{pressure} + \frac{1}{2} \cdot \text{density} \cdot \text{speed}^2 \\ &\quad + \text{density} \cdot \text{acceleration due to gravity} \\ &\quad \cdot \text{height} \\ &= \text{constant (along a streamline).} \end{aligned} \quad (5.2.3)$$

Fluids and Motion

Fluids are fascinating when they move. Stationary water and air may be essential to life, but they're also fairly simple; only their pressures vary from place to place, and even these are determined primarily by gravity. Rushing rivers or gusts of wind, however, with their wonderful variety of simple and complicated behaviors, are much more exciting. The motions of fluids aren't just interesting; they're also important, since our world is filled with objects and machines that work in whole or part because of the behaviors of moving fluids. In this chapter, we look at a variety of situations in which fluid motion contributes to the way things work.

ACTIVE LEARNING EXPERIMENTS

A Vortex Cannon

Fluids are real; they exist independently of any solids that might move through them. That notion is easy to accept in reference to water, since we can see that water doesn't wait for a boat to pass before moving in interesting ways. Air is harder to visualize in this fashion because we seldom see it by itself, apart from its effects on buildings or airplanes or our skin.

To begin seeing air as something tangible and to demonstrate the rich possibilities of motion available to

it, build a vortex cannon, a device that sends rings of air sailing across a room. Although a serious vortex cannon is best constructed from a large drum or crate, you can make a simple one from an empty cardboard cereal box.

Seal the rectangular box on all edges with tape, and cut a circular hole 5 cm. (2 in.) in diameter in the center of one face. To use your cannon, just tap hard on the other face of the box. Rings of air will leap out of the hole and sail across the room at about 5 m/s (11 mph).



Courtesy Lou Bloomfield

Although you can't see these rings, you can follow their motion by looking for their effects on objects they encounter. The rings can easily blow out candles or rustle light window drapes across a small room. If you have a friend blow some rings at you, you'll feel exactly where they hit on your face or shirt.

To see these rings directly, fill the cereal box with smoke, mist, or dust. What do you think the rings will look like when they emerge from the hole? Will the smoke in each ring be stationary, or will it be swirling about the ring in some manner? How will the ring's size

and speed of travel depend on the size of the hole from which it emerges? Will it depend on how hard you tap the cereal box? Will the ring's size and speed change in flight?

Now tap the box and watch the smoke rings. What motions do you see? Change the hole size and see how this change affects the rings and their motion. As you can see, the air can execute some very complicated movements all by itself. In this chapter, we will examine how moving air, water, and other fluids affect our everyday lives.

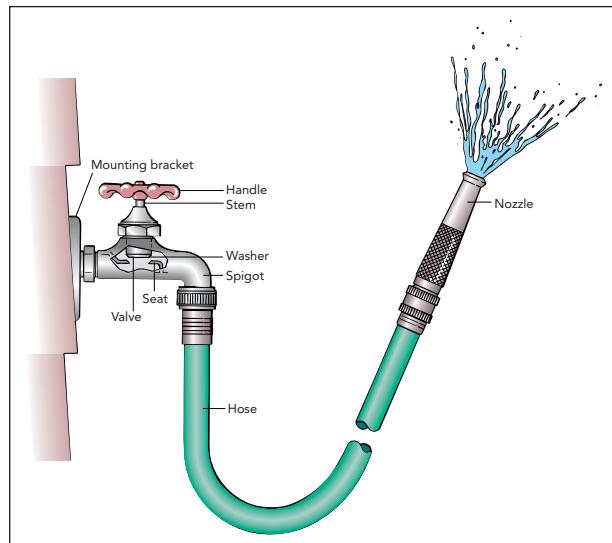
Chapter Itinerary

In particular, we explore (1) *garden watering*, (2) *balls sports: air*, and (3) *airplanes*. In Garden Watering, we look at how water's pressure and speed vary as it flows through a faucet, a hose, and a nozzle. In Ball Sports: Air, we investigate the effects of air on the motions of balls. In Airplanes, we study the ways in which moving air supports and propels airplanes in flight. A more complete preview can be found in the Chapter Summary and Important Laws and Equations at the end of the chapter.

This chapter continues to develop the concept of energy conservation along a streamline that was introduced in Chapter 5. It also brings up several new types of forces that are present when fluids move past one another or past solid objects. These ideas are present in a vast array of commonplace activities, from washing windows with a hose to pumping water with a windmill.

SECTION 6.1

Garden Watering



Tending a flower garden often involves watering. While this once meant walking the garden's paths with a watering can, modern plumbing has made such effort unnecessary. With a hose and nozzle attached to a faucet, you can do your job without leaving your lawn chair. Although the tools involved—

faucets, hoses, and nozzles—are simple and unsophisticated, the principles behind them are not. All three make elegant use of the laws of fluid flow and will introduce us to several important complications, notably friction-like effects and turbulence, that we had previously chosen to ignore.

Questions to Think About: What determines the rate at which water flows through a hose? If honey ran through your hose instead of water, how would that affect the flow? Why does water make noise as you sprinkle your garden? Why does a nozzle make the water spray so fast and so far? Why do the pipes sometimes clank when you abruptly close the faucet?

Experiments to Do: Open a faucet gradually and watch the water begin to flow. What's pushing the water out of the faucet? What happens to the flow rate and speed of the water as you open the faucet further? Look below or behind the faucet and try to determine how the water enters and exits the faucet. When do you hear the water flowing?

Attach a hose to the faucet. Does the water flow as quickly from the open end of the hose as it did from the faucet alone? Cover most of the hose end with your thumb and watch the water spray out into the air. Why does the

water travel so much farther when you almost stop its flow? Do you feel the water pressing against your thumb? Attach a nozzle to the hose and see how the flow rate affects the

strength of the spray. Fill a bucket with the open hose and then again while using the nozzle. Which method fills the bucket faster?

Water's Viscosity

1 Your car's engine is protected by motor oil with a carefully chosen viscosity. If that oil were too thin, it would flow out from between surfaces and wouldn't keep them from rubbing against one another. If that oil were too thick, the engine would waste power moving its parts through the oil. Years ago, you had to change your motor oil for the season. Thick 40-weight motor oil was used in summer because hot weather made it thinner; thin 10-weight oil was used in winter because cold weather made it thicker. Now, a modern multigrade oil maintains a nearly constant viscosity over a wide range of temperatures and need not be changed with the seasons. This oil contains tiny molecular chains that ball up when cold but straighten out when hot. These chains thicken hot oil so that 10W-40 oil resembles 10-weight oil in winter and 40-weight oil in summer.

Having brought water to your home in the previous chapter, we're already well on our way to watering your garden. However, to reach the garden, water must first travel through a long stretch of hose that's lying straight on level ground. Does the length of this hose have any effect on the water delivery process?

The answer is yes, longer hoses generally deliver less water. However, according to what we learned in Chapter 5, water should coast through a straight, level hose at a constant velocity and constant pressure, and the length of that hose shouldn't matter. We evidently overlooked something important—friction. Moving water doesn't slide freely through a stationary hose. In reality, it experiences frictional forces that oppose its motion relative to the hose.

This friction is unusual, though, because most of the water in the hose never actually touches the hose itself. If water deep inside the hose is going to experience any forces due to relative motion, then those forces are going to have to occur *within* the water. Water must exert frictional forces on itself!

Sure enough, water does experience internal frictional forces. They're called **viscous forces**, forces that appear whenever one layer of a fluid tries to slide across another layer of that fluid. Viscous forces oppose relative motion, and you can observe their effects easily when you pour honey out of a jar. The honey at the jar's surface is stuck there by chemical forces and remains stationary. But even honey that's far from the walls can't move easily; it experiences viscous forces as it tries to move relative to nearby honey. Since honey is a "thick," or *viscous*, fluid, viscous forces act quite effectively to keep all the honey moving with nearly the same velocity. Since the honey at the walls can't move, viscous forces tend to prevent any of the honey from moving.

Water isn't as thick as honey (Table 6.1.1), so it's less resistant to relative motion. The measure of this resistance to relative motion within a fluid is called **viscosity**, and water's viscosity is less than that of honey. In fact, if you heat the water up, it will become even less viscous and thus flow more easily. Typical of most liquids, this decrease in viscosity with increasing temperature reflects the molecular origins of viscous forces: the molecules in a liquid stick to one another, forming weak chemical bonds that require energy to break. In a hot liquid, the molecules have more thermal energy, so they break these bonds more easily to move past one another (see **1**).

TABLE 6.1.1 Approximate Viscosities of a Variety of Fluids

Fluid	Viscosity [†]
Helium (2 K)	0 Pa · s
Air (20 °C)	0.0000183 Pa · s
Water (20 °C)	0.00100 Pa · s
Olive oil (20 °C)	0.084 Pa · s
Shampoo (20 °C)	100 Pa · s
Honey (20 °C)	1000 Pa · s
Glass (540 °C)	1012 Pa · s

[†]The pascal-second (abbreviated Pa · s and synonymous with kg/m · s) is the SI unit of viscosity. Only the superfluid portion of ultracold liquid helium exhibits zero viscosity.

Check Your Understanding #1: Keeping Warm on a Windy Day

A loosely woven wool sweater has many tiny air passages between the wool fibers, yet it dramatically reduces the rate at which air flows through to your skin when you stand in a breeze. Why doesn't air flow easily through the gaps between the fibers?

Answer: The air at the surfaces of the fibers is stationary, and the air's viscosity slows the motion of air in the vicinity of the fibers.

Why: Although air trying to pass through the gaps may not directly touch the fibers, viscous forces tend to keep all the air moving together at the same velocity. As soon as some air is held up by a fiber, the air nearby is slowed by viscous forces. The fibers of a sweater are close enough together that viscous forces slow all the air trying to pass through the sweater. Imagine trying to pour honey through a sweater.

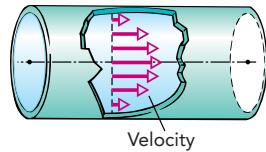


Fig. 6.1.1 The speed of water flowing through a pipe is not constant across the pipe. The water near the walls is stationary, while the water at the center of the pipe moves the fastest. The differences in velocity are the result of viscous forces.

Flow in a Straight Hose: The Effect of Viscosity

Viscosity slows the flow of water through your hose. Chemical forces between the hose and the outermost layer of water hold that layer of water stationary, and this motionless layer exerts viscous forces on the layer of moving water inside it. As this second layer slows, it exerts viscous forces on yet another layer. Layer by layer, viscous forces hold back the moving water until even water at the center of the hose feels viscosity's slowing effects (Fig. 6.1.1). Although water at the center of the hose moves faster than water in any other layer, it's still affected by the stationary hose.

These viscous forces impede water delivery. Instead of coasting effortlessly through the straight, level hose, real water needs a pressure gradient to push it steadily forward. Like the file cabinet sliding on the sidewalk in Section 2.2, water must be pushed through the hose if it's to maintain a continuous flow. Also like that file cabinet, the water becomes hotter as the work done pushing it forward is wasted and becomes thermal energy.

However, unlike the forces of ordinary sliding friction—which don't depend on relative velocities—viscous forces become larger as the relative velocities within a fluid increase. That's because as two layers of water slide past one another faster, their molecules collide harder and more frequently. Since it experiences stronger viscous forces, fast-moving water wastes more energy per meter and needs a larger pressure gradient to keep it moving steadily through a hose than does slow-moving water.

Because of viscous forces, the amount of water flowing steadily through a hose has four properties:

1. It's inversely proportional to the water's viscosity. The more viscous the water, the more difficulty it has flowing through the hose.
2. It's inversely proportional to the length of the hose. The longer the hose, the more opportunity viscous forces have to slow the water down.
3. It's proportional to the pressure difference between the hose's inlet and its outlet. This pressure difference determines the water's pressure gradient and thus how hard the water is pushed forward through the hose.
4. It's proportional to the fourth power of the diameter of the hose. Tripling the hose's diameter provides the water with nine times as much room and also allows water near the hose's center to move nine times faster.

We can turn these four proportional relationships into an equation by adding the correct numerical constant ($\pi/128$). The final relationship is called **Poiseuille's law** and can be written as a word equation:

$$\text{volume} = \frac{\pi \cdot \text{pressure difference} \cdot \text{pipe diameter}^4}{128 \cdot \text{pipe length} \cdot \text{fluid viscosity}}, \quad (6.1.1)$$

in symbols:

$$\frac{\Delta V}{\Delta t} = \frac{\pi \cdot \Delta p \cdot D^4}{128 \cdot L \cdot \eta},$$

and in everyday language:

It's hard to squeeze honey through a long, thin tube.

POISEUILLE'S LAW

The volume of fluid flowing through a cylindrical pipe each second is equal to $(\pi/128)$ times the pressure difference (Δp) across that pipe times the pipe's diameter to the fourth power, divided by the pipe's length times the fluid's viscosity (η).

It's hardly surprising that the flow rate depends in this manner on the pressure difference, pipe length, and viscosity; we've all observed that low water pressure or a long hose lengthens the time needed to fill a bucket with water and that viscous syrup pours slowly from a bottle. However, the dependence of the flow rate on the fourth power of pipe diameter is unexpected. Even a small change in the diameter of a hose significantly changes the amount of water that hose delivers each second. The two most common garden hoses in the United States have diameters of $\frac{5}{8}$ inch and $\frac{3}{4}$ inch, and while these hoses differ by a seemingly insignificant 20% or a factor of 1.2 in diameter, the $\frac{3}{4}$ -inch hose can carry about 1.2⁴ or two times as much water as the $\frac{5}{8}$ -inch hose (see **2** and **3**).

We can also look at viscous forces in terms of ordered energy. By opposing the flow of water through a hose, viscous forces do negative work on it and reduce its ordered energy—the energy considered in Bernoulli's equation, which doesn't include thermal energy. Just how much ordered energy the water retains depends on how fast it moves inside the hose. If you allow lots of water to leave the hose, water will move through it quickly and encounter large viscous forces. In the process, most of the water's ordered energy will be wasted as thermal energy and the water will pour gently out of the end of the hose.

However, if you partially block the hose's opening with your thumb and reduce the flow, water will travel slowly through the hose and encounter smaller viscous forces. As a result, the water will retain most of its ordered energy and will still be at high pressure when it reaches your thumb. This high-pressure water will then accelerate to enormous speed as it passes through the narrow opening and sprays out into the air.

We can now explain why water delivery systems normally use the widest pipes that are practical and affordable. In contrast to a narrow hose, wide pipes can carry large amounts of water while letting that water travel slowly, experience weak viscous forces, and waste little of its ordered energy. In such energy-efficient water delivery systems, friction is insignificant and Bernoulli's equation (Eq. 5.2.4) accurately predicts water's properties throughout its trip.

Check Your Understanding #2: Air Ducts

The long air ducts used to ventilate homes and businesses usually have very large diameters. These ducts are often visible near the ceilings of modern warehouse-style stores and restaurants as pipes roughly 0.5 m across. Why must the ducts be so large in diameter?

Answer: Air's viscosity slows its flow through ductwork. Moving large volumes of air rapidly without a large pressure difference between inlet and outlet requires a large diameter pipe.

Why: The airflow through long ductwork is dominated by viscous forces. The volume of air moved through ductwork is often enormous, and the pressure difference between the inlet and outlet is normally only a fraction of an atmosphere. To allow the air to move quickly through the ducts, their diameters must be large.

2 To deliver large amounts of water at high pressure or velocity, fire hoses must have large diameters. When filled with high-pressure water, these wide hoses become stiff and heavy, making them difficult to handle. Chemical additives that decrease water's viscosity allow firefighters to use narrower, lighter, and more flexible hoses.

3 Very large diameter pipes are required to transport crude oil across the Alaskan wilderness. The distances are long, and the fluid is viscous, although it is heated to lower its viscosity.

Check Your Figures #1: Old Plumbing

When your friend's house was new, the kitchen faucet could deliver 0.50 liter per second (0.50 L/s). Over the years, mineral deposits have built up in the pipes and reduced their effective diameters by 20%. How much water can the faucet deliver now?

Answer: It can deliver about 0.20 L/s.

Why: Water flow in pipes obeys Poiseuille's law (Eq. 6.1.1) and is proportional to the pipe diameter to the fourth power. Reducing the diameter by 20%, a factor of 0.8, while leaving the pipe length, pressures, and viscosity unchanged, will reduce the volume flow rate by a factor of 0.8^4 or 0.41. The pipes now deliver about 0.20 L/s (0.50 L/s times 0.41). It clearly doesn't take much mineral accumulation to dramatically reduce the flow through a pipe.

Flow in a Bent Hose: Dynamic Pressure Variations

Let's suppose that on reaching your garden the hose bends toward the right and the flowing water bends with it. That water is accelerating as it turns, and as we observed in Chapter 5, water accelerates horizontally only in response to unbalanced pressures. Since the hose is motionless, the unbalanced pressures inside it must be caused by the water itself; the water is experiencing dynamic pressure variations.

To understand these dynamic pressure variations, let's follow the streamlines as water traverses this bend. Although we've just introduced viscous forces, we're going to ignore them here for clarity and simplicity. Viscous forces are certainly important in the long narrow hose; however, the bend is so short that viscous forces have little effect on what happens to the water passing through it.

Neglecting viscous forces, the water's ordered energy is constant along each streamline and we can observe the interchanges of energy allowed by Bernoulli's equation (Eq. 5.2.4). However, since the hose rests on level ground, water's gravitational potential energy can't vary and the only interchanges we'll see are between pressure potential energy and kinetic energy.

Figure 6.1.2 shows the water's steady-state flow pattern near the bend. We're looking down on the hose in this calculated drawing and, as indicated by the black streamlines, water that is initially flowing straight ahead arcs rightward at the bend and eventually continues directly toward the right.

Water approaches the bend through a straight section of hose in which its streamlines are straight and parallel. Water flows at constant velocity along these streamlines—the

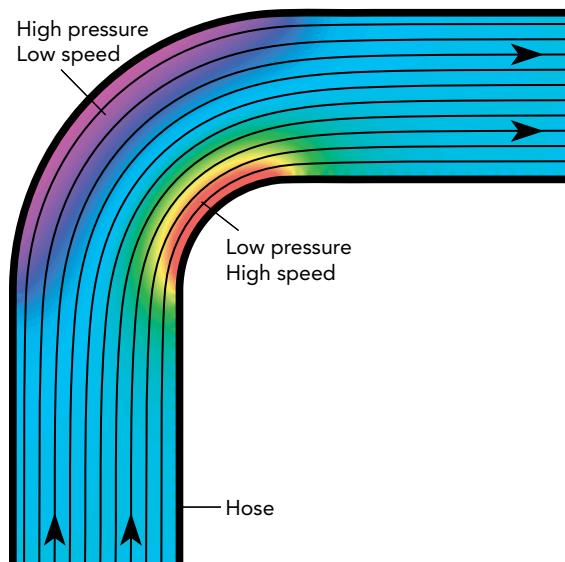


Fig. 6.1.2 Water in a bent hose experiences changes in speed and pressure. The black streamlines show the paths the water takes as it flows around the bend. The spacing between streamlines indicates flow speed (a wider space is a slower flow), and the background color indicates pressure (violet is higher pressure; red is lower pressure).

water's direction is fixed and it can't change its speed. If it tried to speed up, it would leave a gap behind it; if it tried to slow down, it would cause a "traffic jam." The water's pressure must be uniform throughout this straight section because it would accelerate if there were any differences in pressure.

The water's constant velocity and uniform pressure are represented visually in Fig. 6.1.2. Local water velocities are revealed by the directions and separations of the streamlines. Water's velocity always points along its streamline, and its speed is inversely proportional to the streamline spacing. Streamlines that become more widely spaced denote decreasing speed—water that slows down spreads out sideways and its streamlines separate from one another. Streamlines that become more narrowly spaced denote increasing speed—water that speeds up stretches out along its path and its streamlines draw toward one another. Since the streamlines leading up to the bend are straight and evenly spaced, we know that water moves along each streamline at constant velocity.

Local water pressure is displayed in Fig. 6.1.2 as the colors of the rainbow. Colors toward the violet end of the spectrum denote higher pressures, while those toward the red end of the spectrum denote lower pressures. Since the straight section has a uniform blue-green color, the water there has a uniform, moderate pressure.

Once the water starts bending toward the right, however, its velocities and pressures begin to vary. Since the water is accelerating toward the inside of the bend, there must be a pressure imbalance pushing it in that direction. Sure enough, the turning stream of water develops higher local pressure (violet) near the outside of the bend and lower local pressure near the inside of the bend (red). A similar pressure imbalance accompanies any bend in a fluid's path; the pressure is always higher on the outside of the bend than it is on the inside of that bend. After all, that pressure imbalance is what causes the fluid flow to bend!

BENDS AND PRESSURE IMBALANCES

When the path of a fluid in steady-state flow bends, the pressure on the outside of the bend is always higher than the pressure on the inside of the bend.

To keep the ordered energy constant along a streamline, each decrease in the water's local pressure is accompanied by an increase in the water's local speed, and the opposite is also true. Water arcing around the outside of the bend slows down (the streamline spacing widens) as its pressure rises, while the water arcing around the inside of the bend speeds up (the streamline spacing narrows) as its pressure drops.

As the hose straightens out beyond the bend, the water's pressures and speeds return to what they were before the bend. Water from the outside of the bend speeds up and its pressure drops, while water from the inside of the bend slows down and its pressure rises. In the straight section following the bend, water's velocity is once again constant along each streamline and its pressure is uniform.

Odd as these pressure and speed changes may seem, they are quite real and have real consequences. If your hose were clear and you could introduce thin threads of dye into the flowing water, you'd see these dyed streamlines arc around the bend just as they do in Fig. 6.1.2. If the hose were weak and couldn't tolerate excessive pressure, it would be most likely to burst on the outside of the bend, where the local water pressure is highest.

You might wonder which causes which: Does each pressure change cause a speed change, or does each speed change cause a pressure change? The answer is that they occur together and are equally entitled to be called cause and effect. Once the steady-state flow pattern has established itself, water following a particular streamline experiences rises and falls in pressure at the same time that it experiences decreases and increases in speed. The two effects—pressure changes and speed changes—simply go hand in hand.

Check Your Understanding #3: Washing Your Shirt with a Spoon

You're washing dishes when the stream of water falling from the kitchen faucet follows the curve of a spoon and sprays up onto your shirt. As the stream bent upward, where was its pressure greatest?

Answer: Pressure was greatest at the surface of the spoon.

Why: The pressure in the stream of water must be higher on the outside of the bend—at the spoon's surface—than on the inside of the bend.

Flow through a Nozzle: From Pressure to Speed

When water finally flows through the nozzle at the end of your hose, it exchanges its remaining pressure potential energy for kinetic energy and sprays out into the garden. The nozzle's narrowing channel initiates this energy transformation so that low-speed, high-pressure water entering the nozzle becomes fast-moving, atmospheric-pressure water leaving the nozzle.

Figure 6.1.3 shows that, as water passes through the nozzle, the narrowing channel herds all the streamlines together, so the water's local speed increases. The water following each streamline is speeding up to squirt through the bottleneck without causing a backup. This increase in water's local speed is accompanied by a decrease in water's local pressure, as indicated by the color shift toward the red end of the spectrum.

By the time the water leaves the hose nozzle, its pressure has dropped all the way to atmospheric pressure and it has turned all its available pressure potential energy into kinetic energy. It emerges as a narrow stream of fast-moving water and arcs gracefully through the air. No wonder you can reach the farthest parts of your garden with water when you use a nozzle.

COMMON MISCONCEPTIONS: Speed and Pressure in Fluids

Misconception: A fast-moving fluid always has a low pressure.

Resolution: The pressure of a specific portion of fluid depends on its circumstances and can take any value, high or low. However, if a fluid speeds up without descending as it flows along a streamline in steady-state flow, its pressure will decrease. In that special context, the *faster* moving fluid has a *lower* pressure.

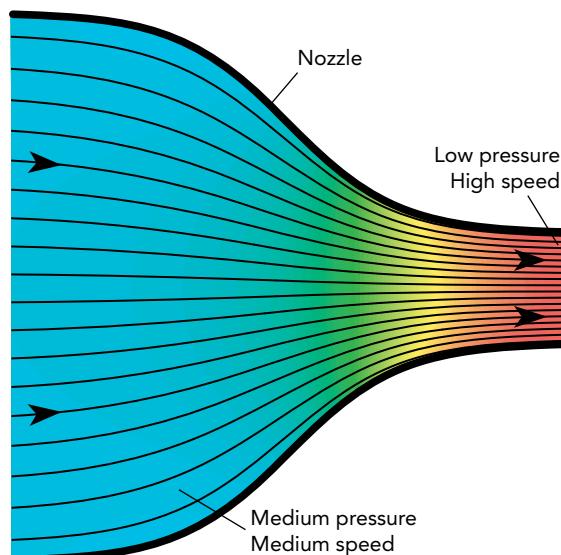


Fig. 6.1.3 Water flowing through a nozzle speeds up and its pressure drops. The narrowing spacing between streamlines indicates that the flow speed is increasing, while the color shift from violet toward red indicates that the pressure is dropping.

Check Your Understanding #4: Cleaning House

As air rushes steadily into the narrow opening of a vacuum cleaner attachment, it accelerates to high speed and its pressure drops well below atmospheric pressure. From where does the air's newfound kinetic energy come?

Answer: Even at atmospheric pressure, air has pressure potential energy. As it accelerates toward even lower pressure in the attachment, air converts some of that pressure potential energy into kinetic energy.

Why: Vacuum cleaners employ nozzles to get air moving very quickly. Fast-moving air rushing into the vacuum cleaner draws dust with it and cleans your home.

The Onset of Turbulence

As you direct the stream of water at the plants in your garden, you notice two interesting phenomena: first, the stream pushes on any surface that slows it down, and second, it tends to break up into fragments as it flows around obstacles. The pushing effect is another Bernoulli result: when the stream collides with a surface and bends away, its pressure at the surface rises to cause the bend and that rising pressure pushes the surface forward.

The breakup effect, however, is something new. In trying to go around the obstacle, the stream of water loses its orderly structure and disintegrates into a swirling, hissing froth. Actually, the hiss you hear is familiar; you heard it as you opened the faucet to start water flowing through the hose. That faucet uses a movable stopper to control the flow of water into the hose; as you opened the faucet, you gradually removed this stopper from the water pipe to allow water to flow more freely into the hose and the faucet hissed. Whether water encounters a plant or a faucet stopper, there is something about high speeds and obstacles that upsets the smoothly flowing fluid.

Up until now, we've discussed only **laminar flow**, the smooth, silent flow that's characterized by simple streamlines. In laminar flow, adjacent regions of a fluid always remain adjacent. For example, if you place two drops of dye near one another in a smoothly flowing stream, they will remain close together indefinitely as they follow streamlines in the laminar flow (Fig. 6.1.4, left). Laminar flow is the orderly result of viscous forces, which tend to bring adjacent portions of fluid to the same velocity. When viscosity dominates a fluid's motion, the flow is usually laminar.

However, as the stream flows swiftly past rocks and obstacles, its streamlines break up into the eddies and churning "white water" that make rafting exciting (Fig. 6.1.4, right). The dye is quickly dispersed in this frenzied **turbulence**. The stream is experiencing **turbulent flow**, a roiling, noisy flow in which adjacent regions of fluid soon become separated from one another as they move independently in unpredictable directions. Turbulent flow is the disorderly consequence of inertia, which tends to propel each portion of fluid independently according to its own momentum. When inertia dominates a fluid's motion, the flow is usually turbulent.



Fig. 6.1.4 Water flows slowly past rocks in the stream on the left, and its viscosity keeps it smooth and laminar. Water flows quickly past rocks in the stream on the right, and its inertia separates it into swirling, splashing pockets of turbulence.



Courtesy Lou Bloomfield

The plants and faucet stopper are evidently initiating turbulence in what had been laminar flows; flows that were dominated by viscous forces are suddenly dominated by inertia instead. Whether a flow is laminar or turbulent depends on several characteristics of the fluid and its environment:

1. The fluid's viscosity. Viscous forces tend to keep nearby regions of fluid moving together, so high viscosity favors laminar flow (Fig. 6.1.5).
2. The fluid's speed past a stationary obstacle. The faster the fluid is moving, the more quickly two nearby regions of fluid can become separated and the harder it is for viscous forces to keep them together.
3. The size of the obstacle the fluid encounters. The larger the obstacle, the more likely that it will cause turbulence because viscous forces will be unable to keep the fluid ordered over such a long distance.
4. The fluid's density. The denser the fluid, the less it responds to viscous forces and the more likely it is to become turbulent.

Rather than keeping track of all four physical quantities independently, English mathematician and engineer Osborne Reynolds (1842–1912) found that they could be combined into a single number that permits a comparison of seemingly different flows. The **Reynolds number** is defined as

$$\text{Reynolds number} = \frac{\text{density} \cdot \text{obstacle length} \cdot \text{flow speed}}{\text{viscosity}}. \quad (6.1.2)$$

The units on the right side of Eq. 6.1.2 cancel one another, so the Reynolds number is dimensionless; that is, it's just a simple number, such as 10 or 25,000. As the Reynolds number increases, the flow goes from viscous-dominated to inertia-dominated and therefore from laminar to turbulent. In his experiments, Reynolds found that turbulence usually appears when the Reynolds number exceeds roughly 2300. You can observe this transition by moving a 1-cm-thick (0.4-in-thick) stick through still water. If you move the stick slowly, about 10 cm/s (4 in/s), the Reynolds number will be about 1000 and the flow around the stick will be laminar. But if you speed the stick up to about 50 cm/s (20 in/s), the Reynolds number will rise to about 5000 and the flow will become turbulent.

Courtesy Lou Bloomfield



Fig. 6.1.5 Honey's large viscosity keeps it flowing smoothly (laminar flow) when you pour it. Water's small viscosity allows it to splash about (turbulent flow) in a fountain. In both cases, the effective obstacle length is the width of the flowing fluid, about 1 cm.

One of the most common features of turbulent flows is the **vortex**, a swirling region of fluid that moves in a circle around a central cavity. A vortex resembles a miniature tornado, with its cavity created by inertia as the fluid spins. Vortices are easily visible behind a canoe paddle, in a mixing bowl, or in a cup of coffee stirred rapidly with a spoon. Once an object moves fast enough through a fluid to create turbulence, these vortices begin to form. Each vortex builds up behind the object but is soon whisked away to form a wake of *shed vortices*, vortices that have broken free of the object that made them.

While laminar flow is fully predictable, turbulent flow exhibits chaotic behavior or **chaos**; you can no longer predict exactly where any particular drop of water will go. The study of chaos is a relatively new field of science. Because a **chaotic system**, that is, a system exhibiting chaos, is exquisitely sensitive to initial conditions, even the slightest change in those conditions may produce profound changes in its situation later on.

Even when you can't see turbulent water flow, you can usually hear it. The churning motion of turbulence converts some of the water's ordered energy into thermal energy and sound. The turbulence near the faucet slightly reduces the water's ordered energy as it enters the hose and therefore its speed as it emerges from the nozzle and sprays toward your garden.

A different sound occurs when you abruptly close the nozzle and stop the flow of water. Moving water has momentum, and stopping it suddenly requires an enormous backward force. Since the slowing flow is not steady state, Bernoulli's equation doesn't apply and the pressure can surge to astronomical values near the front of the moving water. This pressure surge is what accelerates the water backward to slow it down and what leads to the loud "thump" sound you hear as the water stops. Known as **water hammer**, the surging pressure in front of stopping water jerks the nozzle, swells the hose, and may even rattle the pipes in your home.

Check Your Understanding #5: Urban Windstorms

On a windy day in a city with many tall buildings, leaves and papers can be seen swirling about in the air or on the sidewalks. What causes these whirling air currents?

Answer: The air flowing through the "canyons" created by the buildings becomes turbulent and forms vortices that swirl the leaves and papers.

Why: Whether an object moves through a fluid or a fluid moves past an object, a Reynolds number is associated with the situation. When the Reynolds number becomes high enough that viscosity is unable to keep the fluid flowing in an orderly fashion, turbulence appears. In wind blowing through a big city, turbulence and its whirling vortices are everywhere.

Check Your Figures #2: Wind on the Open Road

Is the flow of air around a convertible laminar or turbulent as the convertible cruises down the highway? (Air's viscosity is given in Table 6.1.1.)

Answer: Turbulent.

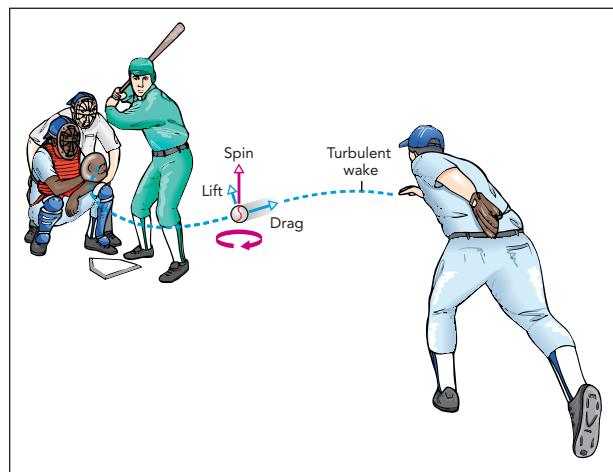
Why: To calculate the Reynolds number for the airflow around the car, you need air's viscosity from Table 6.1.1 (0.0000183 Pa · s, or 0.0000183 kg/m · s), air's density from Section 4.1 (1.25 kg/m³), the car's size (roughly 3 m), and its speed through the air (roughly 55 mph or 25 m/s). You can then calculate its approximate Reynolds number using Eq. 6.1.2:

$$\text{Reynolds number} = \frac{1.25 \text{ kg/m}^3 \cdot 3 \text{ m} \cdot 25 \text{ m/s}}{0.0000183 \text{ kg/m} \cdot \text{s}} = 5.1 \text{ million.}$$

The convertible's Reynolds number is far above the threshold for turbulence (2300), so the air swirls chaotically around the vehicle. That explains why your hair flies around wildly.

SECTION 6.2

Ball Sports: Air



Much of the subtlety and nuance in games such as baseball and golf come from the way balls interact with air. If baseball were played on the moon, which has no air, the only pitches would be the fastball and the not-so-fastball. Moon golfers

wouldn't have to worry about hooks or slices. In this section, we will investigate how air affects the flight of balls and other related objects.

Questions to Think About: Why can you throw a real baseball so much farther than a hollow plastic one? Why does a long fly ball appear to drop straight down when you try to catch it in deep center field? What kind of force could make a curveball curve? What makes a well-hit golf ball hang in the air before falling to the green? How can a knuckleball or spitball jitter about in flight?

Experiments to Do: To make air's effects most apparent, you need a ball that weighs little but has lots of surface area. A beach ball is ideal, but a whiffle ball or hollow plastic ball will also do nicely. See how far you can throw it. How does it stop? Does it slow down and lose height gradually, or does it stop rapidly and fall to the ground? Now make the ball spin as you throw it. Why does the ball curve in flight? Does a faster spin make the ball curve more or less? Which way is the ball spinning, and how is the spin related to the direction of its curve? Change the direction of spin. Which way does the ball curve now?

When a Ball Moves Slowly: Laminar Airflow

One of the first things you might notice if you joined a new baseball franchise on the moon is that pitched balls reach home plate faster than back at home. Since the moon has no atmosphere, there is no air resistance to slow a ball down. In the previous section, we saw how objects affect moving fluids. Now as we study **aerodynamics**, the science of air's dynamic interactions, we'll see how fluids affect moving objects.

In air, a moving ball experiences **aerodynamic forces**—that is, forces exerted on it by the air because of their relative motion. These consist of **drag forces** that push the ball downwind and **lift forces** that push the ball to one side or the other (Fig. 6.2.1). We'll begin our study of ball aerodynamics with drag forces, commonly called *air resistance*, and we'll start with a slow-moving ball. The reason for starting slow is that at low speeds viscous forces are able to organize the air as it flows around the ball; viscosity dominates over inertia, and the airflow around the slow-moving ball is laminar.

Figure 6.2.2 shows the pattern of laminar airflow around a slow-moving ball. Actually, the pattern is the same whether the ball moves slowly through the air or the air moves slowly past the ball. For simplicity, let's move along with the ball and study the airflow from the ball's inertial frame of reference. In that inertial frame, the ball appears stationary with the air flowing past it.

The slow-moving air separates neatly around the front of the ball and comes back together behind it. It produces a **wake**, an air trail behind the ball, that's smooth and free of turbulence. However, the air's speed and pressure aren't uniform all the way around the ball. The airflow bends several times as it travels around the ball and, as we saw in the previous section, such bends always involve pressure imbalances. Since the air pressure far from the ball is steadfastly atmospheric, those pressure imbalances are always caused by pressure variations near the ball's surface. Whenever air bends away from the ball, so that the ball is on the outside of a bend, the pressure near the ball must be higher than atmospheric. Whenever the air bends toward the ball, so that the ball is on the inside of a bend, the pressure near the ball must be lower than atmospheric.

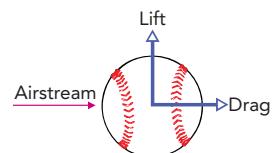


Fig. 6.2.1 The two types of aerodynamic forces exerted on objects by air are drag and lift. Drag is exerted parallel to the onrushing airstream and slows the object's motion through the air. Lift is exerted perpendicular to that airstream so that it pushes the object to one side or the other. Lift is not necessarily in the upward direction.

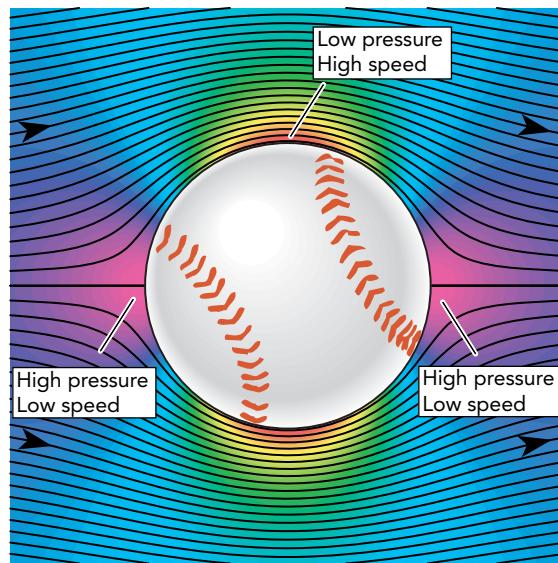


Fig. 6.2.2 The airflow around a slowly moving ball is laminar. Air slows down in front of and behind the ball (widely spaced streamlines), and its pressure increases (shifts toward the violet end of the spectrum). Air speeds up at the sides of the ball (narrowly spaced streamlines), and its pressure decreases (shifts toward the red end of the spectrum). However, the pressure forces on the ball balance one another perfectly, and it experiences no pressure drag. Only viscous drag is present to affect the ball.

With that introduction, let's examine the slow-moving airflow around the ball. Air heading toward the ball's front bends away from it, so the pressure near the front of the ball must be higher than atmospheric. This rise in air pressure is accompanied by a decrease in **airspeed**, the air's speed relative to the ball. Figure 6.2.2 indicates the pressure rise by a color shift toward the violet end of the spectrum and the decrease in airspeed by the widening separation of the streamlines.

The air rounding the ball's sides bends toward it, so the pressure near those sides of the ball must be below atmospheric. This drop in air pressure is accompanied by an increase in airspeed. Figure 6.2.2 indicates the pressure drop by a color shift toward the red end of the spectrum and the increase in airspeed by the narrowing separation of the streamlines.

The laminar airflow continues around to the back of the ball and then trails off behind it. Since the departing air again bends away from the ball, the pressure near the back of the ball must be higher than atmospheric. As before, the shift toward violet in Figure 6.2.2 indicates this pressure rise and the widening separation of the streamlines points out the accompanying decrease in airspeed.

It may seem strange that the air pressure can be different at different points on the ball, but that's what happens in a flowing stream of air. It's particularly remarkable that low-pressure air at the sides of the ball is able to flow around to the back of the ball, where the pressure is higher. This air is experiencing a pressure imbalance that pushes it backward, opposite its direction of travel. But a pressure imbalance causes acceleration, not velocity, and the low-pressure air flowing past the sides of the ball has enough ordered energy and forward momentum to carry it all the way to the back of the ball. Although this air slows as it flows into the rising pressure, it manages to complete its journey.

The airflow around the ball is symmetrical, and the forces that air pressure exerts on the ball are also symmetrical. These pressure forces cancel one another perfectly, so the ball experiences no overall force due to pressure. Most important, the high pressure in front of the ball is balanced by the high pressure behind it. As a result of this symmetrical arrangement, the only aerodynamic force acting on the ball is **viscous drag**, the downstream frictional force caused by layers of viscous air sliding across the ball's surface (see **4**).

4 When the airflow around an object is laminar, the pressure forces on it cancel perfectly and it experiences no drag due to pressure imbalances—no *pressure drag*. The absence of pressure drag was a great puzzle to early aerodynamicists, who knew that the airflow around dust is laminar and that it experiences a drag force. This mystery was named d'Alembert's paradox, after Jean Le Rond d'Alembert (1717–1783), the French mathematician who first recognized it. D'Alembert and his contemporaries didn't know about the viscous drag force, which is what really slows dust's motion through the air.

We'll soon see that viscous drag is only a small fraction of the air resistance experienced by sports balls. It is also the force that suspends dust in the air for hours and is an important issue for airplane wings. Moreover, it's a force that we've encountered before: viscous drag slowed water in your garden hose in the previous section!

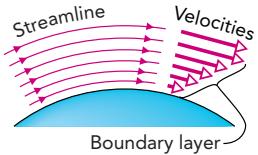


Fig. 6.2.3 As air flows past a surface, a thin layer of it is slowed by viscous drag forces. This boundary layer is laminar at low Reynolds numbers and doesn't become turbulent until the Reynolds number exceeds about 100,000.

Check Your Understanding #1: Smooth Flow in a Stream

When water in a stream flows slowly past a small rock, the water in front of the rock slows down and its increased pressure lifts the water level slightly. The water level behind the rock also rises slightly. Explain.

Answer: The slow flow of water around the rock is laminar, so its pressure is highest at the front and back of the rock. The increased pressure behind the rock lifts the water level there.

Why: Laminar flow around an obstacle tends to create high-pressure regions at the front and back, and low-pressure regions on the sides. In the case of our small rock, the pressure differences are visible as changes in the water level. At the front and back of the rock, the relatively high pressures push the water level upward, while at the sides of the rock, the water level is depressed.

When a Ball Moves Fast: Turbulent Airflow

Balls don't always experience laminar airflow. Turbulence is common, particularly in sports, and brings with it a new type of drag force. When the air flowing around a ball is turbulent, the air pressure distribution is no longer symmetrical and the ball experiences **pressure drag**, the downstream force exerted by unbalanced pressures in the moving air. These unbalanced pressures exert an overall force on the ball that slows its motion through the air.

A ball can experience turbulent airflow and pressure drag when the Reynolds number exceeds about 2000. The Reynolds number, introduced in the previous section, combines the ball's size and speed with the air's density and viscosity to give an indication of whether the airflow is dominated by viscosity or inertia. At low Reynolds numbers, the air's viscosity dominates over its inertia and the airflow is laminar. At high Reynolds numbers, however, air's inertia dominates over its viscosity and the airflow tends to become turbulent. This turbulence, however, won't start until something triggers it, and viscosity provides that trigger.

To understand viscosity's role, we must look at the air near the ball's surface. Even in a strong wind, viscous forces slow down a thin **boundary layer** of air near the ball's surface (Fig. 6.2.3). Discovered by Ludwig Prandtl ⁵ with help from Gustave Eiffel (Fig. 6.2.4), this boundary layer moves more slowly and has less ordered energy than the freely flowing air farther from the surface.

As air flows toward the back of the ball, it travels through an **adverse pressure gradient**, a region of rising pressure that pushes backward on the air and causes it to decelerate. While the freely flowing airstream outside the boundary layer has enough energy and forward momentum to continue onward and reach the back of the ball on its own, air in the boundary layer does not. It needs a forward push.

At low Reynolds numbers, the entire airstream helps to push that boundary layer all the way to the back of the ball and the airflow remains laminar. At high Reynolds numbers, however, viscous forces between the freely flowing airstream and the boundary layer are too weak to keep the boundary layer moving forward into the rising pressure behind the ball.

Without adequate help, the boundary layer eventually **stalls**; that is, it comes to a stop and thereby spoils steady-state flow. More horrible still, this stalled boundary layer air is pushed backward by the adverse pressure gradient and returns all the way to the ball's sides. As it does, it cuts like a wedge between the ball and the freely flowing airstream. The

⁵ Among Ludwig Prandtl's (German engineer, 1875–1953) many pivotal contributions to aerodynamic theory is the concept of boundary layers in fluid motion. Prandtl was so engrossed in establishing Göttingen as the world's foremost aerodynamic research facility that he did not have time to court a wife. Deciding he should be married, Prandtl wrote his former advisor's wife, asking to marry one of her two daughters but not specifying which one. The family selected the eldest daughter, and the wedding took place.

Courtesy Lou Bloomfield



Fig. 6.2.4 Early experiments in aerodynamics were performed by Gustave Eiffel (French engineer, 1832–1923), who designed the tower that bears his name. In the 1890s, Eiffel dropped objects of various sizes and shapes from his tower and measured the drag that they experienced. His work was used by Prandtl to explain the reduction in drag that accompanies the appearance of turbulent boundary layers.

result is an aerodynamic catastrophe—the airstream separates from the ball, leaving a huge turbulent wake or air pocket behind the ball (Fig. 6.2.5).

Because of this turbulent wake, air no longer bends smoothly away from the back of the ball and there is no rise in pressure there. Instead, the pressure behind the ball is roughly atmospheric. The absence of a high-pressure region behind the ball spoils the symmetry of pressure forces on the ball, and those forces no longer cancel. The ball experiences an overall pressure force downwind—the force of pressure drag. In effect, the ball is transferring forward momentum to the air in its turbulent wake and dragging that wake along with it.

Pressure drag slows the flight of almost any ball moving faster than a snail’s pace. The pressure drag force is roughly proportional to the cross-sectional area of the turbulent air pocket and to the square of the ball’s speed through the air. Area times speed is the volume of air the ball affects per second. The second factor of speed recognizes how much the air’s speed changes as the ball drags it along in its air pocket. For a ball moving at a moderate speed, the air pocket is about as wide as the ball and the ball experiences a large pressure drag force.

Check Your Understanding #2: Leaving No Trace

When your canoe coasts extremely slowly across the water of a still lake, it leaves almost no trail in the water behind it. When you paddle it swiftly through the water, however, the canoe leaves a swirling wake. Explain this difference.

Answer: The slow-moving canoe experiences laminar flow in the water, while the fast-moving canoe experiences turbulent flow.

Why: If the canoe’s speed is less than about 1 cm/s, its Reynolds number will be less than 2000 and the water flow around it will be laminar. The water will pass smoothly around the canoe’s sides and join back together behind it. However, when the canoe is moving fast enough that its Reynolds number exceeds 2000, the water flow becomes turbulent and the canoe leaves a churning wake in the water behind it. This wake produces pressure drag on the canoe and extracts energy from it. Anyone who has paddled a canoe knows that overcoming this pressure drag can be exhausting.

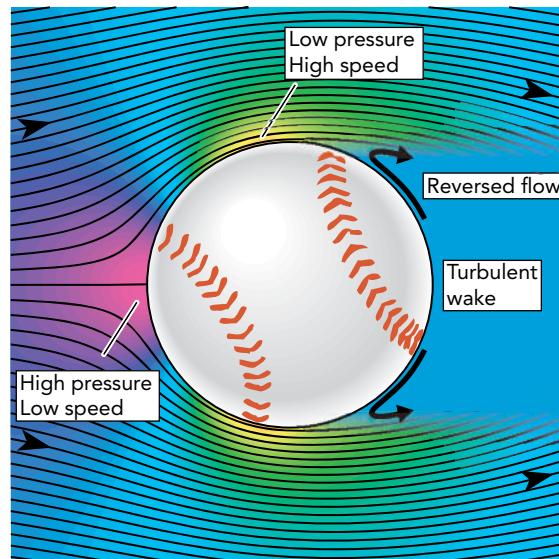


Fig. 6.2.5 When a ball’s speed gives it a Reynolds number between about 2000 and 100,000, its laminar boundary layer stalls in the rising pressure behind the ball. The resulting reversed flow causes the main airflow to separate from the ball’s surface, leaving a large, turbulent wake. The average pressure behind the ball remains low, and the ball experiences a large pressure drag.

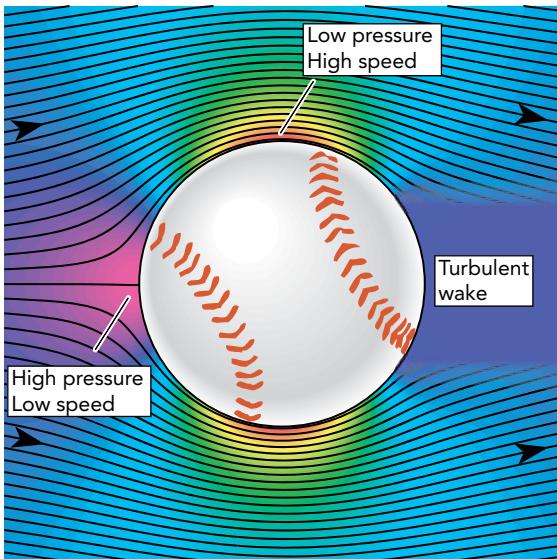


Fig. 6.2.7 When a ball travels fast enough that its Reynolds number exceeds 100,000, its boundary layer becomes turbulent. This turbulent layer travels much of the way around the back of the ball before it separates from the surface. The freely flowing air follows it, and the two leave a relatively small turbulent wake. The ball experiences only a modest pressure drag.

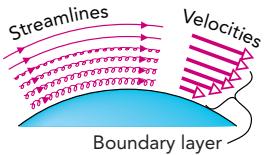


Fig. 6.2.6 When the Reynolds number exceeds about 100,000, the boundary layer of air flowing past a surface becomes turbulent. This whirling fluid brings in extra ordered energy and forward momentum from the freely flowing airstream and can travel deep into a region of increasing pressure.

The Dimples on a Golf Ball

If this were the whole story, you would never hit a home run at a baseball game or a 250-yard drive on the golf course. But inertia has yet another card to play.

At very high Reynolds numbers, the boundary layer itself becomes turbulent (Fig. 6.2.6). It loses its laminar streamlines and begins to mix rapidly within itself and with the freely flowing airstream nearby. This mixing brings additional ordered energy and forward momentum into the boundary layer and makes it both harder to stop and more resistant to reversed flow. Although this turbulent boundary layer still stalls before reaching the back of the ball, the stalled air flows upstream only a short distance. The freely flowing airstream still separates from the ball; however, that separation occurs far back on the ball and the resulting turbulent wake is relatively small (Fig. 6.2.7).

As a result of this smaller air pocket, the pressure drag is reduced from what it would be without the turbulent boundary layer. The effect of replacing the laminar boundary layer with a turbulent one is enormous; it's the difference between a golf drive of 70 yards and one of 250 yards! The effects of the Reynolds number on the airflow around a ball are summarized in Table 6.2.1.

Delaying the airflow separation behind the back of the ball is so important to distance and speed that the balls of some sports are designed to encourage a turbulent boundary layer (Fig. 6.2.8). Rather than waiting for the Reynolds number to exceed 100,000, the point near which the boundary layer spontaneously becomes turbulent, these balls “trip” the boundary layer deliberately (Fig. 6.2.9). They introduce some impediment to laminar flow that causes the air near the ball's surface to tumble about and become turbulent. The drop in pressure drag more than makes up for the small increase in viscous drag, which is why a golf ball has dimples. A tennis ball's fuzz, however, evidently creates more drag than it eliminates; shaving a tennis ball would actually help it maintain its speed.

TABLE 6.2.1 Effects of Reynolds Number on the Airflow around a Ball or Other Object

Reynolds Number	Boundary Layer	Type of Wake	Main Drag Force
<2000	Laminar	Small laminar	Viscous
2000–100,000	Laminar	Large turbulent	Pressure
>100,000	Turbulent	Small turbulent	Pressure



Courtesy Lou Bloomfield

Fig. 6.2.8 Early golf balls were handmade of leather and stuffed with feathers. Golf became popular when cheap balls made of a hard rubber called gutta-percha became available. New, smooth “gutties” (top-left) didn’t travel very far, though; they flew better when they were nicked and worn. Manufacturers soon began to produce balls with various patterns of grooves (top-right) or bumps (bottom-left) on them, and those balls traveled dramatically farther than smooth ones. Modern golf balls (bottom-right) have dimples instead of grooves or bumps.

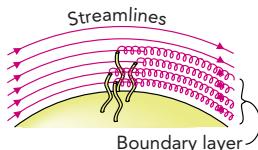


Fig. 6.2.9 The boundary layer can be made turbulent at Reynolds numbers below 100,000 by “tripping” it with obstacles such as fuzz or dimples.

How much does drag affect balls in various sports? For those that involve rapid movements through air or water, the answer is quite a bit. Drag forces increase dramatically with speed; as soon as a turbulent wake and pressure drag appear, the drag force increases as the square of a ball’s speed. As a result, baseball pitches slow significantly during their flights to home plate, and the faster they’re thrown, the more speed they lose. A 90-mph fastball loses about 8 mph en route, while a 70-mph curveball loses only about 6 mph.

A batted ball fares slightly better because it travels fast enough for the boundary layer around it to become turbulent, an effect that appears at around 160 km/h (100 mph). While the resulting reduction in drag explains why it’s possible to hit a home run, the presence of air drag still shortens the distance the ball travels by as much as 50%. Without air drag, a routine fly ball would become an out-of-the-park home run. To compensate for air drag, the angle at which the ball should be hit for maximum distance isn’t the theoretical 45° above horizontal discussed in Section 1.2. Because of the ball’s tendency to lose downfield velocity, it should be hit at a slightly lower angle, about 35° above horizontal (Fig. 6.2.10).

Since the ball loses much of its horizontal component of velocity during its trip to the outfield, a long fly ball tends to drop almost straight down as you catch it. Gravity causes it to move downward, but drag almost stops its horizontal motion away from home plate. Drag also limits the downward speed of a falling ball to about 160 km/h (100 mph). That’s the baseball’s **terminal velocity**, the downward velocity at which the upward drag force exactly balances its downward weight and it stops accelerating. Even if the ball is dropped from a plane, its velocity won’t exceed this value.

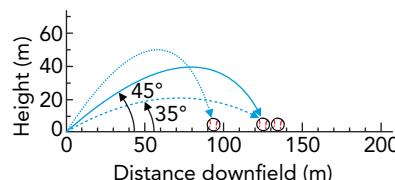


Fig. 6.2.10 Air drag slows the flight of a batted ball so that the ideal angle at which to hit it isn’t the theoretical 45° of Fig. 1.2.7. An angle of roughly 35° above horizontal will achieve the maximum distance.

Check Your Understanding #3: Designing a Great Sports Car

As an automobile designer, your job is to minimize the aerodynamic drag experienced by the car on which you are working. Where should you try to locate the point at which the airflow separates from the car?

Answer: The point should be located as far back on the car as possible.

Why: As for all large, fast-moving objects in air, pressure drag is the main source of air resistance. You want to minimize this drag by keeping the air flowing smoothly over the car until it leaves the rear around a small turbulent wake. The smaller the air pocket behind a car, the better. Aerodynamically designed production cars leave a turbulent wake that is only about one-third as large in area as the thickest cross section of the car. Although there is still room for improvement, these cars experience far less drag than the boxy cars of earlier times.

Curveballs and Knuckleballs

The drag forces on a ball push it downstream, parallel to the onrushing air. In some cases, though, the ball may also experience lift forces—forces that are exerted perpendicular to the airflow (Fig. 6.2.1). To experience drag, the ball only has to slow the airflow down; to experience lift, the ball must deflect the airflow to one side or the other. Although its name implies an upward force, lift can also push the ball toward the side or even downward.

Curveballs and knuckleballs both use lift forces. In each of these famous baseball pitches, the ball deflects the airstream toward one side and the ball accelerates toward the other. Again we have action and reaction—the air and the ball push off one another. Getting the air to push the ball sideways is no small trick. Explaining it isn't easy either, but here we go.

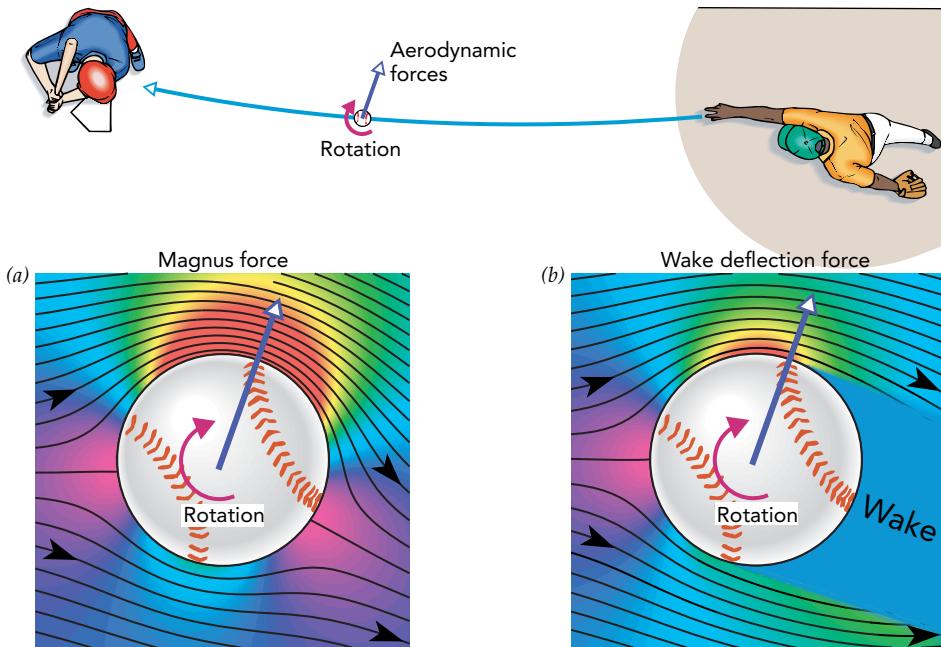
A curveball is thrown by making the ball spin rapidly about an axis perpendicular to its direction of motion. The choice of this axis determines which way the ball curves. In Fig. 6.2.11, the ball is spinning clockwise, as viewed from above. With this choice of rotation axis, the ball curves to the pitcher's right because the ball experiences two lift forces to the right. One is the Magnus force, named after the German physicist H. G. Magnus (1802–1870) who discovered it. The other is a force we will call the wake deflection force.

The **Magnus force** occurs because the spinning ball carries some of the viscous air around with it (Fig. 6.2.11a). The steady-state flow pattern that forms around this ball is asymmetric: the airstream that moves with the turning surface is much longer than the airstream that moves opposite that surface. Since the longer airstream bends mostly toward the baseball, the average pressure on that side of the ball must be below atmospheric. The shorter airstream bends mostly away from the ball, so the average pressure on that side of the ball must be above atmospheric. Because the pressure forces on the ball's sides don't balance one another, the ball experiences the Magnus force toward the low-pressure side—the side turning toward the pitcher—and deflects in that direction. The airflow deflects in the opposite direction.

In laminar flow, the Magnus force is the only lift force acting on a spinning object. However, a pitched baseball has a turbulent wake behind it and is also acted on by the **wake deflection force**. This force appears when the ball's rapid rotation deforms the wake that develops behind it at high Reynolds numbers. When the ball isn't spinning, the freely flowing airstream separates from the ball approximately at its side and that separation is symmetrical all the way around the ball's middle (Fig. 6.2.5). However, when the ball is spinning (Fig. 6.2.11b), the moving surface pushes on the airstream with viscous forces. As a result, airstream separation is delayed on one side of the ball and hastened on the other. The overall wake of air behind the ball is thus deflected to one side, and the ball experiences the wake deflection force toward the opposite side—the side turning toward the pitcher. The wake deflection force and the Magnus force both push the ball in the same direction.

Of these two forces, the wake deflection force is probably the more important for a curveball, although the Magnus force is usually given all the credit. A skillful pitcher can make a baseball curve about 0.3 m (12 in) during its flight from the mound to home plate—

Fig. 6.2.11 A rapidly rotating baseball experiences two lift forces that cause it to curve in flight. (a) The Magnus force occurs because air flowing around the ball in the direction of its rotation bends mostly toward it, while air flowing opposite its rotation bends mostly away from it. (b) The wake deflection force occurs because air flowing around the ball in the direction of its rotation remains attached to the ball longer and the ball's wake is deflected.



the more spin, the more curve. The pitcher counts on this change in direction to confuse the batter. The pitcher can also choose the *direction* of the curve by selecting the axis of the ball's rotation. The ball will always curve toward the side of the ball that is turning toward the pitcher. When thrown by a right-handed pitcher, a proper curveball curves down and to the left, a slider curves horizontally to the left, and a screwball curves down and to the right.

When the pitcher throws a ball with backspin, so that the top of the ball turns toward the pitcher, the ball experiences an upward lift force. In baseball this force isn't strong enough to overcome gravity, but it does make the pitch hang in the air for an unusually long time and appear to "hop." Not surprisingly, a fastball thrown with strong backspin is called a hanging fastball. A fastball thrown with relatively little spin falls naturally and is called a sinking fastball. In golf, where the club can give the ball enormous backspin, the ball really does lift itself upward so that it flies down the fairway like a glider.

There are, however, some interesting cases when a ball's behavior stems from its *lack* of spin. In baseball, for example, a knuckleball is thrown by giving the ball almost no rotation. The ball's seams are then very important. As air passes over a seam, the flow is disturbed so that the ball experiences a sideways aerodynamic force, a lift force. The ball flutters about in a remarkably erratic manner. Releasing the ball without making it spin is difficult and requires great skill. Pitchers who are unable to throw a knuckleball legally sometimes resort to lubricating their fingers so that the ball slips out of their hands without spinning. Like its legal relative, this so-called spitball dithers about and is hard to hit. The same is true for a scuffed ball.



Check Your Understanding #4: Center Court

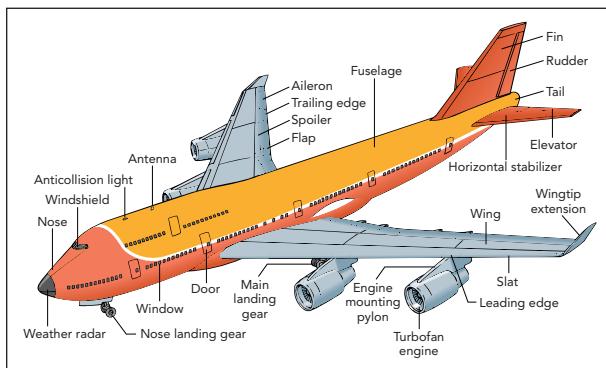
One of the most difficult and effective strokes in tennis is the topspin lob, in which the top of the ball spins away from the player who hit it. Which way is the lift force on this ball directed?

Answer: The lift force is directed downward, so the ball accelerates downward faster than it would by gravity alone.

Why: A ball with topspin falls faster than it would without a spin. In tennis, the topspin strokes appear to dive downward once they cross the net. Their downward curve means that they can travel very fast and still remain inside the court and are thus very hard to return.

SECTION 6.3

Airplanes



We have now set the stage for the ultimate aerodynamic machines—airplanes. Freed from contact with the ground, airplanes are affected only by aerodynamic forces and gravity, hopefully in that order. Despite their complex appearances, airplanes employ physical principles that we have already successfully examined. This section revisits many familiar concepts, but it also explores new territory. For example, you may have already figured out what type of aerodynamic force holds an airplane up, but what type of aerodynamic force keeps it moving forward?

Questions to Think About: Why are the wings of small, propeller-driven aircraft relatively large and bowed compared to those of jets? Why does a commercial airplane extend slats and flaps from its wings during takeoffs and landings? What pushes airplanes forward in flight? How can some airplanes fly

upside down? Why do most fast commercial aircraft employ jet engines and not propellers?

Experiments to Do: The best experiment for this section is to take a plane flight or at least to visit the airport and watch the planes.

As you sit in the plane during takeoff, feel the plane accelerating forward. If you're on a commercial jet, notice that the slats and flaps on the airplane's wings are extended during takeoff, making the wings wider and more curved. How could this increased width and curvature help the plane take off? The pilot holds the plane on the ground until it reaches the proper speed, then quickly tips it upward into the air. An invisible vortex of air peels away from the trailing edge of the wing, and the plane lifts off the ground.

Once airborne, the airplane retracts its landing gear, slats, and flaps. Watch the trailing edge of each wing as the plane turns or changes altitudes; you'll see various surfaces there move up or down. Similar motions occur on the tail. How do these surfaces control the plane's orientation?

Near its destination, the airplane prepares for landing. Again the slats and flaps are extended. Watch as spoilers on the tops of the wings pop up and down noisily. How do these surfaces affect the drag force on the plane? The plane's landing gear extends, and it touches down on the runway. The propellers or jet engines abruptly begin to slow the airplane, assisted by the spoilers on the wings. Feel the plane accelerating backward. Another invisible vortex of air peels away from the trailing edge of the wing, rotating in the opposite direction from the first vortex, and the flight is over.

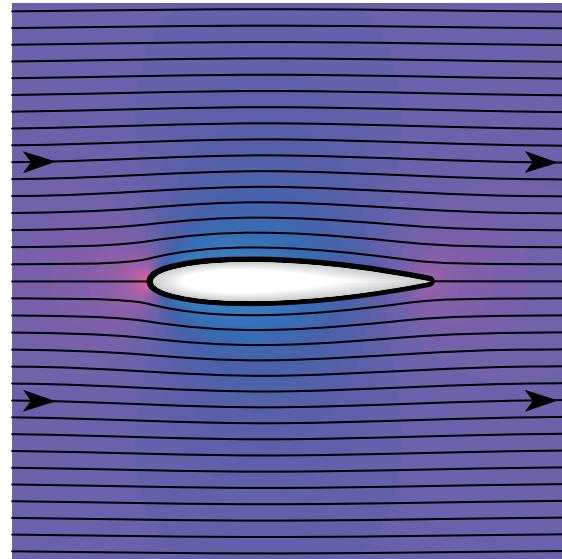
Airplane Wings: Streamlining

By now you've probably realized that an airplane is supported in flight by an upward lift force on its wings and that this lift force comes from deflecting the passing airflow downward. Each wing is an **airfoil**, an aerodynamically engineered surface that's designed to obtain particular lift and drag forces from the air flowing past it. More specifically, each wing is shaped and oriented so that, during flight, the airstream flowing over the wing bends downward toward its top surface while the airstream under the wing bends downward away from its bottom surface. These bends are associated with pressure changes near the wing itself and are responsible for the upward lift force that suspends the airplane in the sky.

However, to get a more complete understanding of how the wing develops this lift, let's go for a flight. Imagine yourself in an airplane that has just begun rolling down the runway. From your perspective, air is beginning to flow past each of the airplane's wings. When this moving air encounters the wing's *leading edge*, it separates into two airstreams: one traveling over the wing and the other under it (Fig. 6.3.1). These airstreams continue onward until they leave the wing's *trailing edge*. Since the airplane's nose is still on the ground, the wing is essentially horizontal and the airflow around it is simple and symmetrical.

Since the wing isn't deflecting the airflow yet, it's experiencing no lift, only drag. However, while this drag pushes the airplane downwind, opposite its forward motion along

Fig. 6.3.1 An airplane wing is a streamlined airfoil, and the airflow around it is laminar. This horizontal wing is symmetric, top and bottom, and the airflow splits evenly into airstreams above and below it. Since it doesn't deflect the airflow, it experiences no lift.



the runway, the effect is weak. The wing produces almost no turbulent wake and thus experiences almost no pressure drag. What little drag it does experience is mostly viscous drag, essentially surface friction with the passing air.

Although the wing's near lack of air resistance should surprise you, you probably take it for granted. That's because you've often observed that such "streamlined" objects cut through the air particularly well. Having a long, tapered tail allows the wing to avoid the flow separation and turbulent wake that occur behind an unstreamlined ball.

What makes the horizontal wing **streamlined** is the extremely gradual rise in air pressure after its widest point. Although this gently rising pressure pushes the wing's boundary layer backward, opposite the direction of flow, the force it exerts is so weak that the layer doesn't stall. Driven onward by viscous forces from the freely flowing airstream, the wing's boundary layer manages to keep moving forward all the way to the wing's trailing edge and never triggers flow separation. The wing produces almost no turbulent wake and experiences almost no pressure drag.

Check Your Understanding #1: Slicing through the Air

The fastest bicycles are those with fairings, streamlined shells that reduce air resistance dramatically. How do those fairings work?

Answer: A fairing delays airflow separation in the airflow around a bicycle and thereby reduces pressure drag.

Why: Without a fairing, a bicycle racer must struggle against severe pressure drag. The sharply rising pressure gradients that follow wide parts of his body trigger flow separation, and he develops a huge turbulent wake. A fairing makes the vehicle streamlined; the gently rising pressure gradient following its widest part doesn't trigger flow separation.

Airplane Wings: Producing Lift

With so little air resistance, the airplane accelerates forward rapidly and soon reaches takeoff speed. The pilot then raises the airplane's nose so that its wings are no longer horizontal, and they begin to experience upward lift forces. The airplane's total lift soon exceeds its weight, and it begins to accelerate upward into the air. The airplane is flying!

Let's take a closer look at the moment of takeoff. If you could see the airflow and were paying close attention, you'd notice a remarkable sequence of events that begins when the wings tilt upward.

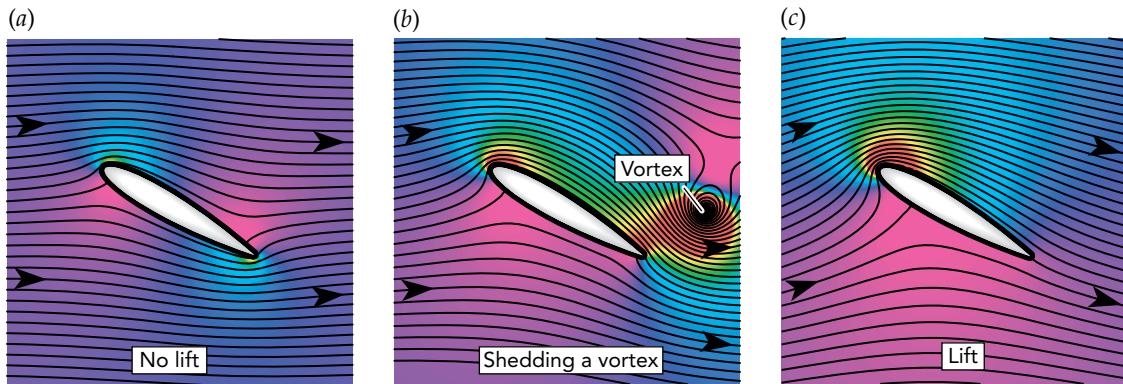


Fig. 6.3.2 (a) Although this wing's leading edge has been tipped upward, giving it a positive angle of attack, the airflow around it is relatively symmetric and produces no lift. (b) The kink at the trailing edge of the wing is unstable and is blown away or shed as a horizontal vortex. (c) The resulting airflow is deflected downward and the wing experiences an upward lift force.

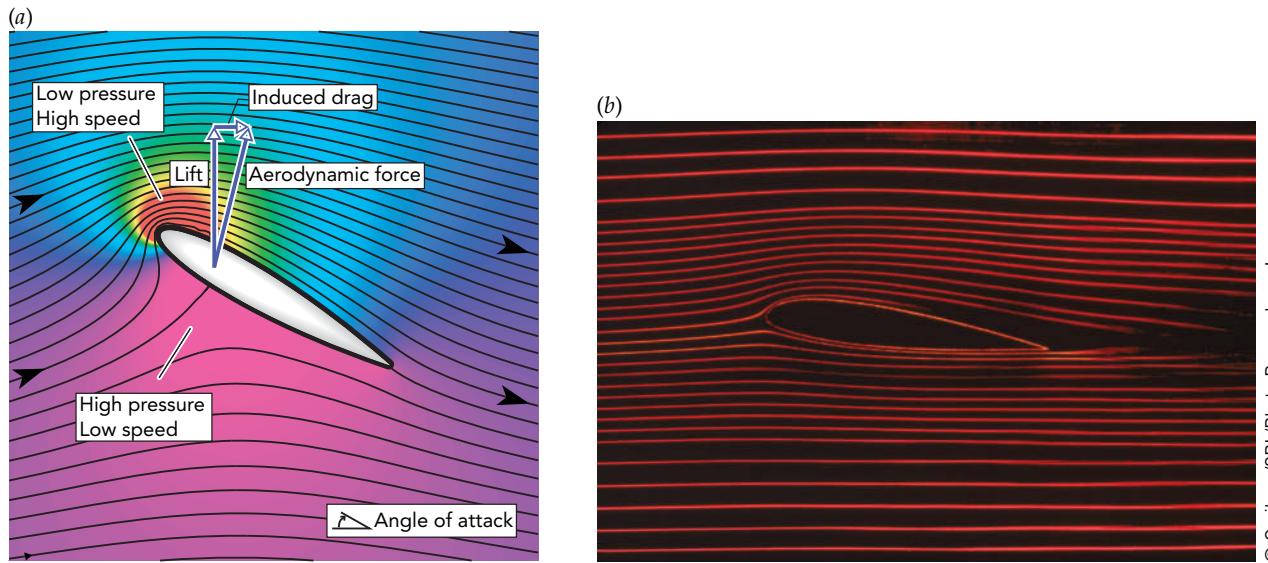
At first, the airflow around the tilted wings continues to travel horizontally on average, although it develops a peculiar shape (Fig. 6.3.2a). The two airstreams, one over the tilted wing and one under it, each bend twice—once up and once down. As we saw while studying balls, when an airstream bends toward the wing, the pressure near the wing is less than atmospheric and when an airstream bends away from the wing, the pressure near the wing is greater than atmospheric. Since each airstream bends equally toward and away from the wing, the average pressures above and below the wing are equal and the wing experiences no lift.

The lower airstream, however, is making a sharp bend around the wing's trailing edge, essentially an upward kink. Air's inertia makes such a kink unstable, and it soon blows away from the wing's trailing edge as a swirling horizontal vortex of air (Fig. 6.3.2b). After shedding that vortex, the wing establishes a new, stable flow pattern in which both airstreams glide smoothly away from the wing's trailing edge (Fig. 6.3.2c), a situation named the *Kutta condition* after the German mathematician M. Wilhelm Kutta (1867–1944).

In this new pattern, the airstream flowing over the wing is longer than the airstream flowing under it and both airstreams bend downward (Fig. 6.3.3). The upper airstream bends primarily toward the wing, so the air's pressure just above the wing is less than atmospheric (a shift toward red) and its speed is increased (narrowly spaced streamlines). In contrast, the lower airstream bends primarily away from the wing, so the air's pressure just below the wing is greater than atmospheric (a shift toward violet) and its speed is decreased (widely spaced streamlines). The air pressure is now higher under the wing than over it, so this new flow pattern produces upward lift. The air now supports your plane and up you go.

Another way to think about this lift is as a deflection of the airflow. Air approaches the wing horizontally but leaves heading somewhat downward. To cause this deflection, the wing must push the airflow downward. In reaction, the airflow pushes the wing upward and produces lift. In other words, the wing transfers downward momentum to the air and is left with upward momentum. These two explanations for lift—the Bernoullian view that lift is caused by a pressure difference above and below the wing, and the Newtonian view that lift is caused by a transfer of momentum to the air—are perfectly equivalent and equally valid.

However, the overall aerodynamic force on the wing isn't quite perpendicular to the onrushing air; it tilts slightly downwind. The perpendicular component of this aerodynamic force is lift, but the downwind component is a new type of drag force—induced drag. **Induced drag** is a consequence of energy conservation; in addition to transferring momentum to the passing air, the wing also transfers some energy to it. The air extracts that energy from the wing by pushing the wing downwind with induced drag and thereby doing negative work on it. Since induced drag is undesirable, the airplane minimizes it by using as much air mass as possible to obtain its lift. A larger mass of air carries away the



© Omikron/SPU/Photo Researchers, Inc.

Fig. 6.3.3 (a) This airplane wing is shaped and oriented so that both airstreams, over it and under it, bend downward. The wing experiences a large aerodynamic force that points upward and slightly downstream. The upward component of this force is lift. The downstream component is induced drag. (b) Smoke trails in a wind tunnel show the airflow past a wing.

airplane's unwanted downward momentum while moving downward less quickly and with less kinetic energy. Since larger wings obtain their lift from larger air masses, they experience less induced drag.

Unfortunately, larger wings also have more surface area and experience more viscous drag, so bigger isn't always better. Also, because wing shape and airspeed affect aerodynamic forces, too, wings must be carefully matched to their airplanes. Small propeller airplanes that move slowly through the air need relatively large, highly curved wings to support them. Those wings are often asymmetrical—more curved on top than on bottom to make maximum use of the limited, low-speed air they encounter each second. Commercial and military jets fly faster and encounter far more high-speed air each second, so they can get by with relatively small, moderately curved wings.

Even at constant airspeed, a wing's lift can be adjusted by varying its **angle of attack**, the angle at which it approaches the onrushing air. The larger the angle of attack, the more the two airstreams bend and the greater the wing's lift. Because the wings are rigidly attached to the plane, the pilot has no choice but to tip the entire plane to adjust its lift. The pilot tips the nose of the plane upward to increase the lift and downward to reduce the lift. That's why raising the plane's nose during takeoff is what finally makes the plane leap up into the air.

Since lift depends so strongly on a wing's angle of attack, some planes can be flown upside down. As long as the inverted wing is tilted properly, it obtains upward lift and supports the plane. This feat is easiest when a plane's wing has the same curvature, top and bottom. That's why stunt fliers who regularly fly upside down often use sport aircraft that have symmetrical or nearly symmetrical wings.

When a plane is in level flight through calm air, the wing's angle of attack is simply its angle above horizontal (Fig. 6.3.4a). As viewed from the plane's inertial frame of reference, the air's velocity is directed horizontally toward the plane and acts as a horizontal *virtual wind* as it approaches the wings. The angle of attack is measured relative to that virtual wind.

When the plane is not in level flight or when the air itself is moving up or down, however, the virtual wind is no longer horizontal. While descending or flying through rising air, the plane encounters a rising virtual wind—the onrushing air's velocity is tilted upward (Fig. 6.3.4b). The wing's angle of attack, still measured relative to that virtual wind, is greater than its angle above horizontal and its lift increases. While ascending or flying through

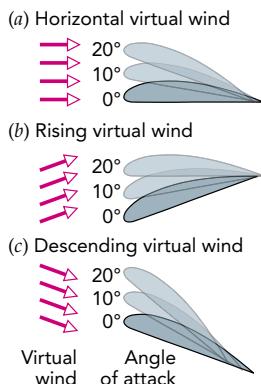


Fig. 6.3.4 (a) When flying horizontally through calm air, an airplane's wing encounters a horizontal virtual wind. The wing's angle of attack is measured relative to that wind. (b) When flying through a rising virtual wind, the wing's angle of attack is greater than its angle above horizontal. (c) When flying through a descending virtual wind, the wing's angle of attack is less than its angle above horizontal.

descending air, the plane encounters a descending virtual wind—the onrushing air's velocity is tilted downward (Fig. 6.3.4c). The wing's angle of attack is less than its angle above horizontal and its lift decreases. When a plane flies through bad weather, with air that is moving alternately up and down, the rapid changes in virtual wind direction lead to fluctuations in lift and exciting accelerations. No wonder there are air sickness bags in the seat pockets.

Check Your Understanding #2: Blowing in the Wind

The sail of a small sailboat bows forward and outward so that wind traveling around the sail's outside surface bends toward the sail, while wind traveling across its inside surface bends away from the sail. How does this arrangement propel the sailboat across the water?

Answer: The air traveling around the outside of the sail has a lower pressure than that traveling across the inside of the sail. The sail experiences a lift force that pushes it and the boat across the water.

Why: Sails experience both lift and drag forces. The sail experiences an aerodynamic force that pushes it outward (lift) and slightly downwind (drag) just as an airplane wing experiences an aerodynamic force upward (lift) and slightly downwind (drag). The sailboat's keel, or centerboard, and its rudder provide additional forces so that the net force on the sailboat can be controlled and it can travel in a variety of directions.

Lift Has Its Limits: Stalling a Wing

There's a limit to how much lift the pilot can obtain by increasing the wing's angle of attack because tilting the wing gradually transforms it from streamlined to **blunt**—that is, to having a rapid rise in air pressure after its widest point. As we saw for balls, blunt objects generally experience airflow separation and pressure drag. Indeed, beyond a certain angle of attack, the airstream over the top of a wing separates from its surface and the wing stalls. This separation starts when air in the upper boundary layer is brought to a standstill by the rapidly rising pressure beyond the wing's widest point. Once this boundary layer stalls, it shaves most of the airstream away from the wing's upper surface.

The separated airstream over the top of the stalled wing leaves a billowing storm of turbulence beneath it (Fig. 6.3.5). This airstream separation is an aerodynamic catastrophe

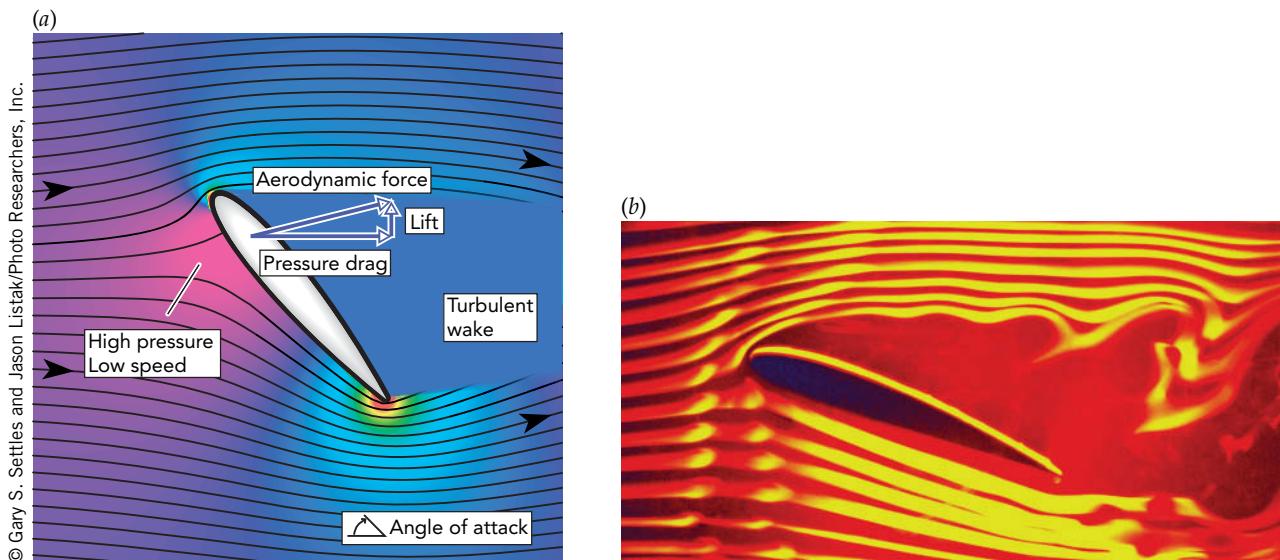


Fig. 6.3.5 (a) A wing stalls when the airstream over the top of the wing separates from its surface. A turbulent air pocket forms above the wing, making it much less efficient. The wing's lift decreases because the average pressure above the thickest part of the wing becomes higher, and the drag increases because the average pressure above the trailing edge becomes lower. (b) Smoke trails in a wind tunnel show that the air separates from the surface and becomes turbulent as it flows over a stalled wing.

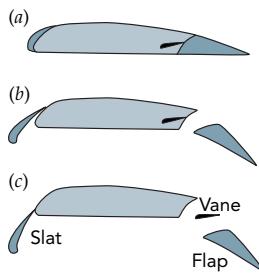


Fig. 6.3.6 At cruising speed, an airplane's wings are moderately curved airfoils (a). During takeoffs (b) and landings (c), slats are extended from the leading edges and flaps from the trailing edges. The airfoils become much more highly curved, generating more lift at low speeds. During landing, a vane is also extended for boundary layer control to prevent stalling.

6 Airplane designers can reduce the dangers of stalling by adding special boundary layer control devices to their aircraft. Narrow metal strips called *vortex generators*, which stick up from the surfaces of wings, introduce turbulence into the boundary layers over the wings. This turbulent flow allows higher-energy air to mix with the boundary layers so that they can continue forward into rising pressure. This process helps keep the airstreams attached to the surfaces.

for the airplane. Because the average pressure above the wing increases, the wing loses much of its lift. Also, the appearance of a turbulent wake heralds the arrival of severe pressure drag. The plane slows dramatically and begins to fall.

To avoid stalling, pilots keep the angle of attack within a safe range. The possibility of stalling also limits the minimum speed at which the airplane will fly. As the airplane slows down, the pilot must increase its angle of attack to maintain adequate lift. Below a certain speed, the airplane can't obtain that lift without tilting its wings until they stall. It can no longer fly.

To avoid stalling, a plane must never fly slower than this minimum speed, particularly during landings and takeoffs. For a small, propeller-driven plane with highly curved wings, the minimum flight speed is so low that it's rarely an issue. For a commercial jet, however, the minimum airspeed is about 220 km/h (140 mph). Airplanes taking off or landing this fast would require very long runways on which to build up or get rid of speed. Instead, commercial jets have wings that can change shape during flight. Slats move forward and down from the leading edges of the wings, and flaps move back and down from the trailing edges (Fig. 6.3.6). With both slats and flaps extended, the wing becomes larger and more curved, similar to the wings of a small propeller plane, and the minimum safe airspeed drops to a reasonable 150 km/h (95 mph). Vanes near the flaps also emerge during landings to direct high-energy air from beneath the wings onto the flaps. These jets of air keep the boundary layers moving downstream and help prevent stalling. (For another approach to stall prevention, see **6**).

Once a commercial jet lands, flat panels on the top surfaces of its wings are tilted upward and cause the airflow to separate from the tops of the wings. The resulting turbulence created by these spoilers reduces the lift of the wings and increases their drag so the plane doesn't accidentally start flying again. Even before landing, the spoilers are sometimes used to slow the plane and help it descend rapidly toward an airport.

In flight, a wing does more than just push the passing air downward; it also twists the air near its tip. Since the air pressure below the wing is greater than the air pressure above it, air tends to flow around the wing's tip from bottom to top. The plane soon leaves this air behind, but not before the air has acquired lots of angular momentum and kinetic energy.

A swirling vortex thus emerges from each wingtip and trails behind the plane for several kilometers, like an invisible tornado. You can occasionally see them behind a plane that's landing or taking off in humid air. A wingtip vortex from a jumbo jet can flip over a small aircraft that flies through it or give passengers in a much larger plane an unexpected thrill. Entered from behind, one of these vortices feels like a horizontal blender; from the side, it feels like a speed hump that you might drive over in a car.

For safety, air traffic controllers are careful to keep planes from flying through one another's wakes and schedule them at least 90 seconds apart on runways. Many modern airplanes have vertical wingtip extensions that reduce these vortices, both to save energy and to diminish the hazard (Fig. 6.3.7).



Check Your Understanding #3: Stunt Flying

A pilot normally tips the plane's nose upward to gain altitude. If the pilot tries to make the plane rise too quickly, the plane will suddenly begin to drop. What is happening?

Answer: The plane's wings are stalling.

Why: Tipping the plane's nose upward increases the wings' angle of attack. While this action increases lift up to a point, it can also cause the airflow to separate from the top surface of the wing. The sudden reduction in lift and increase in drag that accompany stalling can cause the plane to drop. A stall during takeoff or landing is extremely dangerous.



Fig. 6.3.7 This vertical wingtip keeps air from flowing around the end of the wing, a motion that would otherwise leave a powerful vortex in the air trailing behind the plane. Such wingtip vortices waste energy and are hazardous for other aircraft.

Propellers

For a plane to obtain lift, it needs airspeed; air must flow across its wings. And since drag forces push it downwind, a plane in level flight can't maintain its airspeed unless something pushes it upwind. That's why a plane has propellers or jet engines, to push the air backward so that the air pushes the plane forward—action and reaction.

A propeller is an assembly of rotating wings. Extending from its central hub are two or more blades that together form a sophisticated fan (Fig. 6.3.8). These blades have airfoil cross sections and are designed to create forward lift forces when the propeller turns and the blades move through the air.

As a propeller blade slices through the air, the airstreams bending around that blade experience pressure variations (Fig. 6.3.9). The forward airstream bends toward the blade's front surface, so the pressure in front of the blade drops below atmospheric. The rearward airstream bends away from the blade's rear surface, so the pressure behind the blade rises above atmospheric. The resulting pressure difference produces a forward force on the blade and propeller, a **thrust** force.

The propeller blades have all the features, good and bad, of airplane wings. Their thrust increases with size, front-surface curvature, *pitch* (that is, angle of attack), and airspeed; in other words, the larger the propeller, the faster it turns, and the more its blades are angled into the wind, the more thrust it produces. The blades themselves have a twisted shape to accommodate the variations in airspeed along their lengths, from hub to tip.

Like a wing, a propeller stalls when the airflow separates from the front surfaces of its blades; it suddenly becomes more of an air mixer than a propeller. This stalled-wing

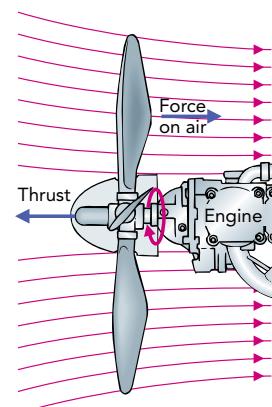


Fig. 6.3.8 A propeller behaves like a rotating wing. As the propeller turns, its blades create lift in the forward direction. This lift pushes the propeller and the aircraft forward through the air, so it's called thrust.

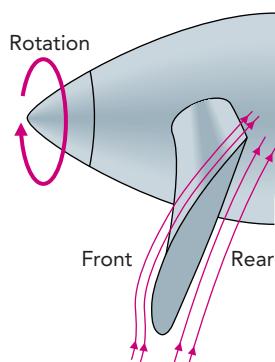


Fig. 6.3.9 As the propeller blade rotates, the flow of air around it creates a low pressure in front of it (left) and a high pressure behind it (right). The blade experiences a lift force that pushes the propeller and plane forward (toward the left). Induced drag tends to slow the rotation of the propeller.

Fig. 6.3.10 The Wright brothers were accomplished aerodynamicists, using this wind tunnel to study and perfect wings and propellers for their airplanes.



The Granger Collection, New York

7 One of the principal sources of noise in submarines is the turbulence created by their propellers. To reduce this turbulence, the propellers of modern nuclear submarines are designed to avoid water flow separation and stalling.

8 In addition to achieving the first self-propelled flight of an airplane in 1903, Orville (1871–1948) and Wilbur (1867–1912) Wright (American aviators) were exceptionally accomplished aerodynamicists. In 1902, Wilbur was the first person to recognize that a propeller is actually a rotating wing. Propellers up until his time were little more than rotating paddles, more effective at stirring the air than propelling the plane. Wilbur's aerodynamically redesigned propeller made flight possible and dominated aircraft design for a decade.

behavior was the standard operating condition for air and marine propellers (see **7**) before the work of Wilbur Wright in 1902 (see **8**). The Wrights were among the first people to study aerodynamics using a wind tunnel (Fig. 6.3.10), and their methodical and scientific approach to aeronautics allowed them to achieve the first powered flight (Fig. 6.3.11). Since the Wrights' work, propellers have experienced almost no pressure drag.

A propeller does, however, experience induced drag. As the propeller's thrust pushes the plane through the air, induced drag extracts energy from the propeller. To keep the propeller turning steadily, an engine must do work on the propeller. Propellers are driven by high-performance reciprocating (piston-based) engines, like those found in automobiles, or by the turbojet engines that we'll discuss later.

Propellers aren't perfect; they have three serious limitations. First, a propeller exerts a torque on the passing air, so that air exerts a torque on the propeller. This reaction torque can flip a small plane. To minimize torque problems, some planes use pairs of oppositely turning propellers and single-propeller planes usually locate their propellers in front so that the spinning air can return angular momentum to them while passing over their wings.

A second problem with a propeller is that its thrust diminishes as the plane's forward speed increases. When the airplane is stationary, a propeller blade moves through motionless air (Fig. 6.3.12a). When the airplane is traveling fast, though, the air approaches that same propeller blade from the front of the plane (Fig. 6.3.12b). To retain its thrust at higher airspeeds, the propeller blade must increase its pitch—that is, its angle of attack. It must swivel forward to meet the onrushing air.

Fig. 6.3.11 The era of powered flight began at 10:35 a.m. on Dec. 17, 1903, when the Wright Flyer lifted Orville Wright into the air over Kitty Hawk, North Carolina. His brother Wilbur stands beside him in this unique photograph of that first powered flight.



The Granger Collection, New York

The third and most discouraging problem with propellers, especially in high-speed aircraft, is drag. To keep up with the onrushing air at high airspeeds, the propellers must turn at phenomenal rates. The tips of the blades must travel so fast that they exceed the **speed of sound**, the fastest speed at which a fluid such as air can convey forces from one place to another. When the blade tip exceeds this speed, the air near the tip doesn't accelerate until the tip actually hits it. Instead of flowing smoothly around the tip, the air forms a **shock wave**, a narrow region of high pressure and temperature caused by the supersonic impact, and the propeller stalls. That's why propellers aren't useful on high-speed aircraft.

Check Your Understanding #4: Circulating the Air

The only difference between a fan and a propeller is what moves—the air or the object. Which side of each fan blade experiences the lowest air pressure, the inlet or the outlet side?

Answer: The inlet side of each fan blade experiences the lowest air pressure.

Why: Air blowing toward you from a fan is like the air blown back by a propeller. The lowest pressures experienced by a propeller are on its forward surfaces. Similarly, the lowest pressures experienced by a fan are on its inlet side surfaces. This pressure imbalance pushes the fan away from you while the fan pushes the air toward you.

Jet Engines

Unlike propellers, jet engines work well at high speeds. While a propeller tries to operate directly in the high-speed air approaching the plane, a jet engine first slows this air down to a manageable speed. To achieve this change in speeds, the jet engine makes wonderful use of the energy transformations allowed by Bernoulli's equation. Strictly speaking, Bernoulli's equation applies only to incompressible fluids such as water. Nonetheless, it's often usable for compressible fluids such as air, especially when the speed and pressure variations are relatively small.

A turbojet engine is depicted in Fig. 6.3.13. During flight, air rushes into the engine's inlet duct or *diffuser* at about 800 km/h (500 mph), the airspeed of the plane. Once inside that diffuser, the air slows down and its pressure increases, leaving its ordered energy unchanged. The air then passes through a series of fanlike compressor blades that push it deeper into the engine, doing work on it and increasing both its pressure and its ordered energy. By the time the air arrives at the combustion chamber, its pressure is many times atmospheric.

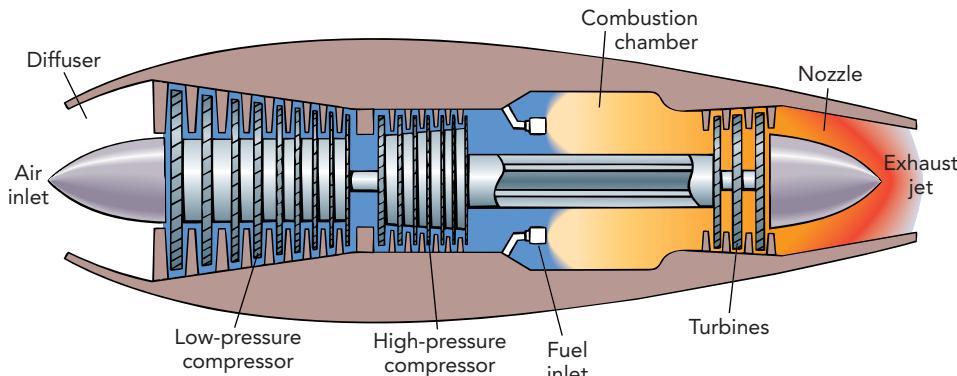


Fig. 6.3.13 The turbojet engine operates by compressing incoming air with a series of fanlike blades. Fuel is mixed with the high-pressure air, and the mixture is ignited. The high-energy, high-pressure air accelerates out the rear of the jet, does work on the turbines, and leaves at a greater speed than it had when it arrived. The engine has accelerated the air backward and experiences a thrust forward.

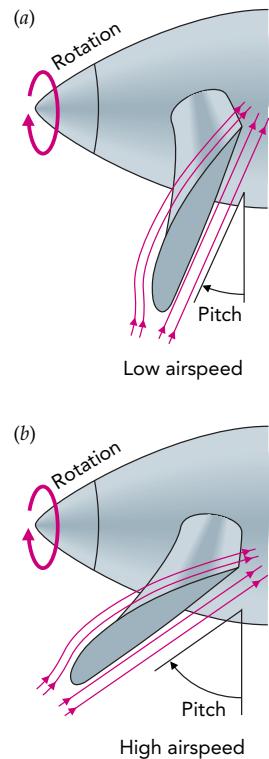


Fig. 6.3.12 At low airspeeds (a), the propeller blade approaches nearly stationary air as it rotates. At high airspeeds (b), air rushes past the propeller, so the blade must swivel forward to meet it. The blade's angle of attack is called its pitch.

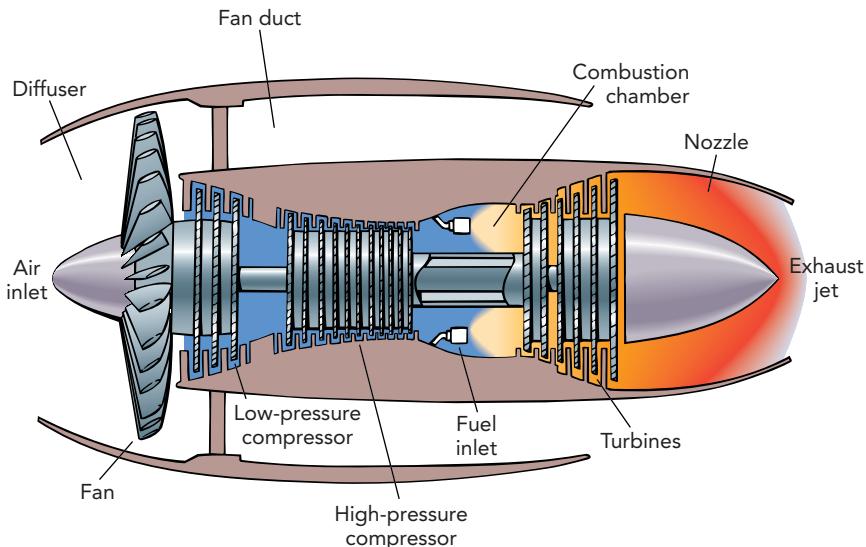


Fig. 6.3.14 The turbofan engine adds a giant fan to the shaft of a normal turbojet engine. Most of the air passing through the fan bypasses the turbojet and returns directly to the airstream around the engine. Because the fan does work on this air, it leaves the engine at a higher speed than it had when it arrived. The air has transferred forward momentum to the engine and the plane.

9 Ramjets are jet engines that have no moving parts. Air that approaches the engine at supersonic speeds interacts with carefully tapered surfaces so that its own forward momentum compresses it to high density. The engine then adds fuel to this pressurized air, ignites the mixture, and allows the hot burned gas to expand out of a nozzle. The engine pushes this exhaust backward, and the exhaust propels the engine and airplane forward. Although the air enters the engine at supersonic speeds, it passes through the combustion chamber much more slowly. In a supersonic combustion ramjet, or “scramjet,” the fuel and air mixture flows through the combustion chamber at supersonic speeds. This motion makes it extremely difficult to keep the fuel burning because the flame tends to flow downstream and out of the engine. The flame can’t advance through the mixture faster than the speed of sound, so it won’t spread upstream fast enough to stay in the engine on its own.

Now fuel is added to the air, and the mixture is ignited. The pressure of this gaseous mixture doesn’t increase as it burns; instead, it expands to occupy more volume. Furthermore, combustion subdivides the fuel molecules into smaller pieces that, therefore, take up still more volume. This extremely hot, high-pressure exhaust gas pours out of the combustion chamber through a wider channel than the one it entered.

The exhaust gas then streams through windmill-like turbine, doing work on that turbine and thereby powering the compressor for the incoming air. After the turbine, the still high-pressure gas finally accelerates and expands through the engine’s outlet nozzle and emerges into the open sky at atmospheric pressure, moderately high temperature, and extraordinarily high speed.

Overall, the engine slows the air down, adds energy to it, and then lets it accelerate back to high speed. Because the engine has added energy to the air, the air leaves the engine traveling faster than when it arrived. The air’s increased backward velocity means that the jet engine has pushed it backward and the air has reacted by exerting a forward thrust force on the jet engine. In other words, the airplane has obtained forward momentum by giving the departing air backward momentum.

The turbojet is less energy efficient than it could be. Since it gives backward momentum to a relatively small mass of air, that air ends up traveling overly fast and with excessive kinetic energy. To make the engine more efficient, it should give backward momentum to a larger mass of air.

The turbofan engine solves this problem by using a turbojet engine to spin a huge fan (Fig. 6.3.14). Since this fan is located in the engine’s inlet duct, the air’s speed decreases and its pressure increases before it enters the fan. The fan then does work on the air and further increases its pressure. While about 5% of this air then enters the turbojet engine, the vast majority of it accelerates out the back of the fan duct and emerges into the open sky at atmospheric pressure and increased speed. Overall, the fan has pushed the air backward and the air has pushed the fan forward, producing forward thrust.

Like a turbojet, the turbofan slows air down, adds energy to it, and then lets it accelerate back to high speed. However, because the turbofan engine moves more air than a turbojet engine, it gives that air less energy and uses less fuel. Because of their increased efficiency, turbofans are used in all modern commercial jets. (For another type of jet engine, see **9**.)

Check Your Understanding #5: Energy and a Jet Engine

A jet engine somehow slows the air down, adds energy to it at low speed, and then returns the air to high speed. Why doesn't slowing the air down waste lots of energy?

Answer: As the air is slowed down, its pressure increases. The air's ordered energy remains constant.

Why: The remarkable result of Bernoulli's effect is that you can slow the air down without squandering its kinetic energy. That energy becomes pressure potential energy, and it passes through the jet engine in that form. As the air leaves the jet engine, its pressure potential energy becomes kinetic energy once again, so no energy is wasted.

Epilogue for Chapter 6

In this chapter, we have looked at a number of objects that use moving fluids to perform their tasks. In Garden Watering, we saw how water moves through openings and channels. We looked at the effects of viscosity and the importance of Bernoulli's equation in describing the conversion of pressure potential energy into kinetic energy and vice versa. Two types of fluid flow appeared: laminar and turbulent. While laminar flow is smooth and predictable, turbulent flow involves unpredictability and chaos, a common behavior in our complicated universe.

In the section on Ball Sports: Air, we examined the ways in which moving air exerts forces on larger objects, both in the downstream direction as drag and in a perpendicular direction as lift. We learned about two different types of drag forces and how these can be controlled or reduced by choosing the shapes or motions of the balls.

Finally, in Airplanes, we explored the aerodynamics of those remarkable machines. We saw how they use air to support and propel themselves. We also examined the limitations of wings and saw what can go wrong if those limitations are exceeded. We studied propulsion in propeller planes and in jet aircraft, and saw that these systems involve Bernoulli's equation and the forward force that comes from pushing air backward.

Explanation: A Vortex Cannon

The vortex cannon creates ring vortices, tiny tornadoes that are bent into loops and so have no beginnings or ends. Air swirls forward in the middle of each ring and backward on its outer edge. This circular tornado structure is created when air flows through the hole of the vortex cannon. Air flows forward in the middle of the hole, while the hole's edges create the backward flow around the outside of the ring. After it leaves the cannon, each ring vortex crawls forward through the surrounding air until its kinetic energy has been exhausted and it slows to a stop. It finishes its existence swirling in place until viscous forces bring its moving air to rest.

Chapter Summary and Important Laws and Equations

How Garden Watering Works: As water flows through a hose, viscous forces in the hose limit its speed, particularly near the walls of the hose. Each time the hose bends, the water pressure rises on the outside of the bend and drops on the inside of the bend. When the water reaches the nozzle, it accelerates through the narrow opening, increasing in speed and decreasing in pressure until it emerges at atmospheric pressure and arcs gracefully into the garden. Whenever the water flows rapidly around obstacles, both in the garden and at the faucet leading the hose, the flow becomes turbulent and noisy.

How Balls and Air Work: A ball traveling through the air experiences two major types of aerodynamic force: drag and lift. For a nonspinning ball traveling very slowly, the only drag

force is viscous drag; it experiences no lift force. As the ball's speed through the air increases, a large turbulent wake appears behind the ball and the ball experiences pressure drag. At a still higher speed, the boundary layer of air near the ball's surface becomes turbulent and the size of the wake behind the ball shrinks. A ball with a turbulent boundary layer experiences less drag than one with a laminar boundary layer, which is why some balls are designed to encourage turbulent boundary layers.

Rotating balls experience lift forces. These forces occur because bending and deflecting airstreams push asymmetrically on these balls. Lift forces can cause a ball to curve in flight or take surprisingly long to fall.

How Airplanes Work: An airplane is supported in flight by air passing across its wings. Air bending toward the top surface of a wing experiences a drop in pressure, while air bending away from the bottom surface of a wing experiences a rise in pressure. As a result of this pressure difference, the wing experiences an upward lift force. The airplane also experiences drag, which tends to slow it down. To keep the airplane moving forward, the plane employs propellers or jet engines. These devices push the air backward, and the air reacts by pushing them forward. A propeller works directly in the oncoming air, increasing the air's energy and pushing it backward with rotating blades. A jet engine first slows the air, then increases its energy by burning fuel in it, and finally lets it accelerate backward to high speed.

1. Poiseuille's law: The volume of fluid flowing through a pipe each second is equal to $(\pi/128)$ times the pressure difference across that pipe times the pipe's diameter to the fourth power, divided by the pipe's length times the fluid's viscosity, or

$$\text{volume} = \frac{\pi \cdot \text{pressure difference} \cdot \text{pipe diameter}^4}{128 \cdot \text{pipe length} \cdot \text{fluid viscosity}}. \quad (6.1.1)$$

2. Bends and pressure imbalances: When the path of a fluid in steady-state flow bends, the pressure on the outside of the

bend is always higher than the pressure on the inside of the bend.

3. Reynolds number: A measure of the relative importance of inertia and viscosity in the fluid flow around an obstacle is given by the Reynolds number:

$$\frac{\text{density} \cdot \text{obstacle length} \cdot \text{flow speed}}{\text{viscosity}}. \quad (6.1.2)$$

We can't see all the motion that takes place around us. Some of it is hidden deep inside each object, where thermal energy keeps the individual atoms and molecules jiggling back and forth in an endless flurry of activity. We're usually aware of this thermal energy only because it determines an object's temperature; the more thermal energy an object contains, the higher its temperature and the hotter it feels.

Nevertheless, thermal energy plays an important role in everyday life. In addition to moving from one place to another, thermal energy can transform a substance from a solid to a liquid to a gas. What you're feeling when you touch a hot object is actually its thermal energy flowing into your comparatively colder hand and raising the temperature of your skin. When thermal energy is flowing in this manner, from a hotter object to a colder one, we call it heat. In this chapter, we examine temperature, heat, and the phases of matter to understand more about our hot and cold world.

ACTIVE LEARNING EXPERIMENTS**A Ruler Thermometer**

One effect that a change in temperature has on a typical object is to change its size. Although this size change is tiny and easily overlooked, you can use mechanical advantage to make it quite visible. Here's how to make a size-change thermometer using only a clear plastic ruler, a pin, a small weight, a piece of stiff paper, and some tape.

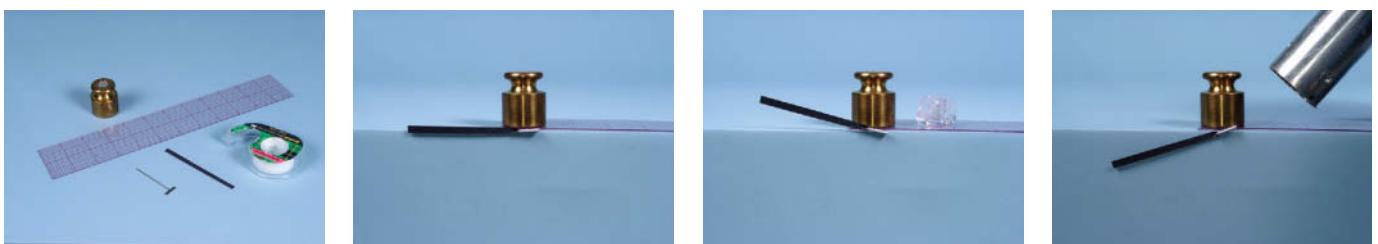
Lay the plastic ruler along the edge of a table and tape one end of it securely to the table. Cut a thin strip of stiff paper, about 3 mm (0.1 in) wide and 15 cm (6 in) long, and push the pin carefully through one end of the strip. Use a dot of tape to stick the pin's head to the paper. When you're done, the paper strip should be securely attached to the pin so that as the pin turns, the strip turns. This strip is your thermometer's pointer.

Now slide the pin under the free end of the ruler, and place the small weight above it. The weight is there to push the ruler and pin together so they experience plenty

of static friction. That way, as the free end of the ruler moves left or right, the pin will rotate and turn the pointer.

Your thermometer is now complete. If you turn the pin and pointer carefully by hand so that the pointer is horizontal, you can "read" the thermometer by its angle relative to the tabletop. If you heat the plastic ruler by breathing on it, laying your hands on it, or warming it gently with a hair dryer, the ruler will become longer. Its free end will move away from its fixed end and will cause the pin to rotate. You will see a small change in the pointer's orientation as your thermometer reports its new temperature.

You can also make the needle turn the other way by cooling the ruler. If you place a few ice cubes on the ruler, the ruler will contract and the needle will turn in the opposite direction. As you can see, the ruler is slightly shorter on a cold day than on a hot one, a fact that limits the ruler's accuracy in measuring distance.



Courtesy Lou Bloomfield

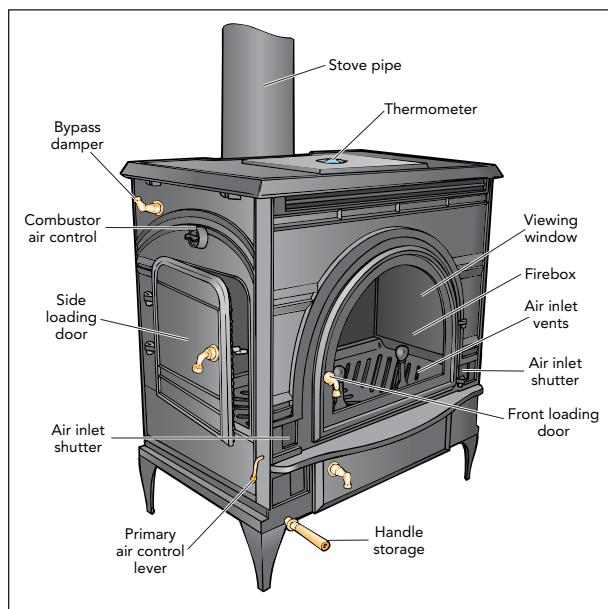
Chapter Itinerary

In this chapter, we examine thermal energy, temperature, and heat in the context of three common types of objects: (1) *woodstoves*; (2) *water, steam, and ice*; and (3) *clothing, insulation, and climate*. In Woodstoves, we look at ways to produce thermal energy and the three principal means by which thermal energy is transferred as heat from hotter objects to colder ones: conduction, convection, and radiation. In Water, Steam,

and Ice, we look at the effects of heat and temperature on the three material phases of water and at the ways transformations occur between those phases. In Clothing, Insulation, and Climate, we examine heat transfer more closely and find out how to control it. For a more detailed preview, look ahead at the Chapter Summary and Important Laws and Equations at the end of the chapter.

SECTION 7.1

Woodstoves



Winter would be pretty unpleasant for most of us were it not for heating. Heating keeps our rooms warm even when the weather outside is cold. One of the most fashionable types of heating is a woodstove, which burns logs in its firebox and sends thermal energy out into the room. In this section, we'll look at how a woodstove produces thermal energy and how this thermal energy flows out of the stove to keep us warm.

Questions to Think About: What happens to the chemical potential energy in a log when you burn it? Why do you get

burned when you touch a hot object? Why does your hand feel hotter when you hold it above a hot surface than when you hold it next to that hot surface? Why does your skin feel warm when you face a campfire, even if the air around you is cold? Why does it take time to warm up a cold object?

Experiments to Do: A burning candle produces thermal energy, providing both light and warmth to a small room. Where does this thermal energy come from and how does it flow into its surroundings?

Light a short candle and look first at how it releases thermal energy. The flame slowly consumes the wax, but it also needs air. Cover the candle with a tall glass, one that won't be touched or damaged by the flame. The glass should prevent room air from reaching the candle. How does the newly sealed environment affect the candle flame? Try to explain this result.

Relight the candle and consider the ways in which heat flows from the flame to you. Carefully pass your hand over the flame, keeping a safe distance above it to avoid being burned. Why does the flame warm your skin so quickly when your hand is directly above it? Now hold your hand beside the flame at a safe distance. You should again feel warmth from the flame. How is heat flowing from the flame to your hand now?

Take a wooden pencil and hold it a few centimeters above the flame for no more than 2 seconds. Then carefully touch the pencil's surface with your fingers. It should warm your fingers. How is heat flowing from the pencil to your hand in this case? Why would it be a painful mistake to try this experiment with a metal pencil? Why should you hold it above the flame for only 2 seconds?

A Burning Log: Thermal Energy

A woodstove produces thermal energy and distributes it as heat to the surrounding room (Fig. 7.1.1). We've encountered thermal energy before: in a file cabinet sliding along the sidewalk, in an old tennis ball bouncing inefficiently off the floor, and in honey pouring slowly from a jar. In each case, ordered energy—energy that could easily be used to do work—became disordered thermal energy and the temperatures of the objects increased. Now that we're going to study devices that are intended to provide heat, let's reexamine thermal energy and temperature to see how thermal energy moves from one object to another.

When you burn a log in the fireplace or woodstove, you're turning the log's ordered chemical potential energy into disordered thermal energy—energy contained in the kinetic and potential energies of individual atoms and molecules. The presence of thermal energy in the log, the woodstove, or the room air is what gives it a temperature; the more thermal energy it has, the higher its temperature.

The nature of thermal energy depends somewhat on what it's in. In the hot burning log, thermal energy is mostly in the wood's atoms and molecules, which jitter back and forth rapidly relative to one another. When each of these particles moves, it has kinetic energy. When it pushes or pulls on its neighbors, it has potential energy.

In the air near the burning log, thermal energy is again mostly in the atoms and molecules. However, since those particles are essentially free and independent, most of this thermal energy is kinetic energy. The air particles store potential energy only during the brief moments when they collide with one another.

In the metal poker that you use to stir the fire, thermal energy is not only in the atoms and molecules but also in the mobile electrons that move about the metal and allow it to conduct electricity. As I noted in Chapter 2, thermal energy is the portion of internal energy that's associated with temperature. Internal energy excludes any energy due to external forces or overall motion, so lifting the poker to increase its gravitational potential energy or waving it to increase its kinetic energy does not affect either its internal energy or its thermal energy. Furthermore, some of the poker's internal energy, including most of its chemical and nuclear potential energy, is not associated with temperature and is not part of its thermal energy.

Check Your Understanding #1: A Warm-Up Pitch

If you drop a ball, will its thermal energy increase as it falls? (Neglect air resistance.)

Answer: No, its thermal energy will remain constant.

Why: As it falls, the ball's gravitational potential energy is transformed into kinetic energy. However, both energies enable the ball to do work, so they aren't included in its thermal energy.

Forces between Atoms: Chemical Bonds

To understand how a burning log produces thermal energy, let's take a look at bonds between atoms and the chemical potential energy that's stored in those bonds. Since both result from the forces between atoms, that's where we'll begin.

As you bring two atoms close together, they exert attractive forces on one another (Fig. 7.1.2a). These chemical forces are primarily electromagnetic in origin and grow stronger as the atoms approach. The attraction diminishes, however, when the atoms start to touch and is eventually replaced by repulsion when the atoms are too close (Fig. 7.1.2b). The separation between atoms at which the attraction ends and the repulsion begins is their *equilibrium separation*—that is, the separation at which the atoms exert no forces on one another (Fig. 7.1.2c). Since atoms are tiny, this equilibrium separation is also tiny, typically only about a ten-billionth of a meter.

Imagine holding two atoms in tweezers and slowly bringing them together. They pull toward one another as they approach, doing work on you and increasing your energy. Since energy is conserved, their energy must be decreasing. They're giving up **chemical potential energy**, energy stored in the chemical forces between atoms.

Once the atoms reach their equilibrium separation, you can let go of them and they won't come apart. Since they've given up some of their chemical potential energy, they can't separate unless that energy is returned to them; it takes work to drag them apart. Without that energy, the atoms are held together by an energy debt known as a chemical bond.

© Peter Anderson/Dorling Kindersley/Getty Images, Inc.



Fig. 7.1.1 This wood-stove transfers heat to the room by conduction through its metal walls, convection of air past its surfaces, and radiation from its black exterior.

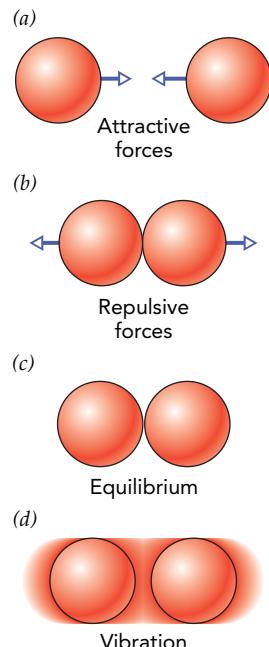


Fig. 7.1.2 (a) Two atoms attract one another at moderate distances but (b) repel when they're too close. (c) In between is their equilibrium separation, at which they neither attract nor repel and are thus in equilibrium. (d) Pairs of atoms with excess energy tend to vibrate about their equilibrium separations.

The bound atoms have become a molecule. The strength of their bond is equal to the amount of work the atoms did when they drew together or, equivalently, the work required to separate them. Bond strengths range from extremely strong in the case of two nitrogen atoms to extremely weak in the case of two neon atoms.

If they have a little extra energy, bound atoms can vibrate back and forth about their equilibrium separation (Fig. 7.1.2d). Whenever the atoms are moving quickly toward or away from one another, most of their energy is kinetic. Whenever the atoms are traveling slowly while turning around, most of their energy is chemical potential. Overall, the molecule's total energy remains constant, and it vibrates back and forth until it transfers its extra energy elsewhere.

Many molecules have more than two atoms. In a large molecule, many pairs of atoms have chemical bonds and equilibrium separations. If you give this molecule excess energy, it will vibrate in a complicated manner as the energy moves among the various atoms and chemical bonds. The atoms in the molecule will continue to jiggle until something removes the excess energy from the molecule.

Like all liquids and solids, our burning log is just a huge assembly of atoms and molecules, held together by chemical bonds of various strengths. These particles push and pull on one another as they vibrate about their equilibrium separations. Their motion is *thermal motion*, and the energy involved in this disorderly jiggling is thermal energy. Because thermal energy is fragmented among the atoms and exchanged between them unpredictably, it can't be used directly to do work.

Check Your Understanding #2: When Atoms Collide . . .

As two independent nitrogen atoms collide with one another, what forces do they experience?

Answer: At first, the atoms experience attractive forces. When the atoms come too close, the forces become repulsive and they bounce. As they subsequently separate, the forces again become attractive.

Why: As they approach one another, the two nitrogen atoms experience attractive forces and a chemical bond begins to form between them. They accelerate toward one another, converting chemical potential energy into kinetic energy. However, when they are very close together, the forces become repulsive and they bounce off one another. They head apart and the forces again become attractive, but their kinetic energy breaks the bond and they separate forever.

Heat and Temperature

Everything contains thermal energy, from the hot burning log to the cold metal poker you use to stir the fire. However, that doesn't mean that the thermal energy is equitably distributed. What happens to thermal energy when you push the log with the poker?

When they touch, the poker and the log begin to exchange thermal energy. In effect, the two objects become one larger object, and thermal energy that has been moving about randomly among the particles in each individual object begins to move about randomly among the particles of both objects. Well, not quite randomly. Although tiny portions of thermal energy move in both directions between the two objects, those exchanges don't cancel. Overall, there is a net flow of thermal energy from the hot log to the cold poker.

To allow us to predict the direction of this flow, we define a temperature for each of the objects. Their temperatures indicate which way, if any, thermal energy will naturally flow between two objects. If no thermal energy flows when two objects touch, then those objects are in **thermal equilibrium** and their temperatures are equal. But if thermal energy flows from the first object to the second, then the first object is hotter than the second.

A temperature scale classifies objects according to which way thermal energy will flow between any pair. An object with a hotter temperature will always transfer thermal

energy to an object with a colder temperature, and two objects with the same temperature will always be in thermal equilibrium. Thus the hot burning log will transfer thermal energy to the cold poker. We say that the burning log is hot because it tends to transfer thermal energy to most objects, while the poker is cold because most objects tend to transfer thermal energy to it.

Energy that flows from one object to another because of a difference in their temperatures is called **heat**. Heat is thermal energy on the move. Strictly speaking, the burning log doesn't *contain* heat; it contains thermal energy. However, when that log transfers energy to the cold poker because of their temperature difference, it's heat that *flows* from the log to the poker. (For a historical note about the understanding of heat, see 1.)

Our present definition of temperature can order the objects around us from hottest to coldest, but it doesn't quantify temperature in any unique way. You could make your own temperature scale by comparing every pair of objects to see which way heat flows between them, but you probably wouldn't enjoy it. You would do better with a standard temperature scale such as Celsius, Fahrenheit, or Kelvin.

Standard temperature scales are based on an object's average thermal kinetic energy per atom. The more kinetic energy each atom has, on average, the more vigorous the object's thermal motion and the more thermal energy it transfers to the atoms in a second object by way of microscopic portions of work. Microscopic work is what actually passes heat between objects—a teeny shove here, a tiny yank there, all at the atomic scale. Since an object with more average thermal kinetic energy per atom will pass heat to an object with less, it makes sense to assign temperatures according to average thermal kinetic energies per atom.

The Celsius, Fahrenheit, and Kelvin scales all measure temperature in this manner. In each scale, a 1-degree, or unit, increase in temperature reflects a specific increase in average thermal kinetic energy per atom. The relationship between average thermal kinetic energy per atom and assigned temperature is based on several standard conditions: absolute zero, water's freezing temperature, and water's boiling temperature. (Recall from Section 5.1 that absolute zero is the temperature at which all thermal energy has been removed from an object.) Once specific temperatures have been assigned to two of these standard conditions, the whole temperature scale is fixed. For example, the Celsius scale is built around 0 °C being water's freezing temperature and 100 °C being water's boiling temperature. Temperatures for the three standard conditions appear in Table 7.1.1.

1 Before the time of Benjamin Thompson, Count Rumford (American-born British physicist and statesman, 1753–1814), heat was believed to be a fluid called caloric that was contained within objects. Thompson disproved the caloric theory by showing that the boring of cannons produced an inexhaustible supply of heat. Among his scientific and technological contributions, Thompson improved cooking and heating methods. He reshaped fireplaces and developed the damper as ways to reduce smoking and improve heat transfer to the room. Thompson also had a life of sensational escapades and great rises and falls in fortune. He fled New Hampshire in 1775 because he was a British loyalist, he fled London in 1782 under suspicion of being a French spy, and he was, at the time of his studies of heat, among the most powerful people in Bavaria.

Check Your Understanding #3: Frozen Fingers

If you pick up an ice cube, your hand suddenly feels cold. Which way is heat flowing?

Answer: From your hotter hand to the colder ice cube.

Why: Heat naturally flows from a hotter object to a colder object. Since the ice cube is colder than your hand, heat flows out of your hand and into the ice cube. Since your hand is losing thermal energy, its temperature drops and you sense cold. Although it's tempting to think of cold as something that flows out of an ice cube, the only thing that really moves about is heat. Ice cubes are wonderful absorbers of heat and cool our drinks by reducing the drinks' thermal energies.

TABLE 7.1.1 Temperatures of Several Standard Conditions, as Measured in Three Temperature Scales: Celsius, Kelvin, and Fahrenheit

Standard Condition	Celsius (°C)	Kelvin (K)	Fahrenheit (°F)
Absolute zero	-273.15	0	-459.67
Freezing water	0	273.15	32
Boiling water	100	373.15	212

Open Fires and Woodstoves

Suppose you needed an easy way to heat your room. The oldest and simplest method would be to start a campfire in the middle of the floor. The burning wood would produce thermal energy, which would flow as heat into the colder room. But how does burning wood produce thermal energy?

This thermal energy is released by a **chemical reaction** between molecules in the wood and oxygen in the air. Recall that atoms do work as they join together in a chemical bond and that the amount of work done depends on which atoms are being joined. For example, while carbon and hydrogen atoms can bond with one another to form *hydrocarbon* molecules, these atoms form much stronger bonds with oxygen atoms. Thus, although it may take work to disassemble a hydrocarbon molecule, the work done by its hydrogen and carbon atoms as they bind to oxygen atoms more than makes up for that investment. As a hydrocarbon molecule burns in oxygen, new, more tightly bound molecules are formed and chemical potential energy is released as thermal energy. The *reaction products* produced by burning hydrocarbons in air are primarily water and carbon dioxide.

Wood is composed mostly of cellulose, a long carbohydrate molecule. *Carbohydrates* contain carbon, hydrogen, and oxygen atoms. Despite the presence of a few oxygen atoms, carbohydrates still burn nicely to form water and carbon dioxide. When you light the wood with a match, you're supplying the energy needed to break the old chemical bonds so that the new bonds can form. This starting energy is called **activation energy**, the energy needed to initiate the chemical reaction. Heat from the match flame gives the wood enough thermal energy to break chemical bonds between various atoms and start the reaction.

Unfortunately, wood isn't pure cellulose. It also contains many complex resins that don't burn well and that create smoke. If you plan to breathe the air in which you burn fuel, wood is an awful choice. You'd be better off with kerosene or natural gas, both of which are nearly pure hydrocarbons and burn cleanly. Actually, wood can be converted to a cleaner fuel by baking it in an airless oven to remove all its volatile resins. This process converts the wood into charcoal, which burns to form nearly pure carbon dioxide, water vapor, and ash.

However, even with clean-burning fuels, the direct fire-in-the-room heating concept has its disadvantages—it consumes the room's oxygen and presents a safety hazard. Nonetheless, fires have heated dwellings for thousands of years. While fireplaces that burn wood or peat have chimneys that carry away their noxious fumes, the rising smoke also takes with it much of the fire's thermal energy and some of the room's air. That's why a room that's heated by a fireplace often feels drafty away from the fireplace itself—cold outside air is seeping in through cracks to replace air drawn up the chimney. Even when clean-burning fuels are used without a chimney, there are no simple solutions to the oxygen-loss or safety problems.

Like a fireplace, a woodstove sends fumes from its burning wood up a chimney. But before its thermal energy can follow the fumes outside, a well-designed woodstove transfers most of that energy into the room. A woodstove is an example of a **heat exchanger**, a device that transfers heat without transferring the hot molecules themselves. Its smoke never enters the room, but heat from that smoke does. Most combustion furnaces, such as those fueled by natural gas, propane, oil, and coal, also employ heat exchangers.

The burning coals and hot gases inside the woodstove contain a great deal of thermal energy and are much hotter than the room air. Because of this temperature difference, heat tends to flow from the fire to the room. What is not so clear yet is how that heat is transferred.

There are three principal mechanisms by which heat moves from the fire to the room: conduction, convection, and radiation. The woodstove makes wonderful use of all three so that most of the thermal energy released by the burning wood is transferred to the room. Let's examine these three mechanisms of heat transport, beginning with conduction.

Check Your Understanding #4: Feeling the Heat

You can make a heat pack by wrapping hot wet towels in a plastic sheet. This pack will warm an injured muscle but will not get it wet. Is thermal energy moving in this case?

Answer: Yes, thermal energy is flowing from the hot towels, through the plastic, to the muscle.

Why: The plastic sheet is acting as a heat exchanger, allowing heat to flow from the hot towels to the cooler muscle but preventing any movement of the hot water itself.

Heat Moving through Metal: Conduction

Conduction occurs when heat flows through a stationary material. The heat moves from a hot region to a cold region, but the atoms and molecules don't. For example, if you place the tip of a metal poker in the fire, the poker's handle will gradually become warm as the metal conducts heat.

Some of this heat is conducted by interactions between adjacent atoms. The vibrating atoms frequently push on one another, doing microscopic work in the process and exchanging minuscule amounts of thermal kinetic energy. In this fashion, thermal energy flows randomly from atom to neighboring atom.

However, when the poker's tip is hotter than its handle, the flow is no longer completely random. The atoms at the hot tip have more thermal kinetic energy to exchange with their neighbors than atoms at the cold handle. The exchanges statistically favor the flow of thermal energy away from the hot tip and toward the cold handle. This flow of thermal energy from hot to cold through the poker is conduction (Fig. 7.1.3).

This atom-by-atom bucket brigade isn't the only way in which materials conduct heat. In a metal, the primary carriers of heat are actually mobile **electrons**, the tiny negatively charged particles that make up the outer portions of atoms. When atoms join together to form a metal, some of the electrons stop belonging to particular atoms and travel almost freely throughout the metal. These mobile electrons can carry electricity (as we'll discuss in Chapter 10) and are also good at transporting heat.

Mobile electrons participate in the bucket brigade of heat conduction because they, too, can push on vibrating atoms and exchange thermal kinetic energy with them. However, while atoms can pass thermal energy only from one neighbor to the next, mobile electrons can travel great distances between exchange partners and can move thermal energy quickly from one place to another.

The ease with which electrons move heat about a metal explains why metals generally have higher thermal conductivities than nonmetals. **Thermal conductivity** is the measure of how rapidly heat flows through a material when it's exposed to a difference in temperatures. The best conductors of electricity—copper, silver, aluminum, and gold—are also the best conductors of heat. Poor conductors of electricity—stainless steel and insulators such as plastic and glass—are also poor conductors of heat. There are a few exceptions to this rule. Diamonds, for example, are terrible conductors of electricity but wonderful conductors of heat.

Conduction is what moves thermal energy from the woodstove's inside to its outside. No atoms move through the metal walls of the stove, just heat. So conduction serves as a filter, separating desirable thermal energy from the unwanted smoke and noxious gases that then go up the chimney.

Thus conduction makes the outside surface of the woodstove hot, allowing heat to flow from it to the colder room. What carries that heat into the room? If you touch the stove, conduction will immediately transfer a huge amount of heat to your skin and you'll be burned. Even without touching the stove, you're aware of its high temperature. It transfers heat into the room by convection and radiation.

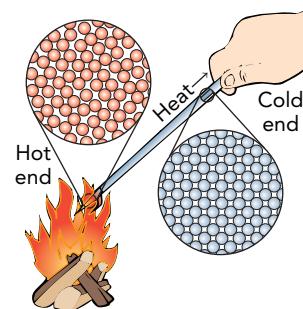


Fig. 7.1.3 When one end of a metal poker is hotter than the other, the atoms at the hot end vibrate more vigorously than those at the cold end. The poker then conducts heat from the hot end to the cold end. Some of this heat is conducted by interactions between adjacent atoms. In the metal poker, however, most of the heat is conducted by mobile electrons, which carry thermal energy long distances from one atom to another.

Check Your Understanding #5: Too Hot to Handle

Some pot handles remain cool during cooking while others become unpleasantly hot. What determines which handles remain cool and which become hot?

Answer: It's determined by the handle's thermal conductivity.

Why: Some handles are made of plastics or stainless steel, which are poor conductors of both electricity and heat. These handles normally remain cool unless hot gases from the stove directly heat them. Other handles are made from aluminum, copper, or cast iron, which are good conductors of electricity and heat. These handles often become unbearably hot.

Heat Moving with Air: Convection

Convection occurs when a moving fluid transports heat from a hotter object to a colder object. The heat moves as thermal energy in the fluid so that the two travel together. The fluid usually follows a circular path between the two objects, picking up heat from the hotter object, giving it to the colder object, and then returning to the hotter object to begin the cycle again.

This circulation often develops naturally. As the fluid warms near the hotter object, its density decreases and it floats upward, lifted by the buoyant force. When the fluid cools near the colder object, its density increases and it sinks downward.

Thus air heated by contact with the woodstove rises toward the ceiling and is replaced by colder air from the floor (Fig. 7.1.4). Eventually, this heated air cools and descends. Once it reaches the floor, it's drawn back toward the hot woodstove to start the cycle over. This moving air is a **convection current**, and the looping path that it follows is a **convection cell**. Within the room, convection currents carry heat up and out from the woodstove to the ceiling and walls. When you put your hand over the stove, you feel this convection current rising as it transfers heat to your hand.

Natural convection is good at heating the air above the woodstove, but most of that hot air ends up near the ceiling. Although some of it will eventually drift downward to where you're standing, convection sometimes needs help. Adding a ceiling fan will help move the hot air around the room and make the woodstove more effective. This forced convection still transfers heat from the hot stove to the colder occupants of the room, but it doesn't rely on the buoyant force to keep the air circulating. The faster the air moves, the more heat it can transport from hot objects to cold objects.

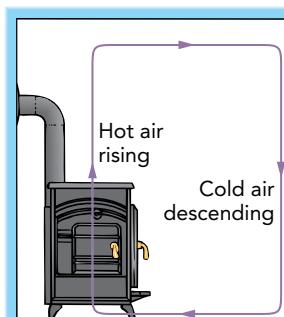


Fig. 7.1.4 When the woodstove is hot, convection carries heat from its surfaces to the ceiling and walls of the room. Warm air rises upward, supported by the buoyant force, and is replaced by cooler air from the floor. The warmed air eventually cools and descends. It then returns toward the stove to repeat the cycle.

Check Your Understanding #6: Heat and Wind

When sunlight warms the land beside a cool body of water, a breeze begins to blow from the water toward the land. Explain.

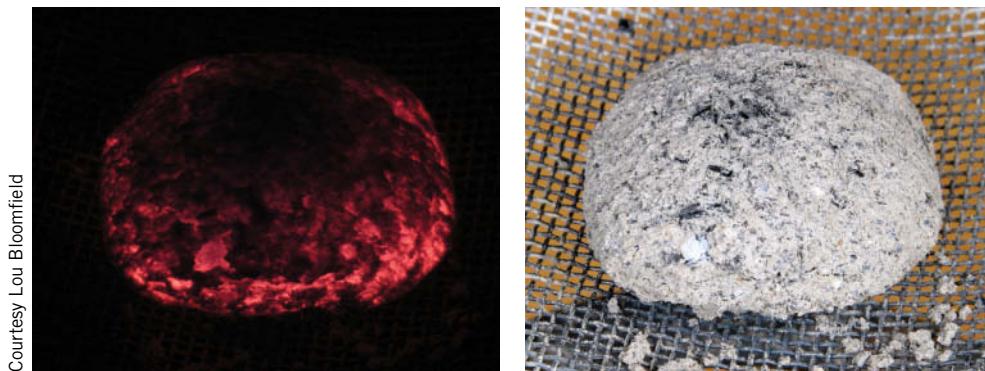
Answer: Convection occurs, with warmed air rising over the land and being replaced by cooler air from above the water. The air moving from over the water to over the land creates the breeze.

Why: Winds are giant convection currents caused by solar heating. Air rises over warm spots on Earth's surface, and surface winds blow toward those warm spots to replace the missing air.

Heat Moving as Light: Radiation

There is one more important mechanism of heat transfer—radiation. As the particles inside a material jitter about with thermal energy, they emit and absorb electromagnetic radiation. This radiation consists of *electromagnetic waves*, which include radio waves, microwaves, and infrared, visible, and ultraviolet light.

We'll study electromagnetic radiation later in this book. For now, what's most important is that this radiation can carry thermal energy. When heat flows from a hot object to a



Courtesy Lou Bloomfield

Fig. 7.1.5 A hot charcoal briquette glows brightly in the dark (left). However, a flash photograph reveals that its surface is actually gray and thus a partial absorber of light (right). If it weren't for the white ash, the briquette's black carbon would make it a nearly ideal emitter of blackbody radiation and absorber of light.

cold object as electromagnetic radiation, we say that heat is being transferred by thermal radiation or simply **radiation**. Unlike conduction and convection, which depend on atoms, molecules, or electrons to carry the heat, radiation occurs directly through space. Radiative heat transfer happens even when two objects have nothing at all between them.

The types of electromagnetic waves in an object's thermal radiation depend on its temperature. While a colder object emits only radio waves, microwaves, and infrared light, a hotter object can also emit visible or even ultraviolet light. The red glow of a hot coal in the woodstove is that coal's thermal radiation.

Since our eyes are sensitive only to visible light, we can't see all the thermal radiation emitted by an object, even when it's hot. However, whether we see it or not, electromagnetic radiation contains energy and transfers heat to whatever absorbs it. While everything emits thermal radiation, the amount of that emission depends on temperature: the hotter an object, the more thermal radiation it emits. When two objects face one another, thermal radiation will travel in both directions between them. However, the hotter object will dominate this radiant exchange of thermal energy, resulting in a net transfer of thermal energy to the colder object. Exchanges of thermal energy via radiation always transfer heat from a hotter object to a colder one.

Radiation transfers a great deal of heat from the woodstove's surface to the surrounding objects. The stove bathes the room in infrared light, which warms everything it reaches. To encourage such radiative heat transfer, the metal woodstove and its chimney are often painted black. Black not only absorbs light well, but it's also particularly good at emitting thermal light (Fig. 7.1.5). If you heat a black poker red hot, it will glow much more brightly than one that's white, silvery, or transparent.

COMMON MISCONCEPTIONS: Black Objects and Light

Misconception: A black object never emits light.

Resolution: Although a black object absorbs all light that strikes it, it still emits thermal radiation and will glow brightly if it's hot enough.

Even if air in the room is cold, you can usually feel the invisible infrared light from a woodstove on your face. When you block this light with your hands, your face suddenly feels colder because less heat is reaching your skin. This thermal radiation effect is even more pronounced with a fireplace or campfire, where thermal radiation from the hot coals and flames is the primary mechanism for heat transfer to the surroundings.

Overall, a modern woodstove is an excellent heat exchanger. As convection draws hot smoke up the long black chimney pipe, the smoke heats the stove and the pipe. These metal components conduct heat to their outer surfaces, which then distribute it around the room by convection and radiation. Although the stove consumes some room air, it controls the airflow with dampers so that it draws in only enough air to completely burn the wood. Overall, the stove extracts heat efficiently, cleanly, and safely from the burning wood.

Check Your Understanding #7: Keeping Warm

When you stand under a heat lamp, you feel warm even though the lamp emits little visible light. How is heat reaching your skin?

Answer: Radiation transfers heat from the lamp's filament to your skin.

Why: A heat lamp emits large amounts of invisible infrared radiation. Although you can't see this radiation, you can feel it on your skin.

Warming the Room

You light the woodstove, and its heat begins flowing out into the cold room. Moving always from a hotter object to a colder object, heat enters each item in the room and gradually raises its temperature. For example, a once-frigid brass bowl near the woodstove is soon pleasantly warm to the touch.

Let's take a look at the relationship between the heat added to that bowl and its temperature rise. Because the bowl's temperature increases steadily when heat is flowing into it steadily, its overall temperature rise must be proportional to the added heat. The constant of proportionality is called the bowl's **heat capacity** and is the amount of heat that must be added to the bowl to cause its temperature to rise by 1 unit. In effect, the bowl's heat capacity is the measure of its thermal sluggishness, its resistance to temperature changes.

Now suppose that you have an assortment of bowls near the woodstove, each a different material. If you keep track of their temperatures, you'll find that some of them warm faster than others. Even when you take into account differences in their masses and in how much heat each one is receiving from the woodstove, you'll find that bowls made of different materials respond differently to added heat. Some materials are more thermally sluggish than others.

We can characterize each material by its heat capacity per unit mass, a quantity known as **specific heat**. To find the specific heat of the material from which a particular bowl is made, you divide that bowl's heat capacity by its mass, or

$$\text{specific heat} = \frac{\text{heat capacity}}{\text{mass}}.$$

The SI unit of specific heat is the **joule per kilogram-kelvin** (abbreviated J/kg · K).

Table 7.1.2 contains the specific heats for a number of common materials. The wide range of values indicates that different materials respond quite differently to added heat. Each material's specific heat depends principally on the number of microscopic ways it can store thermal energy per kilogram. Known as a *degree of freedom*, each independent way of handling thermal energy stores an average thermal energy equal to one-half the Boltzmann constant times the absolute temperature ($\frac{1}{2} kT$). The Boltzmann constant, which we first encountered in Section 5.1, has a measured value of 1.381×10^{-23} J/K, and its units here are equivalent to those that appeared in Section 5.1.

Brass's relatively small specific heat explains why the brass bowl heats up so quickly when you place it directly on the woodstove; the bowl has relatively few degrees of

TABLE 7.1.2 Specific Heats Measured near Room Temperature (293 K) and Atmospheric Pressure

Material	Specific Heat
Lead	128 J/kg · K
Brass	380 J/kg · K
Copper	386 J/kg · K
Air (at constant volume)	715 J/kg · K
Glass	840 J/kg · K
Aluminum	900 J/kg · K
Air (at constant pressure)	1001 J/kg · K
Wood	~1100 J/kg · K
Plexiglas or Lucite	1349 J/kg · K
Steam (at constant pressure)	2027 J/kg · K
Ice	2220 J/kg · K
Water	4190 J/kg · K

freedom in which to store its thermal energy. But if you add even a modest amount of water to the bowl, the water's astonishingly large specific heat will dramatically slow the bowl's temperature rise. Water has a remarkable capacity for thermal energy.

Like brass or water, air also has a specific heat. However, air's specific heat depends on how you measure it. That's because gases tend to expand as they warm up. If you seal the air in a bottle, so that its volume doesn't change, it warms relatively easily; air's specific heat at constant volume is 715 J/kg · K. But if you allow the air to expand as its temperature rises, so that its pressure doesn't change, it's harder to warm. That's because it needs extra energy to push the surrounding air out of its way as it expands; air's specific heat at constant pressure is 1001 J/kg · K.

Since the room isn't perfectly sealed, the air inside it expands as it's heated and the larger specific heat applies. If it's a typical living room, then its volume is about 40 m³ (1400 ft³) and it contains about 50 kg (110 lbm) of air. The heat needed to warm that air is proportional to how much you want to raise its temperature and is equal to air's specific heat times its mass, or about 50,000 J/K. For every kelvin or degree Celsius you want to warm the room air, the woodstove has to provide about 50,000 joules of heat!

When the temperature outside is 0 °C (32 °F), it takes about 1,000,000 joules to warm the room's air to 20 °C (68 °F). That's the heat energy in about 0.07 kg (2 ounces) of wood. Every time you open the front door and let the room fill with outdoor air, the woodstove has to burn that amount of wood to return the room air to 20 °C. You can save energy and help the environment simply by keeping the doors and windows closed whenever you're trying to keep the room air hotter (or colder) than it is outside.

Check Your Understanding #8: Hot Stuff

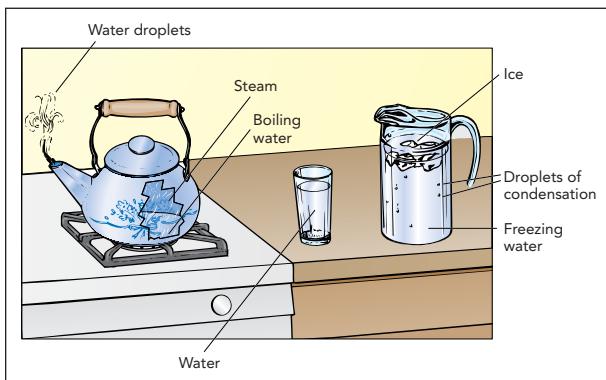
When you pull a metal tray of moist cookies out of the oven, the tray initially feels much hotter than the cookies. A little while later, however, the cookies feel much hotter than the tray. Explain.

Answer: At first, both the tray and the cookies are hot. The tray feels hottest because its metal conducts heat well and can deliver more heat to your skin. But the tray has a relatively small heat capacity, and its temperature plummets in the cool room air. The moist cookies have large heat capacities, so they cool relatively slowly.

Why: Metals often feel hotter or colder to the touch than insulators because they're so good at transferring heat in either direction. However, materials containing water have much larger specific heats than metals and therefore maintain their temperatures better.

SECTION 7.2

Water, Steam, and Ice



Water is probably the most important chemical in our daily lives. It is so crucial to biology, climate, commerce, industry, and entertainment that it merits a whole section of its own. Moreover, it exhibits the three classic phases of matter—solid, liquid, and gas—and illustrates the role that heat plays in transforming one phase into another. Although most of what we can learn from water is applicable to any material, there are a few aspects of water that are almost unique in nature. Water really is a remarkable substance.

Questions to Think About: Why does adding ice make a drink cold? Why do icebergs float on water? How does perspiration cool you off? Why do humid days feel so muggy? What is the difference between evaporating and boiling? How does snow disappear from the ground even when it's very cold outside?

Experiments to Do: Experimenting with water is easy. Pick up an ice cube with wet fingers. What happens if the ice cube has just come out of the freezer? What if it has been melting on the table for a few minutes? Why is there a difference? Put both ice cubes in water. Why do they float?

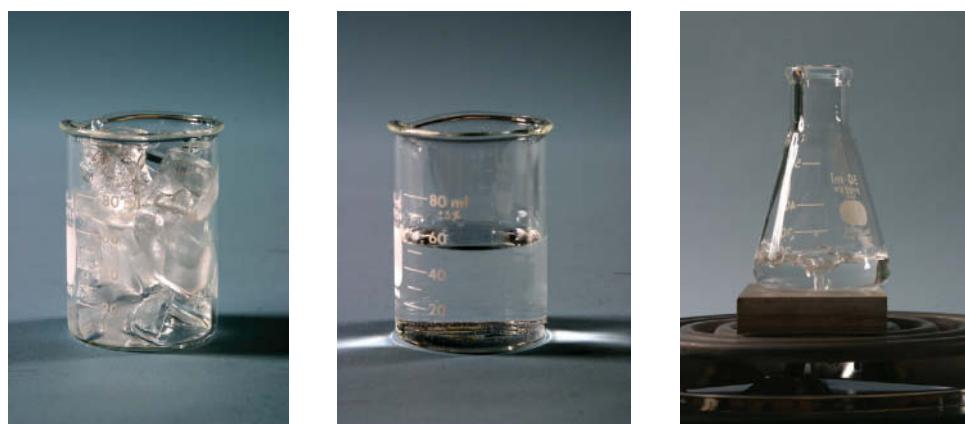
Now heat tap water in a pot. Shortly before the water starts to boil, you will see mist begin to form above it. What is this mist, and why does it form? Carefully feel the mist, but don't burn yourself! The mist feels damp because it contains water. How can water be leaving the pot before the water boils? Notice the small gas bubbles on the walls of the pot. They're not steam; where did this gas come from?

Once the water boils, invisible steam (gaseous water) will push the mist layer upward. Don't touch this steam because it can burn you quickly. Why does steam release so much thermal energy when it touches your skin?

Solid, Liquid, and Gas: The Phases of Matter

Like most substances, water exists in three distinct forms or **phases**: solid ice, liquid water, and gaseous steam (Fig. 7.2.1). These phases differ in how easily their shapes and volumes can change. Ice is **solid**, rigid and incompressible; you can't alter an ice cube's shape or its volume. Water is **liquid**, fluid but incompressible; you can reshape the water in a pitcher but you can't change its volume. Steam is **gaseous**, fluid and compressible; you can vary both the shape and the volume of the steam in a tea kettle.

These different characteristics reflect the different microscopic structures of steam, water, and ice. *Steam, or water vapor, is a gas, a collection of independent molecules kept*



Courtesy Lou Bloomfield

Fig. 7.2.1 The three phases of water: solid (ice), liquid (water), and gas (steam).

in motion by thermal energy. These water molecules bounce around their container, periodically colliding with one another or with the walls. The water molecules fill the container uniformly and can accommodate any changes in its shape or size. Enlarging the container simply decreases the steam's density and lowers its pressure.

When they are independent of one another as gaseous steam, water molecules have a substantial amount of chemical potential energy. They can release some of this energy by joining together to form ordinary water. *Water* is a liquid, a disorderly collection of molecules that cling to one another with chemical bonds. Because these bonds aren't very strong, the molecules in water can use thermal energy to break them temporarily and then change bonding partners. This rebonding process allows water to change shape so that it is a fluid. Despite their flexibility, however, these bonds manage to bunch the water molecules together so snugly that even squeezing can't pack them much tighter. That's why water is incompressible.

The molecules in water can release still more chemical potential energy by linking together stiffly as ice. *Ice* is a solid, a rigid collection of chemically bound molecules. Like most solids, ice is **crystalline**—its water molecules are arranged in an orderly latticework that extends over long distances and gives rise to the beautiful crystal facets seen on snowflakes and frost. Ice's crystalline structure is so constraining that its water molecules can't use thermal energy to change bonding partners, and consequently ice can't change shape.

Just as an orderly stack of oranges at the grocery store takes up less volume than a disorderly heap, a crystalline solid almost always occupies less volume than its corresponding disorderly liquid. The solid phase of a typical substance is thus denser than the liquid phase of that same substance, so the solid phase sinks in the liquid phase.

There is only one common substance that violates that rule—water. Ice's crystalline structure is unusually open, and its density is surprisingly low. Almost unique in nature, solid ice is slightly less dense than liquid water, and ice therefore floats on water. That's why icebergs float on the open ocean and why ice cubes float in your drink. In fact, water reaches its greatest density when it is liquid at about 4 °C (39 °F).

► Check Your Understanding #1: Skating Anyone?

As a pond freezes, where does the ice begin to form?

Answer: It begins to form at the surface of the pond.

Why: Because ice is less dense than water, it floats at the surface of the pond. The pond therefore freezes from its top down. The insulating layer of ice at the top of the pond impedes the freezing of the remainder of the water, so the pond is unlikely to freeze solid. This behavior allows animals to live in the pond during winter. If ice sank, the pond would freeze solid and the animals would die. Thus, ice's tendency to float has profound implications for life on our planet.

Melting Ice and Freezing Water

Ice in a freezer is extremely cold, typically –18 °C (0 °F). When you place this ice on a warm countertop, heat flows into it and its temperature rises. The ice remains solid until its temperature reaches 0 °C (32 °F). At that point, the ice stops getting warmer and begins to melt. **Melting** is a **phase transition**, a transformation from the ordered solid phase to the disordered liquid phase. This transition occurs when heat breaks some of the chemical bonds between water molecules and permits the molecules to move past one another. The melting ice transforms into water, losing its rigid shape and crystalline structure.

Zero degrees Celsius is ice's **melting temperature**, the temperature at which heat goes into breaking bonds and converting ice into water, rather than into making the ice hotter. The ice-water mixture remains at 0 °C until all the ice has melted. When only water remains, heat once again causes its temperature to rise.

The heat used to transform a certain mass of solid into liquid, without changing its temperature, is called the **latent heat of melting** or, more formally, **latent heat of fusion**. The bonds between the water molecules in ice are strong enough to give ice an enormous latent heat of melting; it takes about 333,000 J of heat to convert 1 kg of ice at 0 °C into 1 kg of water at 0 °C. Since water's specific heat is 4190 J/kg · K, that same amount of heat would raise the temperature of 1 kg of liquid water by about 80 °C. Thus it takes almost as much heat to melt an ice cube as it does to warm the resulting water all the way to boiling.

The latent heat of melting reappears when you cool the water back to its melting temperature and it starts to freeze. **Freezing** is another phase transition, a transformation from the disordered liquid phase to the ordered solid phase. As you remove heat from water at 0 °C, the water freezes into ice rather than becoming colder. Because the water molecules release energy as they bind together to form ice crystals, the water releases heat as it freezes. The heat released when transforming a certain mass of liquid into solid, without changing its temperature, is again the latent heat of melting. You must add a certain amount of heat to ice to melt it and you must remove that same amount of heat from water to solidify it.

Check Your Understanding #2: Keeping Crops Warm

Fruit growers often spray their crops with water to protect them from freezing in unusually cold weather. How does liquid water keep the fruit from freezing?

Answer: Liquid water releases a large amount of heat as it solidifies, and this heat helps to protect the fruit from freezing.

Why: Fruit freezes at a temperature somewhat below 0 °C. As cold air removes heat from the water-coated fruit, the water begins to freeze. The liquid water gives up a great deal of heat as it solidifies and prevents the fruit's temperature from dropping below 0 °C until the water has completely turned to ice. At that point, ice's poor thermal conductivity insulates the fruit and slows its loss of heat to the colder air.

Phase Equilibrium: Leaving and Landing

We've seen that ice has a melting temperature, so now let's see *why* it has a melting temperature. To do that, we must look at the interface between solid ice and liquid water. Whenever both phases are in contact, they exchange water molecules across the interface between them. Water molecules regularly break free from the ice to enter the water, and they often drop out of the water to stick to the ice. In other words, water molecules are leaving the ice and landing on it all the time, like airplanes at a busy airport.

Although you can't see the individual leavings and landings, you can observe their net effect. If leavings outpace landings, the ice will gradually transform into water. If the landings outpace leavings, the water will gradually transform into ice. If the two processes balance one another, the ice and water will coexist indefinitely—a situation known as **phase equilibrium**.

Temperature plays a crucial role in this balance because it affects the rate at which water molecules leave the ice. The warmer the ice, the more often water molecules at its surface can gather enough thermal energy to break free and leave. Below ice's melting temperature, water molecules leave the ice too infrequently to balance the landing process and the water transforms completely into ice. Above ice's melting temperature, water molecules leave the ice so often that they outstrip the landing process and the ice transforms completely into water. Only at the melting temperature do the leaving and landing rates balance so that ice and water can coexist in phase equilibrium.

Ice's huge latent heat of melting has a stabilizing effect on the phase equilibrium between ice and water. Whenever ice and water are mixed together, the mixture's temperature will shift rapidly toward 0 °C. That's because if the mixture's temperature is above 0 °C, the ice will melt—a phase transformation that absorbs the heat of melting and thereby

lowers the mixture's temperature toward 0 °C. If the mixture's temperature is below 0 °C, the water will freeze—a phase transformation that releases the heat of melting and thereby raises the mixture's temperature toward 0 °C.

As long as the mixture doesn't run out of ice or water, its temperature will soon reach 0 °C and remain there even as you add or remove heat from it. Any heat you add to the mixture goes into melting more ice, not into raising its temperature. Any heat you remove from the mixture comes from freezing more water, not from lowering its temperature. That's why your glass of ice water remains at 0 °C, even in the hottest or coldest weather (Fig. 7.2.2).

Check Your Understanding #3: Ice Water

At a restaurant, your glass of water contains 25% ice while your friend's glass of water contains 75% ice. Whose water is colder, or are their temperatures equal?

Answer: Their temperatures are equal.

Why: Both glasses contain mixtures of ice and water at 0 °C (32 °F). If either mixture were colder than that, its water would be freezing rapidly into ice, and if either mixture were warmer than that, its ice would be melting rapidly into water.

Courtesy Lou Bloomfield



Fig. 7.2.2 Ice and water can coexist only at 0 °C, ice's melting temperature.

Evaporating Water and Condensing Steam

Water's open surface is another active interface between phases, but this time the liquid water is exchanging molecules with gaseous steam. Water molecules are feverishly leaving the water for the steam and landing on the water from the steam, once again like planes at a busy airport.

While this frantic exchange of molecules is intriguing, what matters most is its net effect. If more molecules leave the water than return to it, then water gradually evaporates into steam. **Evaporation** is the phase transition from liquid to gas. On the other hand, if more molecules are landing on the water than leaving it, the steam gradually condenses into water. **Condensation** is the phase transition from gas to liquid. If landing and leaving are balanced, water and steam coexist in phase equilibrium.

These two phase transitions have enormous thermal consequences. Since the molecules in water cling to one another with chemical bonds, it takes energy to separate them. Although the bonds *between* water molecules are weaker than the bonds *within* water molecules, it still takes a great deal of energy to transform water into steam.

The heat needed to transform a certain mass of liquid into gas, without changing its temperature, is called the **latent heat of evaporation** or, more formally, **latent heat of vaporization**. Water's latent heat of evaporation is truly enormous because water molecules are surprisingly hard to separate. About 2,300,000 J of heat is needed to convert 1 kg of water at 100 °C into 1 kg of steam at 100 °C. That same amount of heat would raise the temperature of 1 kg of water by more than 500 °C!

You are most aware of this latent heat of evaporation on hot summer days when the perspiration that evaporates from your skin draws heat from you and lowers your temperature. As each water molecule leaves your skin to become steam, it gathers up more than its fair share of thermal energy from its environment and carries it away as chemical potential energy. The departing water molecules thus leave you bereft of thermal energy, and you grow colder.

The latent heat of evaporation reappears when steam condenses and the gathering water molecules release their chemical potential energy as heat. The heat released when a certain mass of gas is transformed into liquid, without changing its temperature, is again the latent heat of evaporation. You must add a certain amount of heat to water to evaporate it, and you must remove that same amount of heat from steam to condense it.

The huge amount of heat released by condensing steam is often used to cook food or warm radiators in older buildings. When you steam vegetables, you are allowing steam to condense on the vegetables and transfer heat to them. A double-boiler uses condensing steam to transfer heat from a burner to a cooking container in a controlled manner.

Check Your Understanding #4: Tea Time

A kettle of water heats up rapidly on the stove but takes quite a while to boil away. Why does the water take so long to turn into steam?

Answer: The stove must operate for a long time to provide water's huge latent heat of evaporation.

Why: Converting a kettle of hot water into steam requires an enormous amount of heat. This heat separates the molecules without raising their temperature. Since the stove provides only a certain amount of heat each second, it may take many minutes to boil away the water.

Relative Humidity

Although we've examined the consequences of evaporating and condensing, we haven't yet seen what conditions determine when they'll occur. It all comes down to water molecules leaving and landing, so let's take a look at why one process wins out over the other.

The basic indicator of whether water will evaporate or steam will condense is relative humidity. **Relative humidity** measures the water molecules' landing rate as a percentage of the leaving rate. When the relative humidity is 100%, the two rates are equal and water and steam can coexist—they are in phase equilibrium. When the relative humidity is less than 100%, the landing rate is less than the leaving rate and the water evaporates—only steam will remain at equilibrium. Finally, when the relative humidity is more than 100%, the landing rate is more than the leaving rate and the steam condenses—no equilibrium is possible until the relative humidity has dropped to 100%.

Relative humidity depends on the temperature and on the density of the steam. Temperature affects the leaving rate. The warmer the water, the more thermal energy it contains and the more frequently water molecules leave its surface to become gas. By itself, an increase in temperature will boost the leaving rate and thereby decrease the relative humidity. Rising temperatures thus favor evaporation.

The density of water molecules in the steam affects the landing rate. The denser the steam, the more often water molecules land on the water to become liquid. By itself, an increase in steam density will boost the landing rate and thereby increase the relative humidity. Rising steam densities thus favor condensation. Even when that steam is mixed with air, as it often is, the air molecules act as passive bystanders. The density of water molecules alone determines the air's relative humidity.

Relative humidity plays an important role in countless experiences of everyday life. When the relative humidity is low, water evaporates quickly and the air feels dry. Perspiration cools you effectively. When the relative humidity is high (near 100%), water barely evaporates at all and the air feels damp. Perspiration clings to your skin and doesn't cool you much.

When the relative humidity exceeds 100%, perhaps because of a sudden temperature drop, steam begins to condense everywhere. Water droplets grow on surfaces as dew or form directly in the air as fog, mist, or clouds. If the humidity remains high, these droplets grow larger and eventually fall as rain.

You can measure relative humidity by observing the cooling that accompanies evaporation. The most common scheme involves two thermometers (Fig. 7.2.3), one of which has a wet cloth wrapped around its bulb (right). Water evaporating from the wet cloth cools the thermometer until that water reaches phase equilibrium with steam in the air and evaporation ceases. The dryer the air, the colder the thermometer gets. The temperatures of the two thermometers can then be used to determine the relative humidity, usually with the help of tabulated values.

Courtesy Lou Bloomfield



Fig. 7.2.3 You can determine the air's relative humidity using two thermometers, one of which has a wet cloth wrapped around its bulb (right). Evaporation cools the wet-bulb thermometer by an amount related to the air's relative humidity. The drier the air, the colder the wet-bulb thermometer becomes.

Check Your Understanding #5: Seeing Your Breath

When you breathe out on a cold day, you often see mist appearing from your mouth. Explain.

Answer: As the warm, moist air from your mouth cools, its relative humidity increases above 100%. The water vapor condenses to form a mist of water droplets.

Why: Changes in temperature affect relative humidity. Although the warm air leaving your lungs can contain a large portion of water vapor, its relative humidity increases dramatically as it cools. At lower temperatures, condensation takes over and the invisible water vapor condenses into a mist of visible water droplets.

Subliming Ice and Depositing Steam

We've examined phase transitions between ice and water and between water and steam. That brings us to the phase transitions between ice and steam. Oddly enough, water molecules can leave ice to become steam and can land from steam to become ice. In fact, ice and steam regularly exchange water molecules even when there is no liquid water present at all.

As usual, this exchange of water molecules occurs at the surface of the ice, the interface between ice and steam. Since we're most interested in the net movement of molecules, it comes down to leaving and landing rates on the ice. If molecules leave the ice more often than they land, the ice sublimes. **Sublimation** is the phase transition from solid to gas. If molecules land on the ice more often than they leave, the steam deposits. **Deposition** is the phase transition from gas to solid. Once again, relative humidity measures the landing rate as a percentage of the leaving rate. At 100% relative humidity, ice and steam are in phase equilibrium.

When the relative humidity is below 100%, ice sublimes. This effect gives rise to a number of familiar phenomena. When the weather is cold and dry, snow gradually disappears from the ground without ever melting. In the low relative humidity of a frostless freezer, the ice cubes slowly shrink to miniature size. When you leave food unprotected in that same frostless freezer, it eventually dries out. While this "freezer burn" is a nuisance at home, sublimation from frozen food is used commercially to prepare freeze-dried foods.

When the relative humidity exceeds 100%, steam deposits. This process yields several other familiar effects. Frost forms on cold windows and lawns that are exposed to humid air. In the high relative humidity of a nonfrostless freezer, frost and ice accumulate on the walls and require periodic defrosting. Snowflakes grow in clouds and then descend gracefully to the ground.

Check Your Understanding #6: Disappearing Stink

Mothballs are smelly, crystalline pellets used to protect woolen clothes from attack by insects. If you sprinkle them into a closet and wait a few years, the mothballs will disappear. How do they vanish?

Answer: The mothballs sublime.

Why: Although mothballs do not melt at room temperature, they sublime readily. The air around them quickly becomes saturated with gaseous mothball molecules, and that's why it smells so strongly.

Boiling Water

With three phases and six phase transitions, we seem to be out of possibilities. Where does boiling fit into this picture? Boiling is simply an accelerated form of evaporation in which bubbles of pure steam grow by evaporation inside the water itself. To understand boiling, let's take a look at the interplay between water and steam.

Suppose we seal some water inside an airless container and keep it at a constant temperature. The water will evaporate as steam until the relative humidity inside the container

Courtesy Lou Bloomfield



Fig. 7.2.4 Water boils at 100 °C, when atmospheric pressure can no longer smash the bubbles of water vapor.

reaches 100%. At that point, the water and steam will have reached phase equilibrium; the steam will have just the right density so that its water molecules will land on the water as often as they leave. Steam at its equilibrium density is said to be **saturated**.

Of course, the density of that saturated steam depends on the temperature of the container. If you warm up the container, water molecules will leave the water more frequently and the density of the steam will have to increase to match the landing rate to the leaving rate. The density of saturated steam is thus an increasing function of the temperature.

The saturated steam's density, together with its temperature, determines its pressure—the pressure inside our container. Near room temperature, the pressure of saturated steam is only a few percent of atmospheric pressure. It increases with temperature, however, so that, as the temperature approaches 100 °C (212 °F), the pressure of saturated steam approaches atmospheric pressure.

With that background, suppose that we place a bubble of pure saturated steam in room temperature water that's at atmospheric pressure. Because the pressure inside that bubble is much lower than atmospheric pressure, the surrounding water will rush inward and compress it. As the steam bubble's volume shrinks, the steam will exceed its saturated density and begin to condense. In the blink of an eye, the bubble will be smashed out of existence.

Now suppose we begin warming the water on the stove. As the water temperature increases, steam's saturated density and pressure both increase. At first, saturated steam bubbles remain unstable; they're crushed quickly by atmospheric pressure. However, when the water temperature is near 100 °C, something remarkable happens: saturated steam bubbles suddenly become stable and the water can begin **boiling** (Fig. 7.2.4). At that temperature, water's **boiling temperature**, the pressure of saturated steam reaches atmospheric pressure and bubbles of saturated steam can survive indefinitely within the water. Even more remarkably, these bubbles can grow by evaporation; each bubble's surface is an interface between water and steam, so when heat is added to the water, water can transform into steam and enlarge the bubble. Although the bubbles quickly float to the water's surface and pop, new bubbles can promptly take their place.

Boiling converts water to steam so rapidly that it consumes almost any amount of heat you add to the water. That's why it's so difficult to warm water above its boiling temperature. An open pot of water on the stove warms to water's boiling temperature and then remains at that temperature until all the water has transformed into steam. Only then can the pot's temperature again begin to increase.

The constant, well-defined temperature of boiling water allows you to cook vegetables or an egg at a particular rate. When you place an egg in boiling water, it cooks in 3 minutes because it's in contact with water at its boiling temperature.

► Check Your Understanding #7: Vanishing Bubbles

You are heating water in a pot on the stove. Shortly before the water boils properly, bubbles of steam begin rising from the bottom of the pot but vanish before they reach the water's surface. What happens to those bubbles?

Answer: The rising steam bubbles condense as they pass through water that is below its boiling temperature.

Why: Water at the bottom of the pot reaches its boiling temperature first, so bubbles of steam begin to rise from it. But water nearer the surface of the pot isn't quite hot enough to sustain those bubbles. Instead, the steam inside them condenses and they collapse completely.

Changing Water's Boiling Temperature

Water's boiling temperature depends on the ambient pressure. For an open pot or pan, that pressure is atmospheric pressure. However, atmospheric pressure decreases with altitude and depends somewhat on the weather. Water boils at 100 °C (212 °F) near sea level but at

only 90 °C (194 °F) at an altitude of 3000 m (9800 ft). This reduction in water's boiling temperature with altitude explains why many recipes must be adapted for use at higher elevations. At 3000 m, an egg cooks slowly in boiling water because it's surrounded by 90 °C water, not 100 °C water. The same problem slows the cooking of rice, beans, and many other foods at high altitudes.

At sufficiently low pressures, water boils even at room temperature or below. At the other extreme, high pressures can prevent water from boiling until it's extremely hot. The boilers in steam engines and power plants often operate at such high pressures that water's boiling temperature inside them may exceed 300 °C (572 °F).

One way to decrease cooking times is to use a pressure cooker, a sturdy pot that seals in steam so that the pressure inside it can exceed atmospheric pressure. This increased pressure prevents boiling until the water temperature is well above 100 °C. If you subject water to twice sea-level atmospheric pressure, it won't boil until 121 °C (250 °F). An egg cooks very quickly at that temperature, as do vegetables and other foods.

However, just because water *can* boil, doesn't mean that it *will* boil. Boiling depends on tiny seed bubbles that subsequently grow by evaporation. Without seed bubbles, the water won't boil. Because **nucleation**, or seed-bubble formation, almost never occurs spontaneously in water below 300 °C, something else must create those seeds. Most nucleation occurs at defects or hotspots on the container or at contaminants in the liquid. Those nucleation sites usually trap air or other permanent gases and then serve as nurseries for seed bubbles of steam. Reliance on these specific nucleation sites explains why the-bubbles of boiling water, like those in soda or champagne, often stream upward from-specific spots on their containers.

When you heat water uniformly in a clean, glass container, it may not boil properly at its boiling temperature. Glass has a liquid-like smoothness even at the atomic scale and rarely aids seed-bubble formation. Without any long-lived nucleation sites, the water may stop forming seed bubbles and cease boiling. Once boiling stops, the water's temperature can rise above the boiling temperature so that it becomes **superheated**.

Superheated water, which forms easily and often in a microwave oven, can be extremely dangerous. Touching it with a fork, adding sugar or salt, or even just tapping its container can initiate violent or even explosive boiling (Fig. 7.2.5). The more the water's temperature exceeds its boiling temperature, the more energy it can release suddenly if it abruptly boils. Be careful when you heat water in a microwave oven, particularly in a glass or glazed container. If it doesn't appear to be boiling properly despite being very hot, recognize that it may be superheated. Your safest bet is to stay away from it until it has cooled down.

There is one other interesting way to change water's boiling temperature: dissolve chemicals in it. A dissolved chemical keeps the water molecules busy so that they are less likely to leave the water to become steam, or ice for that matter. Since dissolved chemicals discourage water molecules from leaving water's liquid phase, they suppress any phase transitions that reduce the amount of liquid phase water. That's why dissolving sugar or salt in water slows its evaporation and raises its boiling temperature. It's also why salt water freezes at a lower temperature than freshwater and why salt tends to melt ice. In contrast, sand doesn't melt ice because it doesn't dissolve in water.

Courtesy Lou Bloomfield



Fig. 7.2.5 The water in this glass measuring cup was superheated in a microwave oven and then disturbed with a fork. Explosive boiling blew all the water out of the cup in a fraction of second.

Check Your Understanding #8: Lucky Accident

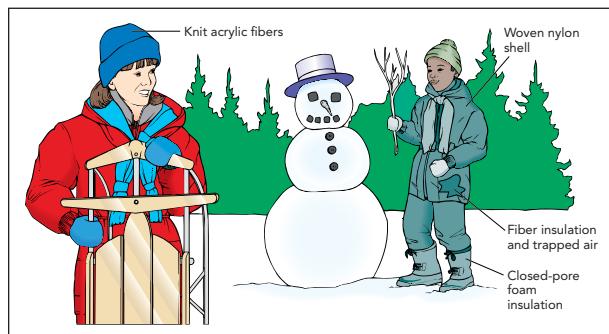
You pull a ceramic mug of hot water out of the microwave oven and add a spoonful of instant coffee. It bubbles amusingly and splatters a little onto the table. What happened?

Answer: The water in your mug was mildly superheated, and adding the powdered coffee nucleated boiling.

Why: Like glass, a glazed ceramic mug has no permanent nucleation sites to assist seed-bubble formation and boiling. You're lucky. The water was only slightly superheated, and when you triggered boiling by adding the powder, it boiled relatively gently and quickly cooled to its normal boiling temperature. Had the water been more severely superheated, you might have been badly scalded.

SECTION 7.3

Clothing, Insulation, and Climate



When you sit in front of a fire on a cold winter day, heat flows from the glowing coals to your skin and you feel warm. When you walk through the snow on your way to the store, however, heat flow is the last thing you want. As the hottest object around, you will become colder, not warmer. So you bundle up in your new down coat. Its thermal insulation keeps you warm in your frigid environment. In this section, we'll examine thermal insulation and see how it prevents heat from moving easily between objects.

Questions to Think About: When you wear a thick coat on a cold day, what provides the heat that keeps you warm? How does your coat keep that heat from leaving you? Your head is exposed, so what keeps your head warm? Why is good cookware

often made of several different materials? Why does wrapping food in aluminum foil help keep it hot or cold? Why do most modern windows have more than one pane of glass in them? Why does increasing the fraction of greenhouse gases in the atmosphere raise the average temperature at Earth's surface?

Experiments to Do: Examine the effects of thermal insulation by touching objects while your hand has or does not have insulation on it. As you pick up a piece of hot toast with your bare hand, why does your skin feel so hot? Now try again with a towel or napkin between your skin and the toast. What has changed? Perform the same experiment with an ice cube. The towel or napkin keeps your skin comfortable when you touch either hot or cold objects. How is that possible? What happens when you use aluminum foil instead of the towel or napkin?

You can do similar experiments with a heavy coat. The coat will obviously keep you warm in a cold environment, but it will also keep you cool (at least briefly) in a hot environment. Sit in front of a fireplace with a heavy coat on. The parts of your body that are covered by the coat will barely notice the fire's presence. Now touch the exposed surface of the coat carefully with your fingers. Why is it so hot? As you can see, you need to be careful when wearing insulated clothes near a fire because it's possible to scorch or even ignite those clothes without feeling the heat. Firefighters use heavy fireproof coats to keep cool as they combat building fires.

The Importance of Insulation

Thermal insulation slows the flow of heat from hot objects to cold and helps keep your body warm, your refrigerator cold, and your home at a comfortable temperature year round. Yet in a world where ordered energy is increasingly precious, good insulation is more than just a convenience; it's a necessity. That's because whenever thermal energy leaves something that must remain hot or enters something that must remain cold, replacing or removing that thermal energy consumes ordered energy. You can save ordered energy and do your part for the environment simply by using good insulation to keep thermal energy where it belongs.

One of the most important examples of thermal insulation is your clothing. Together with your skin, clothing lowers the rate at which heat flows into or out of your body and makes it possible for you to keep your body temperature constant.

Why keep your body temperature constant? After all, cold-blooded animals such as reptiles, amphibians, and fish make no attempts to control their body temperatures. Instead, they exchange heat freely with their surroundings and are generally in thermal equilibrium with their environments. Maintaining a constant body temperature is a goal that's unique to mammals and birds. Why bother?

The chemical processes that are responsible for life are very sensitive to temperature. That sensitivity is due in part to thermal energy's role in initiating chemical reactions; it provides the activation energies that many chemical reactions need in order to proceed. As a cold-blooded animal's temperature decreases, there is less thermal energy per molecule

and these chemical reactions occur less frequently. The animal's whole metabolism slows down, and it becomes sluggish, dimwitted, and vulnerable to predators.

In contrast, warm-blooded animals have temperature regulation systems that allow them to maintain constant, optimal body temperatures. Regardless of its environment, a mammal or bird keeps the core of its body at a specific temperature so that it functions the same way in winter as in summer. The advantages of constant temperature are enormous. On a cold day, a warm-blooded predator can easily catch and devour its slower-moving cold-blooded prey.

There is, however, a cost to being warm-blooded. The thermal energy associated with an animal's body temperature must come from somewhere, and the animal must struggle against its environment to maintain its temperature. Without realizing it, many of our behaviors are governed by our need to maintain body temperature. Our bodies are careful about how much thermal energy they produce, and we strive to control the rate at which we exchange heat with our surroundings.

A resting person converts chemical potential energy into thermal energy at the rate of about 80 calories per hour. Eighty calories per hour is a measure of power, equal to about 100 W, so a resting person generates about 100 W of thermal power. An active person generates even more. This steady production of thermal energy is why a room filled with people can get pretty warm. One hundred watts may not seem like much, but when a hundred people are packed into a tight space, they act like a 10,000-W space heater and the whole room becomes unpleasantly hot.

If you had no way to get rid of this thermal energy of metabolism, you would become hotter and hotter. To maintain a constant temperature, you must transfer heat to your surroundings. Since heat flows naturally from a hotter object to a colder object, your body temperature must generally be higher than the temperature around you. That requirement is one reason why human body temperature is approximately 37 °C (98.6 °F). This temperature is higher than all but the hottest locations on Earth, so heat flows naturally from your body to your surroundings. You produce thermal energy as a by-product of your activities and transfer this thermal energy as heat to your colder surroundings.

The rate at which your resting body generates thermal energy is fairly constant, so the principal way in which you stabilize your body temperature is by controlling heat loss. Like all warm-blooded animals, you have a number of physiological and behavioral techniques for doing just that. In hot weather or when you're exercising hard, you encourage heat loss to avoid overheating. In cold weather, you limit heat loss to stay warm. Since this section is about clothing, we're going to focus primarily on heat loss in cold weather. All three heat-transfer mechanisms are involved in that heat loss, so you must control them all to keep warm.

Check Your Understanding #1: Oven as Space Heater

A toaster oven turns electric energy into thermal energy at a rate of 500 J/s, or 500 W. If the appliance's temperature remains constant, at what rate is it transferring heat to its environment?

Answer: It's transferring heat at the rate of 500 W.

Why: For the oven to maintain a constant temperature, its total thermal energy mustn't change. Since it produces 500 W of thermal power electrically, it must transfer that same amount of thermal power to its environment. It acts as a 500-W space heater.

Staying Warm by Limiting Thermal Conduction

One way in which your body retains heat in cold weather is by impeding conductive heat loss. Heat flows through a material via conduction whenever that material is exposed to a difference in temperature—whenever one side of the material is hotter than the other side. The rate at which heat flows, however, depends on several factors: the overall temperature

difference, the distance separating the hot side from the cold side, the area through which heat can flow, and the material itself.

Remarkably, heat flows in direct proportion to the temperature difference and to the area through which that heat can flow. Thus, if you touch your finger to two different cold surfaces, one 20 °C colder than your internal temperature and the other 40 °C colder, you'll lose heat about twice as fast to the colder surface. If you touch two fingers at once, doubling the area through which heat can flow, you'll lose heat about twice as fast.

On the other hand, further separating the hot and cold sides slows heat flow. The rate of heat flow is inversely proportional to that distance. For example, doubling the thickness of your gloves when you're building a snowman will double the separation between hot and cold and roughly halve the heat flow.

Finally, some materials are better conductors of heat than others; they have different **thermal conductivities**. Skin has a particularly low thermal conductivity, meaning that it conducts relatively little heat compared to materials such as glass or copper. Thermal conductivity is a characteristic of the material itself, and its SI unit is the watt per meter-kelvin (W/m · K). Table 7.3.1 shows the thermal conductivities of a number of common materials.

The overall relationship between heat flow and these four quantities can be written as a word equation:

$$\text{heat flow} = \frac{\text{thermal conductivity} \cdot \text{temperature difference} \cdot \text{area}}{\text{separation}}, \quad (7.3.1)$$

in symbols:

$$H = \frac{k \cdot \Delta T \cdot A}{d},$$

and in everyday language:

When you pick up a hot saucepan, wear a thick oven mitt and grab a small portion of the cooler handle.

TABLE 7.3.1 Approximate Thermal Conductivities of Various Materials

Material	Thermal Conductivity [†]
Argon	0.016 W/m · K
Air	0.025 W/m · K
Oak	0.17 W/m · K
Fat	0.2 W/m · K
Skin	0.21 W/m · K
Hair	0.37 W/m · K
Brick	0.6 W/m · K
Water	0.6 W/m · K
Glass	0.8 W/m · K
Marble	2.5 W/m · K
Stainless steel	16 W/m · K
Steel	50 W/m · K
Aluminum	210 W/m · K
Copper	380 W/m · K
Silver	429 W/m · K
Diamond	1000 W/m · K

[†]The watt per meter-kelvin (abbreviated W/m · K) is the SI unit of thermal conductivity.

Equation 7.3.1 tells us that the amount of heat flowing through your skin depends on its thermal conductivity, surface area, thickness, and the temperature difference across it. Amazingly, your body has optimized these factors to help you minimize heat loss. Even without clothes, you are surprisingly well insulated against conductive heat loss.

Your skin and the layers immediately beneath it contain fats and other thermal insulators. Fat's thermal conductivity is quite low, about a third that of water, so by placing fat in and just beneath your skin, your body improves its heat retention. Furthermore, the presence of a fatty layer beneath your skin effectively thickens your skin and increases the distance separating hot from cold. "Thick-skinned" people retain body heat better than those who are "thin-skinned."

Minimizing surface area means that your body is relatively compact, shaped more like a ball than a sheet. Although other adaptive influences have given you arms, legs, and fingers, which increase your total surface area, you have little superfluous surface through which to lose heat.

Finally, your body tries to lessen conductive heat loss by reducing the temperature difference between your skin and the surrounding air. It does this by letting your hands and feet become much cooler than your core body temperature. That arrangement would be simple were it not for your circulating blood. Your blood must cool down from core body temperature as it approaches your cold fingers and warm back up to core body temperature as it returns to your heart. This change in blood temperature occurs via a mechanism called countercurrent exchange. As the warm blood flows through arteries toward your cold fingers, it transfers heat to the blood returning to your heart through nearby veins (Fig. 7.3.1). The blood heading toward your fingers becomes colder, while the blood returning to your heart becomes warmer.

Check Your Understanding #2: Hot Potatoes

You have been cooking potatoes in a pot of boiling water, and it's time to fish them out. With which are you least likely to burn your hands: copper tongs or stainless steel tongs?

Answer: Stainless steel tongs are less likely to burn you.

Why: The rate at which heat flows from the hotter water to your colder hand is proportional to the tongs' thermal conductivity, and stainless steel has a much lower thermal conductivity than copper.

Check Your Figures #1: Cold Hard Brick

You are visiting a tiny fifteenth-century house that was built entirely of bricks. Its walls have a surface area of 20 m^2 (215 ft^2), and the bricks are 9.2 cm ($3\frac{5}{8} \text{ in}$) thick. When the winter air outside is -20°C and the heated air inside is 20°C , at what rate is heat flowing through the walls of the house?

Answer: 5200 W.

Why: According to Table 7.3.1, the thermal conductivity of brick is $0.6 \text{ W/m} \cdot \text{K}$. The temperature difference between the inside and outside of each brick is 40°C , or, equivalently, 40 K . From Eq. 7.3.1, we obtain:

$$\text{heat flow} = \frac{0.6 \text{ W/m} \cdot \text{K} \cdot 40 \text{ K} \cdot 20 \text{ m}^2}{0.092 \text{ m}} = 5217 \text{ W.}$$

Because of the limited accuracy of our starting values, we round the result to 5200 W. That is an enormous rate of heat loss, and we haven't even considered losses through the roof. This house needs insulation!

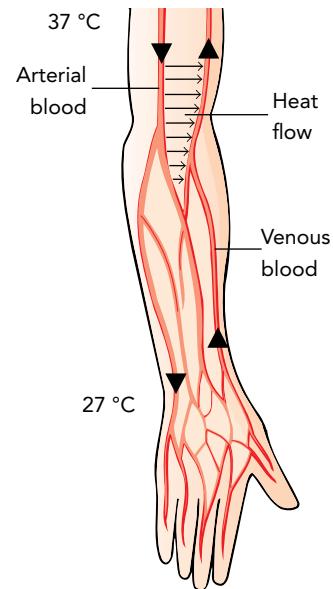


Fig. 7.3.1 Blood flowing toward your hand through arteries exchanges heat with blood returning to your heart through veins. In this fashion, your blood is able to carry oxygen and food to your fingers without warming them all the way up to core body temperature. This adaptation reduces the rate at which you lose heat in cold weather.

Staying Warm by Impeding Convection

Despite its great insulating qualities, skin alone isn't enough to keep you warm in all situations. Some of your body heat inevitably flows to the surface of your skin and from there to your environment. Limiting that second step of heat flow is a job for clothing.

On a cold day, heat leaving your skin warms the nearby air. Since air is a very poor conductor of heat, however, your skin warms only a thin layer of it. If that warmed air never moved, you would have to heat it only once and that would be that. Protected by this insulating layer of air, your skin would no longer have to transport much heat and the temperature difference across your skin would shrink. You would feel warm.

But air moves all too easily. In reality, each time your skin manages to warm the nearby air, natural convection gently lifts that warmed air upward and replaces it with fresh cold air. The temperature difference across your skin thus remains large, and heat flows quickly out through your skin. You feel cold.

Wind worsens this heat loss because it blows away warmed air near your skin even faster than does natural convection. Just as a forced convection oven cooks food faster, so a forced convection freezer (such as a cold, windy day) chills people faster. The enhanced heat loss caused by moving air is called wind chill—you feel even colder on a windy day.

To combat convective heat loss and wind chill, most warm-blooded animals are covered with hair or feathers. These finely structured materials are poor conductors of heat, but their main purpose is actually to trap an even better insulator: air. In the dense tangle of a sheep's wool, air experiences such severe drag forces that it can barely move at all. Since convection requires airflow, the sheep can lose heat only via conduction through the trapped air and wool. That combination of materials is a terrible conductor of heat, so the sheep stays warm.

We humans have relatively little hair and are thus poorly adapted to living in cold, windy climates. Our lack of natural insulation is one of the reasons we wear clothing. Like hair and feathers, our clothing traps the air and reduces convection. Finely divided strands or filaments are particularly effective at stopping the flow of air. Not surprisingly, the best insulating clothing is made of hair (natural or synthetic) and feathers (also natural or synthetic). Since motionless air has a lower thermal conductivity than any of the materials that trap it, the ideal coat uses only enough material to keep a thick layer of air from moving.

This discussion also applies to swimming in cold water. If the water near your skin would stay put, you could warm up a layer of it and then feel relatively warm. That's why some swimmers wear wet suits. The spongy material in a wet suit keeps the layer of water near the swimmer's skin from moving. As long it remains in place, water is a respectable thermal insulator.

Check Your Understanding #3: When a Cold Wind Blows

Why does wearing a thin nylon windbreaker make such a difference in your ability to keep warm on a cool, windy day?

Answer: It traps a layer of insulating air near your skin.

Why: Even though the windbreaker is too thin to provide much insulation itself, it traps air near your skin. With much less circulation of cold air across your skin, you lose heat more slowly and feel relatively warm.

More about Thermal Radiation

When you sit in front of a glowing fire, you feel its thermal radiation warming your skin. But your skin also emits thermal radiation, so what you're really noticing is that your skin is absorbing more thermal radiation from the fire than it's emitting toward your surroundings. You're gaining heat. In contrast, if you sit inside an ice sculpture castle, your skin will absorb less thermal radiation from the ice than it's emitting, so you'll be losing heat and feeling cold.

Short of walking back and forth between fire and ice, how can you control how much heat you gain or lose via thermal radiation? More broadly, what determines how much thermal power you radiate toward your environment and how much you absorb from it? To answer these questions, we need to take a closer look at thermal radiation itself.

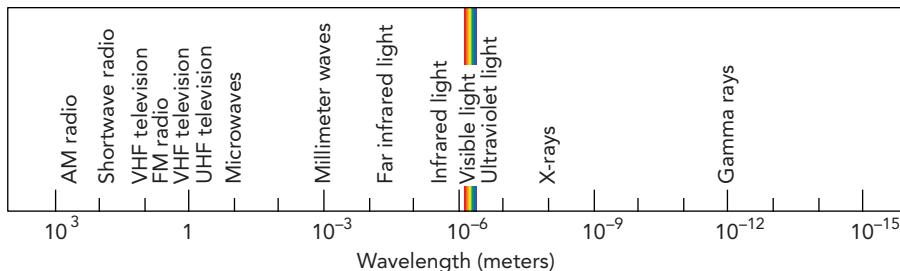


Fig. 7.3.2 The spectrum of electromagnetic radiation, arranged by wavelength. The scale here is logarithmic, meaning that the wavelength decreases by a factor of 10 with each tick mark to the right.

You've probably noticed that you can often judge the temperature of a hot object by the color of the light it emits. An object that's glowing red is too hot to touch, but one that's glowing yellow-white is a serious fire hazard. That connection between temperature and color isn't a coincidence. The light that you see emanating from a hot object is part of that object's thermal radiation. You are quite literally seeing its heat.

However, you're not seeing all of its heat. Although thermal radiation consists of electromagnetic waves and you can see some electromagnetic waves, most are invisible to your eyes. Those that you can see, **visible light**, are part of a continuous spectrum of electromagnetic radiation that extends from radio waves at one extreme to gamma rays at the other (Fig. 7.3.2).

Electromagnetic waves are distinguished by their **wavelengths**, that is, the distances between their wave crests. You can easily recognize the wavelength for the waves on a lake or sea, where the crests are visible and you can directly measure the distance from one to the next. Although the wave crests of electromagnetic waves aren't so easy to observe, they exist and it's possible to measure the distance between them. Moreover, within the visible portion of the **electromagnetic spectrum** (Fig. 7.3.3), you perceive different wavelengths of light as different colors. Light with a wavelength of 630 nanometers (billions of a meter, abbreviated nm) appears red, while light at 420 nm appears blue.

However, an object's thermal radiation isn't a single electromagnetic wave with one specific wavelength. Instead, it's many individual waves that cover a broad range of wavelengths. The distribution of those wavelengths depends on the object's temperature and surface properties, particularly its **emissivity**, the efficiency with which it emits and absorbs thermal radiation. Emissivity is measured on a scale from 0 to 1, with 1 being ideal efficiency. A perfectly black object has an emissivity of 1; it absorbs all light that strikes it and emits thermal light as efficiently as possible.

The distribution of wavelengths emitted by a black object is determined by its temperature alone and is called a **blackbody spectrum**. As shown in Fig. 7.3.4, that spectrum brightens and shifts toward shorter wavelengths as the object's temperature increases. When a black object is colder than about 400 °C, its thermal radiation lies entirely in the invisible long-wavelength portion of the spectrum and you can't see the radiation at all. Even at 400 °C, only your night vision is able to sense the thermal radiation, which appears dim and colorless. At about 500 °C, your color vision begins to detect the glow and your eyes make an average assessment of the distribution of wavelengths. You see dull red. With increasing temperature, you then observe reddish, orangish, yellowish, whitish, and eventually bluish

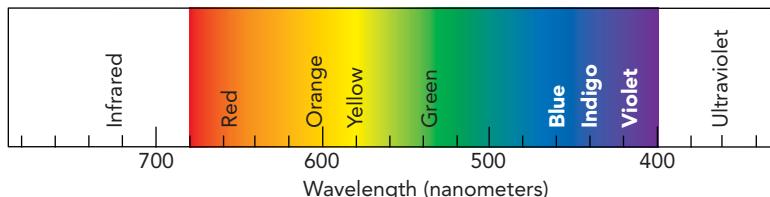


Fig. 7.3.3 A portion of the electromagnetic radiation spectrum around visible light. Wavelengths are measured in nanometers (nm, or billions of a meter).

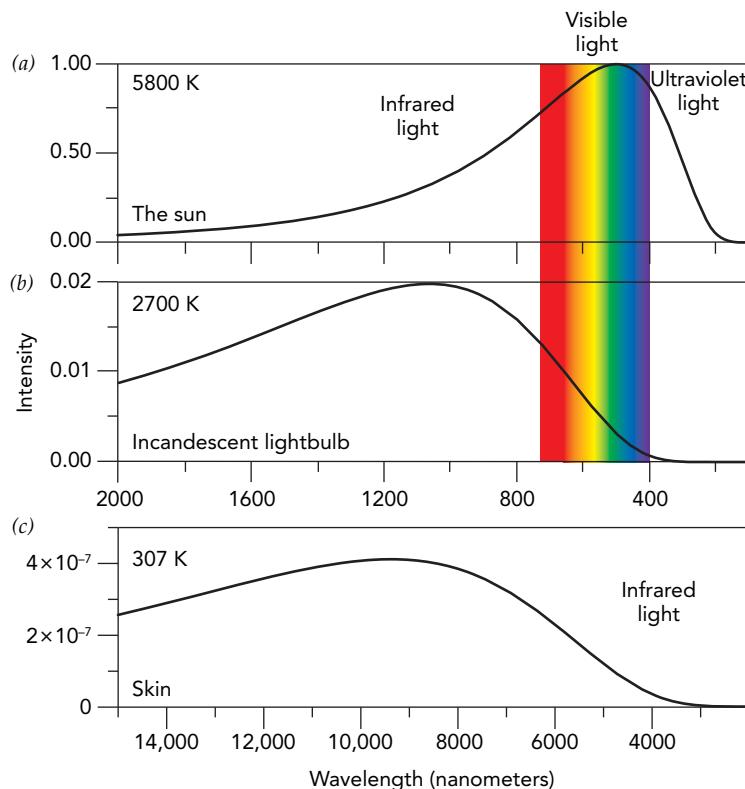


Fig. 7.3.4 The distributions of light emitted by black objects at (a) 5800 K, (b) 2700 K, and (c) 307 K. In addition to containing a larger fraction of visible light, the object at 5800 K is much brighter than the object at 2700 K (note the different intensity scales). The object at 307 K emits only dim, long-wavelength infrared light.

light (Table 7.3.2). The temperature of thermal radiation associated with a particular distribution of wavelengths is the **color temperature** of that light.

Whether visible or not, the entire spectrum of thermal radiation is important to retaining body heat. Your body radiates thermal power and if your radiated power is greater than the thermal power you absorb from your environment, you are losing heat via radiation.

The thermal power you radiate depends primarily on your temperature. It's clear from Fig. 7.3.4 that a black object radiates more power as its temperature increases. What may be surprising, however, is how dramatically that radiated power increases with temperature: it is proportional to the fourth power of the absolute temperature! Thus, although a black pot full of boiling water (373 K) and an identical black pot full of ice water (273 K) have an absolute temperature ratio of only 1.37, the hotter one radiates 1.37^4 , or 3.5, times as much thermal power as the cooler one.

TABLE 7.3.2 Temperatures and Colors of Light Emitted by Hot Objects

Object	Temperature	Color
Skin	34 °C (93 °F)	Invisible
Heat lamp	500 °C (930 °F)	Dull red
Candle flame	1700 °C (3100 °F)	Dim orange
Incandescent bulb	2500 °C (4500 °F)	Bright yellow-white
Sun's surface	5525 °C (9980 °F)	Brilliant white
Blue star	6000 °C (10,800 °F)	Dazzling blue-white

Your radiated power also depends on your surface area and on your emissivity, and it is proportional to each of those quantities. The precise relationship between temperature and radiated power can be written as a word equation:

$$\text{radiated power} = \text{emissivity} \cdot \text{Stefan-Boltzmann constant} \\ \cdot \text{temperature}^4 \cdot \text{surface area}, \quad (7.3.2)$$

in symbols:

$$P = e \cdot \sigma \cdot T^4 \cdot A,$$

and in everyday language:

Hot food radiates away lots of heat, so wrap it compactly in aluminum foil to keep it warm.

This relationship is called the **Stefan-Boltzmann law**, and the **Stefan-Boltzmann constant (σ)** that appears in it has a measured value of $5.67 \times 10^{-8} \text{ J/(s} \cdot \text{m}^2 \cdot \text{K}^4)$. Remember that the temperature must be measured in kelvins.

Your skin temperature and your surface area aren't hard to measure, but what is your emissivity for radiating heat? Prepare for a surprise. Looking at a surface tells you only about its high-temperature emissivity—how well it absorbs visible thermal radiation from a very hot object (such as the 5800 K sun) and would emit visible thermal radiation if it were heated red, yellow, or white hot. Your skin's temperature is only 307 K, however, so its thermal radiation depends on its low-temperature emissivity—how well it absorbs and emits invisible long-wavelength infrared light.

Now for the surprise: regardless of your appearance in visible light, your skin is almost perfectly black at the infrared wavelengths of skin-temperature thermal radiation! Your skin's low-temperature emissivity is approximately 0.97. In fact, most nonmetallic objects have low-temperature emissivities greater than 0.95. If you could see infrared light rather than visible light, all the people and most of the objects around you would look black—you would see that they absorb almost all the infrared light that strikes them and that they glow brightly with their own infrared thermal radiation.

Check Your Understanding #4: Seeing Your Outer Glow

Passive infrared sensors are commonly used to turn on lights when a person walks by. They sense changes in long-wavelength infrared light. How do they recognize a person in total darkness?

Answer: Both the person and his environment radiate infrared light, but because the person's temperature is different from his surroundings, his infrared brightness is also different.

Why: The sensor responds to the thermal radiation emitted by the objects in front of it. As the person walks by, the brightness and pattern of that thermal radiation change and the sensor switches on the light.

Check Your Figures #2: De-Lighted at the Blacksmith's Shop

When a blacksmith pulls a piece of yellow-hot steel out of the furnace, its temperature plummets and it dims rapidly. However, it remains too hot to touch for several minutes. How much faster does the steel cool by radiation at 800 °C than it does at 400 °C? (Assume the steel's emissivity doesn't depend on temperature.)

Answer: The steel cools about 6.46 times as fast at 800 °C.

Why: We first convert the two temperatures to Kelvin: 1073 K and 673 K, respectively. We then find the ratio of the radiated powers at these two temperatures by dividing Eq. 7.3.2 for 1073 K by that same equation for 673 K:

$$\frac{\text{radiated power at } 1073 \text{ K}}{\text{radiated power at } 673 \text{ K}} = \frac{(1073 \text{ K})^4}{(673 \text{ K})^4} = 6.46.$$

Staying Warm by Controlling Thermal Radiation

The average temperature of human skin is about 307 K (34 °C or 93 °F), and a typical person has a body surface area of about 1.8 m^2 (19 ft 2). Assuming an emissivity of 0.97, the Stefan–Boltzmann law predicts that a person will radiate about 850 W of thermal power. If you actually lost heat that fast, you'd cool off quickly and be very uncomfortable. Thus, limiting radiative heat loss is important to maintaining your body temperature.

Fortunately, everything around you is also emitting thermal radiation, and you typically absorb about as much thermal power as you emit. You even absorb some of your own thermal radiation. If you hold your hand in front of your face on a cold day, you'll feel its thermal radiation warming your face.

However, whenever there's a temperature difference between you and your environment, heat will flow from hotter to colder every way it can, including radiation. While conduction and convection transfer heat in proportion to the difference in temperature, radiation transfers heat in proportion to the difference between the *fourth powers* of the absolute temperatures. That exotic relationship between temperatures and heat flow explains why radiative heat transfer to or from your skin is most noticeable when you're exposed to unusually hot or cold objects.

For example, the sun warms your skin quickly because it radiates more heat at you than the rest of your surroundings combined. Measured on an absolute temperature scale, the sun's surface temperature (5800 K) is about 20 times that of your skin (307 K). Though it's very distant and appears small to your eye, the sun radiates about 20^4 , or 160,000, times as much heat at you as you radiate at it.

In contrast, the dark night sky cools you quickly because it's actually colder than your surroundings. The nearly empty space beyond Earth's atmosphere is only a few degrees above absolute zero, and even with the atmosphere's considerable help, the sky's effective radiant temperature is still about 20 °C colder than the air around you. When you lie on the snow in an open field and look up at the clear night sky, your 307-K skin radiates about twice as much heat at the approximately 250-K sky as it radiates back at you. You lose heat quickly and feel cold.

You can improve your situation by moving under a sheltering pine tree. Although the tree is no warmer than the snow, it's about 20 °C hotter than the sky and therefore emits more thermal radiation at you. Although the tree isn't quite a crackling campfire, it will still help to keep you warm.

That said, we still have you running back and forth between hot and cold objects to keep warm. It's time to look at how you can control thermal radiation while staying in one place.

You can start by wearing a heavy coat. The coat's internal insulation allows its outer surface to cool down to the temperature of your environment, so that it radiates only a little more thermal power than it absorbs and doesn't lose much heat. This strategy is common in birds and mammals. Penguins and polar bears let the outsides of their insulated feathers and fur cool to the temperatures of their frozen worlds and therefore lose almost no heat by radiation.

You can also change your emissivity to control radiative heat transfer. We've seen that black surfaces are ideal whenever you want to exchange thermal radiation with other objects, but what about when you don't want to exchange thermal energy? That's where shiny, white, and transparent surfaces come into play.

A perfectly shiny or white surface has an emissivity of zero; it reflects all of the light that strikes it, absorbing none, and is itself unable to emit any thermal radiation. If you put a shiny or white surface between yourself and a snowman, you'll each see your own thermal radiation coming back at you and you won't exchange any heat. This insulating effect works even if one of you wears the shiny or white surface as clothing; whoever is wearing it won't emit thermal radiation at all, and the other will see his or her own thermal radiation returned by that clothing.



Courtesy Lou Bloomfield

Fig. 7.3.5 This silver coffee urn may not seem to be insulated, but it is. Its polished silver surface is so shiny that it is an almost perfect mirror for infrared and visible light. As a result, it neither emits nor absorbs room-temperature thermal radiation well. Moreover, the air near the urn's vertical sides forms tall convection cells that aren't very efficient at carrying heat away from the urn. Overall, the urn loses heat much more slowly than it would if it were black and had fewer vertical surfaces.

Unfortunately, while truly white clothing would make you immune to radiant heat transfer, such clothing doesn't exist. Clothing exemplifies the difference between high-temperature and low-temperature emissivities. Regardless of their colors in visible light, virtually all clothing materials are black to infrared light and have low-temperature emissivities close to 1. In other words, most clothes absorb and emit low-temperature thermal radiation almost perfectly. Thus, although wearing a white robe on a scorching summer day will reduce the amount of thermal radiation you absorb directly from the sun because the robe's high-temperature emissivity is nearly 0, it won't protect you from the sun-warmed objects around you because its low-temperature emissivity is nearly 1.

In contrast, shiny clothing can protect you from those sun-warmed objects, but only if it contains metals. Metals have small low-temperature emissivities because they conduct electricity, a property that allows them to reflect electromagnetic waves extremely well. When you wrap food in aluminum foil, its emissivity drops to about 0.05 and it can barely exchange heat by radiation (Fig. 7.3.5). Wearing metallic clothes has the same effect for people. Lamé fabrics—fabrics with metal threads woven into them—help to keep their typically under-dressed wearers warm. A less flamboyant use of metallic cloth is in the-metal-coated plastic blankets that are part of emergency rescue kits; wrapping yourself in one of these blankets shiny-side out keeps you from exchanging thermal radiation with your surroundings (see 2).

Transparent materials also have emissivities near 0 and neither absorb nor emit thermal radiation well. Unlike shiny or white surfaces, however, transparent materials avoid absorbing thermal radiation by letting it pass through them. For example, when sunlight streams through a window to warm your skin, it's obvious that glass doesn't absorb much of the sun's thermal radiation. As for emission, any artist who sculpts molten glass knows from painful experience that glass provides almost no visible warning that it's extremely hot.

However, emissivity once again varies with temperature. Nearly all the materials that are transparent to the visible light of high-temperature thermal radiation, including glass, are black to the infrared light of low-temperature thermal radiation. That's probably just as well for clothing because clothing that's transparent to thermal radiation wouldn't provide much insulation.

2 Smoke jumpers who battle forest fires occasionally get trapped by the fires they're fighting and must try to survive as the fires burn over them. Their chances improve significantly if they use small personal survival shelters that have shiny metallic surfaces. Huddled against the cool ground underneath one of these tent-like shelters, a firefighter is relatively insulated from the fire overhead. Assuming the firefighter is in a low spot with nothing to burn nearby and makes no contact with the shelter itself, conduction and convection can't convey much heat to the firefighter. And with the shelter's shiny, low-emissivity surface reflecting most of the fire's thermal radiation, radiative heat transfer is also greatly diminished. On August 29, 1985, 73 firefighters were trapped for several hours by a fire in the Salmon National Forrest near Salmon, Idaho, and survived only with the aid of their personal fire shelters.

Check Your Understanding #5: Foil Wrapping

Wrapping a hot dish of food in shiny aluminum foil seems to keep the dish warm longer than wrapping it in clear plastic film. Aluminum is a much better conductor of heat than plastic, so why does it impede the flow of heat so well?

Answer: The aluminum foil prevents the dish from losing heat to its environment via thermal radiation.

Why: Both wraps insulate by trapping air, not by preventing heat from conducting through the wrap itself. They're both too thin to permit much of a temperature difference between their inner and outer surfaces, despite their huge difference in thermal conductivities. In the case of the aluminum foil, however, aluminum's low emissivity also prevents it from exchanging thermal radiation with the food or with the surroundings.

Courtesy Lou Bloomfield

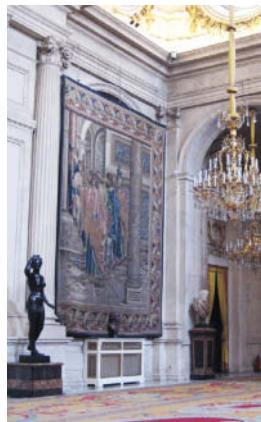


Fig. 7.3.6 Although stone is not a good conductor of heat, it's not nearly as good an insulator as air trapped in a fibrous mat. Medieval stone castles were notoriously cold in winter because heat flowed too easily out of them through their stone walls. This tapestry slows the flow of heat to the outside air and helps to keep the room warm.

Insulating Houses

The goal of housing insulation is to limit heat flow into or out of a house so that the temperature inside the house is nearly independent of the temperature outside. It's based on the same insulating concepts that help keep people and animals warm, but because houses and their contents rarely move, they can employ insulating methods and materials that are heavy, bulky, rigid, or fragile.

While there are many solid materials that are poor conductors of heat, including glass, plastic, wood, sand, and brick, they aren't nearly as insulating as air. Of course, since air tends to undergo convection, it needs to be held in place by porous or fibrous materials such as glass wool, sawdust, plastic foam, or narrow channels. Trapped air is the primary insulation in most buildings.

Glass wool or fiberglass is made by spinning molten glass into long, thin fibers that are then matted together like cotton candy. Solid glass is already a poor conductor of heat, but reducing it to fibers makes it even more insulating. The path that heat must take as it's conducted through the tangled fibers is so long and circuitous that little heat gets through. Most of the volume in glass wool is taken up by trapped air. The glass fibers keep the air from undergoing convection so the air can carry heat only by conduction.

Together, glass wool and the air trapped in it make excellent insulation. Since they're also nonflammable, they are commonly used in ovens, hot-water heaters, and other high-temperature appliances. Most modern houses have about 10–20 cm (4–8 in) of glass wool insulation built into their outside walls, along with a vapor barrier to keep the wind from blowing air directly through the insulation. (For a discussion of older insulating techniques, see Fig. 7.3.6.)

Because hot air rises and cold air sinks, the temperature difference between the hot air below the ceiling of the top floor and the cold air above that ceiling can become quite large during the winter. That ceiling is thus a very important site of unwanted heat transfer and requires heavy insulation. Glass wool inserted above the ceiling of the top floor in a modern house may be more than 30 cm (12 in) thick (Fig. 7.3.7).

While glass wool is an excellent insulator, other materials are used in certain situations. Urethane foam and polystyrene foam are waterproof, windproof, and even better insulators than glass wool. Unfortunately, they're also flammable and fragile. Nonetheless, they are frequently used in buildings and in refrigerators and freezers, where condensation makes waterproof insulation a necessity. Polystyrene foam also makes insulating cups that keep hot coffee hot and cold drinks cold. Unfortunately, polystyrene foam is hard to recycle, so many coffee shops now use alternatives that are less insulating but more environmentally friendly.



Fig. 7.3.7 The space between the sloped ceiling of this building and its actual roof is insulated with a thick mat of fiberglass insulation (left). The importance of that insulation is clearly visible on a snowy day, when the snow melts first where the fiberglass insulation is compressed by the beams and is therefore thinnest (right).



Courtesy Lou Bloomfield

Check Your Understanding #6: Is More Always Better?

Glass wool insulation is easily compressed so that you can put two or three layers into the space that one layer will normally fill. To improve a building's insulation, why not pack as much glass wool insulation into the walls as possible?

Answer: The glass wool's purpose is to trap air so that the air can act as the insulation. Glass is a better conductor of heat than air, so adding more glass wool will actually reduce the insulating effect.

Why: Commercial glass wool is designed to thwart convection in air while minimizing conduction through the glass. The wool is just dense enough to keep the air in place. Overpacking the wool will reduce its insulation and add weight and expense to the construction.

Modern Insulating Windows

Since windows have to be transparent, they can't be filled with fiberglass or foam, and a window made of solid glass is a poor insulator unless it's extremely thick. Insulating windows requires a different approach.

The most common way to insulate a window is to use two thin panes of glass that are spaced a centimeter or two apart. That vertical gap is typically filled with argon gas, which is an even poorer conductor of heat than air (see Table 7.3.1). Although convection occurs inside the gap, the convection cells that form are tall and thin and therefore relatively ineffective at carrying heat from one pane to the other. When properly designed, assembled, and installed, a double-pane window transports relatively little heat by conduction or convection.

What about radiative heat transfer? Alas, this turns out to be a big problem. Although ordinary glass is clear to visible light, it has a low-temperature emissivity of 0.92 and is nearly black to long-wavelength infrared light. Thus the glass panes of a conventional double-pane window are exchanging thermal radiation with almost perfect efficiency. If there's a big temperature difference between the inside and outside panes, heat will flow via thermal radiation from the hotter pane to the colder pane and spoil the window's insulating behavior.

To decrease this radiant heat flow, an energy-efficient low-emissivity (Low-E) window has a thin coating applied to the inner surface of one of its panes. That coating makes the pane's surface behave like a shiny mirror for infrared light; it emits very little thermal radiation, and it reflects back most of what it receives from the other pane. The coating doesn't reflect visible light, though, so the window still looks clear. Known as a "heat mirror" because it reflects only low-temperature thermal radiation, this coating dramatically reduces the flow of radiant heat through the window.

One of the most common Low-E coatings is indium-tin-oxide, a transparent electrical conductor that's also widely used in electronic displays for watches and computers. Like a metal, indium-tin-oxide uses its electrical conductivity to reflect electromagnetic waves. It has a low-temperature emissivity of about 0.10, so it's almost as reflective as aluminum foil. However, because indium-tin-oxide's conductivity is limited, it can respond only to long-wavelength electromagnetic waves. It reflects low-temperature thermal radiation but transmits visible light.

By suppressing all three heat transfer mechanisms, a Low-E window is able to provide surprisingly good insulation. The temperature of its outer pane matches the outdoor temperature and the temperature of its inner pane matches the indoor temperature, yet little heat flows through the gap between those panes.

One potential problem for Low-E windows is leakage. If the argon gas escapes from between the two panes and is replaced by moist air, the heat mirror coating may degrade and the window can become foggy. Keeping the window assembly perfectly sealed for decades is a challenge, particularly when that assembly is subject to large changes in temperature. The metals, plastics, and glass used in the window assembly don't all expand equally as they warm up, and the assembly can literally tear itself apart.

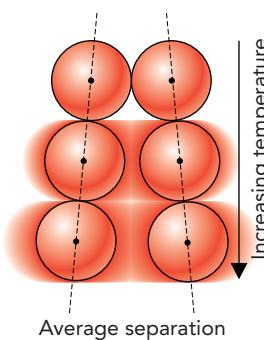


Fig. 7.3.8 The thermal kinetic energy of a solid increases with temperature, causing its atoms to bounce against one another more vigorously. As they vibrate, the atoms repel more strongly than they attract, so their average separation increases slightly.

A material's thermal expansion is caused by atomic vibrations. Because of thermal energy, adjacent atoms vibrate back and forth about their equilibrium separations (Fig. 7.3.8). This vibrational motion isn't symmetrical; the repulsive forces the atoms experience when they're too close together are stiffer than the attractive forces they experience when they're too far apart. As a result of this asymmetry, they push apart more quickly than they draw together and thus spend most of their time at more than their equilibrium separation. On average, their actual separation is larger than their equilibrium separation, and the material containing them is bigger than it would be without thermal energy.

As an object's temperature rises, it grows larger in all directions. The extent to which it expands with increasing temperature is normally described by its **coefficient of volume expansion**, the fractional change in the object's volume per unit of temperature increase. Fractional change in volume is the net change in volume divided by the total volume. Since most materials expand only a small amount when becoming 1 °C (or 1 K) hotter, coefficients of volume expansion are small, typically about $2.5 \times 10^{-5} \text{ K}^{-1}$ for glass, $5 \times 10^{-5} \text{ K}^{-1}$ for metals, and $2 \times 10^{-4} \text{ K}^{-1}$ for plastics.

As the window assembly heats up, its plastics expand more than its metals, which expand more than its glass. Those differences in expansion put stresses on the components and can separate them or break them. Making an assembly that tolerates thousands of warming and cooling cycles without leaking is no small feat.

Check Your Understanding #7: Cool Light

To provide bright white light and almost no unnecessary heat, a special lamp uses a heat mirror. What does that heat mirror do?

Answer: The heat mirror reflects infrared light back toward the actual light source while allowing the visible light to continue on to its destination.

Why: The heat mirror transmits only visible light, so it reflects away infrared light that would simply add unnecessary heat to the lamp's output.

Earth's Temperature and the Greenhouse Effect

Despite variations with time, place, and season, Earth's overall surface temperature has a fairly constant average value of about 15 °C (59 °F). To maintain that constant average temperature, Earth must avoid gaining or losing heat. A net flow of heat to Earth would increase its average temperature, while a net flow from Earth would decrease its temperature. In short, Earth must balance any heat it receives with an equal amount of heat it emits.

Earth's main source of heat is the sun, and this heat reaches Earth almost entirely as electromagnetic radiation. Because the sun's surface temperature is about 5800 K, solar radiation is primarily visible light, although it also includes a substantial amount of infrared and ultraviolet light (Fig. 7.3.4). The total solar power reaching Earth is about $1.74 \times 10^{17} \text{ W}$. For comparison, the world's total electric-generating capacity is roughly $3 \times 10^{12} \text{ W}$.

While about 30% of the sunlight is simply reflected or scattered back from Earth's surface and atmosphere, the remaining $1.21 \times 10^{17} \text{ W}$ is absorbed by Earth's surface and atmosphere. Earth must get rid of that heat by radiating it into cold, dark space as black-body radiation. To radiate $1.21 \times 10^{17} \text{ W}$ of thermal power, Earth's radiating surface—the effective source of its blackbody radiation—must have a temperature of about -18°C (0°F). So while Earth receives heat as mostly visible sunlight, it reemits that heat as invisible infrared light.

If there were no atmosphere, Earth's radiating surface would be the ground itself and the average temperature of that ground would be about -18°C . But Earth does have an atmosphere, and even though that atmosphere is nearly transparent to visible light, it absorbs and emits infrared light fairly well. Earth's principal source of thermal radiation

therefore isn't the ground but the atmosphere. In fact, Earth's radiating surface is located about 5 km (3 miles) above the ground, and it is the air at that altitude that has the -18°C average temperature!

This displacement of Earth's radiating surface from ground level to an elevated layer in the atmosphere gives rise to the **greenhouse effect**. Since the ground itself isn't responsible for radiating away Earth's excess heat, its temperature doesn't have to be -18°C . Also because the atmosphere naturally develops a temperature gradient of about -6.6°C per kilometer of altitude (see 3), the ground 5 km below Earth's radiating surface is about 33°C hotter, or roughly 15°C .

Since life would be difficult at -18°C , we are fortunate to have the greenhouse effect. However, too large a greenhouse effect would be disastrous. If the atmosphere were to become even more effective at absorbing and emitting infrared light, the altitude of Earth's radiating surface would increase and so would the temperature at Earth's surface. It is critical to our climate that this altitude stay about 5 km and not rise to 6 km or more.

Just how effective Earth's atmosphere is at absorbing and emitting infrared light depends on its chemical makeup. Nitrogen and oxygen molecules, though extremely common in the atmosphere, are remarkably transparent to both infrared and visible light. It's the less common, more complicated gas molecules that allow air to absorb and emit infrared light and are thus greenhouse gases.

Although the principal greenhouse gas is water vapor, gases such as carbon dioxide, methane, and nitrous oxide are also extremely important. Each gas absorbs its own characteristic portions of the infrared spectrum, and together they darken the atmosphere to infrared light the way a mixture of watercolors darkens a canvas to visible light. The more greenhouse gases the atmosphere contains, the higher the altitude of Earth's radiating surface and the warmer Earth's surface.

There is now overwhelming scientific evidence that human production of greenhouse gases, particularly carbon dioxide, is causing a rapid warming of Earth's surface and a change in its climate. That climate change will be severe unless we dramatically reduce greenhouse emissions now. Since burning fossil fuels produces carbon dioxide, our current approach to using those fuels must change. Atmospheric methane is generated primarily by digestive bacteria in cows, and reducing it will require changes to the cattle and dairy industries. There are other gases that we have been releasing casually into the atmosphere for years, including refrigerants and aerosol propellants, that turn out to be especially potent greenhouse gases and don't belong in the atmosphere. Alas, maintaining the environment turns out to be much harder than people thought at the dawn of the industrial revolution. We reap what we sow.

3 A competition between gravity and thermal energy structure Earth's atmosphere and give it downward gradients in pressure, density, and temperature (see Section 5.1). One explanation for the atmosphere's decreasing temperature with altitude is that rising air molecules convert thermal energy into gravitational potential energy, so air becomes cooler as it rises upward. Another equivalent explanation is that rising air expands as the surrounding pressure decreases and that expansion cools the air. Although the presence of moisture in the air weakens this effect, the air still gets significantly colder as you go up a mountain.

► Check Your Understanding #8: Canned Heat

When you use a can of spray paint, its propellant (typically propane and butane) immediately enters the atmosphere. As the paint dries, its solvents also enter the atmosphere. How do these gases affect the atmosphere's transparency in the infrared?

Answer: All these gas molecules darken the atmosphere in the infrared.

Why: Most of the chemicals in your spray can end up in the atmosphere, where they increase its efficiency at emitting and absorbing infrared light and therefore contribute to the greenhouse effect.

Epilogue for Chapter 7

This chapter has examined the roles of thermal energy and heat in a variety of common objects. In Woodstoves, we saw how combustion converts ordered chemical potential energy into disordered thermal energy and we studied the ways in which this thermal energy flows into the woodstove's surroundings: conduction, convection, and radiation. In Water, Steam, and Ice, we examined the three phases of matter and looked at how the transformations

between those phases are influenced by temperature, heat, and other characteristics of their environment. In Clothing, Insulation, and Climate, we examined the laws governing heat flow and found out how to limit that flow. We then used that understanding to explain the greenhouse effect and the ways humans are contributing to climate change.

Explanation: A Ruler Thermometer

Like most things, the clear plastic ruler expands when you heat it. Its length increases by an amount proportional to its increase in temperature. When you transfer heat to the ruler, by breathing on it, touching it, or exposing it to a hair dryer, its temperature rises, it expands, and its free end turns the needle and pointer. Since the pointer's movement is proportional to the ruler's length change, it's also proportional to the thermometer's change in temperature.

Chapter Summary and Important Laws and Equations

How Woodstoves Work: A woodstove burns wood in air to obtain hot gas. The burning process is actually a chemical reaction in which molecules in the wood and air are disassembled into fragments and then reassembled into new, more tightly bound molecules such as water and carbon dioxide. This reassembly process releases more energy than was required to disassemble the original wood and oxygen molecules. This extra energy appears as thermal energy within the reaction products, so they're hot. Rather than distributing the hot burned gas directly to the room, a woodstove transfers heat from the burned gas to clean air or water. Heat is conducted through the walls of a woodstove and then flows into the room via convection and radiation.

How Water, Steam, and Ice Work: Water, steam, and ice are the liquid, gaseous, and solid phases of the chemical we call water. The water molecules in liquid water are bound together strongly enough to fix water's volume, but not strongly enough to give water a rigid shape. In steam, the water molecules are independent and the gas has neither a fixed shape nor a fixed volume. The molecules in ice are bound together rigidly in solid crystals, so ice is rigid and its volume is constant.

Ice can transform into water, or vice versa, at its melting temperature. Heat added to ice at this temperature causes it to melt without becoming warmer. Heat removed from water at this temperature causes it to freeze without becoming colder. Similarly, water and ice can transform into steam, or vice versa, through evaporation, condensation, sublimation, and deposition. Turning water into steam requires heat, and this heat is released when the steam turns back into water. Boiling is a special case of evaporation in which evaporation occurs within the body of the water itself. For boiling to occur, the water's vapor pressure must equal the ambient pressure on the water, which is normally atmospheric pressure.

How Clothing, Insulation, and Climate Work: Clothing serves as insulation for people, helping them maintain their body temperatures in inhospitable environments. Most clothes are constructed out of threads and other fibrous materials that not only are poor conductors of heat but also trap air so that it can act as the primary insulation. The outer surface of thick clothing has nearly the same temperature as its surroundings and therefore exchanges little heat via radiation.

Building insulation reduces heat flow between inside and outside. The insulation in walls and ceilings consists primarily of air trapped in fibrous or porous materials such as glass wool and urethane foam, but the insulation in windows is necessarily different. The narrow, gas-filled gap between doubled panes in a modern window limits the heat that can flow between the panes by conduction and convection, while a low-emissivity coating on one of the panes limits radiative heat transfer.

Earth's climate is determined in large part by how it eliminates the heat it receives from the sun. That heat is emitted into space as infrared thermal radiation, mostly by Earth's atmosphere. The average temperature at Earth's surface increases in proportion to the effective altitude from which that thermal radiation is emitted. Since greenhouse gases darken the atmosphere to infrared light and increase that altitude, they warm Earth's surface. Human production of such gases is changing Earth's climate.

1. Heat conduction: The thermal power an object conducts from one surface to the other is equal to its thermal conductivity times the temperature difference between the surfaces times their surface area divided by the distance separating the surfaces, or

$$\text{heat flow} = \frac{\text{thermal conductivity} \cdot \text{temperature difference} \cdot \text{area}}{\text{separation}}. \quad (7.3.1)$$

2. The Stefan–Boltzmann Law: The power an object radiates is proportional to its emissivity times the fourth power of its temperature times its surface area, or

$$\begin{aligned} \text{radiated power} &= \text{emissivity} \cdot \text{Stefan–Boltzmann constant} \\ &\cdot \text{temperature}^4 \cdot \text{surface area}. \end{aligned} \quad (7.3.2)$$

Thermodynamics

ACTIVE LEARNING EXPERIMENTS

Making Fog in a Bottle

Fog occurs naturally when humid air experiences a sudden drop in temperature. To make fog in a bottle, you'll need to do the same thing: cool humid air quickly. In a room where everything is at the same temperature, cooling air sounds impossible. It's not.

To make fog in a bottle, obtain a clean 1- or 2-liter plastic soda bottle and put several spoonfuls of water in it. Cap the bottle tightly, and shake it to help the water evaporate into the air. In a minute or two, the relative humidity in the bottle will reach 100% and you'll be ready to do the experiment.

Lay the bottle on its side on the floor, and step on it with your shoe. You must press down hard enough to dent the bottle significantly, but you don't want to pop it or crack its sides. Wait like this for 20 or 30 seconds. Heat is

flowing out of the bottle, and that takes time. Now step quickly off the bottle, and immediately illuminate its contents with a bright light. You should see a mist of tiny water droplets. Abruptly releasing the compression causes the air's temperature to plummet, and fog forms in the bottle.

If no fog has formed despite a hard squeeze with your shoe and enough patience to let the heat flow out of the bottle, you can help the fog start by dropping a smoking match into the bottle before capping it. You want just enough tiny smoke particles to help the mist droplets form but not enough smoke to obscure the fog itself. It doesn't take much smoke to do the trick. After capping the bottle, squeeze it hard and wait, and then release the compression suddenly and look for the fog.



Courtesy Lou Bloomfield

Chapter Itinerary

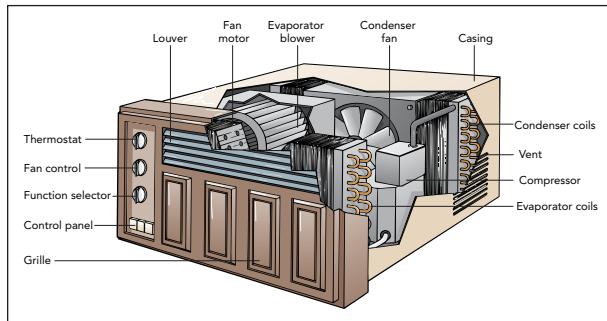
In this chapter, we examine the movement of heat in two everyday machines: (1) *air conditioners* and (2) *automobiles*. In Air Conditioners, we look at the rules governing the movement of thermal energy—the laws of thermodynamics—and see how those rules permit an air conditioner to use ordered electric energy to transfer heat from the cold room air to the hot outdoor air. In Automobiles, we investigate the ways in which thermal energy can be used to do work and see how the engine is able to convert thermal energy into work as heat flows from the hot burning fuel to the cold outdoor air. For additional preview

information, skip ahead to the Chapter Summary and Important Laws and Equations at the end of the chapter.

The principles illustrated by these two machines are also found in many other places. Refrigerators and home heat pumps use ordered energy to transfer heat from cold objects to hot objects, and steam engines and hot-air balloons use the natural flow of heat to make things move. Once you understand the concepts behind these *heat pumps* and *heat engines*, you'll see that they're quite common in the world around you.

SECTION 8.1

Air Conditioners



On a summer day, your problem isn't staying warm; it's keeping cool. Instead of looking for something to burn in your wood-stove, you turn on your air conditioner. An air conditioner is a device that cools room air by removing some of its thermal energy. The air conditioner, however, can't make thermal energy either disappear or become ordered energy. Instead, it transfers thermal energy from the cooler room air to the warmer air outside. Since the air conditioner transfers heat against its natural direction of flow, the air conditioner is a *heat pump*. It's

also a classic illustration of the laws of thermodynamics in action.

Questions to Think About: Why doesn't heat naturally flow from a colder object to a hotter object? An air conditioner removes thermal energy from room air, so why does the air conditioner require electric energy to operate? Where does this electric energy go? Why does an air conditioner always have an indoor component and an outdoor component? If you put a window air conditioner in the middle of a room and turn it on, what will happen to the temperature of the room?

Experiments to Do: Take a look at a window air conditioner. If you can't find one, examine a refrigerator instead because it's basically a powerful air conditioner cooling a food-storage closet. As the air conditioner (or refrigerator) operates, feel the air leaving the indoor air vent (or inside the refrigerator) and compare its temperature to that of air leaving the outdoor air vent (or near the metal coils on the back of the refrigerator). Which way does the cooling mechanism move heat? If that mechanism were absent, which way would heat flow? Turn off the air conditioner or refrigerator, and observe the resulting heat flow. Did you confirm your prediction?

Moving Heat Around: Thermodynamics

On a sweltering summer day, the air in your home becomes unpleasantly hot. Heat enters your home from outdoors and doesn't stop flowing until it's as hot inside as it is outside. You can make your home more comfortable by getting rid of some of its thermal energy. Although we've already looked at ways to *add* thermal energy to room air, we haven't yet learned how to *remove* it. At present, the only cooling method we've discussed is contact with a colder object. Unless you have an icehouse nearby, you need another scheme for eliminating thermal energy. You need an air conditioner.

An air conditioner transfers heat against its natural direction of flow. Heat moves from the colder air in your home to the hotter air outside so that your home gets colder while the outdoor air gets hotter. There's a cost to transferring heat in this manner. The air conditioner requires ordered energy to operate and typically consumes large



Fig. 8.1.1 Although they are touching, no heat is flowing between the red bottle and the blue bottle. Similarly, no heat is flowing between the red bottle and the green bottle. Based on those two observations, the law of thermal equilibrium states that no heat will flow between the blue bottle and the green bottle when they touch.

amounts of electric energy. It's a type of **heat pump**, a device that uses ordered energy to transfer heat from a colder object to a hotter object, against its natural direction of flow.

Before studying how an air conditioner pumps heat, we should first show that pumping is necessary. There are three harebrained cooling alternatives that we need to discredit before turning to air conditioning:

1. Letting heat flow from your home to your neighbor's home.
2. Destroying some of your home's thermal energy.
3. Converting some of your home's thermal energy into electric energy.

We'll find it useful to examine each of these alternatives because, in doing so, we'll learn about the laws governing the movement of thermal energy, the **laws of thermodynamics**.

The first alternative raises an interesting issue. Your home is in thermal equilibrium with the outdoor air, meaning that no heat flows from one to the other and they're at the same temperature. Your neighbor's home is also in thermal equilibrium with the outdoor air. What will happen if you permit heat to flow between your home and your neighbor's home? Nothing. Since both homes are simultaneously in thermal equilibrium with the outdoor air, they're also in thermal equilibrium with one another. All three are at the same temperature.

This observation is an example of the **law of thermal equilibrium** (often called the **zeroth law of thermodynamics**), which says that two objects that are each in thermal equilibrium with a third object are also in thermal equilibrium with one another (Fig. 8.1.1). This seemingly obvious law is the basis for a meaningful system of temperatures. If you had a roomful of objects at 35 °C (95 °F) and some were in thermal equilibrium with one another while others were not, then "being at 35 °C" wouldn't mean much. However, every object that has a temperature of 35 °C is in thermal equilibrium with every other object at 35 °C. The law of thermal equilibrium is observed to be true in nature, so temperature does have meaning. Since your neighbor's home is just as hot as yours, they can relax because you're not going to be sending them any extra heat.

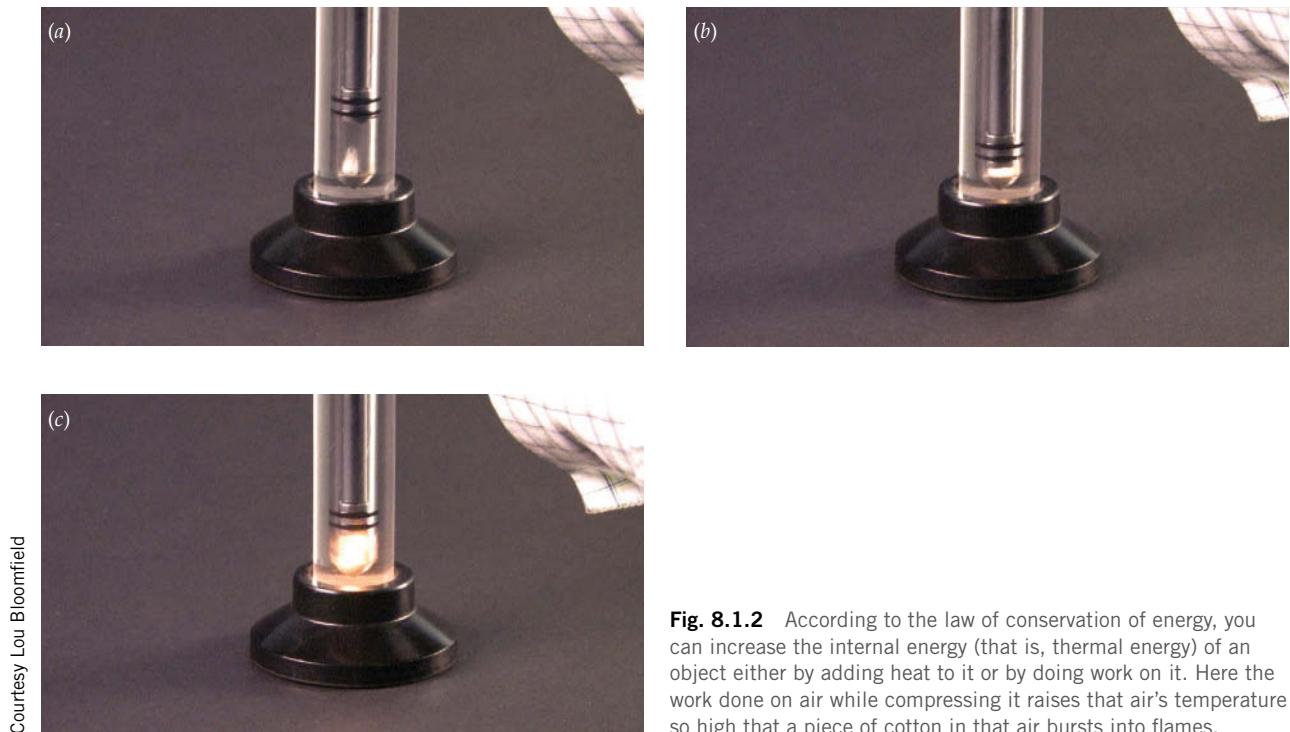
THE LAW OF THERMAL EQUILIBRIUM

Two objects that are each in thermal equilibrium with a third object are also in thermal equilibrium with one another.

The second alternative sounds unlikely from the outset. We've known since the first chapter that energy is special, that it's a conserved quantity. You can't cool your home by destroying thermal energy because energy can't be destroyed. To eliminate thermal energy, you must convert it to another form or transfer it elsewhere.

This concept of energy conservation is the basis for the **law of conservation of energy** (often called the **first law of thermodynamics**), which officially recognizes that there are actually two ways to transfer energy: the mechanical means we know as work and the thermal means we know as heat. You can increase an object's internal energy, which includes its thermal energy, only by doing work on it or by transferring heat to it (Fig. 8.1.2).

In its traditional form, the law of conservation of energy observes that the change in a stationary object's internal energy is equal to the heat transferred into that object minus the work that object does on its surroundings. In other words, heat added to the object increases



Courtesy Lou Bloomfield

Fig. 8.1.2 According to the law of conservation of energy, you can increase the internal energy (that is, thermal energy) of an object either by adding heat to it or by doing work on it. Here the work done on air while compressing it raises that air's temperature so high that a piece of cotton in that air bursts into flames.

its internal energy, while work done by the object decreases its internal energy. The law of conservation of energy can be written as a word equation:

change in object's internal energy = heat added to object – work done by object, (8.1.1)

in symbols,

$$\Delta U = Q - W,$$

and in everyday language:

You can add energy to a ball as heat by cooking it or as work by squeezing it.

THE LAW OF CONSERVATION OF ENERGY

The change in a stationary object's internal energy is equal to the heat transferred into that object minus the work that object does on its surroundings.

Check Your Understanding #1: Stirring Up Trouble

If you put cold water into a blender and mix it rapidly for several minutes, the water will become warm. From where does the additional thermal energy come?

Answer: The blender's blade does work on the water, and this work becomes thermal energy.

Why: The law of conservation of energy states that the change in the water's internal energy is equal to the heat flowing into it minus the work it does on its surroundings. In this case, the water's surroundings are doing work on it by stirring it, so its internal energy increases. Since the water can't store this new internal energy as ordered potential energy, the energy becomes thermal energy and the water gets hotter.

Check Your Figures #1: Powerful You

You have been pedaling the exercise bicycle steadily for 20 min at a power of 200 W. How much thermal power is the bicycle emitting into the room?

Answer: 200 W.

Why: Since the bicycle's internal energy can't increase or decrease forever, it must be emitting as much thermal power into the room as it is receiving mechanical power from you. As required by Eq. 8.1.1, the 200 W of mechanical power you supply to the bicycle is flowing into the room as 200 W of thermal power.

Disorder and Entropy

The third alternative looks much more promising than the first two. It seems as though you should be able to convert thermal energy into electricity (or some other ordered form of energy). You could then sell it back to the electric company and get credit on your bill. Wouldn't that be great?

There's a problem with this idea—ordered energy and thermal energy aren't equivalent. You can easily convert ordered energy into thermal energy, but the reverse is much harder. For example, you can burn a log to convert its chemical potential energy into thermal energy, but you'll have trouble converting that thermal energy back into chemical potential energy to recreate the log.

The basic laws of motion are silent on this issue. The log's original energy and constituent particles still exist, and they could, in principle, act together to reassemble the log. It is the laws of statistics that prevent the log's reassembly. That reassembly process would require a long series of remarkably unlikely events. The particles would all have to move in just the right ways to turn the burned gases back into wood and oxygen, an incredible coincidence that simply never happens. Similarly, all the air particles in your home would have to act together to convert their thermal energy into electricity. Since that coordinated behavior is ridiculously improbable, you're not going to be selling thermal-energy power to the electric company any time soon.

Once ordered energy has been scattered randomly among the individual air particles, you can't collect that energy back together again. Creating disorder out of order is easy, but recovering order from disorder is nearly impossible. As a result, systems that begin with some amount of order gradually become more and more disordered, never the other way around (Fig. 8.1.3). The best they can do is to stay the same for a while so that their disorder doesn't change. From these observations, we can state that the disorder of an isolated system *never decreases*.

This notion of never-decreasing disorder is one of the central concepts of thermal physics. There is even a formal measure of the total disorder in an object, **entropy**. All disorder contributes to an object's entropy, including its thermal energy and its structural defects. Both breaking a window and heating it increase its entropy. Although its name sounds similar to *energy*, don't confuse energy and entropy. Energy is a conserved quantity, while entropy is a quantity that can and generally does increase. It's easy to make more entropy.

Because disorder never decreases, the third cooling alternative is impossible. Turning your home's thermal energy into electric energy would reduce its disorder and decrease its entropy. Our observations about entropy aren't yet complete, though. There is one way to decrease your home's entropy: you can export that entropy somewhere else. In fact, you export entropy every time you take out the garbage, though that action also changes the contents of your home. You can also export entropy without modifying your home's contents by transferring heat somewhere else. Heat carries disorder and entropy with it, so getting rid of heat also gets rid of entropy.

Our rule about entropy never decreasing is weakened by the possibility of exchanging heat and entropy between objects. Before asserting that an object or system of objects can't decrease its entropy, we must ensure that the object is thermally isolated from its surroundings so that it can't export its entropy. With that in mind, the strongest statement that we can make

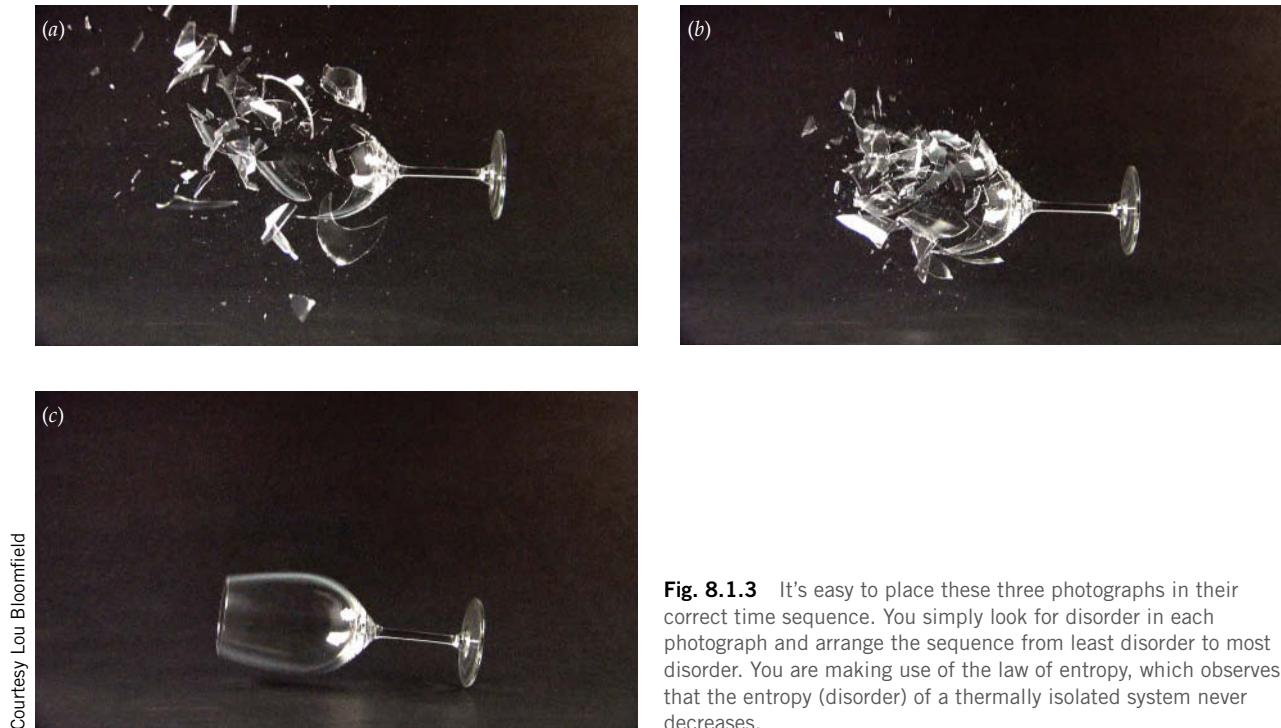


Fig. 8.1.3 It's easy to place these three photographs in their correct time sequence. You simply look for disorder in each photograph and arrange the sequence from least disorder to most disorder. You are making use of the law of entropy, which observes that the entropy (disorder) of a thermally isolated system never decreases.

concerning entropy is that the entropy of a thermally isolated system of objects never decreases. This observation is the **law of entropy** (often called the **second law of thermodynamics**).

THE LAW OF ENTROPY

The entropy of a thermally isolated system of objects never decreases.

Because of the law of entropy, the only way to cool your home is to export its thermal energy and entropy elsewhere. Such a transfer would be easy if you had a cold object nearby to receive the heat. Lacking a cold object, you must use an air conditioner. Like all heat pumps, an air conditioner transfers heat and entropy in such a way that the law of entropy is never violated and the entropy of each thermally isolated system of objects never decreases. As we'll see, the air conditioner lowers the entropy of your home but raises the entropy of the outdoor air even more, so that, overall, the entropy of the world actually increases.

Check Your Understanding #2: Something for Nothing

People have tried for centuries to build machines that provide endless outputs of useful, ordered energy without any inputs of ordered energy. Unfortunately, such perpetual motion machines violate the laws of thermodynamics. If such a machine is thermally isolated, which law does it violate? What if it's not thermally isolated?

Answer: A thermally isolated perpetual motion machine violates the law of conservation of energy, while one that is not thermally isolated violates the law of entropy.

Why: A thermally isolated perpetual motion machine clearly violates the conservation of energy. This isolated machine simply can't export energy forever because it will eventually run out. A perpetual motion machine that is not thermally isolated may not violate conservation of energy because it can absorb heat energy from its surroundings. Instead, it violates the law of entropy. This machine can't endlessly absorb heat energy and then export it as ordered energy. In doing so, the machine will eventually begin to reduce the entropy of the universe and violate the law of entropy. Sad though it may be, perpetual motion machines can't exist.

Pumping Heat Against Its Natural Flow

Although the law of entropy doesn't allow the entropy of a thermally isolated system to decrease, it does permit the objects in that system to redistribute their individual entropies. One object's entropy can decrease as long as the entropy of the rest of the system increases by at least as much. Such entropy redistribution allows part of the system to become colder if the rest of the system becomes hotter.

For example, suppose that there's a pond of cold water behind your home. You pump that water through your bathtub and let it draw heat out of the room air. Your home becomes colder while the pond becomes warmer. This transfer of heat from the hot air in your home to the cold water in the pond satisfies the law of entropy. The entropy of the combined system—your home and the pool of water—doesn't decrease. In fact, it actually increases!

This entropy increase occurs because heat is more disordering to cold objects than it is to hot objects. Each joule of heat that flows from your home to the pool creates more disorder in the pool than it creates order in your home.

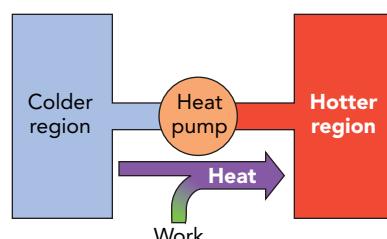
A useful analog for this effect involves two parties taking place simultaneously: the garden society's annual tea party and a 4-year-old's birthday party. The orderly tea party represents the cold pool while the disorderly birthday party represents your hot home. The analogy to letting heat flow from your hot home to the cold pool is to trade one lively 4-year-old from the disorderly birthday party for one quiet octogenarian from the orderly tea party. This exchange will reduce the birthday party's disorder only slightly, but it will dramatically increase the disorder of the tea party. The attendance at each party will be unchanged but their total disorder will increase.

When heat flows from your home to the pool, the overall entropy increases and the law of entropy is more than satisfied. A similar increase in entropy occurs whenever heat flows from a hot object to a cold object, which is why heat normally flows in that direction.

An air conditioner, however, does the seemingly impossible: it transfers heat from a cold object, your home, to a hot object, the outdoor air. This heat flows in the wrong direction, and the disorder it creates by entering the hot outdoor air is less than the disorder it removes by leaving the cold indoor air. It's like returning the tea party's lone 4-year-old to the birthday party in exchange for the elderly garden fancier—the birthday party becomes only a tiny bit more disorderly while the garden party becomes much more orderly, so the net disorder of the two gatherings decreases substantially. Similarly, if nothing else happened when the air conditioner moved heat from the cold indoor air to the hot outdoor air, the entropy of the combined system would decrease and the law of entropy would be violated!

However, we've omitted an important feature of the air conditioner's operation—the electric energy it consumes. The air conditioner converts this ordered energy into thermal energy and delivers it as additional heat to the outdoor air (Fig. 8.1.4). In doing so, the air conditioner creates enough extra entropy to ensure that the overall entropy of the combined system increases. The law of entropy is satisfied after all.

Fig. 8.1.4 A heat pump transfers heat from a colder region to a hotter region. In doing so, it converts some work (ordered energy) into heat (thermal energy in the hotter region). The larger the temperature difference between the two regions, the more work is required to transfer each joule of heat.



The amount of ordered energy the air conditioner consumes depends on the temperatures of the indoor air and outdoor air. If the two are close in temperature, the transfer of heat reduces the entropy only slightly, so the air conditioner doesn't need to convert much ordered energy into thermal energy. If they are far apart in temperature, however, the air conditioner must create lots of extra entropy to make up for the entropy lost in the transfer.

This requirement that entropy not decrease explains why an air conditioner works best when it's cooling your home the least. The greater the temperature difference between the indoor air and the outdoor air, the more electric energy or other form of work the air conditioner must consume to transfer each joule of heat. For an ideally efficient air conditioner or other heat pump, the relationships between the work consumed, the heat removed from the cold object, and the heat added to the hot object can be written as two word equations:

$$\text{heat removed from cold object} = \text{work consumed} \cdot \frac{\text{temperature}_{\text{cold}}}{\text{temperature}_{\text{hot}} - \text{temperature}_{\text{cold}}}$$

$$\text{heat added to hot object} = \text{heat removed from cold object} + \text{work consumed}, \quad (8.1.2)$$

in symbols:

$$\begin{aligned} -Q_c &= W \frac{T_c}{T_h - T_c}, \\ Q_h &= -Q_c + W, \end{aligned}$$

and in everyday language:

The closer the two temperatures are, the easier it is to move heat from cold to hot,

where the temperatures are measured on an absolute scale. The hot object receives not only the heat removed from the cold object but also an amount of heat equal to the work consumed by the transfer. Note also that the work needed to remove heat from your home approaches infinity as its temperature approaches absolute zero; that's why absolute zero is unattainable.

Sadly, a practical air conditioner never reaches ideal efficiency, so it moves less heat than promised by Eqs. 8.1.2. Moreover, heat leaks back into your home through its walls at a rate that's roughly proportional to the temperature difference. No wonder your electric bill soars when you cool your home to arctic temperatures in tropical weather!

Check Your Understanding #3: Heat Pumps in Cold Weather

Homes located in mild climates are often heated by heat pumps during the winter. Home heat pumps are essentially air conditioners run backward. They extract heat from the cold outdoor air and release it to the warm indoor air. Why are heat pumps most effective in mild weather, when the outdoor air isn't too cold?

Answer: heat pump requires more ordered energy to pump heat from a cold object to a hot object when the temperature difference between them is large.

Why: A heat pump becomes less efficient at pumping heat when the temperature of the heat's source becomes much colder than the heat's destination. The colder it is outside, the more ordered energy it takes to move each joule of heat. On bitter cold days, heat pumps aren't able to move enough heat to keep their homes warm, which is why most home heat pumps have built-in electric or gas furnaces to assist them during unusually cold weather.

Check Your Figures #2: A Hot House for Cheap

Your house is heated by an ideally efficient heat pump that uses ordered energy to pump heat from the colder outdoor air to the warmer indoor air. On a winter day, the outdoor temperature is 270 K (-3°C or 26°F) and the indoor temperature is 300 K (26°C or 80°F). To deliver 1000 J of heat to the indoor air, how much electric energy must the heat pump consume?

Answer: The pump must consume 100 J.

Why: The difference in temperature between indoor air and outdoor air is 30 K, so according to Eqs. 8.1.2, the heat pump can remove 9 times as much heat from the outdoor air as it consumes in work and deliver 10 times as much heat to the indoor air. Thus it takes only 100 J of electric energy to remove 900 J of heat from the outdoor air and deliver all 1000 J as heat to the indoor air. What a bargain!

How an Air Conditioner Cools the Indoor Air

Having determined the air conditioner's goals, we can now look at how a real air conditioner meets them. In most cases, the air conditioner uses a fluid to transfer heat from the colder indoor air to the hotter outdoor air. Known as the *working fluid*, this substance absorbs heat from the indoor air and releases that heat to the outdoor air.

The working fluid flows in a looping path through the air conditioner's three main components: an evaporator, a condenser, and a compressor (Fig. 8.1.5). The evaporator is located indoors, where it transfers heat from the indoor air to the working fluid (Fig. 8.1.6). The condenser is located outdoors, where it transfers heat from the working fluid to the outdoor air. The compressor is also located outdoors, where it squeezes the working fluid and does the work needed to move heat against its natural flow. To see how these three components pump heat out of your home, let's look at them individually.

We'll begin with the evaporator, a long metal pipe that's decorated with thin metal fins. The evaporator is a heat exchanger that allows heat to flow from the warm indoor air around it to the cool working fluid inside it. Its fins provide additional surface area through which heat can flow, and a fan blows indoor air rapidly past those fins to encourage that flow.

True to its name, the evaporator is where the working fluid evaporates from liquid to gas. To evaporate, the liquid working fluid needs its latent heat of evaporation, the energy required to separate a liquid's molecules so that it becomes a gas. The working fluid obtains part of that energy from its own thermal energy, and its temperature decreases. Heat from the indoor air then flows into the chilled working fluid and completes the evaporation. By the time the working fluid leaves the evaporator as a gas, it has absorbed a great deal of the indoor air's thermal energy and carries that energy with it as chemical potential energy.

To make the working fluid evaporate, the air conditioner abruptly reduces its pressure. In Chapter 7, we learned that phase transitions such as evaporation depend on molecular landing and leaving rates, and that those rates are sensitive to pressure and temperature. Working fluid

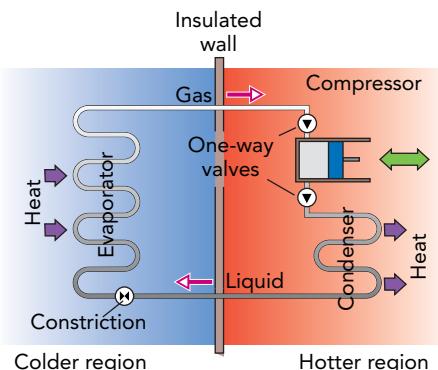


Fig. 8.1.5 A typical air conditioner transfers heat from colder air to hotter air by condensing a gas to a liquid in the hotter region and evaporating the liquid to a gas in the colder region. A compressor provides the necessary input of ordered energy.

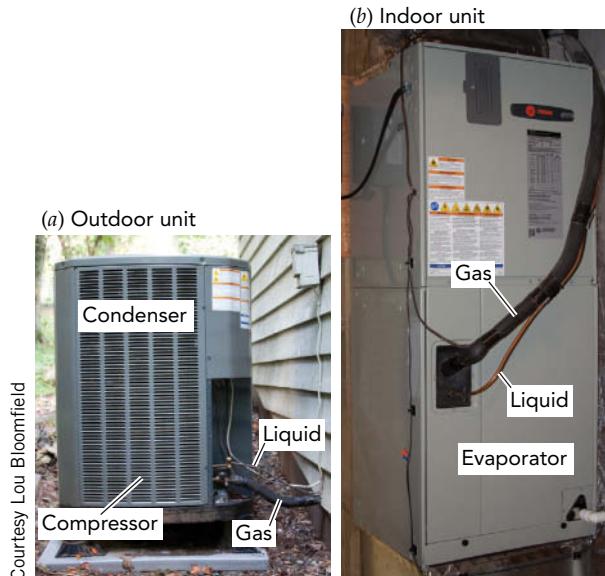


Fig. 8.1.6 The two components of a central air-conditioning system. (a) In the outdoor unit, gaseous working fluid is compressed and condensed into a liquid, releasing heat into the outdoor air. (b) In the indoor unit, liquid working fluid is evaporated into a gas, extracting heat from the indoor air. This system is actually a heat pump, meaning that the roles of the condenser and evaporator can be reversed and the system can heat the indoor air during the winter.

flows toward the evaporator as a warm, high-pressure liquid, but its pressure drops dramatically when it flows into the evaporator through a narrow constriction in the tubing. That pressure drop slows the rate at which working fluid molecules land on the liquid from the gas so that molecules leaving the liquid for the gas dominate and the liquid starts to evaporate.

As the working fluid evaporates, its temperature plummets and its evaporation slows down; molecules rarely gather enough thermal energy from the cold liquid to leave it for the gas. However, heat from the warmer room air now rushes into the cold working fluid and encourages further evaporation. By the time the working fluid emerges from the evaporator, it has evaporated completely and has absorbed considerable thermal energy from the indoor air. It leaves the evaporator as a cool low-pressure gas and travels through a pipe toward the compressor.

Half the air conditioner's job is done: it has removed heat from the indoor air. The remaining half of its job is more challenging; it must add heat to the outdoor air while ensuring that the total entropy of the combined system doesn't decrease. After all, there's no getting around the law of entropy.

Check Your Understanding #4: Cooling at the Gas Grill

When liquid propane evaporates into a gas in the tank of a propane grill, the tank that contains that liquid propane becomes colder. Why?

Answer: The tank's liquid propane needs heat to evaporate into a gas, and it extracts that heat from its surroundings.

Why: Just as in the evaporator of an air conditioner, the evaporating liquid propane absorbs heat.

How an Air Conditioner Warms the Outdoor Air

Satisfying the law of entropy is the task of the compressor, an electrically powered device that squeezes gas into a smaller volume. The compressor receives low-pressure gaseous working fluid from the evaporator, compresses it to much higher density, and delivers it as a high-pressure gas to the condenser. The compressor may use a piston and one-way valves, like the water pump in Fig. 5.2.3, or it may use a rotary pumping mechanism. Regardless of how it functions, the result is the same: the gaseous working fluid undergoes a dramatic increase in density and pressure as it passes through the compressor.



Fig. 8.1.7 The compressor (bottom) and condenser coils (top) are visible on the back of this refrigerator. The compressor squeezes the working fluid into a hot dense gas and delivers it to the condenser. There it gives up heat to the room air and condenses into a liquid. The working fluid evaporates inside the refrigerator, extracting heat from the food.

The compressor does work on the gas while compressing it—pushing the gas inward as the gas moves inward. In accordance with the law of conservation of energy, this work increases the internal energy of the working fluid. The only way that a gas can store additional energy is as thermal energy, so the working fluid leaves the compressor much hotter than when it arrived. There is no getting around that temperature rise; compressing a gas unavoidably raises its temperature.

Hot, high-pressure working fluid then flows into the condenser. Like the evaporator, the condenser is a long metal pipe with fins attached to it. It acts as a heat exchanger, and its fins provide extra surface area to speed the flow of heat from the hotter working fluid inside it to the less hot outdoor air. There may also be a fan to move outdoor air quickly past the condenser and speed up the heat transfer.

As its name suggests, the condenser is where the working fluid condenses from gas to liquid. That condensation starts when the gas begins to cool after compression. While it flows through a pipe toward the compressor, the low-pressure working fluid is stable as a gas. But in the high-pressure, high-density working fluid that emerges from the compressor and starts to cool in the condenser, the landing rate outpaces the leaving rate and the gas begins to condense.

As its molecules bind together to form a liquid, the hot gaseous working fluid releases its latent heat of evaporation. This chemical potential energy becomes thermal energy in the working fluid, keeping it hot so that heat continues to flow out of it through the walls of the condenser to the cooler outdoor air.

By the time the working fluid leaves the condenser as a liquid, it has transformed a great deal of chemical potential energy into thermal energy and released that energy into the outdoor air. The outdoor air receives as heat not only the thermal energy extracted from the indoor air but also the electric energy consumed by the compressor. The working fluid exits the condenser as a warm, high-pressure liquid and travels through a pipe toward the evaporator.

The second half of the air conditioner's job is now complete; it has released heat into the outdoor air and, in the process, converted ordered energy into thermal energy. From here, the working fluid returns to the evaporator to begin the cycle all over again. The working fluid passes endlessly around the loop, extracting heat from the indoor air in the evaporator and releasing it to the outdoor air in the condenser. The compressor drives the whole process and thereby satisfies the law of entropy.

In fact, heat pumps of this type also appear in many common appliances. A refrigerator uses a heat pump to extract heat from food and release that heat to the room air (Fig. 8.1.7). A drinking fountain uses a heat pump to transfer heat from water to the room air (Fig. 8.1.8). Like all heat pumps, these cooling devices release more heat to the warm object (room air) than they extract from the cold object (food or water). That's why when you hold the refrigerator door open and it begins to pump heat from one portion of room air to another portion



Fig. 8.1.8 These drinking fountains have heat pumps built into them to cool the water. Heat removed from the water, along with heat produced by the heat-pumping process itself, is released into the room air through the louvers visible on the sides of the drinking fountains.

of room air, the room gets warmer overall—the electric energy the refrigerator consumes while operating is being turned wastefully into thermal energy in the room. To save energy, keep the refrigerator closed whenever possible.

Many homes in moderate climates are heated by effectively running their air conditioners backward. Called heat pumps rather than air conditioners, these systems are capable of pumping heat against its natural direction of flow in winter as well as in summer.

In summer, a heat pump moves heat from indoor air to outdoor air to cool the home. In winter, however, it moves heat from outdoor air to indoor air to heat the home. Rather than turning electricity directly into thermal energy to heat the home, it leverages that electrical energy by using it to gather the abundant thermal energy from outdoors and to carry that indoors.

Before leaving air conditioners, we should take a moment to look at the working fluid itself. This fluid must become a gas at low pressure and a liquid at high pressure, over most of the temperature range encountered by the air conditioner. For decades, the standard working fluids were *chlorofluorocarbons* such as the various Freons. These compounds replaced ammonia, a toxic and corrosive gas used in early refrigeration.

Chlorofluorocarbons are ideally suited to air conditioners because they easily transform from gas to liquid and back again over a broad range of temperatures. They're also chemically inert and inexpensive. Unfortunately, chlorofluorocarbon molecules contain chlorine atoms and, when released into the air, can carry those chlorine atoms to the upper atmosphere. There they promote the destruction of ozone molecules, essential atmospheric constituents that absorb portions of the sun's ultraviolet radiation. Recently, chlorine-free *hydrofluorocarbons* have replaced chlorofluorocarbons as the working fluids in most air conditioners. Though not as energy efficient and chemically inert as the materials they replace, hydrofluorocarbons do not damage the ozone layer. They are, however, potent greenhouse gases (see Section 7.3) and should not be released to the atmosphere.

Check Your Understanding #5: Air Conditioning or Space Heating?

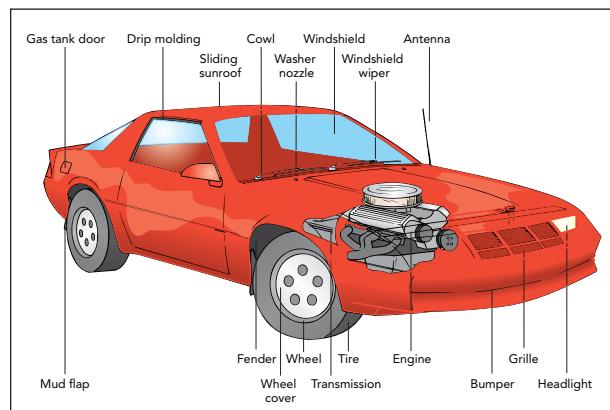
What would happen to the average air temperature in your room if you placed a window air-conditioning unit in the middle of the room and turned it on?

Answer: The room air would become warmer, on average.

Why: The air conditioner would begin to pump heat from its front to its back. The air right in front of the unit would become colder, while the air behind the unit would become hotter. Since the unit would deliver more heat to the hotter air than it would absorb from the colder air, it would increase the total amount of thermal energy in the room. On average, the room would become warmer.

SECTION 8.2

Automobiles



Nothing is more symbolic of freedom and personal independence than an automobile. With its keys in your hand, you can go almost anywhere at a moment's notice. The mechanism that makes this instant transportation possible is the internal combustion engine. Though it has been refined over the years, this engine's basic design has changed little since it was invented more than a century ago. It uses thermal energy released by burning fuel to do the work needed to propel the car forward. That thermal energy can do work at all is one of the marvels of thermal physics and the primary focus of this section.

Questions to Think About: What obstacles stand in the way of using burning fuel's thermal energy to propel a car? Why are two objects, one hot and one cold, required to convert any thermal

energy into useful work? What hot and cold objects does a car have? Why does a car have a cooling system to get rid of waste heat rather than just converting it all into useful work? How does premium, high-octane gasoline differ from regular, low-octane gasoline? Why aren't gasoline and diesel fuel interchangeable?

Experiments to Do: The recent advances in automobile technology and the increasing demands for pollution-control equipment have made automobiles exceedingly complicated. Nonetheless, take a moment to look under the hood of your or

a friend's car. You should be able to identify the engine and its electric support system. You should be able to count four or more spark plug wires heading toward the engine's cylinders. These cylinders convert thermal energy from burning fuel into work to propel the car. Why does the engine need so many cylinders rather than relying on one larger cylinder?

At the front of the engine compartment, you'll find the radiator. How does this device extract waste heat from the engine? Does heat flow naturally into the radiator and then into the outdoor air, or is there a heat pump involved?

Using Thermal Energy: Heat Engines

The light turns green, and you step on the accelerator pedal. The engine of your car roars into action, and in a moment, you're cruising down the road a mile a minute. The engine noise gradually diminishes to a soft purr and vanishes beneath the sound of the radio and the passing wind.

The engine is the heart of the automobile, pushing the car forward at the light and keeping it moving against the forces of gravity, friction, and air resistance. It's not simply a miracle of engineering. It's also a wonder of thermal physics because it performs the seemingly impossible task of converting thermal energy into ordered energy. However, the law of entropy forbids the direct conversion of thermal energy into ordered energy, so how can a car engine use burning fuel to propel the car forward?

The car engine avoids conflict with the law of entropy by being a **heat engine**, a device that converts thermal energy into ordered energy as heat flows from a hot object to a cold object (Fig. 8.2.1). Although thermal energy in a single object can't be converted into work, that restriction doesn't apply to a system of two objects *at different temperatures*. Because heat flowing from the hot object to the cold object increases the overall entropy of the system, a small amount of thermal energy can be converted into work without decreasing the system's overall entropy and without violating the law of entropy.

Another way to look at a heat engine is through the contributions of the two objects. The hot object provides the thermal energy that's converted into work. The cold object provides the order needed to carry out that conversion. As the heat engine operates, the hot object loses some of its thermal energy and the cold object loses some of its order. The heat engine has used them to produce ordered energy. Since the heat engine needs both thermal energy and order, it can't operate if either the hot or the cold object is missing.

In a car engine, the hot object is burning fuel and the cold object is outdoor air. Some of the heat passing from the burning fuel to the outdoor air is diverted and becomes the ordered energy that propels the car. But what limits the amount of thermal energy the engine can convert into ordered energy?

To answer that question, let's examine a simplified car engine. We'll treat the burning fuel and outdoor air as a single, thermally isolated system and look at what happens to their total entropy as the engine operates. In accordance with the law of entropy, this total entropy can't decrease while the engine is transforming some thermal energy into ordered energy.

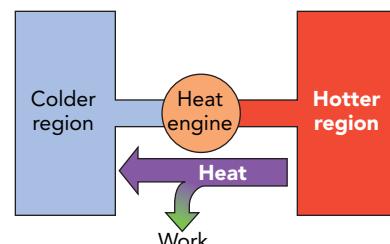


Fig. 8.2.1 A heat engine converts heat (thermal energy from the hotter region) into work (ordered energy) as heat flows from a hotter region to a colder region. The larger the temperature difference between the two regions, the larger the fraction of heat that can be converted into work.

When the car is idling at a stoplight, its engine is doing no work and heat is simply flowing from the hot burning fuel to the cold outdoor air. The system's total entropy increases because this heat is more disordering to the cold air it enters than to the hot burning fuel it leaves. In fact, the system's entropy increases dramatically because the burning fuel is extremely hot compared to the cold outdoor air.

This increase in the system's entropy is unnecessary and wasteful. The law of entropy requires only that the engine add as much entropy to the cold outdoor air as it removes from the hot burning fuel. Since a little heat is quite disordering to cold air, the car engine can deliver much less heat to the outdoor air than it removes from the burning fuel and still not cause the system's total entropy to decrease. As long as the engine delivers enough heat to the outdoor air to keep the total entropy from decreasing, there's nothing to prevent it from converting the remaining heat into ordered energy!

This conversion starts as soon as you remove your foot from the brake and begin to accelerate forward. Instead of transferring all the thermal energy in the burning fuel to the outdoor air, your car then extracts some of it as ordered energy and uses it to power the wheels. The car engine can convert thermal energy into ordered energy, as long as it passes along enough heat from the hot object to the cold object to satisfy the law of entropy.

Obeying the law of entropy becomes easier as the temperature difference between the two objects increases. When the temperature difference is huge, as it is in an automobile engine, a large fraction of the thermal energy leaving the hot object can be converted into ordered energy—at least in theory. For an ideally efficient automobile engine or other heat engine, the relationships among the heat removed from the hot object, the heat added to the cold object, and the work provided can be written as two word equations:

$$\text{work provided} = \text{heat removed from hot object} \cdot \frac{\text{temperature}_{\text{hot}} - \text{temperature}_{\text{cold}}}{\text{temperature}_{\text{hot}}}$$

$$\text{heat added to cold object} = \text{heat removed from hot object} - \text{work provided}, \quad (8.2.1)$$

in symbols:

$$W = -Q_h \frac{T_h - T_c}{T_h}$$

$$Q_c = -Q_h - W,$$

and in everyday language:

The greater the temperature difference between hot and cold, the larger the fraction of heat you can divert and transform into work,

where the temperatures are measured on an absolute scale.

Unfortunately, theoretical limits are often hard to realize in actual machines, and the best automobile engines extract only about half the ordered energy specified by Eqs. 8.2.1. Still, obtaining even that amount is a remarkable feat and a tribute to scientists and engineers who, in recent years, have labored to make automobile engines as energy efficient as possible.

Check Your Understanding #1: Heat Pumps and Heat Engines

An air conditioner uses electric energy to make the air in your home colder than the outdoor air. Could you use this difference in temperatures to operate a heat engine and generate electric energy?

Answer: Yes, you could.

Why: A heat engine is essentially a heat pump operating backward. The air conditioner (a heat pump) uses ordered electric energy to pump heat from the cold air in your home to the hot outdoor air. The heat engine we are considering would use the flow of heat from the hot outdoor air to the cold air in your home to produce ordered electric energy.

Check Your Figures #1: Back to the Steam Age

A train locomotive is powered by an ideal steam engine. If the steam boiler operates at 450 K (177 °C or 350 °F) and the outdoor temperature is 300 K (26 °C or 80 °F), how much work can the steam engine obtain by removing 1200 J of heat from the boiler?

Answer: It can obtain 400 J of work.

Why: Since the hot object is at 450 K and the temperature difference is 150 K, Eqs. 8.2.1 allow $\frac{1}{3}$ of the heat removed from the boiler to be converted into work. The remaining 800 J of heat must flow into the outdoor air.

The Internal Combustion Engine

Invented by the German engineer Nikolaus August Otto in 1867, the internal combustion engine burns fuel directly in the engine itself. Gasoline and air are mixed and ignited in an enclosed chamber. The resulting temperature rise increases the pressure of the gas and allows it to perform work on a movable surface.

To extract work from the fuel, the internal combustion engine must perform four tasks in sequence:

1. It must introduce a fuel-air mixture into an enclosed volume.
2. It must ignite that mixture.
3. It must allow the hot burned gas to do work on the car.
4. It must get rid of the exhaust gas.

In the standard, four-stroke fuel-injected engine found in modern gasoline automobiles, this sequence of events takes place inside a hollow cylinder (Fig. 8.2.2). It's called a *four-stroke* engine because it operates in four distinct steps, or strokes: induction, compression, power, and exhaust. *Fuel-injected* refers to the technique used to mix the fuel and air as they're introduced into the cylinder.

Automobile engines usually have four or more of these cylinders. Each cylinder is a separate energy source, closed at one end and equipped with a movable piston, several valves, a fuel injector, and a spark plug. The piston slides up and down in the cylinder, shrinking or enlarging the cavity inside. The valves, located at the closed end of the cylinder, open to introduce fuel and air into the cavity or to permit burned exhaust gas to escape from the cavity. The fuel injector adds fuel to the air as it enters the cylinder. The spark plug, also located at the closed end of the cylinder, ignites the fuel-air mixture to release its chemical potential energy as thermal energy.

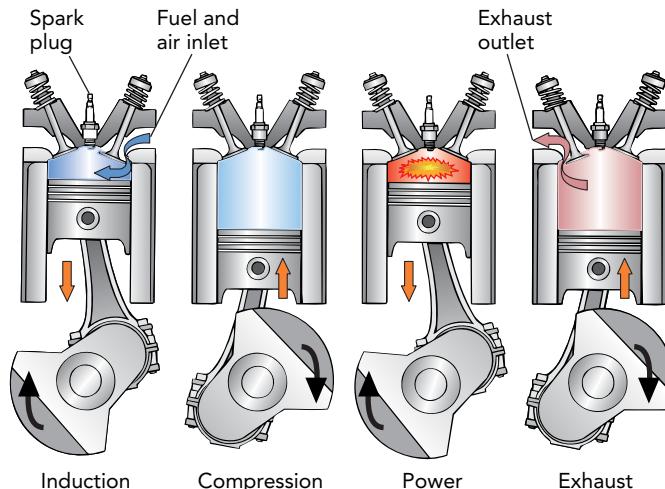


Fig. 8.2.2 A four-stroke engine cylinder. During the induction stroke, fuel and air enter the cylinder. The compression stroke squeezes that mixture into a small volume. The spark plug ignites the mixture, and the power stroke allows the hot gas to do work on the automobile. Finally, the exhaust stroke ejects the exhaust gas from the cylinder.

The fuel-air mixture is introduced into each cylinder during its induction stroke. In this stroke, the engine pulls the piston away from the cylinder's closed end so that its cavity expands to create a partial vacuum. At the same time, the cylinder's inlet valves open so that atmospheric pressure can push fresh air into the cylinder. The cylinder's fuel injector adds a mist of fuel droplets to this air so that the cylinder fills with a flammable fuel-air mixture. Because it takes work to move air out of the way and create a partial vacuum, the engine does work on the cylinder during the induction stroke.

At the end of the induction stroke, the inlet valves close to prevent the fuel-air mixture from flowing back out of the cylinder. Now the compression stroke begins. The engine pushes the piston toward the cylinder's closed end so that its cavity shrinks and the fuel-air mixture becomes denser. The engine does work on the mixture while compressing it—the piston pushes the gas inward as the gas moves inward—so the mixture's internal energy increases. The only way that a gas can store this additional energy is as thermal energy, so the mixture's temperature rises as it's compressed. Since increases in a gas's density and temperature both increase its pressure, the pressure in the cylinder rises rapidly as the piston approaches the spark plug.

At the end of the compression stroke, the engine applies a high-voltage pulse to the spark plug and ignites the fuel-air mixture. The mixture burns quickly to produce hot, high-pressure burned gas, which then does work on the car during the cylinder's power stroke. In that stroke, the gas pushes the piston away from the cylinder's closed end so that its cavity expands and the burned gas becomes less dense. Since the hot gas exerts a huge pressure force on the piston as it moves outward, it does work on the piston and ultimately propels the car. As it does work, the burned gas gives up thermal energy and cools in accordance with the law of conservation of energy. Its density and pressure also decrease. At the end of the power stroke, the exhaust gas has cooled significantly and its pressure is only a few times atmospheric pressure. The cylinder has extracted much of the fuel's chemical energy as work.

The cylinder gets rid of the exhaust gas during its exhaust stroke. In this stroke, the engine pushes the piston toward the closed end of the cylinder while the cylinder's outlet valves are open. Because the burned gas trapped inside the cylinder at the end of the power stroke is well above atmospheric pressure, it accelerates out of the cylinder the moment the outlet valves open. These sudden bursts of gas leaving the cylinders create the “poof-poof-poof” sound of a running engine. Without a muffler on its exhaust pipes, the engine would be loud and unpleasant.

Just opening the outlet valves releases most of the exhaust gas, but the rest is squeezed out as the piston moves toward the cylinder's closed end. The engine again does work on the cylinder as it squeezes out the exhaust gas. At the end of the exhaust stroke, the cylinder is empty and the outlet valves close. The cylinder is ready to begin a new induction stroke.

Check Your Understanding #2: Getting Out More Than You Put In

Why does the burned gas do more work on the piston during the power stroke than the piston does on the unburned fuel-air mixture during the compression stroke?

Answer: The pressure is much higher in the burned gas than in the unburned fuel-air mixture.

Why: The amount of work done on the piston by the gas or done on the gas by the piston depends on the pressure inside the cylinder. The higher that pressure, the more outward force the piston experiences and the more work is done on it as it moves. The sudden rise in pressure that occurs when the fuel-air mixture burns explains why the burned gas does so much work on the piston as it moves outward. That pressure rise is due partly to the rise in temperature and partly to the fragmentation of the fuel and oxygen molecules into more, smaller molecules.

Engine Efficiency

The goal of an internal combustion engine is to extract as much work as possible from a given amount of fuel. In principle, the fuel's chemical potential energy can be converted entirely into work because both are ordered energies. However, it's difficult to convert

chemical potential energy directly into work, so the engine burns the fuel instead. This step is unfortunate, for in burning the fuel, the engine converts the fuel's chemical potential energy directly into thermal energy and produces lots of unnecessary entropy.

But all is not lost. Since the burned fuel is extremely hot, a good fraction of its thermal energy can be converted into ordered energy by diverting some of the heat that flows from the burned fuel to the outdoor air. As we noted earlier, the hotter the burned fuel and the colder the outdoor air, the more ordered energy the engine can extract. To maximize its fuel efficiency, an internal combustion engine obtains the hottest possible burned gas, lets that gas do as much work as it can on the piston, and releases the gas at the coldest possible temperature.

It would be wonderful if, during the power stroke, the burned gas expanded and cooled until it reached the temperature of the outdoor air. The exhaust gas would then leave the engine with the same amount of thermal energy it had when it arrived, and the engine would have extracted all of the fuel's chemical potential energy as work. Unfortunately, that would violate the law of entropy by converting thermal energy completely into ordered energy. As Eqs. 8.2.1 indicate, an operating heat engine always adds some heat to its cold object. In this case, the engine releases the burned gas before it cools to the temperature of the outdoor air. It has no choice; the engine's exhaust must be hot!

An ordinary internal combustion engine, however, is even less fuel efficient than the law of entropy requires. It releases the burned gas while that gas is still well above atmospheric pressure and is therefore still capable of doing work on the piston. By allowing that gas to expand further before release, an improved engine could extract more work from it and thus be more fuel efficient. The Atkinson cycle engine is just such a device, with a power stroke that is longer than its compression stroke. However, even if the engine expanded the burned gas until it reached atmospheric pressure, the temperature of its exhaust would still remain somewhat above the temperature of the outdoor air. Invented in 1882, Atkinson cycle engines have become popular only recently as part of the effort to build more fuel-efficient vehicles.

Regardless of these improvements, real internal combustion engines always waste some energy and extract less work than the law of entropy allows. For example, a fraction of the heat leaks from the burned gas to the cylinder walls and is removed by the car's cooling system. This wasted heat isn't available to produce work. Similarly, sliding friction in the engine wastes mechanical energy and necessitates an oil-filled lubricating system. Overall, real internal combustion engines convert only about 20–30% of the fuel's chemical potential energy into work.

Check Your Understanding #3: Too Much of a Good Thing

What would happen to the pressure of the burned gas in an internal combustion engine if that engine tried to expand it until it cooled to the temperature of the outdoor air?

Answer: The pressure of the burned gas would drop below atmospheric pressure.

Why: As the engine lets the burned gas expand, that gas does work on the piston and its temperature and pressure decrease. By the time the gas reaches atmospheric pressure, it has cooled significantly but it is still hotter than the outdoor air. To cool the burned gas further, the engine would have to expand it further and its pressure would drop below atmospheric. The engine would then be doing work on its piston to create a partial vacuum.

Improving Engine Efficiency

To obtain the hottest possible burned gas, the compression stroke should squeeze the fuel-air mixture into as small a volume as possible. The more tightly the piston compresses the mixture, the higher its density, pressure, and temperature will be before ignition and the hotter the burned gases will be after ignition. Since the efficiency of any heat engine

TABLE 8.2.1 Approximate Ignition Temperatures for the Three Standard Grades of Gasoline During Compression

Octane Number	Approximate Ignition Temperature
87 (regular)	750 °C (1382 °F)
90 (plus)	800 °C (1472 °F)
93 (premium)	850 °C (1562 °F)

increases as the temperature of its hot object increases and since the hot burned gas is the automobile engine's "hot object," its high temperature after ignition is good for efficiency.

The extent to which the cylinder's volume decreases during the compression stroke is measured by its compression ratio, its volume at the start of the compression stroke divided by its volume at the end of the compression stroke. The larger this compression ratio, the hotter the burned gas and the more energy efficient the engine. While normal compression ratios are between 8:1 and 12:1, those in high-compression engines may be as much as 15:1.

Unfortunately, the compression ratio can't be made arbitrarily large. If the engine compresses the fuel-air mixture too much, the flammable mixture will become so hot that it will ignite all by itself. This spontaneous ignition due to overcompression is called pre-ignition or knocking. When an automobile knocks, the gasoline burns before the engine is ready to extract work from it and much of the energy is wasted.

There are two ways to reduce knocking. First, you can mix the fuel and air more uniformly. In a nonuniform mixture, there may be small regions of gas that get hotter or are more susceptible to ignition than others. The fuel-injection technique used in all modern cars provides excellent mixing and also allows a car's computer to adjust the fuel-air mixture for complete combustion and minimal pollution. So unless a car is seriously out of tune, there isn't much room for improvement as far as mixture uniformity is concerned.

Second, you can use the most appropriate fuel. Not all fuels ignite at the same temperature, so you should select a fuel that is able to tolerate your car's compression process without igniting spontaneously. That's exactly what you do when you purchase the proper grade of gasoline. Regular gasoline ignites at a relatively low temperature and is most susceptible to knocking. Premium gasoline ignites at a relatively high temperature and is most resistant to knocking.

Fuels that are more difficult to ignite and more resistant to knocking are assigned higher octane numbers. Regular gasoline has an octane number of about 87, while premium has an octane number of about 93 (Table 8.2.1). Choosing the proper fuel is simply a matter of finding the lowest octane gasoline that your car can use without excessive knocking. A little knocking in the most demanding circumstances is quite acceptable. Most modern well-tuned automobiles work beautifully on regular gasoline. Since only high-performance cars with high-compression engines need premium gasoline, putting anything other than regular gasoline in a normal car is usually a waste of money.

► Check Your Understanding #4: The Only Premium Is the Price

As part of its aggressive advertising campaign, one oil company has begun calling its 93-octane gasoline "the auto elixir." People are flocking to the pumps to fill up even the most ordinary cars with it. Should you join them or merely chuckle?

Answer: Chuckle away.

Why: Like any high-octane gasoline, the elixir is an expensive fuel that has been carefully formulated to be hard to ignite. Although it works wonders for a high-compression engine that would otherwise overheat the fuel-air mixture and cause it to knock, its resistance to ignition is wasted on most ordinary engines.

Diesel Engines and Turbochargers

Since knocking sets the limit for compression ratio, it also sets the limit for efficiency in a gasoline engine. However, diesel engines avoid the knocking problem by separating the fuel and air during the compression stroke (Fig. 8.2.3). Invented by German engineer Rudolph Christian Karl Diesel (1858–1913) in 1896, the diesel engine has no spark plug to ignite the fuel. Instead, it compresses pure air with an extremely high compression ratio of perhaps 20:1 and then injects diesel fuel directly into the cylinder just as the power stroke begins. The fuel ignites spontaneously as it enters the hot compressed air. Unlike an internal combustion engine, which uses a fuel that is difficult to ignite, a diesel engine uses fuel that ignites easily.

Because of its higher compression ratio, a diesel engine burns its fuel at a higher temperature than a standard gasoline engine and can therefore be more energy efficient. It effectively has a hotter “hot object” and can convert a larger fraction of heat into work.

Some gasoline or diesel engines combine fuel injection with a turbocharger. A *turbocharger* is essentially a fan that pumps outdoor air into the cylinder during the induction stroke. By squeezing more fuel-air mixture into the cylinder, a turbocharger increases the engine’s power output. The engine burns more fuel with each power stroke and behaves like a larger engine. The fan of a turbocharger is powered by pressure in the engine’s exhaust system. A nearly identical device called a *supercharger* is driven directly by the engine’s output power.

For an internal combustion engine, turbochargers have a downside: they encourage knocking. As a turbocharger squeezes air into the cylinder, it does work on that air and the air becomes hot. Since the fuel-air mixture enters the engine hot, it may ignite spontaneously during the compression stroke. To avoid knocking in a car equipped with a turbocharger, you may need to use premium gasoline. Some turbocharged cars are equipped with an *intercooler*, a device that removes heat from the air passing through the turbocharger. By providing cool, high-density air to the cylinders, the intercooler reduces the peak temperature of the compression stroke and avoids knocking.

For diesel engines, however, knocking is not an issue. In fact, turbochargers have revolutionized diesel engines, transforming them from sluggish to high performance. Because of their high compression ratios and extremely hot combustion temperatures, diesel engines develop enormous pressures inside their cylinders and tremendous forces on their pistons. As a result, diesel engines can provide large torques and enormous powers, even at low engine speeds.

With higher fuel efficiency and enhanced performance, these modern diesel engines have a clear environmental advantage over gasoline engines. However, they still had to

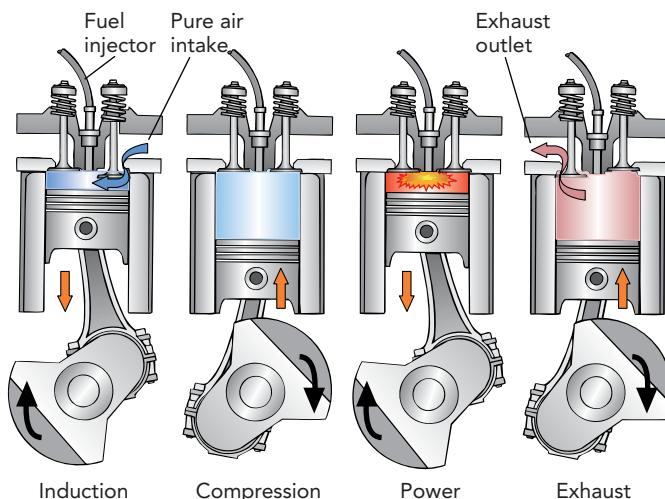


Fig. 8.2.3 A diesel engine cylinder contains pure air during the compression stroke. As the piston does work on it, this air becomes extremely hot. At the start of the power stroke, diesel fuel is injected into the cylinder. The fuel ignites spontaneously, and the hot burned gas does work on the piston and engine during the power stroke.

solve a long-standing problem with diesel engines—particulate pollution. Because a diesel engine must burn its oily liquid fuel in about 1/1000th of a second, it was difficult to burn the fuel completely and some of it became carbon soot. To eliminate that soot, diesel engines were reengineered to swirl their fuel and air vigorously for more complete burning and they now incorporate filters that capture and burn up any soot that does manage to get into the exhaust system. The result is a relatively fuel-efficient vehicle with environmental specifications that rival hybrid vehicles.

COMMON MISCONCEPTIONS: The Hydrogen Economy

Misconception: One day soon, hydrogen will eliminate the need for other sources of energy.

Resolution: Since hydrogen doesn't occur naturally on Earth, it is not a primary energy source. Hydrogen is a secondary energy source, meaning that it is produced using a primary source of energy. Although hydrogen can be made using solar, wind, or hydroelectric energy, most hydrogen is now produced using natural gas or coal. Hydrogen itself is neither renewable nor nonrenewable; what matters is which energy source you use to make it.

Check Your Understanding #5: Steam Heat

How can a steam engine be more energy efficient when it operates on 325 °C steam than when it uses 300 °C steam?

Answer: Like all heat engines, the steam engine can convert more thermal energy into work when the temperature of its hot object (the steam) increases.

Why: The steam engine converts thermal energy into work as heat flows from the hot steam to the outdoor air. The greater the temperature difference between those two objects, the more efficient the steam engine can be at turning thermal energy into work. That is why most steam engines use extremely hot steam.

Multicylinder Engines

Since the purpose of the engine is to extract work from the fuel-air mixture, it's important that each cylinder do more work than it consumes. Three of the strokes require the engine to do work on various gases, and only one of the strokes extracts work from the burned gas. During the induction stroke, the engine does work drawing the fuel-air mixture into the cylinder. During the compression stroke, the engine does work compressing the fuel-air mixture. During the exhaust stroke, the engine does work squeezing the exhaust gas out of the cylinder. Fortunately, the work done on the engine by the hot burned gas during the power stroke more than makes up for the work the engine does during the other three strokes.

Still, the engine has to invest a great deal of energy into the cylinder before each power stroke. To provide this initial energy, most four-stroke engines have four or more cylinders, timed so that there is always one cylinder going through the power stroke. The cylinder that is in the power stroke provides the work needed to carry the other cylinders through the three nonpower strokes, and there is plenty of work left over to propel the car itself.

While the pistons move back and forth, the engine needs a rotary motion to turn the car's wheels. The engine converts each piston's reciprocating motion into rotary motion by coupling that piston to a crankshaft with a connecting rod. The crankshaft is a thick steel bar, suspended in bearings, that has a series of pedal-like extensions, one for each cylinder. As the piston moves out of the cylinder during the power stroke, it pushes on the connecting rod and the connecting rod pushes on its crankshaft pedal. The connecting rod thus produces a torque on the crankshaft. The crankshaft rotates in its bearings and transmits this torque out of the engine so that it can be used to propel the car. So, while each cylinder initially exerts a force, the crankshaft uses that force to produce a torque.

The spinning crankshaft conveys its rotary power to the car's transmission, and from there the power moves on to the wheels. Overall, a significant portion of the heat flowing out of the burning fuel-air mixture is being converted into work and used to spin the car's wheels. Assisted by friction with the pavement, the wheels push the car forward, and you cruise down the highway toward your destination.



Check Your Understanding #6: Hard Starting

Modern cars use an electric motor to start the engine turning, but early cars were started with a hand crank. Why was it so hard to turn the crank?

Answer: The person turning the crank had to do all the work needed to move the engine's pistons through the three nonpower strokes.

Why: Before the engine started running on its own, it couldn't provide any of the energy the cylinders needed during the induction, compression, and exhaust strokes. The person turning the crank had to provide this energy. Once the fuel started burning, the power strokes could take over, but up to that point, turning the crank was hard work.

Epilogue for Chapter 8

This chapter has examined two devices that control the flow of heat to accomplish challenging tasks. In Air Conditioners, we learned how heat pumps use ordered energy to pump heat against its natural flow and observed that the only way to get rid of thermal energy is to transfer it to something else. In Automobiles, we saw that heat engines are able to divert some of the heat flowing from a hot object to a cold object and convert it into useful work. We also examined the roles of the two objects in a heat engine, hot and cold, finding that the hot object provides the energy needed to do the work, while the cold object provides the order that makes the conversion of thermal energy into ordered energy possible.

Explanation: Making Fog in a Bottle

Even when everything in the room is at a single temperature, you can still use thermodynamics to cool the air in the bottle below room temperature. When you step on the bottle, you compress the air inside it and do work on that air. Since air can accommodate added energy only as thermal energy, its temperature rises. The air in the bottle becomes hotter, and heat flows out of the warmer bottle into the cooler room. After half a minute or so, the temperature of air in the bottle returns to the temperature of the room.

When you step off the bottle, the air inside it expands and does work on you. Since air can obtain that work only from its thermal energy, its temperature drops. The air in the bottle becomes colder, and its relative humidity suddenly exceeds 100%. As a result, moisture inside the bottle condenses into droplets and the air becomes foggy.

Smoke particles assist the droplet formation by acting as seeds for the droplets. As we saw in Chapter 7, water can't boil without seed bubbles. Similarly, steam can't condense in air without seed droplets. When the seed droplets fail to form spontaneously, the smoke gives them a little help.

Chapter Summary and Important Laws and Equations

How Air Conditioners Work: An air conditioner moves heat against its natural direction of flow, using a working fluid that passes endlessly through an evaporator, a compressor, and a condenser. The working fluid flows toward the evaporator as a stable high-pressure liquid. Just before pouring into the evaporator, this liquid passes through a constriction in the pipe and

experiences a large drop in pressure. Since the working fluid is not stable as a low-pressure liquid, it evaporates rapidly in the evaporator and thereby absorbs a great deal of heat from the indoor air.

The working fluid then flows to the compressor as a stable low-pressure gas. The compressor squeezes it into a hot, high-density, high-pressure gas and blows it into the condenser. Since the working fluid is not stable as a high-pressure gas, it condenses rapidly in the condenser and thereby releases a large amount of heat to the outdoor air. The resulting liquid working fluid then returns toward the evaporator to begin the cycle again.

To propel this process, the compressor consumes ordered energy and delivers it as heat to the outdoor air. Without this input of ordered energy, the air conditioner could not move heat against its natural direction of flow.

How Automobiles Work: An automobile engine extracts work from its chemical fuel by burning that fuel inside its cylinders and making the resulting burned gas do work on the engine. Most engines have at least four cylinders, each of which requires four strokes to extract work from the fuel. During the induction stroke, a piston moves out of the cylinder and draws a mixture of fuel and air into the resulting cavity. During the compression stroke, the piston moves into the cylinder, compressing this fuel-air mixture to high density, pressure, and temperature. An electric spark then ignites the mixture and it burns to form extremely hot burned gas. During the power stroke, the piston again moves out of the cylinder while the hot gas does work on it. This work is what powers the car. Finally, during the exhaust stroke, the piston moves into the cylinder and squeezes out the burned gas. The cylinder then begins again with fresh fuel and air.

1. Law of thermal equilibrium: Two objects that are each in thermal equilibrium with a third object are also in thermal equilibrium with one another.

2. Law of conservation of energy: The change in a stationary object's internal energy is equal to the heat transferred into that object minus the work that object does on its surroundings or

$$\text{internal energy change} = \text{heat added} - \text{work done}. \quad (8.1.1)$$

3. Law of entropy: The entropy of a thermally isolated system of objects never decreases.

4. Efficiency of a heat pump: The ideal efficiency of a heat pump depends on the temperatures of its hot and cold objects:

$$\text{heat from cold object} = \text{work} \cdot \frac{\text{temp}_{\text{cold}}}{\text{temp}_{\text{hot}} - \text{temp}_{\text{cold}}}$$

$$\text{heat to hot object} = \text{heat from cold object} + \text{work}. \quad (8.1.2)$$

5. Efficiency of a heat engine: The ideal efficiency of a heat engine depends on the temperatures of its hot and cold objects:

$$\text{work} = \text{heat from hot object} \cdot \frac{\text{temp}_{\text{hot}} - \text{temp}_{\text{cold}}}{\text{temp}_{\text{hot}}}$$

$$\text{heat to cold object} = \text{heat from hot object} - \text{work}. \quad (8.2.1)$$

Resonance and Mechanical Waves

Many fascinating motions in the world around us are repetitive ones. Our lives are filled with cycles, from the sun's daily passage overhead to a pond's undulating ripples on a rainy day. These cyclic motions are governed by the physical laws and steadily mark our journey through time and space. Some of these cycles structure our lives out of necessity or tradition, while others are simply there to be observed. Still other cycles have become part of our everyday world because they're useful or enjoyable. In this chapter, we cover those three possibilities by examining cyclic motions in three contexts: in clocks, in musical instruments, and at the seashore.

ACTIVE LEARNING EXPERIMENTS

A Singing Wineglass

One simple experiment with cyclic motion involves a crystal wineglass, a little water, and a delicate touch. A crystal wineglass is hard and thin and easily supports a repetitive mechanical motion called a vibration. The glass's hardness allows it to retain energy in this vibrating motion for a long time so that it rings clearly when you

tap it gently with a spoon and to acquire energy slowly in the manner described next.

The experiment itself is an activity discovered by many a bored youth, sitting too long in a fancy restaurant with only the place settings as entertainment. If you wet your finger slightly and run it gently around the lip of the



Courtesy Lou Bloomfield

wineglass at a slow, steady pace, you should be able to get the wineglass to sing loudly. The walls of the glass will vibrate back and forth and emit a clear tone.

If you can't find a crystal wineglass, you may be out of luck. An ordinary glass doesn't work as well because it converts the energy from your finger into thermal energy

rapidly and doesn't emit much sound. In any case, be sure that the glass has no sharp edges on its lip so that you don't cut your finger.

What will happen to the sound if you add some water to the glass? Try it and see what happens. Is anything visible on the surface of the water as you make the glass vibrate?

Chapter Itinerary

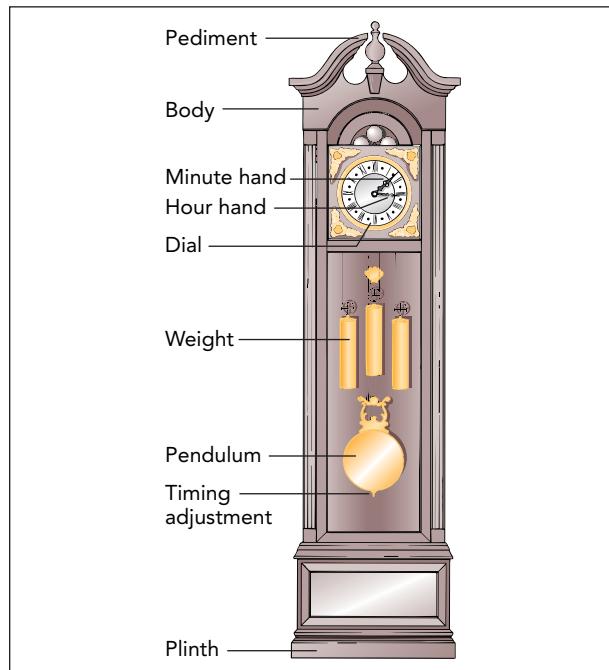
The wineglass's vibration is an example of a natural resonance, a cyclic motion of the glass itself. Whether you tap the glass gently with a spoon or rub it with your finger, you're causing it to undergo this characteristic motion. In this chapter we examine several other objects that exhibit natural resonances: (1) *clocks*, (2) *musical instruments*, and (3) *the sea*.

In Clocks, we study pendulums, balance rings, and quartz crystals to see how their natural resonances are used to measure

the passage of time. In Musical Instruments, we look at the vibrations of strings, air columns, and drumheads to find out how these instruments produce their musical sounds and to learn how those sounds travel to our ears. Last, in The Sea, we look at the nature of waves on the surface of water and explore such issues as tides, tsunamis, and surf. For a more detailed preview, turn to the Chapter Summary and Important Laws and Equations at the end of the chapter.

SECTION 9.1

Clocks



People measure their lives according to the sky, dividing existence into days, months, and years according to the celestial motions of the sun, moon, and stars. On the less romantic scale of daily life, the sky offers little help. Since it provides no easy way to measure short periods of time, people invented clocks.

Early clocks were based on the time it took to complete simple processes—the flow of sand or water, or the burning of candles. However, these clocks had limited accuracy and

required constant attention. Better clocks measure time with repetitive motions such as swinging or rocking. In this section, we'll examine the workings of modern clocks based on repetitive motions. As we do, we'll see that repetitive motions are interesting in their own right and appear throughout nature in countless objects in addition to clocks.

Questions to Think About: What exactly is time? Why do some objects swing or rock back and forth repetitively? How can you use a repetitive motion to measure time? How can you change the rate at which an object swings or rocks? Since repetitive motions don't normally continue forever, how can you keep them going without upsetting their timekeeping ability?

Experiments to Do: You can build the timekeeping portion of a pendulum clock by attaching a small, dense object to the end of a string and hanging that string from a table or doorway. Push the object gently so that it swings back and forth. You'll find that it completes this repetitive motion with great regularity. What limits its regularity?

If the string is about 25 cm (10 in) long, a complete swing (back and forth) will take almost exactly 1 s. Change the length of the string and observe its effect on the swing. Do you think that the object's weight affects the swing? Try a different object and see whether you're right.

Now vary the extent of the swing to find out whether it affects the time each swing takes. How can it be that a small swing takes the same time as a large swing? Notice that you can get the object swinging by pushing on it rhythmically. Randomly timed pushes won't do; you must push the object in synchrony with its motion. At what times should you push it to make it swing farther? To make it swing less far? Rhythmic pushes of this sort keep the timekeeper in a clock moving, hour after hour.

Time

Before examining clocks, we should take a brief look at time itself. Scientists treat **time** as a dimension, similar but not identical to the three spatial dimensions that we perceive in the world around us. In total, our universe has four dimensions: three spatial dimensions and one temporal dimension. Thus it takes four numbers to completely specify when and where an event occurs: three numbers identify the event's location and one identifies its moment in time.

An obvious difference between **space** and time is that, while we can see space stretched out around us, we can observe only the *passage* of time. Although we occupy only one location in space at a given moment, we are somehow more aware of the expanse of space around us. It's much harder to sense the whole framework of time stretching off into the past and future; you must use your imagination.

Another difference between time and space is that, while you can turn around and walk back the way you came, you can't grow any younger. Our travels through space are reversible, but our travels through time are one-way. That irreversibility of time is associated with entropy and the law of entropy (see Chapter 8), particularly the ever-increasing disorder of the universe. Disorder increases, we get older, the weeds grow around our homes, and even clocks run down.

Our perception of space is ultimately based on the need for forces, accelerations, and velocities to travel from one place to another. A city seems far away because we know that traveling there with reasonable forces, accelerations, and velocities will take a long time. Our perception of time is based on the same mechanical principles. If two moments are separated by a long time, then reasonable forces, accelerations, and velocities will permit us to travel large distances between the two moments. In short, our perceptions of space and time are interrelated, and measurements of time and space are connected as well.

We measure space with rulers and time with clocks. How would you make a ruler? You could construct a rather large ruler by driving a car at constant velocity and marking the pavement with paint once each second. Your ruler wouldn't be very practical, but it would fit the definition of a ruler as having spatial markings at uniform distances. You would be using your movement through time to measure space.

How would you make a clock? You could make a rather strange clock by driving a car at constant velocity down your giant ruler and counting each time you see one of your marks go by. You would then be using your movement through space to measure time. Most clocks really do use motion to measure time. As we are about to see, however, they use motions that are a bit more compact than a car ride.



Check Your Understanding #1: The Ultimate Moon Bounce

To measure the distance from Earth to the moon, scientists bounce light from reflectors placed on the moon by the Apollo astronauts. Light travels at a constant speed. How can a measurement of light's travel time to and from the moon be used to determine the distance from Earth to the moon?

Answer: Since light travels at a constant speed, the distance it travels is equal to its speed times its travel time. If you know the travel time and the speed, you can determine that distance.

Why: Many distance measurements are made by measuring time. Surveyors routinely use light's travel time to measure distances. Decorators and architects often use sound's travel time to measure the distances between walls. In general, the motion of an object at constant velocity can be used either to measure the distance traveled if you know the elapsed time or to measure the elapsed time if you know the distance traveled.

Natural Resonances

An ideal timekeeping motion should offer both accuracy and convenience. That rules out some of the obvious choices. The sun, moon, and stars keep excellent time but fail the convenience test. Sure, conservation of energy, momentum, and angular momentum so

dominate their motions that these celestial bodies move steadily and predictably through the heavens, century after century, but what do you do on a cloudy day? And while simple interval timers like sandglasses and burning candles are easy to make and use, they're not very accurate. Besides, who's going to stay up all night lighting fresh candles just to keep the "clock" running? (For an interesting astronomical clock, see 1.)

Instead, practical clocks are based on a particular type of repetitive motion called a **natural resonance**. In a natural resonance, the energy in an isolated object or system of objects causes it to perform a certain motion over and over again. Many objects in our world exhibit natural resonances, from tipping rocking chairs to sloshing basins of water to waving flagpoles, and those natural resonances usually involve motion about a stable equilibrium. Like the bouncing spring scale in Section 3.1, an object that has been displaced from its stable equilibrium accelerates toward that equilibrium but then overshoots; it coasts right through equilibrium and must turn around to try again. As long as it has excess energy, this object continues to glide back and forth through its equilibrium and thus exhibits a natural resonance.

Some resonances, such as those of bouncing balls and teetering bottles, don't maintain a steady beat and aren't suitable for clocks. However, we are about to encounter a group of resonances that are extremely regular and that can be used to measure the passage of time with remarkable accuracy. Those resonances belong to an important class of mechanical systems known as *harmonic oscillators*.

1 The daughter of a planetarium architect, Jocelyn Bell (British astronomer, 1943–) acquired an early interest in radio astronomy. Advised to study physics first, she became the only woman in a class of 50 at Glasgow University. While working on her Ph.D. at Cambridge, Bell discovered an extraterrestrial source of radio bursts, occurring precisely 1.33730113 seconds apart. She had discovered the first pulsar, a collapsed star whose angular momentum keeps it turning at an extraordinarily uniform rate. Each burst coincided with one rotation of the star remnant.

Check Your Understanding #2: An Egg-Timer Clock

Although a sandglass can be made repetitive by turning it over every time the sand runs out, this manual restarting process introduces timing errors. If a 3-min sandglass is always turned over within 10 s after the sand runs out, how accurately will it measure time over the course of a day?

Answer: It may have lost as much as 80 min after 24 h.

Why: If the operator of a sandglass waits 10 s before turning it over each time the sand runs out, the clock will lose 10 s every 3 min, 200 s every h, and 4800 s every day. If you could be sure that the operator would always wait 10 s, you could include it in the clock's design. However, the operator might be faster one time than the next, and this uncertainty makes the clock unreliable and inaccurate. For a repetitive clock to keep accurate time, it must repeat its motion with almost perfect regularity.

Pendulums and Harmonic Oscillators

One of the first natural resonances to find its way into clocks is the swing of a pendulum, a weight hanging from a pivot (Fig. 9.1.1). When the pendulum's center of gravity is directly below its pivot, it's in a stable equilibrium. Its center of gravity is then as low as possible, so displacing it raises its gravitational potential energy and a restoring force begins pushing it back toward that equilibrium position (Fig. 9.1.2). For geometrical reasons, this restoring force is almost exactly proportional to how far the pendulum is from equilibrium. As you displace the pendulum steadily from equilibrium, the restoring force on it also increases steadily.

When you release the displaced pendulum, its restoring force accelerates it back toward equilibrium. Instead of stopping, however, the pendulum swings back and forth about its equilibrium position in a repetitive motion called an **oscillation**. As it swings, its energy alternates between potential and kinetic forms. When it swings rapidly through its equilibrium position in the middle of a swing, its energy is all kinetic. When it stops momentarily at the end of a swing, its energy is all gravitational potential. This repetitive transformation of excess energy from one form to another is part of any oscillation and keeps the oscillator—the system experiencing the oscillation—moving back and forth until that excess energy is either converted into thermal energy or transferred elsewhere.

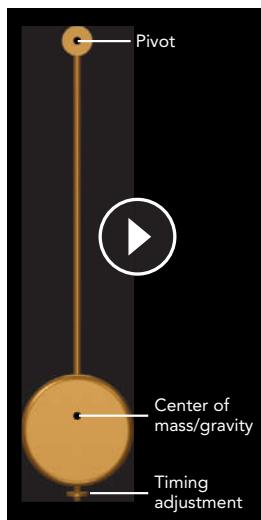
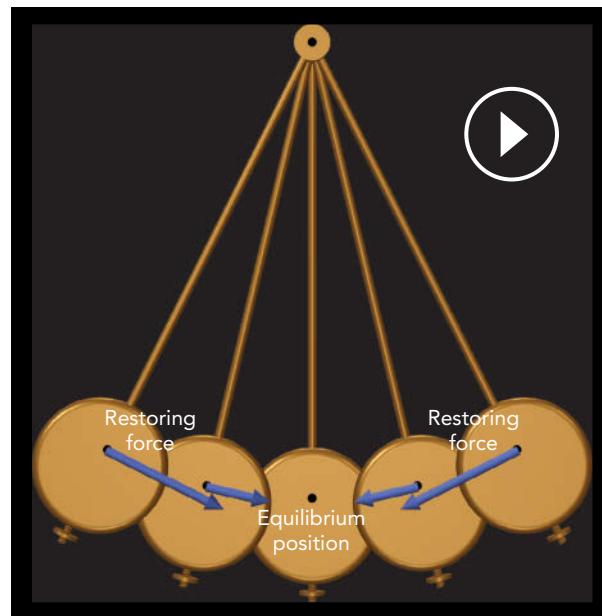


Fig. 9.1.1 A pendulum consists of a weight hanging from a pivot. The pendulum is in a stable equilibrium when its center of gravity is directly below the pivot.

Fig. 9.1.2 If you tilt a pendulum's center of gravity away from its equilibrium position, it experiences a restoring force proportional to its distance from that equilibrium position.



The pendulum isn't just any oscillator, though. Because its restoring force is proportional to its displacement from equilibrium, the pendulum is a **harmonic oscillator**—the simplest and best understood mechanical system in nature. As a harmonic oscillator, the pendulum undergoes **simple harmonic motion**, a regular and predictable oscillation that makes it a superb timekeeper.

The **period** of any harmonic oscillator, the time it takes to complete one full cycle of its motion, depends only on how stiffly its restoring force pushes it back and forth, and on how strongly its inertia resists the accelerations of that motion. **Stiffness** is the measure of how sharply the restoring force increases as the oscillator is displaced from equilibrium; stiff restoring forces are associated with firm or hard objects, while less stiff restoring forces are associated with soft objects. The stiffer the restoring force, the more forcefully it pushes the oscillator back and forth and the shorter the oscillator's period. On the other hand, the larger the oscillator's mass, the less it accelerates and the longer its period.

However, the most remarkable and important characteristic of a harmonic oscillator is not that its period depends on stiffness and mass, but that its period *doesn't* depend on **amplitude**, its furthest displacement from equilibrium. Whether that amplitude is large or small, the harmonic oscillator's period remains exactly the same. This insensitivity to amplitude is a consequence of its special restoring force, a restoring force that is proportional to its displacement from equilibrium. At larger amplitudes, the oscillator travels farther each cycle, but the forces accelerating it through that cycle are stronger as well. Overall, the harmonic oscillator completes a large cycle of motion just as quickly as it completes a small cycle of motion.

Any harmonic oscillator can be thought of as having a restoring force component that drives the back and forth motion and an inertial component that resists that motion. Their competition determines the oscillator's period. Harmonic oscillators with stiff restoring forces and little inertia have short periods, while those with soft restoring forces and great inertia have long periods. Their amplitudes of oscillation simply don't affect their periods, which is why harmonic oscillators are so ideal for timekeeping. Because practical clocks can't control the amplitudes of their timekeeping oscillators perfectly, virtually all of them are based on harmonic oscillators.

HARMONIC OSCILLATORS

A harmonic oscillator is an oscillator with a restoring force proportional to its displacement from equilibrium. Its period of oscillation depends only on the stiffness of that restoring force and on its mass, not on its amplitude of oscillation.

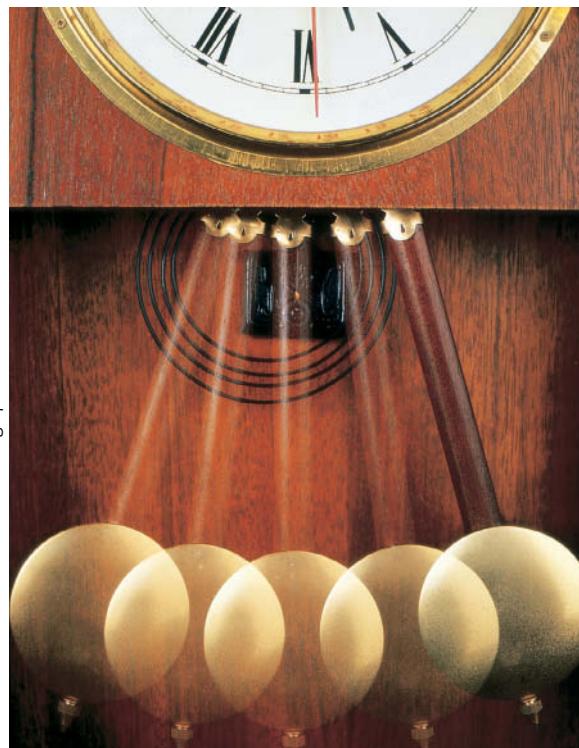
Actually, a pendulum is an unusual harmonic oscillator because increasing its mass doesn't increase its period. That's because increasing the pendulum's mass also increases its weight and therefore stiffens its restoring force. These two changes compensate for one another perfectly so that the pendulum's period is unchanged.

A pendulum's period does, however, depend on its length and on gravity. When you reduce the pendulum's length, the distance from its pivot to its center of mass, you stiffen its restoring force and shorten its period. Similarly, when you strengthen gravity (perhaps by traveling to Jupiter), you increase the pendulum's weight, stiffen its restoring force, and reduce its period. Though we won't try to prove it, the period of a pendulum is

$$\text{period of pendulum} = 2\pi \sqrt{\frac{\text{length of pendulum}}{\text{acceleration due to gravity}}}.$$

Thus a short pendulum swings more often than a long one, and any pendulum swings more often on Earth than it would on the moon.

On Earth's surface, a 0.248-m (10-in) pendulum has a period of 1 s (Fig. 9.1.3), making it suitable for a wall clock that advances its second hand by 1 s each time the pendulum completes a cycle. Since a pendulum's period increases as the square root of its length, a 0.992-m (40-in) pendulum (four times as tall as a 0.248-m pendulum) takes 2 s to complete its cycle and is appropriate for a floor clock that advances its second hand by 2 s per cycle.



© Robert Mathena/Fundamental Photographs

Fig. 9.1.3 The swinging pendulum controls the movement of the clock's hands. The pendulum is 0.248 m long, from pivot to center of mass/gravity, so each cycle takes 1 s to complete and advances the hands by 1 s.

Because a pendulum's period depends on its length and gravity, a change in either one causes trouble. As we learned in Chapter 7, materials expand with increasing temperature, so a simple pendulum slows down as it heats up. A more accurate pendulum is thermally compensated by using several different materials with different coefficients of loudness expansion to ensure that its center of mass remains at a fixed distance from its pivot.

Although gravity doesn't change with time, it does vary slightly from place to place. To correct for differences in gravity between the factory and a clock's final destination, its pendulum has a threaded adjustment knob. This knob allows you to change the pendulum's length to fine-tune its period.

Check Your Understanding #3: Swing Time

A child swinging on a swing set travels back and forth at a steady pace. What determines the period of the child's motion?

Answer: The strength of Earth's gravity and the length of the swing's chains.

Why: A child swinging on a swing set is a form of pendulum. As with any pendulum, the child's period of motion is determined only by the strength of gravity and the length of the pendulum. In this case, the length of the pendulum is approximately the length of the swing's supporting chains. Thus, a tall swing has a longer period than a short swing.

Pendulum Clocks

Although a pendulum maintains a steady beat, it's not a complete clock. Something must keep the pendulum swinging and use that swing to determine the time. A pendulum clock does both. It sustains the pendulum's motion with gentle pushes, and it uses that motion to advance its hands at a steady rate.

The top of the pendulum has a two-pointed anchor that controls the rotation of a toothed wheel (Fig. 9.1.4). This mechanism is called an *escapement*. A weighted cord wrapped around the toothed wheel's shaft exerts a torque on that wheel, so that the wheel would spin if the anchor weren't holding it in place. Each time the pendulum reaches the end of a swing, one point of the anchor releases the toothed wheel while the other point catches it. The wheel turns slowly as the pendulum rocks back and forth, advancing by one tooth for each full cycle of the pendulum. This wheel turns a series of gears, which slowly advance the clock's hands. Although these hands are actually counting the number of pendulum swings since midnight, their movement is calibrated so that their positions indicate the current time.

The toothed wheel also keeps the pendulum moving by giving the anchor a tiny forward push each time the pendulum completes a swing. Since the anchor moves in the direction of the push, the wheel does work on the anchor and pendulum, and replaces energy lost to friction and air resistance. This energy comes from the weighted cord, which releases gravitational potential energy as its weight descends. When you wind the clock, you rewind this cord around the shaft, lifting the weight and replenishing its potential energy.

While these pushes from the toothed wheel can keep even the clumsiest pendulum swinging, a clock works best when its pendulum swings with almost perfect freedom. That's because any outside force—even the push from the toothed wheel—will influence the pendulum's period. The most accurate timekeepers are those that can oscillate without any assistance or energy replacement for thousands or millions of cycles. These precision timekeepers need only the slightest pushes to keep them moving and thus have extremely precise periods. That's why a good pendulum clock uses an aerodynamic pendulum and low-friction bearings.

Finally, the clock must keep the oscillation amplitude of its pendulum relatively constant. From a practical perspective, drastic changes in that amplitude will make the toothed wheel turn erratically. However, there is a more fundamental reason to keep the pendulum's

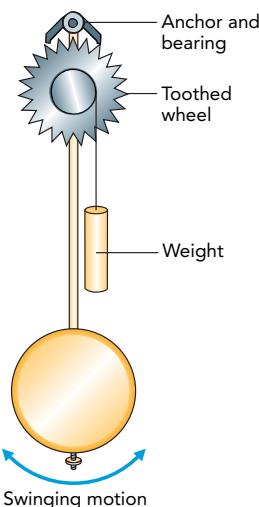


Fig. 9.1.4 A pendulum clock uses a swinging pendulum to determine how quickly a toothed wheel turns and advances a series of gears that control the hands of the clock. The anchor permits the toothed wheel to advance by one tooth each time the pendulum completes a full cycle.

amplitude steady: it's not really a perfect harmonic oscillator. If you displace the pendulum too far, it becomes an **anharmonic oscillator**—its restoring force ceases to be proportional to its displacement from equilibrium, and its period begins to depend on its amplitude. Since a change in period will spoil the clock's accuracy, the pendulum's amplitude must be kept small and steady. That way, the amplitude has almost no effect on the pendulum's period.

Check Your Understanding #4: Swing High, Swing Low

When pushing a child on a playground swing, you normally push her forward as she moves away from you. What happens if you push her forward each time she moves toward you?

Answer: The amplitude of her motion will gradually decrease so that she comes to a stop.

Why: To keep her swinging, you must make up for the energy she loses to friction and air resistance. By pushing her forward each time she moves away from you, you do work on her and increase her energy. However, when you push her as she moves toward you, she does work on you and you extract some of her energy. You are then slowing her down rather than sustaining her motion.

Balance Clocks

Because it relies on gravity for its restoring force, a swinging pendulum mustn't be tilted or moved. That's why there are so few pendulum-based wristwatches. To make use of the excellent timekeeping characteristics of a harmonic oscillator, a portable clock needs some other restoring force that's proportional to displacement but independent of gravity. It needs a spring!

As we saw in Section 3.1, the force a spring exerts is proportional to its distortion. The more you stretch, compress, or bend a spring, the harder it pushes back toward its equilibrium shape. Attach a block of wood to the free end of a spring, stretch it gently, and let go, and you'll find you have a harmonic oscillator with a period determined only by the stiffness of the spring and the block's mass (Fig. 9.1.5). Since the period of a harmonic oscillator doesn't depend on the amplitude of its motion, the block oscillates steadily about its equilibrium position and makes an excellent timekeeper.

Unfortunately, gravity complicates this simple system. Although gravity doesn't alter the block's period, it does shift the block's equilibrium position downward. That

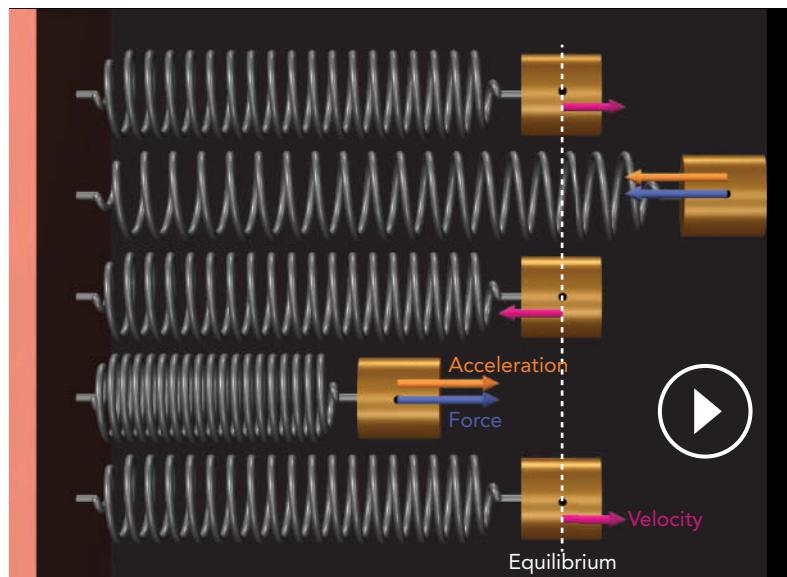


Fig. 9.1.5 A metal cylinder attached to a spring is a harmonic oscillator, shown here in a time sequence (top to bottom). The oscillator's period is determined only by the stiffness of the spring and the cylinder's mass.

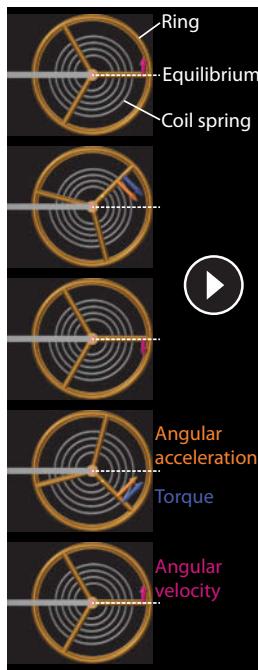


Fig. 9.1.6 A small wheel attached to a coil spring is a harmonic oscillator known as a balance ring, shown here in a time sequence (top to bottom). Its period is determined only by the stiffness of the coil spring and the rotational mass of the wheel.

shift is a problem for a clock that might be tilted sometimes. However, there's another spring-based timekeeper that marks time accurately in any orientation or location. This ingenious device, used in most mechanical clocks and watches, is called a balance ring or simply a balance.

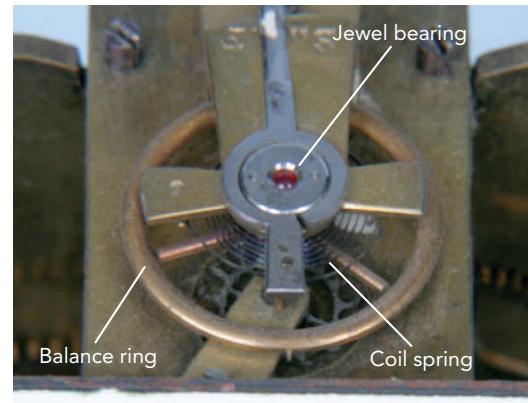
A balance ring resembles a tiny metal bicycle wheel, supported at its center of mass/gravity by an axle and a pair of bearings (Fig. 9.1.6). Any friction in the bearings is exerted so close to the ring's axis of rotation that it produces little torque and the ring turns extremely easily. Moreover, the ring pivots about its own center of gravity so that its weight produces no torque on it. In keeping with its name, the balance is balanced.

The only thing exerting a torque on the balance ring is a tiny coil spring. One end of this spring is attached to the ring, while the other is fixed to the body of the clock. When the spring is undistorted, it exerts no torque on the ring and the ring is in equilibrium. If you rotate the ring either way, however, torque from the distorted spring will act to restore it to its equilibrium orientation. Since this restoring torque is proportional to the ring's rotation away from a stable equilibrium, the balance ring and coil spring form a harmonic oscillator!

Because of the rotational character of this harmonic oscillator, its period depends on the *torsional* stiffness of the coil spring, that is, on how rapidly the spring's torque increases as you twist it, and on the balance ring's *rotational* mass. Since the balance ring's period doesn't depend on the amplitude of its motion, it keeps excellent time. Also, because gravity exerts no torque on the balance ring, this timekeeper works anywhere and in any orientation.

The rest of a balance clock is similar to a pendulum clock (Fig. 9.1.7). As the balance ring rocks back and forth, it tips a lever that controls the rotation of a toothed wheel. An anchor attached to the lever allows the toothed wheel to advance one tooth for each complete cycle of the balance ring's motion. Gears connect the toothed wheel to the clock's hands, which slowly advance as the wheel turns.

Because the balance clock is portable, it can't draw energy from a weighted cord. Instead, it has a main spring that exerts a torque on the toothed wheel. This main spring is a coil of elastic metal that stores energy when you wind the clock. Its energy keeps the balance ring rocking steadily back and forth and also turns the clock's hands. Since the



Courtesy Lou Bloomfield

Fig. 9.1.7 The balance ring in this antique French carriage clock twists back and forth rhythmically under the influence of the spiral spring near its center. The tiny ruby bearings that support the ring minimize friction and permit this clock to keep very accurate time.



main spring unwinds as the toothed wheel turns, the clock occasionally needs winding. (For two interesting examples of balance clocks, see [2](#) and [3](#).)

Check Your Understanding #5: A Little Light Entertainment

When you accidentally strike a chandelier with a broom, this hanging lamp begins to twist back and forth with a regular period. What determines its period of oscillation?

Answer: It's determined by the torsional stiffness of the chandelier's supporting cord and the chandelier's rotational mass.

Why: The hanging chandelier is a harmonic oscillator. Its supporting cord opposes any twists by exerting a restoring force on the chandelier. Once you twist it away from its equilibrium orientation, the chandelier oscillates back and forth with a period determined only by the cord's torsional stiffness (its stiffness with respect to twists) and the chandelier's rotational mass. As with any harmonic oscillator, the amplitude of the chandelier's motion doesn't affect its period.

Electronic Clocks

The potential accuracy of pendulum and balance clocks is limited by friction, air resistance, and thermal expansion to about 10 s per year. To do better, a clock's timekeeper must avoid these mechanical shortcomings. That's why so many modern clocks use quartz oscillators as their timekeepers.

A quartz oscillator is made from a single crystal of quartz, the same mineral found in most white sand. Like many hard and brittle objects, a quartz crystal oscillates strongly after being struck. In fact, it's a harmonic oscillator because it acts like a spring with a metal cylinder at each end (Fig. 9.1.8, left). The two cylinders oscillate in and out symmetrically about their combined center of mass, with a period determined only by the cylinders' masses and the spring's stiffness. In a quartz crystal, the spring is the crystal itself and the cylinders are its two halves (Fig. 9.1.8, right). Since their restoring forces are proportional to displacement from equilibrium, both systems are harmonic oscillators.

Because of its exceptional hardness, a quartz crystal's restoring forces are extremely stiff. Even a tiny distortion leads to huge restoring forces. Since the period of a harmonic oscillator decreases as its spring becomes stiffer, a typical quartz oscillator has an extremely short period. Its motion is usually called a **vibration** rather than an oscillation because vibration implies a fast oscillation in a mechanical system. *Oscillation* itself is a more general term for any repetitive process, and it can even apply to such nonmechanical processes as electric or thermal oscillations.

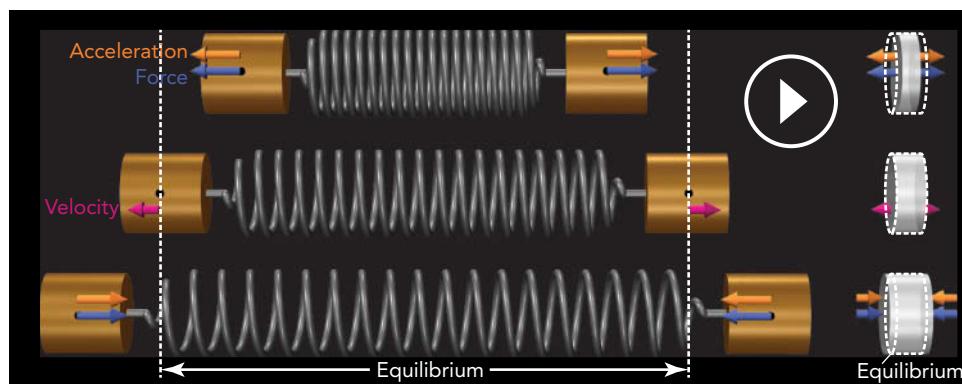


Fig. 9.1.8 A quartz crystal disk (right) acts like a spring with metal cylinders at each end (left). As shown in a time sequence (top to bottom), both systems can oscillate about equilibrium. They accelerate outward when compressed and inward when stretched.

2 The son of freed slaves, Benjamin Banneker (African American mathematician, astronomer, and writer, 1731–1806) grew up on a Maryland tobacco farm and supplemented his limited schooling with borrowed books. Though best remembered for his work in astronomy and for compiling six almanacs, he also produced one of the first clocks made entirely in America. Based on a borrowed pocket watch, Banneker built his wooden balance clock by hand, using a knife to shape the parts. The clock kept accurate time for half a century and even struck the hours.

3 Although ships' navigators could determine latitude (north–south) by measuring the angles of the sun and stars, Earth's rotation prevented them from determining their longitudes (east–west) without help from accurate clocks. Uncertainty in longitude caused so many shipwrecks that, in 1714, the British government offered the Longitude Prize to anyone who could measure longitude to within a prescribed accuracy. One contender for that enormous prize was English clockmaker John Harrison (1693–1776). Harrison spent more than 30 years working on a series of four increasingly accurate clocks. His final prototype, a small balance-ring chronometer known as H4, lost only 5 s during its first transatlantic voyage and could therefore determine longitude to within about 1 mile. Despite H4's repeated successes, the British government took over a decade to fully reward Harrison for his achievements.

Courtesy Lou Bloomfield



Fig. 9.1.9 The quartz crystal in this wristwatch is located inside the silver cylinder at the bottom. Carefully polished to vibrate at a precise frequency, the crystal keeps the watch accurate to a few seconds a month.

Because of its rapid vibration, a quartz oscillator's period is a small fraction of a second. We normally characterize such a fast oscillator by its **frequency**, the number of cycles it completes in a certain amount of time. The SI unit of frequency is the **cycle per second**, also called the **hertz** (abbreviated Hz) after German physicist Heinrich Rudolph Hertz. Period and frequency are reciprocals of one another,

$$\text{frequency} = \frac{1}{\text{period}},$$

so an oscillator with a period of 0.001 s has a frequency of 1000 Hz.

Because the vibrating crystal isn't sliding across anything or moving quickly through the air, it loses energy slowly and vibrates for a long, long time. Also, because quartz's coefficient of thermal expansion is extremely small, the crystal's period is nearly independent of its temperature. With its exceptionally steady period, a quartz oscillator can serve as the timekeeper for a highly accurate clock, one that loses or gains less than a tenth of a second per year.

Of course, a quartz crystal isn't a complete clock. Like the pendulum and balance, it needs something to keep it vibrating and to use that vibration to determine the time. Although these tasks could conceivably be done mechanically, quartz clocks are normally electronic. There are two reasons for this choice. First, the crystal's vibrations are too fast and too small for most mechanical devices to follow. Second, a quartz crystal is intrinsically electronic itself; it responds mechanically to electrical stress and electrically to mechanical stress. Because of this coupling between its mechanical and electrical behaviors, crystalline quartz is known as a *piezoelectric* material and is ideal for electronic clocks.

The clock's circuitry uses electrical stresses to keep the quartz crystal vibrating (Fig. 9.1.9). Just as carefully timed pushes keep a child swinging endlessly on a playground swing, carefully timed electrical stresses keep the quartz crystal vibrating endlessly in its holder. Because the crystal loses so little energy with each vibration, only a tiny amount of work is required each cycle to maintain its vibration.

The clock also detects the crystal's vibrations electrically. Each time its halves move in or out, the crystal experiences mechanical stress and emits a pulse of electricity. These pulses may control an electric motor that advances clock hands or may serve as input to an electronic chip that measures time by counting the pulses.

The quartz crystals used in clocks and watches are carefully cut and polished to vibrate at specific frequencies. In effect, these crystals are tuned like musical instruments to match the requirements of their clocks. Because counting each vibration of the crystal consumes energy, we can prolong clocks' battery lives by using crystals that vibrate relatively slowly—just above the range of human hearing. To make a small crystal vibrate slowly, the manufacturer cuts away most of the center of that crystal to weaken its restoring forces and slow its oscillations. The resulting quartz “tuning fork” oscillator (Fig. 9.1.10) is carefully metalized to permit the watch to interact with it electrically and then it's tuned: a laser beam slowly evaporates the metal from the tips of the fork, decreasing its mass and increasing its frequency, until the desired frequency is reached.

Courtesy Lou Bloomfield



Fig. 9.1.10 Shaped like a tiny tuning fork, this watch crystal vibrates almost exactly 32,768 times per second. Burn marks at its tips were created when its vibrational frequency was tuned by a laser beam.

Check Your Understanding #6: Heavy Metal Music

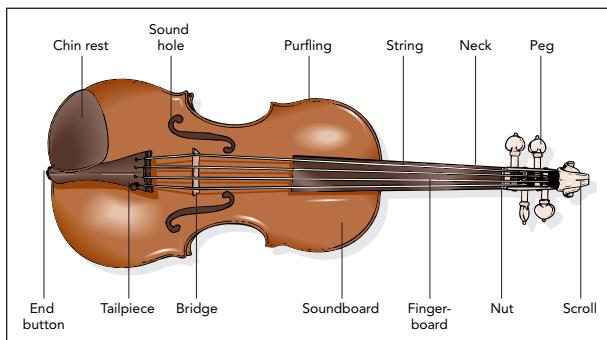
If you drop a metal rod on the floor, end first, you hear a high-pitched tone. What's happening?

Answer: The rod is vibrating as a harmonic oscillator, with its two halves first approaching one another and then moving apart.

Why: The metal rod vibrates in the same manner as a quartz crystal. The body of the rod exerts restoring forces on its two halves. After hitting the floor, these halves move toward and away from one another rapidly, emitting the tone that you hear. Because it's a harmonic oscillator, the frequency (and pitch) of the tone doesn't change as the amplitude of motion decreases.

SECTION 9.2

Musical Instruments



Music is an important part of human expression. Although what qualifies as music is a matter of taste, it always involves sound and often involves instruments. In this section, we'll examine sound, music, and several instruments: violins, pipe organs, and drums. As examples of the three most common types of instruments—strings, winds, and percussion—this trio will help us to understand many other instruments as well.

Questions to Think About: Why are the low-pitch strings on a violin thicker than the high-pitch strings? How does pressing a violin's string against the fingerboard change its pitch? Why does a violin sound different when it's plucked rather than bowed? What purpose does the violin's body serve? What is

vibrating inside a pipe organ? Why are some organ pipes longer than others? Why do most drums sound toneless, providing more rhythm than pitch?

Experiments to Do: Find a violin or guitar, or stretch a strong string between two rigid supports. Even a rubber band will do in a pinch. Pluck the string with your finger and listen to the tone it makes. The string vibrates back and forth at a particular frequency or pitch even as its amplitude of motion gradually decreases. What kind of oscillator has that behavior?

Change the string's frequency of vibration by changing its tension or length. What happens when you shorten the string or prevent part of it from moving? What happens if you increase its tension by pulling it tighter? You can also increase the string's mass by wrapping it with tape. How does that affect its pitch?

You can imitate a pipe organ by blowing gently across the mouth of a bottle or soda straw. If done properly, you'll get the air inside the bottle vibrating up and down rhythmically and you'll hear a tone. What happens to its pitch when you add water to the bottle or pinch off the straw at various points? Why does this tone sound different from that of the string, even when the two have the same pitch?

Finally, a table or plate will act like a drum when you tap it. Compare the sound it makes with those of the previous two "instruments." Does it have a pitch? What distinguishes the sounds of different tables or plates?

Sound and Music

To understand how instruments work, we need to know a bit more about sound and music. In air, **sound** consists of density waves, patterns of compressions and rarefactions that travel outward rapidly from their source. When a sound passes by, the air pressure in your ear fluctuates up and down about normal atmospheric pressure. Even when these fluctuations have amplitudes less than a millionth of atmospheric pressure, you hear them as sound.

When the fluctuations are repetitive, you hear a *tone* with a pitch equal to the fluctuation's frequency. **Pitch** is the frequency of a sound. A bass singer's pitch range extends from 80 to 300 Hz, while that of a soprano singer extends from 300 to 1100 Hz. Musical instruments can produce tones over a much wider range of pitches, but we can hear only those between about 30 and 20,000 Hz, and that range narrows as we get older.

Most music is constructed around *intervals*, the frequency ratio between two different tones. This ratio is found by dividing one tone's frequency by that of the other. Our hearing is particularly sensitive to intervals, with pairs of tones at equal intervals sounding quite similar to one another. For example, a pair of tones at 440 and 660 Hz sounds similar to a pair at 330 and 495 Hz because they both have the interval 3/2.

The interval 3/2 is pleasing to most ears and is common in Western music, where it's called a *fifth*. A fifth is the interval between the two "twinkles" at the beginning of "Twinkle, Twinkle, Little Star." If your ear is good, you can start with any tone for the first twinkle and will easily find the second tone, located at 3/2 the frequency of the first. Your ear hears that factor of 3/2 between the two frequencies.

The most important interval in virtually all music is 2/1, or an *octave*. Tones that differ by a factor of 2 in frequency sound so similar to our ears that we often think of them as

4 In addition to his contributions to mathematics, geometry, and astronomy, the Greek mathematician Pythagoras (ca. 580–500 BC) was perhaps the first person to use mathematics to relate intervals, pitches, and the lengths of vibrating strings. He and his followers laid the groundwork for the scale used in most Western music.

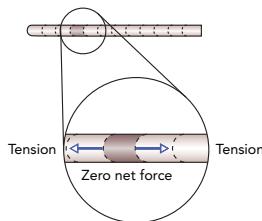


Fig. 9.2.1 A taut violin string can be viewed as being composed of many individual pieces. When the string is straight, the two forces exerted on a given piece by its neighbors cancel perfectly.

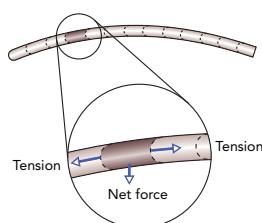


Fig. 9.2.2 When a violin string is curved, the two forces exerted on a given piece by its neighbors don't point in exactly opposite directions and don't balance one another. The piece experiences a net force.

being the same. When men and women sing together “in unison,” they often sing an octave or two apart, and the differences in the tones, always factors of 2 or 4 in frequency, are only barely noticeable.

The octave is so important that it structures the entire range of audible pitches. Most of the subtle interplay of tones in music occurs in intervals of less than an octave, less than a factor of 2 in frequency. Thus most traditions build their music around the intervals that lie within a single octave, such as $5/4$ and $3/2$. They pick a particular standard pitch and then assign notes at specific intervals from this standard pitch. This arrangement repeats at octaves above and below the standard pitch to create a complete scale of notes. (For a history of scales, see **4**.)

The scale used in Western music is constructed around a note called A_4 , which has a standard pitch of 440 Hz. At intervals of $9/8$, $5/4$, $4/3$, $3/2$, $5/3$, and $15/8$ above A_4 lie the six notes B_4 , $C^{\#}_5$, D_5 , E_5 , $F^{\#}_5$, and $G^{\#}_5$. Similar collections of six notes are built above A_5 (880 Hz), which has a frequency twice that of A_4 , and above A_3 (220 Hz), which has a frequency half that of A_4 . In fact, this pattern repeats above A_1 (55 Hz) through A_8 (7040 Hz).

Actually, Western music is built around 12 notes and 11 intervals that lie within a single octave. Five more intervals account for five additional notes, B^{\flat}_4 , C_5 , $D^{\#}_5$, F_5 , and G_5 . It's also not quite true that every note is based exclusively on its interval from A_4 . While A_4 remains at 440 Hz, the pitches of the other 11 notes have been modified slightly so that they're at interesting and pleasing intervals from one another as well as from A_4 . This adjustment of the pitches led to the *well-tempered scale* that has been the basis for Western music for the last several centuries.

Check Your Understanding #1: A Night at the Opera

A typical singing voice can cover a range of about two octaves—for example, from C_4 to C_6 . How broad is this range of frequencies?

Answer: There is a factor of 4 in frequency between the lowest and the highest notes that the typical voice can sing.

Why: Since notes separated by an octave are separated by a factor of 2 in frequency, notes separated by two octaves are separated by a factor of 4.

A Violin's Vibrating String

The tones produced by a violin begin as vibrations in its strings. On their own, these strings are limp and shapeless so they rely on the violin's rigid body and neck for structure. The violin subjects its strings to **tension**, outward forces that act to stretch it, and this tension gives each string an equilibrium shape—a straight line.

To see that a straight violin string is in equilibrium, think of it as being composed of many individual pieces that are connected together in a chain (Fig. 9.2.1). Tension exerts a pair of outward forces on each piece of the string; its neighboring pieces are pulling that piece toward them. Since the string's tension is uniform, these two outward forces sum to zero; they have equal magnitudes but point in opposite directions. With zero net force on each of its pieces, the straight string is in equilibrium.

When the string is curved, however, the pairs of outward forces no longer sum to zero (Fig. 9.2.2). Although those outward forces still have equal magnitudes, they now point in slightly different directions. As a result, each piece experiences a small net force.

The net forces on its pieces are restoring forces because they act to straighten the string. If you distort the string and release it, these restoring forces will cause the string to vibrate about its straight equilibrium shape in a natural resonance. The string's restoring forces are special; the more you curve the string, the stronger the restoring forces on its pieces become. In fact, the restoring forces are **springlike forces**—they increase in proportion to the string's distortion—so the string is a form of harmonic oscillator!

Actually, the string is much more complicated than a pendulum or a balance ring. It can bend and vibrate in many distinct **modes**, or basic patterns of distortion, each with its own period of vibration. Nonetheless, the string retains the most important feature of a harmonic oscillator: the period of each vibrational mode is independent of its amplitude. Thus a violin string's pitch doesn't depend on how hard it's vibrating. Think how tricky it would be to play a violin if its pitch depended on its loudness!

A violin string has one simplest vibration—its **fundamental vibrational mode**. In this mode, the entire string arcs alternately one way and then the other (Fig. 9.2.3). Its kinetic energy peaks as it rushes through its straight equilibrium shape, and its potential energy (elastic potential energy in the string) peaks as it stops to turn around. The string's midpoint travels the farthest (the **vibrational antinode**), while its ends remain fixed (the **vibrational nodes**). At each moment, its shape is the gradual curve of the trigonometric sine function.

In this fundamental mode, the violin string behaves as a single harmonic oscillator. As with any harmonic oscillator, its vibrational period depends only on the stiffness of its restoring forces and on its inertia. Either stiffening the violin string or reducing its mass will quicken its fundamental vibration and increase its fundamental pitch.

A violin has four strings, each with its own stiffness and mass, and therefore with its own fundamental pitch. In a tuned violin, the notes produced by these strings are G₃ (196 Hz), D₄ (294 Hz), A₄ (440 Hz), and E₅ (660 Hz). The G₃ string, which vibrates rather slowly, is the most massive. It's usually made of gut, wrapped in a coil of heavy metal wire. The E₅ string, on the other hand, must vibrate quite rapidly and so needs to have a low mass. It's usually a thin steel wire.

You tune a violin by adjusting the tension in its strings using the pegs in its neck and tension adjusters on the tailpiece. Tightening the string stiffens it by increasing both the outward forces on its pieces and the net forces they experience during a distortion. Since temperature and time can alter a string's tension, you should always tune your violin just before a concert.

A string's fundamental pitch also depends on its length. Shortening the string both stiffens it and reduces its mass, so its pitch increases. That stiffening occurs because a shorter string curves more sharply when it's displaced from equilibrium and therefore subjects its pieces to larger net forces. This dependence on length allows you to raise a string's pitch by pressing it against the fingerboard in the violin's neck and effectively shortening it. Part of a violinist's skill involves knowing exactly where on the string to press it against the fingerboard to produce a particular note.

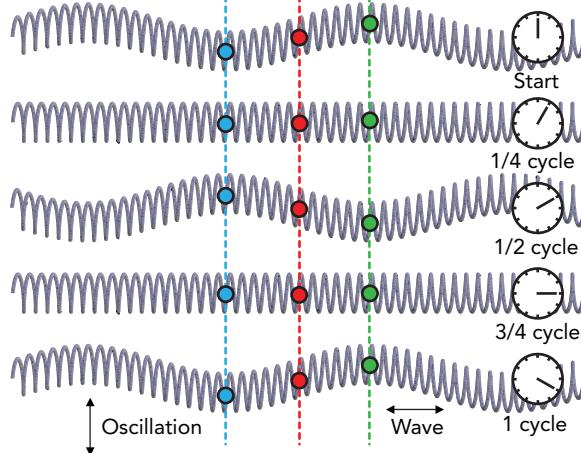
If the arc of a string vibrating in its fundamental mode reminds you of a wave, that's because it is one. It's a **mechanical wave**, the natural motions of an extended object about its stable equilibrium shape or situation. An *extended* object is one like a string, stick, or lake surface that has many parts that move with limited independence. Since its parts influence one another, an extended object with a stable equilibrium exhibits fascinating natural motions that involve many parts moving at once; it exhibits mechanical waves.

With its innumerable linked pieces and its stable equilibrium shape, the violin string exhibits such waves. The string's fundamental mode is a particularly simple wave, a **standing wave**, which is a wave with fixed nodes and antinodes. A standing wave's basic shape doesn't change with time; it merely scales up and down rhythmically at a particular frequency and amplitude (its peak extent of motion). Most important, the standing wave doesn't travel along the string.



Fig. 9.2.3 Pulled outward at its fixed ends, this string's tension gives it a straight equilibrium shape about which it can vibrate. Here, the string vibrates in its fundamental vibrational mode—the whole string moves up and down together as a single harmonic oscillator.

Fig. 9.2.4 In a transverse wave, the underlying oscillation is perpendicular to the wave itself. In this case, a spring is oscillating vertically but forming a horizontal wave. With its fixed nodes and antinodes, this transverse wave is also a standing wave.



Although this wave extends along the string, its associated oscillation is *perpendicular* to the string and therefore *perpendicular* to the wave itself. A wave in which the underlying oscillation is perpendicular to the wave itself is called a **transverse wave** (Fig. 9.2.4). Waves on strings, drums, and the surface of water are all transverse waves.

Check Your Understanding #2: Feeling Tense?

A common way to determine the tension in a cord is to pluck it and listen for how fast it vibrates. Why does this technique measure tension?

Answer: The frequency of a string's fundamental vibrational mode increases with its tension.

Why: Any cord that is drawn taut from its ends will exhibit natural resonances like those in a violin string. The tauter the string, the higher will be the frequencies of those resonances.

The Violin String's Harmonics

The fundamental vibrational mode isn't the only way in which a violin string can vibrate. The string also has **higher-order vibrational modes** in which the string vibrates as a chain of shorter strings arcing in alternate directions (Fig. 9.2.5). Each of these higher-order vibrational modes is another standing wave, with a fixed shape that scales up and down rhythmically at its own frequency and amplitude.

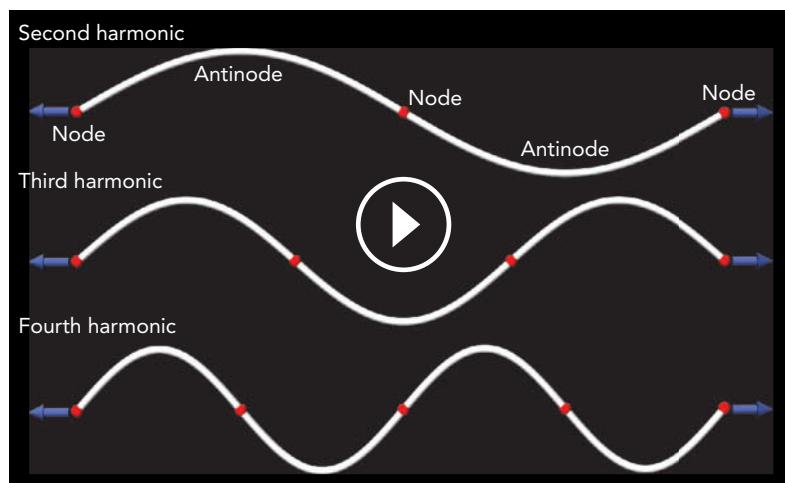


Fig. 9.2.5 A taut string vibrating between two fixed points in its second, third, and fourth harmonic modes. The string vibrates as two, three, or four segments, completing cycles at two, three, or four times the fundamental vibrational frequency, respectively.

For example, the string can vibrate as two half-strings arcing in opposite directions and separated by a motionless vibrational node. In this mode, the violin string not only vibrates as half-strings but has the pitch of half-strings as well. Remarkably, that half-string pitch is exactly twice the whole-string (fundamental) pitch! In general, a string's vibrational frequency is inversely proportional to its length, so halving its length doubles its frequency. Frequencies that are integer multiples of the fundamental pitch are called **harmonics**, so this half-string vibration occurs at the second harmonic pitch and is called the *second harmonic mode*.

A violin string can also vibrate as three third-strings, with a frequency that's three times the fundamental. The interval between this third harmonic pitch and the fundamental pitch is an octave and a fifth ($2/1$ times $3/2$). Overall, the fundamental and its second and third harmonics sound very pleasant together.

While the violin string can vibrate in even higher harmonics, what's more important is that the string often vibrates in more than one mode at the same time. For example, a violin string vibrating in its fundamental mode can also vibrate in its second harmonic and emit two tones at once.

Harmonics are important because bowing a violin excites many of its vibrational modes. The violin's sound is thus a rich mixture of the fundamental tone and the harmonics. Known as **timbre**, this mixture of tones is characteristic of a violin, which is why an instrument producing a different mixture doesn't sound like a violin.

When a violin string is vibrating in several modes at once, its shape and motion are complicated. The individual standing waves add on top of one another, a process known as **superposition**. Each vibrational mode has its own amplitude and therefore its own loudness contribution to the string's timbre.

While these individual waves coexist beautifully on the string, with virtually no effect on one another, the string's overall distorted shape is now the superposition of the individual wave shapes. Not only is that overall shape quite complicated, it also actually changes substantially with time. That's because the different harmonic waves vibrate at different frequencies, and their superposition changes as they change. The string's overall wave is not a standing wave, and its features can even move along the string!

Check Your Understanding #3: Swinging High and Low Together

When two people swing a long jump rope, they can make it swing as a single arc or as two half-ropes arcing in opposite directions. To make the rope swing as two half-ropes, it must be turned faster or with less tension. Why?

Answer: A jump rope is essentially a vibrating string. The two-half-rope pattern is the second harmonic mode, with a vibrational frequency twice that of the normal fundamental arc.

Why: Although it swings around in a circle, the jump rope is actually vibrating up and down at the same time it's vibrating forward and back. Together, these two vibrations create the circular motion. To make the rope vibrate in its second harmonic mode (as two half-ropes) without changing its tension, it must be swung twice as fast as normal.

Bowing and Plucking the Violin String

You play a violin by drawing a bow across its strings. The bow consists of horsehair, pulled taut by a wooden stick and coated with rosin, a sticky substance based on pine sap. This coated horsehair exerts frictional forces on the strings as it moves across them. Most important, however, is that it exerts much larger static frictional forces than sliding ones.

As the sticky bow hairs rub across a string, they grab the string and push it forward with static friction. Eventually the string's restoring force overpowers static friction, and the string suddenly starts sliding backward across the hairs. Because the hairs exert little sliding friction, the string completes half a vibrational cycle with ease. As it stops to reverse direction, however, the hairs grab the string again and begin pushing it forward. This process repeats over and over.



Fig. 9.2.6 Resonant energy transfer makes it possible for sound to shatter a crystal wineglass. When the sound pushes on the glass rhythmically, the sound slowly transfers energy to the glass, until it finally shatters. Because the sound must be extremely loud and at exactly the resonant frequency of the glass, only the most extraordinary opera singers can break a crystal wineglass.

Each time the bow pushes the string forward, it does work on the string and adds energy to the string's vibrational modes. This process is an example of **resonant energy transfer**, in which a modest force doing work in synchrony with a natural resonance can transfer a large amount of energy to that resonance. Just as gentle, carefully timed pushes can get a child swinging high on a playground swing, so too can gentle, carefully timed pushes from a bow get a string vibrating vigorously on a violin. Similar rhythmic pushes can cause other objects to vibrate strongly, notably a crystal wineglass (Fig. 9.2.6) and the Tacoma Narrows Bridge near Seattle, Washington (Fig. 9.2.7). The wineglass's response to a certain tone is also an example of **sympathetic vibration**, the transfer of vibrational energy between two systems that share a common vibrational frequency.

The amount of energy the bow adds to each vibrational mode depends on where it crosses the violin string. When you bow the string at the usual position, you produce a strong fundamental vibration and a moderate amount of each harmonic. Bowing the string nearer its middle reduces the string's curvature, weakening its harmonic vibrations and giving it a more mellow sound. Bowing the string nearer its end increases the string's curvature, strengthening its harmonic vibrations and giving it a brighter sound.

The sound of a plucked violin string also depends on harmonic content and thus on where that string is plucked. However, this sound is quite different from that of a bowed string. The difference lies in the sound's *envelope*, the way the sound evolves with time. This envelope can be viewed as having three time periods: an initial attack, an intermediate sustain, and a final decay. The envelope of a plucked string is an abrupt attack followed immediately by a gradual decay. In contrast, the envelope of a bowed string is a gradual attack, a steady sustain, and then a gradual decay. We learn to recognize individual instruments not only by their harmonic content but also by their sound envelopes.

Check Your Understanding #4: What's the Buzz?

Sometimes a tone from an instrument or sound system will cause some object in the room to begin vibrating loudly. Why does this happen?

Answer: The object has a natural resonance at the tone's frequency, and sympathetic vibration is transferring energy to the object.

Why: Energy moves easily between two objects that vibrate at the same frequency. A note played on one instrument will cause the same note on another instrument to begin playing. Even everyday objects will exhibit sympathetic vibration when the right tone is present in the air.



Fig. 9.2.7 The Tacoma Narrows Bridge collapsed in November 1940, as the result of resonant energy transfer between the wind and the bridge surface. Shortly after its construction, the automobile bridge began to exhibit an unusual natural resonance in which its surface twisted slowly back and forth so that one lane rose as the other fell. During a storm, the wind slowly added energy to this resonance until the bridge ripped itself apart.

An Organ Pipe's Vibrating Air

Like a violin, a pipe organ uses vibrations to create sound. However, its vibrations take place in the air itself. An organ pipe is essentially a hollow cylinder, open at each end and filled with air. Because that air is protected by the rigid walls of the pipe, its pressure can fluctuate up and down relative to atmospheric pressure and it can exhibit natural resonances.

In its fundamental vibrational mode, air moves alternately toward and away from the pipe's center (Fig. 9.2.8), like two blocks on a spring. As air moves toward the pipe's center, the density there rises and a pressure imbalance develops. Since the pressure at the pipe's center is higher than at its ends, air accelerates *away* from the center. The air eventually stops moving inward and begins to move outward. As air moves away from the pipe's center, the density there drops and a reversed pressure imbalance occurs. Since the pressure at the pipe's center is lower than at its ends, air now accelerates *toward* the center. It eventually stops moving outward and begins to move inward, and the cycle repeats. The air's kinetic energy peaks each time it rushes through that equilibrium and its potential energy (pressure potential energy in the air) peaks each time it stops to turn around.

This air is vibrating about a stable equilibrium of uniform atmospheric density and pressure, and it is clearly experiencing restoring forces. It should come as no surprise that those restoring forces are springlike and that the air column is yet another harmonic oscillator. As such, its vibrational frequency depends only on the stiffness of its restoring forces and on its inertia. Either stiffening the air column or reducing its mass will quicken its vibration and increase its pitch.

These characteristics depend on the length of the organ pipe. A shorter pipe not only holds less air mass than a longer pipe, it also offers stiffer opposition to any movements of air in and out of that pipe. With less room in the shorter pipe, the pressure inside it rises and falls more abruptly, leading to stiffer restoring forces on the moving air. Together, these effects make the air in a shorter pipe vibrate faster than the air in a longer pipe. In general, an organ pipe's vibrational frequency is inversely proportional to its length.

Unfortunately, the mass of vibrating air in a pipe also increases with the air's average density, so even a modest change in temperature or weather will alter the pipe's pitch. Fortunately, all the pipes shift together so that an organ continues to sound in tune. Nonetheless, this shift may be noticeable when the organ is part of an orchestra.

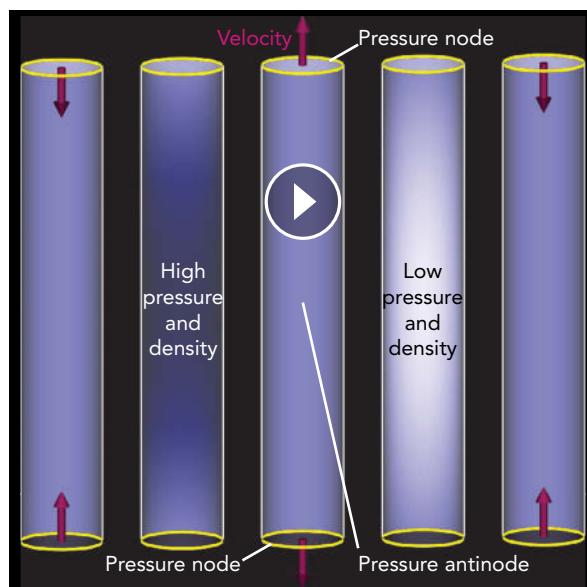
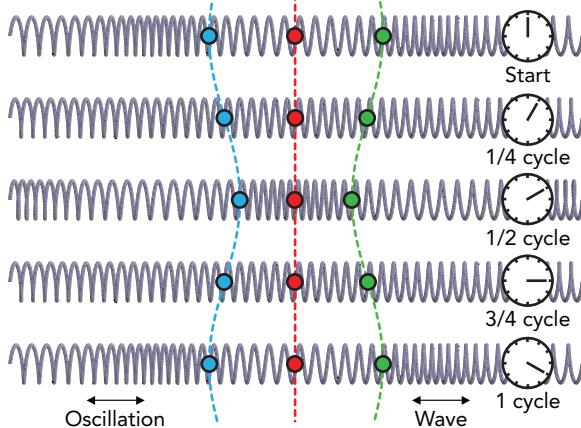


Fig. 9.2.8 A time sequence (left to right) showing a column of air vibrating in a pipe that's open at both ends. Here, the column vibrates in its fundamental vibration mode—the whole column moves in and out together as a single harmonic oscillator.

Fig. 9.2.9 In a longitudinal wave, the underlying oscillation is parallel to the wave itself. In this case, a spring is oscillating horizontally, in the same direction as the wave. With its fixed nodes and antinodes, this longitudinal wave is also a standing wave.



As you may suspect, the fundamental vibrational mode of air in the organ pipe is another standing wave. Air in the pipe is an extended object with a stable equilibrium, and the disturbance associated with its fundamental vibrational mode has a basic shape that doesn't change with time; it merely scales up and down rhythmically.

However, the shape of the wave in the pipe's air now has to do with back-and-forth compressions and rarefaction, not with side-to-side displacements, as it did in the violin string. In fact, all the wave's associated oscillation is *along* the pipe and therefore *along* the wave itself. A wave in which the underlying oscillation is parallel to the wave itself is called a **longitudinal wave** (Fig. 9.2.9). Waves in the air, including those inside organ pipes and other wind instruments and those in the open air, are all longitudinal waves.



Check Your Understanding #5: A Pop Organ

If you blow across a soda bottle, it emits a tone. Why does adding water to the bottle raise the pitch of that tone?

Answer: The water shortens the column of moving air inside the bottle and increases the frequency of its fundamental vibrational mode.

Why: A water bottle is essentially a pipe that is open at only one end. It has a fundamental vibrational mode with a frequency that is half that of an open pipe of equal length. As you add water to the bottle, you shorten the effective length of the pipe and raise its pitch.

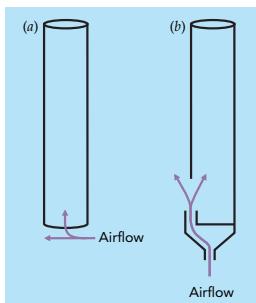


Fig. 9.2.10 (a) Air blown across the bottom of an open pipe will follow any other air that's moving into the pipe. If the air in the pipe is vibrating, this effect will add energy to that vibration. (b) The lower opening in an organ pipe is cut in its side for practical reasons.

Playing an Organ Pipe

The organ uses resonant energy transfer to make the air in a pipe vibrate. It starts this transfer by blowing air across the pipe's lower opening (Fig. 9.2.10a), although for practical reasons that lower opening is usually found on the pipe's side (Fig. 9.2.10b). As the air flows across the opening, it's easily deflected to one side or the other and tends to follow any air that's already moving into or out of the pipe. If the air inside the pipe is vibrating, the new air will follow it in perfect synchrony and strengthen the vibration.

This following process is so effective at enhancing vibrations that it can even initiate a vibration from the random noise that's always present in a pipe. That's how the sound starts when the organ's pump first blows air across the pipe. Once the vibration has started, it grows quickly in amplitude. That amplitude increases until energy leaves the pipe as sound and heat as quickly as it arrives via compressed air. The more air the organ blows across the pipe each second, the more power it delivers to the pipe and the louder the vibration.

Like a violin string, an organ pipe can support more than one mode of vibration. In its fundamental vibrational mode, the pipe's entire column of air vibrates together. In the higher-order vibrational modes, this air column vibrates as a chain of shorter air columns moving in alternate directions. If the pipe has a constant width, these vibrations occur at harmonics of the fundamental. When the air column vibrates as two half-columns, its pitch is exactly twice that of the fundamental mode. When it vibrates as three third-columns, its pitch is exactly three times that of the fundamental. And so on.

Also, the air column inside a pipe can vibrate in more than one mode simultaneously. As with a violin string, the standing waves superpose and the fundamental and harmonic tones are produced together. The shape of the organ pipe and the place where air is blown across it determine the pipe's harmonic content and thus its timbre. Different pipes can imitate different instruments. To sound like a flute, the pipe emits mostly the fundamental tone and keep the harmonics fairly quiet. To sound like a clarinet, its harmonics must be much louder. An organ pipe's loudness always builds slowly during the attack, so it can't pretend to be a plucked string. However, a clever designer can make the organ imitate a surprising range of instruments.

Check Your Understanding #6: An Across Blow

To make the air in a soda bottle vibrate, you must blow *across* the bottle's mouth. Why doesn't blowing *into* its mouth work?

Answer: By blowing across the mouth, you let air that is already vibrating in the bottle redirect your breath so that it enhances the vibration. Blowing into the bottle's mouth merely compresses the air inside the bottle.

Why: Like the bow of a violin moving across its strings, your breath moving across the bottle's mouth enhances the air's vibration via resonant energy transfer. The spontaneous redirection of your breath when you blow across the bottle's mouth leads to rhythmic pushes that are perfectly synchronized with the air's vibration.

5 In 1809, the French Academy of Sciences announced a competition to explain the intricate patterns observed on vibrating surface plates. The only respondent was French mathematician Sophie Germain (1776–1831). As a woman, Germain had been barred from formal education in mathematics and had struggled to learn the subject from books and via correspondence with leading mathematicians, which she conducted under the pseudonym Antoine-August Le Blanc. It took her three tries, but in 1816 she was awarded the prize. Because she was a woman, however, she did not attend the ceremony. Her analysis of surface vibrations, though imperfect, was a visionary effort, made all the more extraordinary by her circumstances. Although her mentor, Carl Friedrich Gauss, managed to convince the University of Göttingen to award her an honorary degree, she died of breast cancer before she could receive it.

A Drum's Vibrating Surface

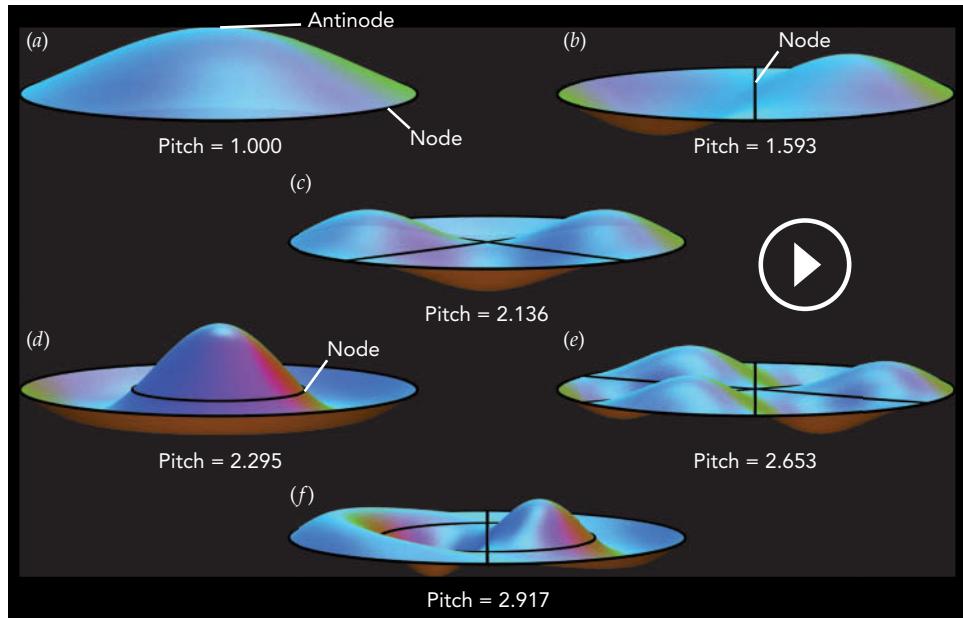
After examining violin strings and organ pipes, you might think that drums have no new physics concepts to show us. But while a drumhead is yet another extended object with a stable equilibrium and springlike restoring forces, its overtone vibrations have an important difference: they *aren't* harmonics.

Violin strings and organ pipes are effectively one-dimensional or linelike objects, dividing easily into half-objects or third-objects that then vibrate at second or third harmonic pitches. Together with the many other one-dimensional instruments in an orchestra or band, they blend seamlessly when they're playing the same fundamental pitch because they share the same harmonics.

However, because a drumhead is effectively two-dimensional or surfacelike, it doesn't divide easily into pieces that resemble the entire drumhead. As a result, the pitches of its overtone vibrations have no simple relationship to its fundamental pitch. A timpani stands out relative to other instruments in part because of the unique overtone pitches.

Figure 9.2.11 illustrates the fundamental (Fig. 9.2.11a) and five lowest-pitched overtone (Fig. 9.2.11b–f) vibrational modes for a drumhead. Each vibrational mode is a standing wave but with vibrational nodes that are curves or lines rather than points. The fundamental mode (Fig. 9.2.11a) has only one node on its outer edge, while the overtone modes have additional nodes within the surface. In each vibrational mode, these nodes remain motionless as the rest of the surface vibrates up and down, its peaks and valleys interchanging alternately. The pitches of the overtone vibrations are indicated relative to the pitch of the fundamental vibration. (For a historical note on the understanding of surface modes, see **5**.)

Fig. 9.2.11 The six lowest-pitch vibrational modes of a drumhead, including (a) the fundamental vibrational mode and (b–f) overtone modes. Pitches are shown relative to the fundamental pitch.



Because striking a drumhead causes it to vibrate in several modes at once, the drum emits several pitches simultaneously. The amplitude of each mode, and consequently its loudness, depends not only on *how hard* you hit the drumhead but also on *where* you hit it. If you hit it at its center, it vibrates primarily in circular modes (Fig. 9.2.11a,d). If you hit it nearer its edge, it also vibrates in noncircular modes (Fig. 9.2.11b,c,e,f).

A timpani sounds most musical when it's struck off-center in such a way that the amplitude of its fundamental vibrational mode is nearly zero and its overtones, particularly Fig. 9.2.11b, dominate its sound. That's because the fundamental vibrational mode emits sound so efficiently that its vibrational energy dissipates before it can produce a discernible tone. Unless all you want is a loud thump, you must hit the timpani off-center so that its long-lived overtone vibrations receive most of the energy and emit most of the sound. The dominant pitch of a properly played timpani is that of its first overtone vibration, and it is tuned with that pitch in mind.

In truth, the pitches shown in Fig. 9.2.11 neglect the effects of air's inertia on the drumhead's vibrations. Since air adds inertia to the drumhead, it lowers the pitches of all the vibrational modes, some more than others. Because of air's influence on pitch, a drum must be tuned to accommodate changes in temperature and weather.

Check Your Understanding #7: Space Kid One

A trampoline is hazardous with several children on it because a child landing on one side of its surface can launch skyward a second child standing on the other side of the surface. How does a downward impact on one side of the trampoline produce a sudden rise of the other side?

Answer: The off-center impact causes the surface to vibrate in its noncircular overtone modes. The simplest such mode, Fig. 9.2.11b, has its two sides moving in alternate directions.

Why: The trampoline is essentially a drumhead, and the children are riding its vibrational modes. Off-center impacts can cause the surface to vibrate in its overtone modes, and these can toss the children in unexpected directions.

Sound in Air

All these vibrations would serve little purpose if we couldn't hear them, so it's time to look at how instruments produce sound. We'll start by looking at sound itself.

We noted at the beginning of this section that sound in air consists of density waves—patterns of compressions and rarefactions that travel outward rapidly from their source. While that observation was mysterious at the time, we can now understand those waves as vibrations in an extended object with a stable equilibrium. That extended object is air.

Neglecting gravity, air is in a stable equilibrium when its density is uniform. If we disturb it from equilibrium, the resulting pressure imbalances will provide springlike restoring forces. These forces, together with air's inertia, lead to rhythmic vibrations—the vibrations of harmonic oscillators. In open air, the most basic vibrations are waves that move steadily in a particular direction and are therefore called **traveling waves**. Like the standing waves inside an organ pipe, the traveling waves in open air are longitudinal—air vibrates along the same direction as the sound wave travels.

As it moves through the open air, a basic traveling sound wave consists of an alternating pattern of high-density regions we'll call **crests** and low-density regions we'll call **troughs** (Fig. 9.2.12). While those names will seem more appropriate when we examine water surface waves in the next section, it's customary to refer to the alternating highs and lows of any wave as crests and troughs, respectively. Whether a wave is standing or traveling, the shortest distance between two adjacent crests is known as the **wavelength**.

A standing wave's crests and troughs merely flip back and forth in place, crests becoming troughs and troughs becoming crests. However, a traveling wave's crests and troughs move steadily in a particular direction at a particular speed. The speed and direction of travel together constitute the traveling wave's **wave velocity**.

Figure 9.2.13 shows five snapshots of a simple sound wave that's heading toward the right. If we watch air's density at the same point in space (green line), it begins as a crest (*a*), decreases (*b*) to a trough (*c*), then rises (*d*) back to a crest (*e*) during one complete vibration cycle. However, if we follow the same crest (red line) over time, it travels one wavelength to the right during one complete vibration cycle (*a–e*). Since a crest moves one wavelength per vibration cycle and frequency is the number of vibration cycles per second, the speed at which the crest moves is equal to the wavelength times the frequency. That relationship can be written as a word equation:

$$\text{wave speed} = \text{wavelength} \cdot \text{frequency}, \quad (9.2.1)$$

in symbols:

$$s = \lambda\nu,$$

and in everyday language:

Broad waves that vibrate quickly travel fast.

Remarkably enough, all sound waves travel at the same speed through air, regardless of wavelength or frequency. That's because as a sound wave's wavelength increases, its crests travel farther during one cycle of vibration but that cycle also takes longer. The two changes balance one another, so the crests travel at the same speed. Longer wavelengths lead to slower vibrations because broadening out the pressure variations in the sound wave weakens its restoring forces. With softer restoring forces and the same inertia (that is, its density), the air vibrates more slowly as the wavelength increases.

Equation 9.2.1 thus yields the same wave speed for any sound wave. Known as the **speed of sound** in air, it's about 331 m/s (1086 ft/s) under standard conditions at sea level (0 °C, 101,325 Pa pressure). Although that's fast, there is still a noticeable delay between when a percussionist strikes the cymbals and when you hear them from across the concert

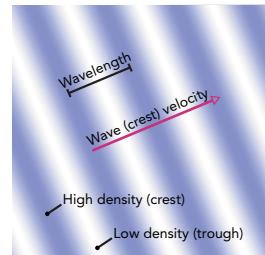


Fig. 9.2.12 A traveling sound wave in air consists of a washboard pattern of high-density (blue) and low-density (white) regions. The distance separating adjacent crests is the wavelength, and the speed and direction of crest motion are the wave velocity.

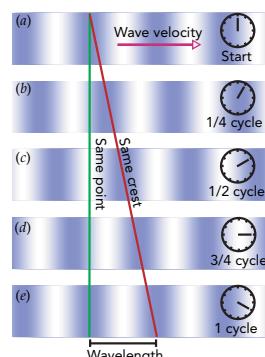


Fig. 9.2.13 A sound wave at five evenly spaced times (*a–e*) showing one complete cycle of oscillation. During that cycle, the pressure at a specific point in space goes from high to low to high (green line), and a specific crest moves 1 wavelength to the right (red line).

hall. Fortunately, because the speed of sound doesn't depend on frequency, when the entire orchestra plays in unison you hear all its different pitches simultaneously.

This discussion of sound assumes that the instruments and listener maintain a constant separation, as they usually do at an orchestra concert. However, when a marching band steps quickly toward or away from the listener, something odd happens: the listener hears its music shifted up or down in pitch. Known as the **Doppler effect**, this frequency shift occurs because the listener encounters sound wave crests at a rate that's different from the rate at which those crests were created. If an instrument and the listener are approaching one another, the listener encounters the crests at an increased rate and the pitch increases. If the two are moving away from one another, the listener encounters the crests at a decreased rate and the pitch decreases. Fortunately, the Doppler effect is subtle at speeds that are small compared to the speed of sound, so you can listen to parades without their sounding flat or sharp.

Check Your Understanding #8: Sounds Faster

Helium in a toy balloon has the same stiffness as ordinary air, but its density and inertia are smaller. How does this difference affect the speed of sound in helium?

Answer: Sound travels faster in helium than it does in ordinary air.

Why: With its reduced density and inertia, helium vibrates faster than air when the two gases carry sound waves of equal wavelengths. Since the speed at which sound of a specific wavelength travels is proportional to the frequency of that sound, the wave speed in helium is greater than that in air.

Check Your Figures #1: Underwater Sound

Although water is about 800 times as dense as ordinary air, water is also about 15,000 times as stiff and its sound vibrations therefore have increased frequencies relative to those in air. When two sound waves have equal wavelengths, the wave in water has a frequency about 4.3 times greater than the wave in air. What is the speed of sound in water?

Answer: The speed of sound in water is about 1420 m/s (4700 ft/s).

Why: From Eq. 9.2.1, the wave speed in water must be 4.3 times the wave speed in air. Since the sound waves have a wave speed of about 331 m/s in air, they must have a speed of about 331 m/s times 4.3 or 1420 m/s in water.

Turning Vibrations into Sound

Anything that disturbs air's otherwise uniform density can produce traveling sound waves. Instruments emit sound by compressing and rarefying the nearby air in sync with their own vibrations. How they accomplish this task differs from instrument to instrument, so we'll have to look at them individually. As we'll see, some instruments find it easier to produce sound than others.

A drum produces sound when its vibrating drumhead alternately compresses and rarifies the nearby air. As portions of that drumhead rise and fall, they upset the air's uniform density and thereby produce sound waves. Whenever it can, though, air simply flows silently out of the drumhead's way, leading to smaller density fluctuations and less intense sound. For example, when the drumhead is experiencing one of the five overtone vibrations shown in Fig. 9.2.11, air flows away from each rising peak in the undulating surface and toward each falling valley. The overtone vibrations still manage to produce sound, but it's less intense and the vibrational energy in the drumhead transforms relatively inefficiently into sound energy.

Air's partial success in dodging the drumhead's overtone vibrations allows those overtones to complete many vibrational cycles before running out of vibrational

energy. Their vibrations are therefore long-lived and have distinct pitches. In contrast, air has difficulty dodging the drumhead's fundamental vibrational mode, which alternately compresses and rarefies the air so effectively that it transfers all its vibrational energy to the air in just a few cycles. That's why the fundamental vibrational mode of a drumhead produces the intense and nearly pitchless "thump" sound that we associate with a drum.

If air can dodge a vibrating surface, it can certainly dodge a vibrating string. Little of a violin's sound comes directly from its vibrating strings; the air simply skirts around them. Instead, the violin creates sound with its top plate, or *belly* (Fig. 9.2.14). The strings transfer their vibrational motions to the belly and the belly pushes on the air to create sound.

Most of this vibrational energy flows into the belly through the violin's *bridge*, which holds the strings away from the violin's body (Fig. 9.2.15). Beneath the G₃-string side of the bridge is the *bass bar*, a long wooden strip that stiffens the belly. Beneath the E₅-string side of the bridge is the *sound post*, a shaft that extends from the violin's belly to its back.

As a bowed string vibrates across the violin's belly, it exerts a torque on the bridge about the sound post. The bridge rotates back and forth, causing the bass bar and belly to move in and out. The belly's motion produces most of the violin's sound. Some of this sound comes directly from the belly's outer surface, and the rest comes from its inner surface and must emerge through its S-shaped holes.

An organ pipe doesn't have to produce sound because that sound already exists. In effect, the pipe's vibrating column of air is a standing sound wave that gradually leaks out of the pipe as a traveling one. Trapped sound is escaping from its container.

This conversion of a standing wave into a traveling wave isn't so remarkable because the two types of waves are closely related. The pipe's standing wave can be thought of as a reflected traveling wave, a traveling wave that's bouncing back and forth between the two ends of the pipe. Because of the reflections, the traveling wave is superposed with itself heading in the opposite direction, and the sum of two equal but oppositely directed traveling waves is a standing wave!

The fact that sound reflects from the open end of an organ pipe is rather surprising. If that end were closed, you'd probably expect a reflection. After all, sound echoes from cliffs and other rigid surfaces. But sound partially reflects from a surprising range of other transitions, including the transition from inside a pipe to outside it. If you don't believe that, clap your hands inside a long pipe and listen for the decaying echoes.

The reflections at the organ pipe's open ends aren't perfect, so the trapped sound wave gradually leaks out and becomes the sound you hear. This process of letting a standing sound wave emerge slowly as a traveling wave is typical of woodwind and brass instruments. The reflection at an open pipe end depends on the shape of that end. Flaring it into the horn shape common in brass instruments reduces the reflection and eases the transition from standing wave to traveling wave. That's why horns project sound so well.



Fig. 9.2.14 A violin's bridge transfers energy from its vibrating strings to its belly. The belly moves in and out, emitting sound. Some of this sound leaves the violin through the S-holes in its body.

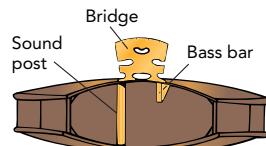


Fig. 9.2.15 The bridge is supported by the bass bar on one side and the sound post on the other. As the strings vibrate back and forth, the bridge experiences a torque that causes the belly of the violin to move in and out and emit sound.

Check Your Understanding #9: Air Guitar

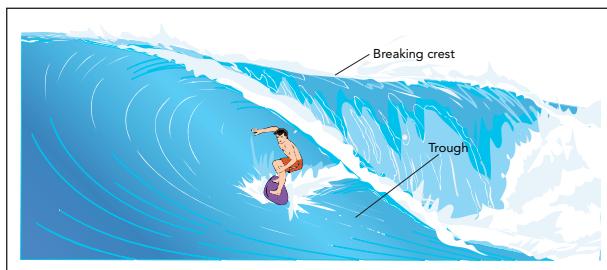
Why does an acoustic guitar have a sound box?

Answer: The sound box transfers the vibrational energy of the strings to the air.

Why: Guitar strings are too narrow to push effectively on the air and emit sound. They do better by transferring their energy to the body of the acoustic guitar so that its flat surfaces can push on the air. An electric guitar avoids the need for a sound box by converting the string's vibrations directly into electric currents and from there into movements of an audio speaker.

SECTION 9.3

The Sea



The sea is never still. If you've visited the seashore, you've probably noticed two of the sea's most important motions: tides and surface waves. In this section, we'll examine the cycle of tides and look at how surface waves travel across water. These water waves can help us understand other wave phenomena, including the electromagnetic waves that are responsible for light and the density waves that are the basis for sound.

Questions to Think About: Why do the tides vary in height from place to place? Why are there no significant tides in a lake or swimming pool? Why does high tide occur about every 12 hours? What moves in a water wave? Do all waves travel at

the same speed? How deep is a wave? Why do waves break near shore? Why do waves always seem to head almost directly toward shore? Why is there often a rhythm to the surf?

Experiments to Do: Although you can make water waves in a basin or tub, they move too quickly for you to see them clearly. You would do better to watch waves at the beach. On a calm day, the sea is in its flat equilibrium shape—but when the wind picks up, look out. How does wind affect the sea's surface? If water waves contain energy, where does that energy come from and what forms does it take in the water?

Watch a floating object move as a wave passes it. Does the object travel with the wave? Does the water itself travel with the wave? Can you think of other cases in which a disturbance moves steadily forward but the material carrying the disturbance remains behind?

Now watch as waves break near the shore. If you look around at different beaches, you'll find waves breaking in two different ways. In some cases a wave will simply collapse into churning froth, while in others its top will dive forward over the water ahead of it. Can you see anything about the beaches or water that might account for this difference?

The Tides

If you watch the sea for a few days, you'll notice the tides. In a cycle as old as the oceans themselves, the water level rises for about $6\frac{1}{4}$ hours to reach *high tide*, drops for about $6\frac{1}{4}$ hours to arrive at *low tide*, and then begins rising again. Once a wonderful mystery, we now know that the tides are caused by Earth's rotation, the moon's gravity, and, to a lesser extent, the sun's gravity.

On Earth, the moon's gravity is so weak that we normally don't notice it. The moon is far away, and as we learned in Section 4.2, gravity diminishes with distance. But this dependence on distance also means that the moon's gravity is stronger on one side of Earth than the other; you experience a stronger pull when you're on the side of Earth nearest the moon than you do when you're on the side opposite it (Fig. 9.3.1). Although you can't feel these variations in moon gravity yourself, Earth's oceans respond to them. The oceans are deformed by the moon's gravity (Fig. 9.3.2), and this deformation produces the tides.

The differences between the moon's gravity at particular locations on Earth and its average strength for the entire Earth give rise to **tidal forces**, residual gravitational forces that act to displace those locations relative to Earth as a whole. The near side of Earth is pulled toward the moon more strongly than average, so it experiences a tidal force toward the moon. The far side of Earth is pulled less strongly than average, so it experiences a tidal force away from the moon.

If Earth were less rigid, these tidal forces would stretch it into an egg shape. The near side of Earth would bulge outward toward the moon, while the far side of Earth would bulge outward away from the moon. But while Earth itself is too stiff to deform much, the oceans are not and they bulge outward in response to the tidal forces. Two *tidal bulges* appear: one closest to the moon and one farthest from the moon (Fig. 9.3.2). The near-side bulge develops because the water there tries to fall toward the moon faster than Earth does as a whole. The far-side bulge forms because water there tries to fall toward the moon

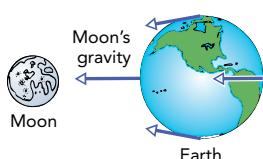


Fig. 9.3.1 The moon's gravity varies over the surface of Earth. The closer a point is to the moon, the stronger the gravity it experiences. This variation in the moon's gravity produces the tides.

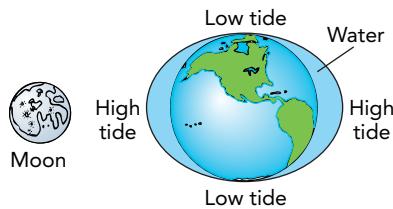


Fig. 9.3.2 Variations in the moon's gravity produce tidal forces on Earth's surface and cause Earth's oceans to bulge outward in two places. These bulges, which are located nearest and farthest from the moon, move over Earth's surface as it rotates.

slower than Earth does as a whole. A beach located in one of these tidal bulges experiences high tide, while one in the ring of ocean between the bulges experiences low tide.

As Earth rotates, the locations of the two tidal bulges move westward around the equator. Since a particular beach experiences high tide whenever it's closest to or farthest from the moon, the full cycle from high to low to high tide occurs about once every 12 h and 24.4 min. The extra 24.4 min reflects the fact that the moon isn't stationary; it orbits Earth and completes a lunar month (the time it takes for the moon to orbit Earth) every 29.53 days. The moon thus passes overhead once every 24 h and 48.8 min, rather than every 24 h.

The moon isn't the only source of tidal forces on Earth's oceans. Although the sun is much farther away from Earth than the moon is, it's so massive that the tidal forces it exerts are almost half as large as those exerted by the moon. The sun's principal effect is to increase or decrease the strength of the tides caused by the moon (Fig. 9.3.3). When the moon and the sun are aligned with one another, their tidal forces add together and produce extra large tidal bulges. When the moon and the sun are at right angles to one another, their tidal forces partially cancel and produce tidal bulges that are unusually small. Twice each lunar month, the tides are particularly strong. These *spring tides* occur whenever the moon and sun are aligned with one another (full moon and new moon). Twice each lunar month the tides are particularly weak. These *neap tides* occur whenever the moon and sun are at right angles to one another (half moon).

Because of this interplay between lunar and solar tidal effects, the cycle of tides varies slightly from day to day. While the average cycle is 12 h and 24.4 min, it fluctuates over a lunar month. Moreover, the exact moment of high or low tide at a particular location is influenced by water's inertia, Earth's rotation, and the environment through which water must flow to form the tidal bulge. This variation is why shore areas often publish tables of the local tides.

Check Your Understanding #1: Seaside Vacations of the Tidal Rich

The Atlantic and Pacific oceans come very close to one another in Central America. If the Atlantic side of this isthmus is experiencing high tide, what tide is the Pacific side experiencing?

Answer: It is experiencing high tide, too.

Why: When the Atlantic side of the Central American isthmus is experiencing high tide, there is a tidal bulge over all of Central America. Both sides of the isthmus experience the same (high) tide.

Tidal Resonances

The sizes of the tides depend on where you are, but high tide is typically a meter or two above low tide. Because the tidal bulges are located near the equator, tides far to the north or south are smaller than that; tides in isolated lakes or seas are smaller still because water can't flow in to create the bulges. However, there are a few special places that have enormous tides. For example, tides in the Bay of Fundy, an estuary between New Brunswick and Nova Scotia, can change the water level by as much as 15 m. How can tides ever get this large?

Giant tides result from natural resonances in channels and estuaries. Just as air in an organ pipe can be made to vibrate strongly by a series of carefully timed pushes from a pump, water in a channel or estuary can be made to oscillate strongly by a series of carefully

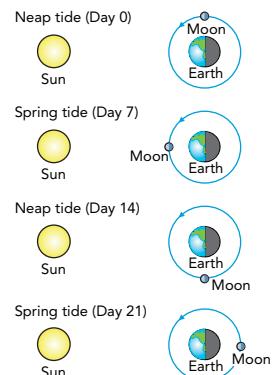


Fig. 9.3.3 The tides vary over a lunar month. They're strongest when the sun and the moon are aligned (spring tides) and weakest when they are at 90° from one another (neap tides).

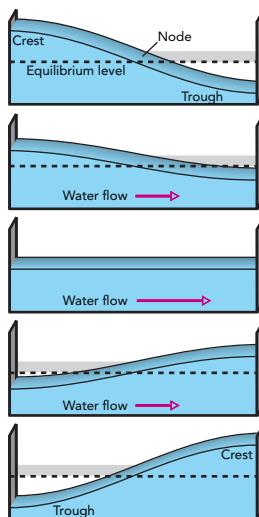


Fig. 9.3.4 Water sloshing in a basin in its fundamental mode. Its surface oscillates up and down, its crest becoming a trough and vice versa, over and over again.

timed pushes from the tides. The water in that channel is another extended object with a stable equilibrium, so it will oscillate about its equilibrium after being disturbed. Giant tides occur through resonant energy transfer when the cycle of the tides gradually builds up the amplitude of a suitable standing wave in a channel.

However, while standing waves in an organ pipe involve the column of air as a whole, standing waves in a channel involve primarily the water's open surface. That water is at equilibrium when its surface is smooth and horizontal, and it experiences springlike restoring forces whenever its surface is disturbed. For the broad waves we're considering in this section, those restoring forces are due to gravity and they are known as *gravity waves*. For the miniature waves in a small drinking glass, however, water's springy, elastic surface contributes significantly to the restoring forces. Waves that involve this *surface tension* are known as *capillary waves*.

You can observe standing gravity waves on the surface of a large cup of coffee. As you walk with the cup in hand and jostle it back and forth rhythmically, you'll produce gravity waves on the coffee's surface. If you time your walking motion so that it's synchronized with the natural rhythm of a particular standing wave, you'll build up the amplitude of that wave by resonant energy transfer and may find yourself needing a change of clothes. Similar gravity-wave resonances occur in washbasins and bathtubs, phenomena that children find particularly amusing—to the dismay of their parents and the delight of people who repair water-damaged floors and ceilings.

Like standing waves on violin strings and drumheads, standing **surface waves** on water are transverse—the water's vertical surface vibrations are perpendicular to the horizontal waves themselves. In a basin, the fundamental vibrational mode (Fig. 9.3.4) has a node along the middle of the basin and antinodes at either end. At one antinode, water arcs up to a crest—a maximum upward displacement from equilibrium. At the other antinodes, water arcs down to a trough—a maximum downward displacement from equilibrium. Over time, the crest drops to become a trough, while the trough rises to become a crest. That process reverses and then repeats, over and over again until the sloshing water turns all its vibrational energy into thermal energy or transfers it elsewhere.

The giant tides at the end of an estuary are simply the antinode of a standing wave fluctuating up and down between a crest and a trough. While the standing waves in an ordinary washbasin have periods measured in seconds or less, large bodies of water can sustain standing waves known as *seiches* that have periods of minutes or even hours. These standing waves appear throughout the oceans and significantly affect the heights and timings of the tides almost everywhere on Earth.

Water in the Bay of Fundy has a fundamental seiche mode with a period of roughly 13.3 hours. Because this period nearly matches the 12.5-hour cycle of the tides, there's a resonant transfer of energy from the moon to the water oscillating in the estuary. The tides drive water back and forth in this estuary until, after many cycles, that water is moving so strongly that its height varies dramatically with time (Fig. 9.3.5).



© Everett Johnson/eStock Photo

Fig. 9.3.5 The giant tides in the Bay of Fundy can cause its water level to change by as much as 15 m between high and low tide.

Check Your Understanding #2: Potential Profits

People occasionally propose using giant tides to generate electric power, but there's a problem with this idea. If you extracted all the gravitational potential energy from the water at high tide in the Bay of Fundy, how long would it take for a giant high tide to appear again?

Answer: It would take many days.

Why: The sloshing water in the Bay of Fundy acquires its energy via resonant energy transfer. The tides do work on the water, slowly increasing its energy over many cycles of its periodic motion. Since each cycle takes half a day, many days are required to build up the energy needed for a giant tide. If you were to extract all this stored energy at once, the bay would have to start sloshing all over again. A small power plant that extracts only a tiny portion of the stored energy each cycle is possible, however, and one such plant is actually in operation.

Traveling Waves on the Surface of Water

As you sit at the seashore on a cloudless day, enjoying a warm, steady breeze, you can't help but notice that the sea in front of you is covered with ridges. These ridges move steadily toward land and finally crash on the beach. Though it's customary to think of each breaking swell as a separate wave, we'll find it useful to view the entire moving pattern of evenly spaced ridges as a single wave—a traveling surface wave on water.

Traveling surface waves are the basic modes of oscillation on the open ocean, the simplest waves on that effectively *limitless* surface. Despite their steady progress across the water, these traveling waves actually involve oscillation. You can see this oscillation by watching a fixed point on the water's surface. That point fluctuates up and down as the crests and troughs of a traveling wave pass through it. The period of this oscillation is the time required for one full cycle of rise and fall, and the frequency is the number of crests passing through that fixed point each second.

The ocean's surface can host an incredible variety of traveling waves, each moving in its own direction with its own period and frequency. Moreover, these basic waves can coexist, adding together on the ocean's surface to create ever more complicated patterns. Like the pure tones of many flutes, which when blended together in proper proportions imitate the richer timbres of oboes and violins and which when blended still further can produce any possible pattern of sound, traveling waves can be superposed in proper proportions to produce any possible surface pattern or wave. When the ocean is rough and its surface features are ripples layered on swells layered on broad undulations, you're seeing this superposition of "pure tone" traveling waves in all its glory.

In contrast, the basic modes of oscillation on a channel or lake are standing surface waves—the simplest waves on that *limited* surface. In a standing wave, the water's surface oscillates up and down vertically, with its crests and troughs interchanging periodically: crests become troughs and troughs become crests. The standing wave's pattern of crests and troughs doesn't move anywhere; it simply flips up and down in place at a certain frequency.

On their limited surface, these standing waves can be superposed to produce any possible surface pattern or wave, so they, too, are like primary colors. Overall, traveling waves constitute the primary palette of waves for a limitless ocean, and standing waves constitute the primary palette of waves for a limited channel or lake.

STANDING AND TRAVELING WAVES

The most basic waves on an extended object of limited dimensions are standing waves. With their different periods and/or patterns, these standing waves can be superposed to form any possible wave on that limited object.

The most basic waves on an extended object of limitless dimensions are traveling waves. With their different periods and/or directions of travel, these traveling waves can be superposed to form any possible wave on that limitless object.

Actually, we've seen these ideas before in the context of musical instruments and sound. Since instruments are limited objects, their basic vibrations are standing waves—the fundamental and overtone vibrations. And because air is effectively limitless, its basic vibrations are traveling waves—the sound waves it carries. The timbre of an instrument reveals the superposition of its many standing waves, while the full sound of an orchestra or band displays the superposition of its many traveling waves.

Both standing and traveling water surface waves carry energy—energy that they typically obtain from the wind, the tide, and occasionally seismic activity. Each wave's energy consists of kinetic energy in moving water and gravitational potential energy in water that has been displaced from its level equilibrium.

In a standing wave, the energy of the entire wave fluctuates back and forth between kinetic and gravitational potential; the wave's kinetic energy peaks as the surface rushes through its level equilibrium, and its potential energy peaks as the surface stops to turn around at its maximum displacement from equilibrium.

A traveling wave's crests and troughs move steadily forward, so the water is never level or motionless. The energy in a traveling wave is therefore always an even mixture of kinetic and potential energies. Also, because of its directed motion, a traveling wave carries momentum pointing in the same direction as the wave velocity.



Check Your Understanding #3: The Frequency of Seasickness

You're sitting in a small boat on the open ocean, rising and falling with each passing wave crest. How could you measure the frequency of the wave passing under your boat?

Answer: Count the number of up and down cycles you complete in a certain amount of time.

Why: Your boat's up and down motion is caused by the oscillation of the ocean's surface. You are rising and falling at the frequency of the passing wave.

The Structure of a Water Wave

We've seen that a traveling surface wave moves across the open ocean with a certain velocity, wavelength, and frequency. But what is the water itself doing as the wave passes?

You can begin to answer that question by watching a bottle floating on the water's surface (Fig. 9.3.6). As a wave passes it, that bottle rises and falls with the crests and troughs, but it makes no overall progress in any direction. Instead, the bottle travels in a circle. Like the bottle, the water itself doesn't actually move along with the passing wave. Although this water bunches up to create each crest and spreads out to create each trough, it returns to its starting point once the wave has left.

Like the bottle in Fig. 9.3.6, a patch of water on the ocean's surface moves in a circular pattern (Fig. 9.3.7) as a traveling wave passes. Water that starts out on top of a crest moves down and forward as the crest departs. It moves down and backward as the trough arrives, then up and backward as the trough departs, and finally up and forward as the next crest arrives. When it reaches the top of the arriving crest, this water is back where it started on

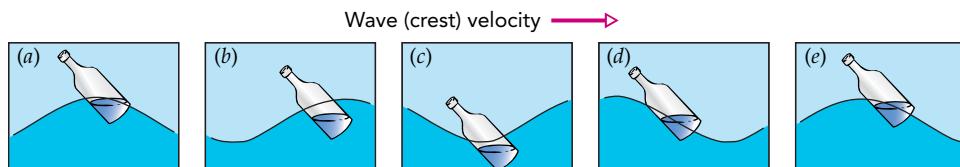


Fig. 9.3.6 You can see that water doesn't move with a wave by watching a bottle floating on the water. The bottle moves in a circle as each crest passes by. Starting at (a) a crest, the bottle moves (b) down and right, (c) down and left, (d) up and left, and then (e) up and right. It returns to its original position just as the next crest arrives.

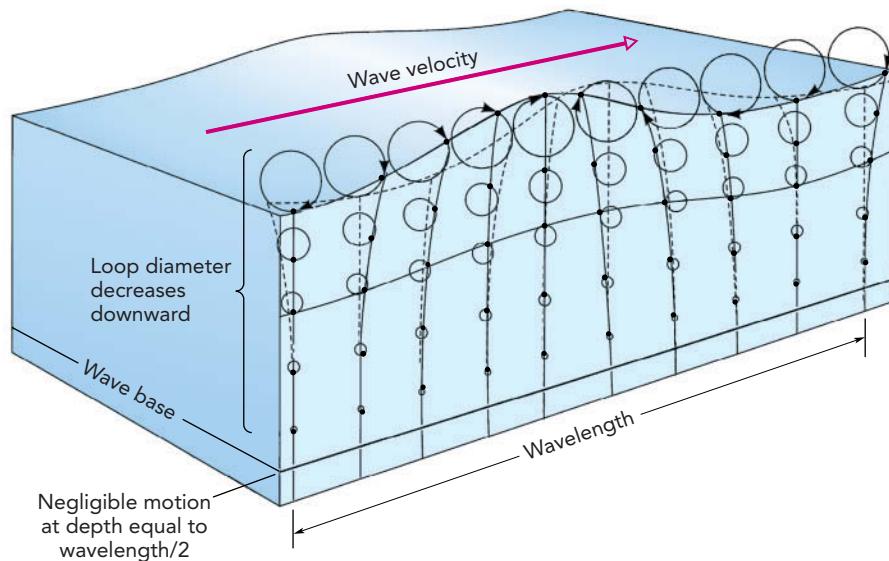


Fig. 9.3.7 Surface water moves in a circular motion as a wave passes. The water currently located at each dark dot will follow the circular path outlined around it as time passes. The circles are largest at the surface and become relatively insignificant once you look more than half a wavelength below the surface. The sense of the circular motion (clockwise or counterclockwise) determines the direction in which the wave travels. This wave travels toward the right.

the ocean's surface. Which way the water circles depends on the wave velocity's direction—the wave's direction of travel. The water at the top of a crest always moves in the same direction as the wave itself.

It isn't only the surface water that moves; water beneath the surface also circles. However, the circles diminish in radius gradually with depth and become negligible at a depth roughly equal to half the wave's wavelength. Thus although it's called a *surface* wave, it has a *depth* to it and therefore a sensitivity to shallow water, as we'll soon see.

These surface waves have another interesting characteristic—their wave speed increases with wavelength. As you may have noticed, long-wavelength swells travel faster than short-wavelength ripples. That's quite different from sound waves, which all have the same wave speed regardless of wavelength.

Such dependence of wave speed on wavelength is known as **dispersion**. It occurs in this case because water's surface is surprisingly stiff when carrying long-wavelength waves. Unlike a taut violin string, which opposes short-wavelength distortions much more stiffly than long-wavelength ones, water's surface uses its weight to oppose long-wavelength disturbances almost as vigorously as it opposes short-wavelength ones. That heightened stiffness for long-wavelength waves boosts their frequencies and therefore increases their wave speeds (see Eq. 9.2.1).

Little ripples have short wavelengths and travel slowly, while large ocean swells have long wavelengths and travel much more rapidly. Giant waves produced by earthquakes and volcanic eruptions, known as *tsunamis*, have extremely long wavelengths and can travel at hundreds of kilometers per hour. Because these giant waves travel so fast and move water so deep in the ocean, they carry enormous amounts of energy and momentum and are potentially disastrous to shore areas (Fig. 9.3.8).

Check Your Understanding #4: Beneath the Waves

If you're swimming and want to dive beneath a wave, how deep must you go to avoid any significant motion in the water?

Answer: You must dive about half a wavelength beneath the water's surface.

Why: A wave causes motion in the water up to a depth of roughly half its wavelength. A typical wave has a wavelength of about 5 m, so you must dive 2.5 m under water before the water will remain essentially still.

Fig. 9.3.8 The Indian Ocean tsunami of December 26, 2004, was initiated by a sudden rise in the ocean floor off the northern coast of Sumatra. This long-wavelength traveling wave moved so quickly through the surrounding Indian Ocean that most shore inhabitants received no warning and were caught unprepared. Moreover, the wave's phenomenal troughs exposed vast stretches of seabed, attracting inquisitive people offshore, where they were then unprotected from the destructive crests that followed. Roughly a quarter million people perished.



© Reuters/Corbis Images

Waves at the Shore

As a wave approaches the shore, it begins to travel through shallow water. Since the water's circular motion extends below the surface, there comes a point at which the wave begins to encounter the seabed. Once the water is shallower than half the wavelength of the wave, the seabed distorts the water's circular motion so that it becomes elliptical.

That change has a number of interesting effects on the wave. First, its wave speed gradually decreases so that the crests begin to bunch together. Second, its amplitude (the height of its crests and depth of its troughs) increases as the wave acts to keep its overall forward momentum constant despite its decrease in speed. These two effects explain why waves that look broad and gradual on the open ocean look quite steep and dangerous nearer the beach. Their crests have bunched together and grown taller, so the slopes between crests and troughs really have become steeper.

A third effect of the shallowing water is a gradual change in the wave's direction of travel. Known as **refraction**, this bending occurs whenever a wave's speed changes as it passes from one environment to another. Since a water surface wave slows as it approaches the shore, that wave refracts—it bends—so as to head more directly toward the shore (Fig. 9.3.9). Because of refraction, waves approach the beach almost perpendicular to it, even if they were traveling at relatively oblique angles far from shore.

A fourth and final effect is the destruction of the wave—it eventually runs out of water and crashes onto the beach. The wave builds each of its crests using local water. When the crest enters very shallow water, there isn't enough water in front of it to construct its forward side. The crest becomes incomplete and begins to “break.”

The form of its demise depends on the slope of the seabed. If that slope is gradual, the wave breaks slowly to form a smooth, rolling, “boiling” surf (Fig. 9.3.10). However, if the slope of the seabed is steep, the wave breaks quickly by having the top of its crest plunge forward over the trough in front of it (Fig. 9.3.11). The steep slope essentially prevents the forward half of the crest from forming. The rearward half continues through its normal circular motion and dives over the missing half-crest in front of it.

The wave can avoid this violent end by colliding with a seawall or cliff instead of a beach. Rather than breaking, the wave will then bounce off the wall and continue on in a new direction. Known as **reflection**, this bouncing effect occurs whenever a wave would have to change significantly to cross a boundary. In this case, the water actually ends at the rigid wall, so the wave simply cannot cross that boundary and must reflect instead. The reflecting wave passes back over itself, and as it does, what had been a traveling wave develops some standing-wave character—a consequence of the wall's limiting the surface of the sea.

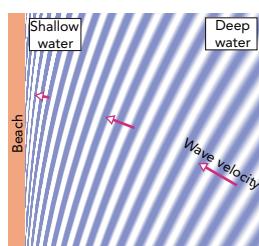


Fig. 9.3.9 When a traveling water surface wave encounters shallow water, it slows down and its direction of travel changes. This refraction process bends the wave velocity so that it points more directly toward the beach.



Fig. 9.3.10 When the slope of the sea bottom is very gradual, the wave crests crumble gently into rolling surf.

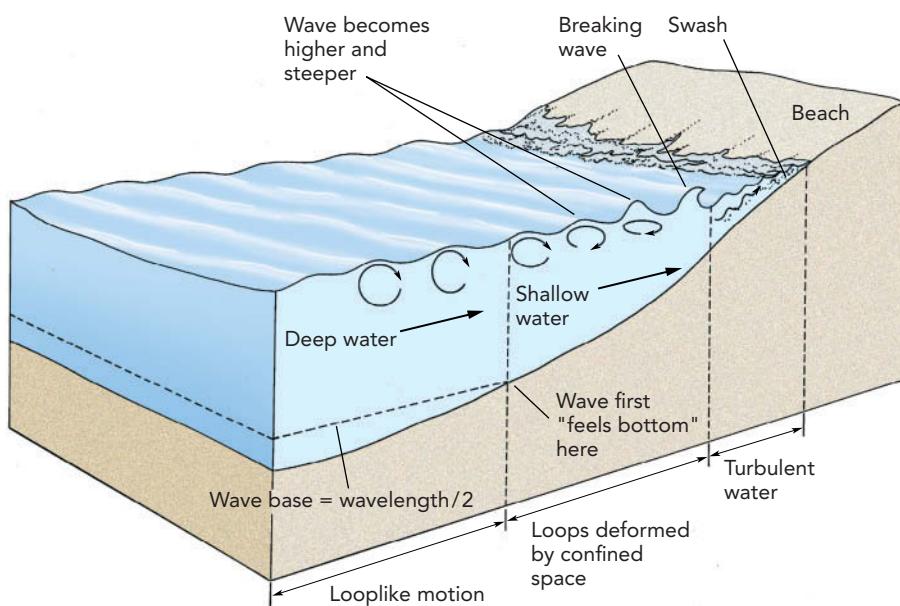


Fig. 9.3.11 When the water becomes too shallow to form a complete crest for the wave, the wave "breaks." If the slope of the shore is steep enough, the crest will be quite incomplete on its shore side and will plunge forward over the trough in front of it.

Less severe changes in how a wave travels on opposite sides of a boundary can also make that wave reflect from the boundary, if only partially. Thus when a water surface wave passes over a sandbar or coral reef and its wave speed changes, it may experience both partial reflection and refraction. These effects contribute to the complicated dynamics of waves near the shore.

Check Your Understanding #5: Out Past the Surf

As you swim away from shore at the ocean, you go through a region where the wave crests are breaking and then reach a more distant region where they don't break. What distinguishes the two regions?

Answer: The breaking region is too shallow for complete crests to form. The nonbreaking region is deep enough to form complete crests.

Why: Wave crests break in shallow regions, where complete crests can't be formed. As long as the water is deep enough, the waves travel intact. However, if there is a shallow region such as a sandbar, even quite far from shore, the crests may break as the wave passes through it.

The Rhythm of the Surf: Wave Interference

If the ocean were carrying only one pure traveling wave toward shore, every breaking swell would look and sound the same. However, there is often a complicated rhythm to the crashing surf; its loudness fluctuates with an overall pattern known as *surf beat*. Surf beat is a sign that the ocean's surface is a busy place; it's actually carrying more than one traveling wave at a time, and these various waves all contribute to the surf.

To understand how multiple traveling waves produce surf beat, let's consider a simple case. Suppose that two traveling waves are heading toward shore and that they have equal amplitudes but different wavelengths (Fig. 9.3.12). Such a situation can easily arise when winds over two portions of the ocean produce two different traveling waves that later overlap. Since they are sharing the ocean's surface, they are superposed on one another.

Because these traveling waves are different, their patterns of crests and troughs can't coincide everywhere. Instead, these waves experience **interference**, their overlapping patterns enhance one another at some locations and cancel one another at other locations. Wherever the two waves are **in phase**, that is, whenever their crests or troughs coincide, they experience **constructive interference**—the waves act together to produce enormous crests or troughs. Wherever the two waves are **out of phase**, that is, whenever the crest of one wave coincides with the trough of the other wave, they experience **destructive interference**—the waves oppose one another to produce muted or absent crests or troughs.

The result is an **interference pattern**, an intricate structure that spreads across space and time when waves are superposed. This interference pattern on the ocean's surface moves and evolves as the traveling waves head toward shore, and it leaves its impression

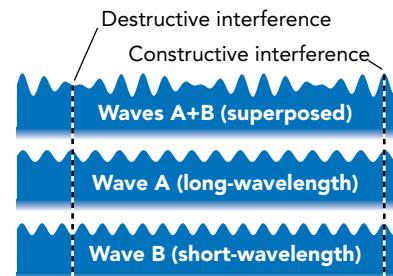


Fig. 9.3.12 When two traveling waves A and B are superposed on the surface of the ocean, they produce interference pattern A + B. As this moving pattern encounters the shore, its varying crest heights give rise to surf beat.

on the crests that eventually break on the beach. Since these crests are no longer equal in height, they exhibit the complicated rhythm of surf beat. When you listen to that beat, you're hearing the consequence of superposition and the interference of waves.

Of course, the real ocean carries many traveling waves, each with its own amplitude, wavelength, and direction of travel. But no matter how complicated the ocean's surface or how intricate the surface beat, you are still just observing the interference of waves.

SUMMARY OF IMPORTANT WAVE PHENOMENA

Reflection: The complete or partial mirror redirection of a wave that occurs when certain dynamical properties of that wave, particularly its speed, change abruptly as it passes from one environment to another.

Refraction: The bending of a wave that occurs when that wave's speed changes as it passes from one environment to another.

Dispersion: The dependence of a wave's speed on its wavelength.

Interference: The interaction of two or more waves when they are superposed; their crests and troughs can enhance one another (constructive interference) or cancel one another (destructive interference) and can produce an interference pattern.

Check Your Understanding #6: Sound Beats

Many notes on a piano are played by two or even three separate strings. If those strings aren't tuned to exactly the same pitch, the resulting sound has a pulsing character—it grows louder and softer rhythmically. What causes this strange pulsation or beat?

Answer: The slightly different sound waves produced by the separate strings are experiencing interference.

Why: Like slightly different water waves, slightly different sound waves exhibit interference effects. The resulting pulsation reflects alternating constructive and destructive interference between the sound waves with slightly different pitches.

Epilogue for Chapter 9

In this chapter, we have looked at natural resonances in a variety of objects. In Clocks, we examined resonances in pendulums, balance rings, and quartz crystals, and we found that these objects are harmonic oscillators—their restoring forces are proportional to displacement. As such, these timekeepers have periods that don't depend on the amplitudes of their motions.

In Musical Instruments, we explored the vibrations of strings and air columns and found that they, too, behave as harmonic oscillators, but with many modes of vibration and thus more complicated behaviors. We found that we could view their motions as waves.

In The Sea, we further explored two different types of waves: standing waves and traveling waves. We saw that both types of waves are found in water, with standing waves appearing in tidal resonances and traveling waves heading out over the open ocean.

Explanation: A Singing Wineglass

Rubbing the lip of the glass is similar to bowing a violin string; both procedures cause resonant energy transfer. Your finger alternately pushes on the glass and then slips,

helping the glass itself to vibrate back and forth beneath your finger in its fundamental vibrational mode. As the glass begins to vibrate, your finger does a little work on it each time the glass is moving in the same direction as your finger. Once the glass is singing loudly, you can keep it singing by circling your finger steadily around the glass. Although the glass continues to emit its energy as sound, you keep giving it more energy by doing work on it with your finger. Adding more water to the glass increases its inertia and slows its vibrations.

Chapter Summary and Important Laws and Equations

How Clocks Work: Clocks are usually based on harmonic oscillators because of their extremely steady periods. Most important, the period of a harmonic oscillator doesn't depend on the amplitude of its motion. Common harmonic oscillator timekeepers include pendulums, balance rings, and quartz crystals.

In a pendulum clock, a swinging pendulum controls the rotation of a toothed wheel, which in turn controls the rotation of the clock's hands. Energy needed to keep the pendulum swinging and to advance the hands comes from the descent of a weighted cord. In a balance clock, a rocking balance ring controls the toothed wheel. Energy for this clock comes from a wound coil spring. A quartz crystal clock detects the vibration of its crystal electronically and uses that vibration to control a motor that advances its clock hands or an electronic circuit that measures time by counting the crystal's vibrations. Small, carefully timed electric pulses from the clock provide the energy that keeps the crystal vibrating.

How Musical Instruments Work: A violin string exhibits natural resonances by vibrating back and forth about its equilibrium shape, a straight line. Bowing the string causes it to vibrate by pushing it away from its equilibrium shape and then allowing it to slip back. The bow transfers energy to the vibrating string by pushing it each time it moves in the bow's direction. The string's fundamental pitch is determined by its mass, tension, and length, and the pitch corresponds to a standing wave on the string. After selecting a string of the correct mass, you tune that string by adjusting its tension. While playing, you change the string's length and pitch by pressing it against the fingerboard. The belly of the violin serves to convert the string's vibration into sound by moving in and out as the string vibrates.

The air inside an organ pipe also exhibits natural resonances. As the organ blows air across the opening at the bottom of a pipe, the air inside that pipe begins to vibrate as a standing wave. The frequency of this vibration is determined principally by the pipe's length, so a short pipe has a higher pitch than a long one. Pipes with different shapes produce different amounts of harmonics and yield different sounds.

The surface of a drum has natural resonances but with a difference: they're not harmonics of the fundamental pitch. The complicated standing waves on the drumhead contribute to its unique sound.

How the Sea Works: The sea exhibits two interesting motions: tides and waves. Tides are caused by bulges in Earth's oceans, created by gravitational tidal forces from the moon and sun. The moon dominates the tides so that bulges appear on the regions of Earth closest to and farthest from the moon. Since Earth is rotating, these bulges move with respect to land and give the tides their 12½-hour cycle.

Waves occur because the water's surface has a stable equilibrium. When it's disturbed from that flat, level equilibrium, this surface will oscillate. Its most basic oscillations are standing waves on confined bodies of water and traveling waves on open water.

The familiar ripples on the ocean's surface are traveling waves that head toward shore at speeds that increase with wavelength. As they approach the shore, these waves form

incomplete crests, which eventually break. The shallowing water also slows the waves, so that they bend more directly toward shore. Interference effects among multiple traveling waves give rise to intricate patterns on the water's surface and to complicated rhythms in the crashing surf.

1. Harmonic oscillator: An oscillator with a restoring force proportional to displacement. Its period of oscillation depends only on the stiffness of its restoring force and on its mass, not on its amplitude of oscillation.

2. Relationship among wave speed, wavelength, and frequency: Wave speed equals wavelength times frequency, or

$$\text{wave speed} = \text{wavelength} \cdot \text{frequency}. \quad (9.2.1)$$

While it's hard to see the electric charges that are responsible for electricity, it's easy to see their effects. They're all around us, in the sparks and shocks of a cold winter day, the imaging process of a xerographic copier, and the illumination of a flashlight when you turn on its switch. Although we often take electricity for granted, it clearly underlies many aspects of our modern world.

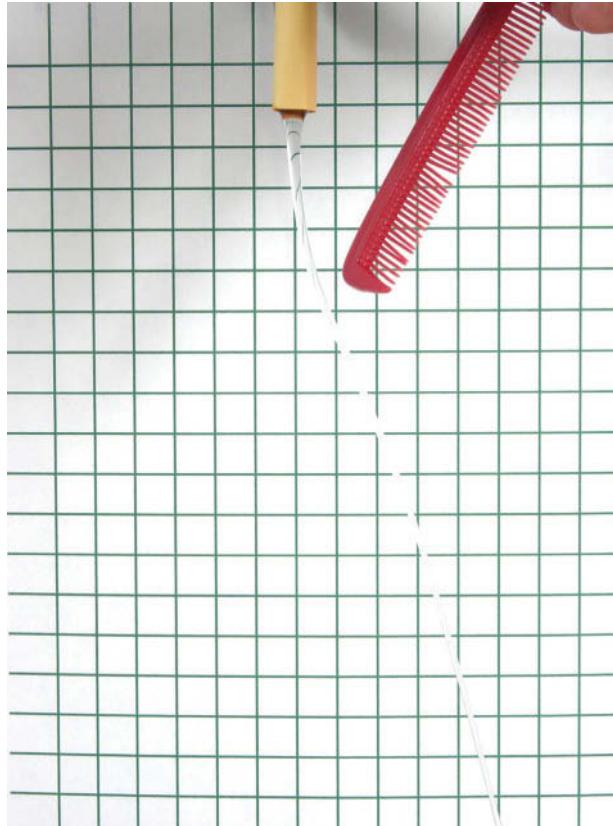
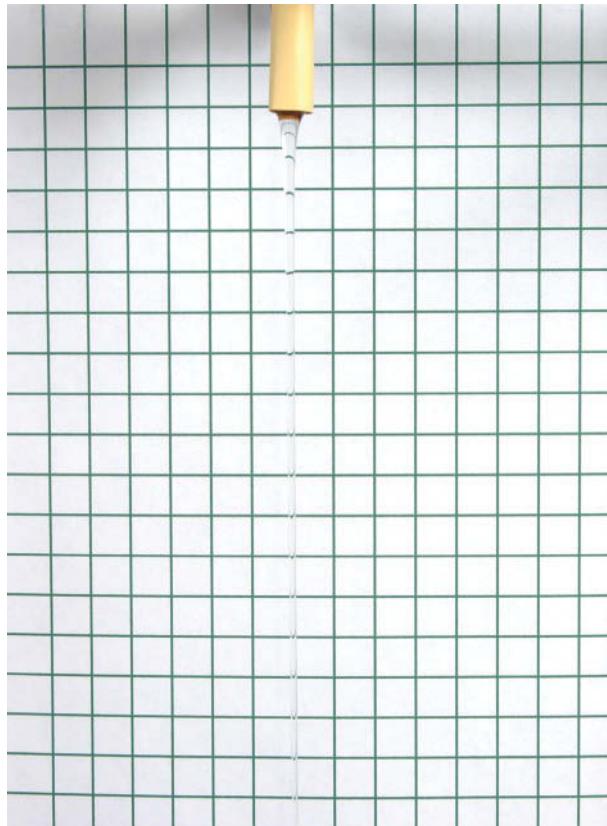
Just imagine what life would be like if there were no electric charges and no electricity. For starters, we'd probably be sitting around campfires at night, trying to think of things to do without television, cell phones, or computer games. But before you remark on just how peaceful such a pre-electronic-age existence would be, let me add one more sobering thought: we wouldn't exist either. Whether it's motionless as static charge or moving as electric current, electricity really does make the world go 'round.

ACTIVE LEARNING
EXPERIMENTS

Moving Water without Touching It

Unlike gravity, which always pulls objects toward one another, electric forces can be either attractive or repulsive. You can experiment with electric forces using a

thin stream of water and an electrically charged comb. First, open a water faucet slightly so that the flow of water forms a thin but continuous strand below the



Courtesy Lou Bloomfield

mouth of the faucet. Next, give your rubber or plastic comb an electric charge by passing it rapidly through your hair or rubbing it vigorously against a wool sweater. Finally, hold the comb near the stream of water, just below the faucet, and watch what happens to the stream. Is the electric force that you're observing attractive or

repulsive? Why does this force change the path of the falling water?

Rubbing the comb through your hair makes it electrically charged. What other objects can acquire and hold a charge when you rub them across hair or fabric? Which works better: a metal object or one that's an insulator? Why?

Chapter Itinerary

Although we often experience electric forces and currents as novelties or nuisances, there are also many devices that depend on them. In this chapter, we examine the mysteries of (1) *static electricity* and study two modern devices based on electricity: (2) *xerographic copiers* and (3) *flashlights*. In Static Electricity, we look at how clothes and other objects acquire charges and how they exert forces on one another as a result. In Xerographic Copiers, we see how these same electric forces work together with light to control the placement of black powder to reproduce images on sheets of paper. In Flashlights, we look at how a

current of electric charges conveys power from batteries to a lightbulb. For a more complete preview of the chapter, turn ahead to the Chapter Summary and Important Laws and Equations at the end of the chapter.

This chapter concentrates on electricity and its charges, but as we will see in Chapter 11, electricity is closely related to magnetism and its poles. While we'll leave the relationships between electricity and magnetism for that next chapter, you may already begin seeing similarities between those two seemingly separate phenomena as you read Chapter 10.

SECTION 10.1

Static Electricity



Electricity may be difficult to see, but you can easily observe its effects. How often have you found socks clinging to a shirt as you remove them from a hot dryer or struggled to throw away a piece of plastic packaging that just won't leave your hand or stay in the trash can? The forces behind these familiar effects are electric in nature and stem from what we commonly call *static electricity*. Static electricity does more than just push

things around, however, as you've probably noticed while reaching for a doorknob or a friend's hand on a cold, dry day. In this section, we'll examine static electricity and the physics behind its intriguing forces and often painful shocks.

Questions to Think About: How does a dryer produce static electricity, and why do some clothes cling while others repel each other? Why does walking across a carpet on a cold, dry day put you at great risk of a shock as you reach for a doorknob? Why do you get only a single brief shock from that knob and not a long sustained one? When you touch a friend and get a shock, did one of you cause that shock or are you both responsible? If rubbing is required to develop static electricity, why does the plastic wrap produce so much of it when you open a package? Why do moist air and antistatic chemicals reduce static electricity?

Experiments to Do: You can study static electricity by rubbing a toy balloon vigorously through your hair or against a wool sweater. Though its appearance won't change, the balloon will begin to attract other things, particularly your hair. What has happened to the balloon? to your hair? Why does the balloon also attract things that weren't rubbed?

Try to get rid of the balloon's attractiveness by letting a thick stream of water flow over its surface. Why does this process return the balloon to normal? What did you "wash" off the balloon? Now rub two identical balloons through your hair and see whether they attract or repel one another. Does the result make sense?

Finally, draw two long strips of transparent tape from a dispenser without rubbing them on anything, and see if they attract or repel. Is rubbing essential to the development of static electricity?

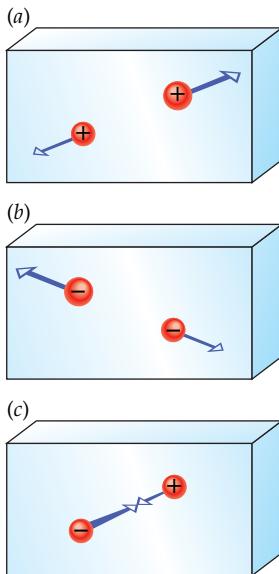


Fig. 10.1.1 (a) Two positive charges experience equal but oppositely directed forces exactly away from one another. (b) The same effect occurs for two negative charges. (c) Two opposite charges experience equal but oppositely directed forces exactly toward one another.

1 Although best remembered for his political activities, American statesman and philosopher Benjamin Franklin (1706–1790) was also the preeminent scientist in the American colonies during the mid-1700s. His experiments, both at home and in Europe, contributed significantly to the understanding of electricity and electric charge. In addition to demonstrating that lightning is a form of electric discharge, Franklin invented a number of useful devices, including the Franklin stove, lightning rods, and bifocals.

Electric Charge and Freshly Laundered Clothes

Unless you have always lived in a damp climate and avoided synthetic materials, you have experienced the effects of static electricity. Seemingly ordinary objects have pushed or pulled on one another mysteriously, and you've received shocks while reaching for light switches, car doors, or friends' hands. Static electricity is more than an interesting nuisance, though; it's a simple window into the inner workings of our universe and worthy of a serious look. It will take some time to lay the groundwork, but soon you'll be able to explain most of the effects of static electricity and even to control it to some extent.

The existence of static electricity has been known for several thousand years. About 600 BC, the Greek philosopher Thales of Miletus (ca. 624–546 BC) observed that when amber is rubbed vigorously with fur, it attracts light objects such as straw and feathers. Known in Greek as *elektron* (*i*ηλεκτρον), amber is a fossil tree resin with properties similar to those of modern plastics. The term *static electricity*, like many others in this chapter, derives from that Greek root.

Static electricity begins with **electric charge**, an intrinsic property of matter. Electric charge is present in many of the **subatomic particles** from which matter is constructed, and these particles incorporate their charges into nearly everything. No one knows why charge exists; it's simply one of the basic features of our universe and something that people discovered through observation and experiment. Because electric charge has so much influence on the objects that contain it, we sometimes refer to those objects as **electric charges**, or simply as **charges**.

Charges exert forces on one another, and these forces are what you observe with static electricity. Next time you're doing laundry, experiment with your clothes as they come out of the dryer. You'll find that some electrically charged garments attract one another, while others repel each other. Evidently, there are two different types of charge. Although this dichotomy has been known since 1733, when it was discovered by French chemist Charles-François de Cisternay du Fay (1698–1739), it was Benjamin Franklin **1** who finally gave the two charges their present names. Franklin called what appears on glass when it's rubbed with silk "positive charge" and what appears on hard rubber when it's rubbed with animal fur "negative charge."

Two like charges (both positive or both negative) push apart, each experiencing a repulsive force that pushes it directly away from the other (Figs. 10.1.1a, b). Two opposite charges (one positive and one negative) pull together, each experiencing an attractive force that pulls it directly toward the other (Fig. 10.1.1c). These forces between stationary electric charges are called **electrostatic forces**.

When you find that two freshly laundered socks push apart, it's because they both have the same type of charge. Whether that charge is positive or negative depends on the fabrics involved (more on that later), so let's just suppose that the dryer has given each sock a negative charge. Since like charges repel, the socks push apart. What does it mean for the dryer to give each sock a negative charge?

The answer to that question has several parts. First, the dryer didn't create the negative charge that it gave to a sock. Like momentum, angular momentum, and energy, electric charge is a conserved physical quantity—it cannot be created or destroyed, only transferred. The negative charge that the dryer gave to the sock must have come from something else, perhaps a shirt.

Second, positive charge and negative charge aren't actually separate entities—they're just positive and negative amounts of the same physical quantity: electric charge. Positive charges have positive amounts of electric charge, while negative charges have negative amounts. Like most physical quantities, we measure charge in standard units. The SI unit of electric charge is the **coulomb** (abbreviated C). Small objects rarely have a whole coulomb of charge, and your sock's charge is only about -0.0000001 C .

Third, the sock's negative charge refers to the sock as a whole, not to its internal pieces. As with all ordinary matter, the sock contains an enormous number of positively

and negatively charged particles. Each of the sock's atoms consists of a dense central core or **nucleus**, containing positively charged **protons** and uncharged **neutrons**, surrounded by a diffuse cloud of negatively charged **electrons**. The electrostatic forces between those tiny charged particles hold together not only the atoms but also the entire sock. However, in giving the sock a negative charge, the dryer saw to it that the sock's **net electric charge**, the sum of all its positive and negative amounts of charge, is negative. With its negative net charge, the sock behaves much like a simple, negatively charged object.

Last, the sock became negatively charged when it contained more electrons than protons. Underlying that seemingly simple statement is a great deal of painstaking scientific study. To begin with, experiments have shown that electric charge is **quantized**, that is, charge always appears in integer multiples of the **elementary unit of electric charge**. This elementary unit of charge is extremely small, only about 1.6×10^{-19} C, and is the magnitude of the charge found on most subatomic particles. An electron has a -1 elementary unit of charge, while a proton has a $+1$ elementary unit of charge. Since the only charged subatomic particles in normal matter are electrons and protons, the sock becomes negatively charged simply by having more electrons than protons.

Returning to the original question, we now know what the dryer did that gave a sock a negative charge. Assuming the sock was electrically **neutral** to start—it had zero net charge—the dryer must have added electrons to the sock or removed protons from the sock or both. These transfers of charge upset the sock's charge balance and gave it a negative net charge.

In keeping with our convention regarding conserved quantities, all unsigned references to charge in this book imply a positive amount. For example, if the dryer gave charge to a jacket, we mean it gave a positive amount of charge to that jacket. We follow this same convention with money: when you say that you gave money to a charity, we assume that you gave a positive amount.

Finally, Franklin's charge-naming scheme was brilliant in concept but unlucky in execution. Although it reduced the calculation of net charge to a simple addition problem, it required Franklin to choose which type of charge to call "positive" and which to call "negative." Unfortunately, his seemingly arbitrary choice made electrons, the primary constituents of electric current in wires, negatively charged. By the time physicists had recognized the mistake, it was too late to fix. Scientists and engineers have had to deal with negative amounts of charge flowing through wires ever since. Imagine the awkwardness of having to carry out business using currency printed only in negative denominations!

Check Your Understanding #1: In Charge of Opening Gifts

The gift you are about to unwrap is electrically neutral. You tear off the clingy wrapper and find that it has a large negative charge. What charge does the gift itself have, if any?

Answer: It has a large positive charge equal in amount to the wrapper's negative charge.

Why: Since charge is a conserved physical quantity, the wrapper and gift must remain neutral overall even after you separate them. The wrapper's negative charge must be balanced by the gift's positive charge.

2 In 1781, after a career as a military engineer in the West Indies, French physicist Charles-Augustin de Coulomb (1736–1806) returned to his native Paris in poor health. There he conducted scientific investigations into the nature of the forces between electric charges and published a series of memoirs on the subject between 1785 and 1789. His research came to a close in 1789 when he was forced to leave Paris because of the French Revolution.

Coulomb's Law and Static Cling

Although your sock and shirt pull together strongly when they're only inches apart, you can put on your shirt and go to the movies without fear of being attacked by your sock from the other side of town. Evidently, the forces between charges weaken with distance.

Over two centuries ago, French physicist Charles-Augustin de Coulomb **2** studied electrostatic forces experimentally and determined that the forces between two electric charges are inversely proportional to the square of their separation (Fig. 10.1.2). For example, doubling the separation between your shirt and sock reduces their attraction by a factor of four, which explains your uneventful night out on the town.

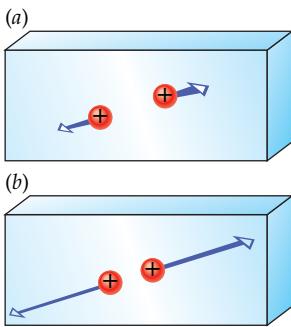


Fig. 10.1.2 The electrostatic forces between two charges increase dramatically as they become closer. As the distance separating two positive charges decreases by a factor of 2 between (a) and (b), the forces those two charges experience increase by a factor of 4.

Coulomb's experiments also showed that the forces between electric charges are proportional to the amount of each charge. That means that doubling the charge on either your shirt or your sock doubles the force each garment exerts on the other. Finally, changing the sign of either charge turns attractive forces into repulsive ones or vice versa. If both garments were either positively charged or negatively charged, they'd repel instead of attracting.

These ideas can be combined to describe the forces acting on two charges and can be written as a word equation:

$$\text{force} = \frac{\text{Coulomb constant} \cdot \text{charge}_1 \cdot \text{charge}_2}{(\text{distance between charges})^2}, \quad (10.1.1)$$

in symbols:

$$F = \frac{k \cdot q_1 \cdot q_2}{r^2},$$

and in everyday language:

When there are enough like charges packed close together on your hair, it'll stand up.

The force on charge₁ is directed toward or away from charge₂, and the force on charge₂ is directed toward or away from charge₁.

This relationship is called **Coulomb's law**, after its discoverer. The **Coulomb constant** is about $8.988 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$ and is one of the physical constants found in nature. Consistent with Newton's third law, the force that charge₁ exerts on charge₂ is equal in amount but oppositely directed from the force that charge₂ exerts on charge₁.

COULOMB'S LAW

The magnitudes of the electrostatic forces between two objects are equal to the Coulomb constant times the product of their two electric charges divided by the square of the distance separating them. If the charges are like, then the forces are repulsive. If the charges are opposite, then the forces are attractive.

In addition to protecting you from distant charged socks, this relationship between electrostatic forces and distance gives rise to another intriguing feature of laundry static: charged clothes can cling to objects that are electrically neutral! For example, a negatively charged sock can stick to a neutral wall.

The origin of this attraction is a subtle rearrangement of charges within the wall. Even though the wall has zero net charge, it still contains both positively and negatively charged particles. When the negatively charged sock is near the wall, it pulls the wall's positive charges a little closer and pushes the wall's negative charges a little farther away (Fig. 10.1.3). Although each individual charge shifts just a tiny distance, the wall contains so many charges that together they produce a dramatic result. The wall develops an **electric polarization**—it remains neutral overall but has a positively charged region nearest the sock and a negatively charged one farthest from the sock.

The wall's positive region attracts the sock, while its negative region repels the sock. Although you might expect those two opposing forces to balance, Coulomb's law says otherwise. Since electrostatic forces grow weaker with distance, the sock is attracted more strongly to the nearer positive region than it is repelled by the more distant negative region. Overall, there is a net electrostatic attraction between the charged sock and the polarized wall, so the sock clings to the wall!

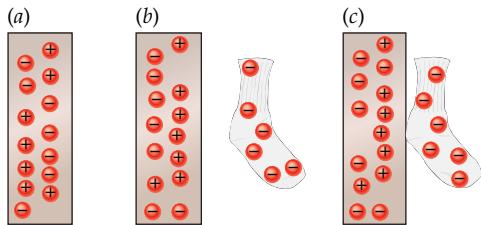


Fig. 10.1.3 (a) A neutral wall contains countless positive and negative charges. (b) As a negatively charged sock approaches the wall, the positive charges move toward it and the negative charges move away from it. (c) The polarized wall continues to attract the sock and holds it in place.

Check Your Understanding #2: Wrapper Recycling

After opening your gift, you try to throw away its negatively charged wrapper. However, the wrapper keeps returning to your hand. What attracts it to your electrically neutral hand?

Answer: Its negative charge polarizes your hand and is then attracted to your hand's nearby positive charge.

Why: Although your hand is neutral, its charges rearrange in response to the nearby wrapper's negative charge. Positive charge in your hand shifts toward the wrapper and attracts it.

Check Your Figures #1: Moving Out

You have two positively charged balls, each of which is experiencing a force of 1 N away from the other. If you halve the distance separating the balls, what force will each exert on the other?

Answer: 4 N.

Why: According to Coulomb's law, the force on each charge varies inversely with the square of their separation. By halving that separation, you increase the electrostatic force by a factor of 4.

Transferring Charge: Sliding Friction or Contact?

While it's clear that the dryer transfers charge between the clothes, why does that charge move and what determines which garments gain charge and which lose it?

You might suppose that sliding friction is responsible for the transfer—that the dryer rubs the clothes together and somehow wipes charge from one garment to the other. After all, friction seems to help you charge a balloon as you rub it through your hair or against a wool sweater. However, be careful—there are other cases of charge transfer that don't involve rubbing at all. For example, the plastic wrap you remove from a store package can acquire a charge no matter how careful you are not to rub it against its contents. And an antique car can build up enough charge to give you a nasty shock even when its pale rubber tires never skid across the pavement.

Charge transfer is less the result of rubbing than it is of contact between dissimilar surfaces. When two different materials touch one another, a few electrons normally shift from one surface to the other. That transfer results from the chemical differences between the two touching surfaces and the associated change in an electron's potential energy when it shifts. In effect, some surfaces are “hungrier” for electrons than others, and whenever two dissimilar surfaces touch, the hungrier surface steals a few electrons from its “less hungry” partner.

The physics behind this theft has to do with **chemical potential energy**, energy stored in the chemical forces that bind together a material's constituent atoms and electrons. To hold onto its electrons, a surface reduces their chemical energies to less than zero, meaning that it would take additional energy to free those electrons from the surface. However, some surfaces reduce the electron chemical potential energies more than others and thus bind their electrons more tightly. If an electron on one surface can reduce its chemical potential energy by shifting to the other surface, it will accelerate toward that “hungrier” surface and eventually stick there. You can picture the electron as “rolling downhill” from a chemical “valley” on one surface to an even deeper valley on the other surface.

This transfer of electrons is self-limiting. As electrons accumulate on the lower energy surface, they begin to repel any electrons that try to follow and the transfer process soon grinds to a halt. It stops altogether when the electrons reach equilibrium—when the forward chemical force an electron experiences is exactly balanced by the backward electrostatic force. The transfer won't resume until you bring fresh uncharged surface regions into contact.

That's where rubbing enters the picture. Rubbing involves lots of surface contact and almost endless opportunities for charge transfer between those surfaces. As clothes tumble about in the dryer, touching one another and often rubbing, some fabrics steal electrons and become negatively charged, while other fabrics lose electrons and become positively charged.

That said, you should be aware that the details of contact charging are messy. For starters, the surfaces that actually touch one another are neither chemically pure nor free of microscopic defects. Although it's generally true that whichever fabric binds electrons most tightly is the one most likely to develop a negative net charge, surface contamination and defects can change the outcome radically. Even your choice of laundry detergent may affect the fabric's surface chemistry and thus how it charges. Furthermore, water molecules cling to most surfaces and influence the contact charging process. Finally, while we've concentrated on the exchange of electrons, it's also possible for certain surfaces to exchange **ions**, that is, electrically charged atoms, molecules, or small particles, along with electrons and acquire net charges as a result.



Check Your Understanding #3: Sticky Tape

When you peel a piece of adhesive tape off a glass window, you find that the tape is attracted toward the spot it left behind. How did the tape and glass acquire electric charges?

Answer: While the tape and glass were in contact, charge was unevenly distributed between their surfaces. Removing the tape merely made that imbalance more obvious.

Why: The tape and glass have different chemical affinities for electrons and become oppositely charged whenever they touch. In fact, the tape's stickiness itself comes from electrostatic attraction.

Separating Your Clothes: Producing High Voltages

The dryer stops, and you take out your favorite shirt. It has several socks clinging to it, so you begin to remove them. As you separate the garments, they crackle and spark. Their attraction is obviously due to opposite charges, but why does separating them make them spark?

To answer that question, let's think about energy as you pull the negatively charged sock steadily away from the positively charged shirt. Since the sock would accelerate toward the shirt if you let go, you are clearly exerting a force on the sock. And because that force and the sock's movement are in the same direction, you are also doing work on the sock. You are transferring energy to it.

That energy is stored in the electrostatic forces; the shirt and sock accumulate **electrostatic potential energy**. Electrostatic potential energy is present whenever opposite charges have been pulled apart or like charges have been pushed together. With the negatively charged sock now far from the positively charged shirt, both attraction and repulsion contribute to the electrostatic potential energy—opposite charges are separated on the two garments, and like charges are assembled together on each garment.

The total electrostatic potential energy in the shirt and sock is the work you did to separate them. However, that potential energy isn't divided equally among the individual charges on these garments. Depending on their locations, some charges have more electrostatic potential energy than others and are therefore more important when it comes to sparks. In recognition of those differences, we need a proper way to characterize the electrostatic potential energy available to a charge at a particular location. The measure we're

seeking is **voltage**, the electrostatic potential energy available per unit of electric charge at a given location, or

$$\text{voltage} = \frac{\text{electrostatic potential energy}}{\text{charge}}.$$

Voltage is a difficult quantity to conceptualize because you can't see charge or sense its stored energy. To help you understand voltage, let's use a simple analogy. In this analogy, the role of charge will be played by water and the role of voltage will be played by pressure. Where voltage is high, visualize water at high pressure. Where voltage is low, picture water at low pressure. Just as water tends to flow from a higher pressure to a lower pressure, so charge tends to flow from a higher voltage to a lower voltage.

This analogy works well because both voltage and pressure measure the energy per unit of something. Voltage is the electrostatic potential energy per unit of charge and pressure is the pressure potential energy per unit of volume (see Section 5.2). Both water at high pressure and charge at high voltage are loaded with energy per unit and likely to do something exciting!

Since the SI unit of energy is the joule and the SI unit of electric charge is the coulomb, the SI unit of voltage is the joule per coulomb, more commonly called the **volt** (abbreviated V). Where the voltage is positive, (positive) charge can release electrostatic potential energy by escaping to a distant neutral place, where the voltage is zero. Charge at positive voltage is analogous to pressurized water, which can release pressure potential energy by flowing into the open air. Where the voltage is negative, charge needs energy to escape to a distant neutral place, where the voltage is zero. Charge at negative voltage is analogous to water at less than atmospheric pressure, which needs energy to flow out into the open air.

In addition to the voltage ↔ pressure analogy, we can also draw a voltage ↔ altitude one. In this second analogy, charges at high voltage are like bicyclists at high altitude. Just as charges tend to flow from a higher voltage to a lower voltage, so bicyclists tend to roll from a higher altitude to a lower altitude. In this analogy, altitude plays the role of voltage and gravitational potential energy plays the role of electrostatic potential energy. The bicyclists release gravitational potential energy as they roll downhill from a mountain, and they need energy to climb uphill from a valley.

Returning to those clothes, you'll find that each point on the shirt or sock has its own voltage. You can determine that voltage by taking a tiny amount of positive charge at that point and moving it to a distant neutral place, where the voltage is zero. The point's voltage is simply the electrostatic potential energy the charge releases during that trip divided by the amount of its charge. If the point you examine is on the positively charged shirt, you'll measure a large positive voltage—probably several thousand volts. If it's on the negatively charged sock, pulling positive charge away from that point will require considerable work, so you'll measure a large negative voltage—probably negative several thousand volts. Whether positive or negative, these large or “high” voltages tend to cause sparks.

We'll look at the physics of sparks and discharges soon, but you can already see why oppositely charged clothes spark as you separate them: that's when the high voltages develop. As long as your sock is clinging tightly to your shirt, there isn't much electrostatic potential energy available. But as soon as you begin to separate them, watch out!

► Check Your Understanding #4: High-Altitude Voltage

Although any cloud may contain opposite charges, only the violent updrafts inside thunderheads are able to separate those charges and produce lightning. Why does such separation lead to lightning?

Answer: That separation takes work, which appears as electrostatic potential energy in the separated charges. The positively charged regions of the thunderhead acquire huge positive voltages, and the negatively charged regions acquire huge negative voltages.

Why: When opposite charges are near each other, they don't necessarily have much electrostatic potential energy per charge and the voltages may be small. Separating those charges to great distances dramatically increases their stored energy and produces high voltages.

Accumulating Huge Static Charges

We've seen that touching two different materials together causes a small transfer of charge from one surface to the other and that separating those oppositely charged surfaces produces elevated voltages and perhaps sparks. However, the quiet crackling and snapping of items in your laundry basket is nothing compared to the miniature lightning bolts you can unleash after walking across a carpet on a dry winter day, stepping out of an antique car, or playing with a static generator. To get a really big spark, you need to separate lots of charge, and that usually requires repeated effort.

Walking across a carpet is just such a repetitive process. Each time your rubber-soled shoe lands on an acrylic carpet, some (positive) charge shifts from the carpet to your shoe. Although the transfer is brief and self-limiting, you now have a little extra charge on your shoe. When you lift that shoe off the carpet, you do work on its newfound charge and your shoe's voltage surges to a high positive value. High-voltage charge tends to leak from one place to another, and the shoe's charge quickly spreads to the rest of your body. By the time your foot lands again on a fresh patch of carpet, the shoe has given away most of its charge and is ready to begin the process all over again.

Each time your foot lands on the carpet, it picks up some charge. Each time it lifts off the carpet, that charge spreads out on your body. By the time you finally reach for the doorknob, you are covered with charge and have an enormous positive voltage. As your hand draws close to the doorknob, it begins to influence the doorknob's charges—pulling the doorknob's negative charges closer and pushing its positive charges away. You are polarizing the doorknob.

As we saw while separating your freshly laundered sock from your shirt, oppositely charged objects that are close but not touching can have both large electrostatic potential energies and strong electrostatic forces. That's the situation here. The closer your hand gets to the doorknob, the stronger the electrostatic forces become until finally the air itself cannot tolerate the forces and a spark forms. In an instant, most of your accumulated electrostatic potential energy is released as light, heat, and sound. And that doesn't include any screams.

As good as walking is at building up charge, though, an antique car is even better. Its pale rubber tires gather negative charge when they touch the pavement and develop large negative voltages as they roll away from it. This charge migrates onto the car body so that, after a few seconds of driving, the car accumulates enough charge to give anyone who touches it a painful shock. Collecting tolls used to be hazardous work! Fortunately, modern tires are formulated to allow this negative charge to return safely to the pavement, so now cars rarely accumulate much charge. Instead, most shocks associated with cars now come from sliding across the seat as you step in or out.

While cars try to avoid static charging, some machines deliberately accumulate separated charge to produce extraordinarily high voltages. The most famous of these static machines is the Van de Graaff generator (Fig. 10.1.4). It uses a rubber belt to lift positive or negative charges onto a metal sphere until the magnitude of that sphere's voltage reaches hundreds of thousands or even millions of volts.

A typical classroom Van de Graaff generator uses a motor-driven rubber belt to carry negative charges from its base to its spherical metal top. Once inside the sphere, the belt's negative charges flow outward onto the sphere's surface, where they can be as far apart as possible. There they remain until something releases them.

Suspended at the top of a tall, insulating column, the Van de Graaff generator's sphere can accumulate an enormous negative charge. You may hear the motor struggling as it pushes the belt's negative charges up to the sphere, a reflection of how much negative voltage the sphere is developing. Eventually it releases its negative charge via an immense spark.

Even without sparks, the Van de Graaff generator is an interesting novelty. If you isolate yourself from the ground and touch the metal sphere while it's accumulating negative

Courtesy Lou Bloomfield

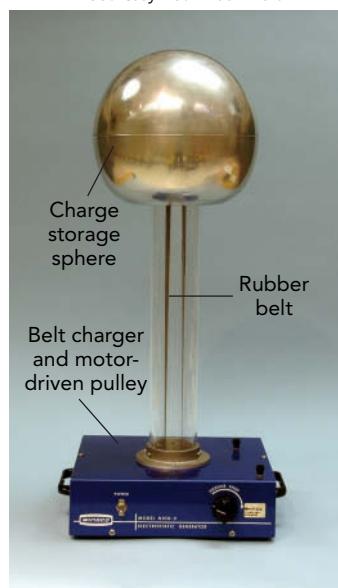


Fig. 10.1.4 Static electricity can be produced by mechanical processes. In this Van de Graaff generator, a moving rubber belt transfers negative charges from the base to the shiny metal sphere. This negative charge creates dramatic sparks as it returns through the air toward the positive charge it left behind.

charges, some of those negative charges will spread onto you as well. If your hair is long and flexible, and permits the negative charges to distribute themselves along its length, it may stand up, lifted by the fierce repulsions between those like charges.

Check Your Understanding #5: Stop the Presses!

The paper in some printing presses moves through the rollers at half a kilometer per minute. If no care is taken, dangerous amounts of static charge can accumulate on parts of the press. How does the moving paper contribute to that charging process?

Answer: Contact between dissimilar materials puts charge on the paper, which then carries that charge with it to isolated parts of the press. Enough charge can accumulate on those parts to be dangerous.

Why: Nonconductive paper is an excellent transporter of electric charge. Once the paper picks up a static charge by touching a dissimilar material, it can carry that charge with it as it moves through the press. Not surprisingly, printing presses use various tools to suppress this static charging.

Controlling Static Electricity: Fabric Softeners and Conditioners

Now that we've seen what static electricity is and how to produce it, we're ready to see how to tame it. Static cling, flyaway hair, and electrifying handshakes aren't everyone's cup of tea. The basic solution to static charge is mobility; if charges can move freely, they'll eliminate static electricity all by themselves. Opposite charges attract, so any separated positive and negative charges will join up as soon as they're allowed to move.

Materials such as metals that permit free charge movement are called **electrical conductors**. Those such as plastic, hair, and rubber that prevent free charge movement are called **electrical insulators**. Since charge movement eliminates static electricity, our troubles with static electricity stem mostly from insulators. If you wore metal clothing, you wouldn't have static problems with your laundry.

The simplest way to reduce static electricity is to turn the insulators into conductors. Even slight conductors, ones that just barely let charges move, will gradually get rid of any accumulations of separated charge. That's one of the main goals of fabric softeners, dryer sheets, and hair conditioners. They all turn insulating materials—fabrics and hair—into slight electrical conductors. The result is the near disappearance of static electricity and all its fashion inconveniences.

How these three items work is an interesting tale. They all employ roughly the same chemical: a positively charged detergent molecule. A detergent molecule is a long molecule that is electrically charged at one end and electrically neutral at the other end. Its charged end clings electrostatically to opposite charges and is chemically "at home" in water. Its neutral end is oil-like, slippery, and "at home" in oils and greases. This dual citizenship is what makes detergents so good for cleaning.

While it might seem that positively and negatively charged detergent molecules would clean equally well, that's not the case. Since cleaning agents shouldn't cling to the materials they're cleaning, it's important that the two not have opposite charges. Fabrics and hair generally become negatively charged when wet—another example of a charge shift when two different materials touch—so negatively charged detergent molecules clean much better than positively charged ones.

Positively charged detergents are still useful, however, although you mustn't apply them until after you've cleaned your clothes or hair. Because they cling so well to wet fibers, these slippery detergent molecules will remain in place long after washing and give fabrics and hair a soft, silky feel. They'll also allow those materials to conduct electricity, albeit poorly, so as to virtually eliminate static electricity!

This conductivity is due principally to their tendency to attract moisture. Water is a slight electrical conductor and damp surfaces allow charges to move around. That's why moist air decreases static electricity. By making fabrics and hair almost imperceptibly damp, the positively charged detergents allow separated charges to get back together and do away with static hair problems and laundry cling. That's why they're the main ingredients in fabric softeners, dryer sheets, hair conditioners, and even many antistatic sprays.

Check Your Understanding #6: No Lightning at Work

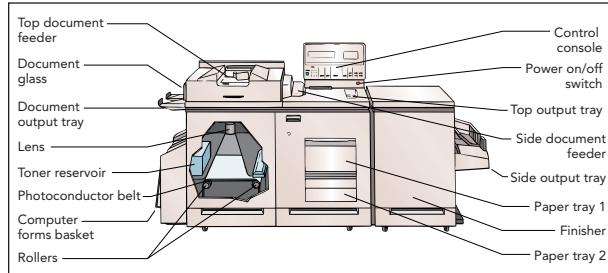
The conveyor belts used to move flammable materials often have metal threads woven into their fabric. Why are such conducting belts important for fire safety?

Answer: An insulating conveyor belt can separate enormous amounts of charge, leading to high voltages, sparks, and possibly fire. A conductive belt can't carry charge with it as it moves, so no charge accumulates.

Why: When an insulating belt has charge on its surface, that charge must move with the belt. However, charges are mobile in a conductive belt and don't normally move with it.

SECTION 10.2

Xerographic Copiers



The days of carbon paper and mimeograph machines are long gone. What modern office could operate without a xerographic copier? Advertisements for copiers are everywhere, and although each manufacturer claims to make the best copiers, that's mostly just salesmanship. In reality, all xerographic copiers are based on the same principles, discovered in 1938 by Chester Carlson. In this section, we'll examine xerographic copiers and the ideas that make them possible.

Questions to Think About: How could you use static electricity to position black powder on a sheet of paper? How would you put that static electricity on the paper? For characters to appear on the sheet, how should its static electricity be distributed? In a copier, what should light do to the static electricity to produce a copy of the original? How can a device spray static electricity onto a surface?

Experiments to Do: To get a feel for how a copier works, cut a small sheet of paper into tiny squares, about 1 mm on a side. Put the squares on a table and suspend a thin plate of clear plastic above them, a few millimeters away. The top of a clear plastic box will do. Now run a plastic comb through your hair or against a sweater several times and touch it to the top of the plastic plate. Squares of paper will leap off the table and stick to the plastic plate. What's holding the squares against the plastic? If the paper were black, how could you form letters on the surface of the plastic?

Xerography: Using Light to Print Copies

The image that a xerographic copier prints on a sheet of paper begins as a pattern of tiny black particles or *toner* on a smooth, light-sensitive surface. The copier uses static electricity and light reflected from the original document to arrange this toner on the surface and then carefully transfers the toner to the paper (Fig. 10.2.1). Invented in 1938 by Chester Carlson ³, this process is basically our old friend static electricity doing something useful.

At the heart of the xerographic copier is a thin, light-sensitive surface made from a **photoconductor**, a normally insulating material that becomes a conductor while exposed to light. Although the darkened photoconductor can keep positive and negative charges apart, these charges quickly draw together when light hits the photoconductor (Fig. 10.2.2).

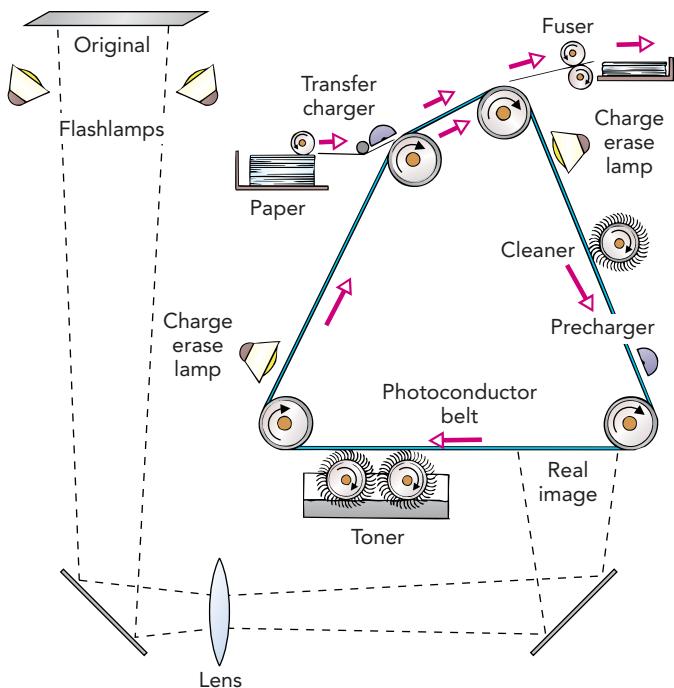


Fig. 10.2.1 This xerographic copying machine uses a photoconductor belt to form black-and-white images of an original document. The copying process begins with the precharger, which coats the photoconductor with charge. The optical system then forms a real image on a flat region of the photoconductor belt, producing a charge image. After the charge image picks up toner particles, the first charge erase lamp eliminates the charge image and weakens the toner's attachment to the belt. The toner is then transferred and fused to the paper.

3 Impoverished as a youth, American inventor Chester F. Carlson (1906–1968) supported his family by washing windows and cleaning offices after school. His work in a print shop as a teenager started him thinking about copying and he began to experiment with electrophotography. After attending Caltech (the California Institute of Technology), he worked for Bell Laboratories but was laid off during the Depression. While attending law school, he continued his experiments and invented the xerographic copying process in 1937–1938. Development of commercial copiers was slow, and it wasn't until 1960 that the Haloid Xerox Corporation produced its first successful copier, Model 914. Carlson became extremely wealthy but gave most of his money away anonymously.

That flexibility allows light from the original document to determine the pattern of static electricity on the photoconducting surface and consequently the placement of toner on the piece of paper.

Each copying cycle begins in the dark with the copier spraying negative charges onto its photoconductor. On the other side of the photoconductor is a grounded metal surface—grounded in the sense that it's electrically connected to Earth so that charges are free to flow between the two. As negative charges land on the open surface of the photoconductor, they attract positive charges onto the metal surface beneath it. When the charge-spraying process is complete, the open surface of the photoconductor is uniformly coated with negative charges while the underlying metal surface is uniformly coated with positive charges (Fig. 10.2.3a).

After this precharging, the copier uses a lens to cast a sharp image of the original document onto the photoconducting surface. We'll examine lenses and the formation of images when we study cameras in Chapter 14. For now, what matters is that light hits the photoconductor only in certain places, corresponding to the white parts of the original document.

There are two standard techniques for exposing the photoconductor to light. Some copiers illuminate the whole original document with the brilliant light of a flash lamp and cast a complete image onto a flattened portion of a photoconductor belt. In other copiers, a moving lamp or mirror illuminates the original a little at a time and the image is cast as a moving stripe on a rotating photoconductor drum. Either way, charges move through any regions of the photoconductor that are exposed to light, leaving these regions electrically neutral (Fig. 10.2.3b). The result is a *charge image*, a pattern of electric charge on the photoconductor's surface that exactly matches the pattern of ink on the original document (Fig. 10.2.3c).

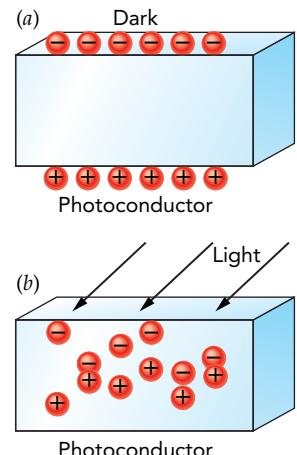


Fig. 10.2.2 (a) In the dark, a photoconductor is an electrical insulator so that separated electric charges on its surfaces remain there indefinitely. (b) When the photoconductor is exposed to light, it becomes an electrical conductor and the opposite electric charges soon join one another.

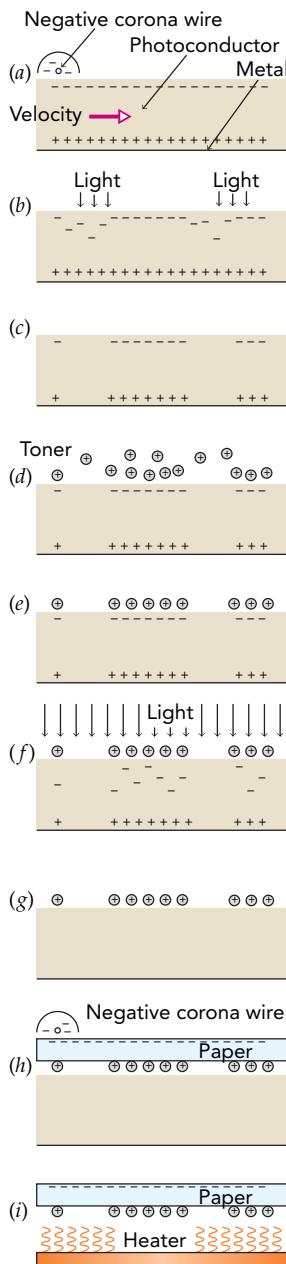


Fig. 10.2.3 The photoconductor is first coated (a) with a uniform layer of negative charge. Exposure to light (b) erases some charge to form a charge image (c). The charge image attracts (d) positively charged toner particles (e). The charge image is erased (f) to release the toner particles (g). The toner is transferred to the negatively charged paper (h) and fused to the paper with heat (i).

To develop this charge image into a visible one, the xerographic copier exposes the photoconductor to positively charged toner particles (Fig. 10.2.3d). This toner is a fine, insulating plastic powder containing a colored pigment, usually black. Applying toner to the photoconductor must be done gently, and it's often accomplished with the help of Teflon-coated iron balls. These tiny balls are held together in long filaments by a rotating magnetic shaft, so that the shaft resembles a spinning brush with extraordinarily soft bristles. These bristles wipe toner particles out of their storage tray and onto the photoconductor. Contact with the Teflon leaves the toner particles positively charged, so they stick to the negatively charged portions of the photoconductor (Fig. 10.2.3e).

The photoconductor now carries a black image of the original document, an image that the copier must transfer to the paper. Before attempting that transfer, the copier first weakens the photoconductor's grip on the toner by exposing it to light from a charge erase lamp. This light eliminates the photoconductor's charge (Fig. 10.2.3f) and leaves the positively charged toner particles clinging only loosely to its surface (Fig. 10.2.3g).

The copier then transfers the toner image to a blank sheet of paper by pressing that paper lightly against the photoconductor while spraying negative charge onto the paper's back (Fig. 10.2.3h). The positively charged toner is attracted to the negatively charged paper, and the two leave the photoconductor together. The copier then heats and presses the copy, permanently fusing the toner onto the paper (Fig. 10.2.3i). Once the image has been transferred to the paper, the copier cleans its photoconducting surface in preparation for the next copy; a second charge erase lamp eliminates any remaining charge, and a brush or squeegee mops up any residual toner.

With that introduction to xerography, you can already explain many things about copiers. For example, while fixing a copier jam, you may find that you have removed unfinished copies—ones bearing toner images that haven't yet been fused onto the paper. The toner of an unfused copy comes off on your hand because it's held in place only by electrostatic forces. When you replace the toner cartridge in a personal copier, in addition to adding new toner, you're also installing a new precharge system, photoconductor drum, and toner applicator (Fig. 10.2.4).

However, we've glossed over three important physics issues. Two we'll leave for later chapters: why a photoconductor becomes conducting when exposed to light (Chapter 13, Light), and how a lens projects an image of the document onto the photoconductor (Chapter 14, Optics and Electronics). The third issue is relevant now, and so we'll examine it carefully—how the copier sprays charges onto surfaces.

Check Your Understanding #1: Sticky Copies

When the copies emerge from a xerographic copier, they tend to stick to things and attract lint. What causes this effect?

Answer: The charge that was placed on the paper to attract the toner isn't always removed completely. Moreover, the toner itself is charged.

Why: The final transfer process, lifting the toner particles from the photoconductor to the paper, is done by charging the paper, and some of this charge remains on the paper when it leaves the copier. Copier transparencies are particularly clingy because plastic retains charge so well.

Discharges and Electric Fields

At the start of the copy cycle, the xerographic copier coats its photoconducting surface uniformly with electric charges. Because this precharging process is done in the dark, while the surface is an electrical insulator, the charges must be sprayed onto it like paint. The copier's charge sprayer is a **corona discharge**, a gentle sustained spark that forms in the air near a needle or fine wire that's kept at high voltage.

It's a type of **discharge**, a flow of electric charge through a gas. Air is normally an insulator because its atoms and molecules are neutral and can't convey charge from one place to another. However, by seeding air liberally with individual charged particles, the copier manages to turn that air into a conductor and then to produce a discharge in it. How does the copier seed the air with charges and produce its discharge? And how does it use that discharge to coat its photoconducting surface? To answer those questions, we need to know more about electrostatic forces and voltages, and about a related concept, electric fields.

Because free charges are hard to come by in the air, the copier begins with just a few charged particles and uses them to generate more. The idea is simple: the copier uses electrostatic forces to accelerate those initial charges to enormous speeds and lets them smash into air's neutral particles. When hit hard enough, a neutral air particle breaks into oppositely charged fragments and thus adds two more free charges to the air. These new charges join the mix, accelerating, colliding, and breaking up still more air particles. A cascade of collisions ensues, and the air "breaks down," transforming from an insulator to a conductor. The copier then uses this conducting air to spray the photoconductor with charges.

Where do those initial charges come from? Surprisingly, they're already there, the products of cosmic rays and natural radioactivity! Every cubic centimeter of ordinary air contains almost 2000 charged particles, roughly half positive and half negative. Considering that this same volume of air contains almost 3×10^{19} neutral particles, that's not many charges. But it's enough to get the discharge started.

To parlay those initial charges into the vast numbers it needs, the copier must accelerate them aggressively. The neutral air particles are so densely packed that it's difficult for the charged ones to pick up much speed before they hit something and slow down. To give each initial charge a good shot at breaking the first neutral particle it hits, the copier must accelerate that charge very quickly.

The copier accelerates its charges using strong electrostatic forces. Up until now, we've associated electrostatic forces with pairs of charges, each charge pushing or pulling on the other. As long as there are only a handful of charges in a given situation, the individual electrostatic forces on a particular charge can be added together to obtain the overall electrostatic force on that charge. However, in the copier's wires and discharge, there are so many individual charges that adding up their forces is virtually impossible. We need some other way to characterize the overall electrostatic force on a particular charge.

Instead of thinking about the many interactions between one particular charge and all the other charges around it, we can view the electrostatic force on our charge as the result of its one interaction with something local: an **electric field**, an attribute of space that exerts an electrostatic force on a charge. The surrounding charges create the electric field and that electric field pushes on our charge. The electrostatic force on our charge depends on the charge's location in space and time, so the value of the electric field also depends on space and time.

The electric field is an example of a **field**, a structure that associates a physical quantity with each point in space and time. The term *field* suggests another structure that extends across space and time—a field of growing wheat. Since the length of the wheat stalks depends on where and when you look, stalk length is a *field*. Moreover, in windy weather, the direction of those wheat stalks also depends on where and when you look, so the stalks are actually vectors and form a **vector field**, a structure that associates a *vector* quantity with each point in space and time. The electric field is also a vector field; at each point, its magnitude is the amount of electrostatic force it exerts per unit of electric charge and its direction is the direction in which it pushes a positive charge.

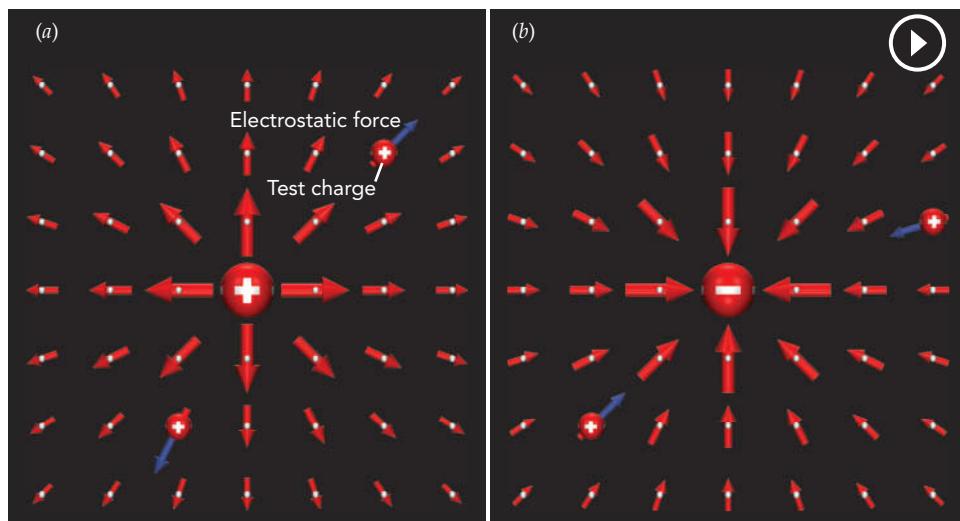
Figure 10.2.5a illustrates the electric field of a motionless positive charge. Each white dot represents a point in space and time, and the red arrow passing through that dot represents the electric field at that point. The direction of the arrow indicates the direction of the field, and its volume is proportional to the field's magnitude. Also shown are the field's effects on two **test charges**, idealized positive charges that have no electric fields of their own and thus no influence on their surroundings.

Courtesy Lou Bloomfield



Fig. 10.2.4 This xerographic copier places the photoconductor drum, toner supply, and a corona wire inside a disposable cartridge. After the paper passes through the cartridge, toner is fused onto its surface and it leaves the copier.

Fig. 10.2.5 (a) The electric field of a motionless positive charge, represented by red arrows at positions indicated by white dots. Each field vector points away from the positive charge and its magnitude is proportional to one over the square of the distance from the charge. Each of two test charges experiences an electrostatic force equal to the test charge's charge times the electric field at its position. (b) The electric field of a motionless negative charge.



Test charges don't really exist, but they're useful conceptual tools for examining electric fields. In the present case, the central positive charge's electric field pushes each test charge away from the positive charge with an electrostatic force that is inversely proportional to the distance from the positive charge. When the central charge is negative (Fig. 10.2.5b), its electric field is reversed and that field pushes each test charge toward the central charge.

From this new perspective, the electrostatic force on a charge is exerted by the electric field itself, not by the source of the electric field. That electrostatic force is equal to the charge times the electric field at the charge's position. We can write this relationship as a word equation:

$$\text{electrostatic force} = \text{charge} \cdot \text{electric field}, \quad (10.2.1)$$

in symbols:

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

and in everyday language:

Charged lint accelerates quickly in a region full of static electricity,

where the electrostatic force is in the direction of the electric field. Note that a particle carrying a negative amount of charge (an electron) experiences a force opposite the electric field. The SI unit of electric field is the **newton per coulomb** (abbreviated N/C).

At present, the electric field may seem like an unnecessary fiction. In later sections, however, we'll see that the electric field is much more than just a new way of thinking about the forces between charges. That's because the electric field truly exists in space, independent of the charges that produce it. In fact, electric fields are often created by things other than charges and can influence things other than charges as well.

The copier employs a very strong electric field to "break down" the air so that it can operate its discharge. That field accelerates charges so rapidly that collision cascades occur and fill the air with free charges. Unfortunately, you can't sense electric fields directly, so it's hard to visualize a strong one. We'll work on that problem, but for now just remember that strong electric fields can initiate discharges in air. That's how thunderstorms produce lightning!

Check Your Understanding #2: Medical Electrons

A medical linear accelerator uses a strong electric field to accelerate electrons forward and give them enormous kinetic energies. These high-energy electrons enter the patient and kill cancer cells. In which direction does the accelerator's electric field point?

Answer: It points backward, away from the patient.

Why: Since electrons are negatively charged, they accelerate in the direction opposite to the field. Since the accelerator must push the electrons forward, toward the patient, its field must point away from the patient.

Check Your Figures #1: Lint Floats

A piece of charged lint refuses to fall because an electric field is exactly supporting its weight. If the lint weighs 10^{-8} N and has a positive charge of 10^{-11} C, what electric field is supporting it?

Answer: The lint is supported by an electric field of 1000 N/C in the upward direction.

Why: The electric field is equal to the force it exerts divided by the charge, or 10^{-8} N divided by 10^{-11} C. It must point upward to support the positively charged lint against the downward pull of gravity.

Conductors and Voltage Gradients

A copier's precharging system uses the gentle corona discharge that develops in the strong electric field just outside a fine high-voltage wire. This discharge ferries charges to the photoconductor surface and coats it uniformly. To understand why a strong electric field exists outside a fine high-voltage wire and why the discharge it produces is "gentle," we need some background. Let's start by looking at electric fields inside and outside electrical conductors.

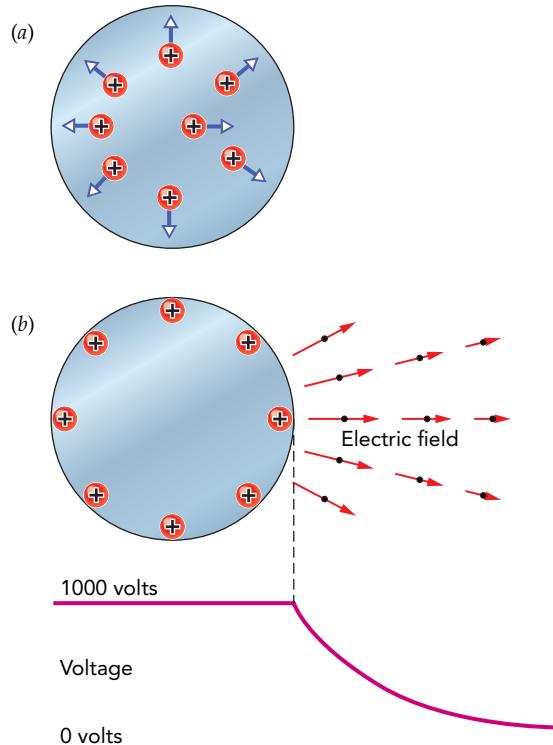
Consider the simplest conducting object—a solid metal ball. If you release some identical positive charges inside that ball (Fig. 10.2.6a), what happens to them? Because they repel one another, those charges accelerate outward and move apart. In fact, they'd leave the ball altogether if they weren't chemically bound to its metal. After spending a moment or two ridding themselves of extra electrostatic potential energy, principally as thermal energy, the charges settle down in stable equilibria on the ball's surface (Fig. 10.2.6b). At its equilibrium point, the outward electrostatic force that each charge experiences from its fellow charges is perfectly balanced by the inward chemical force it experiences from the metal. The net force on it is zero.

At equilibrium, each charge has also minimized its total potential energy. After all, it can't stop accelerating until there is no direction in which it can move to lower its potential energy further. What's amazing about how the charges arrange themselves on the ball's surface is that each one ends up with the *same* total potential energy. That's because if one of them had less total potential energy than the rest, the other charges would accelerate toward it to lower their total potential energies as well!

Since the only potential energy that significantly affects charges in our small homogeneous ball is electrostatic potential energy, every charge in our ball has essentially the same electrostatic potential energy. Because voltage is the electrostatic potential energy per unit of charge, equal potential energies on equal charges means equal voltages—the entire ball has a single, uniform voltage! In our voltage \leftrightarrow altitude analogy, this observation is analogous to the fact that, at equilibrium, the water level in a swimming pool has a single uniform altitude.

Because of the ball's perfect symmetry, charges in equilibrium are spread evenly on its surface. Had we chosen a less symmetric conducting object, such as the copier's fine metal wire, charges in equilibrium would not be spread so evenly. Nonetheless, those charges would still be on the outside of the object, and it would still have a single uniform voltage.

Fig. 10.2.6 (a) When like charges are placed inside a conducting sphere, they repel one another and accelerate toward the sphere's surface. (b) When those charges have settled in equilibrium on the sphere's surface, the sphere has a single, uniform voltage and zero electric field inside it. Outside the sphere, the voltage decreases toward zero and there is an electric field.



VOLTAGE AND CHARGE ON A CONDUCTING OBJECT

With its charges at equilibrium, a homogeneous conducting object has a single uniform voltage and the net charge anywhere in its interior is zero.

Given a conducting object's charges in equilibrium on its surface, we can make one more remarkable observation: the electric field inside that object is zero! To see why, let's place a test charge inside the object. Putting a real charge inside the object would upset the equilibrium, and that charge would be pushed toward the surface. A test charge, however, has no electric field of its own and doesn't influence its surroundings; it leaves the other charges in equilibrium and merely responds to the existing electric field. Because the voltage is uniform throughout the object, the test charge's electrostatic potential energy doesn't depend on its position and it therefore can't reduce its electrostatic potential energy by moving. It doesn't accelerate, so it must be experiencing zero electrostatic force and zero electric field.

ELECTRIC FIELD IN A CONDUCTING OBJECT

With its charges at equilibrium, a homogeneous conducting object has zero electric field in its interior.

Although the voltage is uniform *in* and *on* the copier's fine conducting wire, it varies rapidly with position *outside* that wire (Fig. 10.2.6b). Accompanying this large spatial variation in voltage is a strong electric field, officially called a **voltage gradient**. You can think of a spatial variation in voltage as a “slope” in the voltage. In our voltage ↔ altitude analogy, a voltage gradient is analogous to an altitude gradient, the slope of an ordinary hill.

Just as water accelerates swiftly down a steep slope toward a lower altitude, so (positive) charge accelerates swiftly down a large voltage gradient toward a lower voltage. Both are examples of the accelerations toward lower potential energy that we first examined in Chapter 2. In our voltage \leftrightarrow pressure analogy, a voltage gradient is analogous to a pressure gradient and charge accelerates toward lower voltage in the same way that water accelerates toward lower pressure.

Since both electric fields and voltage gradients cause charges to accelerate, it shouldn't surprise you to learn that a voltage gradient *is* an electric field. Although we'll uncover a second source of electric fields in the next chapter, we'll treat a voltage gradient and an electric field as equivalent for now. Their relationship can be written as a word equation:

$$\text{electric field} = \text{voltage gradient} = \frac{\text{voltage drop}}{\text{distance}}, \quad (10.2.2)$$

in symbols:

$$\mathbf{E} = \mathbf{Gradient}(V),$$

and in everyday language:

Charges rush down steep drops in voltage, much as bicycles rush down steep drops in hill height,

where the electric field points in the direction of the most rapid voltage decrease.

This relationship gives us a second way to look at an electric field. In addition to being the electrostatic force exerted per unit of charge, electric field is also the voltage drop per unit of distance. The SI unit of electric field therefore has a second form: the **volt per meter** (abbreviated V/m). The volt per meter is exactly the same unit as the newton per coulomb. As an example of an electric field produced by a voltage drop, consider the top of a normal 9-V battery. With its two terminals separated by just 0.005 m and a voltage drop of 9 V between them, the space between those terminals contains an electric field of about 1800 V/m, pointing toward the negative terminal.

► Check Your Understanding #3: Don't Get Out of a Hot Car!

During a thunderstorm, a lightning strike places a huge static charge on your car. Why don't you notice this charge as long as you remain inside the car?

Answer: The accumulated charge is all on the outside of the conducting car, so it will affect you only if you step outside and offer it a conducting path to the ground.

Why: Since the car body is an electrical conductor and its charges are in equilibrium, the car has a uniform voltage and there is no electric field inside the car. Outside the car, however, there is a substantial electric field. If your body provides a conducting path to the ground, that electric field will push charges through you and you will experience a shock. Similarly, if electric power lines ever fall on your car, stay inside the car to avoid a potentially lethal shock.

► Check Your Figures #2: Riding the Field

An electric field pushes charged particles through the tube of a fluorescent lamp, allowing it to produce light. To operate properly, a typical fluorescent tube needs an electric field of about 100 V/m. If the average voltage difference between the two ends of a tube is 120 V, how long can that fluorescent tube be?

Answer: It can be about 1.2 m (4 ft).

Why: With a voltage difference of 120 V between its ends and a length of 1.2 m, the tube's electric field will be 120 V divided by 1.2 m, or about 100 V/m. This result explains why so many fluorescent lamps in the United States are about 1.2 m (4 ft) long.

Courtesy Lou Bloomfield

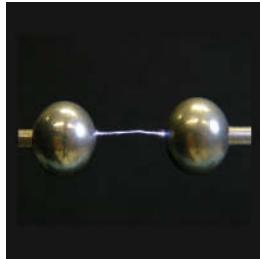


Fig. 10.2.7 These two metal spheres are 1 cm apart. When their difference in voltage is about 30,000 V, the air between them breaks down and forms a spark.

Fine Wires and High Voltages: Corona Discharges

Ordinary air breaks down in an electric field of about 3×10^6 V/m, or, in customary units, about 30,000 V per centimeter. At that field, free charges accelerate so rapidly that a cascade of charge-freeing collisions suddenly transforms air from a nearly perfect insulator into a reasonably good conductor.

You can produce such a strong field all by yourself. On a dry winter day, you can coat yourself with positive charges and raise your voltage to about 30,000 V simply by scuffing your rubber-soled shoes across an acrylic carpet. As you then approach a grounded doorknob at 0 V, the voltage difference between the doorknob and your hand will be 30,000 V. When your hand is about 1 cm from the doorknob, the electric field will reach 30,000 V per centimeter and the air will break down with a brilliant spark (Fig. 10.2.7).

Because your hand and the doorknob are similar in size and shape, the voltage changes smoothly between them (Fig. 10.2.8a). It varies steadily from 0 V on the doorknob to 30,000 V on your hand, so the voltage gradient or electric field is nearly uniform. When two objects differ significantly in size, however, the larger object dominates voltages in the space between them. For example, if you hold a long pin in your hand as you approach the doorknob, the doorknob will control the voltage most of the way to the pin and nearly all the increase in voltage will occur just outside the pin's point (Fig. 10.2.8b). Rather than being uniform, the voltage gradient or electric field will be strongest near that point.

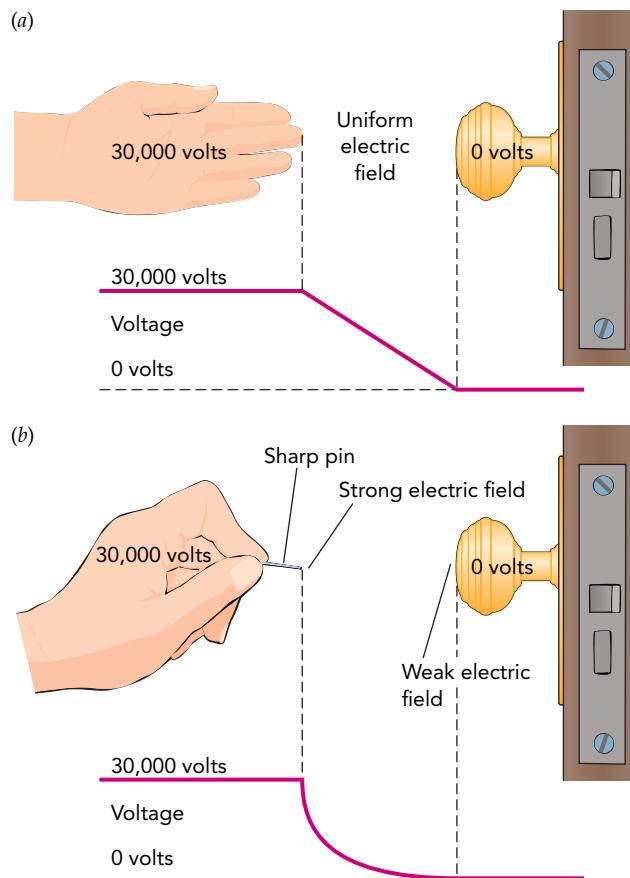
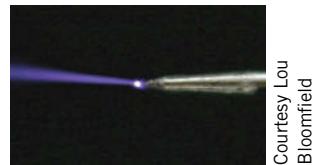


Fig. 10.2.8 Your voltage is 30,000 V when you reach for the 0-V doorknob. (a) Since your hand and the doorknob are similar in size, the voltage decreases steadily between them and the electric field is uniform. (b) When you hold out a pin, the voltage plummets near its sharp point and the electric field there is extremely strong.

The copier makes good use of this nonuniform field. Its fine high-voltage wire is nearly surrounded by a much larger metal shroud. The wire is so thin that its influence fades just a hair's breadth from its surface and the grounded shroud dominates voltage almost all the way to the wire. Although the wire's negative voltage is only -3000 V and it's about 1 cm from the shroud, the voltage changes so rapidly in the air just outside this wire that the electric field there easily exceeds $30,000\text{ V/cm}$ and breaks down the air.

The discharge that forms near the fine wire is a special, self-regulating one—a corona discharge (Fig. 10.2.9). While most discharges can't control how many free charges they produce, a corona discharge automatically maintains a steady production. Because free charges form only in the strong electric field near its thin conductor, their production rate is very sensitive to changes in that conductor's effective thickness. If there are too many free charges in the air near the conductor, their ability to conduct electricity effectively thickens the conductor, weakens the electric field, and slows the production of free charges. The discharge is correcting its own mistake.

Because of this stabilizing effect, the air in a corona discharge maintains a steady electrical conductivity that's ideal for charging a copier's photoconductor. However, corona discharges were common long before copiers. They often occur spontaneously near sharp points or fine wires at high voltages, leading to charge leakage from power transmission lines and occasionally producing a glow called St. Elmo's fire on the masts and rigging of sailing ships (see 4).



Courtesy Lou Bloomfield

Fig. 10.2.9 The electric field near this sharp, high-voltage pin is so strong that it breaks down the air and forms a corona discharge. The resulting glow is produced by air particles that receive energy from the discharge.

Check Your Understanding #4: A Safety Pin

You can avoid the shock of static electricity by holding out a sharp needle as you reach for a metal doorknob or wall. How does that needle protect you from static electricity?

Answer: The needle emits charge into the air via a corona discharge.

Why: The needle acts as your personal lightning rod. When you are carrying a net electric charge, some of that charge settles onto the needle. The strong electric field near the needle's tip initiates a corona discharge, and much of your accumulated charge leaves through it. This discharge limits your net electric charge and thus the size of any shock you experience.

Getting Ready to Copy: Charging by Induction

A corona discharge does more than just turn air into a conductor; it also produces an outward spray of electric charges. Those charges are pushed outward by the electric field surrounding the corona wire. Since the copier's corona wire has a negative voltage, the surrounding electric field points toward that wire. Because negative charges accelerate opposite an electric field, the copier's corona produces a shower of outgoing negative charges. They spray onto the photoconducting surface as it moves steadily past the corona, and the photoconductor thus acquires a uniform coating of negative charges.

As each negative charge lands, it draws a positive charge onto the grounded metal surface beneath the photoconductor and the attraction between those two opposite charges holds them firmly in place. While the photoconductor's open surface is acquiring its uniform negative charge, the metal layer underneath is acquiring an equivalent positive charge (Fig. 10.2.3a). This process, whereby a grounded conductor acquires a charge through the attraction of nearby opposite charge, is called "charging by induction."

The induced positive charge on the metal side of the photoconductor is important to the xerographic process for several reasons. First, it lowers the electrostatic potential energy of the negative charge so that the surface's negative voltage isn't as enormous. Second, without that positive layer nearby, repulsion between like charges would tend to push negative charges on the open surface toward the edges of the photoconductor and distort the resulting images.

4 Contrary to popular belief, lightning rods don't simply attract lightning strikes so as to protect the surrounding roof. Instead, they produce corona discharges that diminish any local buildups of electric charge. By neutralizing the local electric charge, the lightning rod reduces the chances that lightning will strike the house. Similar devices, called static dissipaters, are found near the tips of airplane wings and protect planes from lightning strikes.

Most significantly, however, the positive charge layer gives the negative charge layer somewhere to go when the photoconductor is exposed to light! Wherever light from the original document turns a patch of the photoconductor into a conductor, the negative and positive charge layers rush together and cancel. The resulting uncharged portion of the photoconductor subsequently attracts no toner and produces a white patch on the finished copy.

Having come full circle, we can now see how the copier achieves its goal. It uses a corona discharge to coat a photoconducting surface with negative charge and then selectively erases portions of that charge layer with light from the original document. The remaining charged patches on the photoconductor attract positively charged black toner, which is then transferred permanently to the paper.

It's worth mentioning that, for technical reasons, some copiers precoat their photoconductors with positive rather than negative charges and then use that charge to attract negatively charged toner. These copiers put high positive voltages on their fine wires so that their coronas spray positive charges.

Check Your Understanding #5: Hot Rod

A thick wire connects the lightning rod on the courthouse steeple to the ground and normally ensures that the rod is electrically neutral. However, when a negatively charged cloud floats overhead, what charge does that rod acquire?

Answer: The rod becomes positively charged.

Why: The lightning rod charges by induction—the cloud's negative charge attracts positive charge up the wire and onto the rod. The rod's sharp point initiates a corona discharge, spraying positive charge toward the cloud and gradually decreasing the cloud's charge. In that fashion, the lightning rod acts to suppress lightning strikes.

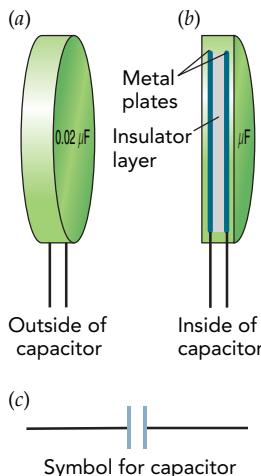


Fig. 10.2.10 (a) A capacitor is usually a disk or cylinder with two protruding wires. Its capacitance is printed on its surface. Inside (b), the wires are connected to two conducting plates that are separated by a thin insulating layer. (c) In a schematic diagram of an electronic device, the capacitor is represented by two parallel lines.

Capacitors

When it's in the dark, the copier's photoconductor system is an example of a **capacitor**, a device that stores separated positive and negative charge. Capacitors are common in modern technology, and most consist of two oppositely charged surfaces separated by a thin insulating layer. To make adding or removing charge easy, those surfaces are usually made of metals or other electrical conductors (Fig. 10.2.10). The copier's capacitor, however, has only one metal surface on its insulating photoconductor layer; the other surface is open and nonconducting (Fig. 10.2.3a). The copier uses its corona discharge to put charge on that open surface, and it uses light to remove that charge.

The two conducting surfaces of an ordinary capacitor are often called *plates*. When one plate is positively charged and the other is negatively charged, the attraction of opposite charges on those plates offsets the repulsion of like charges on each plate. The plates thus manage to store large quantities of separated charge while leaving the overall capacitor electrically neutral.

You can "charge" a capacitor's plates by transferring (positive) charge from its negative plate to its positive one. Since each charge you move experiences an electrostatic force in the opposite direction, you must do work on that charge as you push it from the negative plate to the positive plate. Your work is stored in the capacitor as electrostatic potential energy, and that stored energy is released when you let the separated charge get back together. Since (positive) charge has more electrostatic energy on the positive plate than on the negative plate, the voltage of the positive plate is higher than the voltage of the negative plate. The voltage difference between plates is proportional to the separated charge on them; the more separated charge the capacitor is holding, the larger the voltage difference.

This voltage difference also depends on the physical structure of the capacitor. Increasing the surface area of each plate decreases the repulsion of its like charges. Thinning the insulating layer between the plates increases the attraction of their opposite charges. Both changes lower the separated charge's electrostatic potential energy and, consequently, the

voltage difference between the plates. The bigger and closer the plates, the less energy it takes to store separated charge on them.

Changes that allow the capacitor to store separated charge more easily increase its **capacitance**, the amount of separated charge the capacitor holds divided by the voltage difference between its plates. The SI unit of capacitance is the coulomb per volt, also called the **farad** (abbreviated F). A capacitor with a farad of capacitance stores an incredible amount of separated charge, even at a low voltage difference, but a capacitor with a billionth of a farad of capacitance is much more typical. A capacitor's capacitance is marked on its wrapper, often in an abbreviated form. A Greek letter μ appearing in front of the F means millionths of a farad (μF , or microfarads), a letter n appearing in front of the F means billionths of a farad (nF , or nanofarads), and a letter p appearing in front of the F means trillionths of a farad (pF or picofarads).

Check Your Understanding #6: Recycling Charges

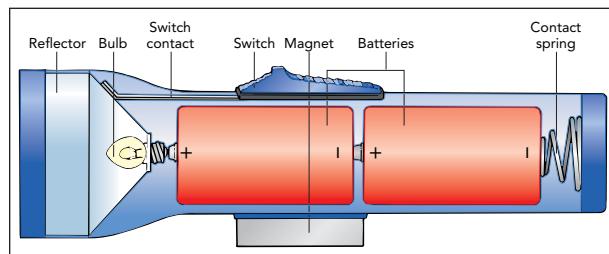
As you recycle old computer parts, you come across a capacitor and wonder if it has separated charge in it. How can you tell?

Answer: If it has separated charge, it will have a voltage difference between its two plates.

Why: If the capacitor contains no separated charge, its two plates will have the same voltage. In other words, the energy per charge on the two plates will be equal. If the plates store separated charge, however, their voltages will be different. The plate containing extra positive charge will have more energy per charge and thus a greater voltage than the plate containing extra negative charge.

SECTION 10.3

Flashlights



There isn't much to a typical flashlight; you can see the few parts it has when you open it to replace its batteries. A flashlight isn't a mechanical device, though; it's electrical. It contains an electric circuit, and most of its components are involved in the flow of electricity. To understand how a flashlight works, we need to understand how an electric circuit works and how electricity carries power from batteries to a lightbulb or LED. As we'll see, flashlights aren't as simple as they appear.

Questions to Think About: Why are some flashlights brighter than others? Why is it important that all the batteries point in the same direction? What is the difference between old batteries and new ones? What makes a flashlight suddenly become dim or bright when you shake it?

Experiments to Do: Find a flashlight that uses two or more removable batteries. Turn it on. What did the switch do to make the flashlight produce light? With the flashlight turned on, slowly open the battery compartment. The lamp will probably become dark. You should be able to turn the flashlight on and off by opening and closing the battery compartment. Why does this method work?

Replace the flashlight's batteries with older or newer ones and compare its brightness. Turn one or more of the batteries around backward, and see how that change affects the flashlight. What happens when you put a piece of paper or tape between two of the batteries? What happens when you carefully clean the metal surfaces of each battery with a pencil eraser before putting the batteries in the flashlight? Does it matter whether the flashlight uses a bulb or an LED?

Electricity and the Flashlight's Electric Circuit

A basic, old-fashioned flashlight has just three components—a battery, a lightbulb, and a switch—connected together by metal strips. When the switch is on, the strips transfer energy from the batteries to the bulb. How does energy move through the strips, and why does the switch start or stop that energy transfer? To answer these questions, we must first understand electricity and electric circuits, so that's where we'll begin.

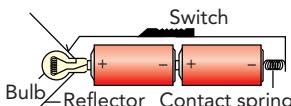


Fig. 10.3.1 A flashlight contains one or more batteries, a lightbulb, a switch, and several metal strips to connect them all together. When the switch is turned on (as shown), the components in the flashlight form a continuous loop of conducting materials. Electrons flow around this loop counterclockwise.

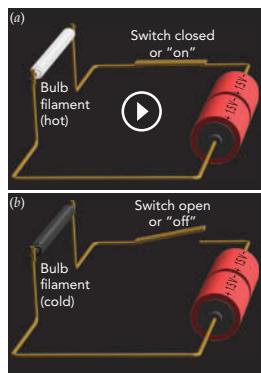


Fig. 10.3.2 (a) When the flashlight's switch is on, it closes the circuit so that current can flow continuously from the batteries, through the bulb's filament, and back through the batteries. It follows this circuit over and over again. (b) When the flashlight's switch is off, it opens the circuit so that current stops flowing.

When you turn on a flashlight, electricity conveys energy from the batteries to the bulb. An **electric current**, a current of electric charges, flows through these components, carrying the energy with it. We'll examine the exact nature of this current soon; for now, you can picture it as a steady stream of tiny positive charges following a circular route that takes them through the batteries, through the bulb, and then back to the batteries for another trip (Fig. 10.3.1). As long as the flashlight is on, charges flow around this loop, receiving energy from the batteries and delivering it to the bulb, over and over again. On route, the charges carry this energy mostly as electrostatic potential energy.

The looping path taken by charges in a flashlight is called an **electric circuit**. Because a circuit has no beginning or end, charges can't accumulate in one place, where their mutual repulsion would eventually stop them from flowing. Circuits are present in virtually all electric devices, and they explain the need for at least two wires in the power cord of any home appliance: one wire carries charges to the appliance to deliver energy, and the other wire carries those charges back to the power company to receive some more.

But what role does the switch play in all this? As part of one conducting path between the batteries and bulb, the switch can make or break the flashlight's circuit (Fig. 10.3.2). When the flashlight is on, the switch completes the loop so that charges can flow continuously around the **closed circuit** (Fig. 10.3.2a). A closed circuit appears in Fig. 10.3.3a.

However, when you turn off the flashlight, the switch breaks the loop to form an **open circuit** (Fig. 10.3.2b). Although one conducting path still connects the batteries and bulb, the loop now has a gap in it and can no longer carry a continuous current. Instead, charges accumulate at the gap and current stops flowing through the flashlight. Since energy can no longer reach the bulb, it goes dark. An open circuit appears in Fig. 10.3.3b.

There's one other type of circuit worth mentioning. A **short circuit** forms when the two separate paths connecting the batteries to the bulb accidentally touch one another (Fig. 10.3.4). This unintended contact creates a new, shorter loop around which the charges can flow. Because the bulb is expected to extract energy from the charges, it's designed to impede their flow and to convert their electrostatic potential energy into thermal energy and light. This opposition to the flow of electricity is called **electrical resistance**. Since the shortened loop offers little resistance, most of the charges flow through it, bypassing the bulb. The bulb dims or goes out altogether.

Since the bulb is the only part of the flashlight that's designed to consume electric energy, a short circuit leaves the charges without a safe place to get rid of their electrostatic potential energy. They deposit it dangerously in the batteries and the metal paths, making them hot. Since short circuits can start fires, flashlights and other electric equipment are designed to avoid them.

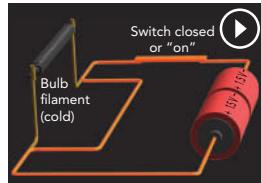


Fig. 10.3.4 When an unwanted conducting path allows current to bypass the flashlight's filament, it forms a short circuit. Because it has no proper place for electrons to deposit their energy, the short circuit becomes hot.

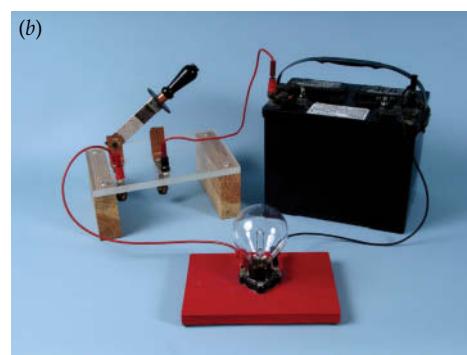
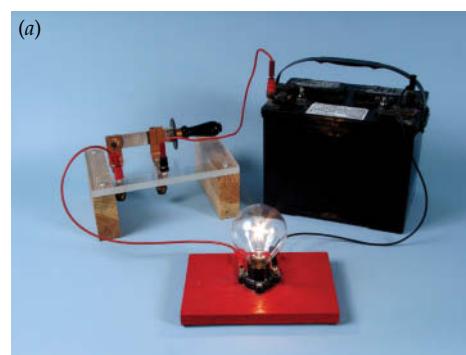


Fig. 10.3.3 When the switch is on (a), current flows around the closed circuit and carries power from the battery to the lightbulb. When the switch is off (b), no current flows through the open circuit.

Many modern flashlights use one or more light-emitting diodes (LEDs) in place of the old-fashioned bulb. However, an LED is more complicated than a bulb, and operating several LEDs at once complicates the flashlight's electric circuit. For simplicity, we'll start with a single bulb in our flashlight and replace it with one or more LEDs later in this section.

Check Your Understanding #1: Cutting the Power

If you remove just one of the wires from an automobile's battery, the vehicle will not start at all. Why doesn't the other wire supply any energy?

Answer: Removing a single wire from the battery breaks the circuit and prevents a steady flow of electric current through the car's electric system.

Why: Neither the battery nor the rest of the car can accumulate charges indefinitely. Without one wire to carry charges from the battery to the car and a second wire to return those charges from the car to the battery, accumulation will occur and charge movement will stop.

The Electric Current in the Flashlight

Each of the tiny charged particles flowing through the flashlight's circuit carries with it just a single elementary unit of electric charge and a minuscule amount of electrostatic potential energy. However, because those charges flow in astonishing numbers, they convey a considerable amount of energy per second—the quantity we know as power (see Section 2.2) and measure in watts (abbreviated W). The bulb needs a certain amount of power to keep its filament glowing brightly, and you can determine how much power is reaching the bulb by multiplying the number of elementary charges passing through the bulb each second by the amount of energy each one delivers.

There are too many elementary charges to count, though. You'll do much better to measure the circuit's **current**, that is, the amount of charge passing a particular point in the circuit per unit of time. The SI unit of current is the coulomb per second, more commonly called the **ampere** (abbreviated A). One ampere corresponds to 1 C of charge passing by the designated point each second. One coulomb is roughly 6.25×10^{18} , or 6,250,000,000,000,000,000, elementary charges, so even a 1-A current involves a tremendous flow of elementary charges.

Using electric current instead of counting charges, you can determine how much power is reaching the bulb by multiplying that current by its electrostatic energy per coulomb—the quantity we already know as voltage. For example, a current of 2 A (2 C/s) at a voltage of 3 V (3 J/C) will bring 6 W of power (6 J/s) to the bulb. Brighter flashlights involve larger currents, greater voltages, or both.

Current has a direction, pointing along the route of positive charge flow. When you turn on the flashlight in Fig. 10.3.1, charge flows around the circuit clockwise, from the battery chain's positive terminal, through the bulb's filament, through the switch, and into the battery chain's negative terminal. However, now it's time to address an awkward issue: the positive charges that flow clockwise around this circuit are fictitious. In reality, the electric current is carried by negatively charged electrons heading in the opposite direction!

As mentioned before, this issue dates back to Franklin's unfortunate choice of which charge to call positive and which to call negative. By the time scientists discovered the electron and realized that these negatively charged particles carry currents in wires, current had already been defined as pointing in the direction of positive charge flow. Since it was far too late to make current and electron flow point in the same direction, scientists and engineers simply pretend that current is carried by fictitious positive charges heading in the current's direction.

This fiction actually works extremely well, as illustrated by a simple example. When negatively charged electrons flow to the right through a neutral piece of wire, the wire's

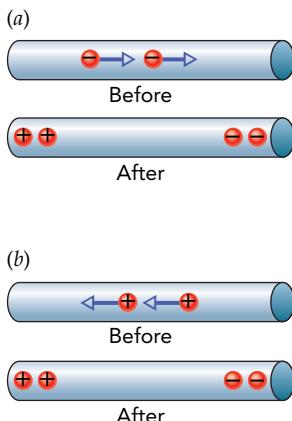


Fig. 10.3.5 A current of negatively charged particles flowing to the right through a piece of wire (a) can't easily be distinguished from a current of positively charged particles flowing to the left (b). The end result of both processes is an accumulation of positive charge on the left end of the wire and negative charge on its right end.

right end becomes negatively charged and its left end becomes positively charged (Fig. 10.3.5a). But exactly the same thing would happen if a current of fictitious positively charged particles were to flow to the left through that same piece of wire (Fig. 10.3.5b). Without sophisticated equipment, you can't tell whether negative charges are flowing to the right or positive charges are flowing to the left because the end results are essentially indistinguishable.

We, too, adopt this fiction and pretend that current is the flow of positively charged particles. In this and subsequent chapters, we'll stop thinking about electrons and imagine that electricity is carried by positive charges moving in the direction of the current. There are only a few special cases in which the electrons themselves are important, and we'll consider those situations separately when they arise.

Check Your Understanding #2: Don't Touch That Pipe

A walk across wool carpeting in rubber-soled shoes has left you covered with negative charges. If you bring your hand near a large piece of metal, the negative charges will leap through the air as a spark to reach the metal. Which way is current flowing in this spark?

Answer: The current flows from the metal toward your hand.

Why: Because current is defined as the flow of positive charges, it points in the direction opposite the flow of negative charges. Thus, the current flows from the metal toward your hand. These charges move only briefly because there is no circuit. For the charges to move continuously, they would have to be recycled and a circuit would be essential.

Batteries

While a battery is basically a portable source of electric power, here are two other interesting ways to think of it. The first is rather abstract: a battery is a type of pump. It “pumps” charge from low voltage to high voltage, much as a water pump pumps water from a low altitude to a high altitude or as a second water pump pumps water from low pressure to high pressure. Once again, our voltage ↔ altitude and voltage ↔ pressure analogies are helpful. Each pump moves something against its natural direction of flow, pushing it forward and doing work on it in the process. The battery increases a charge’s electrostatic potential energy by pushing it up a voltage gradient. The first water pump increases water’s gravitational potential energy by pushing it up an altitude gradient. The second water pump increases water’s pressure potential energy by pushing it up a pressure gradient.

The second perspective on batteries is more mechanical: a battery is a chemically powered machine. It uses chemical forces to transfer charges from its negative terminal to its positive terminal. As positive charges accumulate on the battery’s positive terminal, the voltage there rises, and as negative charges accumulate on the battery’s negative terminal, the voltage there drops. Since the battery does work transferring charges from low voltage to high voltage, it is converting its chemical potential energy into electrostatic potential energy in these separated charges.

A battery’s rated voltage reflects its chemistry, specifically the amount of chemical potential energy it has available for each charge transfer. As the voltage difference between its terminals increases, so does the energy required for each charge transfer. Eventually, the chemicals can’t do enough work on a charge to pull it away from the negative terminal and push it onto the positive terminal, so the transfers stop. The battery is then in equilibrium—the electrostatic forces opposing the next charge transfer exactly balance the chemical forces promoting it. A typical alkaline battery reaches this equilibrium when the voltage of its positive terminal is 1.5 V above the voltage of its negative terminal. Lithium batteries, with their more energetic chemistries, can achieve voltage differences of 3 V or more.

When you turn on the flashlight, you upset its equilibrium by allowing charges to leave the battery’s positive terminal for its negative terminal. With fewer separated charges now

on its terminals, the battery's voltage difference decreases slightly and it begins pumping charges again. That renewed charge transport replenishes the terminals' separated charges and opposes any further decrease in the battery's voltage. In this manner, a 1.5-V alkaline battery maintains a nearly steady voltage difference of 1.5 V between its terminals, whether its flashlight is on or off.

That alkaline battery is powered by an electrochemical reaction in which powdered zinc at its negative terminal reacts with manganese dioxide paste at its positive terminal. This reaction resembles controlled combustion. In effect, the battery "burns" zinc to obtain the energy it needs to pump charges from its negative terminal to its positive terminal. However, as the battery consumes its chemical potential energy, its ability to pump charges diminishes. When its chemicals are nearly exhausted, the battery's increasing disorder reduces its voltage. An aging battery can pump less current than a fresh one, and it provides that current with less voltage. Ultimately, less power reaches the flashlight's bulb and it goes dim.

Most flashlights use more than one battery. When two alkaline batteries are connected together in a chain, so that the positive terminal of one battery touches the negative terminal of the other, the two batteries work together to pump charges from the chain's negative terminal to its positive terminal (Fig. 10.3.6). Each battery pumps charges until its positive terminal is 1.5 V above its negative terminal, so the chain's positive terminal is 3.0 V above its negative terminal. Because charges never leave the flashlight's circuit, only relative voltages matter in that circuit. We'll find it convenient to ignore the flashlight's absolute voltages and define the voltage of the battery chain's negative terminal to be 0 V (Fig. 10.3.6). With that choice, the voltage of its positive terminal becomes 3.0 V.

The more batteries in the flashlight's chain, the more energy a charge receives overall and the more the voltage will increase from the chain's negative terminal to its positive terminal. A flashlight that uses six alkaline batteries in its chain has a positive terminal that is 9 V above its negative terminal. A typical 9-V battery actually contains a chain of six miniature 1.5-V batteries, arranged so that their voltages add up to 9 V (Fig. 10.3.7).

If you reverse one of the batteries in a chain, the reversed battery will extract energy from any charge passing through it (Fig. 10.3.8). While the chain may still pump charge from its negative terminal to its positive terminal, its overall voltage will be reduced because, instead of adding 1.5 V to the chain's overall voltage, the reversed battery will subtract that amount. If the chain has three batteries, two will add energy to the charge while the third will subtract it, and the chain's overall voltage will be only 1.5 V.

As the reversed battery extracts energy from the charges passing through it, at least some of that extracted energy is converted into chemical potential energy. The reversed battery is recharging! Battery chargers follow that concept, pushing current backward through a battery—from its positive terminal to its negative terminal—to restore the chemical potential energy in a rechargeable battery. However, normal alkaline batteries are "nonrechargeable," meaning that they turn most of the recharging current's energy into thermal energy instead of chemical potential energy. Nonrechargeable batteries may overheat and explode during recharging.

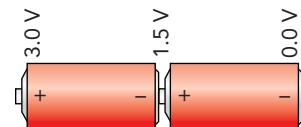


Fig. 10.3.6 When two 1.5-V batteries are connected in a chain, their voltages add so that the chain's positive terminal has a voltage that is 3.0 V higher than the chain's negative terminal. If the chain's negative terminal is at 0 V, then the chain's positive terminal is at 3.0 V.



Courtesy Lou Bloomfield

Fig. 10.3.7 A 9-V battery actually contains six small 1.5-V cells, connected in a chain. Positive charges that enter the chain at the battery's negative terminal pass through all six cells before arriving at the battery's positive terminal.

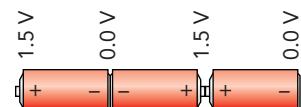


Fig. 10.3.8 When one battery in a chain of three is reversed, the reversed battery's voltage is subtracted from the sum of the others. The chain's positive terminal has a voltage only 1.5 V higher than its negative terminal. The reversed battery recharges.

Check Your Understanding #3: Car Batteries

A lead-acid automobile battery provides 12 V between its negative terminal and its positive terminal. It actually contains six individual batteries, connected in a chain. How much voltage does each individual battery provide?

Answer: Each individual battery provides 2 V.

Why: The voltages of the individual batteries add, so it takes six 2-V batteries to yield a 12-V chain. If one of the batteries becomes weak, through damage, loss of fluid, or consumption of its chemical potential energy, the voltage of the chain will drop below 12 V. Pushing current backward through a lead-acid battery recharges it.

Bulbs and Metal Strips

Whereas a battery gives charges electrostatic potential energy by pushing them *up* a voltage gradient, a bulb releases that electrostatic potential energy by letting charges slide *down* another voltage gradient. Those two devices make a perfect pair: the battery provides electric power, and the bulb consumes it. The bulb uses this power to heat its tungsten filament so hot that it glows yellow-white—but how does electricity heat the filament?

Consider a flashlight with two alkaline batteries (Fig. 10.3.9a). The bulb's filament is a fine wire, and its two ends are electrically connected to the battery chain's terminals. With one end at 3.0 V and the other at 0.0 V, the filament has a voltage gradient across it and therefore an electric field. How is that possible? While discussing copiers, we observed that a conductor has a uniform voltage throughout. Isn't the filament violating that rule?

No, it isn't. While a conductor has a uniform voltage when its charges are in *equilibrium*, the charges in the bulb are in equilibrium only when the flashlight is off. When you switch the flashlight on, you impose a 3.0-V difference between the two ends of the filament and the filament's charges immediately begin accelerating down the voltage gradient toward the 0.0-V end.

In our voltage \leftrightarrow altitude analogy, it's as though you suddenly tilted a level field to create a hill and water that had been lying motionless on that field now accelerated downhill. However, a better analog to the individual charges that we're considering at the moment would be bicyclists. Picture hundreds of bicyclists on a level field that suddenly tilts to form a hill. All the bicyclists that were at equilibrium on the level field now accelerate downhill.

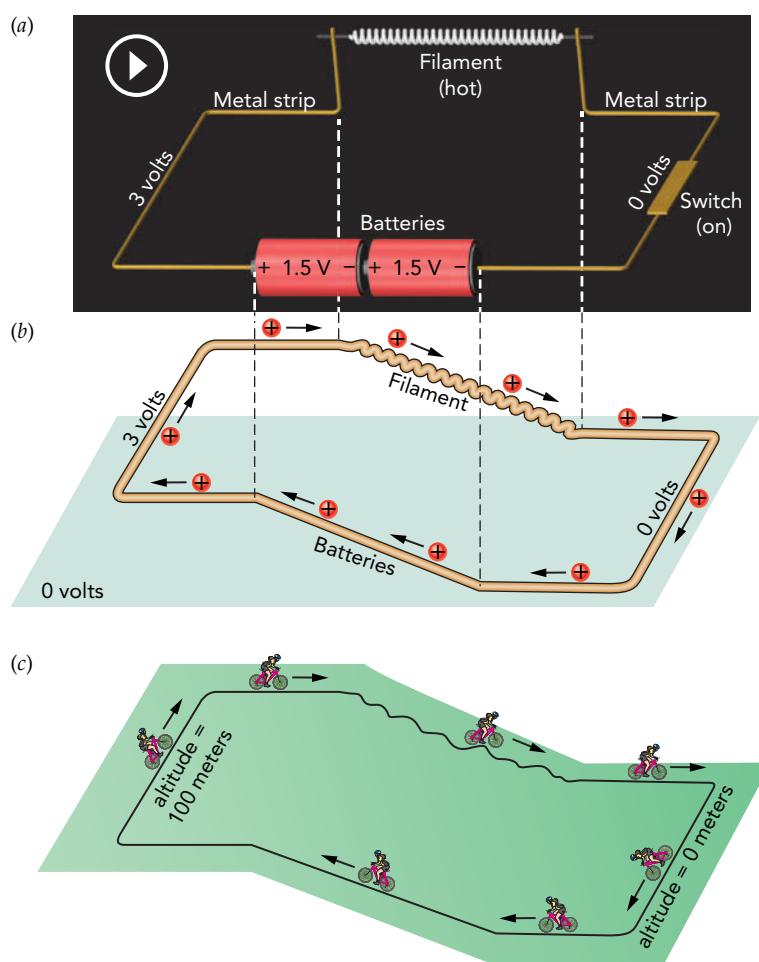


Fig. 10.3.9 (a) Current in a flashlight's circuit conveys power from the batteries to the filament. (b) Its voltage rises in the batteries and decreases in the filament. The filament's voltage drop leads to an electric field that keeps the charges moving forward at constant velocity despite many collisions in the filament. (c) This behavior is analogous to bicyclists pedaling up a smooth hill and then rolling down a rough one at constant velocity.

If the filament were a perfect conductor of electricity, each charge would accelerate steadily down the voltage gradient and convert its electrostatic potential energy into kinetic energy. However, the filament has a large electrical resistance—it significantly impedes the flow of electric current. Instead of accelerating smoothly from one end of the filament to the other, each charge bounces its way down the voltage gradient, colliding frequently with the filament's tungsten atoms and giving up kinetic energy with each collision (Fig. 10.3.9b). What began as ordered electrostatic potential energy in the charges becomes thermal energy in the tungsten atoms, and the filament glows brightly. Referring again to our voltage \leftrightarrow altitude analogy, picture the bicyclists riding down a rough hill strewn with rocks and trees (Fig. 10.3.9c). They pick up bruises instead of speed.

What about the metal strips connecting the flashlight's batteries and bulb? These strips are excellent conductors, but like all ordinary electrical wires, they're not quite perfect. Each strip has a small electrical resistance, so charges can't coast through it because of inertia alone. Instead, the charges need a small voltage gradient to keep them moving forward, and they emerge from the strip at a slightly lower voltage than where they entered. The missing energy has become thermal energy, slightly heating the metal strip. In general, the less electrical resistance in the strips carrying current to and from the bulb, the less power is wasted on route and the more power reaches the bulb. That's why it's so important to use thick metal strips or even the flashlight's metal case in the connections.

A poor connection anywhere in the circuit can spoil this efficient transfer of power. If there is dirt or grease on a battery terminal or worn materials in the switch, the current will have to pass through a large electrical resistance and waste power. Improving that connection, either by shaking the flashlight or by cleaning the metal surfaces, will increase the current flow through the circuit, reduce the wasted power, and brighten the flashlight.

Check Your Understanding #4: You Get What You Pay For

The battery in your car is dead, so you use cheap jumper cables to connect the electric system in your car to the electric system in your friend's car. When you try to start your car, too little power reaches it to start its engine. What's wrong with those cables?

Answer: The cables have relatively large electrical resistances.

Why: Starting your car requires a huge current, and the wires supplying that current must not limit it or waste its energy. Cheap jumper cables have too much electrical resistance to fulfill those requirements. There is no substitute for good, thick jumper cables—they're worth the extra money.

Voltage, Current, and Power in Flashlights

When you turn on the flashlight, an electric current carries power from its two alkaline batteries to its bulb. Let's suppose that a current of 1 A is flowing through the flashlight's circuit and take a look at how much power is being transferred.

A bulb consumes electric power because the current passing through it slides down the voltage gradient and there is a drop in voltage between where the current arrives at the filament and where it leaves the filament. This **voltage drop** measures the electrostatic potential energy each unit of charge loses while struggling through the filament. Multiplying the voltage drop by the current passing through the bulb gives you the power consumed by the bulb. This observation can be written as a word equation:

$$\text{power consumed} = \text{voltage drop} \cdot \text{current}, \quad (10.3.1)$$

in symbols:

$$P = V \cdot I,$$

and in everyday language:

A large electric current dropping from high voltage to low voltage is like the torrent of water dropping from high to low over Niagara Falls: both release a lot of power.

Since the voltage drop across the bulb is 3.0 V and the current passing through it is 1.0 A, it's consuming 3.0 W of power.

A battery chain produces electric power because the current passing through it is pushed up a voltage gradient and there is a rise in voltage between where the current arrives at the battery chain and where it leaves the battery chain. This **voltage rise** measures the electrostatic potential energy each unit of charge gains while being pumped through the batteries. Multiplying the voltage gain by the current passing through the batteries gives you the power provided by the batteries. This observation can be written as a word equation:

$$\text{power provided} = \text{voltage rise} \cdot \text{current}, \quad (10.3.2)$$

in symbols:

$$P = V \cdot I,$$

and in everyday language:

Raising a large current from low voltage to high voltage is like pumping a huge stream of water from low to high to fight a fire at the top of a skyscraper: both require a lot of power.

Since the voltage rise across the chain is 3.0 V and the current passing through it is 1.0 A, it's providing 3.0 W of power.

Check Your Understanding #5: Current Trends in Music

A large battery powers your portable radio. Current enters the radio through one wire and leaves through another. Which wire has a higher voltage?

Answer: The wire through which current enters the radio.

Why: The radio is consuming power, so the current passing through it is experiencing a voltage drop. The current has a higher voltage when it enters the radio than when it leaves.

Check Your Figures #1: When You Turn the Ignition Key

When you first start your car, its cold engine is difficult to turn and a 200-A current flows through its starter motor. If there is a voltage drop of 12 V between where the current enters the starter and where it leaves the starter, how much power is being consumed?

Answer: 2400 W of power is being consumed.

Why: The voltage drop across the starter motor is 12 V (each coulomb of charge loses 12 J of energy in passing through it), and the current through it is 200 A (200 C of charge pass through it each second). We can use Eq. 10.3.1 to determine the power the motor is consuming:

$$\begin{aligned} \text{power consumed} &= \text{voltage drop} \cdot \text{current} \\ &= 12 \text{ V} \cdot 200 \text{ A} = 2400 \text{ W}. \end{aligned}$$

Check Your Figures #2: When You Turn the Ignition Key Again

When you restart the car, its warm engine is easier to turn and a smaller current of 150 A flows through the car's starter motor. The battery is supplying power to this current, and there is a voltage rise of 12 V between where the current arrives at the battery and where it leaves the battery. How much power is the battery providing?

Answer: It is providing 1800 W of power.

Why: The voltage rise across the battery is 12 V (each coulomb of charge gains 12 J of energy in passing through it), and the current through it is 150 A (150 C of charge pass through it each second). We can use Eq. 10.3.2 to determine the power the battery is providing:

$$\begin{aligned}\text{power provided} &= \text{voltage rise} \cdot \text{current} \\ &= 12 \text{ V} \cdot 150 \text{ A} = 1800 \text{ W}.\end{aligned}$$

Choosing the Bulb: Ohm's Law

Our flashlight's bulb is designed to operate properly with a voltage drop of 3.0 V. Subjected to that voltage drop, it will carry a current of 1 A and thus consume 3 W of electric power—just enough to make it glow properly. If you were to use the wrong bulb in this flashlight, one designed for a different voltage drop, its filament would carry the wrong amount of current and receive the wrong amount of power. Too much power would quickly burn out its filament, while too little power would make the filament glow dimly.

The bulb's filament must clearly match the flashlight, particularly the voltage of the flashlight's battery chain. For example, flashlights that use many batteries require bulb filaments that are designed to operate with large voltage drops. Why is the current carried by a particular bulb filament related to the voltage drop across it, and why do different bulbs respond differently to a particular voltage drop?

The relationship between current and voltage drop is the result of collisions. Charges effectively stop each time they crash into metal atoms, so they need the push of an electric field to keep them moving forward (Fig. 10.3.10). Doubling that electric field doubles each charge's average speed and, because the number of mobile charges in the filament is fixed, also doubles the overall current flowing through the filament. Since the electric field that propels this current is the filament's voltage gradient, doubling the voltage drop through the filament doubles the current as well.

Returning to our voltage \leftrightarrow altitude analogy, picture bicyclists riding on extremely rocky terrain without pedaling. These lazy bicyclists effectively stop each time they crash into rocks, so they need the push of a slope to keep them moving forward. Doubling the

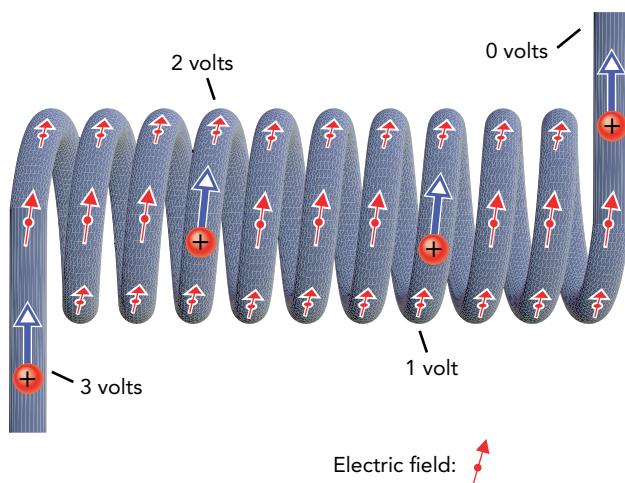


Fig. 10.3.10 This filament has a voltage drop of 3 V between its two ends. Charges moving through this filament are pushed forward by the resulting electric field. They maintain a constant speed despite frequent collisions with tungsten atoms.

⁵ German physicist Georg Simon Ohm (1787–1854) served as a professor of mathematics, first at the Jesuits' college of Cologne and then at the polytechnic school of Nuremberg. His numerous publications were undistinguished, with the exception of one pamphlet on the relationship between current and voltage. This extraordinary document, written in 1827, was initially dismissed by other physicists, even though it was based on good experimental evidence and explained many previous observations by others. In despair, Ohm resigned his position at Cologne, and it was not until the 1840s that his work was accepted. He was finally appointed professor of physics at Munich only two years before his death.

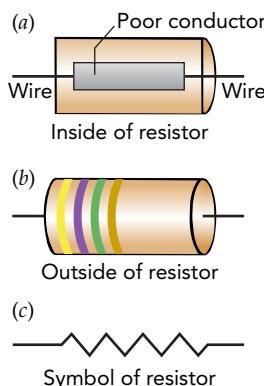


Fig. 10.3.11 (a) A resistor is two wires with an imperfect conductor of electricity between them. (b) It's usually encased in a cylindrical shell, with colored stripes to indicate its resistance. (c) In a schematic diagram of an electronic device, the resistor is represented by a zigzag line.

slope's altitude gradient—the altitude drop per meter of downhill travel—doubles each bicyclist's average speed and, because the number of bicyclists who can fit on the hill at one time is fixed, also doubles the overall current of bicyclists rolling down the hill. Since the slope that propels this bicyclist current is the hill's altitude gradient, doubling the hill height doubles the current of bicyclists as well.

The influence of filament choice on current flow reflects the different electrical resistances of those filaments. Anything that increases the number of mobile electric charges across the filament's width or allows those charges to maintain a higher average speed for a given voltage drop will decrease the filament's electrical resistance and increase the current flowing through it. In fact, electrical resistance is defined as the voltage drop through the filament divided by the current that arises as a result. Making the filament thicker or shorter will lower its resistance, as will changing its composition to make collisions less frequent.

Again our voltage ↔ altitude analogy with bicyclists on a hill is helpful. Anything that increases the number of bicyclists across the hill's width or allows those bicyclists to maintain a higher average speed for a given hill height will decrease the hill's "bicycle resistance" and increase the current of bicyclists rolling down it. In fact, "bicycle resistance" is defined as the hill height divided by the current of bicyclists it produces. Making the hill wider or shorter will lower its bicycle resistance, as will changing its rockiness to make collisions less frequent.

Combining these observations, we see that the current flowing through the filament is proportional to the voltage drop through it and inversely proportional to the filament's electrical resistance, which can be written:

$$\text{current} = \frac{\text{voltage drop}}{\text{electrical resistance}}. \quad (10.3.3)$$

This relationship is called **Ohm's law**, after its discoverer Georg Simon Ohm ⁵. Structuring the relationship this way separates the causes (voltage drop and electrical resistance) from their effect (current flow). However, this equation is often rearranged to eliminate the division. The relationship then takes its customary form, which can be written as a word equation:

$$\text{voltage drop} = \text{current} \cdot \text{electrical resistance}, \quad (10.3.4)$$

in symbols:

$$V = I \cdot R,$$

and in everyday language:

Long, skinny jumper cables have large resistances. When you connect them to a battery to jumpstart your car, they'll carry a relatively small current and your car probably won't start.

The SI unit of electrical resistance, the volt per ampere, is called the **ohm** (abbreviated Ω , the Greek letter omega). Despite its simplicity, Ohm's law is extremely useful in physics and electrical engineering. It applies to so many systems that nearly everything is **ohmic**, that is, can be characterized by an electrical resistance. Once an object's electrical resistance is known, the current flowing through it can be calculated from its voltage drop or its voltage drop can be calculated from the current flowing through it. Devices that obey Ohm's law are common in modern technology, often as simple electronic components known as **resistors** (Figs. 10.3.11). A resistor carries a current equal to its voltage drop divided by its resistance, and it experiences a voltage drop equal to the current it carries times its resistance.

OHM'S LAW

The voltage drop through a wire is equal to the current flowing through that wire times the wire's electrical resistance.

Finally, an object's electrical resistance is typically temperature dependent. Rising temperature increases the number of mobile charges in an object but also makes them collide more frequently with the jiggling atoms. If the increasing collision frequency dominates, as it does in metals, an object's resistance increases with temperature. For example, a filament carries less and less current as it approaches operating temperature, a behavior that helps it avoid overheating. However, if the increase in mobile charges dominates, as it does in semiconductors, an object's resistance decreases with temperature. This explains why semiconductor-based computer chips carry more and more current as they get hotter and will self-destruct at excessive temperatures.

Check Your Understanding #6: Skin Protection

Your skin has a much larger electrical resistance than your tissues. If you touch the two terminals of a battery with your fingers, where is the larger voltage drop—in your skin or in your tissues?

Answer: The larger voltage drop is in your skin.

Why: The fluids in your body resemble saltwater when exposed to voltages; they conduct current relatively well. If it weren't for your skin's large electrical resistance, even battery voltages would be capable of pushing large currents through you and might disrupt your heart and other functions. However, your skin's high resistance protects you from battery voltages. As long as your skin is dry and intact, it usually takes higher voltages to push enough current through you to cause injury.

Check Your Figures #3: Light Resistance

Two 3-W flashlight bulbs have different resistances: $3\ \Omega$ and $12\ \Omega$. Which bulb is meant to operate from two 1.5-V alkaline batteries?

Answer: It's the $3\ \Omega$ bulb.

Why: Equation 10.3.3 indicates that, with the voltage drop of 3 V supplied by the two alkaline batteries, the $3\ \Omega$ bulb will carry:

$$\begin{aligned} \text{current} &= \frac{\text{voltage drop}}{\text{electrical resistance}} \\ &= \frac{3\ \text{V}}{3\ \Omega} = 1\ \text{A}. \end{aligned}$$

Equation 10.3.1 then shows that this bulb consumes the specified 3 W:

$$\begin{aligned} \text{power consumed} &= \text{voltage drop} \cdot \text{current} \\ &= 3\ \text{V} \cdot 1\ \text{A} = 3\ \text{W}. \end{aligned}$$

With a voltage drop of 3 V, the $12\ \Omega$ bulb will carry a current of 0.25 A and consume just 0.75 W of power. It needs a voltage drop of 6 V to consume 3 W.

LED Flashlights: Series and Parallel Circuits

Unfortunately, lightbulbs drain flashlight batteries quickly and burn out at inconvenient times. Since LEDs are much more energy efficient than bulbs and last almost forever, it's no surprise that LED flashlights are rapidly replacing bulb flashlights.

LEDs are sophisticated semiconductor devices that we'll examine in Chapter 13. I have ignored them until now because they're *nonohmic*; that is, they don't obey Ohm's law and cannot be characterized by resistance. Instead, an LED conducts zero current when the voltage drop across it is less than a threshold voltage of about 2–4 V, depending on color, and it conducts a large, potentially damaging current when the voltage drop is significantly more than that threshold. Because small changes in voltage drop across the LED dramatically change the current passing through it, the LED is difficult to control on its own. That's why an LED is often paired with a resistor.

In a simple LED flashlight, current leaving the positive terminal of the batteries flows sequentially through a resistor and an LED before returning to the negative terminal of the batteries (Fig. 10.3.12a). That arrangement, in which the same current flows sequentially through each component, is known as a **series circuit**. Although all the components in a series circuit carry the same current, they share the circuit's overall voltage drop (Fig. 10.3.12b). To apply the voltage ↔ altitude analogy, a current of charge flowing through a series of components is like a current of water flowing down a series of waterfalls (Fig. 10.3.12c).

The overall voltage drop available to the LED flashlight's power-consuming components is equal to the voltage rise provided by its batteries. Because they are in series, the resistor and the LED must share that overall voltage drop (neglecting the small voltage drops in the metal strips and switch). Since they also carry the same current, the resistor's ohmic behavior limits the current passing through the entire circuit, including the LED. In effect, the resistor and LED negotiate—each component takes enough of the overall voltage drop to allow it to conduct the same current as the other component. Following the negotiation, the voltage drop across the LED is slightly more than its threshold voltage and the rest of the overall voltage drop is taken by the resistor. The resistor is chosen so that, when subject to that voltage drop, it conducts the right amount of current to power the LED properly.

When an LED flashlight has more than one LED, the flashlight must supply each with current. This is usually done by dividing the current from the batteries into parts and sending one part through each of the LEDs (Fig. 10.3.13a). That arrangement, in which the current is divided into parts that flow simultaneously through each component, is known as a **parallel circuit**. Although each component in a parallel circuit carries its own fraction of the overall current, all the components experience the same voltage drop (Fig. 10.3.13b). In the voltage ↔ altitude analogy, a current of charge flowing through components in parallel is like a current of water flowing down waterfalls in parallel (Fig. 10.3.13c).

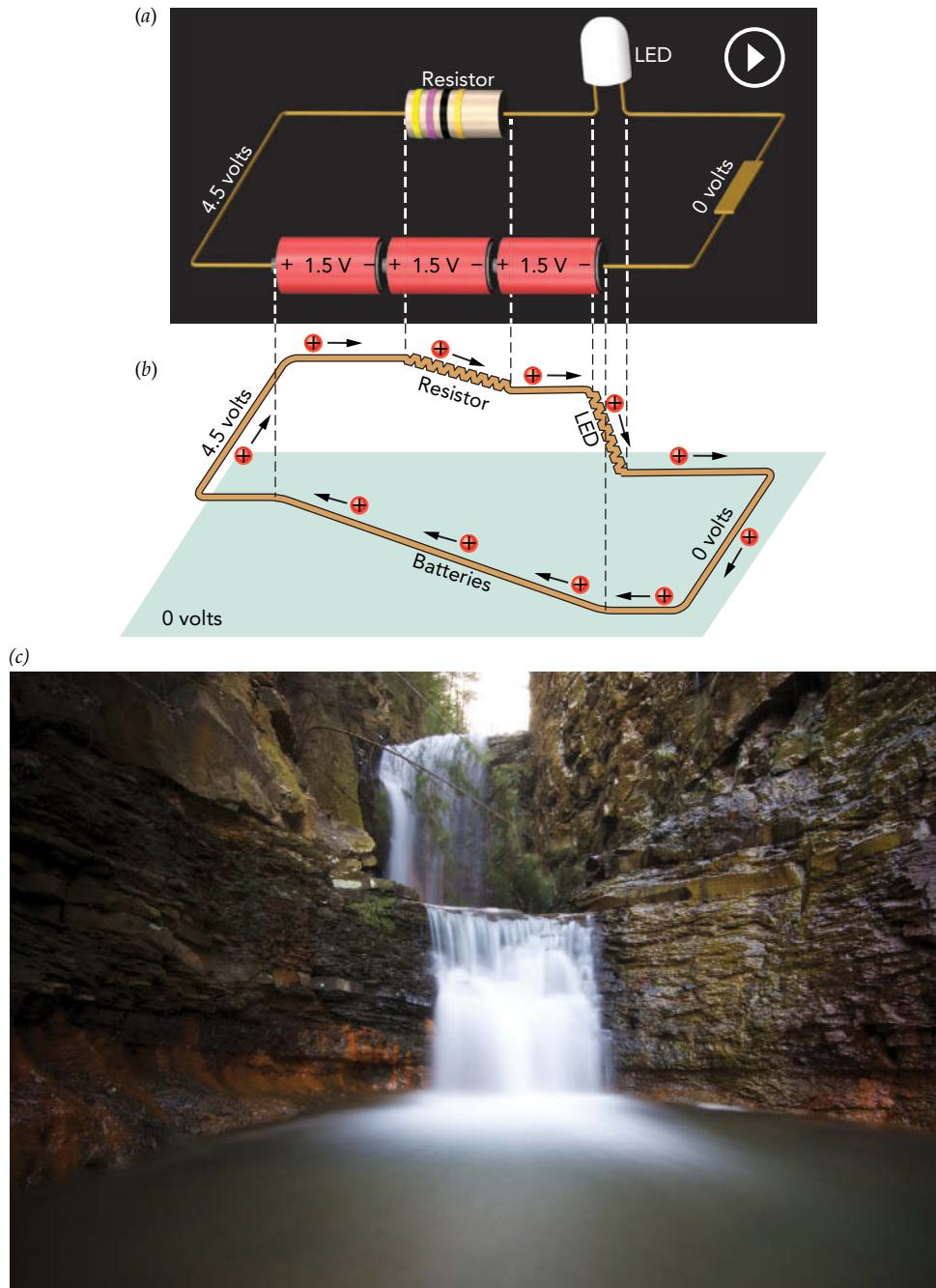
In a typical multiple-LED flashlight, each LED is actually paired with a resistor to regulate its current. The flashlight thus combines parallel and series circuits. Each LED and its resistor are connected in series, so they carry the same current but share the overall voltage drop. The multiple LED-resistor pairs are connected in parallel, so they carry separate portions of the circuit's overall current but experience the same voltage drop.

Check Your Understanding #7: Electric Defrosters

Your car has an electric defroster on its rear window. That defroster consists of 12 thin metal strips bonded to the inside surface of the glass. All the driver-side ends join together and all the passenger-side ends join together. When you switch on the defroster, current flows from the positive terminal of the battery to the driver-side end of the defroster and from the passenger-side end of the defroster to the negative terminal of the battery. Do the individual defroster strips form a series circuit or a parallel circuit?

Answer: The defroster strips form a parallel circuit.

Why: Current from the battery's positive terminal is divided into 12 portions, one for each of the 12 defroster strips. After flowing through the strips, those portions rejoin into a single current to return to the battery's negative terminal. Because the strips are connected in parallel, they carry separate currents but experience the same voltage drop.



© JasonMichael/Flickr/Getty Images, Inc.

Fig. 10.3.12 (a) Current in an LED flashlight's series circuit flows sequentially through the resistor and LED. (b) The voltage rises in the batteries and drops in the resistor and the LED. Because the resistor and LED are in series, they carry the same current but share the overall voltage drop. (c) These waterfalls are in series. The waterfalls carry the same current of water, but they share the overall altitude drop.

Epilogue for Chapter 10

This chapter has dealt with three venues in which charge and electricity play important roles. In Static Electricity, we introduced the concept of electric charge and discussed the attractive and repulsive forces that charged particles exert on one another. We studied the electrostatic potential energies stored in those charge forces and the relationship between this energy and the voltages of various locations. We learned how different objects acquire

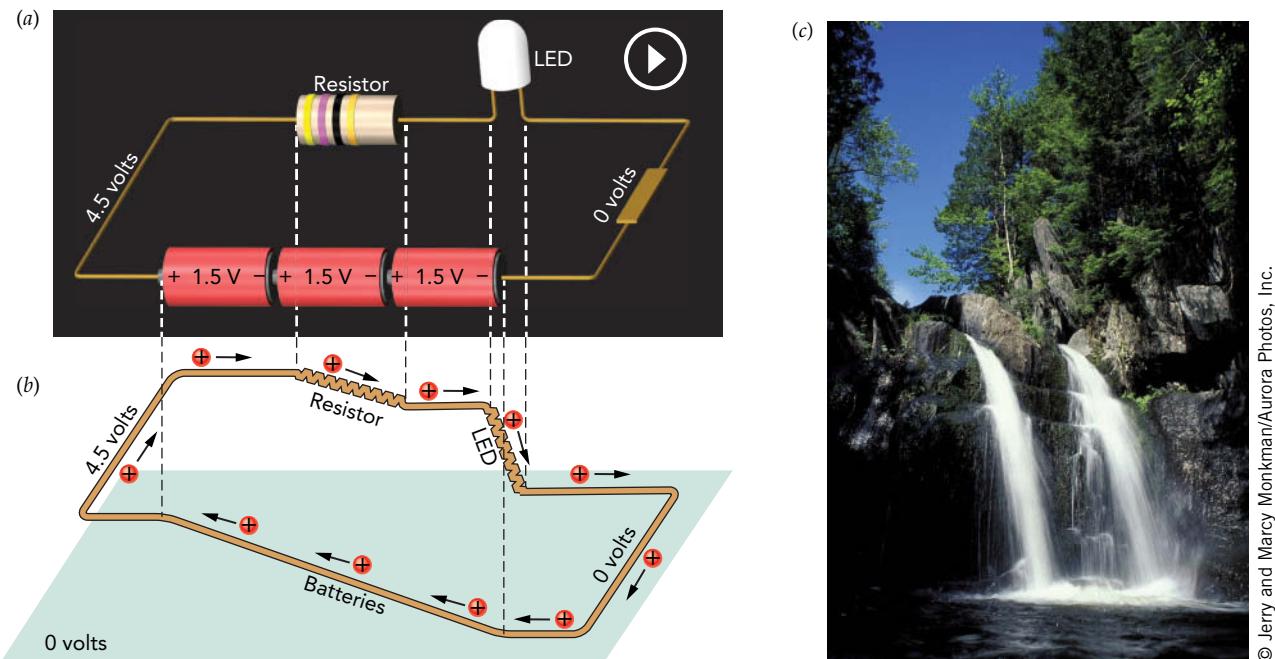


Fig. 10.3.13 (a) Current in a two-LED flashlight's parallel circuit divides into two portions that flow in parallel through the two LEDs, each paired with its own resistor. (b) Because the two LED-resistor pairs are in parallel, they experience the same voltage drop but share the overall current. (c) These waterfalls are in parallel. The waterfalls experience the same altitude drop, but they share the overall current of water.

net charges through contact and how high voltages can be produced by separating large quantities of opposite charge.

In Xerographic Copiers, we studied electric fields and saw how those fields can be used to move charges from one place to another. We examined the corona discharges that form in the strong electric fields around sharp points or thin wires at high voltages. We also saw how placing a charge on one object can induce an opposite charge on a grounded object nearby.

In Flashlights, we examined circuits to see how a current of electric charges can transfer power from batteries to a lightbulb. We found that both voltage and current contribute to this power transfer, and we learned how the bulb's electrical resistance governs the power it consumes.

Explanation: Moving Water without Touching It

This experiment uses contact between dissimilar materials to give the comb a net electric charge. Electrons shift from your hair to the comb, leaving your hair positively charged and the comb negatively charged. Because the comb is an electrical insulator, its negative charge remains trapped on its surface for a long time.

As you hold the negatively charged comb near the water stream, the comb attracts positive charges present in the water and repels negative charges. Because the water conducts electricity somewhat, positive charges can travel down the water stream from the faucet while negative charges travel the other way. The water thus acquires a net positive charge in the presence of the negatively charged comb, an example of charging by induction. Since the oppositely charged water and comb attract one another, the water accelerates toward the comb and arcs sideways as it falls.

Chapter Summary and Important Laws and Equations

How Static Electricity Works: When clothes tumble about in the dryer, contact between dissimilar materials transfers negatively charged electrons from one item to the other. As a result of this contact-charging effect, the various garments acquire net charges, some positive and others negative. When the clothes are subsequently separated, the work done in pulling them apart becomes electrostatic potential energy and the clothes develop high voltages. High voltages can also develop when you walk across a carpet or drive your car along the road. Since high voltage tends to push charges into the air as leaks and sparks, static charging can be a nuisance. You can control it with the help of conducting materials; allowing charge to move spontaneously from high voltage to low voltage prevents large quantities of separated charge from accumulating so that high voltages can't develop.

How Xerographic Copiers Work: The photoconductor in a xerographic copier allows light to control the distribution of electric charge on its surface. This photoconductor is uniformly precoated with charge by passing it near a corona discharge. An optical image of the original document is then projected onto this charged photoconductor. Wherever light hits the photoconductor, the surface charge escapes. The result is a charge image on the photoconductor. Tiny black toner particles, oppositely charged from the unilluminated portions of the photoconductor, are brought near the charge image. These toner particles stick to the charged portions of the photoconductor, forming a visible image of the document. The toner particles are then transferred to and fused onto a sheet of paper to create a finished copy.

How Flashlights Work: A flashlight produces light when a current flows through its electric circuit. This circuit consists of a chain of batteries, a lightbulb, a switch, and several metal strips, all connected in a continuous loop. The current obtains power as it passes through the battery, and it releases this power as it passes through the filament of the lightbulb, heating that filament white hot.

The switch provides a way to control the flow of current in the circuit. When the switch is off, it breaks the circuit and prevents current from flowing completely around the circuit. Without a steady current to carry power from the batteries to the bulb, the bulb is dark. When the switch is on, it completes the circuit so that power can reach the bulb and the bulb lights up.

1. Coulomb's law: The magnitude of the electrostatic force between two objects is equal to the Coulomb constant times the product of their two electric charges divided by the square of the distance separating them, or

$$\text{force} = \frac{\text{Coulomb constant} \cdot \text{charge}_1 \cdot \text{charge}_2}{(\text{distance between charges})^2}. \quad (10.1.1)$$

If the charges are like, then the forces are repulsive. If the charges are opposite, then the forces are attractive.

2. Force exerted on a charge by an electric field: A charge experiences a force equal to its charge times the electric field, or

$$\text{force} = \text{charge} \cdot \text{electric field}, \quad (10.2.1)$$

where the force points in the direction of the electric field.

3. Electric field due to a voltage drop: A voltage drop produces an electric field equal to that voltage drop divided by the distance over which the drop occurs, or

$$\text{electric field} = \text{voltage gradient} = \frac{\text{voltage drop}}{\text{distance}}, \quad (10.2.2)$$

where the field points in the direction of the most rapid voltage decrease.

4. Power consumption: The electric power consumed by a device is the voltage drop across it times the current flowing through it, or

$$\text{power consumed} = \text{voltage drop} \cdot \text{current}. \quad (10.3.1)$$

5. Power production: The electric power provided by a device is the voltage rise across it times the current flowing through it, or

$$\text{power provided} = \text{voltage rise} \cdot \text{current}. \quad (10.3.2)$$

6. Ohm's law: The voltage drop through an ohmic object is equal to the current flowing through it times its electrical resistance, or

$$\text{voltage drop} = \text{current} \cdot \text{electrical resistance}. \quad (10.3.4)$$

This equation does not apply to nonohmic devices.

Magnetism and Electrodynamics

Like electricity, magnetism is an important part of daily life. We use it to post notes on the refrigerator and to figure out which way is north. The story of magnetism wouldn't be complete, though, without including electricity. As we'll see, these two topics are related to one another through change and motion. For example, moving electric charges give rise to magnetism, and changing magnetism gives rise to electricity.

In this chapter, we'll examine magnetism itself, as well as several objects that use the relationships between electricity and magnetism to perform useful tasks. Since the word *dynamics* covers change and motion, these relationships are part of a field of physics known as *electrodynamics*. Brevity isn't the only reason for omitting reference to magnetism in the title of this field; the other reason is that most magnetism is actually produced by electricity. In other words, most magnetism is actually electromagnetism.

ACTIVE LEARNING EXPERIMENTS

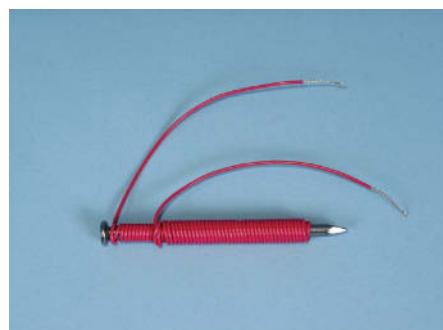
A Nail and Wire Electromagnet

To explore the relationships between electricity and magnetism, try building a simple electromagnet. For this project, you need a large steel nail or bolt, a meter or so of insulated wire, a fresh 1.5-V AA battery, and some small steel objects such as paper clips. The wire's metal conductor should be at least 0.65 mm in diameter (22 gauge or larger) to carry the current you'll send through it without becoming too hot.

Wind the wire tightly around the nail or bolt to form a coil. You should complete at least 50 turns of wire, all in the same direction. The exact number of turns isn't important, and you can make several layers. Be sure that the two ends of the wire are still accessible and remove the insulation from each end so that you can connect them to the battery.

WARNING

The electromagnet that you construct in this experiment will become hot during use. Be prepared to drop the electromagnet if it becomes uncomfortably hot. Don't work near flammable materials.



Courtesy Lou Bloomfield

Now test your electromagnet. Connect one uninsulated end of the coil to each terminal of the battery. You can either hold the wires on the terminals with your fingers or use tape. A 1.5-V battery can't give you a shock unless your skin is broken, but you should be prepared for the wire to get hot as current flows through it. If it gets too hot to hold, let go and make sure that the wire detaches from the battery so that it doesn't start a fire. Don't use a battery larger than AA or the wire may get dangerously hot.

While current is flowing through the wire, the nail will act as a strong magnet, an electromagnet. Try picking up steel objects with this electromagnet. As you touch each object with the nail, it should stick to the nail's surface. Your electromagnet will temporarily magnetize the steel object and attract it. What happens when you touch this magnetized steel object to a second object? What happens when you stop the flow of electric current through the coil of the electromagnet? Why does the coil get hot while current is flowing through it?

Chapter Itinerary

This chapter will examine (1) *household magnets* and (2) *electric power distribution*. In Household Magnets, we look at the forces that bind magnets to refrigerators and see why compasses point northward. We also examine electromagnets to explain how electric doorbells work. In Electric Power Distribution, we see how electricity and magnetism are used to transport electric power from a distant power

plant to your home and how that electric power differs from the power supplied by batteries. Although there are many other magnetic and electromagnetic objects that we encounter daily, these two topics are representative of most of the basic electromagnetic phenomena. For a more complete preview, turn ahead to the Chapter Summary and Important Laws and Equations at the end of the chapter.

SECTION 11.1

Household Magnets



How would a family stay organized without refrigerator magnets? How would the doorbell ring if it couldn't use a magnet to strike its bells? How would a scout navigate in the woods without a compass? How would you get cash or charge purchases without magnetic strips on plastic cards?

We're so used to having magnets around that we take them for granted. Along with being useful, household magnets also let us experiment with another of the basic forces in nature. Although we'll see that magnetism is so intimately related to electricity that the two are ultimately a single unified whole, we'll find it helpful to begin our study of magnetism by treating it as a separate phenomenon and bring electricity into the picture gradually.

Questions to Think About: Why do two magnets attract or repel each other, depending on how they're oriented? If a magnet is attracted to two different refrigerators, why don't those two refrigerators attract or repel one another? How can two strong magnets grip one another from opposite sides of your hand? Why isn't your hand involved in that magnetic attraction? How can some magnets be turned on or off using electricity?

Experiments to Do: Find two button-type refrigerator magnets—simple cylinders that resemble small hockey pucks. If you try to stack these magnets, you'll find that they either attract or repel one another. How do those forces depend on the orientations of the magnets? on their separation? See whether you can float one magnet on top of the other using the repulsive force. What happens when you let go of the top magnet?

Now hold one of the button magnets near a refrigerator or another steel object and study the forces that arise. Can you find a way to make the two objects repel one another? Will a magnet stick to things that aren't made of steel? What about stainless steel?

Now find two identical sheet-type refrigerator magnets—flexible strips that may have advertising printed on their surfaces. Experiment with their interactions. You'll find that simply flipping one over doesn't change the forces from attraction to repulsion. Instead, you'll have to slide the strips across one another. As they slide, they'll alternately attract and repel. How is that possible?

Finally, obtain iron powder or make some by filing down a piece of steel (steel is mostly iron, and they are magnetically

quite similar). Sprinkle some of this powder on your collection of magnets. Notice that it forms strands that seem to bridge the gaps between various points on the magnets. What is the powder bridging? If you sprinkle your powder on a credit card or magnetic ID card, you'll find that it also forms bridges. However, those bridges are tiny and spaced at irregular intervals. Could there be information encoded in these erratically spaced magnetic features?

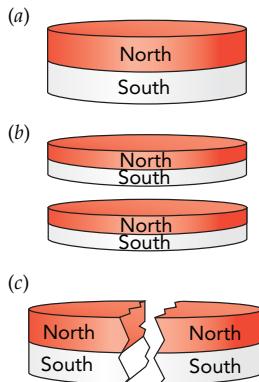


Fig. 11.1.1 (a) A typical button magnet has a north pole on one face and a south pole on the other. Its net pole is zero. (b) Slicing it between its poles or (c) breaking it through its poles always yields a pair of magnets, each with zero net pole.

1 Since magnetic monopoles apparently don't exist, magnets and magnetic materials must obtain their magnetic poles through other means. As we'll see later in this section, electric currents are magnetic and electric current loops act as magnetic dipoles. Circulating currents exist deep within all magnetic materials and are responsible for their magnetism. Some of those currents are associated with charged electrons orbiting atomic nuclei, but most are associated with the rotating nature of electrons, a fundamental characteristic known as *spin*. Every electron is a tiny magnetic dipole.

Button-Shaped Refrigerator Magnets

Refrigerator magnets come in all shapes and sizes, and some are more magnetically complicated than others. It's always best to start simple, so we'll begin with button-shaped magnets.

As you bring two button magnets together, they begin exerting forces on one another. You'll find that those forces can be either attractive or repulsive, depending on how the magnets are oriented, but they always grow weaker with greater distance. Such magnetic forces resemble the electric ones you encounter while removing clothes from a hot dryer, but there are at least two important differences. First, reorienting two electrically charged garments won't turn their attraction into repulsion or vice versa. Second, no matter how you arrange two button magnets, you can't get a magnetic spark to jump from one to the other. Electricity and magnetism are evidently similar yet different. What's the story?

Magnetism is a phenomenon that closely resembles electricity. Just as there are two types of electric charges that exert electrostatic forces on one another, so there are two types of **magnetic poles** that exert **magnetostatic forces** on one another. The word *pole* serves to distinguish magnetism from electricity; **poles** are magnetic, while charges are electric.

The two types of poles are called north and south, and in keeping with this geographical naming, they're exact opposites of one another. Both types of poles carry just one physical quantity: **magnetic pole**. North poles carry positive amounts of magnetic pole, while south poles carry negative amounts. It should come as no surprise that like poles repel each other, while opposite poles attract. Furthermore, the magnetostatic forces between two poles grow weaker as they move apart and are inversely proportional to the square of the distance between them. So far, the similarities between electricity and magnetism are striking.

However, we now come to a crucial difference between electricity and magnetism: while subatomic particles that carry pure positive or negative electric charges are common, particles that carry pure north or south magnetic poles have never been found. Called **magnetic monopoles**, such pure magnetic particles may not even exist in our universe. That cosmic omission explains why there are no magnetic sparks: without monopoles, there is no magnetic equivalent of an electric charge that can leap from one place to another as a magnetic current, let alone a magnetic spark.

Although isolated magnetic poles aren't available in nature, *pairs* of magnetic poles are. These pairs consist of equal north and south poles, spatially separated from one another in an arrangement called a **magnetic dipole**. Since the two opposite poles have equal magnitudes, they sum to zero and the magnetic dipole has zero **net magnetic pole**.

A simple button magnet has both a north pole and a south pole, usually on opposite faces of the button (Fig. 11.1.1a). There are no purely north buttons or purely south buttons. Amazingly enough, even slicing that button magnet in half won't yield separated north and south poles (Fig. 11.1.1b). Instead, new poles will appear at the cut edges and each piece of the original magnet will end up with zero net pole! Breaking the button magnet in half (Fig. 11.1.1c) will also produce pieces with zero net pole. For a discussion of why button magnets always have zero net magnetic pole, see 1.

We can now explain why two of these magnets sometimes attract and sometimes repel. With two poles on each magnet, we have to consider four interactions: two repulsive interactions between like poles (north-north and south-south) and two attractive interactions

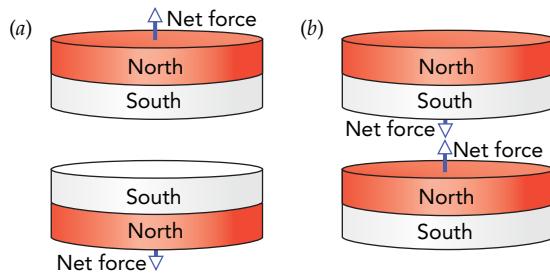


Fig. 11.1.2 (a) When like poles of two button magnets are turned toward one another, the magnets repel. (b) When opposite poles are turned toward one another, they attract.

between opposite poles (north-south and south-north). Although it might seem that all these forces should cancel, the distances separating the various poles and therefore the forces between them depend on the magnets' orientations. Since the closest poles experience the strongest forces, they dominate. If you turn two like poles toward one another, the two magnets will push apart (Fig. 11.1.2a). If you turn their opposite poles toward one another, they'll pull together (Fig. 11.1.2b). If you tip them at an angle, they'll experience torques that tend to twist opposite poles together and like poles apart.

Because there are no monopoles, we're going to need some imagination to understand magnetism well. Let's start with units. Even though we can't collect a unit of pure north pole, we can still define such a unit and understand its behavior. The SI unit of magnetic pole is the **ampere-meter** (abbreviated A·m). That astonishing choice, an *electric* unit appearing in a *magnetic* unit, foreshadows the profound connections between electricity and magnetism that we'll soon encounter.

Just as there is a Coulomb's law for electric charges, there is a **Coulomb's law for magnetic poles**. Coulomb's magnetic experiments, which were complicated by the fact that he had to work with magnetic dipoles rather than individual poles, showed that the forces between magnetic poles are proportional to the amount of each pole and inversely proportional to the square of their separation. The exact relationship can be written as a word equation:

$$\text{force} = \frac{\text{permeability of free space} \cdot \text{pole}_1 \cdot \text{pole}_2}{4\pi \cdot (\text{distance between poles})^2}, \quad (11.1.1)$$

in symbols:

$$F = \frac{\mu_0 \cdot p_1 \cdot p_2}{4\pi r^2},$$

and in everyday language:

Don't hold two strong magnetic poles near one another unless you're prepared to be pushed around hard as they attract or repel each other.

The force on pole₁ is directed toward or away from pole₂, and the force on pole₂ is directed toward or away from pole₁. The **permeability of free space** is $4\pi \times 10^{-7}$ N/A². Consistent with Newton's third law, the force that pole₁ exerts on pole₂ is equal in amount but oppositely directed from the force that pole₂ exerts on pole₁.

COULOMB'S LAW FOR MAGNETISM

The magnitudes of the magnetostatic forces between two magnetic poles are equal to the permeability of free space times the product of the two poles divided by 4π times the square of the distance separating them. If the poles are like, then the forces are repulsive. If the poles are opposite, then the forces are attractive.

Check Your Understanding #1: Two Halves Make a Whole

You have a disk-shaped permanent magnet. The top surface is its north pole and the bottom surface is its south pole. If you crack the magnet into two half circles, the two halves will push apart. Why?

Answer: The top surfaces of both halves are still north poles, and the bottom surfaces are still south poles. The two tops repel, as do the two bottoms.

Why: This puzzling phenomenon, in which a shattered permanent magnet opposes attempts to reassemble it, is an illustration of the potential energy contained in a permanent magnet. The magnet is a collection of many tinier magnets, all aligned with their north poles together and their south poles together. Like poles repel one another, so the tiny magnets are difficult to hold together. Given a chance, the magnet will push apart into fragments. Very strong permanent magnets release so much potential energy when they break that they practically explode when cracked.

The Refrigerator: Iron and Steel

Two button magnets can push or pull on one another, but what if you have only one? The easiest way to observe magnetic forces in that case is to hold that single magnet near your refrigerator or another piece of iron or steel. The magnet is attracted to the refrigerator. However, if you flip the button magnet over, thinking that it will now be repelled by the refrigerator, you'll be disappointed. Although the refrigerator is clearly magnetic, its magnetism somehow responds to the button magnet so that the two always attract.

Actually, the refrigerator's behavior isn't all that mysterious. Its steel is composed of countless microscopic magnets, each with a matched north pole and south pole (Fig. 11.1.3). Normally those individual magnetic dipoles are oriented semirandomly (Fig. 11.1.3a), so the refrigerator exhibits no overall magnetism. However, as you bring one pole of a button magnet near the refrigerator, its tiny magnets evolve in size, shape, and orientation (Fig. 11.1.3b). Overall, opposite poles shift closer to the button magnet's pole and like poles shift farther from the button magnet's pole. The steel develops a **magnetic polarization** and consequently attracts the pole of the button magnet.

This polarization remains strong only as long as the button magnet's pole is nearby. When you remove the button magnet, most of the tiny magnets in the steel resume their semirandom orientations and the steel's magnetic polarization shrinks or disappears. When you then bring the button magnet's other pole close to the refrigerator, its steel develops the opposite magnetic polarization and again attracts the button magnet's pole. No matter which pole or assortment of poles you bring near the refrigerator, its steel will polarize in just the right way to attract those poles.

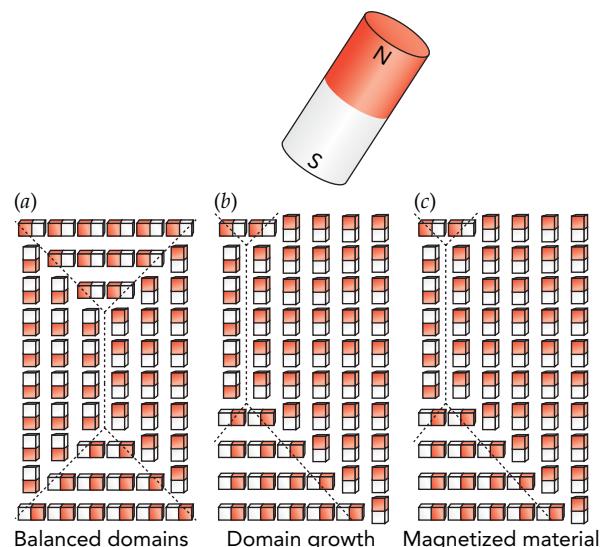


Fig. 11.1.3 (a) The countless microscopic magnets in iron or steel are normally oriented somewhat randomly. (b) But when a strong magnetic pole is nearby, those tiny magnets reorient to attract it. In soft magnetic materials, this reorientation is only temporary. (c) Hard magnetic materials, however, remain magnetized long after the external pole has departed.

If you try this trick with a plastic or aluminum surface, the button magnet won't stick. What's special about steel that allows it to develop such a strong magnetic polarization? The answer is that ordinary steel, like its constituent iron, is a **ferromagnetic** material—it is actively and unavoidably magnetic on the size scale of atoms.

To understand ferromagnetism, we must start by looking at atoms and the subatomic particles from which they're constructed: electrons, protons, and neutrons. For complicated reasons, all those subatomic particles have magnetic dipoles, particularly the electrons, and the atoms they form often display this magnetism. Despite a tendency for the subatomic particles to pair up with opposite orientations so that their magnetic dipoles cancel one another, most isolated atoms have significant magnetic dipoles.

Although most atoms are intrinsically magnetic, most materials are not. That's because another round of pairing and canceling occurs when atoms assemble into materials. This second round of cancellation is usually so effective that it completely eliminates magnetism at the atomic scale. Materials such as glass, plastic, skin, copper, and aluminum retain no atomic-scale magnetism at all, and your button magnet won't stick to them. Even most stainless steels are nonmagnetic.

However, there are a few materials that avoid this total cancellation and thus manage to remain magnetic at the atomic scale. The most important of these are the ferromagnets, a class of magnetic materials that includes ordinary steel and iron. If you examine a small region of ferromagnetic steel, you'll find that it is composed of many microscopic regions, or **magnetic domains**, that are naturally magnetic and cannot be demagnetized (Fig. 11.1.3a). Within a single domain, all the atomic-scale magnetic dipoles are aligned and together they give the overall domain a substantial net magnetic dipole.

While common steel always has these magnetic domains, magnetic interactions orient nearby domains so that their magnetic dipoles oppose one another and cancel. The microscopic magnets balance one another so well that the steel appears nonmagnetic. That's too bad; the appliance showroom would be a much more exciting place to visit if the cancellation weren't so good.

However, when you bring a strong magnetic pole near steel (Fig. 11.1.3b), the individual domains grow or shrink, depending on which way they're oriented magnetically. The steel undergoes magnetization and becomes **magnetized** (Fig. 11.1.3c). The atoms themselves don't move during this process; the change is purely a reorientation of the atomic-scale magnetic dipoles. Domains that attract your button magnet's pole grow while those that repel it shrink, and the button magnet sticks to the refrigerator.

Courtesy Lou Bloomfield



Fig. 11.1.4 (a) Although these paper clips were initially unmagnetized, the pole of a strong permanent magnet magnetizes them as a chain. (b) After the magnetizing magnet is removed, the clips retain some of their magnetization.

Courtesy Lou Bloomfield



Fig. 11.1.5 Iron powder forms bridges between the magnetic poles of this plastic sheet magnet.

Check Your Understanding #2: Chain Links

If you touch the north pole of a permanent magnet to one end of a steel paper clip, the clip's other end will become magnetic. What pole will that other end have?

Answer: It will have a north pole.

Why: The paper clip will become magnetically polarized with its south pole nearest the permanent magnet's north pole. The other end of the paper clip will have a north pole and will be able to polarize other paper clips. These polarized clips attract one another strongly enough to cling together in a long chain.

Plastic Sheet Magnets and Credit Cards

When you remove your button magnet from the refrigerator, the steel returns to its original nonmagnetic state—it undergoes demagnetization and becomes **demagnetized**. Well, *almost* demagnetized. The demagnetization process isn't perfect because some of the domains get stuck. Although magnetic forces within the steel favor a complete return to apparent nonmagnetism, chemical forces can make it hard for the domains to grow or shrink. Adjacent domains are separated by **domain walls**, boundary surfaces between one direction of magnetic orientation and another. These domain walls must move if the domains are to change size. However, flaws and impurities in the steel can interact with a domain wall and keep it from moving. When that happens, the steel fails to demagnetize itself completely (Fig. 11.1.4). To remove the last bit of residual magnetism from steel, you must help the domain walls move, typically with heat or mechanical shock.

A **soft magnetic material** is one that demagnetizes itself easily when all nearby poles are removed. Chemically pure iron, which has few flaws or impurities, is a soft magnetic material—easy to magnetize and easy to demagnetize. A **hard magnetic material** is one that does not demagnetize itself easily and that tends to retain whatever domain structure is imposed on it by its most recent exposure to strong nearby poles (Fig. 11.1.3c). Your button magnet is made from a hard magnetic material!

Like steel, the material in your button magnet is ferromagnetic (or closely related to ferromagnetic). Unlike steel, however, your button magnet's domains do not shrink or grow easily. During its manufacture, the button magnet was magnetized by exposing it to such strong magnetic influences that its domains rearranged to give it permanent magnetic poles. It now has a north pole on one face and a south pole on the other. Unless you expose the button to extremely strong magnetic influences or heat it or pound it, it will retain its present magnetization almost indefinitely. In that respect, the button is a **permanent magnet**.

Not all permanent magnets are as simple as button magnets. Depending on how they were magnetized, they can have their north and south poles located in unexpected places and even have more than one pair of poles. Plastic sheet magnets are a good example of multiple-pole magnets: each has a repeating pattern of alternate poles along its length. The exact patterns vary, but most have poles that form alternating parallel stripes. You can find these stripes by letting them polarize and attract iron powder (Fig. 11.1.5) or by sliding two identical sheet magnets across one another. The sheets will attract and bind together most strongly when opposite poles are aligned across from each other. They'll repel when you shift one of the magnets so that like poles are aligned.

A hard magnetic material's ability to "remember" its magnetization can be useful for saving information. Once magnetized in a particular manner so as to represent a piece of information, the material will retain its magnetization and the associated information until something magnetizes it differently. Information retention in hard magnetic materials forms the basis for most magnetic recording and storage, including the magnetic strips on credit cards, magnetic tapes, computer disks, and magnetic random access memory (MRAM) (Fig. 11.1.6).

Courtesy Lou Bloomfield



Fig. 11.1.6 Iron powder discloses the locations of magnetic poles on this credit card's magnetic strip. Information is stored by choosing where those poles are located.

Check Your Understanding #3: Now for the Flip Side

If you bring the north pole of a large strong magnet near the north pole of a small weak magnet that you are holding in place, what will happen to that small magnet?

Answer: Its magnetic poles will interchange.

Why: Even though the small magnet can't move, its magnetic poles can. When the repulsion between the two north poles becomes strong enough, the poles of the small magnet will interchange and it will then present its south pole to the north pole of the large magnet. You will have permanently reversed the small magnet's poles.

Compasses

If you've spent time hiking, you may well own a magnetic compass. Like a button magnet, the needle of that compass is a simple permanent magnet with one north magnetic pole and one south magnetic pole. This needle aids navigation because Earth itself has a magnetic dipole and that dipole affects the orientation of the needle—the needle's north magnetic pole tends to point northward.

Already, we can guess what must be located near Earth's *north* geographic pole—a *south* magnetic pole. Attraction from that south magnetic pole is what draws the compass's north magnetic pole toward the north. However, the full story is more complicated. To begin with, Earth's magnetic poles are actually located far beneath its surface and aren't perfectly aligned with the geographic poles. To make matters worse, magnetically active materials in everything from distant mountains to nearby buildings assert their own magnetic influences on the compass needle. Overall, the compass needle is responding to the influences of countless magnetic poles, both near and far. Given how difficult it would be to sum up all those separate influences, it is easier to view the compass needle as interacting with something local—a **magnetic field**, an attribute of space that exerts a magnetostatic force on a pole. According to this new perspective, the compass needle responds to the local magnetic field, a field that's created by all the surrounding magnetic poles.

As with an electric field, the magnetic field here appears to be acting merely as an intermediary; various poles produce the magnetic field, and this magnetic field affects our compass needle. As we'll see, however, a magnetic field is more than just an intermediary or a fiction. It is quite real and can exist in space, independent of the poles that produce it. Just as electric fields can be created by things other than charge, so magnetic fields can be created by things other than pole.

The magnetic field at a given location measures the magnetostatic force that a unit of pure north pole would experience if it were placed at that point. More specifically, the magnetostatic force is equal to the pole times the magnetic field at the pole's position. We can write this relationship as a word equation:

$$\text{magnetostatic force} = \text{pole} \cdot \text{magnetic field}, \quad (11.1.2)$$

in symbols:

$$\mathbf{F} = p\mathbf{B},$$

and in everyday language:

If you place a strong magnet in a big magnetic field, expect to be pushed around,

where the magnetostatic force is in the direction of the magnetic field. Note that a negative amount of pole (a south pole) experiences a force opposite the magnetic field. The SI unit of magnetic field is the **newton per ampere-meter**, also called the **tesla** (abbreviated T).

2 Earth's magnetic poles are not particularly well aligned with its geographical poles and have actually moved about 1100 km over the past century. Earth's south magnetic pole is presently drifting northwest across the Canadian arctic at about 40 kilometers per year. Moreover, Earth's magnetic poles have been trading places about once every 700,000 years for the past 330 million years. The most recent reversal was about 780,000 years ago.

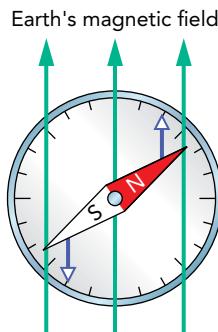


Fig. 11.1.7 A compass needle aligns with the local magnetic field. Its north pole experiences a magnetostatic force in the direction of the field, and its south pole experiences a force opposite the field.

Earth's magnetic field is relatively weak, about 0.00005 T in a roughly northward direction (see **2**). (For comparison, the field near your button magnet may be 0.1 T or more.) Earth's field pushes the compass needle's north pole northward and south pole southward (Fig. 11.1.7). Unless the compass needle is perfectly aligned with that field, it experiences a torque and undergoes angular acceleration. Since its mount allows the needle to rotate only horizontally and it experiences mild friction as it does, the needle soon settles down with its north pole pointing roughly northward. If its mount allowed it to rotate vertically as well as horizontally, the needle's north pole would dip downward in the northern hemisphere and upward in the southern hemisphere. In general, the needle minimizes its magnetostatic potential energy by pointing along the direction of the local magnetic field and is thus in a stable equilibrium when orientated that way. After a few swings back and forth, your compass needle points along the local magnetic field, which, we hope, points northward.

Because Earth's magnetic field is so uniform in the vicinity of your compass, its northward push on the needle's north pole exactly balances its southward push on the needle's south pole and the needle experiences zero net force. However, if you bring your compass near a button magnet, the local magnetic field will not be uniform and the needle may experience a net force. The magnetic field gets stronger near one of the button's poles and the compass needle will experience a net force toward or away from that pole, depending on which way it's orientated.

When the needle is aligned with a nonuniform field—its north pole pointing in the same direction as the local field—the forces on its two opposite poles won't balance and it will experience a net force in the direction of increasing field. If it is aligned against the field, it will experience a net force in the direction of decreasing field. In practice, as you bring the compass near your button magnet, its needle will first pivot into alignment with the local field and then find itself pulled toward increasing field, toward the nearest pole of the button magnet. The same thing happens when you bring two button magnets together; each pivots into alignment with the other's magnetic field, and the two then leap at each other. Watch out for your fingers!

A piece of steel exhibits similar behavior when you hold it near a button magnet: it becomes magnetized along the direction of the local magnetic field and then finds itself pulled toward increasing field, toward the button magnet's nearest pole. That's how the button magnet holds your notes to the refrigerator!

Check Your Understanding #4: Crazy Compass

You lock the needle of your compass and move its north pole near the north pole of a powerful button magnet. Will the needle experience a magnetostatic force toward strong field or weak field?

Answer: The needle will experience a force toward weak magnetic field (away from the button magnet).

Why: With its magnetic poles aligned opposite to the button's magnetic field, the compass needle experiences a force toward weaker magnetic field. Actually, if you continue to push the needle closer to the button magnet, you can accidentally remagnetize that needle; its poles will permanently interchange and it will subsequently point south rather than north! Not to worry, though, because you can simply repeat this procedure to restore the compass to normal.

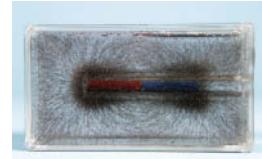
Check Your Figures #1: Careful with That Wrench!

You mistakenly place a long steel wrench in the 1-T field near a strong magnet. The field magnetizes the wrench, and it develops a north pole of 1000 A·m at its near end and an equal south pole at its distant end. Only the near end of the wrench is in the 1-T field and experiences a magnetic force. What force does the field exert on the wrench and its north pole?

Answer: It exerts almost 1000 N (225 lb) in the direction of the field.

Why: According to Eq. 11.1.2, the force exerted on the wrench's north pole is equal to its 1000-A·m pole times the 1-T magnetic field. Since 1 T = 1 N/A·m, that product is 1000 N and points in the direction of the field. Such large forces are not unusual when steel or iron objects are exposed to a strong magnetic field, so be careful working near big magnets!

Courtesy Lou Bloomfield



Iron Filings and Magnetic Flux Lines

Magnetic fields seem abstract; it would be helpful if you could see them. Remarkably enough, you can—just sprinkle iron filings into the field! Although you'll need to support their weight with paper or a liquid, an interesting pattern will form. Like tiny compass needles, the iron particles magnetize along the local magnetic field and then stick together, north pole to south pole, in long strands that delineate the magnetic field (Fig. 11.1.8)!

These strands map the magnetic field in an interesting way. First, at each point on a strand, the strand points along the local magnetic field. Second, the strands are most tightly packed where the local magnetic field is strongest. In other words, the strands follow along the local magnetic field direction and have a density proportional to that local field. The lines highlighted by these strands are so useful that they have their own name, **magnetic flux lines**.

Flux lines are often helpful when exploring a magnetic field. If you're studying the magnetic field in a large area, you probably don't want to use iron filings. Instead, you can hold a compass in your hand and walk in the direction its needle is pointing—the direction of the magnetic field. The path you'll follow in this compass-guided walk is a magnetic flux line. If you repeat this trip from many different starting points, you'll explore the whole magnetic field, flux line by flux line. Since a magnetic field tends to point away from north poles and toward south poles, these tours will typically take you from north poles to south poles. In fact, for our permanent magnets, every magnetic flux line begins at a north pole and ends at a south pole.

That last observation about flux lines is quite general: they never start at or end on anything other than a magnetic pole. While flux lines emerge in all directions from a north pole and converge from all directions on a south pole, that's it; flux lines never begin or end in empty space. If you're following a magnetic flux line with your compass, you will either reach a south pole or walk forever!

The possibility of that endless walk is somewhat disconcerting; if the flux line you're following doesn't end at a pole, what created its magnetic field? The answer reveals a deep connection between magnetism and electricity. Some magnetic fields aren't produced by magnetic poles at all; they're produced by electricity! To see how that's possible, let's take a look at another common household magnet—the electromagnet in an ordinary doorbell.

Check Your Understanding #5: Building Bridges

If you sprinkle iron filings onto the magnetic strip of a credit card, a pattern of tiny iron bridges will form. Where are the magnetic poles relative to those bridges?

Answer: The poles are at the ends of the bridges.

Why: The iron filings follow the magnetic flux lines, which extend from north poles to south poles. Thus, one end of each bridge is a north pole and the other end is a south pole.

Electric Doorbells and Electromagnets

A classic electric doorbell uses a magnet and a spring to drive a piece of iron into two chimes, “ding-dong.” When you press the doorbell button, you close an electric circuit and the resulting electric current pushes the iron *magnetically* into the first chime, “ding.” When you release the button, you open the circuit, stopping the current and its magnetism so that the spring can push the iron back into the second chime, “dong.”

The big news here is that electric currents can produce magnetic forces. In fact, there is nothing optional about this connection—electric currents *are* magnetic. More specifically, moving electric charge produces a magnetic field.

3 Self-educated before the French Revolution, during which his father was executed, French physicist André-Marie Ampère (1775–1836) became a science teacher in 1796. He served as a professor of physics or mathematics in several cities before settling at the University of Paris system in 1804. In 1820, only a week after learning of Oersted's experiment showing that an electric current causes a compass needle to deflect, Ampère published an extensive treatment of the subject. Evidently, he had been thinking about these ideas for a long time.

Courtesy Lou Bloomfield



Fig. 11.1.9 Iron powder shows that flux lines around a current-carrying wire form concentric rings around that wire.

Courtesy Lou Bloomfield

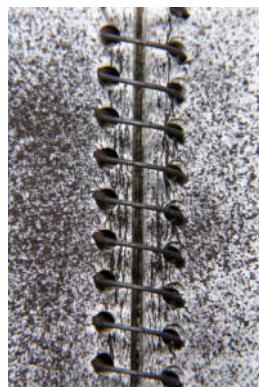


Fig. 11.1.10 Iron powder shows that flux lines pass straight through a current-carrying coil and return outside it, much like the flux lines around a similarly shaped bar magnet.

FIRST CONNECTION BETWEEN ELECTRICITY AND MAGNETISM

Moving electric charge produces a magnetic field.

Imagine the surprise of Danish physicist Hans Christian Oersted (1777–1851) when he observed in 1820 that current in a wire caused a nearby compass needle to rotate. Until that moment, electricity and magnetism had appeared to be independent phenomena. Inspired by Oersted's experiment, French physicist André-Marie Ampère **3** undertook a 7-year study of the relationships between electricity and magnetism, and started the revolution that eventually unified them within a single overarching conceptual framework.

When we use iron powder to disclose the magnetic flux lines surrounding a long, straight, current-carrying wire, we too are in for a surprise (Fig. 11.1.9). Those flux lines circle the wire like concentric rings, growing more widely separated as the distance from the wire increases. The wire is an **electromagnet**, a device that becomes magnetic when it carries an electric current. However, because an electromagnet has no true magnetic poles, the magnetic flux lines can't stretch from north pole to south pole. Instead, each flux line of an electromagnet is a closed loop. If you took a compass-guided walk along one of these flux lines, you'd retrace your steps over and over.

Since the flux lines are packed tightest near the surface of the current-carrying wire, that's where the magnetic field is strongest. Recalling that a piece of iron is pulled toward increasing magnetic field, we see that the wire will attract iron to it whenever it's carrying a current.

The magnetic field around a current-carrying wire is fairly weak, however, and a practical doorbell winds that wire into a coil to concentrate and strengthen its field. Although the magnetic field around a current-carrying coil is complicated, we can use iron powder to make it visible (Fig. 11.1.10). Remarkably enough, the flux lines outside the coil resemble those outside a button magnet of similar dimensions (Fig. 11.1.11). It's as though the coil has a north pole at one end and a south pole at the other! Because there are no true poles present, however, the flux lines don't end anywhere. Instead, they continue right through the middle of the coil and form complete loops.

When current flows through the coil, nearby iron finds itself magnetized along the local magnetic field and then pulled toward increasing field—toward the tightly packed flux lines at the coil's end. But why stop there? Since the flux lines continue right into the coil and grow even more tightly packed inside, the iron will be pulled inward toward the very center of the coil!

That's how the doorbell works. When you press the doorbell button, current flows through a coil of wire and the resulting magnetic field yanks an iron rod into the center of

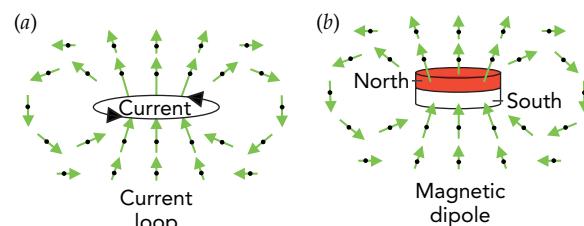


Fig. 11.1.11 (a) The magnetic field around a loop of current-carrying wire points up through the loop and down around the outside of the loop. The magnetic field arrow passing through each black dot indicates the magnitude and direction of the force a north test pole would experience at the dot's location. (b) The field produced by a two-pole button magnet is almost identical to that of the loop.

that coil. About the time the rod reaches the center, part of it hits the first chime. When you then open the switch, stopping the current and its magnetism, a spring pushes the iron rod back out of the coil and it hits the second chime. These two chimes make the familiar ding-dong!

While current is flowing through the coil and the iron rod is inside it, the two objects act as a single powerful electromagnet. The magnetic field surrounding the pair is the sum of the coil's modest magnetic field and the magnetized iron's much stronger field. In effect, the current in the coil magnetizes the iron and the iron then creates most of the surrounding magnetic field. Practical electromagnets, which control switches and valves in your furnace or air conditioner and can lift cars at the scrap yard, generally use iron or related materials to dramatically enhance the magnetic field produced by a current in a coil of wire (Fig. 11.1.12).

Courtesy Lou Bloomfield



Fig. 11.1.12 This electric switch is controlled by an electromagnet and is called a *relay*.

Check Your Understanding #6: Current Technology

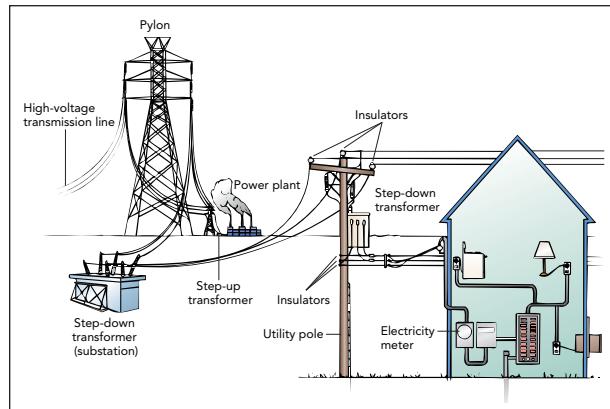
In magnetic resonance imaging (MRI), a patient is immersed in an intense magnetic field. That field is created entirely without permanent magnets or even iron. How is that possible?

Answer: The magnetic field is created by the current in a coil of wire.

Why: MRI requires a magnetic field that is intense, uniform, and spacious enough for a patient to fit inside. The best way to create such a colossal field is with a current-carrying coil. In fact, the field is so enormous that its flux lines extend far from the magnet and can attract iron or steel objects from across the room. Understandably, magnetic materials are forbidden near MRI magnets.

SECTION 11.2

Electric Power Distribution



Electricity is a particularly useful and convenient form of ordered energy. Because it's delivered to our homes and offices as a utility, we barely think about it, except to pay the bills. The wires that bring it to us never plug up or need cleaning and work continuously except when there's a power failure, blown fuse, or tripped circuit breaker.

Just how does electricity get to our homes? In this section, we'll look at the problems associated with distributing electricity far from the power plant at which it's generated. To understand these problems, we'll examine the ways in which wires affect electricity and see how electric power is transferred and rearranged by devices called transformers.

Questions to Think About: Why do power distribution systems use alternating current? What is the purpose of high-voltage wires? Why does the power company place large electric devices on the utility poles near homes or on the ground near neighborhoods? What are the advantages and disadvantages of 120-V versus 230-V electric power?

Experiments to Do: Experiments with electric power distribution are rather dangerous, but you can observe the ways in which your local electric power is distributed. If your area is connected to a major electric power network, you'll be able to find an entire hierarchy of power-conversion facilities. The power should travel from the power plant to your town on high-voltage wires, normally located overhead on tall pillars or pylons. These wires should end at a large power-conversion facility, where enormous devices transfer power to the lower-voltage power lines that fan out across your town. In some places, these wires are overhead; in others, they're underground. However, this power is still not ready for household use. It goes through at least one more stage of conversion before it reaches individual homes.

All these conversion steps are performed by transformers. You can find transformers as boxes or cylinders on utility poles or on the ground outside homes. In cities, transformers are often located inside buildings, out of sight. Even though you may have trouble finding these transformers, they're there. In this section, we'll see why they're necessary.

WARNING

Electricity is dangerous, particularly when it involves high voltages. The principal hazard is that an electric current will pass through your body in the vicinity of your heart and disrupt its normal rhythm. Very little current is needed to cause trouble, but your skin is such a poor conductor that it normally prevents harmful currents from passing through you. Be careful around high voltages, however, because they can propel dangerous currents through your skin and put you at risk. Although your body usually has to be part of a closed circuit for you to receive a shock, don't count on the absence of an identifiable circuit to protect you from injury—circuits have a tendency to form in surprising ways whenever you touch an electric wire. Be especially careful whenever you're near voltages of more than 50 V, even in batteries, or when you're near any voltages if you're wet or your skin is broken. That said, voltages of 12 V or less are so unlikely to injure you that they are generally considered safe. Nearly all household batteries provide voltages in that safe range and can be handled individually with little risk of shock. Similarly, most power adapters for electronic devices provide 12 V or less and thus rarely cause shocks.

Direct Current Power Distribution

Batteries may be fine for powering flashlights, but they're not very practical for lighting homes. Early experiments that placed batteries in basements were disappointing because the batteries ran out of energy quickly and needed service and fresh chemicals all too frequently.

A more cost-effective source for electricity was coal- or oil-powered electric generators. Like batteries, generators do work on the electric currents flowing through them and can provide the electric power needed to illuminate homes. But although generators produce electricity more cheaply than batteries, early ones were large machines that required fresh air and attention. These generators had to be built centrally, with people to tend them and chimneys to get rid of the smoke.

This was the approach taken by the American inventor Thomas Alva Edison (1847–1931) in 1882 when he began to electrify New York City following his development of a practical incandescent lightbulb ⁴. Each of the Edison Electric Light Company's generators acted like a mechanical battery, producing **direct current** (DC) that always left the generator through one wire and returned to it through another. Edison placed his generators in central locations and conducted the current to and from the homes he served through copper wires. However, the farther a building was from the generator, the thicker the copper wires had to be. That's because wires impede the flow of current and making them thicker allows them to carry current more easily.

Wire thickness is important because, like the filament of the flashlight bulb we studied in Chapter 10, wires have electrical resistance. In accordance with Ohm's law (Eq. 10.3.4), the voltage drop through a wire is equal to its electrical resistance times the current passing through it. In the case of a wire conducting current from a generating plant to a home, our primary concern is how much power the wire wastes as thermal power. We can determine this wasted power by combining Ohm's law with the equation for power consumed by a device (Eq. 10.3.1):

$$\begin{aligned} \text{power consumed} &= \text{voltage drop} \cdot \text{current} \\ &= (\text{current} \cdot \text{electrical resistance}) \cdot \text{current} \\ &= \text{current}^2 \cdot \text{electrical resistance.} \end{aligned} \tag{11.2.1}$$

⁴ Lewis Howard Latimer (African American scientist and inventor, 1848–1928) was only 8 years old when the U.S. Supreme Court's *Dred Scott* decision made his escaped-slave father a fugitive and forced him to disappear. Left behind with his mother, Latimer did well in school and became a skilled draftsman and engineer. While working for Edison's rival, Hiram Maxim, Latimer became an expert in fabricating carbon filaments for incandescent lamps. When Latimer later joined Edison's team of inventors, the "Edison Pioneers," his sturdy carbon filaments quickly replaced Edison's own fragile bamboo ones and provided the crucial ingredient necessary for Edison's lamps to become a commercial success.

The wire's wasted power is proportional to the square of the current passing through it! This relationship became all too clear to Edison when he tried to expand his power distribution systems. The more current he tried to deliver over a particular wire, the more power it lost as heat. Doubling the current in the wire quadrupled the power it wasted.

Edison tried to combat this loss by lowering the electrical resistances of the wires. He used copper because only silver is a better conductor of current. He used thick wires to increase the number of moving charges. He also kept the wires short so that they didn't have much chance to waste power. This length requirement forced Edison to build his generating plants within the cities he served. Even New York City contained many local power plants. (For an interesting tale about the early days of electric power, see [5].)

Edison also tried to minimize waste by decreasing the currents in the wires while increasing the voltage differences between those wires. Delivered power is equal to the current times the voltage drop (see Eq. 10.3.1), so while less current flowed through each home, the larger voltage drop left the delivered power unchanged.

The voltage difference between the two wires carrying current to and from a home is often referred to as the voltage of the electric power. Using that terminology, Edison needed to provide large-voltage or "high-voltage" DC electric power to his customers. High voltage power is dangerous, however, because it can cause fire-starting sparks and nasty shocks. It could be handled safely outside a home, but bringing it inside was another matter. Edison used the highest voltage power that safety allowed.

Although scientists have discovered a number of materials that lose their electrical resistance at extremely low temperatures and become perfect electrical conductors, or **superconductors** (Fig. 11.2.1), these superconductors are still too impractical for power-distribution systems. Their use is limited to local applications such as large electromagnets and specialized electronic devices.

► Check Your Understanding #1: The Trouble with DC Electric Power

If Edison doubled the length of his delivery wires, while keeping the currents through them the same, what would happen to the power they consumed?

Answer: The power would roughly double.

Why: Doubling the length of a wire is like placing two identical wires one after the next. If each wire uses 1 unit of power, then two wires should use roughly 2 units of power. Electrical resistance is proportional to the length of a wire and inversely proportional to the cross-sectional area of that wire. Shortening and thickening a wire reduce its resistance.

[5] Love Canal is the most famous U.S. toxic waste dump. The dump was created in the 1920s at an abandoned section of canal constructed in 1892 by William T. Love. Love intended his canal to connect the upper and lower Niagara Rivers so that the descending water could be used to generate DC electric power for the citizens of Niagara Falls, New York. The advent of AC power transmission systems in 1896 made the canal less useful, and it was never finished.

Courtesy Lou Bloomfield



Introducing Alternating Current

The real problem with distributing electric power via direct current is that there's no easy way to transfer power from one DC circuit to another. Because the generator and the light-bulbs must be part of the same circuit, safety requires that the entire circuit use low voltages and large currents. DC power distribution therefore wastes much of its power in the wires connecting everything together.

However, as we'll soon see, alternating current (AC) makes it easy to transfer power from one AC circuit to another so that different parts of the AC power-distribution system can operate at different voltages with different currents. Most significantly, the wires that carry the power long distances are part of a high-voltage, low-current circuit and therefore waste little power.

An **alternating current** is one that periodically reverses direction—it alternates. For example, when you plug a toaster into an AC electrical outlet and switch it on, the current that flows through the toaster's heating element reverses its direction of travel many times each second.

The power company propels this alternating current through the toaster by subjecting it to an alternating voltage drop, a voltage drop that periodically reverses direction. As you

Fig. 11.2.1 A magnetic cylinder floats above the surface of a superconductor at 78 K. Currents flowing freely in the superconductor make it magnetic and cause it to repel the magnetic cylinder.



Fig. 11.2.2 This electric outlet follows the U.S. standard for 120-V AC, 15-A service. The wide slots (left) are neutral, the narrow slots (right) are hot, and the curved holes (center) are ground. This outlet provides ground-fault circuit interruption (GFCI) protection; if any current leaving hot fails to return to neutral, or vice versa, the outlet shuts off instantly until it is reset. The test button simulates a current leak and will shut off the outlet if its protection is functioning properly.

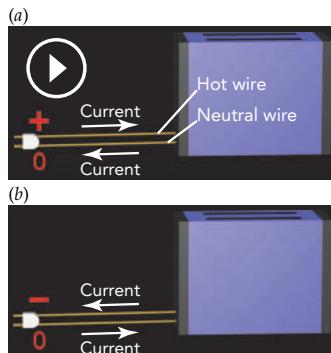


Fig. 11.2.3 When a toaster is plugged into a U.S. 120-V AC outlet, the voltage of its hot wire varies with time. The voltage of its neutral wire remains at 0 V. Since current always flows through the toaster from higher voltage to lower voltage, that current reverses each time the voltage of the hot wire reverses.

may recall from Section 10.3 on flashlights, current in a filament, heating element, or any other device that obeys Ohm's law flows down a voltage gradient, from a higher voltage to a lower voltage, much as bicyclists roll down an altitude gradient from a higher altitude to a low altitude or as water flows down a pressure gradient from a higher pressure to a lower pressure. While a flashlight's battery subjects the flashlight's filament to a steady voltage drop and obtains a direct current, the power company subjects the toaster's heating element to an alternating voltage drop and obtains an alternating current.

Alternating voltages are present at any AC electrical outlet. In the United States, an ordinary AC outlet offers three connections: *hot*, *neutral*, and *ground* (Fig. 11.2.2). In a properly installed outlet, the absolute voltage of *neutral* remains near 0 V, while the absolute voltage of *hot* alternates above and below 0 V. *Ground*, which is an optional safety connection that we'll discuss later, also remains near 0 V absolute.

One of the toaster's wires is connected to *hot* and the other to *neutral* (Fig. 11.2.3). Since current always flows through the toaster's heating element from higher voltage to lower voltage, it flows from *hot* to *neutral* when *hot* has a positive voltage (Fig. 11.2.3a). and from *neutral* to *hot* when *hot* has a negative voltage (Fig. 11.2.3b).

In normal AC electric power, the *hot* voltage varies sinusoidally—it's proportional to the trigonometric sine function with respect to time (Fig. 11.2.4). This smoothly alternating voltage propels a smoothly alternating current through the toaster. During each reversal, the current in the heating element gradually slows to a stop before gathering strength in the opposite direction. In the United States, AC voltages reverse every 120th of a second, yielding 60 full cycles of reversal (back and forth) each second (60 Hz). In Europe, the reversals occur every 100th of a second, so AC voltages complete 50 full cycles of reversal each second (50 Hz).

Fortunately, these reversals have little effect on many household devices. Toasters, electric heaters, and incandescent lightbulbs consume power because of their electrical resistances and don't care which way current passes through them. In fact, power consumption in such simple ohmic devices is used to define an effective voltage for AC electric power. An outlet's nominal AC voltage—technically, its **root mean square (RMS) voltage**—is defined to be equal to the DC voltage that would cause the same average power consumption in an ohmic device. Thus, 120-V AC power delivers the same average power to a toaster as 120-V DC power.

However, the reversals of AC power aren't without consequence. First, some electrical and most electronic devices are sensitive to the direction of current flow and must handle the reversals carefully. Second, the power available from an ordinary AC outlet rises and falls with each voltage reversal and is momentarily zero at the reversal itself. The toaster's heating element actually varies slightly in temperature because of these power fluctuations, and devices that can't tolerate even an instant without power must store energy to avoid shutting down during the reversals.

Finally, AC power's sinusoidally varying voltages peak well above their nominal values, exceeding those values by a factor of the square root of 2 (about 1.414). For example, the voltage of the *hot* connection in an ordinary 120-V AC power outlet actually swings between +170 V and -170 V. Those higher peak voltages are important for insulation and electrical safety.

Check Your Understanding #2: Timing Is Everything

Sticking your fingers into an electrical outlet is never a good idea, but is there a moment when you could do it without getting a shock?

Answer: Yes, you could do it at the moment the voltages are reversing.

Why: While the ground and neutral wires of the electrical outlet are normally without charge and therefore relatively safe, the hot wire is usually charged and dangerous. That hot wire's voltage alternates rapidly between high positive voltage and high negative voltage. Only when it's passing through 0 V can you touch it without risking a shock. However, that safe moment is so brief that you can't realistically avoid a shock. Don't try it!

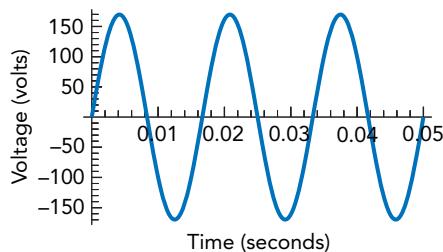


Fig. 11.2.4 The voltage of the hot wire of a U.S. 120-V AC outlet varies sinusoidally in time and completes 60 full cycles per second. Although it peaks at 6170 V, its effective time average or root mean square (RMS) voltage is 120 V. The voltage of the neutral wire is always 0 V.

Magnetic Induction

Edison was adamantly opposed to alternating current, which he viewed as dangerous and exotic. Indeed, its fluctuating voltages and moments without power don't make alternating current look attractive at all.

The champion of alternating current was Nikola Tesla (1856–1943), a Serbian American inventor who was backed financially by the American inventor and industrialist George Westinghouse (1846–1914). The advantage that Tesla and Westinghouse saw in alternating current was that its power could be transformed—it could be passed via electromagnetic action from one circuit to another by a device called a transformer.

A **transformer** uses two important connections between electricity and magnetism to convey power from one AC circuit to another. The first is familiar: moving electric charge creates magnetic fields. This connection allows electricity to produce magnetism. The second connection, however, is something new: magnetic fields that change with time create electric fields. Discovered in 1831 by Michael Faraday **6**, this relationship allows magnetism to produce electricity!

SECOND CONNECTION BETWEEN ELECTRICITY AND MAGNETISM

Magnetic fields that change with time produce electric fields.

Whether you wave a permanent magnet back and forth, or switch an electromagnet on and off, you are changing a magnetic field with time and thereby producing an electric field. If there are mobile electric charges around to respond to that electric field, they'll accelerate and you'll have created or altered an electric current and possibly done work on it as well. This process, whereby a time-changing magnetic field initiates or influences an electric current, is called **magnetic induction**.

A transformer combines these two connections in sequence—electricity produces magnetism produces electricity. However, rather than returning electric power to where it started, the transformer moves that power from the current in one coil of wire through a magnetic field to the current in a second coil of wire.

Check Your Understanding #3: Electric Phonographs

In the days of vinyl records, a phonograph reproduced sound by sliding a diamond stylus through a record's undulating groove. A magnet attached to that stylus moved up and down with each undulation and produced a current in a nearby coil of wire. Why did the magnet's motion affect the coil?

Answer: The moving magnet produced an electric field, which pushed mobile charges through that wire coil.

Why: The tiny vibrating magnet affected the coil's current via magnetic induction.

6 With only a primary education, English chemist and physicist Michael Faraday (1791–1867) apprenticed with a bookbinder at 14. At 21, he became a laboratory assistant to Humphry Davy, a renowned chemist. Faraday's experiments with electrochemistry and his knowledge of work by Oersted and Ampère led him to think that, if electricity can cause magnetism, then magnetism should be able to cause electricity. Through careful experimentation, he found just such an effect. Toward the end of his career, Faraday became a popular lecturer on science and made a particular effort to reach children.

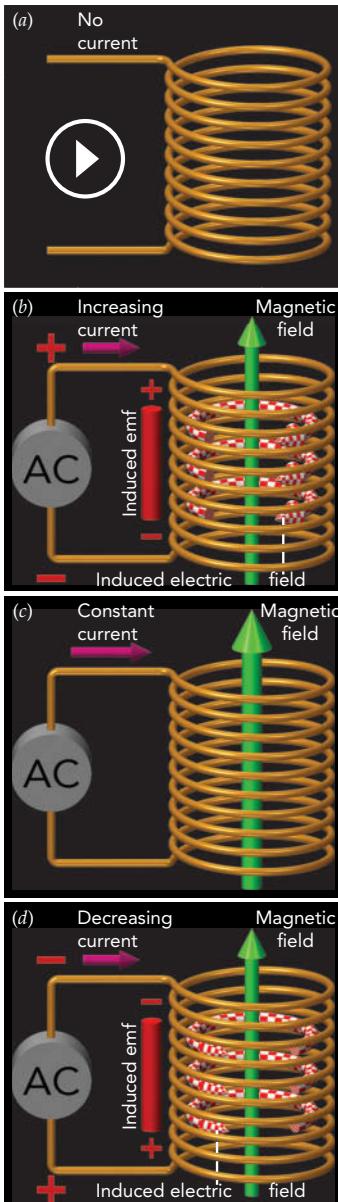


Fig. 11.2.5 (a) A current-free inductor has no magnetic field. (b) As the inductor's current and magnetic field increase, its induced electric field opposes the current increase and its induced emf takes energy from the current. (c) A constant current has a constant magnetic field and zero induced emf. (d) As the inductor's current and magnetic field decrease, its induced electric field opposes the current decrease and its induced emf gives energy to the current.

Alternating Current and a Coil of Wire

To help us understand the power transfer that takes place in a transformer, let's start with a simpler case. What happens when you send an alternating current through a single coil of wire?

Because currents are magnetic, the coil becomes an electromagnet. However, since the current passing through it reverses periodically, so does its magnetic field. Also, because a magnetic field that changes with time produces an electric field, the coil's alternating magnetic field produces an alternating electric field.

This induced electric field has a remarkable effect—it pushes on the very alternating current that produces it! Although it's not obvious how this electric field should affect that current, the result turns out to be simple (Fig. 11.2.5). As the coil's current increases, the induced electric field pushes that current backward and thereby opposes its increase (Fig. 11.2.5b). As the coil's current decreases, the induced electric field pushes that current forward and thereby opposes its decrease (Fig. 11.2.5d). No matter how the coil's current changes, the induced electric field always opposes that change!

This opposition to change is universal in magnetic induction, where it's known as **Lenz's law**: when a changing magnetic field induces a current in a conductor, the magnetic field from that current opposes the change that induced it. In other words, the effects of magnetic induction oppose the changes that produce them. In the present case, self-directed magnetic induction or “self-inductance” leads our coil to oppose its own changes in current. A wire coil's natural opposition to current change makes it quite useful in electrical equipment and electronics, where it's called an **inductor** (Fig. 11.2.6).

LENZ'S LAW

When a changing magnetic field induces a current in a conductor, the magnetic field from that current opposes the change that induced it.

However, our coil's induced electric field does more than just push the current around; it can also do positive or negative work on that current and thereby shift that current from one voltage to another. The coil's overall voltage shift, from one end of the coil to the other, is known as its **induced emf** (short for electromotive force). Current enters the coil at one voltage and exits at another voltage, courtesy of the coil's induced emf.

For comparison, a battery has an *electrochemical emf*; current enters the battery at one voltage and exits at another voltage, thanks to the battery's electrochemical emf. But while a battery's electrochemical emf is fixed, a coil's induced emf can change with time. When an alternating voltage difference is applied to the coil, its induced emf alternates and exactly matches the applied voltage difference.

A coil's ability to shift an alternating current gracefully from one voltage to another makes it possible to plug a properly designed coil into an AC electric outlet without causing disaster. As the outlet's voltage difference alternates, the coil's induced emf follows that voltage difference perfectly. Magnetic induction keeps the coil's current small and, if we neglect the coil's tiny electrical resistance, none of the AC electric power is wasted as thermal power.

As the coil's voltage difference and induced emf alternate, the coil's current keeps trying to flow from higher voltage to lower voltage. The coil's opposition to current change, however, delays the current's response so that its alternating current lags a quarter of an AC cycle behind its alternating voltage difference. For example, the current reaches its peak flow toward the top of the coil a quarter of an AC cycle after the voltage at the top of the coil reached its peak positive value. Although the coil's voltage difference and current both vary sinusoidally with time, the current has a phase shift or phase delay of 90° relative to the voltage difference.

Because of this phase shift, current flows through the coil from higher voltage to lower voltage only half the time. The other half, current flows from lower voltage to higher voltage. When current flows toward lower voltage, the induced emf removes electrostatic potential energy from the current. When current flow toward higher voltage, the induced emf returns electrostatic potential energy to the current. Energy alternately leaves the current and returns, but where does that energy reside when it's not in the current?

The missing energy is in the coil's magnetic field! Magnetic fields contain energy. The amount of energy in a uniform magnetic field is half the square of the field strength times the volume of the field divided by the permeability of free space. We can write this relationship as a word equation:

$$\text{energy} = \frac{\text{magnetic field}^2 \cdot \text{volume}}{2 \cdot \text{permeability of free space}}, \quad (11.2.2)$$

in symbols:

$$U = \frac{B^2 \cdot V}{2 \cdot \mu_0},$$

and in everyday language:

Strong permanent magnets store so much magnetic energy that they can be dangerous if you break them. The pieces will flip around violently, and you may get pinched.

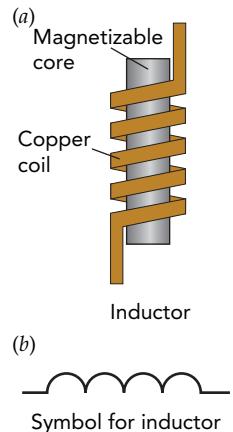


Fig. 11.2.6 (a) An inductor is a coil of wire that stores energy in its magnetic field. To increase its inductance, the coil may contain a magnetizable core of iron or ferrite. (b) In a schematic diagram of an electronic device, the inductor is represented by a stylized coil.

COMMON MISCONCEPTIONS: Magnets as Limitless Sources of Energy

Misconception: Magnets are infinite sources of energy that could provide electric or mechanical power forever!

Resolution: Although a magnet's field does contain energy, that energy is limited and was invested in it during its magnetization. To extract that energy, you'd have to demagnetize and thus destroy the magnet.

In effect, our coil is playing with the alternating current's energy, storing it briefly in the magnetic field and then returning it to the current. The coil stores energy while the magnitude of the current increases—the field strengthens and the current loses voltage. The coil returns energy while the magnitude of the current decreases—the field weakens and the current gains voltage. Because the coil's self-induced emf is responsible for bouncing this energy back to the current, it's frequently called a **back emf**.

Check Your Understanding #4: Slow Fall

If you drop a strong magnet onto a nonmagnetic but highly conducting surface, the magnet will descend remarkably slowly. What's delaying the magnet's fall?

Answer: The falling magnet is inducing currents and magnetism in the surface. In accordance with Lenz's law, that induced magnetism opposes the change that produces it; it acts to slow the magnet's descent.

Why: A strong magnet induces such powerful magnetic opposition in a good conductor that moving the magnet is difficult. This effect is most evident with a superconductor, that is, a material that conducts electricity perfectly and can sustain induced currents forever. A superconductor can slow a falling magnet to a stop and hold it suspended indefinitely (Fig. 11.2.1).

Check Your Figures #1: Field the Energy

An MRI diagnostic unit fills about 0.1 m^3 of space with a 4-T magnetic field. How much energy is contained in that field?

Answer: The field contains about 640,000 J.

Why: Since 1 T is equivalent to $1 \text{ N/A} \cdot \text{m}$, Eq. 11.2.2 gives us the energy of a 4-T field occupying 0.1 m^3 as:

$$\begin{aligned}\text{energy} &= \frac{(4 \text{ N/A} \cdot \text{m})^2 \cdot 0.1 \text{ m}^3}{2 \cdot (4\pi \times 10^{-7} \text{ N/A}^2)} \\ &= 640,000 \text{ N} \cdot \text{m} = 640,000 \text{ J}.\end{aligned}$$

Two Coils Together: A Transformer

A single coil experiences only self-inductance. Any energy removed from the coil's current by its induced emf must eventually be returned to that same current; it has nowhere else to go. But when two coils share the same electromagnetic environment, they experience mutual inductance and can exchange energy via magnetic induction. Energy removed from one coil's current by its induced emf can be given to the other coil's current by its induced emf.

That possibility is the basis for a transformer, a device that transfers electric power from one circuit to another. In its simplest form, a transformer consists of two coils, primary and secondary, wrapped around a magnetizable core that enhances magnetic induction and allows them to share the same electromagnetic environment. When alternating current flows through the primary coil, it produces an induced electric field that affects both coils and both coils develop induced emfs. The induced emf in the primary coil removes energy from its alternating current while the induced emf in the secondary coil gives that energy to its alternating current.

I'll explain how this energy transfer works by starting with the transformer shown in Fig. 11.2.7a. A generator provides its primary coil with 120-V AC electric power, while its secondary coil is an open circuit.

With the generator subjecting the primary coil to an alternating voltage difference, the primary coil behaves like an inductor. It carries an alternating current and develops an induced emf that exactly matches the voltage difference imposed by the generator. Although the coil's current naturally tries to flow from higher voltage to lower voltage, the coil's self-inductance opposes current changes and delays the current's response. As a result, the coil's alternating current lags a quarter of an AC cycle or 90 degrees in phase behind the coil's alternating voltage difference.

Because of its 90-degree phase lag, this alternating current conveys no average power from the generator to the primary coil. Nonetheless, it plays an important role in the transformer: it alternately magnetizes and demagnetizes the transformer's core and thereby produces the induced emfs in both coils. For that reason, it is known as the *magnetizing current*.

The transformer's secondary coil is identical to its primary coil, except for being upside down on the right side of the magnetizable core. The secondary coil's induced emf is therefore identical to that of the primary coil, except for being upside down. Though it's an open circuit and can't carry a significant current, the secondary coil has a voltage difference across it due to the induced emf and it can thus act as a source of 120-V AC electric power!

In Fig. 11.2.7b, the transformer's secondary is connected to a lightbulb. Because the lightbulb's filament obeys Ohm's law, its current remains proportional to the voltage difference across it, even if that voltage difference alternates. With the secondary coil imposing an alternating voltage difference on the filament, that filament carries an alternating current

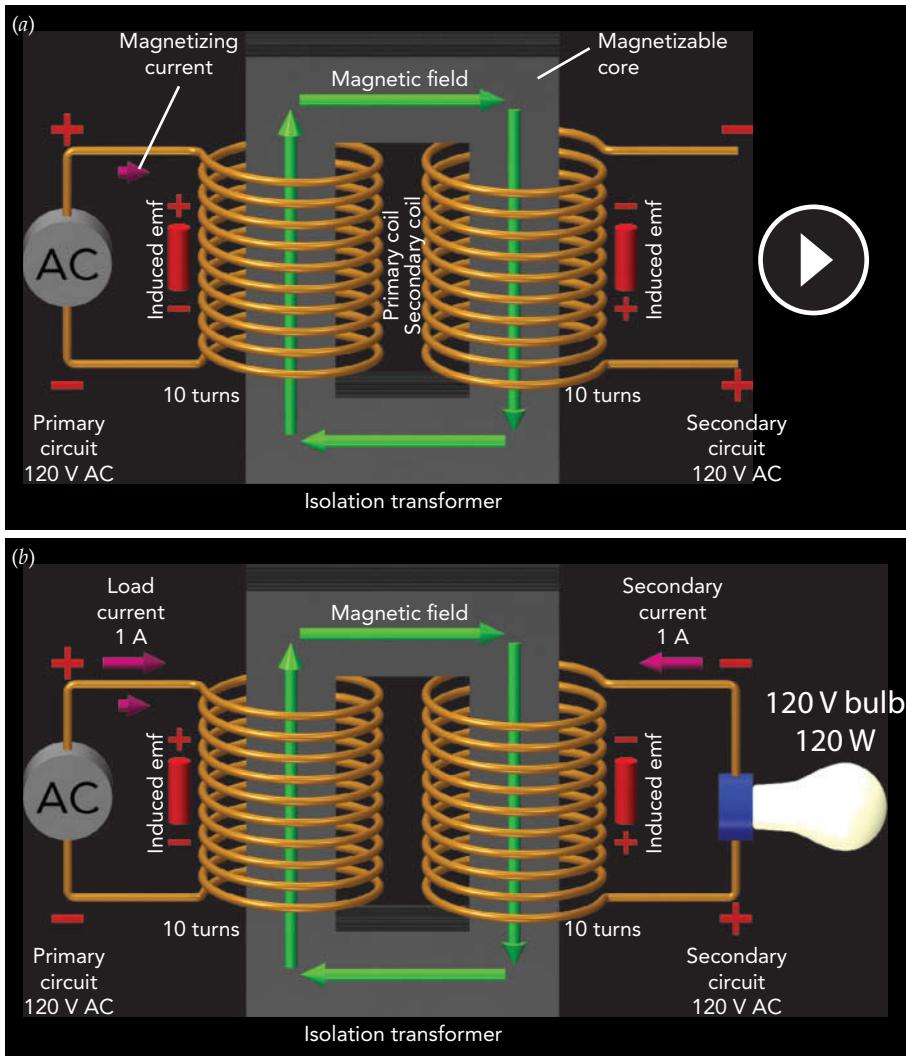


Fig. 11.2.7 (a) When 120-V AC electric power causes a magnetizing current to flow in the primary coil of this isolation transformer, both coils develop 120-V AC induced emfs. (b) When a lightbulb completes the secondary circuit, the resulting secondary current causes an additional load current to flow in the primary circuit. Overall, the transformer transfers power from its primary circuit to its secondary circuit.

that is synchronized with the voltage difference. That current always flows toward lower voltage in the filament, dropping off energy, and toward higher voltage in the secondary coil, picking up energy. Provided with 120-V AC power by the secondary coil, this particular lightbulb filament carries an AC current of 1 A.

The current that is now flowing through the secondary coil has its own magnetic effect on the transformer. Fortunately, that effect is surprisingly simple: it causes an additional current to flow through the primary coil.

Known as the *load current*, this additional current shares the same conducting path as the magnetizing current, like two fleets of cars sharing the same two-lane road. Unlike the magnetizing current, however, the load current is synchronized with the voltage difference across the primary coil and its two effects are different. It produces a magnetic field that exactly cancels the magnetic field of the secondary current and it conveys electric power from the generator to the primary coil.

I'll start with the magnetic cancellation. When a current travels in a circle, the magnetic field inside the circle is proportional to the current. A wire coil guides its current in a circle multiple times, effectively multiplying that current by the number of turns in the coil. The magnetic field inside a coil of wire is thus proportional to the current in the wire times the number of turns in the coil.

Since the transformer's primary and secondary coils are identical, they have the same number of turns. If the currents in those two coils are equal in amount but opposite in direction, their magnetic fields will cancel.

The transformer naturally adjusts its load current to achieve that magnetic cancellation. Since the secondary coil has a 1-A alternating current flowing through it toward higher voltage, the primary coil acquires a 1-A alternating load current flowing through it toward lower voltage. Their magnetic fields cancel perfectly, leaving the magnetizing current solely responsible for the transformer's magnetic field and its induced emfs.

Because the coils are identical, their induced emfs are both 120-V AC. The negative work that the primary coil's induced emf does on its 1-A current is therefore equal in amount to the positive work that the secondary coil's induced emf does on its 1-A current. Overall, the transformer is transferring an AC power of 120 W from the current in its primary circuit to the current in its secondary circuit.

To maintain their magnetic cancellation, the load current automatically mirrors any changes in the secondary current. For example, if we add a second lightbulb to the secondary circuit and thereby double the current in the secondary coil, the load current will also double. Because of this mirroring effect, the transformer always consumes as much power from its primary circuit as is consumed by its secondary circuit.



Check Your Understanding #5: Use Only with AC Power

If you send direct current through the primary coil of a transformer, no power will be transferred to the secondary circuit. Explain.

Answer: When direct current flows through the transformer's primary coil, it creates a constant magnetic field around the iron core. Since that field doesn't change, it doesn't create any electric fields and doesn't induce current in the transformer's secondary coil.

Why: The current through the primary coil must change so that the magnetic field in the coils will change and current will be induced in the secondary coil. Transferring power from one circuit to another is so useful that there are many DC-powered devices that switch their power on and off to mimic alternating current so that they can use transformers.

Changing Voltages

When a source of AC electric power is connected to a transformer's primary coil, the transformer's primary and secondary coils both develop induced emfs. If the secondary coil is identical to the primary coil (Fig. 11.2.7), their induced emfs are equal and the secondary coil becomes a source of AC power at the same voltage as that received by the primary coil. For example, if you plug the transformer's primary coil into a 120-V AC outlet, its secondary coil will provide 120-V AC electric power to the secondary circuit.

A transformer with identical coils is known as an *isolation transformer* and it provides an important measure of electrical safety. Since its primary and secondary circuits are electrically isolated, charge can't move between those circuits and cause trouble. For example, when lightning strikes the power company's wires, the resulting burst of charge on the primary circuit can't pass to any appliances or instruments that are part of the secondary circuit. Not surprisingly, hospitals often employ isolation transformers to protect patients from shocks.

Most transformers, however, have unequal coils and therefore different induced emfs in those coils. They provide AC power at voltages that are different from those they receive.

The primary coil's induced emf naturally matches the voltage difference applied to it, but the secondary coil's emf can vary. It depends on the number of secondary turns—the number of times the secondary coil's wire encircles the magnetizable core. The more loops the secondary current makes around the core, the more positive or negative work the transformer's induced electric field does on that current and the larger the secondary coil's induced emf.

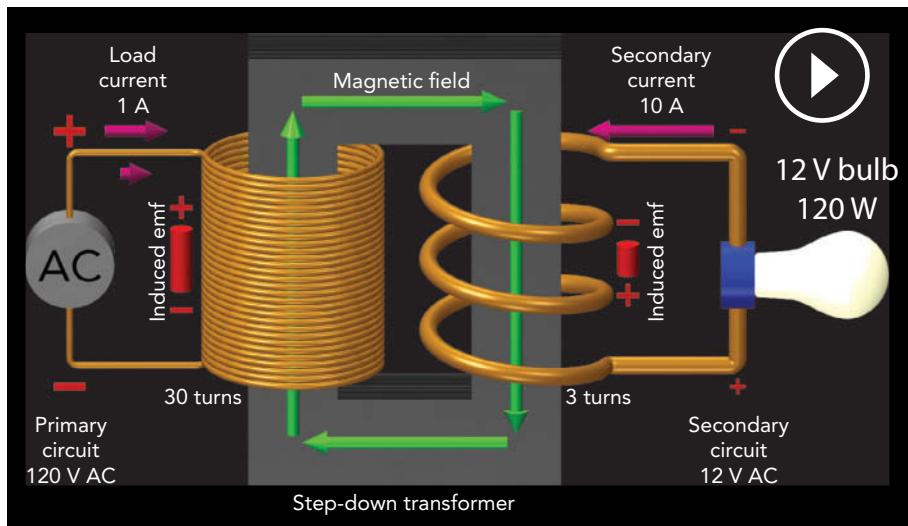


Fig. 11.2.8 Because this step-down transformer has 1/10th as many turns in its secondary coil as in its primary coil, its secondary coil provides 1/10th of the AC voltage provided to its primary coil. It transforms 120-V AC power into 12-V AC power for the low-voltage bulb. Its secondary current is 10 times its primary load current.

Since the secondary coil's induced emf is proportional to its number of turns, it acts as a source of AC power with a voltage equal to the voltage applied to the primary coil times the ratio of secondary turns to primary turns, or

$$\text{secondary voltage} = \text{primary voltage} \cdot \frac{\text{secondary turns}}{\text{primary turns}} \quad (11.2.3)$$

An isolation transformer is simply the special case in which the turn numbers are equal and their ratio is 1.

When a transformer has fewer secondary turns than primary turns (Fig. 11.2.8), it provides a secondary voltage that is less than the primary voltage and is called a *step-down transformer*. Step-down transformers are common in electronic devices and power adapters, where they step household AC voltages down to much smaller AC voltages. For example, if a transformer's ratio of secondary turns to primary turns is 0.1 and you supply 120-V AC power to its primary coil, its secondary coil will act as a source of 12-V AC power.

Not surprisingly, there are also *step-up transformers* that have more secondary turns than primary turns and that provide secondary voltages that are greater than their primary voltages (Fig. 11.2.9). The transformer that powers a neon sign typically has 100 times as many turns in its secondary coil as in its primary coil. When its primary coil is supplied by 120-V AC power, its secondary coil provides the 12,000-V AC power needed to illuminate the neon tube.

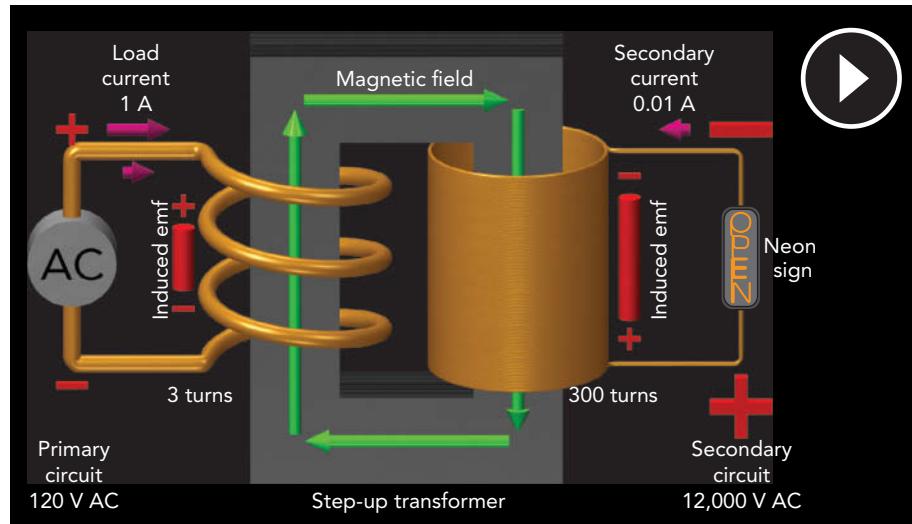
Even when a transformer has unequal coils, the magnetic fields produced by the load current in its primary coil and the secondary current in its secondary coil must still cancel. Since each coil's magnetic field is proportional to its current times its number of turns, the coil with fewer turns must compensate by carrying a larger current. As a result, a transformer's secondary current is equal to its primary load current times the ratio of primary turns to secondary turns, or

$$\text{secondary current} = \text{primary load current} \cdot \frac{\text{primary turns}}{\text{secondary turns}}, \quad (11.2.4)$$

where those currents flow in opposite directions; the primary load current flows toward lower voltage in the primary coil while the secondary current flows toward higher voltage in the secondary coil.

As you can see, changing a transformer's ratio of secondary turns to primary turns affects both the voltage and the current in the secondary circuit. The secondary voltage is proportional to that ratio, while the secondary current is inversely proportional to that ratio.

Fig. 11.2.9 Because this step-up transformer has 100 times as many turns in its secondary coil as in its primary coil, its secondary coil provides 100 times the AC voltage provided to its primary coil. It transforms 120-V AC power into 12,000-V AC power for the neon sign. Its secondary current is 1/100th its primary load current.



The product of those two quantities, secondary voltage and secondary current, no longer depends on the turn ratio and it's the power provided to the secondary circuit. It's also equal to the primary voltage times the primary load current, which is the power received from the primary circuit. The transformer is providing the same power to its secondary circuit as it's receiving from its primary circuit!

With that in mind, we can look again at the three types of transformers. An isolation transformer (Fig. 11.2.7) transfers AC power without any change in voltage or current; the secondary circuit has the same voltage and the same current as in the primary circuit. With a step-down transformer (Fig. 11.2.8), the secondary circuit has a smaller voltage and a larger current than in the primary circuit. With a step-up transformer (Fig. 11.2.9), the secondary circuit has a larger voltage and a smaller current than in the primary circuit.

Check Your Understanding #6: Travel Trouble

Your portable lava lamp operates on 120-V AC power, but you're visiting a country with 240-V AC power. You plug a travel adapter into the 240-V AC outlet and its transformer provides your lamp with the 120-V AC power it expects. Compare the numbers of turns in the transformer's two coils.

Answer: The transformer's secondary coil has half as many turns as its primary coil.

Why: To step down the voltage, a transformer must have fewer turns in its secondary coil than in its primary coil. Fewer turns leads to a smaller emf in the secondary coil and a smaller output voltage for the transformer.

Real Transformers: Not Quite Perfect

Although we've been pretending that inductors and transformers are flawless and that their wires conduct electricity perfectly, that's not quite true. In reality, the wires used in those devices have electrical resistances and waste power in proportion to the squares of the currents they carry. To minimize this wasted power, real inductors and transformers are designed to minimize their resistances. To the extent it is practical, they employ thick wires made of highly conducting metals and those wires are kept as short as possible.

Unfortunately, inductors and transformers built from wires alone can't develop strong magnetic fields and large induced emfs unless they carry large magnetizing currents or have long, many-turn coils. To avoid those current or coil problems, many inductors and virtually all transformers wrap their coils around magnetizable cores. Those cores respond magnetically to the alternating currents around them, enhancing their magnetic fields

and increasing the induced emfs. Aided by those magnetizable materials—typically iron or iron alloys—cored inductors and transformers work well even with short, few-turn coils.

A core provides another crucial benefit to a transformer; it guides the transformer's magnetic flux lines so that nearly all of them pass through both coils, even when those coils are somewhat separated in space. Sharing their flux lines in that manner gives the coils a common electromagnetic environment and permits them to exchange electric power easily.

Making two separate coils share their flux lines isn't easy. Since a coil has no net magnetic pole, each flux line that emerges from it must ultimately return to it. Without a core, however, most flux lines leaving a coil return to it almost directly and remain nearby throughout their trip. Those unadventurous flux lines are unlikely to pass through a second, separate coil. Not surprisingly, a coreless transformer works well only when its two coils are wound so closely together that they can't help but share the same flux lines.

Winding both coils around a ring-shaped magnetic core makes it easy for the flux lines to pass through both coils because those flux lines are drawn into the core's soft magnetic material and follow it as if in a pipe. Although the flux lines leaving a coil must still return to it eventually, most of them complete that trip by way of the core—a journey that then takes them through the other coil. With nearly all the flux lines channeled by the core through both coils, power can flow easily from one coil to the other.

A core thus provides a transformer with great flexibility; its coils can be practically anywhere as long as they encircle that core. However, cores aren't quite perfect pipes for flux; they leak slightly. Therefore, the most efficient transformers have coils that are wound nearby or on top of one another.

Although magnetizable cores make small efficient transformers practical, they also introduce a few problems. First, the cores must magnetize and demagnetize easily to keep up with the magnetizing current in the primary coil. If they lag behind, they'll waste power as thermal power. Sadly, perfect magnetic softness is unobtainable and all cores waste at least a little power through delays in their magnetizations.

Second, because these cores are subject to the same electric fields that push currents around in the coils, they shouldn't conduct electricity. If they do, they'll develop useless internal currents known as *eddy currents* and thereby waste power heating themselves up. Since most soft magnetic materials are electrical conductors, transformer cores are frequently divided up into insulated particles or sheets so that little or no current can flow through them. Despite best efforts at minimizing resistive heating in their coils, and magnetization and eddy current losses in their cores, all transformers still waste some power. Even the best transformers are only about 99% energy efficient.

Check Your Understanding #7: Winds of Change

Large power transformers have cooling fins and often fans to blow air across them. Why does a transformer need this cooling?

Answer: Its magnetic core converts some of the electric power into thermal power. Unless that thermal power is eliminated, the transformer will overheat.

Why: Transformers aren't perfectly energy efficient; they convert a small fraction of the electric power they handle into thermal power. Their magnetic cores contribute to that inefficiency because their limited magnetic softness and electric conductivity cause them to heat up. Fins and fans are essential for keeping large transformers cool.

Alternating Current Power Distribution

We're finally prepared to deal with the basic conflicts of power transmission. To minimize resistive heating in the power lines connecting a power plant with a distant city, electric power should travel through those lines as small currents at very high voltages. To be practical, though, as well as to avoid shock and fire hazard, electric power should be delivered to homes as large currents at modest voltages.

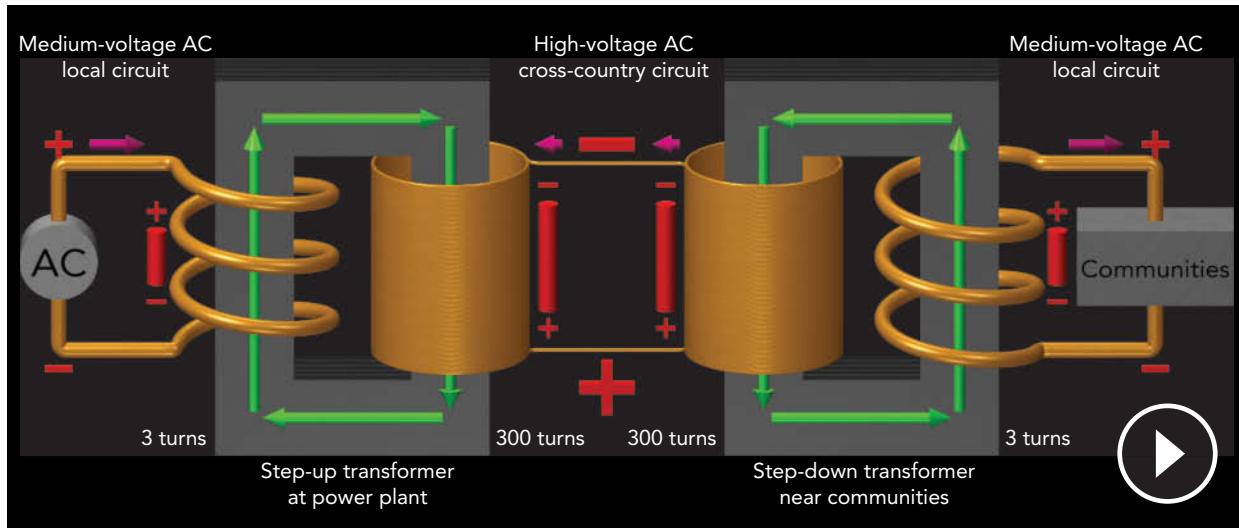


Fig. 11.2.10 Power from an AC generator is transmitted cross-country by stepping it up to very high voltage at the power plant, transmitting it long distance as a small current at very high voltage, and stepping it back down to medium voltage near the communities that are to be served. The three circuits used in this delivery system are electrical insulated from one another.

Although there is no simple way to meet both requirements simultaneously with direct current, transformers make it easy to satisfy them both with alternating current. We can use a step-up transformer to produce the very-high-voltage AC electric power suitable for cross-country transmission and a step-down transformer to produce the low-voltage AC electric power that's appropriate for delivery to communities (Fig. 11.2.10).

At the power plant, the generator pushes a huge alternating current through the primary circuit of a step-up transformer at a supply voltage of about 5000 V AC. The current flowing through the secondary circuit is only about 1/100 the current in the

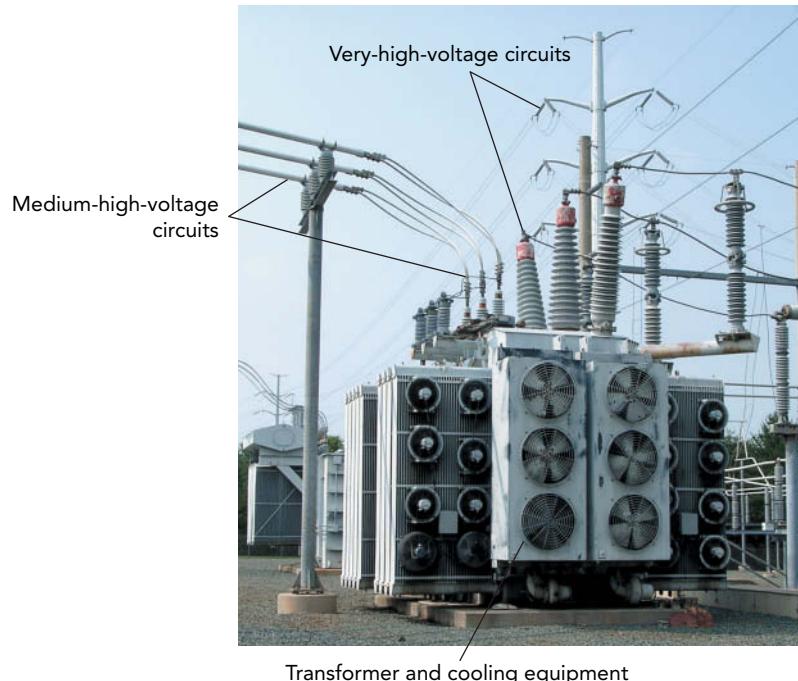


Fig. 11.2.11 This giant transformer transfers millions of watts of power from the very-high-voltage cross-country circuits above it to the medium-high-voltage neighborhood circuits to its left. Fans keep the transformer from overheating.

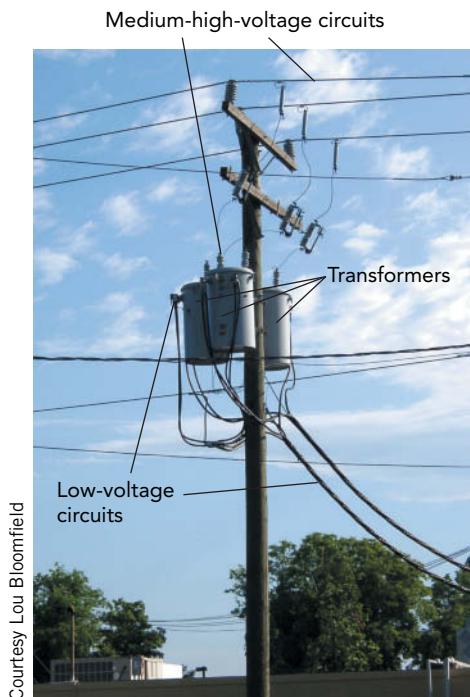


Fig. 11.2.12 The three metal cans on this utility pole are transformers. They transfer power from the medium-high-voltage neighborhood circuits above them to the low-voltage household circuits at the lower right.



Fig. 11.2.13 This transformer transfers power from a medium-voltage underground circuit to a low-voltage underground circuit used by nearby homes. It handles 50 kV·A or 50,000 W of power.

 Check Your Understanding #8: High-Voltage Wires

If a power utility were able to increase the voltage of its transmission line from 500,000 to 1,000,000 V, how would that affect the power lost to heat in the wires?

Answer: It would reduce the amount of heat produced to only 25% of the previous value.

Why: At 1,000,000 V, the transmission line would be able to carry the same power as a 500,000-V transmission line with only half the current. Since the power wasted by the transmission line itself is proportional to the square of the current, halving the current would reduce the power waste to 25%.

AC Electric Generators and Motors

As we've seen, a transformer "converts" electric power into electric power—it extracts electric power from a primary circuit and delivers electric power to a secondary circuit. However, electric power and mechanical power are physically equivalent, and mechanical power can substitute for electric power. What if we replace one of the electric circuits in a transformer with a mechanical system?

If we replace the transformer's primary circuit with a mechanical system, we obtain a generator. A generator is a device that extracts mechanical power from machinery and delivers electric power to a circuit. Figure 11.2.14a shows a simple generator, one that

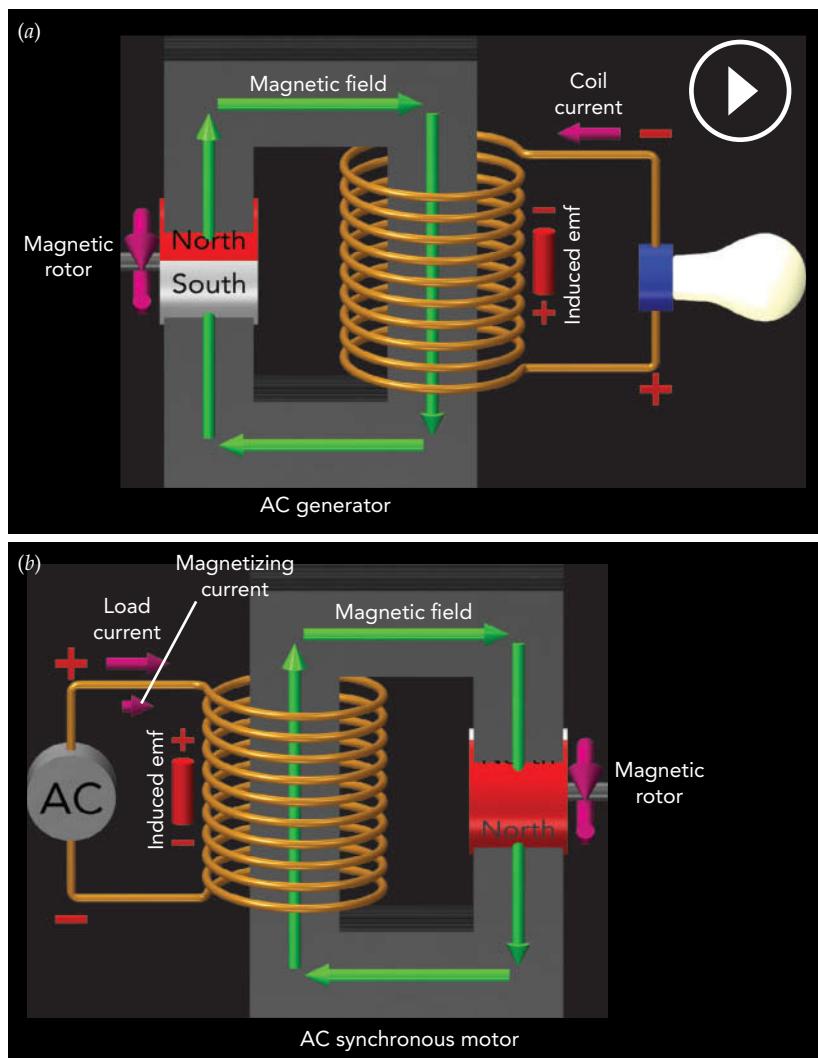


Fig. 11.2.14 (a) This AC generator resembles the transformer in Fig. 11.2.7, except that it receives power from the motion of its spinning magnetic rotor rather than from an electric current in a primary coil. (b) This AC synchronous motor also resembles the transformer, except that it provides power in the motion of its spinning magnetic rotor rather than in an electric current in a secondary coil.

looks strikingly like the transformer in Fig. 11.2.7. Both devices have a (secondary) coil wrapped around a magnetizable core. However, in place of the transformer's primary coil, the generator has a spinning magnet, or rotor. As the generator's magnetic rotor spins, it produces a sinusoidally alternating magnetic field in the core. This alternating magnetic field, in turn, produces an alternating electric field and an alternating induced emf in the coil. That emf propels an alternating current through the circuit that delivers electric power to the lightbulb.

The current in the generator's coil has consequences for its rotor. The rotor turns slightly ahead of the generator's alternating magnetic field, so that the rotor's magnetic field cancels the current's magnetic field. Because the rotor is a little ahead the generator's alternating magnetic field, it experiences a backward torque and has mechanical power extracted from it. To keep the rotor spinning, the machinery must continue to supply mechanical power to the generator. Overall, the generator is converting mechanical power into electric power.

If we replace the transformer's secondary circuit with a mechanical system, we obtain a motor. A motor is a device that extracts electric power from a circuit and delivers mechanical power to machinery. Figure 11.2.14b shows a simple motor, one that again resembles the transformer in Fig. 11.2.7. Like the transformer, the motor has a (primary) coil wrapped around a magnetizable core. In place of the transformer's secondary coil, however, the motor has a spinning magnetic rotor. As an alternating current flows through the motor's circuit and coil, it produces a sinusoidally alternating magnetic field in the coil. That alternating magnetic field interacts with the magnetic rotor and delivers mechanical power to it.

The rotor's motion has consequences for the current in the motor's coil. If the rotor is spinning freely and remains perfectly synchronized with the core's alternating magnetic field, it has no effect on the core's magnetic field or the current in the coil. If the rotor is doing mechanical work, however, it needs a forward torque to keep it spinning and it obtains that torque by turning slightly behind the motor's alternating magnetic field. The rotor's magnetic field then causes a load current to flow through the motor's coil. That load current delivers power to the motor and its magnetic field cancels the rotor's magnetic field. Overall, the motor is converting electric power into mechanical power.

Notice that Figs. 11.2.14a and 11.2.14b are almost mirror images of one another. That's because generators and motors are wonderfully similar devices. In fact, a single device can often act as either a generator or a motor. If you supply electric power to its circuit, its rotor will spin and provide mechanical power. If you supply mechanical power to its rotor, current will flow through its circuit and provide electric power.

Check Your Understanding #9: Electric Biking

When you pedal a high-tech exercise bicycle, you are probably spinning the rotor of an electric generator. That generator supplies power to a heating filament with an adjustable electrical resistance. How should the bicycle alter that electrical resistance to make pedaling more difficult?

Answer: It should reduce the heater's resistance.

Why: By lowering the heater's resistance, the bicycle increases the current flowing through the circuit. That increased current carries more power from the generator to the heater, so the generator extracts more mechanical work from the bicyclist.

Epilogue for Chapter 11

In this chapter, we studied magnetism and the ways in which magnetism relates to electricity. In Household Magnets, we looked at the concept of magnetic pole and the attractive or repulsive forces that poles exert on one another. We examined magnetic materials and saw how their magnetic properties make them useful for various purposes. We also encountered electromagnets and began to see that magnetism isn't independent of electricity. In Electric

Power Distribution, we saw how alternating electric currents make it possible to transfer power from one circuit to another by way of a transformer and its electromagnetic properties. We learned that transforming electric power to extremely high voltages and small currents minimizes the power wasted between power plants and cities.

Explanation: A Nail and Wire Electromagnet

When you connect the wire from one terminal of the battery to the other, a current flows from the positive terminal to the negative terminal through the wire. (In reality, negatively charged electrons move from the battery's negative terminal, through the wire, to its positive terminal, but we've adopted a fiction that positive charges are heading the other way.) This current produces a magnetic field around the wire. Because the wire is coiled around the nail, this magnetic field passes through the nail and causes its magnetic domains to resize until most of them are aligned with the field. Without any current in the wire, the magnetic domains in the steel point in many different directions, so the nail appears non-magnetic. However, with the current orienting the domains, they together produce a large magnetization. The nail becomes magnetic and exerts strong magnetic forces on other nearby objects.

Chapter Summary and Important Laws and Equations

How Household Magnets Work: Common refrigerator magnets are composed of hard magnetic materials that were permanently magnetized by their manufacturers. Simple button magnets have a single pair of magnetic poles, one north and one south, but plastic sheet magnets usually have many poles. These magnets stick to a refrigerator's surface by temporarily magnetizing that surface's soft magnetic materials and then becoming attracted to the opposite poles on that surface.

A compass is another permanent magnet, but one designed to align with Earth's magnetic field. In fact, that magnetic field can be mapped out using a compass. The magnetic fields around smaller magnets can be made visible with iron filings instead. However, permanent magnets aren't the only sources of magnetic fields; we found that when current flows through the coil in a doorbell, it becomes a magnet as well—an electromagnet.

How Electric Power Distribution Works: To minimize power losses in the transmission lines between power plants and cities, power distribution systems use alternating currents and transformers. Near the power plant, relatively low-voltage, high-current electric power is transformed into very-high-voltage, low-current power for transmission through cross-country power lines. Because the power consumed by these high-voltage wires depends on the square of the currents they carry, the power losses are greatly reduced by this technique. When the power arrives at a city, it's transformed into medium-voltage, high-current power for distribution to neighborhoods. Finally, in neighborhoods, step-down transformers transform this power to low-voltage, very-high-current power for distribution to individual homes and offices.

1. Coulomb's law for magnetism: The magnitudes of the magnetostatic forces between two magnetic poles are equal to the permeability of free space times the product of the two magnetic poles divided by 4π times the square of the distance separating them, or

$$\text{force} = \frac{\text{permeability of free space} \cdot \text{pole}_1 \cdot \text{pole}_2}{4\pi \cdot (\text{distance between poles})^2}. \quad (11.1.1)$$

If the charges are like, then the forces are repulsive. If the charges are opposite, then the forces are attractive.

2. Force exerted on a pole by a magnetic field: A pole experiences a force equal to its pole times the magnetic field, or

$$\text{magnetostatic force} = \text{pole} \cdot \text{magnetic field}, \quad (11.1.2)$$

where the force points in the direction of the field.

3. Lenz's law: When a changing magnetic field induces a current in a conductor, the magnetic field from that current opposes the change that induced it.

4. Power consumed by a wire or other ohmic device:

$$\text{power consumed} = \text{current}^2 \cdot \text{electrical resistance}. \quad (11.2.1)$$

5. Energy in a magnetic field: The energy in a magnetic field is equal to the square of that field times its volume, divided by twice the permeability of free space, or

$$\text{energy} = \frac{\text{magnetic field}^2 \cdot \text{volume}}{2 \cdot \text{permeability of free space}}. \quad (11.2.2)$$

6. Transformer voltages: A transformer's secondary coil acts as a source of AC power with a voltage equal to the AC voltage

applied to its primary coil times the ratio of secondary turns to primary turns or

$$\text{secondary voltage} = \text{primary voltage} \cdot \frac{\text{secondary turns}}{\text{primary turns}}. \quad (11.2.3)$$

7. Transformer currents: The AC current in a transformer's secondary coil is equal to the AC load current in its primary coil times the ratio of primary turns to secondary turns or

$$\text{secondary current} = \text{primary load current} \cdot \frac{\text{primary turns}}{\text{secondary turns}}. \quad (11.2.4)$$

Electromagnetic Waves

Electric and magnetic fields are so intimately related that each can create the other even in empty space. In fact, the two fields can form electromagnetic waves, in which they recreate one another endlessly and head off across space at an enormous speed. These electromagnetic waves are all around us and are the basis for much of our communications technology, for radiative heat transfer, and for our ability to see the universe in which we live.

ACTIVE LEARNING EXPERIMENTS

A Disc in the Microwave Oven

You can experiment with electromagnetic waves using a microwave oven. As we'll see in Section 12.2, microwaves are a type of electromagnetic wave that falls between radio waves and light. Because electromagnetic waves consist of electric and magnetic fields, they can propel electric currents through metal objects. Those currents can do some interesting things.

In this experiment, you'll put metal in a microwave oven. There is always some risk associated with this activity, so if you cannot fully accept that risk yourself, skip this experiment and look at the photographs instead. If you're young enough to require adult supervision or consent, obtain it or skip the experiment.

If you choose to conduct this experiment, make sure that there is nothing flammable in or near the oven and that the area around you is well ventilated. The plastic in the disc will heat up and release a mildly unpleasant smell that will dissipate. Do not leave the oven on for more than 4 s or that smell will become truly unpleasant.

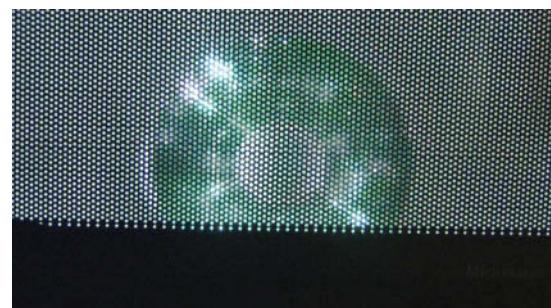
The metal for this experiment is the ultrathin reflective film located inside a CD or DVD. Although that metal layer's electrical conductivity makes it reflective, it wasn't designed to carry large electric currents. When

exposed to the microwave electric fields in the oven, that metal will carry large currents, heat up, tear, and spark.

To perform the experiment, you'll need a CD or DVD that you are comfortable destroying. Place a microwave-safe ceramic mug near the center of the microwave oven. Lean the CD or DVD against the mug so that you can see the disc's shiny face through the oven door. If the oven has a rotating tray, you may want to remove that tray temporarily for the experiment. You may also want to block the oven's light temporarily with black tape. Be sure that you replace the tray and unblock the light when you're done.

Close the oven door, and check that you can see the shiny surface of the CD or DVD clearly. Now prepare to turn the microwave oven on for no more than 4 s. You should know what to expect: 2 s during which the oven will build up its microwave power and 2 s during which the disc's metal layer will spark wildly.

When you're prepared and everything is safely arranged, turn the oven on for 4 s. You should see lots of sparking in that metal layer. Even if you don't, turn the oven off after 4 s or the overheated plastic smell will become a serious problem. Let the disc cool and the smell dissipate. Discard the disc safely.



Courtesy Lou Bloomfield

Why did the microwaves heat the thin metal layer in the disc? What happened to the temperature of the plastic disc as the metal layer heated up? Which expands faster with increasing temperature: the metal or the plastic?

Why did the metal layer tear? Once the metal layer had fragmented into many sharp islands, why did sparks jump between those islands?

Chapter Itinerary

In this chapter, we'll discuss how electromagnetic waves are formed and detected in two common appliances: (1) *radio* and (2) *microwave ovens*. In Radio, we examine the ways in which charge moving in an antenna can emit or respond to electromagnetic waves and how those waves can be controlled to send sound information from a radio transmitter to a radio receiver. In Microwave Ovens, we see how microwaves affect water molecules and metals, and also how they are produced in the oven's magnetron tube. Although we can't see the electromagnetic waves that these two devices use, these waves clearly play important roles in our world. For more about what we'll study, turn to the Chapter Summary and Important Laws and Equations at the end of the chapter.

SECTION 12.1 ↓ Radio

This section describes how radio works. We'll look at how radio waves are transmitted and how they're received. We'll also examine the common ways in which sound is represented by radio waves so that it can travel through space to a receiver far away.

Questions to Think About: How might the movement of electric charge in one metal antenna affect electric charge in a second antenna nearby? What about when the second antenna is far away from the first antenna? What does it mean when a radio station claims to transmit 50,000 W? How does your radio select one channel from among all the possibilities?

Experiments to Do: Listen to a small AM radio and notice that the volume of the sound depends on the radio's orientation or location. Radio waves are pushing electric charges back and forth along the radio's hidden internal antenna. You can sometimes find an orientation in which the radio is silent because in that orientation the radio waves are unable to move charges along the antenna. If you put the radio inside a metal box, it will also become silent. Can you explain why?

You can try similar experiments with a cordless telephone—actually a radio transmitter and receiver. See how far you can go with the handset before you lose contact with the base unit. Notice that the antenna's size and orientation affect its range. What happens to the reception if you stand behind a large metal object?

A fluctuating electric current can represent sound information, and that's what it does each time you speak into a microphone or listen to music through earphones. Currents need wires, however, so how can we send sound information to someone who is moving? We need a way to represent sound that doesn't involve wires. We need radio.

A Prelude to Radio Waves

Before we can examine radio and radio waves, let's take a moment to finish the introduction to electrodynamics that we began in Chapters 10 and 11. Although we've already learned most of the fundamental relationships between electricity and magnetism, the remaining one is about to become important. To refresh your memory, we have observed so far that electric fields can be produced by electric charges, subatomic particles, and changing magnetic fields and that magnetic fields can be produced by subatomic particles and

TABLE 12.1.1 Sources of Electric and Magnetic Fields

Sources of Electric Fields	Sources of Magnetic Fields
Electric charges and subatomic particles	Magnetic poles and subatomic particles
Moving magnetic pole	Moving electric charge
Changing magnetic fields	Changing electric fields

moving electric charges (Table 12.1.1). If isolated magnetic poles exist, they produce magnetic fields and, when moving, electric fields.

In 1865, Scottish physicist James Clerk Maxwell (1831–1879) discovered one additional source of magnetic fields: *changing electric fields*. That effect is subtle and scientists overlooked it for most of the nineteenth century. It wasn't until Maxwell was trying to formulate a complete electromagnetic theory that he uncovered this additional connection between electricity and magnetism. This final relationship completed the set shown in Table 12.1.1. Taken together, these relationships allowed Maxwell to understand one of the most remarkable phenomena in nature—electromagnetic waves!

THIRD CONNECTION BETWEEN ELECTRICITY AND MAGNETISM

Electric fields that change with time produce magnetic fields.

Since electric fields can create magnetic fields and magnetic fields contain energy, it's clear that electric fields must contain energy, too. The amount of energy in a uniform electric field is the square of the field strength times the volume of the field divided by 8π times the Coulomb constant. We can write this relationship as a word equation:

$$\text{energy} = \frac{\text{electric field}^2 \cdot \text{volume}}{8\pi \cdot \text{Coulomb constant}}, \quad (12.1.1)$$

in symbols:

$$U = \frac{E^2 \cdot V}{8\pi \cdot k},$$

and in everyday language:

When you charge a large capacitor, it stores a great deal of energy in the electric field between its plates.

With these observations, we have finished the prelude and are ready to see how radio works.



Check Your Understanding #1: A Real Flux Capacitor

When a capacitor has separated charge on its plates, there is a strong electric field between those plates. Connecting the plates with a wire will discharge the capacitor, and its electric field will suddenly vanish. As the electric field disappears, what other field is present between the plates?

Answer: There is now a magnetic field between the plates.

Why: Since a changing electric field produces a magnetic field, the plates have a magnetic field between them while their electric field is disappearing.

Check Your Figures #1: Lightning in the Fields

During a thunderstorm, the charged clouds produce an electric field of about 10,000 V/m near the ground. How much energy is contained in 1.0 cubic meter of that electric field?

Answer: The electric field in 1.0 m³ is about 0.00045 J.

Why: Since 10,000 V/m is equivalent to 10,000 J/C · m, Eq. 12.1.1 gives us the energy of a 10,000 V/m field occupying 1.0 m³ as:

$$\begin{aligned} \text{energy} &= \frac{(10,000 \text{ J/C} \cdot \text{m})^2 \cdot 1.0 \text{ m}^3}{8\pi \cdot 8.988 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2} \\ &= \frac{0.00045 \text{ J}^2}{\text{N} \cdot \text{m}} = 0.00045 \text{ J}. \end{aligned}$$

While that isn't much energy per cubic meter, a single lightning strike may release the electric field energy in almost a billion cubic meters. No wonder it produces such a bang!

Antennas and Tank Circuits

A radio transmitter communicates with a receiver via radio waves. These waves are produced by electric charge as it moves up and down the transmitter's antenna and are detected when they push electric charge up and down the receiver's antenna. What exactly are radio waves, and how does charge on the antenna produce them?

We've already seen that electric charge produces electric fields and that moving charge produces magnetic fields. However, something new happens when charge *accelerates*. Accelerating charge produces a mixture of changing electric and magnetic fields that can reproduce one another endlessly and travel long distances through empty space. These interwoven electric and magnetic fields are known generally as **electromagnetic waves**. In the case of radio, the electromagnetic waves have low frequencies and long wavelengths, and are known as **radio waves**.

Before we look at the structure of a radio wave and at how it travels through space, let's start with a much simpler situation. We'll look at how two nearby metal antennas affect one another. Figure 12.1.1 shows a radio transmitter and a radio receiver, side by side. Because of their proximity, electric charge on the transmitter's antenna is sure to affect charge on the receiver's antenna.

To communicate with the nearby receiver, the transmitter sends charge up and down its antenna. This charge's electric field surrounds the transmitting antenna and extends all the way to the receiving antenna, where it pushes charge down and up. Unfortunately, the resulting charge motion in the receiving antenna is weak and the receiver may have difficulty distinguishing it from random thermal motion or from motion caused by other electric fields in the environment. Therefore the transmitter adopts a clever strategy—it moves charge up and down its antenna rhythmically at a particular frequency. Since the resulting motion on the receiving antenna is rhythmic at that same frequency, it's much easier for the receiver to distinguish from unrelated motion.

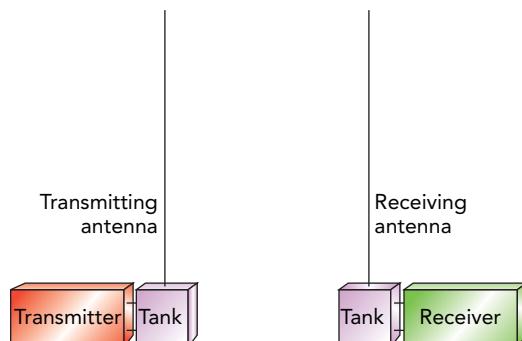


Fig. 12.1.1 Electric charge rushing on and off the transmitting antenna causes a similar motion of electric charge in the receiving antenna.

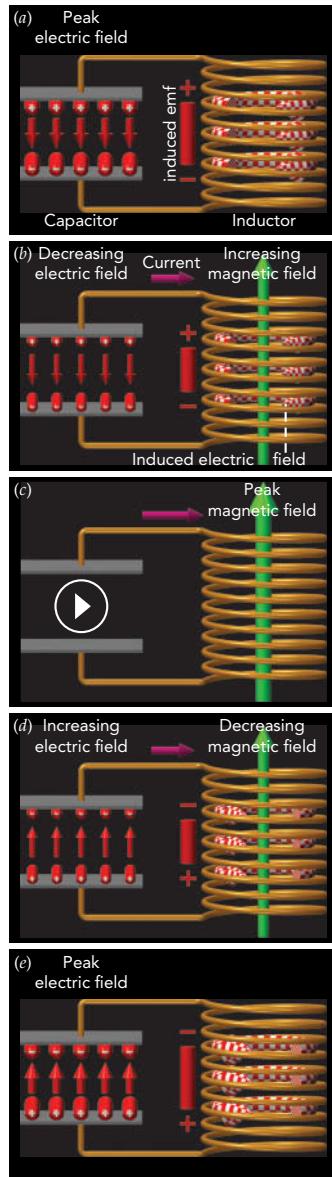


Fig. 12.1.2 A tank circuit is an electromagnetic harmonic oscillator consisting of a capacitor and an inductor. This sequence shows half a cycle of its oscillation at five equally spaced moments in time. Energy sloshes rhythmically back and forth between the capacitor's electric field and the inductor's magnetic field.

Using this rhythmic motion has another advantage: it allows the transmitter and receiver to use **tank circuits**, resonant electronic devices consisting only of capacitors and inductors (Fig. 12.1.2). Charge can “slosh” back and forth through a tank circuit at a particular frequency, much as water can slosh back and forth in a water storage tank at a particular frequency (see Fig. 9.3.4). Just as you can get the water sloshing strongly by giving it gentle pushes that are synchronized with its rhythmic motion, so the transmitter can get charge sloshing strongly through its tank circuit by giving that charge gentle pushes that are synchronized with its rhythmic motion. Both are examples of resonant energy transfer (see Section 9.2). By helping the transmitter move larger amounts of charge up and down the antenna, the tank circuit dramatically strengthens the transmission.

A second tank circuit attached to the receiving antenna helps the receiver detect this transmission. Gentle, rhythmic pushes by fields from the transmitting antenna cause more and more charge to move through the receiving antenna and its attached tank circuit. While the motion of charge on this antenna alone may be difficult to detect, the much larger charge sloshing in the tank circuit is unmistakable.

We can understand how a tank circuit works by looking at how charge moves between its capacitor and its inductor. Let’s imagine that the tank circuit starts out with separated charge on the plates of its capacitor (Fig. 12.1.2a). Since the inductor conducts electricity, current begins to flow from the positively charged plate, through the inductor, to the negatively charged plate. The current through the inductor must rise slowly and, as it does, it creates a magnetic field in the inductor (Fig. 12.1.2b).

Soon the capacitor’s separated charge is gone and all the tank circuit’s energy is stored in the inductor’s magnetic field (Fig. 12.1.2c). However, the current keeps flowing, driven forward by the inductor’s opposition to current changes. The inductor uses the energy in its magnetic field to keep the current flowing, and separated charge reappears in the capacitor (Fig. 12.1.2d). Eventually, the inductor’s magnetic field decreases to zero and everything is back to its original state—almost. While all the tank circuit’s energy has returned to the capacitor, the separated charge in that capacitor is now upside down (Fig. 12.1.2e).

This whole process now repeats in reverse. The current flows backward through the inductor, magnetizing it upside down, and the tank circuit soon returns to its original state. This cycle repeats over and over again, with charge sloshing from one side of the capacitor to the other and back again.

A tank circuit is an electronic harmonic oscillator, equivalent to the mechanical harmonic oscillators we examined in Chapter 9. Like all harmonic oscillators, its period (the time per cycle) doesn’t depend on the amplitude of its oscillation. Thus, no matter how much charge is sloshing in the tank circuit, the time it takes that charge to flow over and back is always the same.

The tank circuit’s period depends only on its capacitor and its inductor. The larger the capacitor’s capacitance, the more separated charge it can hold with a given amount of energy and the longer it takes that charge to move through the circuit as current. The larger the inductor’s **inductance**, its opposition to current changes, the longer that current takes to start and stop. A tank circuit with a large capacitor and a large inductor may have a period of a thousandth of a second or more, while one with a small capacitor and a small inductor may have a period of a billionth of a second or less.

Inductance is defined as the voltage drop across the inductor divided by the rate at which current through the inductor changes with time. This division gives inductance the units of voltage divided by current per time. The SI unit of inductance is the volt-second per ampere, also called the **henry** (abbreviated H). While large electromagnets have inductances of hundreds of henries, a $1\text{-}\mu\text{H}$ (0.000001-H) inductor is more common in radio.

Its resonant behavior makes the tank circuit useful in radio. That’s because small, rhythmic pushes on the current in a tank circuit can lead to enormous charge oscillations in that circuit. In radio, these rhythmic pushes begin when the transmitter sends an alternating current through a coil of wire. Fields from this coil push current back and forth through the nearby transmitting tank circuit, causing enormous amounts of charge to slosh back and

forth in it and travel up and down the transmitting antenna. That charge's electric field then pushes rhythmically on charge in the receiving antenna, causing substantial amounts of charge to travel down and up it and slosh back and forth in the receiving tank circuit. The receiver can easily detect this sloshing charge.

Energy flows from the transmitter to the receiver via resonant energy transfer: from the transmitter, to the transmitting tank circuit and antenna, to the receiving antenna and tank circuit, and finally to the receiver. This sequence of transfers can work efficiently only if all the parts have resonances at the same frequency. Tuning a radio receiver to a particular station is largely a matter of adjusting its capacitor and inductor so that its tank circuit has the right resonant frequency.

Check Your Understanding #2: No Tanks

Why doesn't the radio transmitter simply push electric charge directly on and off the antenna, without using a tank circuit?

Answer: The amount of charge that the transmitter can move directly on and off the antenna is too small to create a strong radio wave.

Why: The tank circuit is useful because it allows the transmitter to move much more charge. Just as a tuning fork is inefficient at emitting sound waves by itself, so a radio antenna is inefficient at emitting radio waves by itself. You can make the tuning fork much louder by coupling it to an object that resonates at its frequency. Similarly, you can make the radio antenna emit a much stronger radio wave by coupling it to a tank circuit that resonates at its frequency.

Radio Waves

When the two antennas are close together, charge in the transmitting antenna exerts electrostatic force directly on charge in the receiving antenna. However, when the antennas are far apart, the interactions between them are more complicated. Charge in the transmitting antenna must then emit a radio wave to push on charge in the receiving antenna. Like a water wave, a radio wave is a disturbance that carries energy from one place to another. But unlike a water wave, which must travel in a fluid, a radio wave can travel through otherwise empty space, from one side of the universe to the other.

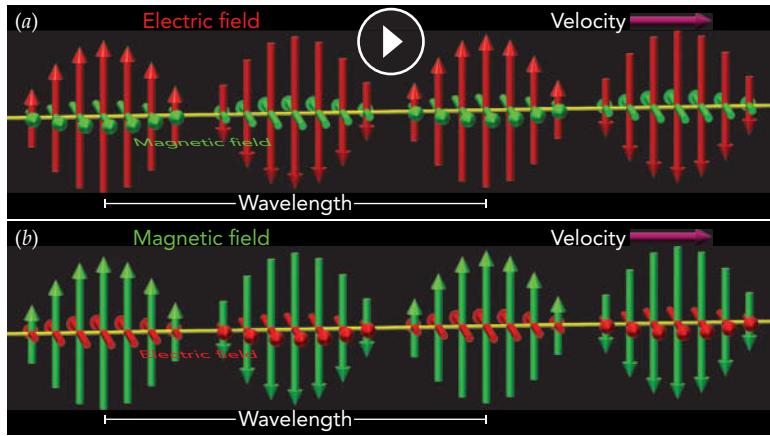
Like all electromagnetic waves, a radio wave consists only of a changing electric field and a changing magnetic field. These fields re-create one another over and over again as the wave travels through empty space at the speed of light—exactly 299,792,458 m/s (approximately 186,282 miles per second).

The radio wave is created when electric charge in the antenna accelerates. Whereas stationary charge or a steady current produces constant electric or magnetic fields, accelerating charge produces fields that change with time. As charge flows up and down the antenna, its electric field points in alternating directions vertically, and its magnetic field points in alternating directions horizontally. These changing fields then re-create one another again and again, and sail off through space as an electromagnetic wave. At each point along its path, the wave's electric field direction, magnetic field direction, and direction of travel are mutually perpendicular.

The wave emitted by a vertical transmitting antenna has a **vertical polarization**, that is, its electric field points alternately up and down (Fig. 12.1.3a). We identify those “ups” as crests, and the distance between adjacent crests is its wavelength. For radio waves, that wavelength is usually 1 m (3.3 ft) or more. The wave's magnetic field is perpendicular to its electric field and points in alternating directions horizontally.

Had the transmitting antenna been tipped on its side, the wave's electric field would have pointed in alternating directions horizontally and the wave would have had a **horizontal polarization** (Fig. 12.1.4b). The wave's magnetic field would then point alternately up and down. Whatever the polarization, the electric and magnetic fields move forward together as a traveling wave, so the pattern of fields moves smoothly through space at the speed of light.

Fig. 12.1.3 Electromagnetic waves traveling toward the right at the speed of light. The red arrows represent electric fields at points on the yellow path; the green arrows represent magnetic fields at those same points. At each point, the wave's electric field, magnetic field, and direction of travel are mutually perpendicular. (a) In a vertically polarized electromagnetic wave, the electric field is directed vertically and the magnetic field is directed horizontally. (b) In a horizontally polarized electromagnetic wave, the electric field is directed horizontally and the magnetic field is directed vertically.



COMMON MISCONCEPTIONS: Electromagnetic Waves and Undulations

Misconception: Since the fields of an electromagnetic wave appear wavy (Fig. 12.1.3), the light wave itself undulates; it actually undulates up and down or back and forth as it heads rightward!

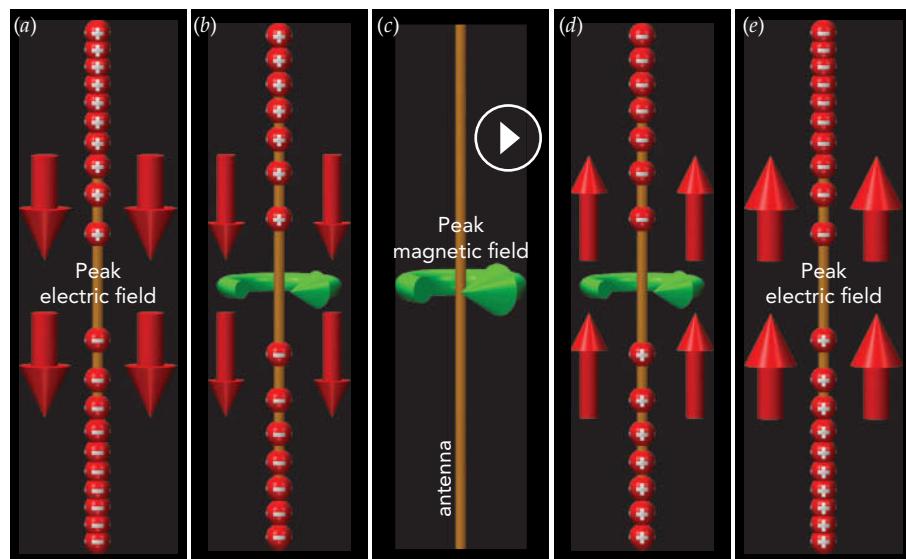
Resolution: The arrows drawn to represent the fields in an electromagnetic wave are associated with points along the straight path of the wave. Each wave in Fig. 12.1.3 is heading directly rightward along the yellow path and the arrows indicate field values at points along that path.

If you stood in one place and could watch this wave pass, you'd notice its electric field fluctuating up and down at the same frequency as the charge that created it. When the wave passes a distant receiving antenna, it pushes charge up and down that antenna at this frequency. If the receiving tank circuit is resonant at this frequency, the amount of charge sloshing in it should become large enough for the receiver to detect.

A radio station can optimize its transmission by using a resonant transmitting antenna. A straight antenna is another electronic harmonic oscillator, with a period that depends only on its length. When that length is half the wavelength of the radio wave it's transmitting, charge sloshes up and down the antenna in a natural resonance at the frequency of the radio wave.

Surprisingly, the antenna is actually a tank circuit (Fig. 12.1.4)—its tips act as the plates of a capacitor and its middle acts as the inductor. Despite its linear shape, the antenna has the

Fig. 12.1.4 A straight antenna is a linear tank circuit, with ends that act as capacitor plates and a middle that acts as an inductor. This sequence shows half a cycle of its oscillation at five equally spaced moments in time; compare a–e to Fig. 12.1.2a–e. Charge oscillates up and down the antenna and the antenna's energy alternates between its electric field and its magnetic field.



same resonant behavior as the coiled tank circuit shown in Fig. 12.1.2. When the transmitting tank circuit and antenna are resonant at the same frequency, there's a resonant energy transfer from one to the other. These resonant effects help to produce a powerful radio wave.

Because it's half a wavelength long and its ends are oppositely charged, this antenna is known as a *halfwave dipole antenna*. Many radio stations use such antennas. It's possible, however, to omit the bottom half of the dipole antenna by placing the top half above an electrically conducting surface. The conducting surface is a mirror (Fig. 12.1.5) and its reflection of the top half acts like the missing bottom half of the dipole antenna. Known as a *quarter-wave monopole antenna*, this shorter antenna is often more convenient, particularly when the antenna projects upward from a metal surface or the ground.

The transmitting antenna sends the strongest portion of its radio wave out perpendicular to its length. That's not unexpected because the motion of charge on the antenna is most obvious when viewed from a line perpendicular to its length. Thus, a vertical antenna sends most of its wave out horizontally, where people are likely to receive it. No wave emerges from the end of an antenna.

Both electric and magnetic fields contain energy, so as the electromagnetic wave travels through space, it carries energy away from the transmitter. When a radio station advertises that it "transmits 50,000 W of music," it's claiming that its antenna emits 50,000 J of energy per second or 50,000 W of power in its electromagnetic wave. The receiving antenna must absorb enough of this power to detect the wave. However, the farther the wave gets from the transmitting antenna, the more spread out and weaker it becomes. Trees and mountains also absorb or reflect some of the wave and hinder reception.

For the best reception, a listener should be located where the radio wave is strong and where there's an unobstructed path from the transmitting antenna to the receiving antenna. To be resonant, the receiving antenna should be a half-wave dipole or a quarter-wave monopole, and it should be oriented along the radio wave's polarization: vertical for a vertically polarized radio wave or horizontal for a horizontally polarized radio wave. Aligning the receiving antenna with the wave's polarization makes certain that the wave's electric field pushes charge *along* the antenna, not *across* it.

To ensure good reception regardless of receiving antenna orientation, many radio stations transmit a complicated *circularly polarized* wave that combines both vertical and horizontal polarizations. To form this wave, they need several antennas. For wavelengths under a few meters, these antennas can all be attached inexpensively to a single mast. That's why commercial FM and TV broadcasts, which use short-wavelength radio waves, are usually transmitted with circular polarization. However, commercial AM broadcasts, which use long-wavelength radio waves, are transmitted only with vertical polarization.

Because commercial FM radio waves usually include both polarizations, FM receiving antennas can be vertical or horizontal. Portable FM receivers often use vertical telescoping antennas, while home receivers frequently use horizontal wire antennas. All these antennas are approximately half-wave dipole antennas or quarter-wave monopole antennas.

A quarter-wave monopole antenna for commercial AM radio would have to be about 100 m (330 ft) long, so straight AM antennas (such as those on cars) are much shorter than optimal. That's why many AM antennas are designed to respond to the radio wave's horizontal magnetic field rather than to its vertical electric field. These magnetic antennas are horizontal coils of wire that experience induced currents when exposed to fluctuating magnetic fields.

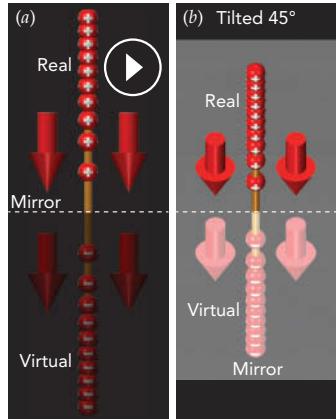


Fig. 12.1.5 (a) Side view of a quarter-wave monopole antenna above a conducting surface or mirror. The mirror reflects a virtual image of the antenna, so the system acts like a half-wave dipole antenna. (b) The same monopole antenna and its reflection viewed from a 45° angle above the mirror.

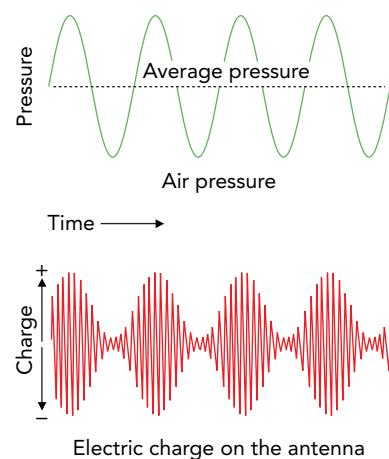
Check Your Understanding #3: There's No Place Like Home

Why do cordless telephones work only when they're close to their base units?

Answer: When the cordless telephone is too far from its base unit, the electromagnetic waves become so spread out that they have trouble communicating.

Why: The powers emitted by the base unit and the handset are small, so their waves are relatively difficult to detect. As long as the handset and base unit are close, they are able to detect each other's waves. When the distance between them becomes too great, the waves become too spread out to detect and the handset and base unit lose contact with one another.

Fig. 12.1.6 When sound is transmitted using amplitude modulation, air pressure is represented by the strength of the radio wave. A compression is represented by strengthening the radio wave and a rarefaction is represented by weakening it.



Representing Sound: AM and FM Radio

A radio transmitter does more than simply emit a radio wave. It uses that radio wave to represent sound. Because sound waves are fluctuations in air density and radio waves are fluctuations in electric and magnetic fields, a radio wave can't literally "carry" a sound wave. However, a radio wave can carry sound information and instruct the receiver how to reproduce the sound.

To convey sound information, the radio station alters its radio wave to represent compressions and rarefactions of the air. The receiver then recreates those compressions and rarefactions. There are two common techniques by which a radio wave can represent those density fluctuations. One is called amplitude modulation and involves changing the overall strength of the radio wave. The other is called frequency modulation and involves small changes in the frequency of the radio wave.

In the **amplitude modulation** (AM) technique, air density is represented by the strength of the transmitted wave (Fig. 12.1.6). To represent a compression of the air, the transmitter is turned up so that more charge moves up and down the transmitting antenna. To represent a rarefaction, the transmitter is turned down so that less charge moves up and down the antenna. The frequency at which charge moves up and down the antenna remains steady, so only the amplitude of the radio wave changes. The receiver measures the strength of the radio wave and uses this measurement to re-create the sound. When it detects a strong radio wave, it pushes its speaker toward the listener and compresses the air. When it detects a weak radio wave, it pulls its speaker away from the listener and rarefies the air.

In the **frequency modulation** (FM) technique, air density is represented by the frequency of the transmitted wave (Fig. 12.1.7). To represent a compression of the air, the

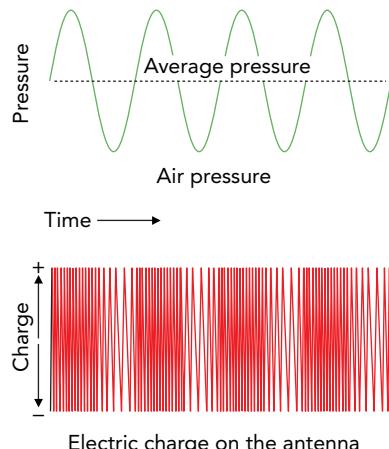


Fig. 12.1.7 When sound is transmitted by frequency modulation, air pressure is represented by changing the frequency of the radio transmitter slightly. A compression is represented by increasing that frequency and a rarefaction by decreasing it.

transmitter's frequency is increased slightly so that charge moves up and down the transmitting antenna a little *more* often than normal. To represent a rarefaction, the transmitter's frequency is decreased slightly so that the charge moves up and down a little *less* often than normal. These changes in frequency are extremely small—so small that charge continues to slosh strongly in all the resonant components and reception is unaffected. The receiver measures the radio wave's frequency and uses this measurement to re-create the sound. When it detects an increased frequency, it compresses the air, and when it detects a decreased frequency, it rarefies the air.

Although the AM and FM techniques for representing sound can be used with a radio wave at any frequency, the most common commercial bands in the United States are the AM band between 550 kHz and 1600 kHz (550,000 Hz and 1,600,000 Hz) and the FM band between 88 MHz and 108 MHz (88,000,000 Hz and 108,000,000 Hz). Elsewhere in the spectrum of radio frequencies are many other commercial, military, and public transmissions, including TV, shortwave, amateur radio, telephone, data, police, and aircraft bands. These other transmissions use AM, FM, and a few other techniques to represent sound and information with radio waves.

Check Your Understanding #4: Another Volume Control

When you are listening to the AM radio in a car and drive through a tunnel, the volume becomes very low. Explain.

Answer: The tunnel blocks most of the radio wave. Since only a small fluctuating wave reaches your radio, the radio produces only small fluctuations in air density with its speaker.

Why: An AM radio has trouble distinguishing between a distant transmission representing loud music and a nearby transmission representing soft music. In both cases, the receiver detects only small variations in the current moving up and down its antenna. That's why you must turn up the volume of an AM radio as you move farther from the transmitting antenna or as you enter a tunnel.

Bandwidth and Cable

A pure, single-frequency radio wave doesn't carry any information. To represent sound, video, or any other form of information, the radio wave must vary with time. Think of smoke signals—a steady stream of smoke carries no information, but carefully timed puffs of smoke can send a message.

Once a radio wave is varying with time to carry information, it no longer has a single pure frequency. Regardless of which aspects of the radio wave are varying, that wave now includes a range of frequencies. The more information the radio wave carries each second, the broader that range of frequencies becomes.

When it's representing sound, a radio wave has a range of radio frequencies that stretches from somewhat below the official frequency of the radio wave, the *carrier frequency*, to somewhat above that frequency. The wider the audio frequency range of the sound, the more sound information must be sent each second and the broader the range of radio frequencies needed to represent that sound. The range of frequencies needed to transmit such a stream of information is known as the transmission's **bandwidth**.

By international agreement, an AM radio station may use 10 kHz of bandwidth, 5 kHz above and below its carrier frequency. To stay within that bandwidth, the sound being represented can't contain frequencies above 5 kHz. Although this restricted frequency range is bad for music, it allows competing stations to function with carrier frequencies only 10 kHz apart, so 106 different stations can operate between 550 kHz and 1600 kHz.

An FM radio station may use 200 kHz of bandwidth, 100 kHz on each side of its carrier frequency. This luxurious allocation permits FM radio to represent a very broad range of audio frequencies, in stereo, which is why an FM radio station can do a much better job of sending music to your radio than an AM station can. In recent years, FM stations have

begun using their 200-kHz bandwidths to carry digital information representing several programs of “high-definition” sound. (We’ll examine digital audio in Chapter 14.)

Because high-frequency radio waves travel in straight lines between antennas, it’s hard to receive a commercial FM station from more than about 100 km (60 mi) away. Even when the transmitting antenna sits on top of a tall tower, Earth’s curvature and surface terrain severely limit the range of FM reception.

Low-frequency radio waves, such as those used by commercial AM stations, are reflected by charged particles in Earth’s outer atmosphere, so portions of the radio wave that would otherwise be lost to space bounce back toward the ground. This returning power allows you to receive AM stations over a considerable distance, even when you have no direct line of sight to the transmitter’s antenna. At sundown, these atmospheric layers become so effective at reflecting AM radio that you can hear a transmission from thousands of kilometers away as clearly as if it were a hometown station.

The spectrum of electromagnetic waves is a limited resource, and if it could only be used once, it would quickly run out of bandwidth. Fortunately, distance and enclosures make it possible to reuse the spectrum many times. Cell phones that are far from one another can share the same carrier frequencies and bandwidth because their radio waves weaken with distance and essentially don’t overlap. However, even nearby radio transmissions can use the same carrier frequencies by enclosing their electromagnetic waves inside cables.

Cable radio, television, and data networks are similar to broadcast networks except that they send electromagnetic waves through cables rather than through empty space. A typical radio or television cable consists of an insulated metal wire inside a tube of metal foil or woven metal mesh. This wire-inside-a-tube arrangement is called *coaxial cable* because its two metal components share the same centerline or axis. In contrast, a typical computer-data cable consists of a number of insulated metal wires that are twisted into several pairs.

Electromagnetic waves can propagate easily through a coaxial or twisted-pair cable, following its twists and turns from the transmitter that produces the waves to the receiver that uses them. The fact that wires are assisting these waves in their travels makes them more complicated than waves in empty space. However, they still involve electric and magnetic fields and still propagate forward at nearly the speed of light.

Because the electromagnetic waves inside a cable don’t interact with those outside it, the transmitter and receiver can use whatever parts of the spectrum they choose, without concerns about sharing. A typical coaxial cable can handle frequencies up to about 1000 MHz and typical twisted-pair cable can reach 350 MHz, so either one can carry a great deal of information each second.

However, coaxial cables must now compete with optical fiber cables that guide light from one place to another. We’ll examine optical fibers in Section 14.2. Like radio waves, light is an electromagnetic wave and can be amplitude or frequency modulated to represent information. However, light’s frequency is extremely high; the frequencies of visible light range from 4.5×10^{14} to 7.5×10^{14} Hz. If we were to allocate FM radio channels 200 kHz apart throughout the visible spectrum, there would be about 1.5 billion channels available!



Check Your Understanding #5: Beaten by the Bandwidth

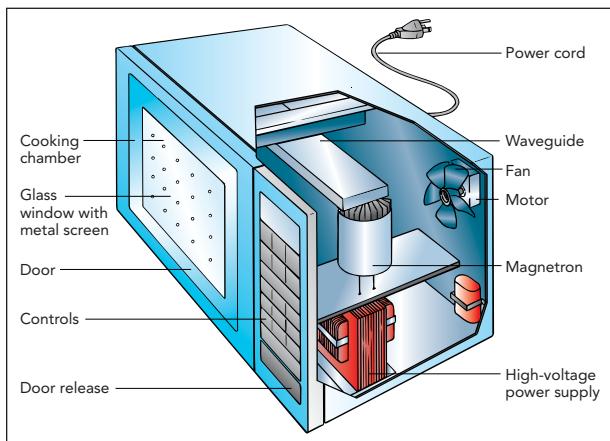
You’re playing piano for an AM radio station, and you strike the highest key on the piano. The pitch of the resulting sound is 4186 Hz. Can the listeners hear that note from their radios?

Answer: Yes, they can hear it, at least in principle.

Why: The bandwidth of an AM radio station extends to 5000 Hz above and below its carrier frequency, so the station is officially permitted to represent sound frequencies as high as 5000 Hz. However, in practice the station probably begins to filter out sounds well below that frequency to avoid accidentally violating their license.

SECTION 12.2

Microwave Ovens



In addition to carrying sounds from one place to another, electromagnetic waves can carry power. One interesting example of such power transfer is a microwave oven. It uses relatively high-frequency electromagnetic waves to transfer power directly to the water molecules in food so that the food cooks from the inside out. This section discusses both how those waves are created and why they heat food.

Questions to Think About: Why do microwave ovens tend to cook food unevenly if you don't move the food during cooking? How can part of a frozen meal become boiling hot while another part remains frozen? Why must you be careful with metal objects placed inside the oven? Why do some objects remain cool in the microwave oven, while other objects become extremely hot? How does microwave popcorn work?

Experiments to Do: A microwave oven transfers power primarily to the water in food. You can see this effect by placing completely water-free food ingredients such as salt, baking powder, sugar, or salad oil on a microwave-safe ceramic dish in a microwave oven. Cook the ingredients briefly. You will find that the ingredients and dish remain relatively cool. Add just a little water to the collection. What happens when you cook them this time?

Now try cooking a very cold ice cube. The cube should come directly from the freezer on an ice-cold plate, so that its surface is solid and dry. What happens? If ice contains water and water is what absorbs power in a microwave oven, why doesn't the ice absorb power and melt?

Microwaves and Food

When studying thermal radiation in Section 7.3, we discussed the *wavelengths* of electromagnetic waves. While examining radio, we concentrated on the *frequencies* of electromagnetic waves. However, we know from Eq. 9.2.1 that the wavelength and frequency of a wave aren't independent. A basic electromagnetic wave in empty space has both a wavelength and a frequency, and their product is the speed of light. That relationship can be written as a word equation:

$$\text{speed of light} = \text{wavelength} \cdot \text{frequency} \quad (12.2.1)$$

in symbols:

$$c = \lambda \cdot \nu,$$

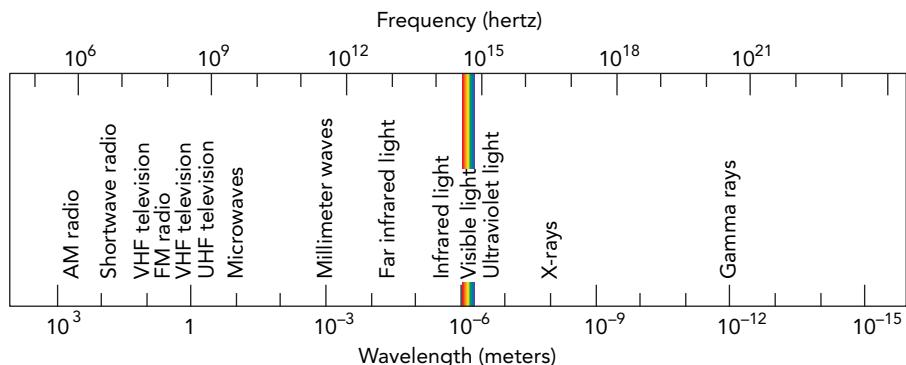
and in everyday language:

The higher the frequency of an electromagnetic wave, the shorter its wavelength becomes.

Like Fig. 7.3.2, Fig. 12.2.1 shows the approximate wavelengths of many types of electromagnetic waves, but it also shows their frequencies.

Radio broadcasts use the low-frequency, long-wavelength portion of the electromagnetic spectrum. Commercial AM radio broadcasts at frequencies of 550 to 1600 kHz (wavelengths of 545 to 187 m) and commercial FM radio broadcasts at frequencies of 88 to 108 MHz (wavelengths of 3.4 to 2.8 m). As long as their wavelengths are 1 m (3.3 ft) or longer, electromagnetic waves are called *radio waves*. Electromagnetic waves that have

Fig. 12.2.1 The electromagnetic spectrum. Microwaves have wavelengths between about 1 m and 1 mm, corresponding to frequencies from 300 MHz up to 300 GHz.



Although he was orphaned as a child and never completed grade school, American Percy Lebaron Spencer (1894–1970) had a brilliant career as a scientist and microwave engineer. In 1945, while visiting a magnetron testing laboratory, he leaned over an operating magnetron and the candy bar in his shirt pocket melted. Immediately recognizing what had happened, he soon had popcorn popping about the lab and even cooked an egg until it exploded. Cooking has never been the same since.

wavelengths of 1 mm or longer, but less than 1 m, are called **microwaves**. Microwave ovens usually cook food with 0.122-m electromagnetic waves, so their name is appropriate.

To explain how a microwave oven heats food 1, let's begin by looking at water molecules. Water molecules are electrically polarized—that is, they have positively charged ends and negatively charged ends. This polarization comes about because of quantum physics and the tendency of oxygen atoms to pull electrons away from hydrogen atoms. The water molecule is bent, with its two hydrogen atoms sticking up from its oxygen atom like Mickey Mouse's ears. When the oxygen atom pulls the electrons partly away from the hydrogen atoms, its side of the molecule becomes negatively charged, while the hydrogen atoms' side becomes positively charged. Water is thus a polar molecule.

In ice, these polar water molecules are arranged in an orderly fashion with fixed positions and orientations. However, in liquid water, the molecules are more randomly oriented (Fig. 12.2.2). Their arrangements are constrained only by their tendency to bind together, positive end to negative end, to form a dense network of coupled molecules. This binding between the positively charged hydrogen atom on one water molecule and the negatively charged oxygen atom on another molecule is known as a *hydrogen bond*.

If you place liquid water in a strong electric field, its water molecules will tend to rotate into alignment with the field. That's because a misaligned molecule has extra electrostatic potential energy and accelerates in the direction that reduces its potential energy as quickly as possible. In this case, the water molecule will experience a torque and will undergo an angular acceleration that makes it rotate into alignment. As it rotates, the molecule will bump into other molecules and convert some of its electrostatic potential energy into thermal energy.

A similar effect occurs at a crowded party when everyone is suddenly told to face the front of the room. People brush against one another as they turn, and sliding friction converts some of their energy into thermal energy. If the people are told to turn back and forth repeatedly, they will become quite warm. The same holds true for water. If the electric field reverses its direction many times, the water molecules will turn back and forth and become hotter and hotter.

A microwave's fluctuating electric field is well suited to heating water. A microwave oven uses 2.45-GHz (2.45-gigahertz or 2,450,000,000-Hz) microwaves to twist the food's water molecules back and forth billions of times per second. As the water molecules turn, they bump into one another and heat up. The water absorbs the microwaves and converts their energy into thermal energy. This particular microwave frequency was chosen not because of any resonant effect but because it was not in use for communications and because it cooks food uniformly. If the frequency were higher, the microwaves would be absorbed too strongly by food and wouldn't penetrate deeply into large items. If the frequency were lower, the microwaves would pass through food too easily and wouldn't cook it efficiently.

This twisting effect explains why only foods or objects containing water or other polar molecules cook well in a microwave oven. Microwave-safe ceramic plates, glass cups, and plastic containers are water-free and usually remain cool. Even ice has trouble absorbing microwave power because its crystal structure constrains the water molecules so they can't turn easily.

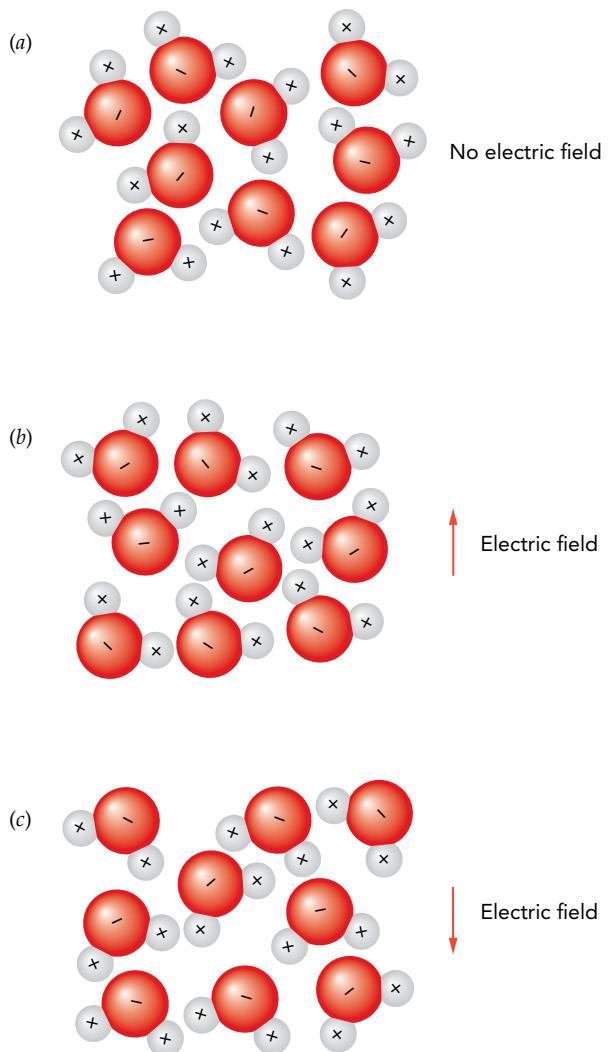


Fig. 12.2.2 (a) The water molecules in liquid water are randomly oriented when there's no electric field. (b, c) But an electric field tends to orient them with their positive ends in the direction of the field.

Although ice melts slowly in a microwave oven, the liquid water it produces heats quickly. This peculiar heating behavior explains why it's so easy to burn yourself on frozen food heated in a microwave oven. The portions of the food that defrost first absorb most of the microwave power and overheat, while the rest of the food remains frozen solid. You never know whether your next bite will break your teeth or sear the roof of your mouth. To address this problem, many microwave ovens have defrost cycles in which microwave heating is interrupted periodically to let heat flow naturally through the food to melt the ice. Once the frozen parts have melted, all the food can absorb microwaves.

► Check Your Understanding #1: Microwave Popcorn

A popcorn kernel contains moist starch trapped inside a hard, dry hull. You can scorch this hull by cooking the corn in hot oil but not when you cook it in a microwave oven. How can the microwave oven pop the corn without risk of overheating the hull?

Answer: The microwave oven transfers heat to water molecules in the starch so that the hull never becomes hotter than the material inside it.

Why: A corn kernel cooked in oil is heated by contact with the hot oil and the pot. You can easily overheat the outer hull and burn it. However, microwaves transfer heat to the water molecules inside the kernel. The hull can't overheat because the hottest thing it touches is the starchy insides of the kernel. When the pressure of steam inside the kernel becomes high enough, the hull breaks and the kernel "pops."

Check Your Figures #1: Shopping for Food

The red light used by many grocery store checkout stations to scan product codes is produced by a helium–neon laser. This light is an electromagnetic wave with a wavelength of approximately 633 nm. What is its frequency?

Answer: The frequency is about 4.74×10^{14} Hz.

Why: Since the product of frequency and wavelength for an electromagnetic wave is equal to the speed of light, its frequency is equal to the speed of light divided by its wavelength:

$$\frac{299,792,458 \text{ m/s}}{0.00000633 \text{ m}} = 4.74 \times 10^{14} \text{ Hz.}$$

Metal in a Microwave Oven

Contrary to popular lore, metal objects and microwave ovens aren't always incompatible. In fact, the walls of the oven's cooking chamber are metal, yet they cause no trouble when exposed to microwaves during cooking. Like most metal surfaces, the walls reflect microwaves. They do this by acting as both receiving and transmitting antennas. Electric fields in the microwaves cause mobile charges in the metal surfaces to accelerate and absorb the original microwaves. As these charges accelerate, they emit new microwaves. The emitted microwaves have the same frequencies as the original ones, but they travel in new directions. The original microwaves have been reflected by the surface.

The cooking-chamber walls reflect the oven's microwaves and keep them bouncing around inside. Even the metal grid covering the window reflects microwaves. That's because charge has enough time during a microwave cycle to flow around each hole in the

grid and compensate for the hole's presence. As long as the wavelength of an electromagnetic wave is much larger than the holes in a metal grid, the wave reflects perfectly from that grid. In fact, if there's nothing inside the oven to absorb the microwaves, they'll bounce around inside it until they return to their source, a vacuum tube called a magnetron (Fig. 12.2.3), and eventually cause it to overheat.

While metal surfaces help confine the microwaves inside the oven, cooking your food and not you, extra metal inside the microwave can cause trouble. If you wrap food in aluminum foil, the foil will reflect the microwaves and the food won't cook. However, food placed in a shallow metal dish cooks reasonably well because microwaves enter the open top, pass through the food, reflect, and pass through the food again.

Sometimes metal's mobile charges do more than just reflect microwaves. If enough charge is pushed onto the sharp point of a metal twist-tie or scrap of aluminum foil, some of it will jump right into the air as a spark. This spark can start a fire, particularly when the twist-tie is attached to something flammable, like a plastic or paper bag. As a rule of thumb, never put a sharp metal object in the microwave oven.

Some metal objects heat up in a microwave oven. When microwaves push charge back and forth in a metal, the metal experiences an alternating current. If the metal has a substantial electrical resistance, this alternating current will experience a voltage drop and heat up the metal. While thick oven walls and cookware have low resistances and remain cool, thin metal strips quickly overheat. Metallic decorations on porcelain dinnerware are particularly susceptible to damage in a microwave oven, so warming up coffee in Grandma's gold-rimmed teacup is sure to be a disaster. When you put metal in a microwave oven, make sure that it is thick enough to conduct electricity well and that it has no sharp points.

Courtesy Lou Bloomfield

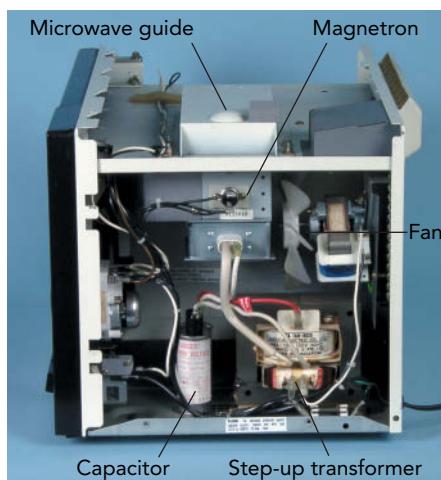


Fig. 12.2.3 This oven's magnetron microwave source is located in the middle of the picture, just to the left of its cooling fan. Microwaves travel to the cooking chamber through the rectangular metal duct on top of the oven. The high-voltage transformer at the bottom right provides power to the magnetron.

Resistive heating in conducting objects can actually be useful at times. Since microwave ovens cook food inside and outside at the same time, the food's surface never gets particularly hot, and the food doesn't brown or become crisp. To improve their textures and appearances, some foods come with special wrappers that conduct just enough current to become very hot in a microwave oven. These wrappers provide the high surface temperatures needed to brown the foods.

Another peculiar feature of microwave ovens is that they don't always cook evenly. That's because the amplitude of the microwave electric field isn't uniform throughout the oven. As the microwaves bounce around the cooking chamber, they pass through the same spot from several different directions at once. When they do, they exhibit interference effects (see Section 9.3). At one location, the individual electric fields may point in the same direction and experience constructive interference so that food there heats up quickly. At another location, however, those fields may point in opposite directions and experience destructive interference so that food there doesn't cook well at all.

If nothing is moving in the microwave oven, the pattern of microwaves inside it doesn't move either. There are then regions in which the electric field has very large amplitudes and regions in which the amplitudes are very small. The larger the amplitude of the electric field, the faster it cooks food.

To heat food uniformly in such a microwave oven, you must move the food around as it's cooking. Many ovens have turntables inside that move the food automatically. Another solution to this problem is to stir the microwaves around the oven with a rotating metal paddle. The pattern of microwaves inside the chamber changes as the paddle turns, and the food cooks more evenly. Still other microwave ovens use two separate microwave frequencies to cook the food. Because these two frequencies cook independently, it's unlikely that a portion of the food will be missed by both waves.

Check Your Understanding #2: Half and Half

You place a thick metal divider into your microwave oven so that it divides the cooking chamber exactly in half. The oven sends its microwaves into the right half of the chamber. If you put food in the left half of the chamber, will it cook?

Answer: No, it won't cook.

Why: The metal divider will reflect the microwaves and keep them from entering the left half of the oven.

Creating Microwaves with a Magnetron

Clearly, changing electric fields cook the food as microwaves bounce around the inside of an oven. But how are these microwaves created? From the previous section on radio, you might guess that the oven creates an alternating current at 2.45 GHz and that this current causes charge to slosh in a tank circuit and move up and down an antenna. That's pretty much what actually happens inside a magnetron tube.

A magnetron is a special vacuum tube—a hollow chamber from which all the air has been removed. Composed primarily of metal and ceramic parts, the magnetron uses beams of electrons to make charge slosh in a number of microwave tank circuits. These tank circuits have resonant frequencies of 2.45 GHz, the operating frequency of the oven. With the help of a tiny antenna, the magnetron emits the microwaves that cook the food.

The microwave tank circuits are arranged in a ring around the magnetron's evacuated chamber. For one of these tank circuits to oscillate naturally at 2.45 GHz, its capacitor must have an extremely small capacitance and its inductor must have an

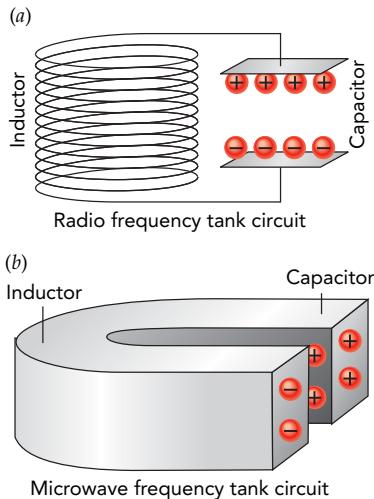


Fig. 12.2.4 (a) At radio frequencies, a tank circuit's inductor is a coil of wire and its capacitor is a pair of separated plates. (b) At microwave frequencies, a tank circuit's inductor is merely the curve of a C-shaped strip and its capacitor is the tips of that strip.

extremely small inductance. These requirements can be met by a C-shaped strip of metal (Fig. 12.2.4). Its curve is the inductor, and its tips are the capacitor.

Electric charge sloshes back and forth on the C-shaped strip just as it does in a conventional tank circuit (Fig. 12.1.2). Known as a **resonant cavity**, this strip is another electronic harmonic oscillator and therefore has a period that doesn't depend on the amount of charge that's sloshing.

The magnetron of a microwave oven typically contains eight of these resonant cavities, each carefully adjusted in size and shape so that its natural resonance occurs exactly at 2.45 GHz. Because these cavities are arranged in a ring and each one shares its tips with those of its neighboring cavities, they tend to oscillate alternately (Fig. 12.2.5). At the start of an oscillatory cycle, half the metal tips are positively charged and half are negatively charged (Fig. 12.2.5a). Currents begin to flow through the ring and produce magnetic fields in the resonant cavities (Fig. 12.2.5b). These magnetic fields propel the currents around the ring even after the charge separations have vanished. Soon the charge separations reappear but with the positive and negative tips interchanged (Fig. 12.2.5c).

Check Your Understanding #3: Large Economy Size

If a manufacturer made a magnetron that was slightly larger than normal in every dimension, how would that magnetron behave?

Answer: It would operate at a frequency below 2.45 GHz.

Why: The frequency of the microwaves emitted by a magnetron is determined exclusively by the natural resonances of its cavities. If those cavities are enlarged, both the inductances of their curves and the capacitances of their tips will increase. Their resonant frequencies will decrease, and the magnetron will emit lower-frequency microwaves.

Powering the Magnetron: The Lorentz Force

As currents oscillate back and forth around the cavities at 2.45 GHz, they fill the magnetron with alternating electric and magnetic fields. However, as the energy in these fields is extracted to cook the food or is lost to the imperfect conductivities of the cavities themselves, something must continuously replenish it. That replacement power is supplied to the cavities by four streams of energetic electrons.

At the center of the magnetron tube, surrounded only by empty space, is an electrically heated cathode that tends to emit electrons (Fig. 12.2.6a). A high-voltage power supply pumps negative charge onto this cathode so that a strong electric field points toward it from the positively charged cavity tips. If there were no other fields present in the magnetron, negatively charged electrons would emerge from the hot cathode and accelerate toward the positively charged tips as four beams of electrons (Fig. 12.2.6b).

However, the magnetron also includes a large permanent magnet. Why else would it be called a *magnetron*? This magnet creates a strong, steady magnetic field that points upward along the axis of the magnetron, parallel to the cathode itself (Fig. 12.2.6c). The purpose of this magnetic field is to alter the motions of

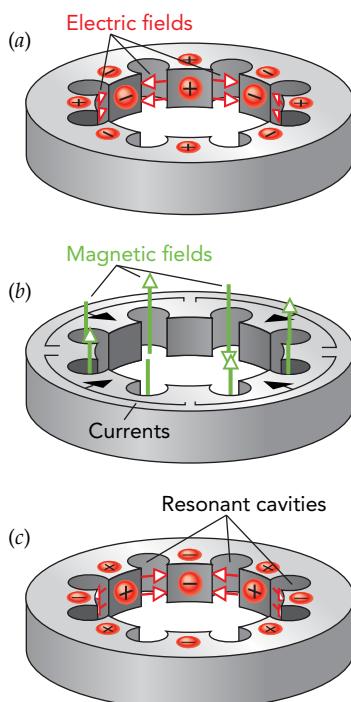


Fig. 12.2.5 A typical magnetron has eight C-shaped resonant cavities arranged in a ring. (a) Separated charge on the tips of the cavities (b) flows as currents through the ring and (c) becomes reversed. As the currents flow, magnetic fields appear in the eight cavities, pointing alternately up and down.

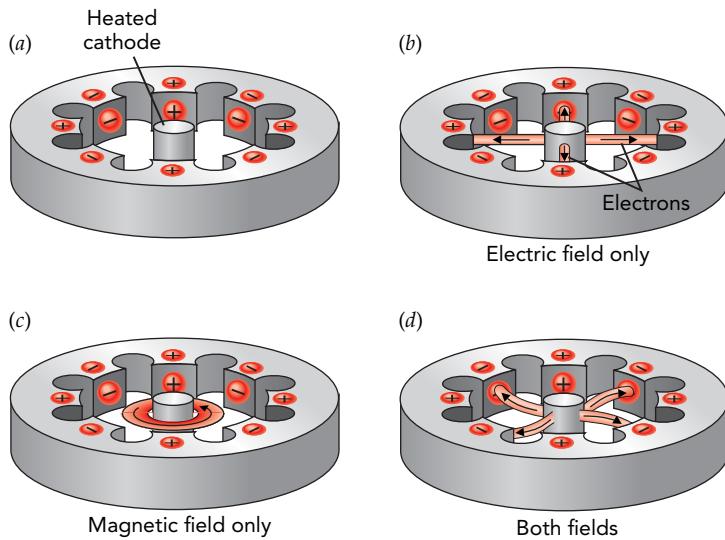


Fig. 12.2.6 (a) Electrons are emitted by the hot cathode in the center of a magnetron's ring of resonant cavities. (b) Electric fields alone would accelerate the electrons toward the positively charged cavity tips. (c) A magnetic field alone (pointing upward) would make the electrons orbit the cathode in counterclockwise loops. (d) Together, these fields create spokelike electron beams that circle the cathode counterclockwise and always strike negatively charged tips of the cavities.

the electrons. An electron has an electric charge but not a magnetic pole, and a stationary charge experiences a force from an electric field but not from a magnetic field. How then can the magnetron's magnetic field affect the motion of the electrons?

The key word in that last paragraph is *stationary*. Once a charge is *moving* through a magnetic field, it does experience a force—the **Lorentz force**. Named after its discoverer, Dutch physicist Hendrik Antoon Lorentz (1853–1928), the Lorentz force affects a charge that is moving through a magnetic field. This force pushes the charge at right angles to both the charge's velocity and the magnetic field (Fig. 12.2.7). The strength of the Lorentz force is proportional to the charge, to the velocity, to the magnetic field, and to the sine of the angle between the velocity and the magnetic field. Last, the direction of the Lorentz force on a positive charge follows a right-hand rule: when the extended index finger of your right hand points along the charge's velocity and your bent middle finger points along the magnetic field, the force on the charge points along your extended thumb. A negative charge experiences a force in the opposite direction. This relationship can be written as a word equation,

$$\text{Lorentz force} = \text{charge} \cdot \text{velocity} \cdot \text{magnetic field} \cdot \text{sine of angle}, \quad (12.2.2)$$

in symbols:

$$F = qvB \cdot \sin(\text{angle}),$$

and in everyday language:

When charged particles from the sun encounter Earth's magnetic field, they get pushed into spiral paths, producing the aurora borealis and the aurora australis,

where the angle involved is between the velocity and the magnetic field, and the direction of the Lorentz force follows the right-hand rule.

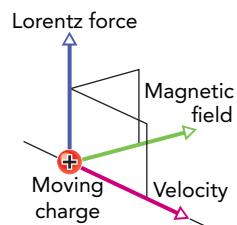


Fig. 12.2.7 A positive charge moving through a magnetic field experiences a Lorentz force that's perpendicular to both its velocity and the magnetic field. A negatively charged particle experiences a Lorentz force in the opposite direction.

The Lorentz force dramatically changes the paths of electrons in the magnetron. If there were no other fields inside the magnetron, electrons would experience only Lorentz forces perpendicular to their velocities and would circle around the magnetic flux lines in counterclockwise loops—a behavior known as **cyclotron motion**. The circling electrons would remain near the cathode and would never go near the cavities.

In a real magnetron, however, the electric field of Fig. 12.2.6b and the magnetic field of Fig. 12.2.6c are present simultaneously. Because both of these fields exert forces on moving electrons, the paths the electrons follow are extremely complicated (Fig. 12.2.6d). The outward-directed and circulating motions merge together into four electron beams that arc outward and rotate counterclockwise, like the spokes of a spinning bicycle wheel. An electron beam reaches each cavity, not at its positively charged tip, as it would without the magnetic field, but at its negatively charged tip. The electron beams actually add to the charge separations in the cavities!

The electron beams sweep around the cathode in perfect synchronization with the oscillating charge on the cavities. The beams sweep from one tip to the next in the same amount of time it takes for the charge separation on the tips to reverse. As a result, the beams always arrive on the negatively charged tip. By adding to the charge separations, the electron beams provide power to the oscillations in the cavities, keeping them going and allowing them to transfer power to the food. The electron beams actually initiate the oscillation in the cavities by adding energy to tiny random oscillations that are always present in electric systems.

How does the oscillating charge inside the magnetron create microwaves inside the oven's cooking chamber? There are many ways to extract microwaves from the ring of cavities. One extraction method is to insert a single-turn wire coil into one of the magnetron's cavities. As the magnetic field in that cavity changes, it induces a 2.45-GHz alternating current in the coil. One end of this coil is attached to the ring, but the other end passes out of the magnetron through an insulated, air-tight hole in the ring and connects to a quarter-wave monopole antenna. This 3-cm (1.2-in) antenna emits microwaves into a metal pipe attached to the cooking chamber. These microwaves reflect their way through the pipe and into the cooking chamber, where they cook the food.

Check Your Understanding #4: Lorentz Speaks

An ordinary audio speaker contains a wire coil immersed in a strong magnetic field. When an audio system sends currents through the coil, the coil experiences a force proportional to that current. What force is pushing on the coil?

Answer: The pushing force is the Lorentz force.

Why: The moving charges in the coil's current experience the Lorentz force as they pass through the magnetic field. This force is conveyed to the wire coil, which is attached to a movable surface. That surface moves back and forth as the current fluctuates and produces sound. Speakers are clearly an elegant and practical application of the Lorentz force in everyday life.

Check Your Figures #2: Lorentz Speaks with Precision

If the coil in an audio speaker experiences a Lorentz force of 1 N when it carries a current of 1 A, what force will it experience when it carries a current of 2 A and all the charges in it are thus traveling twice as fast as before?

Answer: It will experience a force of 2 N.

Why: As indicated in Eq. 12.2.2, the Lorentz force is proportional to the velocity of a charge. Doubling the current in the coil doubles the velocities of its mobile charges, and they experience twice the Lorentz force.

Epilogue for Chapter 12

This chapter examined two common devices that are based on electromagnetic waves. In Radio, we saw that electromagnetic waves can be created by accelerating electric charge and that these waves can be detected by looking for their effects on other electric charges. We also examined the techniques that are used to send sound information through space by having them control either the amplitude or the frequency of electromagnetic waves.

In Microwave Ovens, we explored the ways in which electromagnetic waves can interact directly with polar water molecules and can transfer energy to those molecules. We saw how interactions between microwaves and a metal object can lead to reflection, sparking, or heating. We also examined the technique used in ovens to create powerful microwave radiation.

Explanation: A Disc in the Microwave Oven

The fluctuating electric field in the microwave oven propels large currents back and forth through the disc's metal layer. That layer is so thin that it has a large electric resistance and it heats up as the current flows through it. The temperature of the plastic also rises because of its intimate contact with the metal layer. Since the plastic has a larger coefficient of volume expansion than the metal, the expanding plastic tears the metal layer and reduces it to islands with sharp points and narrow bridges.

Once the metal layer has fragmented, the microwave-driven currents can put substantial electric charges on the sharp points. Those charges can jump between islands as sparks. The currents passing through narrow conducting bridges can heat those bridges so hot that their metal vaporizes and they form glowing, current-carrying plasma arcs.

Chapter Summary and Important Laws and Equations

How Radio Works: A radio transmitter creates a radio wave when electric charge accelerates up and down its antenna. To get as much charge moving as possible, the transmitter attaches a tank circuit to the antenna and slowly adds energy to that tank circuit until an enormous amount of charge is flowing up and down the antenna. If the antenna is one-quarter wavelength long, it's also resonant at the transmission frequency and boosts the amount of charge sloshing up and down.

The radio receiver detects this radio wave when it causes charge to accelerate up and down the receiving antenna. If both the receiving antenna and the receiver's tank circuit are resonant at the transmission frequency, large amounts of charge will slosh back and forth in the receiver's tank circuit and the receiver will detect the transmission.

This radio wave can represent sound using either the AM or FM technique. In the AM technique, the strength of the wave is increased or decreased to represent compressions and rarefactions of the air, respectively. In the FM technique, the precise frequency of the transmission is increased or decreased to represent those compressions or rarefactions.

How Microwave Ovens Work: A microwave oven uses microwaves to cook food. These microwaves bounce around the cooking chamber, where they transfer energy to water molecules in the food. Because a water molecule is polar, having a positive end and a negative end, it tends to align with an electric field. The microwave's fluctuating electric field causes the tightly packed water molecules to twist back and forth rapidly, and the ensuing collisions heat the water and cook the food.

The oven's microwaves are produced by a magnetron, a vacuum tube containing resonant cavities and a heated cathode. By combining strong electric and magnetic fields, the magnetron produces powerful beams of electrons that add energy to the charge oscillating in the cavities. A loop of wire and a short antenna extract power from the resonant cavities and emit the microwaves that then cook the food.

- 1.** Energy in an electric field: The energy in an electric field is equal to the square of that field times its volume, divided by 8π times the Coulomb constant, or

$$\text{energy} = \frac{\text{electric field}^2 \cdot \text{volume}}{8\pi \cdot \text{Coulomb constant}}. \quad (12.1.1)$$

- 2.** Relationship between wavelength and frequency: The frequency of an electromagnetic wave times its wavelength equals the speed of light, or

$$\text{speed of light} = \text{wavelength} \cdot \text{frequency}. \quad (12.2.1)$$

- 3.** Lorentz force: When an electric charge moves through a magnetic field, it experiences a force equal to its charge times its velocity times the magnetic field times the sine of the angle between the velocity and the magnetic field, or

$$\begin{aligned} \text{Lorentz force} &= \text{charge} \cdot \text{velocity} \cdot \text{magnetic field} \\ &\cdot \text{sine of angle}, \end{aligned} \quad (12.2.2)$$

where that force is at right angles to both the velocity and the magnetic field and follows a right-hand rule.

13

Light

Although radio waves and microwaves are useful for communications and energy transfer, there's another portion of the electromagnetic spectrum that we find far more important—light. Light consists of very-high-frequency, very-short-wavelength electromagnetic waves. Light's frequencies are so high that normal antennas can't handle it. Instead, it's absorbed and emitted by the individual charged particles in atoms, molecules, and materials. Because of its special relationship with the charged particles in matter, light is important to physics, chemistry, and materials science. Moreover, it's one of the principal ways by which we interact with the world around us.

ACTIVE LEARNING EXPERIMENTS

Splitting the Colors of Sunlight

We see light because it stimulates cells in our eyes. This stimulation is an example of light's ability to influence chemistry. And because our eyes are able to distinguish among the different wavelengths of light, we perceive colors. Sunlight normally appears uncolored because it contains a rich mixture of wavelengths that our eyes interpret as whiteness. However, there are situations in

which sunlight becomes separated into its constituent colors.

You can observe this separation of colors by looking at sunlight passing through a cut crystal glass or bowl, or by reflecting sunlight from a CD or DVD. Hold the object in direct sunlight and observe the light that it redirects toward your eyes or projects onto a white sheet of paper



Courtesy Lou Bloomfield

nearby. While some of the light you see will still be white, you should see colors as well.

Turn the object slowly in your hand and observe how the colors change. You will see gradual progressions from one color to the next. What should that sequence of colors

be? How does this sequence relate to the colors of the rainbow? What is the relationship between this sequence and the wavelengths of light? Can you get some sense for the relative wavelength spacings of the classic rainbow colors: red, orange, yellow, green, blue, indigo, and violet?

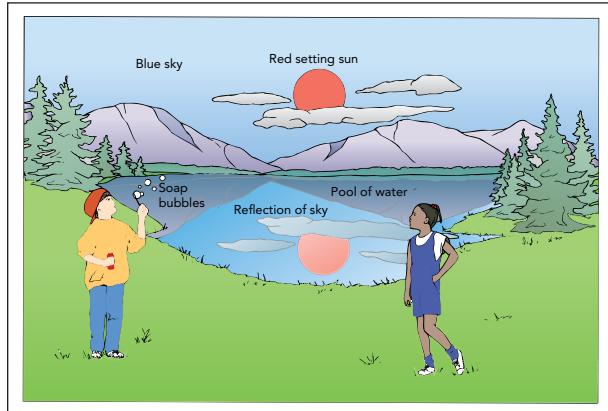
Chapter Itinerary

In this chapter, we'll examine three sources of light: (1) *sunlight*, (2) *discharge lamps*, and (3) *LEDs and lasers*. In Sunlight, we see how sunlight travels to our eyes and how its passage through the atmosphere, raindrops, and soap bubbles can separate it into its constituent colors. In Discharge Lamps, we explore the ways in which atoms and molecules emit and absorb light, and how different atoms and molecules can be used to produce light of different colors. In LEDs and Lasers, we look at how

electronic devices can produce light and at how atoms, molecules, and solids can duplicate or amplify the light passing through them to produce intense beams of highly ordered light. In the process of studying these three light sources, we'll also learn about three different types of light: thermal light, atomic resonance light, and coherent light. For additional preview information, flip to the Chapter Summary and Important Laws and Equations at the end of the chapter.

SECTION 13.1

Sunlight



For thousands of years, people have marked the passage of time by the rising and setting of the sun over the horizon. The sun first appears as a red disk in the east every morning, rises white in the blue sky, and then sets once again as a red disk in the west. The sunlight that we see takes about 8 minutes to travel the 150,000,000 km (93,000,000 mi) from the sun to our eyes and provides most of the energy and heat that make

life on Earth possible. Although the light in sunlight is really just another electromagnetic wave, and could be considered part of the previous chapter, it's so important to everyday life that it deserves special attention. Therefore we'll begin by looking at how sunlight interacts with our world.

Questions to Think About: Why is the sky blue during the day? Why is the sun red at sunrise and sunset? Why do you see rainbows only when the sun is relatively low in the sky? Why do we see colors when sunlight passes through cut crystal or through a soap bubble?

Experiments to Do: Sunlight is actually composed of many different electromagnetic waves. These waves differ in frequency and wavelength like the radio waves from your two favorite stations. You don't need a machine to help you distinguish among the various wavelengths of light; you can use your eyes. Take a look at a soap bubble on a bright sunny day. You see the bubble because it reflects light. In fact, the clear bubble appears colored, even though the sunlight hitting it is white. That's because the bubble separates sunlight according to wavelength and sends only certain wavelengths toward your eyes.

Sunlight and Electromagnetic Waves

Electromagnetic waves can have any wavelength, from thousands of kilometers to a fraction of the width of an atomic nucleus. The radio waves and microwaves that we examined in Chapter 12 have wavelengths longer than 1 mm. In the present chapter, we turn our attention to shorter-wavelength radiation. In particular, we'll study electromagnetic waves with wavelengths between 400 nm and 750 nm (recall that 1 nm, or 1 nanometer, is 10^{-9} m). These are the electromagnetic waves that we perceive as **visible light** and the principal components of sunlight.

Because the electromagnetic waves in sunlight have such short wavelengths, their frequencies lie between 10^{14} and 10^{15} Hz (Fig. 13.1.1). As one of these waves of sunlight passes by, its electric field fluctuates back and forth almost 1,000,000,000,000 times each second. Since producing microwaves, which have much longer wavelengths and much lower frequencies, already requires specialized components and tiny antennas, what can possibly emit or absorb light waves? The answer is the individual charged particles in atoms, molecules, and materials. These tiny particles can move extremely rapidly, often vibrating about at frequencies of 10^{14} Hz, 10^{15} Hz, or even higher. As these charged particles accelerate back and forth, they emit light waves. Similarly, passing light waves cause individual charged particles in atoms, molecules, and materials to accelerate back and forth, thereby absorbing the light waves as well.

Sunlight originates at the outer surface of the sun, in a region called the photosphere. There, atoms and other tiny charged systems (mostly atomic ions and electrons) jostle about at 5800 K. Since these charged particles accelerate as they bounce around, they emit electromagnetic waves.

Because the sun's surface emits light through the random, thermal motions of its charged particles, the distribution of wavelengths it emits is determined only by its temperature. It emits a blackbody spectrum, as we discussed in Section 7.3. Because the photosphere's temperature is 5800 K, the jostling motions are extremely rapid and most of the sunlight falls in the visible portion of the electromagnetic spectrum (Fig. 13.1.2).

However, not all sunlight is visible. On the long-wavelength, low-frequency side of visible light is **infrared light**. We can't see infrared light with our eyes, but we feel it when we stand in front of a hot object. In sunlight, infrared light is produced by charges that are accelerating back and forth more slowly than average.

On the short-wavelength, high-frequency side of visible light is **ultraviolet light**. We can't see ultraviolet light either, but we're aware of its presence because it induces chemical damage in molecules. It causes sunburns and encourages skin to tan. In sunlight, ultraviolet light is produced by charges that are accelerating back and forth more rapidly than average.

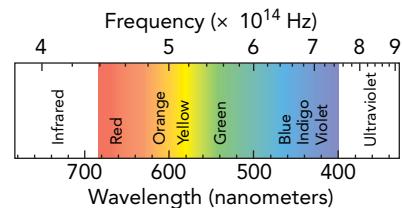


Fig. 13.1.1 The visible portion of the spectrum of sunlight. Each wavelength of visible light has a particular frequency and is associated with a particular color. At the ends of the visible spectrum are invisible infrared and ultraviolet lights.

Check Your Understanding #1: The Rosy Glow of Candlelight

Why does a burning candle emit reddish or yellowish light?

Answer: Charged particles in the hot flame accelerate back and forth and emit electromagnetic waves that include the low-frequency end of the visible spectrum.

Why: The accelerating charged particles in hot objects emit light. The hotter the object, the faster the charged particles move and accelerate and the higher the frequencies of light they emit. A candle isn't hot enough to emit the whitish light of the sun, so it emits mostly reddish or yellowish light.

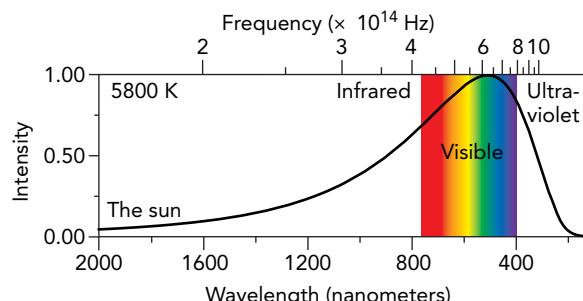


Fig. 13.1.2 Sunlight comes from the sun's photosphere, where the temperature is 5800 K. This light has a blackbody distribution of wavelengths, with much of its intensity concentrated in the visible portion of the overall electromagnetic spectrum.

Sunlight's Passage to Earth

Sunlight travels from the sun to Earth at the speed of light—but what sets the speed of light? Actually, as we learned in Section 4.2, it's one of the fundamental constants of nature, with a defined value of 299,792,458 m/s in empty space. Although we could argue that the speed of light is set by the relationships between the electric and magnetic fields, that observation simply passes the buck. If you were then to ask what sets the relationships between the electric and magnetic fields, the answer would be the speed of light.

Rather than justifying why sunlight travels as fast as it does in empty space, let's look at what happens to it when it enters a region that's not empty. After all, sunlight eventually reaches Earth's atmosphere and, when it does, several interesting things happen.

First, the sunlight slows down as its electric and magnetic fields begin to interact with the electric charges and magnetic poles in the atmosphere. Light polarizes the molecules it encounters—its electric field displaces positive charges from negative charges, and its magnetic field displaces north poles from south poles. These polarization effects delay light's passage so that it travels more slowly. Since most transparent materials respond much more strongly to light's electric field than to its magnetic field, we'll concentrate on only electric effects.

The factor by which light slows down in a material is known as the material's **index of refraction**. Light travels particularly slowly through materials that are easy to polarize, and some of them have indices of refraction of 2 or even 3. Because air near sea level is only slightly polarizable, however, its index of refraction is just 1.0003. Although this reduction in light's speed is too small to notice directly, we do notice the polarized air particles that cause it. These polarized air particles are what makes the sky blue (Fig. 13.1.3).

The particles in air consist of individual atoms and molecules, small collections of atoms and molecules, water droplets, and dust. As a wave of sunlight passes through one of these particles, the particle becomes polarized. Its electric charges accelerate back and forth as the sunlight's electric field pushes them around, and they reemit a new electromagnetic wave of their own.

This new wave draws its energy from the original wave. In effect, the particle acts as a tiny antenna, temporarily receiving part of the electromagnetic wave and immediately retransmitting it in a new direction. This process, whereby a tiny particle redirects the path of a passing light wave, is called **Rayleigh scattering**, after the English physicist Lord Rayleigh (John William Strutt, 1842–1919), who first understood it in some detail.

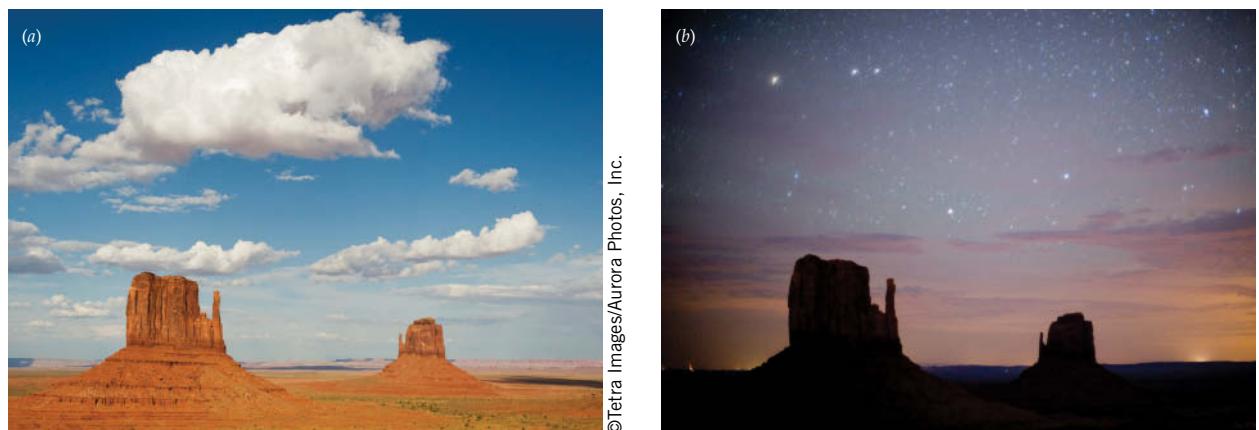


Fig. 13.1.3 (a) During the day, the sky above Monument Valley appears blue because Earth atmosphere scatters mostly blue sunlight toward us. (b) At night, the scattered sunlight is gone and the atmosphere is a clear window through which to observe the stars.

Although most sunlight travels directly to our eyes, some of it undergoes Rayleigh scattering and reaches us by more complicated paths. We see the direct light as coming from the brilliant disk of the sun, but the scattered light gives the entire sky a fairly uniform blue glow (Fig. 13.1.4). Why is this glow blue?

The sky's blue color comes about because the tiny air particles that Rayleigh-scatter sunlight are too small to make good antennas for that light. We observed in Section 12.1 that a dipole antenna works best when it is about half as long as the wavelength of its electromagnetic wave. The air particles make particularly bad antennas for long-wavelength red light, so very little red sunlight undergoes Rayleigh scattering on its way through the atmosphere. However, the air particles make less poor antennas for short-wavelength blue light. Some of the violet, indigo, and blue sunlight does Rayleigh-scatter and reaches our eyes from all directions. We see this Rayleigh-scattered light as the bluish glow of the sky.

Rayleigh scattering not only makes the sky blue; it also makes the sunrises and sunsets red. As the sun rises or sets, its light must travel long distances through Earth's atmosphere to reach your eyes. Its path is so long that most of the violet, indigo, and blue light Rayleigh-scatters away miles to your east or west and all you see is the remaining red light. Sometimes the whole local sky appears reddish because there simply isn't any bluish light left to scatter toward you. Sunrises and sunsets are particularly colorful when extra dust or ash is present in the atmosphere to enhance the Rayleigh scattering. Air pollution, forest fires, and volcanic eruptions tend to create unusually red sunrises and sunsets.

In contrast, clouds and fog appear white because they're composed of relatively large water droplets. These droplets are larger than the wavelengths of visible light and scatter all of sunlight's wavelengths equally well. Although this scattering is often so effective that you can't see the sun's disk through a cloud, it doesn't give the cloud any color. The cloud simply looks white.

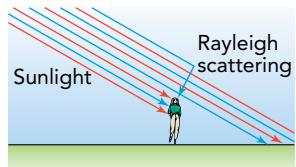


Fig. 13.1.4 As sunlight passes through the atmosphere, some of its violet, indigo, and blue light undergoes Rayleigh scattering from particles in the air. We see this redirected light as the diffuse bluish sky. The remaining light reaches our eyes directly from the sun and tends to be reddish, particularly at sunrise and sunset.

Check Your Understanding #2: Seeing the Blues

The air in a dark, smoky room often looks bluish when illuminated by white light. What creates this bluish appearance?

Answer: Rayleigh scattering from the microscopic smoke particles scatters more blue light than red light and gives the air a bluish glow.

Why: When smoky air in a dark room is illuminated by a bright spotlight or another white light source, the tiny particles in the air Rayleigh-scatter some of the light. When you look at the air against a dark background, you can see this glow. Since Rayleigh scattering affects blue light most strongly, the air appears bluish.

Rainbows

Sometimes water droplets do separate the colors of sunlight. When sunlight shines on clear round raindrops as they fall during a storm, these raindrops can create a rainbow. To understand how clear spheres of water can bend sunlight's path and separate it according to wavelength, we must understand three important optical effects: refraction, reflection, and dispersion. We encountered those same wave phenomena while studying water surface waves in Section 9.3, but now they appear in a new context—light waves!

Let's begin by looking at what happens when a wave of sunlight passes directly through a raindrop. Because water is more polarizable than air, the wave slows down inside the raindrop and its cycles bunch together (Fig. 13.1.5). This bunching effect reduces the light's wavelength inside the drop. Light's frequency remains unchanged because the cycles can't disappear, but they move more slowly.

If a narrow wave of sunlight is aimed directly through the center of the raindrop, it will follow a straight path and emerge essentially unaffected from the other side (Fig. 13.1.6a). However, if that wave strikes the raindrop near the top, it will bend as it enters the water (Fig. 13.1.6b). This occurs because the lower edge of the wave will reach the water first and

Fig. 13.1.5 As an electromagnetic wave enters a material, its speed decreases and the waves bunch up together. Its wavelength decreases.

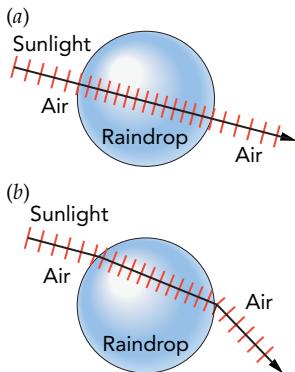
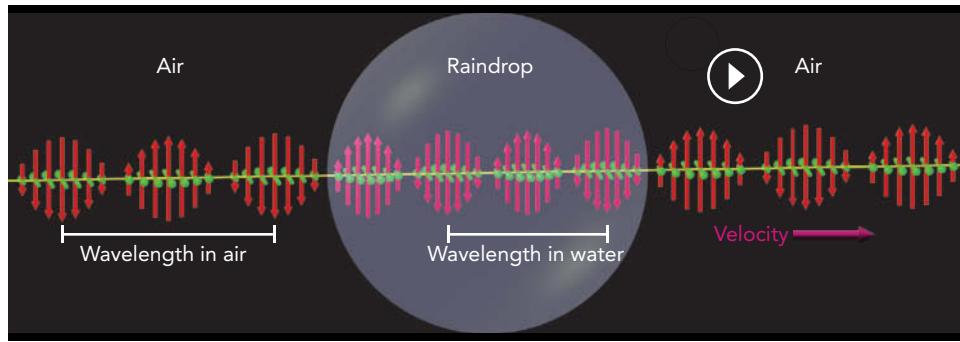


Fig. 13.1.6 A side view of two narrow waves of sunlight entering and leaving raindrops. The lines drawn across each light wave represent upward electric field maxima and are bunched together as light slows down in water.

slow down, the upper edge will then overtake it and the wave will bend downward. The wave will head more directly into the water.

As the wave in Fig. 13.1.6b leaves the raindrop, its upper edge emerges first and speeds up while the lower edge lags behind. The wave bends downward even further and heads less directly into the air and away from the water.

This bending of sunlight at the boundaries between materials is *refraction*. It occurs whenever sunlight changes speeds as it passes through a boundary at an angle. If sunlight slows down at a boundary (Fig. 13.1.7a), it bends toward the line perpendicular to the boundary and heads more directly into the new material. If sunlight speeds up at a boundary (Fig. 13.1.7b), it bends away from the line perpendicular to the boundary and heads less directly into the new material. The amount of the bend increases as the speed change increases.

However, part of the sunlight striking a boundary doesn't pass through the boundary at all. Instead, it *reflects*. In Section 9.3, we attributed wave reflection specifically to changes in wave speed at a boundary. However, the more general cause of wave reflections is an **impedance mismatch**, an abrupt change in how the wave moves through its environment. In general, **impedance** measures a system's opposition to the passage of a current or a wave. For an electric current, impedance measures how much voltage is needed to propel a given current. For an electromagnetic wave, impedance measures how much electric field is needed to produce a given magnetic field. Impedance effects are common in nature (see **1**) and also apply to mechanical waves and mechanical currents. When sound and water waves encounter impedance mismatches, they partly reflect as echos.

1 When electromagnetic waves travel through cables and wires, they reflect from impedance mismatches. These mismatches occur whenever the relationship between electric and magnetic fields changes, and should be avoided in television and data communication systems. Using cables, wires, or adapters with the wrong impedances or leaving open ports on cable splitters will produce reflections that may interfere with reception.

Fig. 13.1.7 When a light wave is incident on a glass surface, part of it reflects and the rest refracts. The incident wave and reflected wave are at equal but opposite angles from the line perpendicular to the surface. The transmitted wave refracts (bends) toward or away from the perpendicular. (a) Light that slows down when moving from air to glass refracts toward the perpendicular. (b) Light that speeds up when moving from glass to air refracts away from the perpendicular.

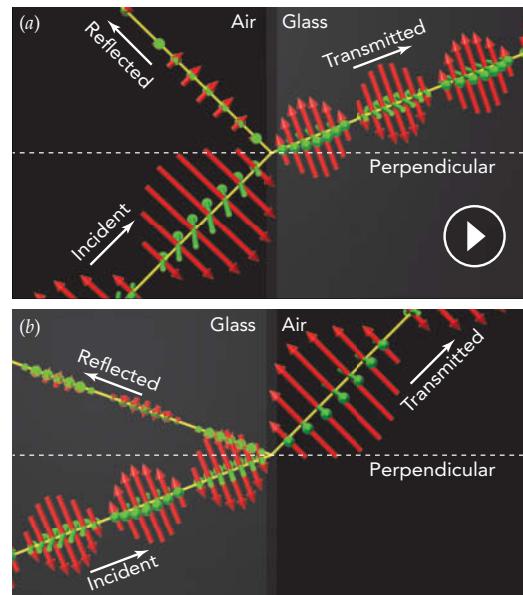




Fig. 13.1.8 A rainbow forms when water droplets reflect sunlight back toward your eyes. Because the different wavelengths of light follow slightly different paths, we see the different colors coming from slightly different directions and observe bands of color.

2 Sand appears white because it redirects sunlight in all directions. One explanation for this effect is that the sand grains act as tiny antennas that respond to and reemit light's electromagnetic waves. A second explanation is that the sand grains present the sunlight with thousands of air-sand boundaries from which to reflect. However, both explanations are descriptions of exactly the same physics—the charged particles in the sand grains are electrically polarized by the waves passing through them. These waves are randomly redirected without being absorbed, and they give the sand its white appearance.

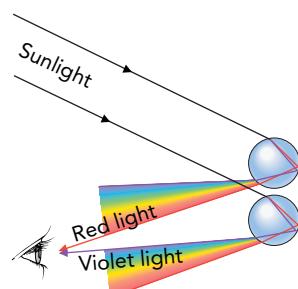


Fig. 13.1.9 As sunlight passes through spherical raindrops, its colors separate. Violet light bends more at each air-water boundary than does red light, and the two emerge from the raindrops heading in different directions. You see red light coming toward you from the upper raindrops and violet light from the lower raindrops.

The impedance of empty space is high because there is nothing to aid it in producing a magnetic field. Inside most materials, however, the electric field has help. The electric field polarizes the material, which then helps to create the magnetic field. Because of this assistance, the impedance of most materials is much less than that of empty space. Since air is almost empty space, the boundary between air and water is an impedance mismatch for light.

Passing through an impedance mismatch upsets the balance between a light wave's electric and magnetic fields. To compensate for this imbalance, part of the incoming wave reflects off the boundary. Thus some sunlight reflects each time it enters or leaves a water droplet. The fraction of light that reflects depends on the severity of the impedance mismatch, but it is typically 4% between air and most transparent materials, including water (for reflection from sand, see **2**). In contrast, metals polarize so easily that their impedances are essentially zero and they reflect light almost perfectly.

There is one more important point about sunlight's passage through water: red light travels about 1% faster through water than violet light does. That's because higher-frequency violet light polarizes the water molecules a little more easily than lower-frequency red light does and that increased polarization slows down the violet light. This frequency dependence of light's speed in a material is *dispersion*. Dispersion affects refraction. The more light slows as it enters a raindrop, the more it bends at the boundary. Since violet light slows more than red light, violet light also bends more and the different colors of sunlight follow somewhat different paths through the raindrop.

A rainbow is created when raindrops separate sunlight according to color (Fig. 13.1.8). To see the rainbow, you stand with the sun at your back and look up at the sky. When sunlight hits the raindrops, they redirect some of that light back toward you. Since each raindrop redirects light only in a narrow range of angles, you can't see light from every raindrop. Only the raindrops in a narrow arc of the sky redirect visible light toward you. This arc appears brightly colored because raindrops at the inner edge of the arc send violet light toward you while raindrops at the outer edge of the arc send red light toward you. In between, you see all the colors of the rainbow.

Figure 13.1.9 shows how a raindrop redirects different colors of light in different directions. Although there are many possible paths light can take through the raindrop, this path is the one that produces rainbows. Sunlight enters near the top of the raindrop and bends inward. Violet light bends more than red light, so the sunlight begins to separate according to color. Some sunlight is also reflected from the raindrop but doesn't contribute to rainbows.

When the light inside the raindrop strikes the back surface, most of it leaves the drop and is lost. A small fraction of the light reflects from that surface, however, and continues to travel

through the raindrop. When this light reaches the raindrop's front surface, most of it leaves the drop. Violet light bends more strongly than red light as they reenter the air, so the different colors of light leave the drop heading in different directions. Since violet light is redirected more upward than red light, you see violet light coming toward you from the lower raindrops. Red light is redirected more downward so you see it coming toward you from the upper raindrops. Thus the upper arc of the rainbow is red, while the lower arc is violet.

Check Your Understanding #3: The Look of Diamonds

A diamond pendant sparkles with color when you look at it in sunlight. From where do the colors come?

Answer: A diamond exhibits dispersion, so different frequencies of sunlight follow somewhat different paths through the diamond's polished facets. The different colors of sunlight emerge from the diamond traveling in slightly different directions, so you can see them individually.

Why: One of the delightful aspects of a diamond is its strong dispersion. It bends violet light much more than red light, so sunlight is separated into its different colors as it passes through the stone. Cleverly cut facets help to isolate the individual colors.

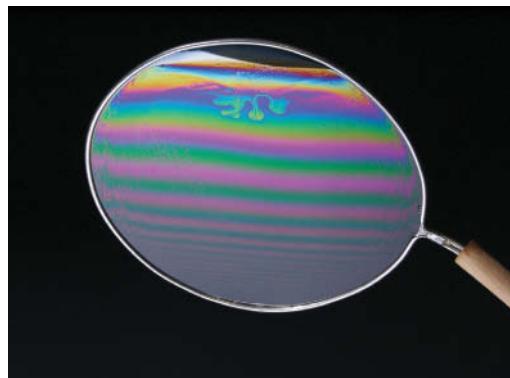
Soap Bubbles

Soap bubbles also separate sunlight into its various colors (Fig. 13.1.10), but they use another wave phenomenon—interference. We encountered interference of mechanical waves in Section 9.3 when we studied wave beat at the seashore and interference of electromagnetic waves in Section 12.2 when we considered the unevenness of microwave cooking. Now we'll look at interference in another type of electromagnetic wave, light.

Light's interference effects stem from the summing or *superposition* of its electromagnetic waves. When several light waves overlap at a particular location, their electric fields sum together and so do their magnetic fields. If their individual fields all point in the same direction, the waves experience *constructive interference*—they sum together in a mutually assisting way and the light intensity at that location is enhanced (Fig. 13.1.11a). However, when their individual fields point in opposing directions, the waves experience *destructive interference*—they sum together in a canceling way and the light intensity at that location is reduced (Fig. 13.1.11b).

Both forms of interference occur when sunlight reflects from the outer skin of a soap bubble. As each wave of sunlight hits that thin film of soapy water, the film's front surface reflects about 4% of the wave and the film's back surface reflects another 4% (Fig. 13.1.12). Since both reflections travel in the same direction, the reflected light that you see reaches your eyes via two different paths, one from each reflection. If these two waves arrive in phase—that is, with their electric fields synchronized and assisting one another—you see the particularly bright reflection of constructive interference. If the two waves arrive out of

Fig. 13.1.10 Light reflected by the front and back surfaces of this soap film interferes with itself and gives the film its colorful appearance. Since the colors are determined by film thickness and since the film's thickness increases in the downward direction, the film displays horizontal bands of color.



Courtesy Lou Bloomfield

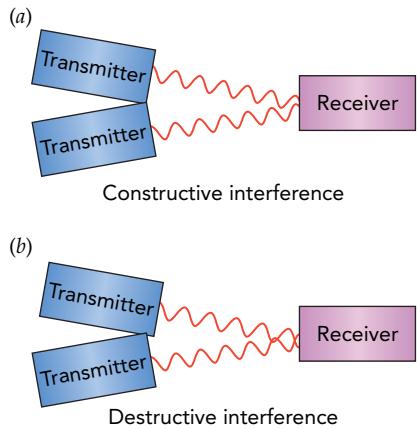


Fig. 13.1.11 (a) When waves from two separate paths arrive in phase at the receiver, constructive interference produces a particularly large effect on the receiver. (b) When waves arrive out of phase, destructive interference produces a particularly weak effect on the receiver.

phase—with their electric fields canceling one another—you see the particularly dim reflection of destructive interference.

Whether you see constructive or destructive interference depends on the wavelength of the sunlight. The back-surface reflection has to travel twice through the soap film, so it's delayed relative to the front-surface reflection. If the delay is just long enough for the wave to complete an integral number of cycles, then the two reflected waves are in phase with one another as they head toward your eye and you see a bright reflection. If the delay allows the back-surface reflection to complete an extra half cycle, then the two reflected waves are out of phase with one another and you see a dim reflection.

Sunlight contains many different wavelengths of light, and these wavelengths behave differently during the reflection process. You see a colored reflection, consisting mainly of those wavelengths of light that experience constructive interference. Because the delay experienced by the back-surface reflection depends on the thickness of the soap film, you can actually determine the film's thickness by studying its colors.

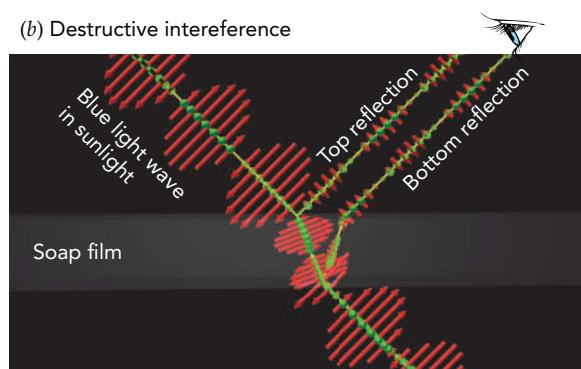
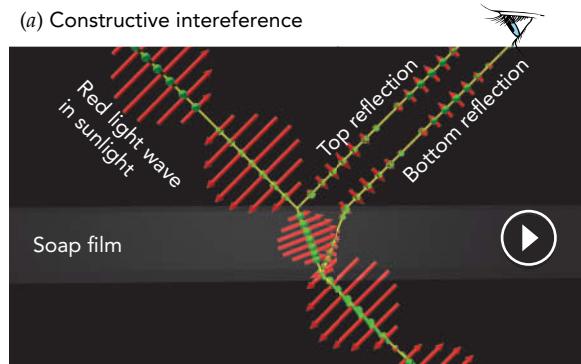


Fig. 13.1.12 Sunlight partially reflects from both the top and bottom surfaces of a soap film, but the bottom-surface reflection is delayed relative to the top-surface reflection. Phase differences between those two partial reflections lead to interference effects when they both enter your eye. For this soap film and viewing angle, (a) a red light wave's reflections enter your eye in phase and experience constructive interference, so you see the red light. However, (b) a blue light wave's reflections enter your eye out of phase and experience destructive interference, so you see no blue light.

Check Your Understanding #4: A Slick-Looking Oil Slick

A thin layer of oil or gasoline floating on water appears brightly colored in sunlight. From where do these colors come?

Answer: They come from interferences between light reflected by the top and bottom surfaces of the floating layer of oil.

Why: A thin film of almost anything on water will appear colored because of interference. Part of each light wave striking the film reflects from the top surface, and part reflects from the bottom surface. These two reflected waves interfere with one another in a wavelength-dependent manner and make the layer appear brightly colored. Different colors correspond to different thicknesses of the thin films.

Sunlight and Polarizing Sunglasses

All sunglasses absorb some of the sunlight passing through them, but the best ones absorb horizontally polarized light much more strongly than vertically polarized light. These polarizing sunglasses dramatically reduce glare by eliminating most of the light reflected from horizontal surfaces.

When light strikes a transparent surface at right angles, about 4% of that light is reflected, regardless of its polarization. When light strikes a horizontal surface at a shallow angle, however, its polarization profoundly affects its reflection. A light wave that has a horizontal electric field (essentially horizontally polarized light) reflects strongly from the surface (Fig. 13.1.13a). In contrast, a light wave that has a horizontal magnetic field (essentially vertically polarized light) barely reflects at all (Fig. 13.1.13b).

When the wave's electric field is horizontal, it pushes electric charges back and forth along a horizontal surface. Charges shift relatively easily in that direction, so the surface polarizes well and reflects the wave strongly. The shallower the angle between the surface and the incident light wave, the larger the surface's polarization and the greater the reflection. Horizontally polarized light thus reflects strongly from horizontal surfaces and its reflection increases as its angle to the surface becomes shallower.

When the wave's electric field is vertical or nearly vertical, it pushes electric charges in and out of a horizontal surface. Since the charges can't leave the surface, the surface

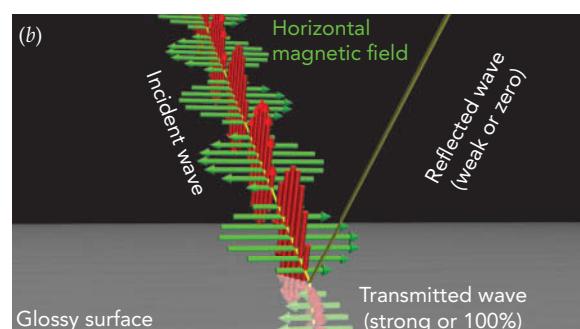
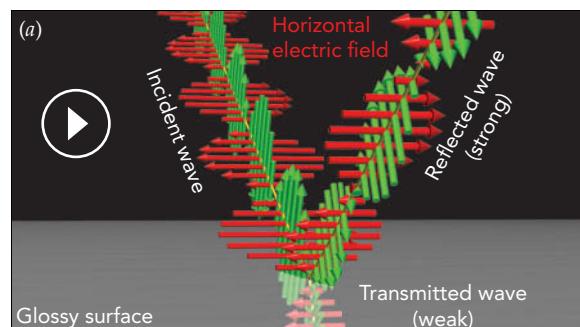
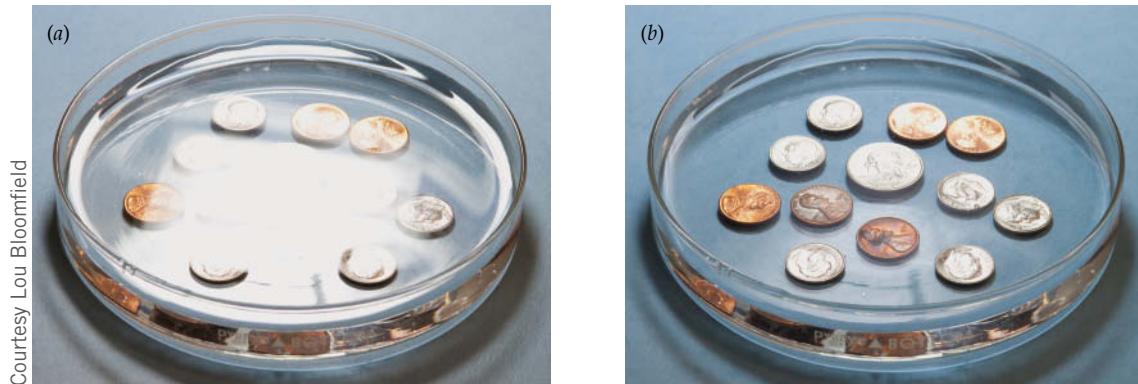


Fig. 13.1.13 When a light wave heading toward you encounters a glossy horizontal surface at a shallow angle, the strength of its partial reflection depends on its polarization. (a) When the incident wave's electric field is horizontal, the reflection is strong. (b) When the incident wave's magnetic field is horizontal, the reflection is weak. This wave is incident at precisely Brewster's angle, so there is zero reflection and all of the wave enters the glossy surface.



Courtesy Lou Bloomfield

Fig. 13.1.14 (a) The horizontal surface of water reflects light strongly at shallow angles, making it difficult to see the coins on the bottom of this dish. (b) When you look through a filter that blocks horizontally polarized light, the main component of glare from horizontal surfaces, you can see the coins lying beneath the water's surface.

polarizes ineffectually and reflects the wave weakly. The polarization and reflection become even weaker as the angle becomes shallower until, at **Brewster's angle**, the wave doesn't reflect at all. Vertically polarized light thus reflects weakly from horizontal surfaces and that reflection drops to zero at Brewster's angle. For angles shallower than Brewster's angle, however, the reflection becomes stronger as the angle becomes shallower.

Although direct sunlight is an even mixture of vertically and horizontally polarized waves, the shallow-angle glare reflecting from the horizontal surfaces of water, pavement, or painted car hoods is mostly horizontally polarized waves. Polarizing sunglasses are designed to absorb horizontally polarized light, which is why they are so effective at reducing glare (Fig. 13.1.14).

Check Your Understanding #5: Looking into the Reflecting Pool

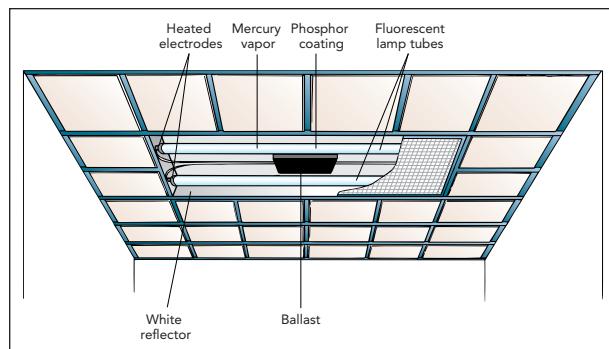
When you look into a pool of water, you see mostly a reflection of the sky. When you wear polarizing sunglasses, however, you see into the water clearly. Explain.

Answer: Most of the light that reflects from the water is horizontally polarized. The sunglasses block horizontally polarized light, so you see mostly light from within the pool of water.

Why: Sunlight reflecting from a horizontal surface is mostly horizontally polarized waves. By blocking horizontally polarized light, polarizing sunglasses virtually eliminate this reflected light, and you see mostly light from below the surface. If you turn the sunglasses sideways, they will block the wrong polarization of light and you will see mostly the reflected light. Try it.

SECTION 13.2

Discharge Lamps



Energy efficiency is crucial to modern lighting. Incandescent lamps provide pleasant, warm illumination, but most of the power they consume is wasted as invisible infrared light. Fluorescent and other gas discharge lamps produce far more visible light with the same amount of electric power and now dominate office, industrial, and street lighting. In this section we'll explore several types of discharge lamps—fluorescent, mercury vapor, sodium vapor, and metal-halide lamps—that all share a common theme: current passing through a gas.

Questions to Think About: Why are incandescent lightbulbs being phased out? Why do different fluorescent lamps often

have slightly different colors? Why is the light from a neon sign red? Why shouldn't you turn a fluorescent lamp on and off frequently? Why are streetlights so dim when they first turn on? Why are some highway lights orange?

Experiments to Do: Examine the discharge lamps around you, particularly the white fluorescent tube lamps. Where do their lights originate? In a fluorescent lamp, light comes from the white phosphor coating on the tube's inner surface. What about in a neon lamp or a mercury vapor streetlight? Compare

the colors of various lamps, including several different fluorescent lamps. Are their lights identical?

Both fluorescent and neon lamps start almost immediately—but watch how long it takes a streetlight to start. Fluorescent and neon lamps remain cool during operation, but the mercury, sodium, and metal-halide lamps used in street lighting get hot. Watch a streetlight warm up. Does its color change? If the streetlight loses power even for a moment, it must wait about 5 minutes before it can start again. Why must it wait?

How We See Light and Color

Before examining discharge lamps, let's look at how our eyes recognize color. It might seem that they actually measure the wavelengths of light, but that's not the case. Instead, our retinas contain three groups of light-sensing *cone cells* that respond to three different ranges of wavelengths. One group of cone cells responds to light near 600 nm and lets us see red, another responds to light near 550 nm and lets us see green, and a third responds to light near 450 nm and lets us see blue (Fig. 13.2.1). These cone cells are most abundant at the center of our vision. Our retinas also contain *rod cells*, which are more light-sensitive than cone cells, but rod cells can't distinguish color. They are most abundant in our peripheral vision and provide us with night vision.

Having only three types of color-sensing cells doesn't limit us to seeing just three colors. We perceive other colors whenever two or more types of cone cells are stimulated at once. Each type of cell reports the amount of light it detects, and our brains interpret the overall response as a particular color.

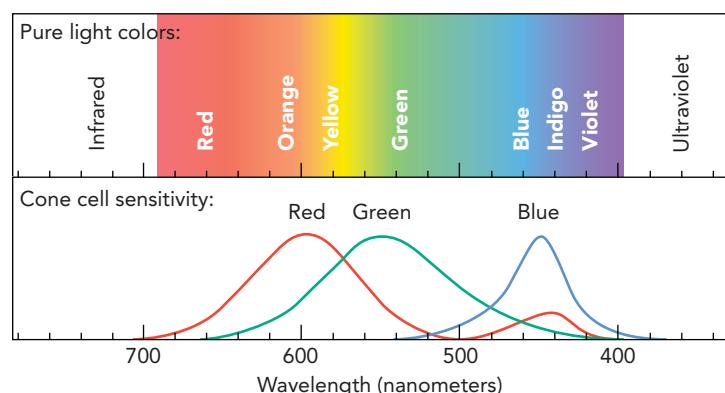
Although light of a certain wavelength will stimulate all three types of cone cells simultaneously, the cells don't respond equally. If the wavelength is 680 nm, the red cone cells will respond much more strongly than the green or blue cells. Because of this strong red response, we see the light as red.

Other wavelengths of light stimulate the three types of cells somewhat more evenly. Light with a wavelength of 580 nm is in between red and green light. Both the red-sensitive and the green-sensitive cone cells respond about equally to this light, and we see it as yellow.

We also see yellow when looking at an equal mixture of 640-nm light (red) and 525-nm light (green). The 640-nm light stimulates the red-sensitive cone cells and the 525-nm light stimulates the green-sensitive ones. Even though there is no 580-nm light entering our eyes, we see the same yellow color as before.

In fact, mixtures of red, green, and blue light can make us see virtually any color. For that reason, these three are called the **primary colors of light** or the *primary additive*

Fig. 13.2.1 The red-sensitive cells in our retina detect light near 600 nm, the green-sensitive cells near 550 nm, and the blue-sensitive cells near 450 nm. The red-sensitive cells also respond near 440 nm, so that we see violet.



colors (Fig. 13.2.2a). Color televisions and computer screens use tiny sources of red, green, and blue light to produce their full-color images.

Although the idea of mixing primary colors also applies to paints, inks, and pigments, the palette is different. The **primary colors of pigment** or the *primary subtractive colors* are cyan, magenta, and yellow (Fig. 13.2.2b). When you apply one of these primary pigments to a white surface, it absorbs or subtracts one of the primary colors of light from the surface's reflection. Cyan subtracts the reflection of red, magenta subtracts the reflection of green, and yellow subtracts the reflection of blue. Color printers, photographs, magazines, and books use tiny patches of cyan, magenta, and yellow pigments to produce their full-color images.

Check Your Understanding #1: Mixed Up Light

If you look at a mixture of 70% red light and 30% green light, what color will you see?

Answer: You will see orange.

Why: Your red-, green-, and blue-sensitive cone cells have the same response to this light mixture as they would have to pure 600-nm light. Such light appears orange, so you see orange when viewing the mixture.

More Light, Less Heat: Gas Discharges

When all three of our color-sensing cells respond about equally, we see white light. That's because our vision evolved under a single incandescent light source—the sun. Sunlight stimulates the red-, green-, and blue-sensitive cells in our eyes about evenly, so any other source of "white light" must do the same.

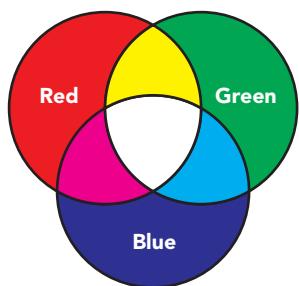
Most of the twentieth century's white-light illumination came from incandescent lightbulbs, but that won't be true for the twenty-first century. Although an incandescent lightbulb makes a good attempt at producing white light, it suffers from two serious drawbacks. First, because its filament can't reach the temperature of the sun's surface (5800 K), its blackbody spectrum contains too little blue light and appears redder than sunlight. Second, because most of its electromagnetic radiation is invisible infrared light, it doesn't use energy efficiently to produce visible light. It produces only about 15 lumens per watt, where the **lumen** is a standard measure of usable illumination. (For comparison, the discharge lamps that we're about to discuss can easily exceed 100 lumens per watt.)

Fortunately, modern science has given us some alternative sources of light, sources that don't use heat and thermal radiation to produce their light. Among these sources are gas discharge lamps, which emit light when electric currents pass through their gases. Some discharge lamps are colored, and others do excellent jobs of producing white light. Also, many of them are far more energy efficient at producing visible light than are incandescent lightbulbs.

To get an understanding of gas discharges and the lights they emit, let's start with one of the simplest examples, a neon sign tube. Although neon's rich red glow isn't suitable for lighting and isn't particularly energy efficient, it's great for signs and also relatively easy to understand.

A neon sign tube is a sealed glass tube that contains pure neon gas at a density less than 1% that of the atmosphere outside the tube. It has metal electrodes at each end so that electric current can enter the gas through one electrode and leave through the other. Of course, gases are normally electrical insulators, and neon is no exception. To transform the tube's neon into a conductor, a large voltage difference is applied between its two electrodes. As we saw in Section 10.2, a gas breaks down when exposed to large voltage gradients; its few naturally occurring ions initiate a cascade of ionizing collisions that quickly fill the gas with charged particles and render it conducting.

(a) Additive colors (light)



(b) Subtractive colors (pigment)

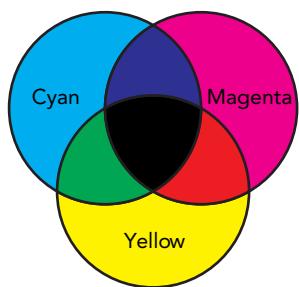


Fig. 13.2.2 (a) The primary colors of light or additive colors, red, green, and blue, can be combined to form any colors of light. (b) The primary colors of pigment or subtractive colors, cyan, magenta, and yellow, subtract red, green, and blue, respectively, from reflected or transmitted light and can be combined to form any colors of pigment.

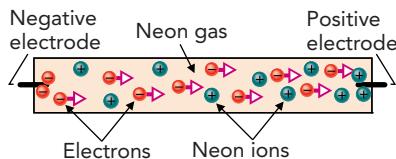


Fig. 13.2.3 A neon sign tube sends electrons through low-pressure neon gas. These electrons collide with neon atoms, transferring energy to them and causing them to emit primarily red light. Positively charged neon ions, created by particularly energetic collisions, keep the electrons from repelling one another to the walls of the tube.

Whereas ordinary air breaks down at a voltage gradient of about 30,000 V/cm, low-density neon breaks down at a much lower voltage gradient. That's because in a low-density gas, charged particles can travel farther and accumulate more kinetic energy before each collision. When about 10,000 V is applied between the electrodes of a neon sign tube, the gas breaks down and begins to emit its familiar red glow.

The lamp is then experiencing a **discharge**; that is, current is flowing through the neon gas. This current consists mostly of electrons, flowing from the tube's negatively charged electrode to its positively charged electrode (Fig. 13.2.3). Although these electrons collide frequently with neon atoms, they have so little mass that they usually just bounce off the atoms without losing much energy. Like Ping-Pong balls rebounding from elephants, the electrons do most of the bouncing and then continue on their way.

Every so often, however, an electron will collide with a neon atom and something different will happen; the neon atom will rearrange internally and absorb part of the electron's kinetic energy. The electron will rebound with less energy than it had before, and the neon atom will go on to emit light, probably red light. To understand why that light is probably red, however, we'll have to look at quantum physics and the structure of the neon atom.

Check Your Understanding #2: Making Whiter Light

Some photographic lamps simulate sunlight by placing a blue filter in front of an incandescent bulb. This filter absorbs some of the red light so that the lamp appears whiter. Does this filter increase or decrease the lamp's energy efficiency?

Answer: It decreases the lamp's energy efficiency.

Why: Although the bluish filter changes the spectrum of wavelengths so that the amount of blue light leaving the lamp increases relative to the red light, it does this by absorbing some of the red light. The filter gets hot, and less electric power used by the bulb leaves as visible light.

Particles, Waves, and Quantum Physics

As quantum physics gradually revealed itself to the scientists of the early twentieth century, they found the experience both exhilarating and disorienting. Prior to that era, the physical world seemed to divide neatly into particles and waves: scientists viewed an electron only as a particle and light only as a wave. However, one of the most basic observations of quantum physics, and the one most relevant to our present topic, is that everything has both particle and wave characteristics. Put simply, everything begins and ends as a particle, but travels as a wave.

For an electron, the quantum surprise is that it travels as a wave. For light, the quantum surprise is that it is emitted and absorbed as a particle. Called the **wave-particle duality**, this observation that everything in nature has both particle and wave characteristics has left few areas of physics unaffected. But although quantum physics is now a basic and essential part of nearly all modern physics research, its effects are subtle and often nonintuitive. They are most apparent in the microscopic world and are visible to us only indirectly. No wonder they seem so strange.

We'll encounter quantum physics and its effects throughout the remainder of this book. It figures prominently in the light emitted by atoms, the electronic properties of semiconductors, the operation of lasers, and the radioactive decays that release nuclear energy. Our examination of atoms will acquaint us with the wave nature of electrons and show us how the wave phenomena that we studied in Chapter 9 reappear in quantum physics. In discharge lamps, LEDs, and lasers, we'll explore the particle nature of light and see how the collision effects we studied in Chapters 1–3 are influenced by quantum physics. In our examination of radioactivity, we'll uncover particle and wave effects that we would not

even have anticipated without quantum physics. With each encounter, we'll take another small bite of the quantum apple—looking at how quantum effects manifest themselves in our everyday world.

Check Your Understanding #3: Wavy Atoms

Do atoms ever exhibit wave properties?

Answer: They certainly do.

Why: Like everything else in nature, atoms have both particle and wave properties. Atoms travel as waves and have recently been shown to exhibit many of the wave effects we studied in Chapter 9, including refraction, reflection, and interference.

Electron Standing Waves in Atoms: Orbitals

According to the wave-particle duality, electrons have both particle and wave characteristics. Like all objects in our universe, electrons travel as *waves* when they move from place to place, and it's only when you go looking for them that you find them as *particles* at particular locations (see 3). When they reside in atoms and you leave them alone, electrons are best understood as waves.

In a nonquantum world, an electron would exist only as a particle and would be able to move at any speed along any path, even in an atom. However, ours is a quantum world and the electron does not travel as a particle at all; it travels as a wave. The electron is therefore constrained by the rules that govern waves, some of which we encountered in Chapter 9. Like the confined mechanical waves on a violin string, drumhead, or basin of water, the confined electron waves in an atom have limited possibilities.

We observed in Chapter 9 that the most basic mechanical waves on a limited object are all standing waves—waves that effectively oscillate in place. This rule also applies to quantum waves in limited objects. An atom is a limited object, and the electrons in an atom are best understood as standing waves in that atom. Each electron wave extends across the atom and has such wave characteristics as wavelength and frequency.

Unlike a vibrating string or drumhead, however, the electron is a three-dimensional wave and its oscillation is internal. Instead of vibrating back and forth, each point of the wave cycles invisibly around a quantum phase. Since that quantum phase is rather mysterious, you can think of an electron standing wave as a fuzzy cloud in which each point cycles endlessly through a series of colors: red, yellow, green, blue, red, yellow, and so on (Fig. 13.2.4). All those points cycle at the same quantum frequency, but they're not necessarily the same color at the same time. When the electron is in a three-dimensional standing wave, the electron cloud's spatial structure is constant as the wave oscillates internally and the electron exhibits no motion at all.

This wave character has profound effects on the electronic structure of atoms. Most significantly, it limits what electrons can do in those atoms. As we saw in Section 9.2, a violin string's one-dimensional standing waves consist only of its fundamental mode (Fig. 9.2.3) and its harmonic modes (Fig. 9.2.5), and a drumhead's two-dimensional standing waves consist only of its fundamental mode and overtones (Fig. 9.2.11). Similarly, an electron's three-dimensional standing waves in an atom consist only of a fundamental mode (Fig. 13.2.4a) and overtone modes (Figs. 13.2.4b–d). Although there are a great many overtone modes available to the electron, their possibilities are nonetheless limited.

The electron standing waves in atoms are known as **orbitals**—a nod to the orbits that electrons would make around atomic nuclei if this were a nonquantum world. Electrons in solids are less focused on specific atomic nuclei, so the standing waves in solids are instead called **levels**—a recognition that each standing wave has an amount, or level, of energy associated with it. We'll see when we examine LEDs in Section 13.3 that each solid's

3 In 1927, American physicists Clinton Joseph Davisson (1881–1958) and Lester H. Germer (1896–1972) showed that electrons travel as waves by observing interference effects when electrons were reflected from different atomic layers in a crystal of nickel metal. When the various electron waves arrived at a detector in phase, the detector found many electrons. When the waves arrived out of phase, the detector found few electrons. Their work was aided by a fortuitous accident in which air entered the experiment's glass vacuum tube. While carefully eliminating oxygen from their nickel sample after that leak, they managed to perfect its crystalline structure, making it possible to observe the interferences.

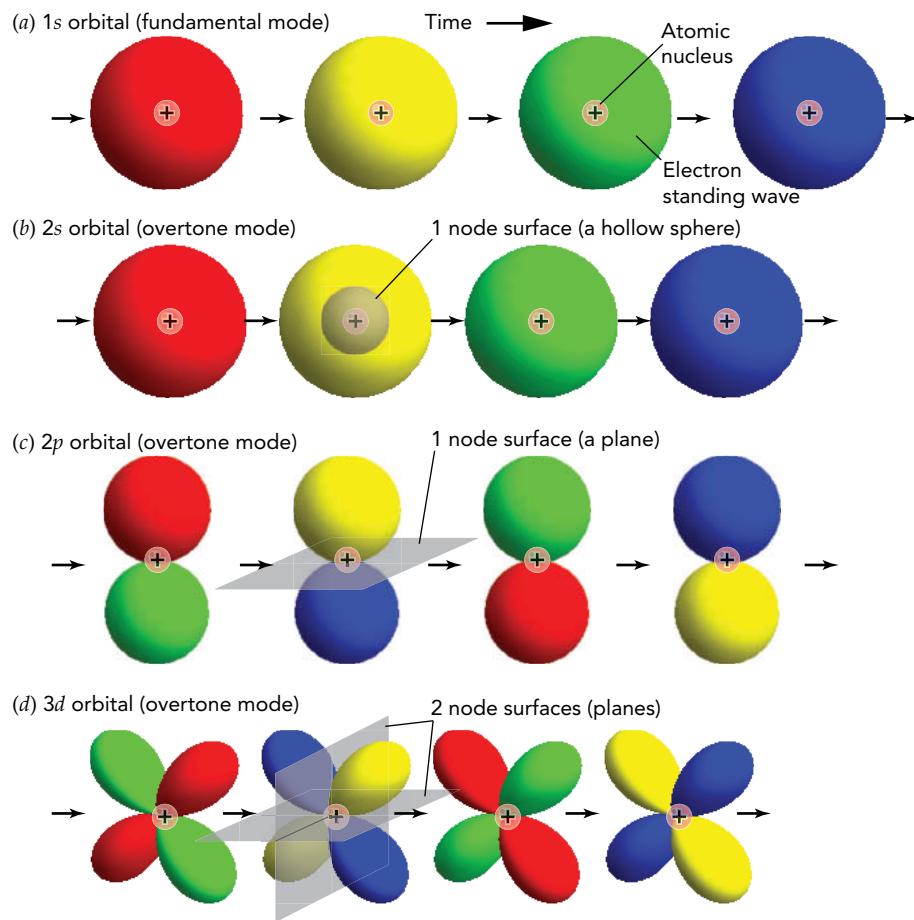


Fig. 13.2.4 In an atom, an electron normally exists in a three-dimensional quantum standing wave or orbital. As time passes, that wave undergoes an internal oscillation, depicted here as the color sequence red, yellow, green, blue, and so on. (a) The 1s orbital is the electron's fundamental mode in an atom. It has the lowest energy and slowest quantum frequency of all the atom's electron modes. (b) The 2s orbital is an overtone mode with somewhat more energy and a faster quantum frequency than the 1s orbital. It has one node surface—a hollow sphere. (c) One of three possible 2p orbitals; these are overtone modes that have energies and quantum frequencies similar to those of the 2s orbital. Each 2p orbital has one node surface—a plane. (d) One of five possible 3d orbitals; these are overtone modes with still greater energies and faster quantum frequencies. Each 3d orbital has two node surfaces—both are planes.

limited level choices determine whether it's a conductor, an insulator, or a semiconductor—the basis for modern electronics. In the present section, we'll see that an atom's limited orbital choices determine the colors of light it can emit or absorb and therefore the characteristics of most discharge lamps.

Check Your Understanding #4: Going Nowhere Fast

When an electron is in an atomic orbital, how often does it shift from one side of the atom to the other?

Answer: An electron in an atomic orbital doesn't move at all, so it never shifts from one side of the atom to the other.

Why: An atomic orbital is an electron standing wave that is spread across the atom. Although that wave is oscillating internally, it makes no spatial motion at all and therefore doesn't shift from one place to another. The electron wave simply oscillates in place.

Assembling Atoms

Another remarkable observation of quantum physics is that every indistinguishable electron must have its own orbital or level, its own unique quantum wave. This law is called the **Pauli exclusion principle**, after its discoverer, Wolfgang Pauli ⁴. The principle applies to a whole class of subatomic particles, the **Fermi particles**, that includes all the basic constituents of matter: electrons, protons, and neutrons. For reasons that lie deep in what is known as quantum field theory, two indistinguishable Fermi particles can never be in the same quantum wave.

THE PAULI EXCLUSION PRINCIPLE

No two indistinguishable Fermi particles ever occupy the same quantum wave.

However, a peculiar property of electrons allows two of them to share an orbital or level. Electrons have two possible internal states, usually called spin-up and spin-down. Because a spin-up electron is distinguishable from a spin-down electron, one spin-up electron and one spin-down electron can share a single orbital or level. However, two electrons is the absolute maximum allowed by quantum physics and the Pauli exclusion principle.

Despite being a wave, an electron in an orbital has a specific total energy—the sum of its kinetic and potential energies. That energy, which depends on the shape and structure of the electron wave, also determines the wave's oscillation frequency. According to quantum physics, the electron's total energy and the oscillation frequency of its wave are exactly proportional to one another; a low-energy electron oscillates slowly, while a high-energy electron oscillates quickly. Each orbital, each quantum standing wave in an atom, has a specific frequency and energy. The electron's fundamental mode (the $1s$ orbital, Fig. 13.2.4a) has the lowest frequency and energy, while the overtone modes have progressively higher frequencies and energies (Figs. 13.2.4b–d).

Although these orbitals are three-dimensional standing waves, they resemble the two-dimensional standing waves of the drumhead shown in Fig. 9.2.9. Like the drumhead's vibrational modes, the orbitals are distinguished from one another by their patterns of nodes—surfaces along which the electron wave has no amplitude. An electron's fundamental mode has no nodes (Fig. 13.2.4a) and has the lowest quantum frequency. Each of the overtone modes, however, has at least one node (Figs. 13.2.4b–d). In general, the more nodes an orbital has, the greater its quantum frequency.

Physicists have come to view these standing waves as abstract placeholders, independent of the electrons that may or may not exhibit them at a given moment. The orbitals are then analogous to seats in an stadium, each of which may or may not be occupied by electrons. Instead of saying that the *electron* exhibits a particular standing wave or orbital, we often say that the *orbital* is occupied by an electron.

At sufficiently low temperature, electrons occupy those orbitals that have the least energy and fill the atom's orbitals from lowest energy on up. In accordance with the Pauli exclusion principle, each orbital accommodates two electrons, one spin-up and one spin-down, until all the atom's electrons have been placed. The chemical nature of a particular atom, and its location in the periodic table of the elements (Fig. 13.2.5), is determined by how many electrons it has and how those electrons fill the available orbitals. Atoms are electrically neutral, so the number of negatively charged electrons in an atom is the same as the number of positively charged protons in its nucleus, its **atomic number**.

There are patterns to orbitals; orbitals may share the same number of nodes or be rotated versions of one another. These patterns among orbitals give rise to the **shells**, that is, groups of orbitals that have similar energies and tend to fill with electrons at about the same time. Although the orbital patterns and shells are complicated by the fact that charged

⁴ Austrian physicist Wolfgang Pauli (1900–1958) rose to fame at 21 by writing an article on relativity that impressed even Einstein. He went on to discover the exclusion principle, a fundamental part of quantum theory. He was renowned for his intensely critical attitude toward new ideas, considering them all “rubbish” until convinced otherwise. Pauli was also quite interested in psychology, corresponding with Carl Gustav Jung and writing a number of articles on the subject.

Fig. 13.2.5 The periodic table of the elements is organized by the filling of electron orbitals. Major shells of orbitals fill with electrons from left to right. The number of electrons in a neutral atom is equal to its atomic number.

The periodic table illustrates the filling of atomic orbitals across different elements. The orbitals are filled in a specific sequence: 1s, then 2s, followed by 2p, then 3s, 3p, 3d, 4s, 4p, 4d, 4f, 5s, 5p, 5d, 6s, 6p, 6d, 7s, 7p, 7d, and finally 8s, 8p, 8d, 9f, 10g, and 11h.

Orbital Filling Order:

- 1s orbital:** Hydrogen (H) is shown filling the 1s orbital.
- 2s orbital:** Helium (He) is shown filling the 2s orbital.
- 2p orbitals:** Boron (B) through Neon (Ne) fill the 2p orbitals.
- 3s orbital:** Lithium (Li) and Beryllium (Be) fill the 3s orbital.
- 3p orbitals:** Sodium (Na) and Magnesium (Mg) fill the 3p orbitals.
- 3d orbitals:** Potassium (K) and Calcium (Ca) fill the 3d orbitals.
- 4s orbital:** Scandium (Sc) through Zinc (Zn) fill the 4s orbital.
- 4p orbitals:** Rubidium (Rb) and Strontium (Sr) fill the 4p orbitals.
- 4d orbitals:** Yttrium (Y) and Zirconium (Zr) fill the 4d orbitals.
- 4f orbitals:** Lanthanum (Ce) through Holmium (Ho) fill the 4f orbitals.
- 5s orbital:** Francium (Fr) and Radium (Ra) fill the 5s orbital.
- 5p orbitals:** Actinium (Ac) and Thorium (Th) fill the 5p orbitals.
- 5d orbitals:** Americium (Am) and Neptunium (Np) fill the 5d orbitals.
- 6s orbital:** Curium (Cm) and Uranium (U) fill the 6s orbital.
- 6p orbitals:** Bismuth (Bi) and Astatine (At) fill the 6p orbitals.
- 6d orbitals:** Rhenium (Re) and Osmium (Os) fill the 6d orbitals.
- 7s orbital:** Oganesson (Og) fills the 7s orbital.
- 7p orbitals:** Nobelium (No) and Lawrencium (Lr) fill the 7p orbitals.

- Alkali Metals: Highly reactive, tend to donate lone s electron so as to empty outer shell
 - Alkaline Earth Metals: Moderately reactive, tend to donate s electrons
 - Transition Metals: Common metals with similar properties, differ only in numbers of d electrons
 - Poor metals: Additional metals with differing properties
 - Metaloids: Intermediate between metals and nonmetals
 - Non-metals: Semiconductors and insulators
 - Halogens: Highly reactive, tend to steal one p electron so as to complete outer shell
 - Noble gases: Non-reactive gases with completed outer shells
 - Lanthanides: Moderately reactive metals, differ only in number of $4f$ electrons
 - Actinides: Moderately reactive metals, differ only in number of $5f$ electrons

electrons influence one another and distort one another's standing waves, many atomic properties are determined simply by which orbitals are occupied by an atom's electrons.

The atomic orbitals are identified primarily by an integer and a letter, both of which relate to their node patterns (Fig. 13.2.4). The integer (1, 2, 3,...) is one more than the number of node surfaces in the orbital (0, 1, 2,...) and the letter (*s*, *p*, *d*, *f*, *g*, *h*,...) indicates how many of those node surfaces pass through the atom's center (0, 1, 2, 3, 4, 5,...). Although *s* orbitals appear one at a time, *p* orbitals appear in groups of three, *d* orbitals in groups of five, *f* orbitals in groups of seven, and so on.

Check Your Understanding #5: **Three's a Crowd**

Suppose that electrons had three different internal states: spin-up, spin-down, and spin-off. How would that change affect the way electrons arrange themselves in atoms?

Answer: Each orbital could then accommodate up to three electrons, so atoms would have fewer occupied orbitals.

Why: The Pauli exclusion principle prevents two indistinguishable electrons from occupying the same quantum wave. With three distinguishable electron states, each orbital could accommodate up to three electrons: a spin-up electron, a spin-down electron, and a spin-off electron. With more electrons settling into the low-energy orbitals, atoms would have fewer occupied orbitals and quite different behaviors.

Neon's Red Glow

A neon atom has 10 electrons, so it takes 5 orbitals to accommodate them. The first two electrons go into the $1s$ orbital, the fundamental mode with zero nodes (Fig. 13.2.4a). Filling the $1s$ orbital completes the first major shell. The next two electrons go into the $2s$ orbital, which has one node surface, a hollow sphere (Fig. 13.2.4b). Finally, neon's last six electrons go into the three $2p$ orbitals, each of which has one node surface, a plane passing through the atom's center (Fig. 13.2.4c). Since filling the $2s$ and $2p$ orbitals completes the second major shell, the neon atom (Ne in Fig. 13.2.5) is chemically inert. That is why neon exists as a gas of individual atoms!

An arrangement of occupied orbitals is called a **state**, and the state we have just described is neon's **ground state**, the state with the lowest possible total energy. The neon atom has other states available, but they all involve additional energy. That's where the discharge enters the picture. When a charged particle collides with a ground-state neon atom, there's a chance that the collision will knock an electron out of its usual orbital into one of the empty orbitals. The neon atom will then be in an **excited state**—it will have extra energy.

Suppose, for example, that a collision has just shifted one of the neon atom's electrons from a $2p$ orbital to a $3p$ orbital. The atom will quickly undergo a series of state-to-state transitions that eventually return it to its ground state. With each transition, the atom will drop to a lower energy state and emit a **photon**, a particle, or **quanta**, of light. The photon carries away the energy released in transitioning from the higher-energy state to the lower-energy state. In all likelihood, one of those transitions will shift the electron from its $3p$ orbital to the $3s$ orbital and produce a photon of red light, the red of a neon sign!

As long as the electron remains in the $3p$ orbital—a particular standing wave—it can't emit an electromagnetic wave. That standing wave has a quantum oscillation; however, the oscillation is internal to the electron and neither the electron wave nor its charge has any overall motion. Since charge must accelerate to emit electromagnetic waves (see Section 12.1), without such overall motion there can be no emission of electromagnetic waves.

When the $3p$ electron begins its transition to the empty $3s$ orbital, however, its quantum wave starts to move. That wave is no longer one motionless standing wave but is, instead, composed of two partial waves: part $3p$ orbital and part $3s$ orbital. Because those two partial waves have different energies and different quantum frequencies, they experience time-dependent interference effects. The patterns of constructive and destructive interference change so that the overall electron wave moves rhythmically back and forth and the atom begins to emit an electromagnetic wave. By the time the transition is complete and the electron is entirely in the $3s$ orbital, the atom has emitted a photon of red light. Like an electron, light travels as a *wave* but behaves as a *particle* when you try to locate it. While it's being emitted or absorbed by an atom, light exhibits its particle nature.

We've seen that hot objects emit light, a process called **incandescence**, but here the neon atom emits light without heat, a process called **luminescence**. That luminescence is the result of a **radiative transition**, a transition between quantum states in which a photon of light is emitted or absorbed. In this case, the radiative transition emits a photon that carries away the energy released when the excited neon atom's $3p$ electron shifts into the empty $3s$ orbital. The difference in energy between those two states determines the photon energy, which in turn determines the frequency and color of the photon's light wave.

According to quantum physics, a photon's frequency is proportional to its energy. Specifically, the photon's energy is equal to the frequency of its electromagnetic wave times a fundamental constant of nature known as the *Planck constant*. This relationship between energy and frequency can be written as a word equation:

$$\text{energy} = \text{Planck constant} \cdot \text{frequency}, \quad (13.2.1)$$

in symbols:

$$E = h \cdot v,$$

and in everyday language:

Ultraviolet light and X-rays can injure your skin and tissues because each particle of high-frequency light carries a great deal of energy.

First used by German physicist Max Planck (1858–1947) in 1900 to explain the light spectrum of a hot object, the **Planck constant** has a measured value of $6.626 \times 10^{-34} \text{ J} \cdot \text{s}$. Also, although we have just encountered the Planck constant and Eq. 13.2.1 in the context of light waves, it actually applies to all quantum waves. For example, the quantum frequency of an electron is related to its energy by Eq. 13.2.1.

The value of the Planck constant is so tiny that even a photon of 10^{15} -Hz ultraviolet light has an energy of only $6.626 \times 10^{-19} \text{ J}$. A typical beam of light thus contains so many photons that you can't see that they're arriving as particles. However, the energy in a single ultraviolet photon is substantial on a molecular scale. It can damage a molecule in your skin, contributing to a sunburn and inducing your skin to tan as a defensive response. X-rays have even higher frequencies, and their energetic photons can cause more severe molecular damage.

A neon atom can only emit photons corresponding to energy differences between two of its states, a constraint that severely limits its light spectrum. Neglecting nuclear issues that we'll discuss in Chapter 15, all neon atoms are identical and emit the same characteristic spectrum of light. The visible part of that spectrum is dominated by the warm red glow of photons emitted when electrons shift from $3p$ to $3s$ orbitals, which is why a neon sign glows red.

Since atoms of different elements have different numbers of electrons and different states, each emits its own unique spectrum of light after being excited. Copper atoms emit a blue-green spectrum, strontium atoms a deep red, and sodium a bright yellow-orange. Chemists, astronomers, and manufacturers rely on those emission spectra for information and applications. As we will soon see, so do scientists and engineers working on illumination.

Check Your Understanding #6: Colorful Atoms

Many fireworks involve brilliantly colored lights. How do the atoms in burning chemicals produce particular colors of light?

Answer: When fire adds energy to an atom and shifts it to an excited state, it emits photons in order to return to its ground state. Each photon's energy, frequency, and color are determined by energy differences between the atom's states.

Why: Each color in fireworks is produced by a particular type of atom. Some atoms, such as strontium and lithium, emit mostly red light as they return to their ground states. Barium atoms emit green light, copper atoms emit blue-green light, and sodium atoms emit yellow-orange light.

Check Your Figures #1: The Particles in a Radio Wave

How much energy is carried by a photon from a 1000-kHz AM radio station?

Answer: The photon carries $6.626 \times 10^{-28} \text{ J}$.

Why: Since the Planck constant is $6.626 \times 10^{-34} \text{ J} \cdot \text{s}$ and the frequency of the radio wave is 10^6 Hz or 10^6 cycles/second, the energy per photon is given by Eq. 13.2.1 as

$$\text{energy} = (6.626 \times 10^{-34} \text{ J} \cdot \text{s}) \cdot (10^6 \text{ cycles/s}) = 6.626 \times 10^{-28} \text{ J}.$$

This energy is so small that it's virtually impossible to observe the particulate character of radio waves.

Fluorescent Lamps

If neon tubes appeal to you for illumination, you probably march to your own drummer. Most people opt for a somewhat better simulation of sunlight in their discharge lamps. As an energy-efficient source of artificial sunlight, though, it's hard to beat fluorescent lamps.

At the heart of a fluorescent lamp is a narrow glass tube filled with argon, neon, and/or krypton gases at about 0.3% of atmospheric density and pressure. The tube also contains a few drops of liquid mercury metal, some of which evaporates to form mercury vapor. About 1 in every 1000 gas atoms inside the tube is a mercury atom, and it's these mercury atoms that are responsible for the light.

Like a neon sign tube, a fluorescent lamp uses a discharge in its gas to produce light. Although fluorescent lamps occasionally use high voltages to initiate their discharges, most operate their discharges at household voltages and must rely on alternative techniques to render their gases conducting. These low-voltage lamps normally heat their electrodes so that thermal energy ejects electrons from their surfaces into the gas. Regardless of how the discharge is started, however, the result is current flowing through the gas and the emission of light.

The fluorescent lamp has a problem, though. While its mercury atoms emit most of the light in its discharge, that light is almost entirely ultraviolet. The final radiative transition that returns each mercury atom to its ground state ($6p \rightarrow 6s$) releases a large amount of energy and produces a photon with a wavelength of 254 nm. This light can't go through the glass walls of the tube, and you couldn't see it if it did. So the fluorescent lamp converts it into visible light with the help of phosphor powder on the inside of the glass tube.

Phosphors are solids that *luminesce*—they emit light—when something transfers energy to them. Their behavior is similar to that of an atom; an energy transfer shifts the phosphor from its ground state to an excited state, and it then undergoes a series of transitions that return it to its ground state. Some of those transitions are radiative ones and emit light.

In a fluorescent lamp, the phosphor is excited by ultraviolet light. This excitation or energy transfer is actually a radiative transition, but one in which the photon is absorbed by the phosphor; one of its electrons makes a transition from a lower-energy level to a higher-energy level. During this absorption, the light's electric field pushes the electron's wave back and forth rhythmically until it shifts to the new level. The photon disappears, and the phosphor receives its energy.

Once the phosphor is in an excited state, its electrons begin to make transitions back to their ground-state levels. Most of those transitions radiate visible light, the light that you see when you look at the lamp. However, some of the transitions radiate invisible infrared light or cause useless vibrations in the phosphor itself. Despite this wasted energy, phosphors are relatively efficient at turning ultraviolet light into visible light, a process called **fluorescence**.

The phosphors in a fluorescent lamp are carefully selected and blended to fluoresce over a broad range of visible wavelengths. That blending is necessary because, like atoms, each phosphor fluoresces with a characteristic spectrum of light that's determined by the energy differences between its levels. Several different phosphors are needed to produce the rainbow spectrum of white light. While some advertising and novelty lamps use brightly colored, unblended phosphors, fluorescent lighting tubes use phosphor mixtures that imitate white thermal radiation sources such as the sun or incandescent lightbulbs.

Because compact fluorescent lamps typically produce 60 to 100 lumens per watt, about 4 to 6 times the luminous efficiency of incandescent lightbulbs, you can save a great deal of energy by replacing incandescent lightbulbs with compact fluorescents. In fact, incandescent lightbulbs are being phased out because of their poor energy inefficiency. To ease the acceptance of compact fluorescent lamps by a reluctant public, scientists and engineers have had to improve the quality of their light.

Early fluorescent lamps used a phosphor blend known as "daylight," which emitted far too much blue light and made everything look cold and medicinal. Daylight lamps rarely

found their way into homes. The first improvement came with the developments of the “cool white” and “warm white” phosphor blends. Cool white resembled sunlight, and warm white resembled light from an incandescent lightbulb. However, even these improved lamps didn’t emit true white-light spectra—the blackbody spectra of the hot sun or the hot filament of an incandescent lightbulb. Objects looked a little off-color when illuminated by these lamps, so people were hesitant to use them in place of incandescent lightbulbs. Photographers and cinematographers wanting to capture true colors were particularly troubled by fluorescent lighting of that era.

However, the latest generation of fluorescent lamps uses phosphor blends that produce nearly perfect blackbody spectra. Many of those blends are even sold according to their color temperatures (see Section 7.3). For example, a 5100 K compact fluorescent lamp emits light that is almost indistinguishable from the thermal radiation emitted by a black object heated to 5100 K. That 5100 K color temperature is approximately the color temperature of sunlight averaged over a day. (Sunlight’s color temperature is affected by the atmosphere and is highest at noon and lowest at dawn and dusk.) If you like the color of sunlight, use 5100 K compact fluorescents.

People who prefer the warm look of incandescent lighting should choose compact fluorescent lamps with a 2700 K color temperature. These lamps imitate the blackbody spectrum of a 2700 K tungsten filament almost perfectly. Because this 2700 K blackbody spectrum is richer in reds and oranges, it has a comforting, warm character to it. If you can’t make up your mind between 5100 K sunlight and 2700 K incandescent light, you can choose many color temperatures in between. Even photographers and cinematographers are now using fluorescent lights, to the delight of actors and news anchors who used to sweat under the intense heat of powerful, energy-inefficient incandescent lamps.



Check Your Understanding #7: Sorry, We Forgot the Coating

If there were no phosphor coating on the inside of a fluorescent tube, what would you see when the lamp operated?

Answer: You would see only a dim (blue-white) glow from the lamp because most of the light produced by the mercury discharge is invisible ultraviolet light.

Why: Without its phosphor coating, a fluorescent tube produces very little visible light. The phosphor coating is needed to convert the discharge’s ultraviolet light into visible light. Even if you could see ultraviolet light, you would see very little of it leaving the tube. The tube is made of glass, which absorbs virtually all ultraviolet light with wavelengths shorter than 350 nm. Even in a normal fluorescent tube, any ultraviolet not converted to visible light by the phosphors is absorbed by the glass tube.

A Few Practical Issues

A fluorescent tube needs a substantial electric field inside it to keep its electrons moving forward. Since this electric field is proportional to the voltage drop through the tube, longer tubes need higher voltages. Power-line voltages (110 to 240 V) are appropriate for tubes up to 3 m (10 ft) in length, but the longer and frequently colored fluorescent tubes used in artwork or advertising require much higher voltages.

To keep electrons from pushing one another into its walls, a fluorescent tube must also contain positively charged mercury ions. The discharge naturally produces these ions during particularly energetic collisions. The result, a gaslike mixture of positively charged ions and negatively charged electrons, is called a **plasma**. A plasma is distinct from a gas because its charged particles exert forces on one another over considerable distances. All operating discharge lamps contain plasmas, including the neon sign that we discussed earlier.

Since a typical fluorescent lamp must heat its electrodes red-hot to form its plasma, it runs current through filaments at each end of its tubes (Fig. 13.2.6). Once the discharge is

operating, impacts by electrons may keep the filaments hot enough to sustain the plasma and allow the heating current to be switched off. However, some fluorescent fixtures can be dimmed, in which case electron heating alone won't sustain the plasma. Dimmable fluorescent fixtures must continue to pass heating currents through their filament/electrodes.

Each filament/electrode is coated with a material that helps it emit electrons into the gas. Unfortunately, that coating is thin and easily damaged by a process called **sputtering**, in which positive mercury ions from the plasma collide with the coating and chip away its atoms. Once enough of the coating has been removed, the lamp won't be able to sustain a discharge and it will need to be replaced. Because sputtering is particularly severe during start-up, most fluorescent lamps fail after 10,000 to 40,000 starts. Since it takes energy to construct each fluorescent lamp, turning a lamp off for only a short period, a few minutes or less, won't actually save energy in the long run. When you'll be returning to the room in a couple of minutes, leave the fluorescent lights on. Otherwise, turn them off.

Fluorescent lamps operate most efficiently and are at their brightest when their internal temperatures reach about 40 °C. Below that temperature, the vapor pressure of mercury is too low and too much of the lamp's tiny store of mercury is liquid rather than gas. That's why most fluorescent lamps are somewhat dim when you first turn them on and they brighten as they warm up from room temperature to about 40 °C. It takes time for the mercury vapor to reach its optimum density. Although this warm-up period is objectionable to some people, it's a small price to pay for a vastly more energy-efficient lighting.

It would be great if fluorescent lamps could be based on a substance less toxic than mercury. However, mercury is uniquely well-suited to those lamps: its vapor pressure near room temperature is nearly ideal for an electric discharge, it converts electric energy into ultraviolet light with great efficiency, and it isn't damaged or consumed as the lamp operates. Much more important than eliminating mercury from fluorescent lamps is recycling those lamps after they fail. It's not hard to extract and reuse the mercury, and it should be done routinely.

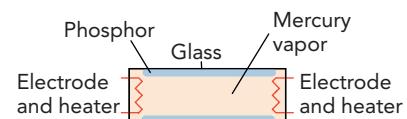


Fig. 13.2.6 In a hot-electrode fluorescent tube, the electrodes are actually filaments and are heated by running currents through them.

Check Your Understanding #8: Getting Red-y

Why do the ends of some fluorescent lamps glow red during starting?

Answer: To start many lamps, their electrodes must be heated red-hot. You can often see light emitted by these hot filaments/electrodes.

Why: In most fluorescent fixtures, the filaments are heated to high temperatures just before the discharge starts. Thermal energy then ejects electrons from the filament so that they can carry current through the gas.

Mercury, Metal-Halide, and Sodium Lamps

Whereas a low-pressure mercury discharge emits mostly ultraviolet light, a high-pressure mercury discharge emits more visible light than ultraviolet light. This change occurs because ultraviolet light becomes trapped in densely packed mercury atoms and only visible light is able to escape from the discharge.

Known as **radiation trapping**, this effect occurs because mercury atoms absorb 254-nm photons just as well as they emit them. The same radiative transition that causes a mercury atom to emit a 254-nm photon ($6p \rightarrow 6s$) can also run backward to absorb that photon ($6s \rightarrow 6p$). In a dense gas of mercury, whenever one mercury atom emits a 254-nm photon, another mercury atom snaps that photon up. So while the discharge keeps pouring energy into the mercury atoms, they can't get rid of it as 254-nm photons. Instead, they emit most of their energy through radiative transitions between other excited states. Since this light is much less likely to be captured by other mercury atoms, it emerges from the lamp as bluish visible light.

When you first turn on a high-pressure mercury lamp, most of the mercury is liquid and the pressure is low. The lamp starts like a small fluorescent tube without phosphor, so you see very little visible light. However, the tube is designed to heat up during operation so that the liquid mercury evaporates to form a dense gas. As the gas pressure rises, the tube's color changes. The 254-nm photons become trapped and photons emitted by many other radiative transitions begin to dominate. The lamp emits brilliant, blue-white light.

To make a lamp that's a little less bluish, some high-pressure mercury lamps contain additional metal atoms. These atoms are introduced into the lamps as metal-iodide compounds, so they are often called metal-halide lamps. Sodium, thallium, indium, and scandium iodides all contribute their own emission spectra to the outgoing light and help to strengthen the red end of its spectrum. They give metal-halide lamps a warmer color than pure mercury lamps. Both high-pressure mercury and metal-halide lamps produce about 50 to 60 lumens per watt, making them almost as energy efficient as fluorescent lamps.

Pure sodium lamps resemble mercury lamps, except that they use sodium atoms. Sodium is a solid at room temperature, so both low- and high-pressure sodium lamps must heat up before they begin to operate properly. A low-pressure sodium lamp is extremely energy efficient because its 590-nm light comes directly from a sodium atom's strongest radiative transition. That transition is from sodium's lowest excited state to its ground state ($3p \rightarrow 3s$). Low-pressure sodium lamps produce almost 200 lumens per watt, which is why highways were frequently illuminated by their yellow-orange glow.

However, this monochromatic illumination is unpleasant and permits no color vision at all. Although it's marginally acceptable on a highway, you wouldn't want it near your home. That's why people buy high-pressure sodium lamps for home use (Fig. 13.2.7). For a small decrease in energy efficiency, to about 150 lumens per watt, you get a big improvement in color.

Remarkably enough, the 590-nm emission itself smears out at high pressure to cover a wide range of wavelengths, from yellow-green to orange-red. This spreading occurs because of the many collisions suffered by the densely packed sodium atoms as they try to emit 590-nm light. These collisions distort the atomic orbitals so that the photons emerge with somewhat shifted energies. Overall, a high-pressure sodium lamp emits remarkably little light exactly at 590 nm because the ground-state sodium atoms trap that light; they run the ($3p \rightarrow 3s$) transition backward and absorb the photons ($3s \rightarrow 3p$). This trapping is so effective that there is actually a hole in the lamp's spectrum right at 590 nm.

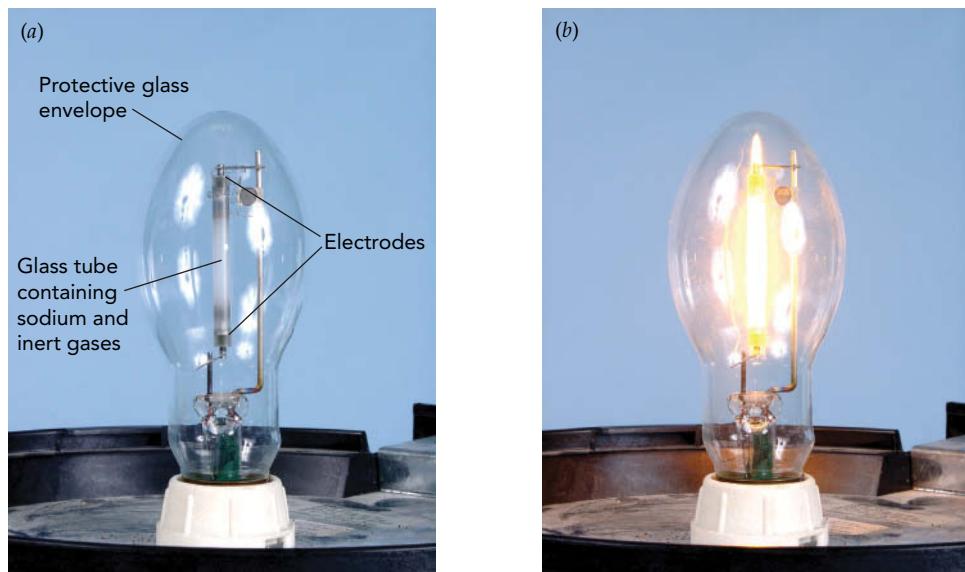


Fig. 13.2.7 (a) The active component of a high-pressure sodium vapor lamp is a small translucent tube. (b) As the lamp warms up, sodium metal in the tube evaporates to form a brilliant yellow discharge. The dense vapor of sodium atoms in the tube traps the 590-nm light so that the lamp emits a richer spectrum of wavelengths and a less monochromatic glow than a low-pressure sodium lamp.

Courtesy Lou Bloomfield

High-pressure discharge lamps suffer from a problem not found in low-pressure lamps: they're difficult to start when hot. It's much harder to initiate a discharge in a high-pressure gas than in a low-pressure gas, so they all start at low pressure and then evolve to high pressure. If the discharge in a high-pressure mercury, sodium vapor, or metal-halide lamp is interrupted and loses its plasma, the lamp must cool down before it can be restarted.

Check Your Understanding #9: Slow Glowing

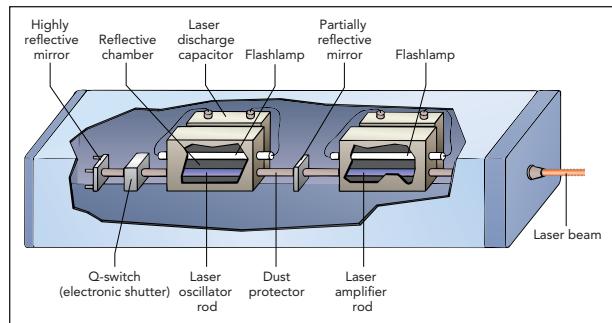
At dusk, a mercury streetlight turns on. It glows dimly at first and gradually increases in brightness. What is happening during this warm-up period?

Answer: The small high-pressure discharge tube is heating up, vaporizing more mercury.

Why: When the streetlight first turns on, the pressure of mercury atoms inside it is low and it emits very little visible light. As the discharge heats the tube, more mercury atoms evaporate until eventually all the mercury inside the tube is gaseous.

SECTION 13.3

LEDs and Lasers



Today's electronic devices have so many LEDs (light-emitting diodes) that our homes are ablaze with tiny colored lights at night. LEDs are quickly replacing lightbulbs in flashlights, and LED illumination is on the rise. LEDs combine the light-producing quantum transitions of discharge lamps with the solid-state physics of modern electronics. In this section, we'll see how electrons behave in solids and how those solids can be tailored to emit light.

We'll also take a look at lasers. Since their invention in the late 1950s, lasers have found countless uses, from cutting metal and clearing human arteries to surveying land and playing CDs and DVDs. However, lasers are more than just novel applications of old ideas. Instead, they bring together quantum and optical physics to produce a new type of light. This light is radically different from the light produced by incandescent and

fluorescent lamps, and its properties make it particularly useful for many applications. In this section, we'll examine the nature of this new light and the ways in which lasers produce it.

Questions to Think About: Why is laser light usually brightly colored? Why does laser light often appear as a narrow beam? In movies, "lasers" are often shown to emit bright streaks of light that can be dodged if one jumps quickly enough; is that view realistic?

Experiments to Do: Although lasers are household objects, the ones in CD or DVD players and laser printers are relatively inaccessible. If you don't own a laser pointer, look at a store barcode scanner. The scanning system contains a gas or solid-state laser that emits a very narrow beam of bright red light. This system also contains a rotating mirror or holographic disk that directs the beam as a pattern of thin stripes onto anything that passes through the scanner. A light sensor inside the window watches for this moving beam of light to travel across a label. If you look down at the laser light emerging from the scanner's window or observe the spot of a laser pointer on the wall, you'll see both the purity of its color and its strange speckled character; the light appears to consist of tiny light and dark speckles. These speckles are caused by interference effects, which are extremely pronounced in the ordered laser light. This light also appears unusually bright and, as when looking at the sun, you should keep your gaze brief to avoid eye injury.

WARNING

Laser light is dangerous. Because laser light can be extremely bright and can focus so tightly, it presents a serious eye hazard. A laser beam entering your eye will focus to a tiny spot on your retina and may cause rapid and permanent injury. Although lasers up to Class III are relatively eye-safe, they rely on your natural blink reflex to protect you. You should never stare into any laser. Class IV lasers are not eye-safe, and their light should never enter your eyes at all.

Electrons in Solids

An LED is a *solid-state* device—that is, it's a solid object that luminesces. Like light from a neon sign, light from an LED is produced when an electron shifts from a higher-energy quantum state to a lower-energy quantum state by way of a radiative transition. In the neon sign, those quantum states involve *orbitals*, the electron standing waves of an atom. In an LED, however, those quantum states involve *levels*, the electron standing waves of a solid. To understand how an LED emits light, we must start by understanding levels.

Solids are limited objects, and the basic quantum waves for electrons in limited objects are always standing waves. The electron standing waves in atoms are called orbitals because the positively charged nucleus wraps the negatively charged waves around itself in a manner suggesting orbits. The electron standing waves in solids are less focused on specific atomic nuclei and are called levels instead. Although the nuclei are present, of course, each electron is attracted to all the nuclei at once and that changes the nature of the standing waves.

As we did with orbitals, we'll view levels as abstract placeholders, independent of the electrons that may or may not exhibit them at a given moment. The levels in a solid are then analogous to seats in a theater, each of which may or may not be occupied right now. Instead of saying that the *electron* is experiencing a particular standing wave or level, we say that the *level* is occupied by an electron. In this reversed perspective, the level plays the more important role.

Like an electron in an orbital, an electron in a level has a specific total energy—the sum of its kinetic and potential energies. That energy, which depends on the shape and structure of the electron wave, also determines the wave's quantum frequency (see Eq. 13.2.1).

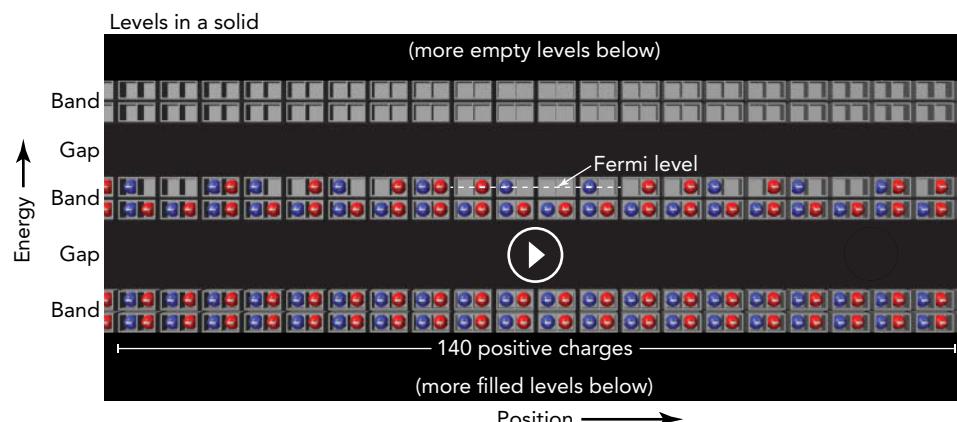
A solid contains an enormous number of electrons, and there are always plenty of levels around to accommodate those electrons. But which levels do they occupy?

At sufficiently low temperature, electrons occupy those levels that have the least energy. For thermodynamic reasons, the electrons settle into the lowest-energy levels available, two electrons per level (recall the Pauli exclusion principle). By the time all the electrons have been accommodated, they fill the levels up to a certain maximum energy. Halfway between the highest filled level and lowest unfilled level is the **Fermi level**, a hypothetical level that defines the top of this *Fermi sea* of electrons. The energy an electron would have in that hypothetical level is called the **Fermi energy**.

Our theater analogy can again provide insight into this level-filling process. It is analogous to what happens at a popular show: the seats fill from the orchestra level on up—everyone wants to sit in the lowest (and closest) seat. When the show starts, people have filled all the seats up to a certain highest seat. Halfway between that last filled seat and the next unfilled seat is the hypothetical Fermi seat.

If we represent the levels graphically by boxes and arrange them vertically according to energy (Fig. 13.3.1), then the levels (boxes) below the Fermi level contain two electrons

Fig. 13.3.1 A conceptual representation of the levels (electron quantum standing waves) in a solid. Each level is a box holding at most one spin-up electron (blue) and one spin-down electron (red). The horizontal axis is position and the vertical axis is the energy of an electron in a level. Because levels form out of atomic orbitals, levels made from similar orbitals have similar energies and group into bands. For this portion of the solid to be electrically neutral, 140 positive charges in its nuclei must be cancelled by electrons in these levels, so 140 electrons settle into them.



each, while those above the Fermi level are empty. Although thermal energy complicates this picture somewhat by shifting electrons about near the Fermi level, we can ignore that detail near room temperature or below.

Since levels are standing waves, they don't have sharply defined locations in space. However, we can safely imagine that each level places its electrons near a particular location in the solid, as shown in Fig. 13.3.1. Although this picture is somewhat oversimplified, it's accurate enough to illustrate much of the physics of charge motion in materials.

Of course, electrons aren't the only charged particles in a solid. The atoms also have positively charged nuclei. But those nuclei are essentially immobile and rarely participate in the flow of electricity.

Check Your Understanding #1: Taking It to a Higher Level

If you add one extra electron to an otherwise neutral metal ball, into which level will that electron go?

Answer: The electron will go into the lowest-energy empty level, the one just above the Fermi level.

Why: Since each electron goes into the lowest-energy level that's available, this new electron will fill the level just above the Fermi level. Actually, if the neutral metal ball has an odd number of electrons, then the Fermi level contains only one electron. In that case, the new electron will fill the other opening in the Fermi level.

Metals, Insulators, and Semiconductors

The levels in a solid occur in groups called **bands**. Each band corresponds to standing waves with a particular type of structure. Since the levels in a band involve similar waves, they also involve similar energies. Between these bands of levels there are sometimes **band gaps**, ranges of energy in which no levels exist. The solid does not and cannot contain electrons with energies that lie within a band gap.

Bands and band gaps are what distinguish metals, insulators, and semiconductors. When the Fermi level is located in a band gap, it can prevent the electrons in a solid from responding to outside forces. To see how that happens, let's examine first a metal and then an insulator.

In a **metal**, the Fermi level lies in the middle of a band (Fig. 13.3.2). Because the band's empty levels are just above its filled levels, very little energy is needed to shift electrons from the filled levels to empty levels. This feature allows the metal to conduct electricity. When you subject the metal to a voltage difference, giving its left side a higher voltage than its right side, the resulting electric field points toward the right. That field pushes the metal's negatively charged electrons toward the left and they begin to move left. They move by shifting from filled levels to empty levels (Fig. 13.3.2), obtaining the energy needed to reach those empty levels from the work done on them by the electrostatic forces. Overall, electrons enter the metal from its right and leave from its left, so the metal conducts electricity!

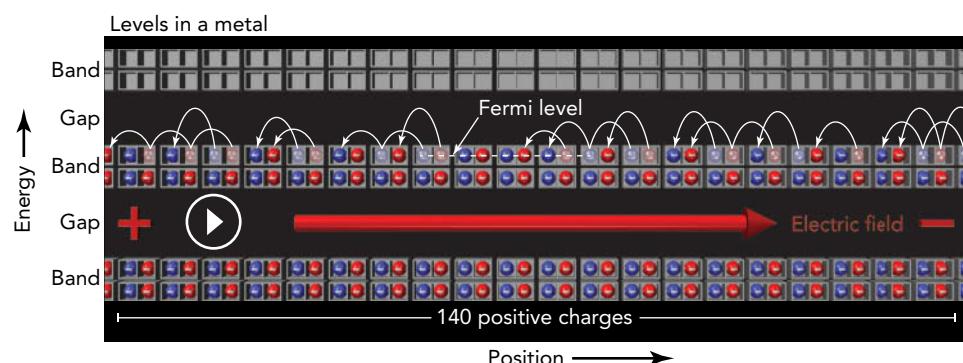
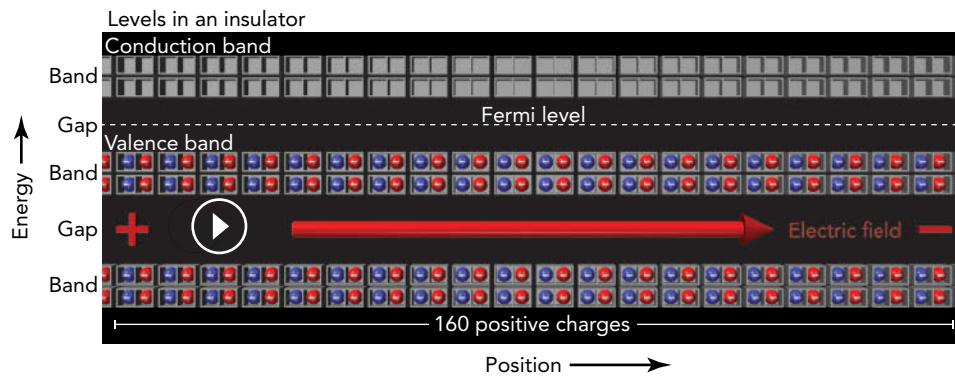


Fig. 13.3.2 In a metal, electrical neutrality is reached while electrons are still filling a band of levels and the Fermi level lies in that partially filled band. Electrons near the Fermi level can shift easily to nearby empty levels. When a voltage gradient produces an electric field in the metal, electrons in the partially filled band move opposite the field and the metal conducts electric current.

Fig. 13.3.3 In an insulator, electrical neutrality is reached just as electrons finish filling a band of levels and the Fermi level lies between that filled conduction band and the empty valence band well above it in energy. Electrons in the conduction band don't have enough energy to shift into empty levels in the valence band. When a voltage gradient produces an electric field in the insulator, electrons in the full valence band can't move and the insulator doesn't conduct electric current.



In our theater analogy, a metal is a theater in which only about half the ground-floor seats are filled. If you ask people in the theater to begin shifting left, those near the top of the occupied seats can shift about easily. Each finds an empty seat nearby on the left and moves over. New people are then able to enter the theater from the right while others leave the theater from the left. This metal theater is conducting people.

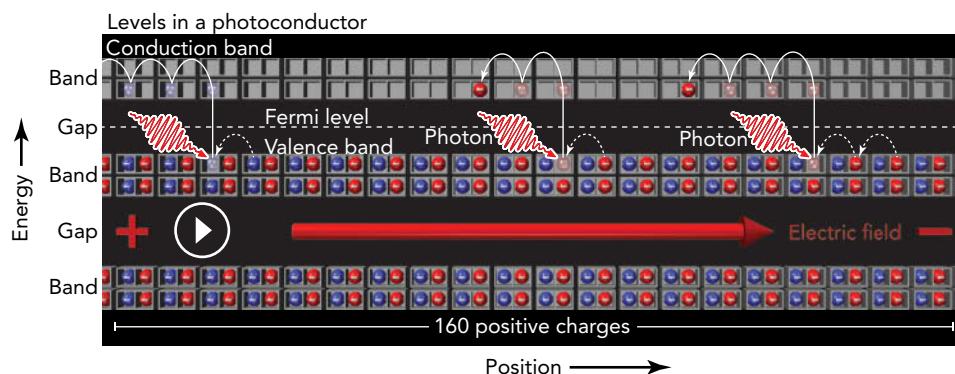
Unlike the situation in a metal, in an **insulator** the Fermi level lies in the middle of a band gap, between the top of one band and the bottom of another band (Fig. 13.3.3). With no easily accessible empty levels available, a great deal of energy is required to shift electrons from the filled levels to empty levels. When you subject the insulator to a voltage difference, higher voltage on the left, the resulting electric field points toward the right and pushes the insulator's electrons toward the left. Those electrons, however, have no empty levels nearby and are thus unable to move. To shift into one of the empty levels in the upper band, an electron in the lower band would need more energy than it can get from the electrostatic forces. Since no net charge flows through the insulator, it doesn't conduct electricity!

In our theater analogy, an insulator is a theater in which all the ground-floor seats are full and the balcony seats are empty. When you ask people in this theater to begin shifting left, they can't do it. All the ground-floor seats to the left are filled, and people can't reach the balcony to make use of its empty seats. This insulator theater is unable to conduct people.

In a metal, the band of levels containing the Fermi level is only partially filled, and electrons can easily shift from filled to unfilled levels. In an insulator, the band below the Fermi level—the **valence band**—is full and the band above the Fermi level—the **conduction band**—is empty, making such shifts extremely difficult.

Even in an insulator, however, an electron can shift from a **valence level** (a level in the valence band) to a **conduction level** (a level in the conduction band) if something provides the necessary energy. One such energy source is light. When an insulator is exposed to the right type of light, that light can shift electrons from the material's valence band to its conduction band (Fig. 13.3.4).

Fig. 13.3.4 When light strikes an insulator, its photons may carry enough energy to shift electrons from the insulator's valence band to the conduction band. If so, electrons can then use the two partially filled bands to move through the material in response to an electric field. The illuminated insulator becomes an electrical conductor, photoconductor.



Once electrons appear in the normally empty conduction band and empty levels appear in the normally full valence band, electrons can respond to electric fields. They can shift from filled levels to nearby empty levels and thus travel through the material. Electrons can then enter the material from one side and leave from the other, so the material conducts electricity. Because light has made this insulator a conductor, we call the material a **photoconductor**.

Turning again to our analogy, light's role in the insulator theater is performed by a playful gorilla that walks about the ground floor, tossing patrons into the balcony. With some of the ground-floor seats suddenly empty and some of the balcony seats suddenly occupied by dazed theatergoers, the crowd can now respond to your request to move left. The gorilla has made the insulator theater a conductor of people—what you might call a “gorillaconductor.”

Not all light causes photoconductivity in an insulator. As we saw in Section 13.2, light is emitted and absorbed as particles called photons. Each photon's energy is proportional to its frequency; the higher the frequency of light, the more energy each of its photons contains. To shift an electron across the large band gap in a typical insulator, high-energy, high-frequency light is needed; the insulator must be exposed to violet or even ultraviolet light.

Nature also provides materials with smaller band gaps that can be crossed with the help of low-energy, low-frequency red or even infrared light. These materials are called **semiconductors** because their properties lie somewhere between those of conductors and insulators. Semiconductors have small band gaps, making it relatively easy for light, heat, or other types of energy to shift electrons between valence and conduction levels. In our analogy, a semiconductor theater is an insulator theater with a low balcony, so that even a baby gorilla can toss people into it.

For more than half a century, scientists and engineers have worked with semiconductors to produce an astonishing array of electronic devices. By carefully tailoring the shapes and chemical compositions of semiconducting materials such as silicon, germanium, and gallium arsenide, they have created virtuoso instruments for electron waves in solids that are every bit as remarkable as the instruments for musical waves found in great orchestras. Of all these electronic instruments, the simplest is the semiconductor diode.

Check Your Understanding #2: Stop and Go Shopping

In many grocery checkout counters, a conveyor belt carries food to the register but stops when the food reaches the end and blocks a beam of light. How might a photoconductor be used to sense this blockage?

Answer: The beam of light shines on the photoconductor, allowing it to carry current and turn on the conveyor belt's motor. When food blocks the light beam, the photoconductor becomes an electrical insulator and the belt stops moving.

Why: Photoconductors are commonly used in light sensors. Light allows the photoconductor to carry current, and that current can be used to operate machinery, trigger a burglar alarm, or turn lights on or off. In the present case, it turns on the conveyor belt's motor.

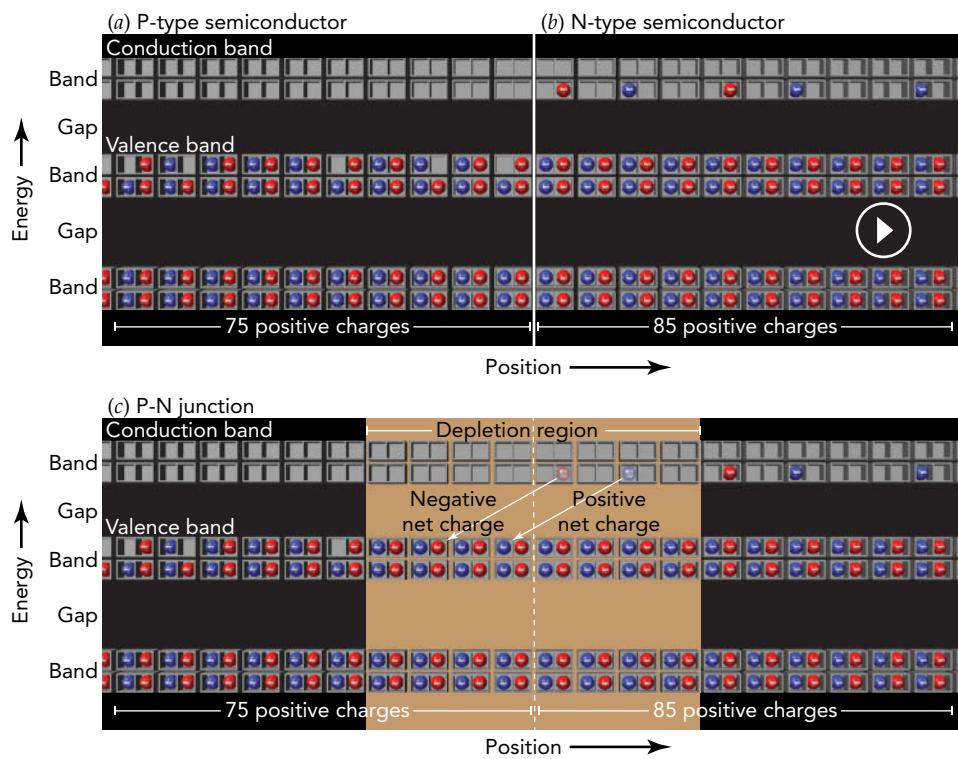
Diodes

A **diode** is a one-way device for current; it allows current to flow through it in one direction but not in the other. Although diodes have taken many forms over the years, the diodes in virtually all modern electronic devices are built from semiconductors.

A basic semiconductor diode is made by joining together two different semiconductors. These two semiconductors have been modified so that they don't have perfectly filled valence levels and perfectly empty conduction levels. Instead, they're **doped** with atomic impurities that either create a few empty valence levels (**p-type semiconductor**, Fig. 13.3.5a) or place a few electrons in the conduction levels (**n-type semiconductor**, Fig. 13.3.5b). These empty valence levels or conduction-level electrons allow p-type and n-type semiconductors to conduct electricity.

Fig. 13.3.5 (a) A p-type semiconductor has slight shortage of positive charges in its nuclei and becomes electrically neutral before its valence band fills completely with electrons. (b) An n-type semiconductor has a slight excess of positive charges in its nuclei and becomes electrically neutral after partly filling its conduction band with electrons.

(c) When the p- and n-type semiconductors touch, conduction-level electrons migrate from the n-type semiconductor to the p-type semiconductor, creating a thin, electrically polarized depletion region at the p-n junction.



When a piece of p-type semiconductor touches a piece of n-type semiconductor, however, something remarkable happens—a **p-n junction** forms at the place where the two meet (Fig. 13.3.5c). To reduce their potential energies, higher-energy conduction-level electrons from the n-type semiconductor migrate across the p-n junction to fill empty lower-energy valence levels in the p-type semiconductor. It's an example of static electricity (see Section 10.1); when two dissimilar materials touch, the material that is hungrier for electrons (the p-type semiconductor) steals some electrons from its partner (the n-type semiconductor).

As the electron migration proceeds, the n-type semiconductor acquires a positive net charge because it now has fewer electrons than positive charges. The p-type semiconductor acquires a negative net charge because it now has more electrons than positive charges. Electrostatic forces from this separated charge oppose the further migration of electrons across the junction and gradually bring that migration to a halt. Everything is then in equilibrium.

Near the p-n junction, there is now a **depletion region**, an area in which electron migration has emptied all the conduction levels and filled all the valence levels. With no conduction-level electrons or empty valence levels left, the depletion region can't conduct electricity and charge can't move across the p-n junction. The depletion region is an insulator, and the two pieces of semiconductor have become a diode.

In our theater analogy, the p-n junction is analogous to a theater with two halves. In the left or “p-type” half, the balcony is empty and even the ground floor has some empty seats. In the right or “n-type” half, the ground floor is filled and there are even a few people in the balcony. Since these two halves touch, people in the right balcony notice the empty seats in the left ground floor, and a few of them near the center of the theater clamber down from the right balcony to the left ground floor to take advantage of the better seats. Near the center of the theater, the ground floor is now filled and the balcony is empty, forming a depletion region in which no one can move left or right. The central part of the theater can't conduct people!

Let's now look at what happens when we attach wires to each semiconductor half and use a battery to impose a voltage difference between those two halves. In Fig. 13.3.6a, the voltage of the p-type side (left) is higher than the voltage of the n-type side (right), so the resulting electric field points toward the right and pushes electrons toward the left. Electrons migrate toward the p-n junction in the n-type side's conduction levels and away from the p-n junction in the p-type side's valence levels. At the same time, some electrons enter the n-type side from its wire and some electrons exit the p-type side to its wire. The depletion region becomes thinner and, when the voltage difference is large enough, vanishes altogether. Once that happens, electrons begin to cross the p-n junction and the entire device and it conducts electric current.

In the theater analogy, we're adding people on the right to the n-type balcony and removing them on the left from the p-type ground floor. The new people in the n-type balcony can move about the empty seats and migrate toward the center of the theater. Similarly, the empty seats in the p-type ground floor allow people to shift about so that empty seats become available near the center of the theater. At that point, people in the n-type balcony can cross over to the p-type balcony and then jump down to the ground floor. There is a net leftward flow of people through the theater; the theater is conducting people from right to left.

What happens if we reverse the battery? In Fig. 13.3.6b, the voltage of the n-type side (right) is higher than the voltage of the p-type side (left), so the resulting electric field points toward the left and pushes electrons toward the right. In this case, the depletion region becomes thicker as the electrons fill in the empty valence levels in the p-type side and leave the conduction levels in the n-type side. The widening depletion region prevents charge from moving and no current flows across the p-n junction. The device remains an insulator.

In the theater analogy, we're removing people on the right from the n-type balcony and adding them on the left to the p-type ground floor. Soon the n-type balcony is virtually empty and the p-type ground floor is essentially full. The entire theater is now a depletion region and behaves like the insulator theater. No one can move and the theater can't conduct people.

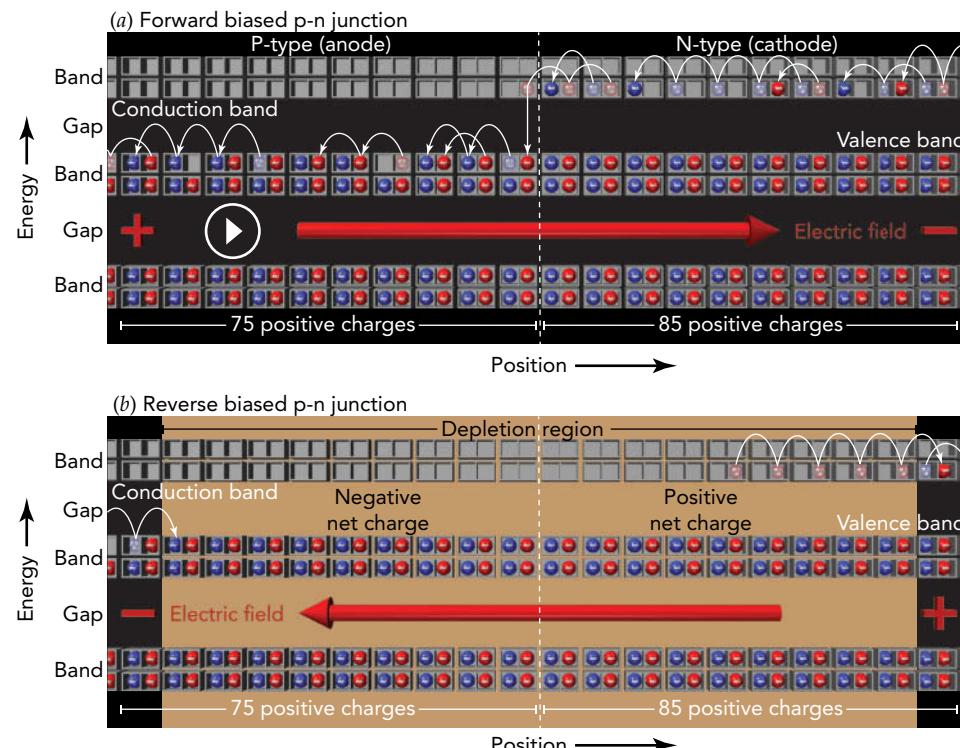


Fig. 13.3.6 (a) When a p-n junction's p-type anode has a higher voltage than its n-type cathode, the resulting electric field causes electrons to migrate toward the junction in the cathode and away from the junction in the anode. The depletion region shrinks or vanishes, electrons cross the junction, and current flows through the device. (b) When a p-n junction's p-type anode has a lower voltage than its n-type cathode, the resulting electric field causes electrons to migrate away from the junction in the cathode and toward the junction in the anode. The depletion region grows, no electrons cross the junction, and no current flows through the device.

Since it allows current to flow in one direction but not the other, the p-n junction is a diode. For historical reasons, the diode's p-type side is called the *anode* and its n-type side is called the *cathode*. Current, which is the flow of positive charge, can pass through a diode only from its anode to its cathode. Since current naturally flows from higher voltage to lower voltage, a diode carries current only when it is **forward biased**, that is, when its anode has a higher voltage than its cathode. When it is **reverse biased**, when its anode has a lower voltage than its cathode, no current flows through the diode.

Check Your Understanding #3: It's a One-Way Street

What will happen if you include a p-n junction (a diode) in the AC circuit that connects the power company to your table lamp?

Answer: Current will flow through the circuit only half the time, and your lamp will be dim.

Why: If you include a diode in an AC circuit, it will prevent current from flowing in one direction. Current will flow through the circuit only during the half of each power-line cycle when the diode's cathode is negatively charged and its anode positively charged. Since the lamp will receive only about half of its normal power, its filament won't reach normal operating temperature. The lamp will glow dimly but the bulb will last an extraordinarily long time. Diodes are often used in this manner to create dim light levels in lamps or appliances.

Light-Emitting Diodes

Even when a diode is forward biased, its depletion region doesn't vanish until the voltage of its anode is significantly higher than the voltage of its cathode. For the ordinary silicon diodes commonly used in AC adapters and computer power supplies, a voltage drops of 0.6 V is required before they'll conduct current. Since the current passing through the diode is then experiencing a voltage drop, the diode is consuming electric power. Because the diodes used in adapters and power supplies serve only as switches to control the direction of current flow, the electric power consumed by these diodes is simply converted into thermal power. But some specialized diodes are designed to convert much of the electric power they consume into optical power—they emit light!

When a diode is forward biased and current flows from its anode to its cathode, conduction-level electrons in the n-type cathode travel across the p-n junction to become conduction-level electrons in the p-type anode. In effect, the anode is then in an excited state; it has conduction-level electrons and empty valence levels (Fig. 13.3.6a).

What happens next depends on the characteristics of the diode. In a normal silicon diode, those conduction-level electrons shift to empty valence levels without producing significant light. Silicon's band structure has characteristics that discourage light emission, so most of these electron transitions produce internal vibrations and heat the diode instead of producing light.

In specialized diodes made from more exotic semiconductors, conduction-level electrons injected from the n-type cathode into the p-type anode frequently undergo radiative transitions to empty valence levels and thereby emit light (Fig. 13.3.7). Composed primarily from combinations of gallium, indium, aluminum, arsenic, phosphorus, and nitrogen, they are known as light-emitting diodes, or LEDs. LEDs now come in just about any color of the rainbow, including infrared, red, orange, yellow, green, blue, violet, and ultraviolet (Fig. 13.3.8). Although white LEDs are also common, they're actually violet or ultraviolet LEDs with built-in phosphors that fluoresce white.

An LED's color is directly related to the energy released when an electron in its p-type anode shifts from a conduction level to a valence level. The most convenient unit in which to measure that energy is the **electron volt** (abbreviated eV), the energy released when 1 elementary unit of electric charge experiences a 1-V decrease in voltage (1 eV is equal to 1.6021×10^{-19} J). In a typical red LED, an electron releases 1.9 eV as it shifts from a conduction level to a valence level and can produce a photon with an energy of 1.9 eV. Since

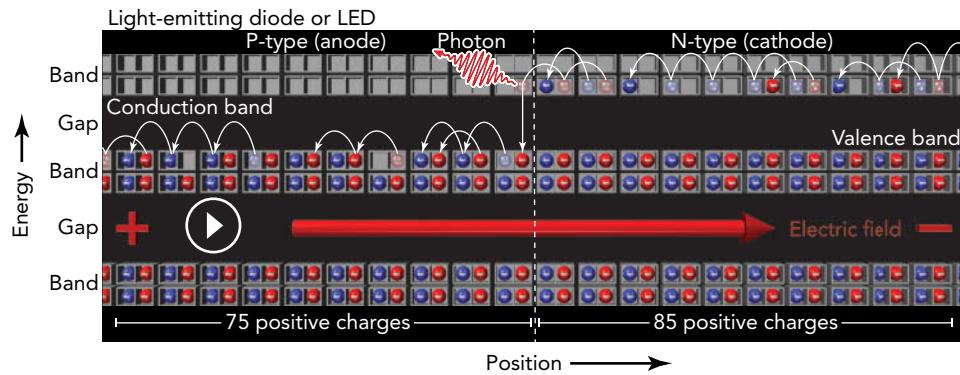


Fig. 13.3.7 When current flows through the p-n junction of light-emitting diode, conduction-level electrons from the n-type cathode are injected into conduction levels of the p-type anode. In the anode, those conduction-level electrons can shift to empty valence levels and release energy. About half the time, a radiative transition converts that energy into a photon of light.

energy and frequency are related by Eq. 13.2.1 and frequency and wavelength are related by Eq. 9.2.1, that 1.9-eV photon has a frequency of 4.6×10^{14} Hz and a wavelength of 650 nm.

To operate and produce these 1.9-eV photons, the red LED must be forward biased with a voltage drop of at least 1.9 V. The current-carrying diode uses that voltage drop to inject electrons into the anode's conduction band, where they have energies 1.9 eV above the valence band. Many of those electrons subsequently release their excess energies as 1.9-eV photons of light.

The shorter the wavelength of the light an LED emits, the more energy each electron must release as it shifts from a conduction level to a valence level and the larger the semiconductor's band gap must be. A violet LED that emits 400-nm light requires a band gap of about 3.1 eV to produce its 3.1-eV photons. That LED also needs a forward-bias voltage in excess of 3.1 V. The larger voltage drops required by LEDs near the violet end of the spectrum explain why those LEDs need higher-voltage power sources or more batteries.

Unfortunately, at most about half of the electrons sent across an LED's p-n junction succeed in lighting the room. Although a substantial fraction of those electrons emit photons, most of the photons are reabsorbed by the semiconductor before they leave the LED; the same radiative transitions that emit this light (conduction level \rightarrow valence level) can also absorb it (valence level \rightarrow conduction level). Despite those difficulties, modern LEDs can produce visible light with energy efficiencies comparable to those of fluorescent lamps. LED efficiencies continue to rise along with their operating lifetimes, and they are gradually becoming a primary form of illumination.



Courtesy Lou Bloomfield

Fig. 13.3.8 These LEDs are connected in series so that the same current flows sequentially through each of them. However, their different band gaps cause them to emit different colors of light.

Check Your Understanding #4: Lighten Up

When you increase the current flowing through an incandescent lightbulb, it brightens and its color shifts toward the blue end of the spectrum. If you increase the current flowing through an LED, what happens to its brightness and color?

Answer: The diode becomes brighter, but its color doesn't change significantly.

Why: The diode's brightness depends on how many electrons cross its p-n junction each second, and increasing that number increases the brightness. However, the color of light the diode emits depends mostly on its band gap and therefore doesn't change much when you increase the current in the diode.

Lasers and Laser Light

To understand lasers, you must understand how laser light differs from the normal light emitted by hot objects or by individual atoms in an electrical discharge. Each particle of normal light, each photon, is emitted willy-nilly without any relationship to the other light particles being emitted nearby. Because of this light's independent and unpredictable character, it's called *spontaneous light* and its creation is called **spontaneous emission of radiation**.

Theoretical work by Albert Einstein and others in the 1920s and 1930s predicted the existence of a second type of light, *stimulated light*, that can be created when an excited atom or atomlike system duplicates a passing photon. Although this **stimulated emission of radiation** can occur only when the excited atom is capable of emitting the duplicate photon spontaneously, the copy that it produces is so perfect that the two photons are absolutely indistinguishable. Together, these two photons form a single electromagnetic wave.

To get a slightly better picture of how such stimulation occurs, think about an isolated atom in an excited state. That atom will eventually return to its ground state, but it must emit one or more photons for this to happen. That atom waits around in the excited state until it begins a spontaneous radiative transition. During the transition, one of the atom's electrons accelerates back and forth, and the atom emits a photon.

If a similar photon passes through the atom as it's waiting in the excited state, that photon's electric field can stimulate the radiative transition process through sympathetic vibration. The field pushes and pulls on the atom's electrons and makes them accelerate back and forth. Although this effect is small, it may be enough to trigger the emission of light. If the atom does emit light, the photon it produces will be a perfect copy of the stimulating photon.

When this stimulated emission process was first discovered, people immediately recognized that it made light amplification possible. If enough excited systems could be assembled together, a single passing photon could be duplicated exactly over and over again. Instead of a single particle of light, you would soon have thousands, or millions, or even trillions of identical light particles.

Implementing this idea had to wait until the late 1950s, however, when the technical details of how to actually achieve light amplification were worked out. In 1960, the first laser oscillators were constructed. These were devices that emitted intense beams of light in which each particle of light was identical to every other particle of light. A single particle of light had been duplicated by the stimulation process into countless copies.

When individual excited atoms or atomlike systems emit light through spontaneous emission, the particles of light head off separately as many independent electromagnetic waves (Fig. 13.3.9a). Light consisting of many independent electromagnetic waves is called **incoherent light**.

However, when that same collection of excited atoms or atomlike systems emit light by stimulated emission, all of the light particles are *absolutely* identical and form a single electromagnetic wave (Fig. 13.3.9b). Unlike electrons, which are Fermi particles, many identical photons can have the same quantum wave because photons are Bose particles and

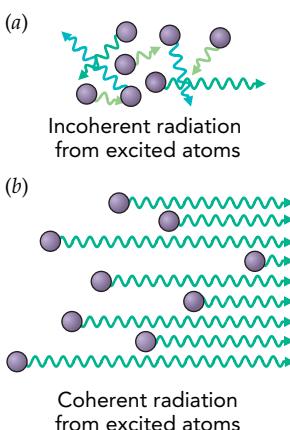


Fig. 13.3.9 (a) Photons of incoherent light are created independently and have somewhat different wavelengths and directions of travel. (b) Photons of coherent light are produced by stimulated emission and are identical in every way.

Bose particles don't obey the Pauli exclusion principle. Light consisting of many identical photons and a single electromagnetic wave is called **coherent light**. Because of its single-wave nature, coherent light exhibits remarkable interference effects. These effects are easily seen in the coherent light emitted by lasers.

Check Your Understanding #5: A Light Dusting of Photons

If you were to measure the electric field in the light from a flashlight, you would find that it fluctuates about randomly. Why is the light's electric field so disorderly?

Answer: The flashlight produces incoherent light, with many independent light waves contributing their individual electric fields to the overall electric field. When you were to measure that overall field, you find it very disorderly.

Why: Because the photons of incoherent light are independent, their individual electric fields don't fluctuate together. At any place and time, the individual electric fields will add in a complicated, random fashion. As time passes, this overall field will fluctuate randomly. In contrast, coherent light, in which each photon is the same as all the others, has a much more orderly electric field because all of the photons contribute equally.

Light Amplifiers and Oscillation

Producing coherent light requires amplification. You must start with only one particle of light and duplicate it many times. The basic tool for this duplication of light is a **laser amplifier** (Fig. 13.3.10). When weak light enters an appropriate collection of excited atoms or atomlike systems—the **laser medium**—that light is amplified and becomes brighter. The new light has exactly the same characteristics as the original light, but it contains more photons.

When we think of lasers, however, we rarely imagine a device that duplicates photons from somewhere else. We usually picture one that creates light entirely on its own. To do that, the laser must produce the initial particle of light that it then duplicates to produce others. A **laser oscillator** is a device that uses the laser medium itself to provide the seed photon, which it then duplicates many times (Fig. 13.3.11). If a laser medium is enclosed in a pair of carefully designed mirrors, it's possible for the stimulation process to become self-initiating and self-sustaining. However, the mirrors must be curved properly and must have the correct reflectivities. One mirror must normally be extremely reflective, while the other must transmit a small fraction of the light that strikes its surface.

When the laser medium is placed between the two mirrors, there is a chance that a photon, emitted spontaneously by one of the excited systems, will bounce off a mirror and return toward the laser medium. As that returning photon passes through the laser medium, it's amplified. Because the photon was emitted by one of the excited systems, it has the right wavelength to be amplified by other excited systems. (For a discussion of a photon's properties, see 5.)

By the time the original photon leaves the laser medium, it has already been duplicated many times. This group of identical photons then bounces off the second mirror and returns for another pass through the laser medium. It continues to bounce back and forth between the mirrors until the number of identical photons in the collection is astronomical.

Eventually there are so many identical photons that the laser medium is no longer able to amplify them. The laser medium has only so much stored energy and only so many excited systems in it. If the laser medium continues to receive additional energy, it may continue to amplify the light somewhat. If it doesn't receive more energy, light amplification will eventually cease.

To let the light out of this laser oscillator, one of its mirrors is normally *semitransparent*—that is, some of the photons that strike the surface of the mirror travel through it rather than reflecting. (The one-way glass used for surveillance is actually a semitransparent mirror.) This transmission creates a beam of outgoing light, a *laser beam*. The laser beam continues to emerge from the mirror as long as the amplification process can support it.

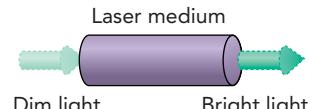


Fig. 13.3.10 A laser amplifier uses excited atoms or atomlike systems to increase the number of light particles leaving the laser medium. Incoming light is duplicated by stimulated emission.

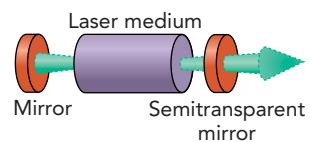


Fig. 13.3.11 A laser oscillator is a laser amplifier enclosed in mirrors. Oscillation occurs when the laser medium spontaneously emits one photon in just the right direction. This photon bounces back and forth between the two mirrors and is duplicated many times. Some of the light is extracted from this laser by making one of the mirrors semitransparent.

5 Although it might seem that a photon should have an exact wavelength and frequency, and travel in only one direction, that's not the case. Photons travel as electromagnetic waves and spread in more than a single direction. Also, because each photon has a beginning and an end, its wave contains more than a single wavelength or a single frequency. Thus, while lasers can produce some of the most perfect electromagnetic waves imaginable, those waves still spread outward slightly and still have a range of wavelengths and frequencies.

Because this laser beam consists of duplicates of one original photon, it is coherent light. For technical reasons, many lasers duplicate more than one original photon simultaneously, so their laser beams are a little less coherent than they might be. However, with suitable fine-tuning, one original photon can usually be made to dominate the laser beam.

When you focus a flashlight's beam with a lens, its independent photons won't end up exactly together at the focus of the lens. That's because the photons leave the flashlight heading in somewhat different directions and because their broad range of wavelengths leads to dispersion problems in the lens. However, since virtually all the photons in a laser beam are identical—they travel along the same path, and their wavelengths are the same—they can all focus together to an extremely small spot. That's why a laser printer employs a laser; a laser beam can illuminate a very small spot on the photoconductor drum that is used in the xerographic process (Section 10.2) to produce a printed image.

COMMON MISCONCEPTIONS: Laser Beams

Misconception: A laser beam is a narrow glowing cylinder that travels through space at the speed of an arrow.

Resolution: A laser beam is light, and it travels through space at the speed of light. To be visible from the side, a laser beam must scatter off something in its path, such as dust, mist, or air molecules. In empty space, a laser beam is invisible except when it directly illuminates your eyes.

Check Your Understanding #6: More of a Good Thing

If you take the laser beam from a particular laser oscillator and send it through a similar laser amplifier, what will happen to the laser beam?

Answer: It will become even brighter.

Why: A laser oscillator normally emits as intense a beam of light as it can, based on the amount of stored energy in the laser medium. This beam of light can be amplified further by sending it through a separate laser amplifier. Most high-powered lasers use a laser oscillator and one or more laser amplifiers to create particularly bright beams of light.

How a Laser Medium Works

Obtaining the excited systems needed to amplify light is a critical issue for lasers. Ideally, a laser involves four different states of an atom or atomlike system: the ground state, an excited state, the upper laser state, and the lower laser state. The reason for having four separate states should become clear in a moment.

Let's consider an atom that acts as an ideal laser amplifier (Fig. 13.3.12). The atom starts in its ground state. A collision or the absorption of a photon shifts it to the excited state, giving it the energy it needs to amplify light. The atom then shifts to the upper laser state, either by emitting a photon or as the result of a collision. This preliminary shift is important because it prevents the excited atom from returning directly to the ground state and avoiding the amplification process. Once it has shifted to the upper laser state, the atom is stuck there and will wait around long enough to amplify light.

The atom is poised to duplicate a passing photon. But not any old photon will do; it must match the photons that the atom is capable of emitting. For example, a neon atom with an excited $3p$ electron (Section 13.2) can duplicate a red photon but not a blue photon. This color selectivity of a laser amplifier is why most laser beams have a single, pure color.

When a suitable photon passes through the atom, that photon stimulates the emission of a duplicate photon and the atom undergoes a radiative transition to the lower laser state. So far, so good. However, if the atom remains in the lower laser state, it might

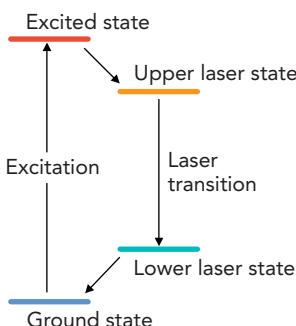


Fig. 13.3.12 An ideal laser system passes through four different states during the laser's operation.

absorb a photon of the laser light and return to the upper laser state. To prevent this sort of radiation trapping, the atom must quickly shift to the ground state, either by emitting a photon or as the result of another collision. The atom is then ready to begin the cycle all over again.

This four-state cycle, or something close to it, is found in nearly all lasers. The cycle helps the laser develop a **population inversion** between its upper and lower laser states, a situation in which there are more atoms in the upper laser state prepared to *emit* the laser light than there are atoms in the lower laser state prepared to *absorb* that light. Developing a population inversion is critical to laser amplification because, without it, the laser medium is more absorbing than amplifying and there can't be a buildup of light intensity.

In each laser, something provides the energy needed to shift atoms or atomlike systems in the laser medium from their ground states to their excited states in order to develop a population inversion. This transfer of energy into the laser medium to prepare it for amplifying light is called *pumping*. How a particular laser medium is pumped depends on the laser.

The most common pumping mechanisms are electronic and optical. In electronic pumping, currents of charged particles use their kinetic or electrostatic energies to excite the medium's atoms or atomlike systems from their ground states to their excited states. In optical pumping, intense light is shone on the laser medium, causing a similar excitation.

The most important examples of optical pumping are ion-doped solid-state lasers. These lasers are based on atomic ions embedded in transparent solids. Common ions are titanium (Ti), neodymium (Nd), and erbium (Er), and they are often embedded in sapphire, yttrium aluminum garnet (YAG), or glass. Ti:sapphire, Nd:YAG, and Er:glass lasers are important in modern research, technology, and optical communications systems. When these laser media are exposed to extremely bright light, their ions become excited and they can act as laser oscillators or amplifiers (Figs. 13.3.13 and 13.3.14).

A laser diode is quite similar to an LED, except that a laser diode uses its radiative transitions to amplify light. Since that amplification can occur only when light emission exceeds light absorption, the laser diode must produce a population inversion between an upper laser state and a lower laser state.

The laser diode achieves such an inversion by concentrating current into a very narrow p-n junction made from heavily doped semiconductors. The intense current injects an enormous density of electrons into the anode's conduction band, where they quickly settle into the lowest-energy conduction levels—the upper laser state. The heavy doping empties most of the anode's highest-energy valence levels—the lower laser state. With many electrons in the upper laser state and few in the lower laser state, the diode has a population inversion and can amplify light.

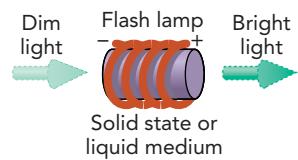
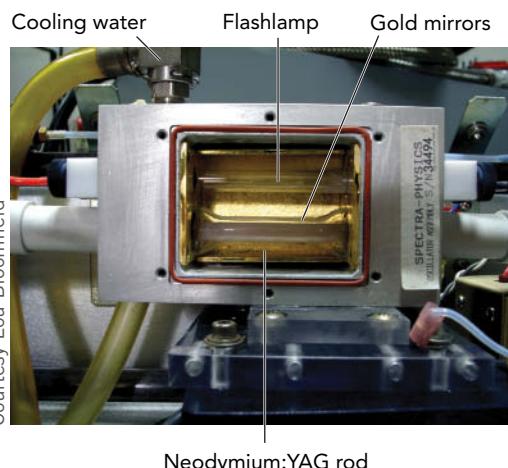


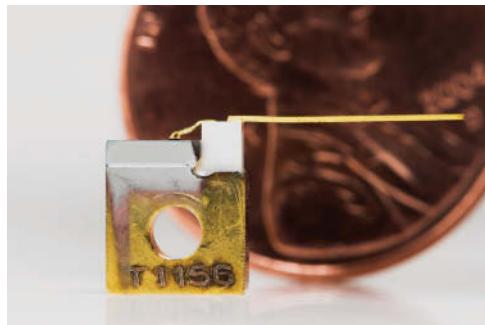
Fig. 13.3.13 In an optically pumped laser, intense light from a flashlamp, an arc lamp, or even another laser transfers energy to the laser medium. Atoms or atomlike systems inside the medium store this energy and use it to amplify light.



Courtesy Lou Bloomfield

Fig. 13.3.14 This flashlamp-pumped laser amplifier contains a purple neodymium:YAG rod. The rod is in the bottom half of the opened, gold-lined amplifier box and is protected by a glass tube. Light from a long flashlamp in the top half of the box excites the neodymium ions so that they can amplify infrared light passing horizontally through the rod.

Fig. 13.3.15 This tiny semiconductor chip is a diode laser that emits an intense beam of coherent light when a current flows through it.



© GIPhotoStock Z/Alamy

Most laser diodes act as laser oscillators (Fig. 13.3.15), amplifying their own spontaneously emitted light until it forms an intense coherent beam. The ends of the anode itself are usually reflective enough to act as mirrors and form a complete laser oscillator. However, to concentrate the laser light in one direction and to control its beam characteristics, many laser diodes have complicated structures and coatings.

Check Your Understanding #7: Not So Fast, 007

A secret agent in a movie uses a tiny hand-held laser to burn a hole through a thick metal plate. From the point of view of power, why is such a laser essentially impossible to make?

Answer: The light in such a laser beam would carry an enormous amount of power. Something must transfer that power to the laser medium, an impossible task in a tiny hand-held unit.

Why: The power in a laser beam comes from the laser medium. The laser medium must have gotten it from somewhere else. Since batteries aren't up to delivering thousands of watts of electric power, it's unlikely that powerful hand-held lasers will ever be developed. Even if a suitable power source were available, these lasers would tend to overheat. The conversion of power to light is not perfectly efficient, and most of the energy ends up as thermal energy in the laser's components. Lasers require cooling to remove this wasted energy.

Epilogue for Chapter 13

In this chapter, we explored the creation and movement of light. In Sunlight, we looked at how light is scattered during its passage through the atmosphere and at how it bends and reflects when it moves from one material to another. We also examined the interference effects that occur when a light wave follows more than one path to a particular destination and learned how polarizing sunglasses can diminish glare.

In Discharge Lamps, we examined electrical discharges in gases. We saw that atoms excited by collisions with charged particles can subsequently emit light through radiative transitions. We studied the primary colors of light to see how the phosphors on the inside surface of a fluorescent tube are able to produce a reasonable facsimile of white sunlight.

In LEDs and Lasers, we looked at how electrons can emit light after crossing a junction between two different types of semiconductor, and we looked at the difference between common incoherent light and the unusual coherent light emitted by a laser. We saw that lasers use stimulated emission to duplicate photons so that a small number of initial photons can be amplified into an enormous number.

Explanation: Splitting the Colors of Sunlight

Light bends as it passes through the cut facets of the crystal glass or bowl. Because of dispersion, the angle of each bend depends slightly on the wavelength of the light involved, so the different colors of sunlight follow slightly different paths through the crystal. When the

light emerges from the crystal, its various wavelengths head in somewhat different directions; you see colors. Color sequences progress from long wavelength to short wavelength, or vice versa. You see red, orange, yellow, green, blue, indigo, and violet, or the reverse—the colors of the rainbow.

Chapter Summary and Important Laws and Equations

How Sunlight Works: Sunlight originates at the 5800 K outer surface of the sun when electrically charged particles there accelerate back and forth rapidly and emit electromagnetic waves. This sunlight travels at the speed of light through empty space until it reaches Earth's atmosphere. There it slows slightly and some of it Rayleigh-scatters. Short-wavelength light is Rayleigh-scattered more strongly than long-wavelength light, so the sky appears blue.

As the sunlight passes through various objects, it slows down and its colors separate. When sunlight passes through falling raindrops, its different wavelengths follow different paths and create rainbows. As sunlight reflects from thin films such as soap bubbles, its waves are divided and may follow several different paths to the same destination. These light waves then interfere with one another so that some waves appear strong and bright while others are weak and dim. Interference depends on the wavelengths of light, so thin films appear brightly colored.

How Discharge Lamps Work: Discharge lamps produce light by passing electric currents through gases. Those gases are turned into electrical-conducting plasmas by filling them with charged particles, either by exposing them to strong voltage gradients or by injecting electrons into them from heated electrodes. Once a plasma is formed, current can pass through it and collisions within that current-carrying plasma cause the gas particles to emit light.

A fluorescent lamp emits visible light when the phosphor coating on the inside of its tube is exposed to 254-nm ultraviolet light produced inside the tube by a low-pressure mercury vapor discharge. This ultraviolet light excites the phosphor coating and causes it to emit visible light. In contrast, mercury, metal-halide, and sodium lamps use discharges to produce visible light directly and don't employ phosphors. By operating at high pressures, those discharge lamps produce relatively broad light spectra and provide energy-efficient illumination.

How LEDs and Lasers Work: In an LED, light is emitted when conduction-level electrons from the diode's cathode cross the p-n junction into the anode and undergo radiative transitions to empty valence levels. The color of light produced by the LED is determined primarily by the anode's band gap.

Lasers amplify light through the process of stimulated emission. In this process, energy is transferred to atoms or atomlike systems contained in a laser medium. These excited systems might emit light spontaneously, and they can be stimulated into emitting duplicates of a passing photon. When a photon with just the right wavelength passes through an excited system, that system is likely to give up its stored energy by emitting an exact copy of the initial photon. In a laser oscillator, two mirrors cause light to bounce back and forth through the laser medium. An initial photon is duplicated endlessly to produce coherent light. One of the mirrors is semitransparent so that part of the light emerges from the laser as a laser beam.

1. Pauli exclusion principle: No two indistinguishable Fermi particles ever occupy the same quantum wave. frequency of the light wave, or
energy = Planck constant • frequency. (13.2.1)
2. Relationship between energy and frequency: The energy in a photon of light is equal to the Planck constant times the

Many of the devices around us perform useful tasks by manipulating light, charge, or both. The techniques of optics deal with light and allow cameras to record images of the objects in front of them, our eyes to observe those objects directly, and eyeglasses and magnifying glasses to help us see details we'd miss with our eyes alone. The techniques of electronics deal with charge and permit an audio player's memory to store sound information, its computer to retrieve that information, and its amplifier and headphones to re-create the sound at the push of a button.

Optical tools such as lenses and prisms have been around for hundreds of years, and electronic devices such as resistors, capacitors, and inductors also have a long history. Advances of modern technology, however, have accelerated developments in both fields. The invention of lasers has sped the growth of the optics industry and the invention of transistors has revolutionized the world of electronics. Rapid progress in both fields, optics and electronics, has brought them closer together and has given birth to the combined field of optoelectronics. There is even hope that one day computers will be as much optical devices as they are electronic.

ACTIVE LEARNING
EXPERIMENTS

Magnifying Glass Camera

There are many household devices that manipulate light, and one of the most familiar is a magnifying glass. A magnifying glass bends light rays toward one another as they pass through it. In this chapter, we'll see how a simple converging lens of this sort can magnify an object or cast

its image onto a light-sensitive surface. For the moment, we'll use it to cast the image of a window onto a wall.

Take a magnifying glass to a room with a bright window and turn off the lights. Hold the magnifying glass near the wall opposite the window and move the glass



Courtesy Lou Bloomfield

toward or away from the wall until you see a window-shaped pattern of light appear on the wall. Once that pattern is visible, carefully adjust the magnifying glass's orientation and distance from the wall to obtain the sharpest image of the window. You'll probably also see images of objects outside that window, but you'll have to move the magnifying glass to sharpen those images. Which way must you move the glass, and why can't all the images be sharp at the same time?

Chapter Itinerary

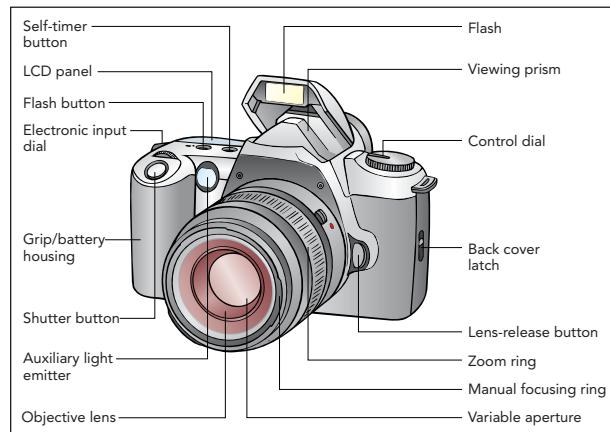
This process of bringing light together to form a small spot or an image of a distant object is a common theme in optics. Representing information as charge and then storing, manipulating, and using that information is typical of electronics. In this chapter, we'll examine a number of systems that are based on these sorts of manipulation of light and charge: (1) *cameras*, (2) *optical recording and communication*, and (3) *audio players*. In Cameras, we see how lenses bend light to form images and how those images are used to create photographs. In Optical

You can project images of other brightly illuminated objects on a sheet of white paper. What features of the lens determine the sizes of those images? What determines their orientations? Block part of the lens, and notice how it affects the images. Try to form images of objects at different distances from you. Do they all form images simultaneously, or do you have to adjust the lens somehow to bring each image into sharp focus?

Recording and Communication, we explore the roles of lasers in optics while investigating several novel optical effects. In Audio Players, we see how a small assortment of basic electronic components is brought together to build a computer and an audio amplifier, and how those two devices have been merged together into a single unit so that you can listen to thousands of songs as you lounge on the beach. For additional preview information, turn to the Chapter Summary and Important Laws and Equations at the end of the chapter.

SECTION 14.1

Cameras



In the two centuries since their invention, cameras have become extremely easy to use. What started as a hobby for a few dedicated enthusiasts has evolved into an everyday activity. Despite all the technological improvements, however, photography still employs many of the same principles it did in the 1800s. Cameras still use lenses to project images onto light-sensitive surfaces, and photographers still have to worry about getting the exposure right, focusing properly, and avoiding the blur of rapid motion. In this section, we'll explore some of the principles that make cameras work.

Questions to Think About: Why are expensive camera lenses so complicated, with so many separate pieces of glass? Why

does a longer lens seem to bring the objects nearer to you? What does a camera's aperture do? Why do nearsighted and farsighted people wear different eyeglasses?

Experiments to Do: The basic activity of a camera—projecting an image of the scene in front of you onto its image sensor—requires nothing more than a simple magnifying glass. In fact, you can use almost any simple lens that's bowed outward in the middle, including the eyeglasses of a farsighted person with no astigmatism or drugstore reading glasses. Stand in a darkened room across from a single bright lamp. Hold a sheet of white paper so that it faces the lamp and move the lens back and forth in front of the paper. Make sure that the lens itself is also facing the lamp.

You should find a distance at which the lens casts a clear image of the lamp onto the sheet of paper. You'll find that the lamp appears upside down and backwards and that the image becomes fuzzy if you move the lens toward or away from the paper. You'll also find that, when the lamp's image is sharp, the images of things in front of or behind the lamp are fuzzy, and vice versa. Which way must you move the lens to bring more distant or less distant objects into focus?

Cut a hole in a piece of cardboard and use it to cover all but the center portion of the lens. The image of the lamp will now be substantially darker, but it will also be sharper over a wider range of distances from lens to paper. If the hole is narrow enough, virtually everything is in focus at once. Evidently, the diameter of the lens and its ability to focus on several things at once are closely related. Why?

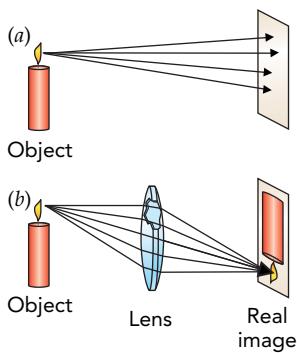


Fig. 14.1.1 (a) Without a lens, light from a candle uniformly illuminates an image sensor. (b) When a lens is introduced between the candle and the sensor, it brings light from each point on the candle back together on the sensor's surface, forming an upside down and backward real image of the candle. The distance between the lens and the sensor must be chosen correctly or the image will be blurry.

Lenses and Real Images

When you take a picture of the scene in front of you, the lens of your camera bends light from that scene into a real image on a light-sensitive surface. A **real image** is a pattern of light, projected in space or on a surface, that exactly reproduces the pattern of light in the original scene. Since the real image that's projected looks just like the scene you're photographing, recording the light in that image is equivalent to recording the appearance of the scene itself.

That light-sensitive surface was at one time always photographic film, but now digital cameras have almost completely replaced film with electronic image sensors. Fortunately, the two light-sensing surfaces are essentially interchangeable, so we can refer to them both as *image sensors*: one is electronic, and one is photochemical.

Real images don't occur without help. When light from a candle falls directly on an image sensor, it produces only diffuse illumination (Fig. 14.1.1a). Similarly, you can't tell by looking at a sheet of paper what a candle looks like because the light that leaves the candle travels in all directions and is as likely to hit the top of the paper as it is to hit the bottom (Fig. 14.1.1b).

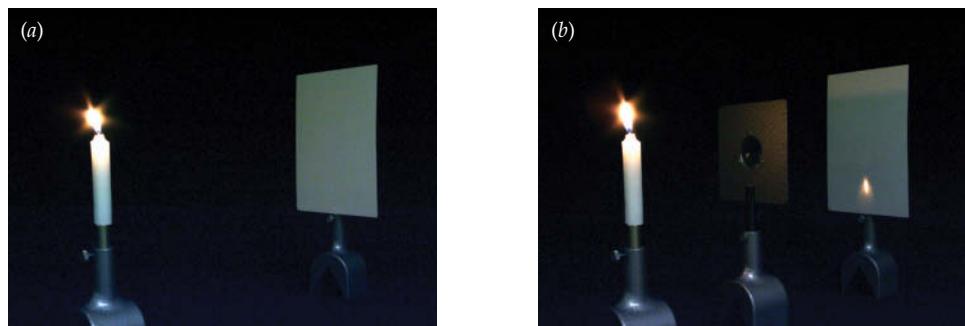
That's why a camera needs a **lens**, a transparent object that uses refraction to form images. The light passing through a lens bends twice, once as it enters the glass or plastic and again as it leaves. In a camera lens, this bending process brings much of the light from one point on the candle back together at one point on the sensor. As you can see in Fig. 14.1.2, the real image that forms is upside down and backward. This inversion of the real image relative to the object always happens when a single lens creates a real image.

The curved shape of the camera lens allows it to form a real image. Light passing through the upper half of the lens is bent downward, while light passing through the lower half is bent upward. Because the camera lens bends light rays toward one another, it's a **converging lens**. You can see how it forms an image in Fig. 14.1.2b by following some of the rays of light leaving one point on the candle.

The upper ray from the candle flame travels horizontally toward the top of the lens. As it enters the lens and slows down, this ray of light bends downward. It bends downward again as it leaves the lens and travels downward toward the bottom of the image sensor.

The lower ray from the candle flame travels downward toward the bottom of the lens and bends upward as it enters the lens. It bends upward again as it leaves the lens and travels horizontally toward the bottom of the image sensor.

These two rays of light reach the image sensor at the same point. They are joined there by many other rays from the same part of the candle flame so that a bright spot forms on the sensor. Overall, each part of the candle illuminates a particular spot on the image sensor, so the lens creates a complete image of the candle on the sensor.



Courtesy Lou Bloomfield

Fig. 14.1.2 (a) When candlelight falls directly on a sheet of paper, it produces no image. (b) A lens inserted between the candle and paper forms an inverted real image of the flame on the paper.

However, the lens can bring the light back together to form a sharp image on the sensor only if the lens and sensor are separated by just the right distance (Fig. 14.1.3). If the sensor is too close to the lens, then the light doesn't have room to come together. If the sensor is too far from the lens, then the light begins to come apart again before reaching the sensor. In either case, the image on the sensor is blurry. The candle's real image is only *in focus* at one distance from the lens.

If the candle moves toward or away from the camera lens, the distance between the lens and the image sensor must also change (Fig. 14.1.4). When the candle is far away, all its light rays that pass through the lens arrive traveling almost parallel to one another and the inward bend caused by the lens makes those rays converge together quickly. The rays come into focus relatively near the lens, and that's where the sensor must be (Fig. 14.1.4a). The candle's image on the sensor is much smaller than the candle itself because the light rays have only a short distance over which to move up or down after leaving the lens.

When the candle is nearby, its light rays that pass through the lens are diverging rapidly and the inward bend caused by the lens is just barely enough to make those rays converge at all. As a result, the rays come into focus relatively far from the lens (Fig. 14.1.4b). The candle's image on the sensor is quite large because the light rays have considerable distance over which to move up or down after leaving the lens.

Because distant and nearby objects form real images at different distances from the camera lens, they can't both be in focus on the same image sensor. When you take a picture of a person standing in front of a mountain, only one of them can be in sharp focus. However, if you're willing to compromise a little bit on sharpness, a lens can sometimes form acceptable images of both objects.

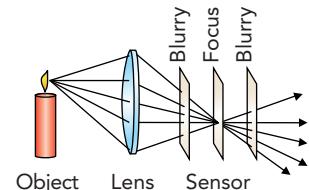


Fig. 14.1.3 The real image is in focus only when the image sensor is just the right distance from the lens. If the sensor is too near or too far from the lens, the image is blurry.

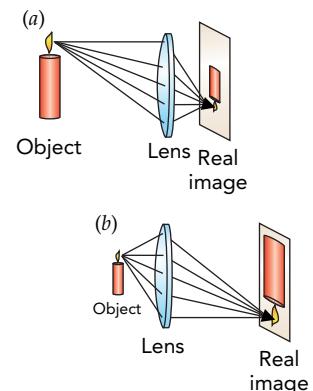


Fig. 14.1.4 (a) The light from a distant candle is traveling in almost the same direction, and the lens focuses it easily. The real image forms close to the lens. (b) The light from a nearby candle diverges, and the lens has more difficulty bending it back together. The real image forms far from the lens. If the candle is too close to the lens, no real image forms at all.

Check Your Understanding #1: Seeing the Lights

If you hold a magnifying glass at the proper distance above a sheet of white paper, you will see an image of the overhead room lights on the paper. Explain.

Answer: It is a real image of the room lights, created by the converging magnifying glass.

Why: A magnifying glass is a converging lens and can form a real image. It brings light spreading outward from the room lights together onto the sheet of paper as a real image.

Focusing and Lens Diameter

A disposable camera is little more than a box with a lens. The lens projects a real image of the scene in front of it onto the camera's image sensor. Light in the real image exposes the image sensor, which records the image permanently. While it may also have a shutter that starts and stops the exposure, a flash to provide extra light, and a mechanism that prepares for the next photograph, there's little else to this simple camera.

However, there are limitations to the disposable camera design. One of the most severe limitations is that you can't focus it—the camera has a fixed distance between the lens and image sensor. Nonetheless, it manages to form relatively sharp real images on the sensor, even when there are objects at various distances from the camera. These simple cameras work because they use narrow (small-diameter) lenses. A narrow lens gathers less light than a wide lens, but it doesn't require *focusing*, adjusting the distance between the lens and the image sensor.

Because a wide (large-diameter) lens brings rays together from many different directions, you must focus it (Fig. 14.1.5a); if the image sensor is even slightly too near or too far from the lens, the recorded image will be blurry. In contrast, a narrow lens forms a reasonably clear image even without focusing. Any rays from one part of the scene that succeed in passing through the narrow lens must already be fairly close together. Their initial closeness means that these converging rays illuminate only a small part of the image sensor even when the sensor isn't exactly the right distance from the lens (Fig. 14.1.5b).

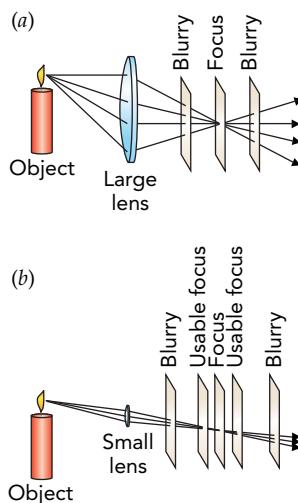


Fig. 14.1.5 (a) A large lens collects lots of light but its focus is critical. The image is blurry except at the actual focus. (b) A small lens collects less light but its image is relatively sharp anywhere near the focus.

Courtesy Lou Bloomfield

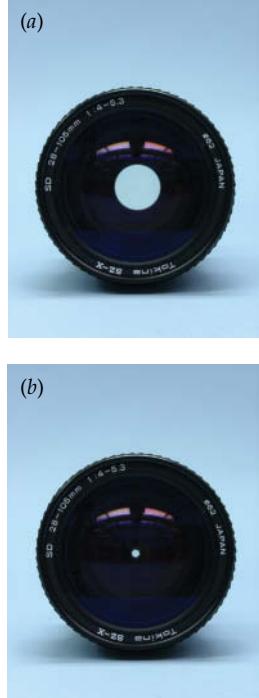


Fig. 14.1.6 The aperture of this lens can be reduced by closing its internal diaphragm, dimming its image but increasing its depth of focus.

Since the image sensor can't record every minute detail anyway, the image that forms on it doesn't have to be in absolutely perfect focus. As a result, a camera with a narrow lens and no focus adjustment manages to take pretty good pictures.

Unfortunately, these simple cameras collect very little light and need extremely light-sensitive image sensors. These high-speed sensors can't record photographs that are as sharp as low-speed sensors can. Furthermore, the pictures produced by simple cameras lack fine details; although everything is almost in focus, most things are a bit fuzzy if you look carefully or make enlargements.

More sophisticated cameras use wider lenses that gather more light and expose the image sensors much more rapidly. They also automatically adjust the distance between the lens and the sensor. They identify the object you are photographing and position the camera lens so that it projects a sharp image on the sensor. As the camera focuses, you can usually see components in the lens moving backward or forward to arrive at the correct distance from the sensor.

Even a camera with a wide lens can take advantage of the narrow lens trick for focusing. Its lens contains an internal diaphragm that reduces its **aperture**, or effective diameter. The **diaphragm** is a ring of metal strips with a central opening. These strips can swing in or out, changing the diameter of the diaphragm's opening and with it the aperture of the lens (Fig. 14.1.6).

When its lens aperture is narrow, the sophisticated camera imitates a simple camera—nearly everything is essentially in focus simultaneously. In such a situation, the camera has a large *depth of focus*. Actually, the sophisticated camera can bring the most important object into perfect focus, so it produces pictures that are superior to those from simple cameras. However, narrowing the aperture of the lens also reduces the amount of light reaching the image sensor. The scene in front of the camera must either be very bright or the exposure must be relatively long. You don't get something for nothing.

Although widening the aperture of a large lens makes full use of its light-gathering capacity, focusing then becomes crucial. Even a small error in the lens-to-sensor distance produces a blurry picture, so the depth of focus is very small. This trade-off between light gathering and depth of focus is a continual struggle for photographers. However, photographers sometimes take advantage of the small depth of focus in wide lenses to blur the background or foreground of a photograph deliberately. A camera's portrait setting adopts this strategy to produce sharp images of people against blurred backgrounds.

At other times, photographers choose long exposures at narrow apertures to bring an entire scene into sharp focus. A camera's landscape setting takes this route, so everything in the photograph shows full detail. To capture fast motion while retaining a large depth of focus, photographers use a flash to brighten the scene and shorten the exposure. Unfortunately, a camera flash is ineffective at brightening a distant scene, and it can produce unpleasant reflections from windows and eyes. A camera's sports setting emphasizes brief exposures to avoid speed blur, even though that may require a wide aperture and small depth of focus.

Check Your Understanding #2: Portrait Photos

While taking a photograph of your friend, with the diaphragm of your large camera lens wide open, you notice that the background is blurry. Explain.

Answer: Light from the more distant background focuses closer to the lens than light from your friend. Since the camera is focusing on your friend, the background appears out of focus.

Why: When the aperture of your camera lens is wide open, focusing becomes critical. Objects in front of or behind your friend are out of focus on the image sensor and appear blurry.

Focal Lengths and f-Numbers

Lenses are characterized by two quantities: focal length and f-number. The **focal length** of a lens is the distance between the lens and the real image it forms of a *very distant object*. For example, if a real image of the moon forms 100 mm (4 in) behind a particular lens, then

that lens has a focal length of 100 mm. The focal lengths of camera lenses range from less than 10 mm (0.4 in) in many cell phones and compact cameras to about 2 m (7 ft) in cameras used for nature photography.

When light from a scene passes through a short-focal-length lens, it comes to a focus near that lens and produces a relatively small image on the image sensor. Because a long-focal-length lens permits the light passing through it to spread out more before coming to a focus, it produces a larger real image on the sensor.

The “normal” lens for a particular camera has a focal length that allows all the objects in your central field of vision to fit onto the image sensor (see Table 14.1.1). When you hold the finished photograph about 30 cm (1 ft) from your eyes, the objects in the picture appear about the same size they did when the photograph was taken. The focal length of a camera’s normal lens is about 1.5 times the horizontal width of its image sensor.

A wide-angle lens has a shorter focal length than the normal lens (Fig. 14.1.7a). The image it projects onto the image sensor is smaller but brighter, and most of the objects in your entire field of vision appear in the photograph. A telephoto lens has a longer focal length than the normal lens (Fig. 14.1.7b). The image it projects onto the sensor is larger but dimmer, with only objects at the center of the scene appearing in the photograph.

In addition to indicating where the image of a distant object forms, the focal length of the camera lens relates the object distance to the image distance. The **object distance** is the distance between the lens and the object you’re photographing. The **image distance** is the distance between the lens and the real image it forms (Fig. 14.1.8). The relationship is called the **lens equation** and can be written in a word equation:

$$\frac{1}{\text{focal length}} = \frac{1}{\text{object distance}} + \frac{1}{\text{image distance}}, \quad (14.1.1)$$

in symbols:

$$\frac{1}{f} = \frac{1}{o} + \frac{1}{i},$$

and in everyday language:

The farther away an object is, the closer to the lens its image forms.

THE LENS EQUATION

One divided by the focal length of a lens is equal to the sum of one divided by the object distance and one divided by the image distance.

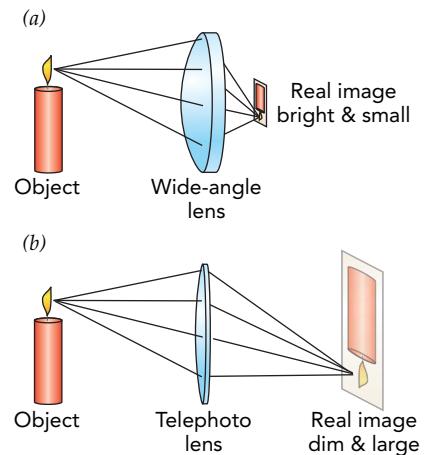


Fig. 14.1.7 (a) A wide-angle lens has a short focal length and forms a small, bright real image near the lens. (b) A telephoto lens has a long focal length and forms a large, dim real image far from the lens.

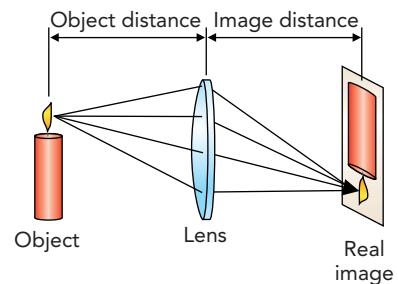


Fig. 14.1.8 The relationship between the object distance, the image distance, and the focal length of the lens is given by the lens equation.

TABLE 14.1.1 Several Cameras, the Widths of the Image Sensor They Use, and Their Normal Lenses

Type of Camera	Sensor Width	Normal Lens
Typical digital camera	8 mm	12 mm
35-mm camera	36 mm	50 mm
2 1/4-inch medium-format camera	2 1/4 inches	80 mm
5-inch portrait camera	5 inches	180 mm

According to the lens equation, the image distance for a distant object is equal to the focal length of the lens. That agrees with our earlier discussion of focal length. However, when the object is nearby, the image distance becomes larger than the focal length. That's why a camera lens moves away from the image sensor as you focus closer. When the object distance becomes less than the focal length, the image distance becomes negative and no real image forms at all. That's why you can't focus on an object that's too close to the lens.

A lens's **f-number** characterizes the brightness of the real image that it forms on the image sensor, with smaller f-numbers indicating brighter images. The f-number is calculated by dividing the lens's focal length by its diameter, or

$$\text{f-number} = \frac{\text{focal length}}{\text{diameter}}.$$

Since long focal length lenses naturally produce larger and dimmer images on the image sensor, the f-number takes into account both the light-gathering capacity of the lens and its focal length. Increasing a lens's diameter increases its light-gathering capacity and decreases its f-number. Increasing a lens's focal length decreases the brightness of its real image and increases the f-number. Doing both at once, increasing the lens diameter and focal length equally, leaves the brightness and f-number unchanged.

Most sophisticated cameras use large-diameter lenses so that their f-numbers are generally less than 4. Since it's difficult to fabricate a lens that's larger in diameter than its focal length, the smallest practical f-number is about 1. Also, because long-focal-length lenses need large apertures to keep their f-numbers small, some telephoto lenses are huge.

The diaphragm inside a lens allows you to decrease the lens's aperture and thus increase its f-number. A factor of 2 increase in f-number corresponds to a factor of 2 decrease in the lens's effective diameter and a factor of 4 decrease in the lens's light-gathering area. Thus when you double the f-number of the lens, you must compensate by quadrupling the exposure time. Although closing the aperture increases the lens's depth of focus, it requires a longer exposure.

Check Your Understanding #3: Bright and Sharp Photographs

On bright, sunny days, your automatic camera takes photographs with large depths of focus, while on dark, overcast days its pictures have much smaller depths of focus. What causes this difference?

Answer: On a bright day, your camera needs only a small aperture to gather enough light for an exposure, so the depth of focus is large. On a dark day, it uses the largest aperture available to gather light and has a small depth of focus.

Why: Light gathering and depth of focus go hand in hand. If you must open the aperture of your camera's lens to gather enough light for an exposure, focusing will become critical and your photographs will have small depths of focus.

Check Your Figures #1: The Image of an Apple

If the distance from an apple to a converging lens is twice the focal length of that lens, where will the real image of the apple form?

Answer: The image will form at a distance twice the focal length behind the lens.

Why: Since the object distance is twice the focal length, we can use Eq. 14.1.1 to find the image distance:

$$\begin{aligned}\frac{1}{\text{image distance}} &= \frac{1}{\text{focal length}} - \frac{1}{\text{object distance}} \\ &= \frac{1}{\text{focal length}} - \frac{1}{2 \cdot \text{focal length}} \\ &= \frac{1}{2 \cdot \text{focal length}}.\end{aligned}$$

The image distance is twice the focal length of the lens.

Improving the Quality of a Camera Lens

A high-quality camera lens isn't a single piece of glass or plastic. Instead, it's composed of many separate elements that function together as a single lens. This complexity improves the quality of the real image. To begin with, dispersion in a single-element lens causes different colors of light to bend differently and focus at different distances behind that lens. Known as *chromatic aberration*, this problem can be fixed by using several lens elements made of different types of glass or plastic with different amounts of dispersion. These elements compensate for one another so that the overall lens, known as an *achromat*, has very little dispersion and almost no color-focusing problems.

After correcting for color and other technical image problems, a sophisticated camera lens may contain more than 10 individual elements. For the purposes of the lens equation, this complicated lens has an effective center from which to calculate object and image distances. However, having so many separate elements creates reflection problems; each time light passes from air to glass or vice versa, some of it reflects. To avoid fogging the photographs with this bouncing stray light, the individual elements are *antireflection coated* with thin layers of transparent materials. The best coatings use interference effects to cancel out the reflected light waves and give the lens only a weak violet reflection.

Many modern cameras are equipped with zoom lenses. A zoom lens is a complicated lens that can change the size of the real image it projects onto the image sensor. By carefully moving its lens elements relative to one another, the zoom lens can adjust its effective focal length.

A common type of zoom lens contains three separate groups of lens elements and produces a sequence of three images (Fig. 14.1.9). The first lens group forms a first image of the scene in front of the camera. The second lens group forms a second image of that first image. The third lens group projects a third, real image of the second image onto the image sensor. Zooming the lens—that is, changing its focal length— involves altering the spacings between the lens groups to vary the second lens group's object and image distances and thus the relative sizes of the first and second images.

As the zoom lens changes from short focal length to long focal length, the image it projects on the sensor becomes larger. This effect allows you to compose the picture so that the scene fills the photograph completely. A lens that can change its focal length while retaining the same f-number and still keep the real image in focus on the sensor is a truly remarkable achievement.

Check Your Understanding #4: Not Such a Good Picture

If you use a large magnifying glass to form a real image of an overhead fluorescent light fixture, you will find that the corners of the light fixture are blurry and have rainbow colors in them. Why is the image quality so poor?

Answer: The magnifying glass has only one glass element and suffers from chromatic aberration (and many other image imperfections).

Why: A single converging lens can't form a high-quality image on a flat surface because it has chromatic aberration, spherical aberration, coma, and astigmatism. If the lens is small, these failings are often invisible. In a large magnifying glass, however, they are all readily apparent and you can't form a sharp image of the entire light fixture.

The Viewfinder and Virtual Images

SLR (single lens reflex) cameras permit you to change their lenses so that you can choose a lens that's optimized for the task at hand. When you peer through the viewfinder of an SLR camera (Fig. 14.1.10), you're looking at the same real image that will be projected

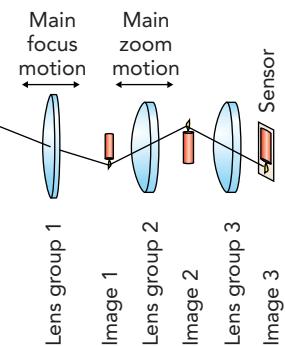


Fig. 14.1.9 A common type of zoom lens uses three lens groups to project a variable-size real image on the image sensor. Zooming is done mostly by moving the second lens group to change its image and object distances. The first lens group is responsible for focusing, and the third lens group projects the real image on the sensor.

Fig. 14.1.10 The mirror in the center of this reflex camera directs light from the lens (removed for this photograph) onto the focusing screen above it. During the exposure, the mirror swings upward to permit light from the lens to strike the image sensor at the back of the camera.



Courtesy Lou Bloomfield

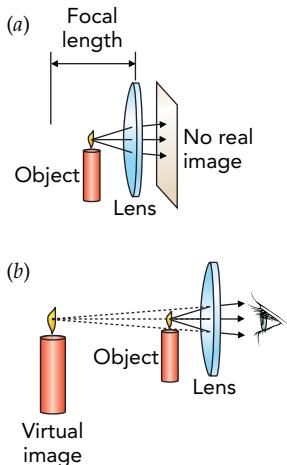


Fig. 14.1.11 (a) Light from an object very near a converging lens diverges after passing through the lens and no real image forms. (b) Your eye sees a virtual image that is large and far away.

onto the image sensor during the exposure. The light you see travels through the camera's main lens, reflects off a mirror, and projects onto a translucent screen inside the top of the camera. You're simply inspecting this screen and the real image through a magnifying lens in the eyepiece. During the exposure, the mirror flips out of the way and the real image projects briefly onto the image sensor.

Since the screen and real image are only an inch or two from your eye, you can't focus on them without the help of the eyepiece lens. The eyepiece lens is converging, but in this case it doesn't form a real image. Instead, it forms a **virtual image**, an image located at a negative image distance—that is, on the wrong side of the lens!

The screen displaying the scene that you're photographing is so close to the eyepiece lens that the object distance is less than that lens's focal length. According to the lens equation, the image distance should be negative, and it is; the image is located on the screen side of the eyepiece lens (Fig. 14.1.11). You can't put your fingers in the light and project this image on your skin because the image is virtual rather than real.

You can, however, see this image through the eyepiece. It's located farther away than the screen itself, so your eye can comfortably focus on it. Also, the image is magnified—the eyepiece lens is acting as a magnifying glass (Fig. 14.1.12). This lens provides magnification because, when you look at the screen through it, the screen image covers a wider portion of your field of vision. This magnification increases as the eyepiece lens's focal length decreases. That's because a shorter-focal-length eyepiece lens must be quite close to the screen, where it can bend light rays coming from a smaller region so that they fill your field of vision. The eyepiece lens in a typical camera has been chosen so that the screen fills a comfortable portion of your visual field, allowing you to examine the virtual image in great detail and adjust the lens and camera settings until you have just the right picture in your view. Then all you have to do is take the photograph.

Cameras with fixed lenses often have two separate viewfinder systems. A typical digital camera has an electronic viewfinder, which displays the real image being projected onto its image sensor. Many digital and most film cameras also have optical viewfinders.

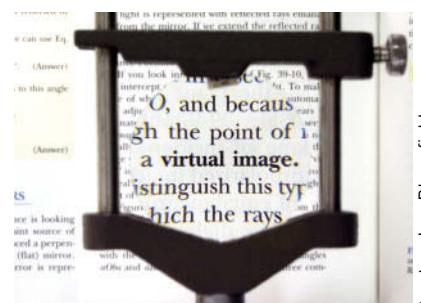


Fig. 14.1.12 This magnifying glass creates an enlarged virtual image located far behind the printed text. You can't touch the image or put your fingers in its light, but you can see it clearly with your eyes.

Courtesy Lou Bloomfield

Although optical viewfinders vary in style and sophistication, the best combine real and virtual images. In a *real-image optical viewfinder*, a system of lenses, mirrors, and/or prisms produces an erect real image of the scene and you then examine that real image through an eyepiece magnifying glass. The lenses projecting the real image zoom along with the camera's main lens so that what you see through the viewfinder is similar to what the camera's image sensor will record.

Check Your Understanding #5: Real and Virtual Images

As you move a magnifying glass slowly toward the photograph in front of you, you see an inverted image that grows larger and nearer to your eye. This image eventually becomes blurry, and then a new upright and enlarged image appears on the photograph's side of the lens. What's happening?

Answer: The lens is initially creating a real image near your eye. This real image moves past you as the lens approaches the photograph. Finally, the lens is close enough to create an enlarged virtual image of the photograph.

Why: A magnifying glass forms either a real or virtual image, depending on object distance and the lens's focal length. If the lens and photograph are separated by more than the focal length, the image is real and you see it near your eye. You can touch this inverted image or project it on a piece of paper. If the lens and picture are separated by less than the focal length, the upright image is virtual and you see it on the far side of the lens.

Image Sensors

Once the lens has projected its real image onto the image sensor, it's the image sensor's job to record that pattern of light. Interestingly enough, both film and electronic image sensors use semiconductors, and both detect light when its photons shift electrons from valence levels to conduction levels. However, how those two image sensors act on the electron transitions is quite different.

Photographic film detects light photochemically. Embedded in the film are tiny crystals of silver salts. Composed primarily of silver and halogen atoms, these semiconductor crystals are extremely sensitive to light. When a silver halide crystal absorbs a photon of visible light, it can undergo a radiative transition that shifts an electron from a valence level to a conduction level and eventually frees one silver atom from a silver halide molecule. After several nearby silver atoms have been freed by light, they can form a tiny particle of silver metal. When the film is developed, this silver particle transforms the entire silver halide crystal into metallic silver. The microscopically rough structure of that silver makes it appear black rather than shiny.

In black-and-white photography, the silver particles themselves form a negative image on the developed film. Wherever the film was struck by light, it acquires a dense, black pattern of silver particles. Wherever light was absent, the film becomes clear once the unexposed silver salts are washed away. Although the image on the developed film itself is negative—light is dark and dark is light—the process of preparing photographic prints reverses light and dark a second time so that the image on the prints is positive.

In color photography, the silver halide crystals are exposed to light through color filters and sensitizers, so that the film separately records its exposure to the three primary colors of light (see Section 13.2). During development, the silver itself is washed away, but a negative color image remains in the film. For example, wherever blue light struck the film, it acquires a yellow tint and therefore absorbs blue light. Again, the photographic printing process reverses the colors a second time so that the prints have positive images.

Electronic image sensors are based on **photodiodes**, diodes that are optimized to detect light. They combine the light-sensing behavior of photoconductors (see Fig. 13.3.4) with the current-controlling behavior of diodes (see Fig. 13.3.6). A vast array of these

photodiodes record the pattern of light in the real image, and the camera subsequently reads that pattern by measuring the accumulated charges on each of its photodiodes. To obtain color information, the image sensor's photodiodes are covered with a pattern of red, green, and blue filters so that each photodiode measures the intensity of only one primary color of light.



Check Your Understanding #6: Foggy Photos

Airport screening devices often use X-rays to search for hidden items. These high-energy photons can damage film. How?

Answer: X-rays cause radiative transitions in the silver salt crystals.

Why: X-rays can penetrate through normally opaque objects and reach the film. If a silver salt crystal in the film absorbs an X-ray, it will respond as though it was exposed to visible light. Although most modern screening devices use such weak X-ray sources that the effects are minimal, repeated screenings of high-speed film will gradually ruin it.

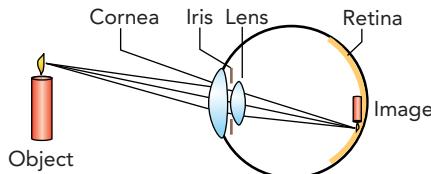


Fig. 14.1.13 An eye is a camera, with its cornea and lens forming a real image on the retina. The eye focuses by changing the curvature of its lens. The iris changes the eye's f-number.

Eyes and Eyeglasses

Not all cameras involve modern technology. Most people are born with two of them—their eyes. Like the cameras we have discussed, each eye consists primarily of a converging lens and an image sensor (Fig. 14.1.13). In this case, the lens is a combination of the front surface of the eyeball, its *cornea*, and the internal lens just beneath the cornea. The image sensor is the *retina*, a vast pattern of light-sensitive cells and nerves at the back of the eyeball.

When you look at the scene in front of you, the cornea and lens of your eye project a real image of that scene onto your retina and your retina reports the resulting pattern of light to your brain. As usual, the real image is inverted and reversed left to right, but your brain compensates for that effect.

Since your eyeball can't alter the distance between the lens and the image sensor, it focuses the real image by adjusting the focal length of the lens. When you look at nearer objects, the lens in your eye becomes more highly curved and its focal length decreases. The light rays from that nearer object thus converge more sharply and form a real image on your retina. When you view a more distant object, the lens becomes less curved and its focal length increases.

Like a sophisticated camera, your eye has an iris within its lens system. When you view a bright scene, that iris shrinks to limit the amount of light striking your retina. As a side effect, your depth of focus increases and everything appears sharper. It's easier to focus when you read or work in a well-lit environment.

However, not all eyes are perfect, and many need help forming sharp real images on their retina. Although modern laser surgical techniques can reshape corneas to improve image sharpness, the classic approach to better vision is to wear eyeglasses or contact lenses. An eye's lens system already consists of two components, the cornea and the lens, so adding a third component, eyeglasses, is no big deal.

A person who is farsighted can't see nearby objects sharply because her lens system has too long a focal length (Fig. 14.1.14a). Although it can project real images of distant objects on her retina, nearby objects focus too far away from the front of her eye and the light strikes her retina before it forms a real image.

To compensate for farsightedness, she wears eyeglasses with converging lenses (Fig. 14.1.14b). These lenses begin the task of bending light rays together even before they

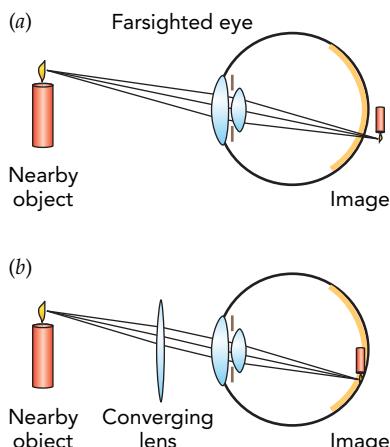


Fig. 14.1.14 (a) A farsighted eye bends light too weakly to focus on a nearby object. The real image forms beyond the retina. (b) A converging lens shifts the real image forward so that it focuses on the retina.

enter her eyes. Her own lens system completes the bending, and the real images form closer to the front of her eyes. She is thus able to see nearby objects clearly.

In contrast, a person who is nearsighted is unable to focus on distant objects because his lens system has too short a focal length (Fig. 14.1.15a). The real images of those distant objects form too close to the front of his eye, and the light has already begun to spread apart by the time it reaches his retina.

To compensate for nearsightedness, he wears eyeglasses with diverging lenses (Fig. 14.1.15b). A **diverging lens** is one that bends light rays apart and therefore has a negative focal length. Typically thinner at its middle than at its edge, a diverging lens bends the nearly parallel rays of light from a distant object so that they diverge more rapidly. Those rays then appear to come from a much nearer object, actually a nearby virtual image, and his eyes are able to focus them properly on his retina.

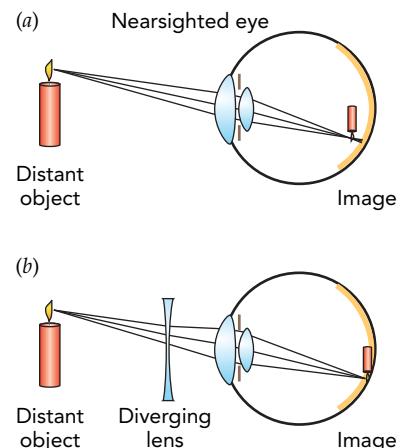


Fig. 14.1.15 (a) A nearsighted eye bends light too strongly to focus on a distant object. The real image forms before the retina. (b) A diverging lens shifts the real image backward so that it focuses on the retina.

Check Your Understanding #7: Keeping Up Your Image

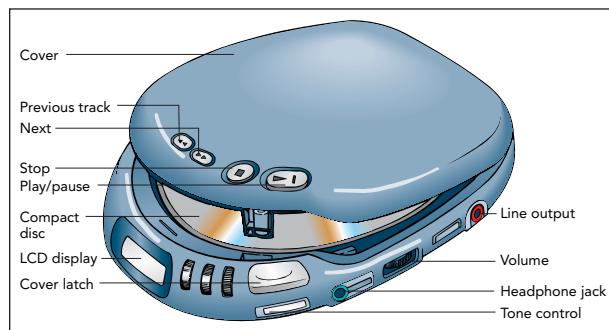
The eyeglasses of a farsighted person can project a real image of a distant scene on a white wall. However, the eyeglasses of a nearsighted person produce no real image. Why not?

Answer: Nearsighted eyeglasses use diverging lenses, which bend light rays apart and don't focus them together into a real image.

Why: To form a real image of a distant scene, you need to bring light rays together with a converging lens. Diverging lenses spread light rays apart and can't form real images on their own.

SECTION 14.2

Optical Recording and Communication



Using light to convey information is as old as signal fires and as natural as sight itself. Over the years, advances in light sources, optical materials, and electronics have radically increased the possibilities for optical information systems. Optics and information go so well together that they're partly responsible for the current information revolution. The introduction of compact disc players in the early 1980s transformed the music industry virtually overnight, and optical fibers are knitting our world together at an astonishing pace. In

this section, we'll look at how optical devices use light to manipulate information.

Questions to Think About: Where is a CD's or DVD's information stored? How do CD or DVD players ignore fingerprints, dust, and other surface contamination during playback? Why is a laser involved in recording and playing back a CD or DVD? Why can a Blu-ray disc hold more information than a DVD and a DVD more than a CD? Why are CDs and DVDs so free of noise? How can a glass fiber direct light in a curving path? How can light travel through kilometers of glass fiber without becoming dim?

Experiments to Do: Hold a prerecorded CD, DVD, or Blu-ray by its edges, and look at its unlabeled surface. Beneath the clear plastic face is a smooth, shiny layer that reflects a rainbow of colors. Tiny pits in this layer cause this coloration. These pits form a spiral track around the center of the disc, and adjacent arcs of this track are so closely spaced that light waves reflecting from them interfere, as from a soap film. The reflective layer is so thin that you can see through it if you hold it in front of a bright light. Can you see the reflections of dust particles on the unlabeled surface? How deep below the plastic surface is the shiny layer?

Representing Sound and Light: Analog and Digital

A CD doesn't store sound any more than a DVD or Blu-ray stores flickering light. Instead, these discs store *representations* of sound and light that it can use to recreate them on demand.

Embedded in each disc is enough information to reproduce a concert or a movie, and to do so almost perfectly. Between the microphones or camera that originally collected that information and the headphones or home theater that finally reconstructs the sound or light are a number of fascinating processes, some optical and some electronic. In this section, we concentrate on the optical processes. In Section 14.3, we'll look at the electronic ones.

In Section 12.1, we saw that a radio wave can be used to represent sound. Sound waves are fluctuations in air's density, and the AM technique represents those density changes by changes in the amplitude of a radio wave. The FM technique is similar, but it represents those density changes by small changes in the frequency of a radio wave. Both techniques are **analog representations** of the sound, meaning that a continuously variable physical quantity (a radio wave's amplitude or frequency) represents another continuously variable physical quantity (air's density). Like any analog representation, the radio transmitter is drawing an *analogy* between two continuously variable physical quantities—amplitude or frequency is serving as an *analog* for air density. The radio receiver draws the reverse analogy and thereby recreates the sound itself.

Although parts of the CD player use analog representations for sound, the optical recording and playback part uses a different representation—digital. In **digital representations**, a continuously variable physical quantity is first represented by a series of numbers and then each of those numbers is represented by a set of physical quantities—a set of digits—each of which can have only a limited number of discrete values. Those discrete values are known as symbols, and symbols can be just about anything, as long as you can tell them apart. As we'll see in a moment, the numerals 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9 are such symbols, and the set of numerals 392 is a digital representation of the number *three hundred ninety-two*.

Suppose that you're recording your band and want a digital representation of its sound. The first step is to represent that sound's density fluctuations as a series of numbers. With the help of a microphone and some other equipment, you measure the air density many times per second and obtain the series of numbers. Let's assume that one of those numbers is *one hundred twenty-four*, meaning that the air density during that measurement was 124 units above the normal density.

The second step is to represent each number by a set of digits. We'll allow each digit to hold 1 of 10 possible symbols. Those symbol choices could be ¥, ☺, ☻, ☻, ☻, ☻, ☻, ☻, ☻, and ☻, but let's use a more familiar collection, the numerals 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9. Note that the symbol 5 isn't actually a number; it's just a shape consisting of two straight lines and partial circle.

With a single digit and our set of 10 symbols, we can represent the numbers *zero* through *nine*. With 2 digits and those same 10 symbols, we can represent the numbers *zero* through *ninety-nine*. With three digits, we can represent *zero* through *nine hundred ninety-nine*. Since we're trying to represent the number *one hundred twenty-four*, we clearly need at least three digits. When we place the symbols 1, 2, and 4 in those three digits, as 124, we are representing the number *one hundred twenty-four*.

The convention that we are following when letting 124 represent *one hundred twenty-four* is known as **decimal**. In decimal, we break numbers into *ones*, *tens*, *hundreds*, *thousands*, and so on—the powers of 10—and use 1 digit and a set of 10 symbols to represent how much of each power of 10 is present in the number we're representing. The decimal representation of *one hundred twenty-four* is 124, meaning that it contains 1 *hundred* (10^2), 2 *tens* (10^1), and 4 *ones* (10^0). When these pieces are added together, they sum to *one hundred twenty-four*.

What if, instead of 10 symbols, our collection had only 2 symbols: 0, 1? We could still represent large numbers, but we would need many more digits than before. We would have to break numbers into *ones*, *twos*, *fours*, *eights*, *sixteens*, and so on. Instead of using the powers of 10 (as in decimal), we would be using the powers of 2. This system for representing numbers using the powers of 2 is called **binary**.

In binary, *one hundred twenty-four* is written as **1111100**, meaning that it contains 1 *sixty-four* (2^6), 1 *thirty-two* (2^5), 1 *sixteen* (2^4), 1 *eight* (2^3), 1 *four* (2^2), 0 *twos* (2^1), and 0 *ones* (2^0). When these pieces are added together, they again sum to *one hundred twenty-four*. This apparently complicated way to represent even a fairly small number is actually quite useful. The number has been broken into pieces that have only two possible values; there is either a *thirty-two* in the number being represented or there isn't.

Binary representations require only two different symbols, and those symbols can be any distinguishable objects. We've been using (**1** and **0**), but we could use (**heads** and **tails**), (**smoke** and **no-smoke**), (**dot** and **dash**), (**shiny** and **dull**), (**charged** and **uncharged**), and so on. Inside an optical disc, for example, there is a surface covered with shiny and dull spots, and those spots are a binary representation of numbers. Similarly, in a computer, there are capacitors that are charged or uncharged, and those capacitors are a binary representation of numbers.

There are good reasons for using digital representation to record your band's sound and for using binary. Analog representations are inherently noisy because every accidental change in the physical quantity doing the representing is interpreted as a change in the physical quantity being represented. That's why a phonograph record, which uses height fluctuations in a groove to represent the density fluctuations of sound, can't re-create that sound perfectly. Whenever dust settles into the groove on the record and introduces its own height fluctuations, the phonograph reproduces inaccurate sound, full of pops and snaps.

In contrast, digital representations are noise free. Dust and other imperfections have no effect on digital representations as long as the symbols can still be distinguished for one another. For example, even without brushing the cracker crumbs off your book, laptop, or tablet, you can still read **5049** or **101001** with perfect accuracy, unless you're a total slob. Even when some of the symbols are so badly obscured that they cannot be read, extra symbols can be incorporated into the digital representation to correct for such reading errors.

The reason that binary is so useful in both optical and electronic devices is that working with 2 symbols is much easier than working with 10 of them. The symbols used in these devices aren't shapes on a sheet of paper or computer screen, they're things like the shininess of a surface or the charge on a capacitor. It's easy to distinguish a dull spot (**0**) from a shiny spot (**1**) inside an optical disc, but it's much harder to distinguish between dull (**0**), not quite dull (**1**), slightly shiny (**2**), . . . , very shiny (**8**), and extremely shiny (**9**) spots. Although there are modern technologies that use more than two symbols, notably digital television transmission, binary is much more common.

Check Your Understanding #1: New Math

What number does binary **10000001** represent?

Answer: One hundred twenty-nine.

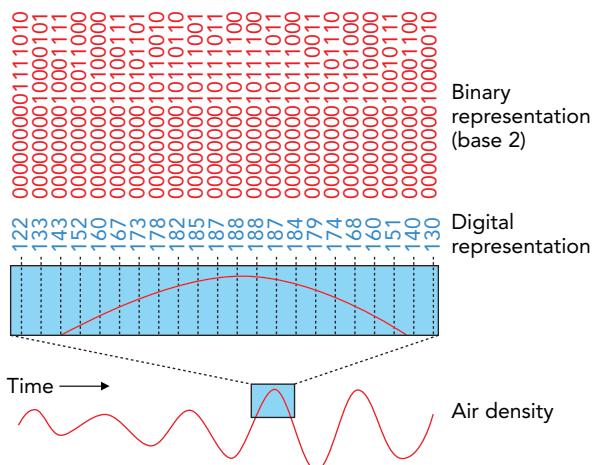
Why: Binary **10000001** contains only 1 *one hundred twenty-eight* (2^7) and 1 *one* (2^0). All the other powers of 2 are not present. Since $128 + 1$ is 129, that is the number represented by this binary value.

Digital Recording

In the conventional CD format, density measurements are made 44,100 times each second for two independent audio channels and these measurements are recorded on the disc in binary form, using 16 bits for each measurement (Fig. 14.2.1). Since the air density can go down as well as up, these bits represent the positive and negative integers from -32,768 to 32,767, which in turn represent how much the air density is above or below the average density. Density measurements with 16 bits of precision are sufficient to reproduce both loud and soft music with almost perfect fidelity.

Like CDs, DVDs record information digitally. But DVDs are a newer technology and therefore more sophisticated. Audio DVDs can choose from several measurement rates, bits per

Fig. 14.2.1 Sound can be represented as a series of numbers. Each number corresponds to the air density at a particular moment in time.



measurement, and numbers of channels. A typical DVD might have five audio channels: left-front, center-front, right-front, left-rear, and right-rear. The three front channels might have 96,000 density measurements per second at 24 bits per measurement, and the two rear channels might have 48,000 measurements per second at 20 bits per measurement. All these samples, bits, and channels involve far more information than is stored on a CD, and a DVD compresses that information before storing it. In contrast, a conventional CD's information is uncompressed although some more modern formats (e.g., mp3) do employ compression techniques.

In either case, air density measurements aren't simply recorded one after another on a disc's surface. Instead, these numbers are extensively reorganized before they're stored. This reorganization allows the player to reproduce the sound perfectly even if the disc can't be read completely. As we'll see shortly, reading these discs is a technological tour de force and susceptible to various failures. To be sure that the sound (and video) can be reproduced completely and without interruption, the numbers are recorded in an encoded manner. They appear redundantly so that, even if one copy of a number is illegible, there is still enough legible information along the same arc of the disc's spiral track to completely re-create that missing number. This duplication of information reduces the playing time of both CDs and DVDs, but it is essential for reliability.

Its encoding scheme leaves a CD or DVD almost completely immune to all but the most severe playback problems. In principle, you can damage or obscure a 2-mm-wide swath of the disc, from its center to its edge, and the player will still be able to reproduce the sound (and video) perfectly. However, damage along an arc of the spiral track is far more threatening to the data. If the player can't read a long stretch of a single arc, it won't be able to recover the information. That's why you should always clean a CD or DVD from its center outward to its edge.

Check Your Understanding #2: Cordless Clarity Away from Home

Cell phones transmit your voice in digital form. In general terms, how does this digital transmission work?

Answer: Your cell phone's microphone converts the sound of your voice into a fluctuating current. This current is measured periodically and represented as a sequence of numbers. These numbers are then represented by radio waves and transmitted through space. The numbers eventually work their way to a receiving cell phone, where the radio waves are converted back into numbers and the numbers are converted back into a fluctuating current. Finally, the current is amplified and sent through a speaker to create sound.

Why: Many modern communication devices use a digital representation of sound. In radio-wave communication, digital representations are particularly useful because they are immune to the noise that plagues analog radio transmissions and also permit more efficient use of the radio bandwidth.

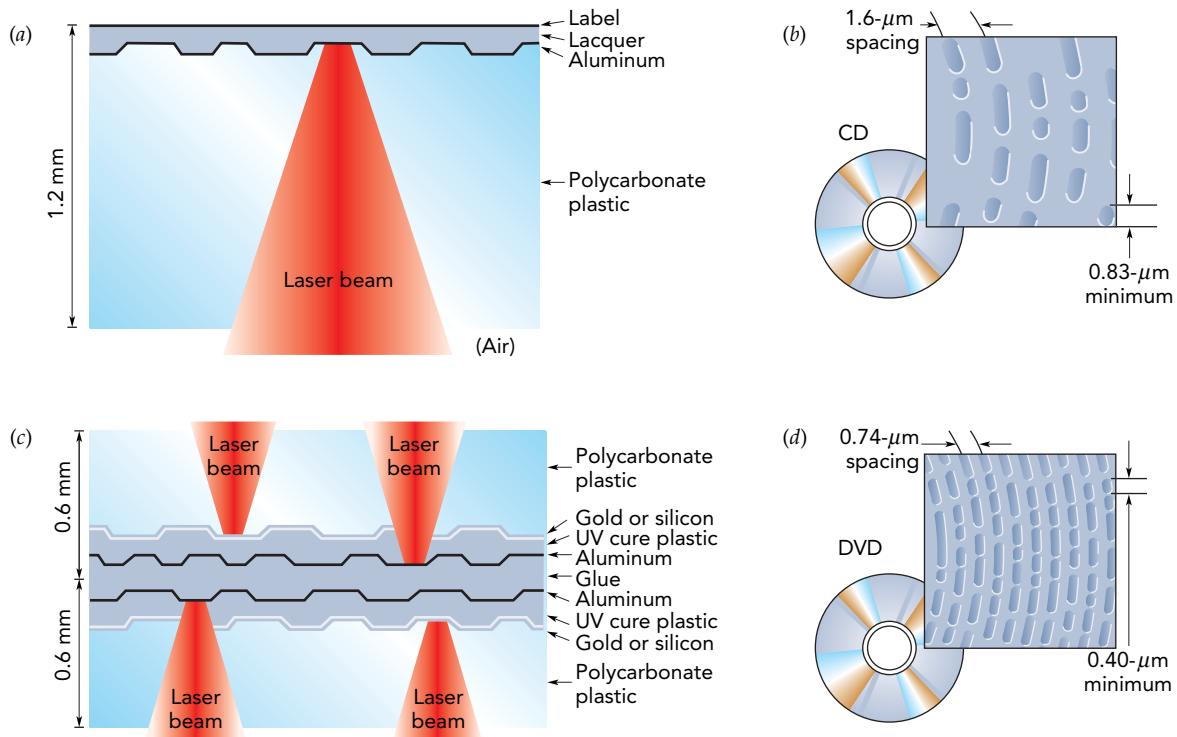


Fig. 14.2.2 (a,b) A CD contains a thin layer of aluminum on one side of a clear plastic disc. The aluminum layer has tiny pits that are detected by a 780-nm laser beam. (c,d) A DVD contains up to four reflective layers of aluminum, gold, or silicon sandwiched between two clear plastic discs. The gold or silicon layers are semireflective. Pits in those layers are detected by a 650-nm laser beam. The beam can focus either on the semireflective layer or, by passing through that layer, on the aluminum layer beyond it.

The Structure of CDs, DVDs, and Blu-rays

Standard CDs, DVDs, and Blu-rays are 120 mm (4.72 in.) in diameter and 1.2 mm (0.05 in.) thick. One side of a CD is clear and smooth but the other side contains a sandwich of layers: a thin film of aluminum, a protective lacquer, and a printed label (Fig. 14.2.2a). In contrast, a DVD is laminated from two 0.6-mm-thick clear plastic discs, with one, two, or four reflective layers of aluminum, gold, or silicon stacked up in between them (Fig. 14.2.2c). The more layers in the DVD, the more information it holds. A Blu-ray disc (BRD) also has 1 to 4 reflective layers, but they are so near the BRD's surface that a hard protective coating is applied to prevent scratches.

The reflective layers are the recording surfaces. These layers are so thin that they actually transmit a small amount of light. In the gold or silicon DVD layers, this semitransparency is essential because it allows the optical system that reads information to send light through the semitransparent layer to the aluminum layer beyond it. The aluminum layers also transmit some light. Although aluminum's electrons accelerate in response to the light's electric field and normally reflect that light completely, there aren't enough electrons in these 50- to 100-nm-thick layers to do the job and some light gets through.

The reflective layers aren't perfectly smooth. Instead, each has a narrow spiral track formed in its surface (Figs. 14.2.2b,d). This track is a series of microscopic pits, as short as 0.83 μm long on a CD, 0.40 μm long on a DVD, and 0.15 μm long on a Blu-ray. Adjacent arcs in the spiral track are only 1.6 μm apart on a CD, 0.74 μm apart on a DVD, and just 0.32 μm apart on a Blu-ray. The lengths of the pits and the flat "lands" that separate them represent numbers. The player examines these pits and lands as the disc turns and converts their lengths into numbers, sound, and video.

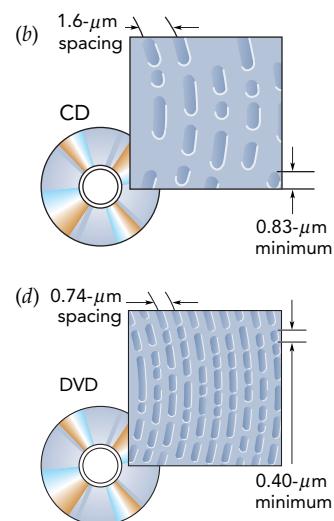


Fig. 14.2.2 (repeated)
 (b) A CD's aluminum layer has tiny pits that are detected by a 780-nm laser beam. (d) A DVD's reflective and semireflective layers have pits that are detected by a 650-nm laser beam. Using a shorter-wavelength laser allows the DVD player to focus its light to a smaller spot and observe the DVD's narrower, more closely spaced reflective features.

The pit lengths and the spacings between arcs weren't chosen arbitrarily. Since electromagnetic waves are unable to detect structures much smaller than their wavelengths (see Section 12.2), the laser beam's wavelength limits the size of the smallest features on a disc. In a CD player, that beam's wavelength is 780 nm in air and 503 nm in polycarbonate plastic—short enough to detect the pits of a CD easily. In a DVD player, the laser beam's wavelength is between 635 and 650 nm in air and between 410 and 420 nm in plastic—just short enough to detect the pits in a DVD. The wavelength reduction inside the disc occurs because polycarbonate plastic has an index of refraction of 1.55, meaning that the light's speed in that plastic is reduced from its vacuum speed by a factor of 1.55. Its wavelength is reduced by the same factor. In a Blu-ray player, the laser beam's wavelength is 405 nm in air and about 260 nm in plastic, and it is just able to observe the tiny pits.

The player detects a pit by bouncing light from the disc and determining how much of it reflects. As the focused laser beam passes over a pit, the reflection becomes dim, in part because the curved pit scatters light in all directions and in part because of interference effects. Light that's reflected back from a pit travels farther than light that's reflected from the flat region around it, so electric and magnetic fields in the two waves are shifted relative to one another. The pit depth was chosen so that the two reflected waves are approximately out of phase and they interfere destructively. Overall, the player's light sensors detect relatively little light when the laser beam is located over a pit.

A CD, DVD, or Blu-ray player uses a laser diode to produce its light. The 780-nm standard for CD players was adopted in 1980, when 780-nm infrared laser diodes were reliable but still fairly expensive. Technology advanced by the mid-1990s, however, and the 635- to 650-nm standard for DVD players reflects the development of inexpensive red laser diodes. New standards follow technology, so with the development of reliable blue laser diodes, new optical recording systems appeared. Blu-ray players and discs, which first appeared in 2006, are based on a 405-nm laser. Because a Blu-ray player can focus its blue laser beam to a much smaller spot than either a CD or a DVD player, a Blu-ray disc holds far more information than either of the earlier formats.

Check Your Understanding #3: The Blue-Light Special

Blu-ray disc players use blue lasers with a wavelength of 405 nm in air or vacuum. A CD player uses an infrared laser with a wavelength of 780 nm. How must the pit depths in the reflective layer of a Blu-ray disc compare to the pit depths in a CD?

Answer: The pits in a Blu-ray disc must be about half as deep as those in a CD.

Why: The pits in the reflective layer should be about a quarter of a wavelength deep so that light reflected from the bottom of a pit interferes destructively with light reflected from an adjacent flat region. Halving the wavelength of the laser light requires halving the depth of the pits to achieve destructive interference.

The Optical System of a CD, DVD, or Blu-ray Player

A CD, DVD, or Blu-ray player's optical system measures the lengths of the tiny pits as they move by on a spinning disc. That reading process requires incredible precision. Not only must the player focus its spot of laser light exactly on the reflective layer, but it must also follow the spiral track as it moves by. The disc itself is neither perfectly flat nor perfectly round, so the player must continuously adjust its reading unit during playback. The optical system must keep its laser beam focused on the reflective layer (autofocusing) and must follow the track as it passes (autotracking). These two automatic processes are beautiful examples of the use of feedback.

The basic structure of a typical CD, DVD, or Blu-ray player is shown in Fig. 14.2.3. Light from a laser diode passes through several optical elements on its way to the disc's reflective layer. It comes to a tight focus on that layer, where it illuminates only a single

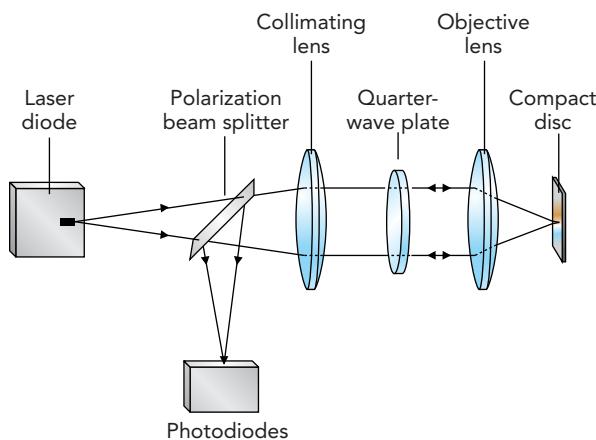


Fig. 14.2.3 In the optical system of a CD, DVD, or Blu-ray player, light from a laser diode passes through a polarization beam splitter, a collimating lens, a quarter-wave plate, and an objective lens before focusing on the reflective layer inside the disc. Reflected light turns 90° at the polarization beam splitter and focuses on an array of photodiodes.

track. Some light reflects from the layer and returns through the optical elements. Finally, the reflected light turns 90° at a special mirror called a polarization beam splitter and focuses on an array of light detectors. The player measures the electric currents flowing through the detectors and uses those measurements both to obtain data from the disc and to control the focusing and tracking systems.

Let's examine this optical system one element at a time. After leaving the laser diode, light passes through a *polarization beam splitter*. This device analyzes the light's polarization. As we saw in Section 13.1, different polarizations of light reflect differently when they strike a transparent surface at an angle. In this case, polarized light from the laser passes through the 45° surface, but light of the other polarization reflects. The beam splitter is specially coated to separate the two polarizations almost perfectly.

Light from the laser diode diverges rapidly as it passes through the beam splitter. It's not that the laser diode is broken or poorly designed; it's that a light wave emerging from a small opening naturally spreads outward, like ripples on a pond. This spreading is known as **diffraction** and occurs whenever a light wave is truncated by passing through an opening. The smaller the opening, the worse is the spreading. Because the emitting surface of the laser diode is essentially a very small opening, the laser beam spreads rapidly as it heads away from the diode. The player uses a converging lens located after the beam splitter to stop this spreading. At that point, the light beam is already wide enough that diffraction causes little additional spreading. The beam leaves the lens *collimated*, meaning that it maintains a nearly constant diameter after passing through the lens.

The laser light then passes through a *quarter-wave plate*. This remarkable device performs half the task of converting horizontally polarized light into vertically polarized light or vice versa. Horizontally and vertically polarized lights are said to be **plane polarized** because their electric fields always oscillate back and forth in one plane as they move through space. The quarter-wave plate turns plane polarized light into circularly polarized light. We encountered circular polarization before in the radio transmissions from FM stations. In **circularly polarized** light, the electric field actually rotates about the direction in which the light is traveling.

Now the light passes through an objective lens that focuses it onto the reflective layer of the disc. On its way to the reflective layer, the light enters the plastic surface of the disc. At its entry point, the beam is still more than 0.5 mm in diameter, which explains why dust or fingerprints on the disc's surface don't cause much trouble. While contamination may block some of the laser light, most of it continues onward to the reflective layer.

The light comes to a tight focus just as it arrives at the reflective layer. Although it might seem that all the light should converge together to a single point on that surface, it actually forms a spot roughly 1 wavelength in diameter. This spot size is limited by the

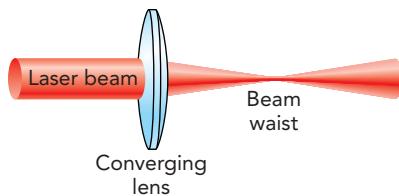


Fig. 14.2.4 When a converging lens focuses a laser beam, the light doesn't meet at a single point in space. Instead, it reaches a narrow waist with a diameter roughly equal to the wavelength of the light.

wave nature of light. No matter how perfectly you try to focus light, you can't make a spot that's much smaller than the light's wavelength. Instead, the beam forms a narrow waist and then spreads apart (Fig. 14.2.4). This *beam waist* is about 1 wavelength of the light in diameter and a few wavelengths long, depending on the f-number of the converging lens. Since that waist is less than $2 \mu\text{m}$ long, the player's autofocus system must keep the objective lens just the right distance from the reflective layer.

This fundamental limitation on how tightly a beam of light can be focused is another example of diffraction; the focusing lens truncates the light wave and inevitably introduces spreading. However, even reaching this ideal focusing limit requires careful design and fabrication of the optical elements. Although most other optical systems fall short of their ideal limits, a CD, DVD, or Blu-ray player's optical system does as well as can be done within the constraints of diffraction itself. Its optics are essentially perfect and are said to be *diffraction limited*.

The amount of light that reflects from the layer depends on whether or not the laser spot hits a pit. This reflected light follows the optical path in reverse. The light is collimated by the objective lens and then passes through the quarter-wave plate again. The plate now finishes the job it started earlier; the light ends up plane polarized but with the opposite polarization it had when it left the laser. Horizontally polarized light is now vertically polarized, and vice versa.

The reflected light then passes through the collimating lens, which makes the light converge, and then strikes the polarization beam splitter. Because the light's polarization has changed, the beam splitter no longer allows the beam to pass directly through. Instead, it turns this reflected beam 90° and directs it toward the detector array. This clever redirection scheme is important for two reasons. First, it conserves laser light by allowing most of it to travel from the laser diode to the detector. Second, it prevents reflected light from returning to the laser diode, where that light would be amplified and cause the laser diode to misbehave.

The light comes to a focus on an array of photodiodes. This array allows the player to detect pits via the reflected light intensity and also to determine whether the objective lens is properly positioned relative to those pits. For generality, Fig. 14.2.3 omits optical elements involved in autofocus and autotracking. However, because of those elements, the pattern of light hitting the detector array indicates which way the objective lens should move, if necessary. That lens is attached to coils of wire that are suspended near permanent magnets. By varying the currents flowing through those coils, the player uses Lorentz forces to move its objective lens about rapidly and keep it in the right place over the disc.



Check Your Understanding #4: That's Why It's Called a Laser Disc

Why must the CD, DVD, or BRD player use a laser diode rather than a more conventional light source such as an incandescent bulb?

Answer: The light in a CD, DVD, or BRD player must be coherent so that it can be focused to a single diffraction-limited spot on the reflective layer and experience destructive interference when it encounters a pit.

Why: You can't focus light from an incandescent bulb to a diffraction-limited spot. Each photon leaving the bulb is independent and will focus at a somewhat different point in space. The bulb's incoherent light also doesn't experience the strong interference effects of laser light.

Optical Fibers

Optical playback of prerecorded discs is fine for music and movies, but it isn't of much use when you want up-to-the-minute information. The Internet and World Wide Web require communication links that operate at lightning speed. Yet, even here, optics and light have an important role to play. The fastest way to send enormous amounts of information is to use optical fibers.

An optical fiber is a glass conduit that guides light from one place to another. Nearly every photon that enters the fiber at one end emerges from the other end moments later. In its simplest form, the fiber is made from two different glasses: a solid core of one glass surrounded by a cladding of the other glass. Both glasses are so incredibly transparent that light can travel through them for kilometers with little loss. For comparison, look through the edge of an ordinary piece of window glass and you'll see how dark the glass looks. It absorbs far too much light to be suitable for optical fibers. They're made of the purest glasses known.

If both glasses are almost perfectly transparent, what keeps the light from leaking out of the sides of the fiber? The answer is a phenomenon known as *total internal reflection*. As light tries to move from the inner glass core to the outer glass cladding, it's reflected perfectly and thus can't escape.

Total internal reflection is an extreme case of refraction. When light encounters the boundary between two materials with different indices of refraction, refraction causes that light to bend (Fig. 14.2.5). If the material it enters has a smaller index of refraction than the one it leaves, the light bends away from a line perpendicular to the boundary. The amount of this bend depends on the two indices of refraction and on the angle at which the light approaches the boundary. As long as the approach angle is steep enough, light will succeed in entering the second material. However, if the approach angle is too shallow, the light won't enter the second material at all. Instead, it will reflect perfectly from the boundary. In fact, total internal reflection is far more efficient at reflecting light than a conventional metal mirror.

To keep the light inside the fiber's core, the core glass must have a higher index of refraction than the cladding glass. As light in the high-index core encounters the boundary with the low-index cladding, it experiences total internal reflection and bounces back into the core (Fig. 14.2.6a). As long as the fiber doesn't bend too sharply, the light bounces back and forth inside the core and can't escape. As a result, light entering the fiber core through one of its cut ends follows the fiber all the way to its other cut end.

A large-diameter fiber (typically 50 μm or more) has a problem. Light rays bouncing through it at slightly different angles travel different distances during their passage through the fiber. Light heading almost straight down the center of the fiber rarely bounces and takes less time to complete its trip than light that bounces many times. Because this wide fiber has many bouncing paths or "modes" in which light can travel through it, a short pulse of light going through the fiber gets stretched out in time (Fig. 14.2.6a). This pulse-broadening severely limits the rate at which information can be sent through a multimode fiber.

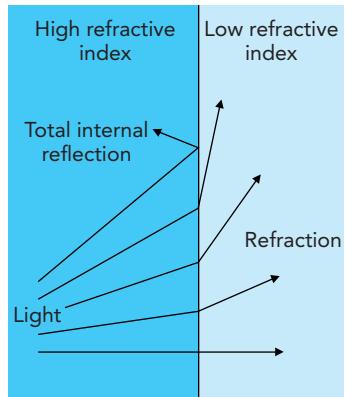


Fig. 14.2.5 When light traveling through one material enters a second material with a lower refractive index, it bends toward the boundary between those materials. If its approach angle is too shallow, the light will bend so much that it will simply reflect from the boundary. This effect is called total reflection.

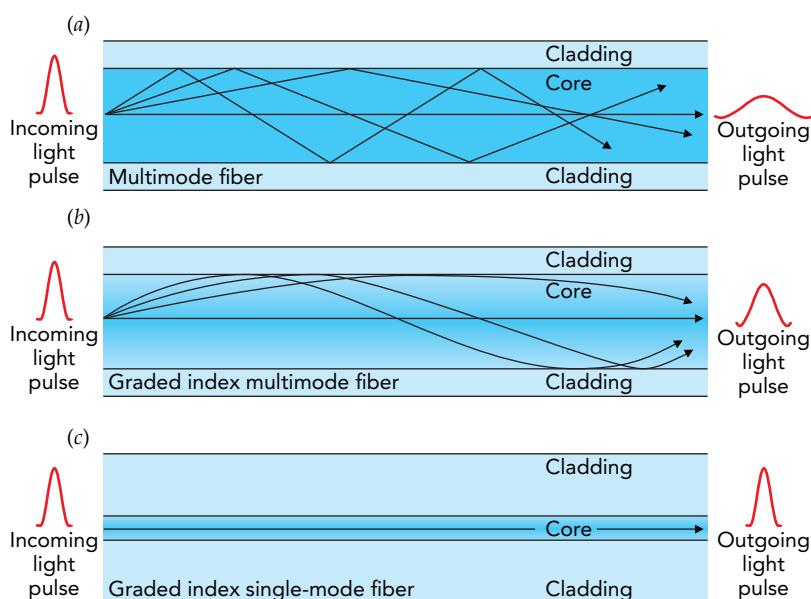


Fig. 14.2.6 Light trying to leave the high-index core of an optical fiber at a shallow angle undergoes total internal reflection at the boundary with the low-index cladding. (a) In an ordinary multimode fiber, a pulse of light can follow many paths through the core and becomes spread out in time. (b) A core with a graded refractive index exhibits less temporal spreading because the reflection process is more gradual. (c) The least spreading occurs in a single-mode fiber. The small core diameter of this fiber provides light with only one mode of travel so that the only spreading occurs because of ordinary dispersion.

To reduce stretching problems, the core of a better-performance optical fiber has a graded refractive index. The core glass is specially treated so that its index of refraction decreases smoothly away from its center toward the cladding. Instead of bouncing abruptly when it reaches the boundary between core and cladding, light in this graded-index environment turns smoothly back toward the core (Fig. 14.2.6b). The path differences between the modes in a graded-index multimode fiber aren't so different, and a short pulse of light isn't stretched very much in a fiber of moderate length.

In very long fibers, however, even small path differences add up, and short pulses become blurred in multimode fibers. Therefore, the highest-performance optical conduits are single-mode fibers. These fibers have very narrow graded-index cores that permit light to travel only in one mode—effectively right down the center of the fiber (Fig. 14.2.6c). The core is typically only 9 μm in diameter. A pulse of light entering this narrow core broadens very little in time during its passage.

What little broadening occurs in a single-mode fiber isn't caused by the light taking different paths; it's caused by ordinary dispersion in the glass. To carry information, the light wave must change with time and thus must have a range of frequencies and wavelengths. As usual, the shorter wavelengths travel slower than the longer wavelengths and the pulses get stretched out in time. To minimize these dispersion effects, the highest-speed optical fibers operate at wavelengths that minimize dispersion. They also operate at wavelengths that minimize the loss of light through absorption in the glass. These two wavelengths coincide at 1550 nm in dispersion-shifted fibers, so this infrared wavelength is commonly used in long-haul optical communication.

Check Your Understanding #5: If It Looks like a Mirror . . .

When you look into an aquarium from the front, the sides of the aquarium often look like mirrors. Those sides are actually clear glass, so why do they reflect light so well?

Answer: The light you see is experiencing total internal reflection as it tries to leave the glass at a shallow angle. You see perfect reflections of the fish inside the aquarium.

Why: Light has trouble passing from water or glass into the air at a shallow angle. When it tries, it experiences total internal reflection and bounces off the surface without losing any intensity at all.

Optical Communication

A typical optical communication transmitter uses a 1550-nm laser diode to generate short pulses of light. These pulses carry information from the transmitter to a receiver somewhere far away. The transmitter produces its pulses by varying the current passing through the laser diode. Light emerging from the laser diode is focused into the exposed core of a single-mode optical fiber, and it follows the core all the way to the other end of the fiber. When the light emerges, it's gathered by a lens and focused onto the receiver's photodiode. Each pulse of light causes a pulse of current to flow through the photodiode, allowing the receiver to begin processing the information.

A laser diode and a single-mode optical fiber can send billions of bits of data per second for 50 km or 100 km without significant errors. At longer distances, the gradual absorption of light in the glass makes it difficult to receive the information reliably. The easiest solution to this problem is to receive the data before the light becomes too weak and then to retransmit it with another laser diode.

Instead of interrupting the optical transmission with a receiver and retransmitter, some long-haul communication systems employ erbium-doped fiber amplifiers (EDFAs). An EDFA is a piece of optical fiber that has about 0.01% erbium ions in its glass core. When the EDFA is exposed to 980-nm or 1480-nm light, it becomes a laser amplifier for 1550-nm light. As the weakened pulses of light from a long fiber pass through the fiber amplifier, the amplifier duplicates photons and brightens the pulses. These amplified pulses then continue

through ordinary fiber before being amplified again. Undersea optical cables often splice fiber amplifiers into the fibers every 50 km or so. These amplifiers allow light to travel thousands of kilometers through a continuous path, from one side of an ocean to the other.

To get the most out of a single optical fiber, many communication systems use several laser diodes operating at somewhat different wavelength ranges around 1550 nm. Light from these diodes is merged together and focused into the fiber. When the light emerges at the far end of the fiber, its different wavelength ranges are split apart and directed onto individual receivers. The different wavelength ranges are like different channels, so that this wavelength-division multiplexing allows one fiber to carry far more information than it could with light from a single laser. Remarkably enough, an EDFA can amplify all of these different channels at once because erbium ions can copy a wide range of wavelengths.

Check Your Understanding #6: Making Something Out of Nothing

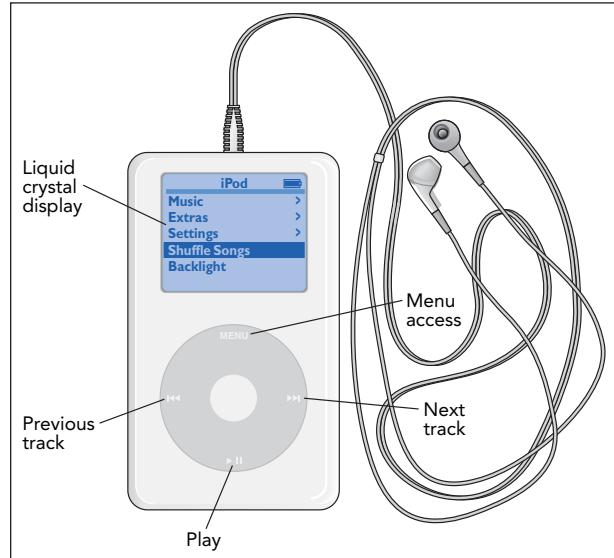
If light weakens gradually as it passes through a long optical fiber, why can't you simply wait until the very end of the fiber to amplify what little light remains and then send the amplified light into a receiver?

Answer: As the light is absorbed, the number of photons in a pulse gradually diminishes. When that number gets to zero, there's nothing left to amplify.

Why: Each pulse of light in the fiber contains a limited number of photons. The glass absorbs many of these photons during their long passage through a fiber, so they must be amplified before their numbers become so small that there's a chance that none of them makes it at all.

SECTION 14.3

Audio Players



Audio players have revolutionized portable music systems. Everywhere you look, people are sporting earpieces and listening to their favorite tunes with these little electronic marvels or their smartphone siblings. Part computer and part stereo system, an audio player is a spectacular synthesis of some of the highest forms of modern electronic technology. Because they contain such a broad range of electronic components, audio players offer an excellent introduction to much of modern electronics.

To understand how audio players work, we need to look at how sound can be represented electronically and at how those electronic representations can be stored, retrieved, and ultimately used to re-create the sound itself. That exploration will take us all the way from the digital world of computers to the analog world of amplifiers and headphones.

It will also expose us to that workhorse of modern electronics—the transistor. Early audio electronic devices were built with vacuum tubes, which were relatively bulky components that wasted power and aged quickly. Transistors have made audio electronics much more practical. They have also made computers so small and inexpensive that every audio player can have its own computer.

Questions to Think About: What does it mean to store songs? How can a computer use numbers to represent music? Why does an audio player need electric power to operate? Why does an audio player become warm as it operates? Does the volume of the sound affect the battery life? What does an audio amplifier's power rating mean? How are electric power and sound volume related? How do the treble and bass controls of an audio player affect the sound?

Experiments to Do: Find an audio player or a smartphone. Turn it on, and notice that it takes a short time to “wake up.” You’re observing the boot process of a computer. But if all of the computer’s information is already inside it, why does it need to do so much work as it starts? As we’ll soon see, the device’s computer uses different types of memory, some of

which are wiped clean when the audio player is turned off. With that tidbit in mind, try to imagine what is happening during the boot process.

Play some music and experiment with the sound controls. The volume control determines how much power reaches the headphones. Recall that a flashlight with a poor connection produces dim light. If you create a poor connection between

the player and the earpiece, what happens to the volume? Where did the lost power go?

Now experiment with the bass and treble controls. If the player has any other audio effects, try them. Are you still hearing sound as it was originally performed? How do the player's tone settings influence the sound reproduction? Is the most perfect imitation of the original sound always what you want to hear?

Transistors

The story of audio players begins with the story of **transistors**. Invented in 1948 by three American physicists, William Shockley (1910–1989), John Bardeen (1908–1991), and Walter Brattain (1902–1987), transistors are key elements in nearly all modern electronic equipment. Like the diodes we examined in Section 13.3, transistors are built from doped semiconductors—semiconductors such as silicon to which chemical impurities have been added. However, unlike diodes, which operate in only a single circuit, transistors allow the current in one circuit to control the current in another circuit.

There are many types of transistors, but the simplest and most important type is the *field-effect transistor*. Actually, even here there are several varieties, so we'll focus on the

one that's most widely used in audio players, cell phones, video equipment, and computers: the *n-channel metal-oxide-semiconductor field-effect transistor* (or n-channel **MOSFET**). Despite its complicated name, the n-channel MOSFET is a relatively simple device, consisting principally of three semiconductor layers and a nearby metal or metal-like surface (Fig. 14.3.1). The three layers are called the *drain*, the *channel*, and the *source*, and the metal surface is called the *gate*.

The drain and the source consist of a strongly doped n-type semiconductor (many conduction-level electrons) and the channel between them consists of a lightly doped p-type semiconductor (a few empty valence levels) (Fig. 14.3.2a). When those three layers touch, they form two back-to-back p-n junctions, and conduction-level electrons from the drain and source migrate into the channel to fill its few empty valence levels (Fig. 14.3.2b). The completed transistor is thus left with a vast depletion region (a region devoid of empty valence levels or occupied conduction levels) extending all the way from its drain to its source. With nothing to convey charge through its channel, the transistor can't conduct current between its drain and source. The MOSFET is effectively Off.

However, if more electrons could be coaxed into the channel somehow, those electrons would have to go into the channel's conduction levels and the channel would then behave like an n-type semiconductor. With an n-type channel sandwiched between an n-type drain and an n-type source, the p-n junctions would vanish and so would the depletion region. The three layers would

become, in effect, a single piece of n-type semiconductor, and the transistor would then be able to conduct current between its drain and source. The MOSFET would be effectively On.

Drawing extra electrons into the channel is the task of the metal-like gate. Separated from the channel by an incredibly thin insulating layer, the gate controls the channel's ability to carry current. When a tiny positive charge is placed on the gate through a wire, it attracts extra electrons into the channel's conduction levels from the source, drain, and their wires and the transistor begins to conduct current (Fig. 14.3.2c). The more positive charge there is on the gate, the more extra electrons are drawn into the channel and the more current can flow through the transistor. In effect, the transistor behaves like an adjustable resistor with a resistance that decreases as the positive charge on its gate increases.

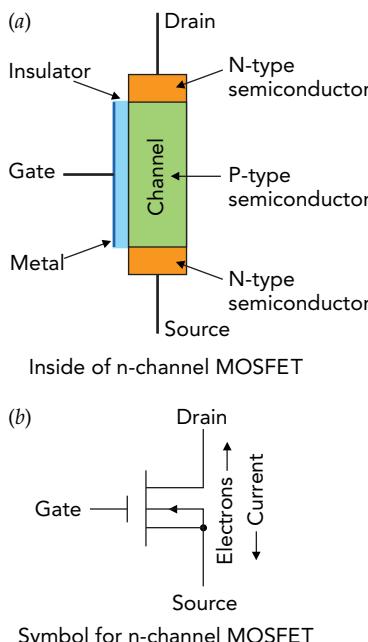


Fig. 14.3.1 (a) In an n-channel MOSFET, the channel is normally a depletion region that cannot carry current between the source and drain. But when positive charge on the gate attracts electrons into the channel, however, the channel becomes an n-type semiconductor and allows current to flow. (b) The symbol representing an n-channel MOSFET in a schematic diagram.

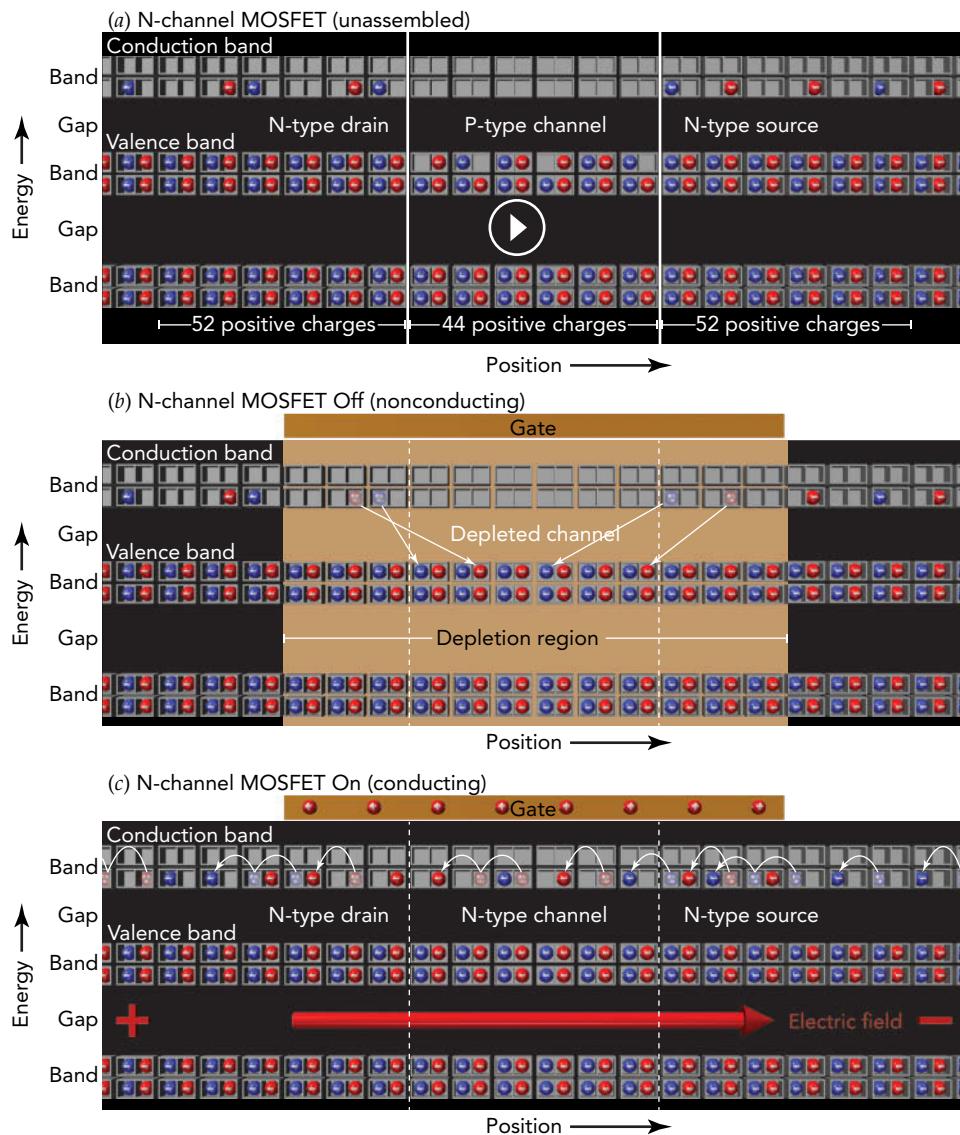


Fig. 14.3.2 (a) An n-channel MOSFET is formed from three pieces of semiconductor: a n-type drain, p-type channel, and n-type source. (b) When those pieces touch, conduction-level electrons from the n-type source and drain fill the empty valence levels of the p-type channel and form a vast insulating depletion region. The depleted channel is nonconducting and the MOSFET is Off. (c) When positive charge is placed on the nearby gate, however, it attracts extra electrons into the channel's conduction levels and the negatively charged channel then behaves as n-type semiconductor. The entire structure is conducting and the MOSFET is On.

We can now understand the n-channel MOSFET's name. *N-channel* refers to the channel's n-type behavior when its gate is positively charged and the transistor can carry current. Although the channel is chemically p-type (Fig. 14.3.2a), it becomes electrically n-type when extra electrons are drawn into it and it acquires a negative net charge (Fig. 14.3.2c). *Metal-oxide-semiconductor* indicates that the metal or metal-like gate is separated from the semiconductor channel by a thin insulating layer of oxide. This insulator can be as little as 1.2 nm thick and agonizingly easy to puncture, so modern electronic devices are heavily protected against the damaging effects of static electricity. *Field-effect transistor* indicates that the electric field from charge on the gate is what draws electrons into the channel and controls the current flow through the transistor.

Check Your Understanding #1: Power and Control

Widening the channel of an n-channel MOSFET allows it to handle more current between its source and drain. However, the enlarged transistor needs more positive charge on its gate to control that current. Explain.

Answer: The larger transistor also has a larger gate. With more surface over which to spread its charge, the gate needs more positive charge to draw conduction-level electrons into the channel.

Why: MOSFETs range in size from remarkably small (less than $0.01 \mu\text{m}^2$) to relatively large (several square millimeters). The smallest ones are used in computer chips, where millions of MOSFETs are created on a single wafer of silicon only a centimeter square. A tiny charge on the gate of one of these MOSFETs will allow it to conduct current. The largest MOSFETs are used in power-control devices such as amplifiers and power supplies. These transistors have large gates, and much more charge is needed to allow one of them to conduct current.

Storing Digital Sound Information

An audio player is half computer and half stereo system. It stores and manipulates sound information in digital form like a computer but then amplifies that information for the headphones in analog form like a stereo system. We'll mirror that sequence when examining the player's electronics: we'll start with its digital memory and processing systems and finish with its audio amplifier.

Inside the digital portion of the audio player, air pressure measurements and other numbers are represented in binary form. How large or precise those numbers are determines how many binary digits are needed to represent them. Each binary digit is called a **bit**, and using more bits allows you to represent larger or more precise numbers. In general, the more detailed the information, the more bits are needed to represent it.

Eight bits can be used to represent any number from 0 (which is **00000000**) to 255 (which is **11111111**). Since there are fewer than 256 of many common objects, these objects can be identified by groups of 8 bits. For example, the symbols used in ordinary text have been assigned numbers between 0 and 255, with 65 denoting the letter A. Since the 8 bits **01000001** represent 65, they also specify an A. Groups of 8 bits are so common and useful that they are called **bytes**.

Although it's possible to store sound information using 1 byte per air pressure measurement, a byte usually doesn't provide enough precision for quality sound reproduction. More often, digital audio is saved using 2 bytes per pressure measurement. These pressure measurements are made tens of thousands of times per second, usually from several microphones simultaneously to provide stereo or surround sound. Even when sophisticated data compression techniques are used to eliminate redundant or inconsequential information, a great many bits are still needed to represent an album. Thus an audio player needs a lot of memory.

There are several ways in which the audio player, like any computer, stores a bit. In its main working memory (often called random access memory, or RAM), each bit is a tiny capacitor that uses the presence or absence of separated electric charge to denote a **1** or a **0**. The player stores a bit by producing or removing separated charge and recalls the bit by checking for that charge.

Each capacitor is built right at the end of its own n-channel MOSFET. That MOSFET controls the flow of charge to or from the capacitor. To store or recall a bit, the audio player places positive charge on the gate of the MOSFET so that the MOSFET becomes electrically conducting. The memory system can then transfer charge to or from the bit's capacitor.

Storing the bit is relatively easy; the player simply sends the appropriate charge through the MOSFET to the capacitor. Recalling the bit is harder because the charge on the capacitor is extremely small. Sensitive amplifiers in the memory system detect any charge flowing through the MOSFET from the capacitor and report what they find to the audio player. Since this reading process removes charge from the capacitor, the memory system must immediately store the bit again.

Unfortunately, these tiny capacitors can't hold separated charge forever because it leaks out to their surroundings. Memory that uses charged capacitors to store bits is called dynamic memory and must be refreshed (read and restored) hundreds of times each second to ensure that a 1 doesn't accidentally switch to a 0 or vice versa.

Dynamic memory is also volatile—its contents are lost when the audio player turns off. To conserve its batteries, the player keeps its music information in nonvolatile memory, memory that doesn't need power to retain its information. New possibilities for nonvolatile memory appear almost every year, but at present the three leading forms are flash, magnetic disk, and optical disc memories.

Flash memory resembles dynamic memory in that each bit is stored as the presence or absence of charge associated with a MOSFET. In flash memory, though, that charge resides on the MOSFET's floating gate—a second, unattached gate located in the insulating layer between the channel and the normal gate. Since this floating gate is surrounded by insulator, it can keep its charge for decades. As long as that charge is present, it will determine the MOSFET's conductivity and whether the bit is a 0 or a 1.

Reading bits from flash memory is easy, but storing them is a challenge. The same isolation that traps charge on the floating gate for years makes that charge difficult to change. To add or remove electrons from the floating gate, the memory system applies relatively large voltages to the MOSFET's source, drain, and normal gate and the resulting strong electric fields permit electrons to cross through the insulation separating the channel from the floating gate.

To add electrons to the floating gate, the electric fields are arranged so they accelerate channel electrons to such high speeds that those electrons simply burrow right through the insulating layer to the floating gate. To remove electrons from the floating gate, the electric fields are arranged so that the floating gate's electron standing waves are distorted into the insulator. When these distorted waves reach far enough into the insulator, electrons begin to leak through it into the channel via a process known as *quantum tunneling*. (We'll return to explore quantum tunneling in Chapter 15.)

Flash memory is fast to read but relatively slower to write. Moreover, the electron burrowing process causes cumulative damage to the insulating layer and limits the number of times flash memory can be written. An audio player uses a mixture of dynamic memory and flash memory; it does its computational work in dynamic memory but retains its long-term information in flash memory.

However, there's another memory concept that remains more cost-effective than flash memory for storing vast amounts of information—magnetic disk memory. Although flash memory has largely replaced magnetic disks in audio players, computers still depend extensively on magnetic disks.

Just as the magnetic strip on a credit card (Fig. 11.1.6) can store information in the locations of its magnetic poles, the surface of a magnetic disk can store information in the orientations of its magnetic poles. Actual hard disks are smooth aluminum platters that have been coated with hi-tech hard magnetic materials. Using microscopic electromagnets to write magnetic poles and sophisticated semiconductor magnetic sensors to read them, modern hard disks can pack over 1 billion bits into a square millimeter of surface (over 80 gigabytes per square inch). Simply locating those microscopic bits on platters that rotate over 100 times per second is an electromechanical tour de force, yet these disks do it routinely even while you are moving your laptop around the room.

Check Your Understanding #2: Like Sending a Letter to Yourself

Every few thousandths of a second, a computer's memory system stops briefly to read and then rewrite every bit in its dynamic memory. What is going on?

Answer: The computer is making sure that the charge stored on each capacitor in the dynamic memory adequately represents that bit's contents.

Why: Since charge leaks quickly from the capacitors in dynamic memory, the contents of each bit must be refreshed many times a second. This refreshing process, reading each bit and storing it back into memory, slows the computer down slightly.

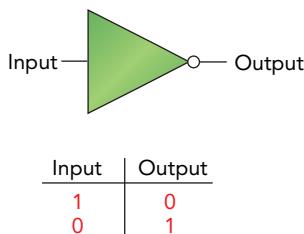


Fig. 14.3.3 An inverter, shown here symbolically, produces 1 output bit that is the inverse of its 1 input bit.

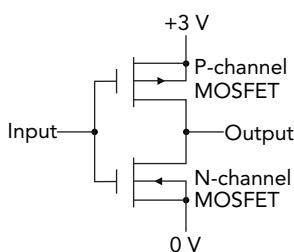


Fig. 14.3.4 When negative charge arrives at the input of a CMOS inverter, its p-channel MOSFET (top) permits positive charge to flow to the output. When positive charge arrives at the input, the n-channel MOSFET (bottom) sends negative charge to the output.

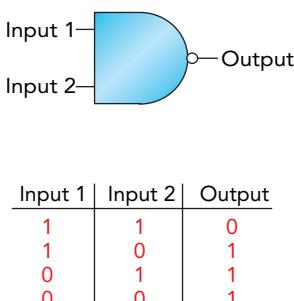


Fig. 14.3.5 The output bit of a Not-AND or NAND gate, shown here symbolically, is a 1 unless both input bits are 1s.

The Audio Player's Computer

We've seen how sound information can be represented and stored as bits, so now let's look at how an audio player's computer works with those bits. This digital processing is done by electronic devices that take groups of bits as their inputs and produce new groups of bits as their outputs. Since their output bits are related to their input bits by the rules of logic, these electronic devices are called logic elements.

The simplest logic element is the inverter, which has only one input bit and one output bit. Its output is the inverse of its input (Fig. 14.3.3). If an inverter's input bit is a 1, then its output bit is a 0, and vice versa. Inverters are used to reverse an action—turning a light on rather than off or starting a song rather than stopping it. Inverters are also used as parts of more complicated logic elements.

But inverters aren't just abstract logic elements; they're real electronic devices. They act on electrical inputs and create electrical outputs. In an audio player's computer, inverters and other logic elements represent input and output bits with electric charge. Positive charge represents a 1, and negative charge represents a 0. Thus when positive charge arrives at the input of an inverter, the inverter releases negative charge from its output.

Inverters and other logic elements are usually constructed from both n-channel and p-channel MOSFETs. We've already seen that n-channel MOSFETs conduct current only when their gates are positively charged. P-channel MOSFETs do just the reverse, conducting current only when their gates are negatively charged. The drain and source of a p-channel MOSFET are made from p-type semiconductor, and the channel is made from n-type semiconductor. Since n-channel and p-channel MOSFETs are exact complements to one another, logic elements built from them are called complementary MOSFET or CMOS elements. An audio player's computer is built almost entirely from CMOS elements.

A CMOS inverter consists of one n-channel MOSFET and one p-channel MOSFET (Fig. 14.3.4). The n-channel MOSFET is connected to the negative terminal of the computer's power supply and controls the flow of negative charge to the inverter's output. The p-channel MOSFET is connected to the power supply's positive terminal and controls the flow of positive charge to the output. When negative charge arrives at the inverter's input and moves onto the gates of the MOSFETs, only the p-channel MOSFET conducts current and the output becomes positively charged. When positive charge arrives at the input, only the n-channel MOSFET conducts current and the output becomes negatively charged.

A computer, however, needs logic elements that are more complicated than inverters. One such element is the Not-AND or NAND gate. This logical element has 2 input bits and 1 output bit, and its output bit is 1 unless both input bits are 1s (Fig. 14.3.5). It's called a Not-AND gate because it's the inverse of an AND gate. An AND gate produces a 0 output unless both input bits are 1s. Simple memoryless logic elements are often called gates.

A CMOS NAND gate uses two n-channel MOSFETs and two p-channel MOSFETs (Fig. 14.3.6). The two n-channel MOSFETs are arranged in series—one after the next—so that current passing through one must also pass through the other. If either transistor has negative charge on its gate, no current can flow through the series. Components arranged in series all carry the same current, but they may experience different voltage drops.

The two p-channel MOSFETs are arranged in parallel—one beside the other—so that current can flow through either one of them to the output. If either transistor has negative charge on its gate, current can flow from one side of the pair to the other. Components arranged in parallel share the current they receive through one wire and deliver it together to the second wire. Although parallel components may share the current unevenly among themselves, they all experience the same voltage drop.

If negative charge arrives at either input of the CMOS NAND gate, the series of n-channel MOSFETs will be nonconducting and one of the p-channel MOSFETs will deliver positive charge to the output. However, if positive charge arrives at both inputs, both p-channel MOSFETs will be nonconducting and the series of n-channel MOSFETs will deliver negative charge to the output. Thus the CMOS NAND gate has the correct logic behavior.

These two logic elements, inverters and NAND gates, can be combined to produce any conceivable logic element. For example, they can be used to build an adder, a device that sums the numbers represented by two groups of input bits and produces a group of output bits representing that sum. These adders can themselves be used to build multipliers, and multipliers can be built into still more complicated devices. In this fashion, the simplest logic elements can be used to construct an entire computer.

Actually, a computer isn't built exclusively from NAND gates and inverters. To improve its speed and reduce its size, it uses a few other basic logic elements as well. Like the CMOS NAND gate and inverter, these elements are constructed directly from n-channel and p-channel MOSFETs.

All these logic elements are wired together in an intricate pattern to create a complete computer (Fig. 14.3.7). In an audio player, this computer retrieves and organizes music information and prepares it for the playback electronics, which are not digital. The computer's last act is to deliver the digital music information, the air pressure measurements, to a *digital-to-analog converter*, or DAC. This electronic device is the interface between the two representations of information: digital and analog. The music information leaves the DAC as a voltage that's proportional to air pressure. This voltage is the input for the audio player's main analog component, its audio amplifier. Actually, the player has two complete analog audio systems so that it can produce stereo sound. But since those systems are identical, we'll focus on only one of them.

Check Your Understanding #3: Getting It Together

How could you use inverters and NAND gates to create an AND gate, a logic element that has two inputs and one output, with the output being 1 only if both of the inputs are also 1s?

Answer: You could connect an inverter to a NAND gate so that the output signal of the NAND gate is the input signal of the inverter. When both of the NAND gate's input signals are 1s, it will produce an output of 0. This 0 will arrive at the inverter, which will invert it and yield an output of 1.

Why: Connecting logic elements together one after the next is the standard method for producing more complicated logic elements. In this case, two elements connected together produce a third.

The Audio Player's Audio Amplifier

The fluctuating voltage provided by the audio player's DAC is often called an *audio signal* because it represents audio information. Many analog or digital representations of information are called **signals**, including video signals, data signals, and even turn signals. But although the player's audio signal contains all the information needed to reproduce the original sound, in convenient analog format, it doesn't have the power that the headphones need to produce that sound at a reasonable volume. Something must first enlarge the audio signal; it needs to be *amplified*.

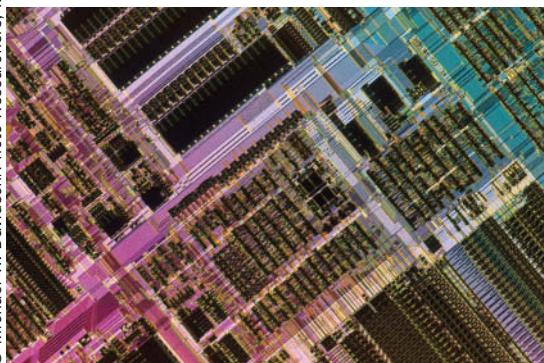


Fig. 14.3.7 This microscope photograph shows an integrated circuit microprocessor—approximately a computer on a chip. Aluminum strips connect millions of MOSFETs and other components that have been formed by photographic techniques on the surface of a thin wafer of silicon.

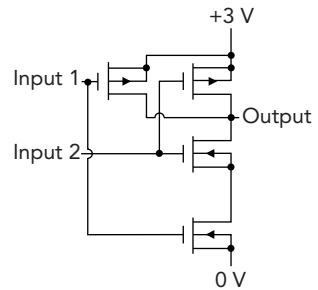


Fig. 14.3.6 A CMOS NAND gate has two input bits. When negative charge arrives through either input, the chain of n-channel MOSFETs (*bottom*) stops conducting current and one of the two p-channel MOSFETs (*top*) permits positive charge to reach the output. Only if both inputs are positively charged will negative charge reach the output.

Fig. 14.3.8 A simple audio amplifier can be built with one n-channel MOSFET, two resistors, and two capacitors. A 9-V battery powers the device.

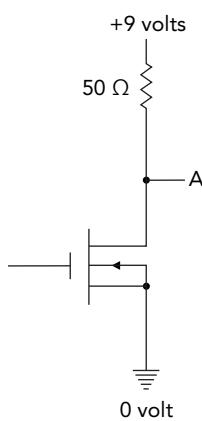
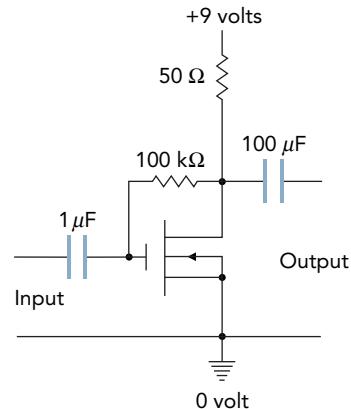


Fig. 14.3.9 The voltage at A depends on the resistance of the MOSFET. The lined triangle at the bottom signifies connection to ground (often Earth itself).

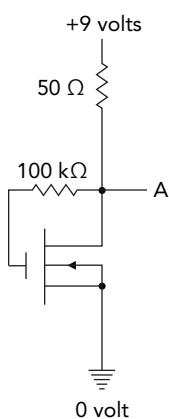


Fig. 14.3.10 The 100-k Ω resistor transfers positive charge to the gate until the voltage at A drops to about 5 V.

Devices that enlarge various characteristics of signals are called **amplifiers**. An audio amplifier is an amplifier that's designed to boost signals in the frequency range that we hear or feel (20 to 20,000 Hz). It has two separate circuits—an input circuit and an output circuit—and it uses the small current passing through its input circuit to control a much larger current passing through its output circuit. In this manner, the amplifier provides more power to its output circuit than it receives from its input circuit. Since energy is conserved, an amplifier needs a separate power source to provide its amplification—in this case, the 9-V battery.

Figure 14.3.8 shows the schematic diagram for a simple audio amplifier, built from the components we've just studied. This amplifier has only five components: an n-channel MOSFET, two resistors, and two capacitors. It draws power from a 9-V battery (or an equivalent power adapter) and amplifies a tiny alternating current in its input circuit into a large alternating current in its output circuit.

To understand how this amplifier works, let's first remove everything but the MOSFET and the 50- Ω resistor (Fig. 14.3.9). These two components are in series, so any current that passes through one must also pass through the other. When the MOSFET doesn't conduct current, no current flows through the 50- Ω resistor and it experiences no voltage drop. So the voltage at A is 9 V. However, if the transistor does conduct current, a voltage drop will appear through the 50- Ω resistor and the voltage at A will decrease.

The transistor will conduct current only if positive charge is put on its gate. That can be done by connecting the gate to A with a 100-k Ω resistor (Fig. 14.3.10). Since A is at 9 V, it is positively charged and pushes charge toward anything at lower voltage. Current flows slowly through the resistor from A to the gate. However, as positive charge accumulates on the gate, the transistor begins to conduct current and the voltage at A drops. When the voltage at A reaches the voltage on the gate, current stops flowing through the resistor.

The amplifier is then in a stable equilibrium; A has a voltage of approximately 5 V and the transistor's gate has a modest amount of charge on it. The 100-k Ω resistor provides the transistor with **feedback**; that is, it provides the transistor with information about the present situation at A that the transistor can use to correct or improve that situation. Although the feedback is slowed by the resistor's large electrical resistance, it perpetually acts to return the voltage at A to its equilibrium value. If the transistor conducts too little current, charge flows onto its gate and makes it conduct more. If the transistor conducts too much current, charge flows off its gate and makes it conduct less.

The amplifier is now exquisitely sensitive to small changes in the charge on the transistor's gate. If you add just a tiny bit more positive charge to the gate, down goes the voltage at A. If you remove just a tiny bit of positive charge from the gate, up goes the voltage at A. Although current in the feedback resistor tries to undo these changes, it acts too slowly to oppose short-timescale variations. The amplifier's input signal successfully adds or subtracts positive charge from the gate, and the amplifier's output signal emerges from A.

The amplifier has two input wires. An analog audio signal's current flows into the amplifier through one wire and returns through the other. However, the audio signal is not connected directly to the gate. Instead, it's connected to the gate through a capacitor (Fig. 14.3.11). In addition to storing charge and energy, a capacitor can transfer current between two wires that have different voltages. Such voltage flexibility is important in battery-powered audio amplifiers that must do their amplifying exclusively with positive voltages. Aided by input and output capacitors, our audio amplifier can have an average operating voltage of about +5 V while it also has average input and output voltages of 0 V.

To see how a capacitor passes along current, let's watch input current flow rightward into our amplifier's input capacitor. As that current's positive charge accumulates on the capacitor's left plate, it attracts negative charge onto the capacitor's right plate and away from the gate. The capacitor remains electrically neutral, but the gate becomes more positively charged. Overall, the capacitor has conveyed input current to the gate even though no charge has actually passed through its insulating layer and its two plates remain at different voltages.

With the help of the input capacitor, the amplifier's fluctuating input current produces a fluctuating charge on the transistor's gate, and the voltage at A fluctuates as a result. Even a tiny fluctuating current on the input wires creates a large fluctuating voltage at A.

This fluctuating voltage is responsible for sending a fluctuating current through the headphones. Although headphones are not truly ohmic devices, they respond to fluctuating voltages by carrying fluctuating currents. By applying a fluctuating voltage drop across the headphone's two wires, the amplifier can cause it to carry a fluctuating current and produce corresponding pressure fluctuations and sound.

However, the voltage at A averages about 5 V, while the headphones expect an average voltage drop of 0 V. To convey the fluctuating voltages and currents from A to the headphones, while eliminating their large voltage difference, our amplifier connects them via an output capacitor (Fig. 14.3.12). As before, the fluctuations in current and voltage on the output capacitor's left plate are mirrored by current and voltage fluctuations on its right plate. Even though the amplifier operates at a high average voltage, the output signal for the headphones has an average voltage of 0 V.

Tiny fluctuating currents in our amplifier's input circuit produce large fluctuating currents in its output circuit. This amplifier works remarkably well, given its simplicity. If you connect a microphone to the input wires and headphones to the output wires, the headphones will do a surprisingly good job of reproducing the sound in the microphone.

However, our simple amplifier isn't perfect. It distorts the sound somewhat, and it doesn't handle all frequencies or amplitudes of sound equally. It also wastes a large amount of electric power heating the $50\text{-}\Omega$ resistor. The amplifiers in audio players carefully correct for these problems. Most use feedback to make sure that their output signals are essentially perfect replicas of their input signals, only larger. They sense their own shortcomings and correct for them.

Perfect replication of the input signal isn't always desirable. Sometimes you want to boost the volume for part of the sound. The treble and bass controls on an audio player allow you to selectively change the loudnesses for the high- and low-frequency portions of the sound, respectively.

An amplifier is typically rated according to the peak power it can supply to the headphones (or speakers) and its average power should never reach that value. Yet the amplifier can reach its peak power during a passage that's not particularly loud. That's because sound waves often interfere with one another (see Section 9.3 to review wave interference), and when their crests and troughs coincide at the microphone, constructive interference can briefly produce enormous pressure fluctuations.

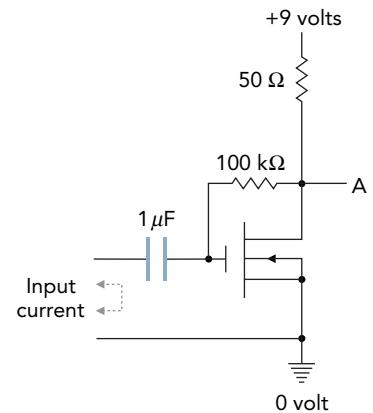


Fig. 14.3.11 Because current flowing back and forth through the two input wires affects the charge on the transistor's gate, it also affects the voltage at A.

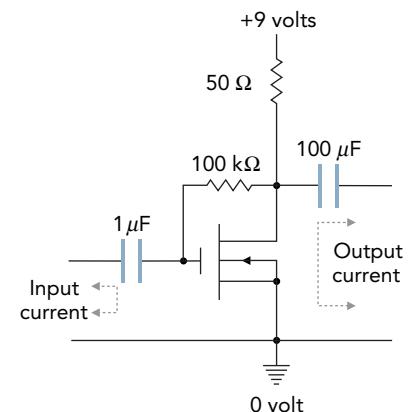


Fig. 14.3.12 The amplifier causes currents to flow back and forth through the two output wires. The alternating current in the output wires is a good replica of the alternating current in the input wires, only larger.

To reproduce those overlapping waves properly, the audio amplifier must be able to provide several times its average power, though only for a moment. If the amplifier can't deliver that much power, the audio signal it sends to the headphones or speakers will be distorted and the sound will be unpleasant. That's why audiophiles often use powerful amplifiers even when they're playing quiet music. Most modern amplifiers are designed cleverly so that they consume only a little more electric power than they actually need to produce their outputs. They consume little power during quiet passages but can still provide great power during the loud moments.

As for the headphones themselves, they generally use electromagnetic effects to move a surface back and forth in sync with the amplifier's current fluctuations. In most cases, the amplifier's current is sent through a coil of wire that's immersed in a strong magnetic field and attached to a movable surface. This current experiences a Lorentz force due to the magnetic field, and that force drives the current, coil, and surface back and forth as the current fluctuates. The moving surface alternately compresses and rarefies the air, thereby reproducing the original sound. The electric power in the circuit becomes sound power for your ears.

Check Your Understanding #4: Sound Control

When you connect a microphone to the input of a MOSFET-based amplifier, the microphone sends current back and forth through the input wires. As a result, charge moves onto or off what critical control element in the amplifier?

Answer: The critical control element is the gate of a MOSFET.

Why: In all likelihood, the input current adds or removes charges from the gate of a MOSFET in the amplifier's first stage of amplification.

Epilogue for Chapter 14

In this chapter we looked at several common objects and studied the ways in which they use optics. In Cameras, we saw how a converging lens can form a real image of an object and how that real image can be used to record a scene on a piece of film or an image sensor chip. We learned about the roles of focal length and f-number in determining the size and brightness of the real image and explored some of the complications that must be considered when designing lenses.

In Optical Recording and Communication, we studied the ways in which sound and other information can be represented in digital form and stored as structural features on optical discs. We looked at how laser light is affected by various optical devices and looked at what happens to it when it's focused to a small spot. We learned about total internal reflection and how this effect makes it possible to send light hundreds or thousands of kilometers through a fiber of ultraclear glass.

In Audio Players, we looked at the electronic techniques used to represent sound and other information in digital and analog forms. We also examined the most important modern electronic component—the transistor, a semiconductor device that allows the current in one circuit to control the current in another circuit. We saw how an audio player uses transistors and other electronic components in both its digital computer and its analog amplifier.

Explanation: Magnifying Glass Camera

The magnifying glass forms a real image of the window on the wall. All the light rays passing through the lens from one point on the window converge together to one point on the wall. Since each point on the window illuminates one point on the wall, you see a pattern

of light on the wall that looks just like the window itself. The real image on the wall, however, is flipped upside down and its sides are reversed.

Because the window is closer to the lens than any of the objects outside the window, the real image of the window itself forms farthest from the lens. When you hold the magnifying glass so that the window appears sharp on the wall, the objects outside the window are blurry. To bring those more distant objects into sharp focus, you must move the lens toward the wall. Very distant objects, such as a mountain or the moon, will form sharp images on the wall when the distance between the magnifying glass and wall is equal to the focal length of the lens.

Chapter Summary and Important Laws and Equations

How Cameras Work: A camera lens projects a real image of the scene in front of the camera onto the image sensor inside the camera. This real image is formed when the lens bends all the light reaching it from one part of the scene onto one part of the sensor. For this imaging process to work well, the distance between the lens and the image sensor must be adjusted so that the light converges together (focuses) just as it reaches the sensor. If the sensor is too close to or too far from the lens, the image is blurry. The depth of focus depends on the effective diameter of the lens—its aperture. The smaller the lens's aperture, the less critical the focus but the less light the lens gathers. A long-focal-length lens brings the light to a focus far behind it, forming a relatively large but dim real image on the image sensor. To brighten this image, the long-focal-length lens must have a large aperture so that it gathers lots of light. The f-number, the ratio of a lens's focal length to its aperture, characterizes the brightness of the lens's image.

How Optical Recording and Communication Work: A CD, DVD, or Blu-ray represents sound or video information in digital form as a pattern of pits on a thin reflective layer inside the disc. The player uses a beam of laser light to read that pattern. As the disc rotates, the pits pass through the focused laser beam and the amount of reflected light fluctuates up and down. The player monitors the reflected light and uses it both to recreate the music and to keep its optical system properly aligned. This continual realignment allows the player to follow the spiral track of pits as the disc turns and to keep the laser beam tightly focused on the reflective layer. The optical system, which includes a laser diode, several photodiodes, and a variety of lenses and other optical elements, is so well designed and executed that the spot it creates on the reflective layer is limited in how small it can be only by diffraction effects due to the wave nature of light.

Laser light can also carry information long distances through an optical fiber. Made from a core of extremely transparent glass that is clad in a second glass, this fiber confines light via total internal reflection. As light attempts to leave the core at a shallow angle, it is perfectly reflected from the boundary with the cladding and continues through the fiber for great distances.

How Audio Players Work: An audio player combines a computer and an audio amplifier into a single unit. The player's computer stores and retrieves the sound information in digital form, with each air pressure measurement represented by a collection of binary bits. These bits can take only values of 0 or 1. For long-term storage, the player places its digital sound information in flash memory and/or magnetic disk memory. For temporary storage while playing a song or working on its song collection, it uses dynamic memory.

After the player's computer has manipulated that information digitally, using logic elements built primarily from MOSFETs, it passes the digital information to a digital-to-analog converter and that information then takes analog form. This analog signal

represents sound as a current that's proportional to the sound's shift in air pressure away from atmospheric pressure.

The analog sound signal enters the input circuit of the player's audio amplifier, which makes use of MOSFETs. This amplifier uses power obtained from the battery to produce an output signal that represents the same sound but with increased voltage and current. This amplified output signal has enough power to produce loud sound in the headphones.

-
- 1.** Lens equation: One divided by the focal length of a lens is equal to the sum of one divided by the object distance and one divided by the image distance, or

$$\frac{1}{\text{focal length}} = \frac{1}{\text{object distance}} + \frac{1}{\text{image distance}}. \quad (14.1.1)$$

In recent years, scientists have been looking deeper into the atom to see how it's made, farther into space to see how the universe works, and more carefully into matter and motion to see how complicated things can be understood in terms of simple laws. Among the most important tools that these scientists have to work with are quantum theory and the theory of relativity. This chapter will look at some of the ways in which modern physics affects our lives.

ACTIVE LEARNING EXPERIMENTS**Radiation-Damaged Paper**

One path of modern physics involves the control and use of high-energy radiation. Although our access to most forms of high-energy radiation is restricted, there's one source that anyone can use—the sun. Because the sun's ultraviolet light is energetic enough to damage chemical bonds and rearrange molecules, it can provide us with a glimpse of the effects that occur with X-rays and beyond.

To see sun damage for yourself, expose some sheets of colored construction paper to direct sunlight for a few days. Cover the paper with some opaque objects such as coins and place it outdoors. Don't cover the paper with glass because glass absorbs enough ultraviolet light to slow the damage process. After a day or two, you should

find that the exposed portions of the paper have lightened; the sun's ultraviolet radiation has destroyed some of the dye molecules in the paper. If you find that nothing happens, the dye is evidently robust enough to tolerate ultraviolet light for a while.

Try the experiment again with different papers and different colors. Which papers fade fastest? Do they seem to have a half-life? How could you tell?

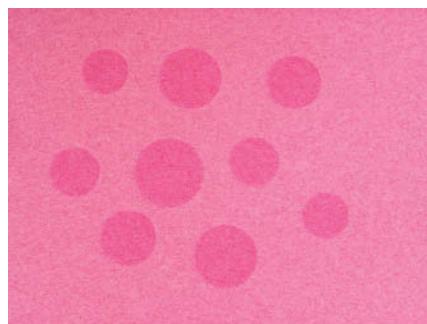
This same optical bleaching appears on items that are displayed in shop windows and on outdoor furniture. It was once the only method people had for whitening fabrics. The sun's ultraviolet light also damages your skin when you sit in the sun—a sunburn isn't thermal damage; it's radiation damage.

Chapter Itinerary

Fortunately ultraviolet light can't penetrate far into your body. In Nuclear Weapons, we look at more penetrating forms of radiation. We also explore the structures of atomic nuclei and see how taking them apart or joining them together can release enormous amounts of energy. In Nuclear Reactors, we study the approaches

used to extract nuclear energy and convert it into electric energy. We also look at the safety and health issues relating to nuclear energy. In Medical Imaging and Radiation, we examine high-energy radiation and see how it's used to help rather than hurt. We study the ways in which X-rays and gamma rays are produced

Courtesy Lou Bloomfield

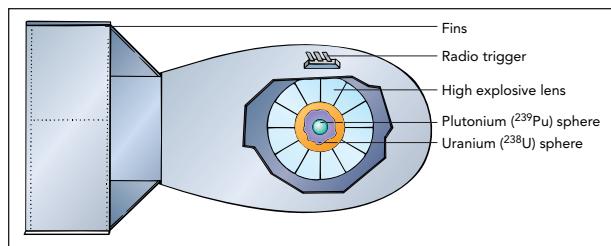


and how they interact with the atoms and molecules in a patient. We also look at the particle accelerators that produce high-energy particles for radiation therapy. Finally, we discuss the bases for computed tomography (CT) imaging and magnetic resonance

imaging (MRI), which make it possible to prepare detailed maps of patients' insides without ever touching their bodies. For additional preview information, see the Chapter Summary and Important Laws and Equations at the end of the chapter.

SECTION 15.1

Nuclear Weapons



The atomic bomb is one of the most remarkable and infamous inventions of the twentieth century. It followed close on the heels of various developments in the understanding of nature, developments that in many ways made the invention of nuclear weapons inevitable. By the late 1930s, scientists had discovered most of the principles behind nuclear energy and were well aware of how those principles might be applied. The onset of World War II prompted concern that Germany would choose to follow the military path of nuclear energy. Propelled by fear, curiosity, and temptation, the scientists, engineers, and politicians of that time brought nuclear weapons into existence. The world has lived in the shadow of these terrible devices ever since.

Questions to Think About: Where is nuclear energy stored in the atoms? From where did this nuclear energy come? How do

nuclear weapons release nuclear energy? Why are nuclear weapons so difficult to build? Why do we associate uranium and plutonium with nuclear weapons? How much uranium or plutonium does it take to build a bomb?

Experiments to Do: Since uranium and plutonium aren't sold in hardware stores, you won't be able to build your own bomb. However, you can get a feel for the way a chain reaction works by playing with a box of dominoes. If you stand the dominoes on end on a level table, each of them will have extra gravitational potential energy that it can release by tipping over. If you spread the dominoes widely about the table and then give the table a gentle shake, they'll tip over one by one.

However, if you pack the dominoes tightly together, so that one falling domino can knock over others, they'll no longer be independent. As you jiggle the table, nothing will happen until the first domino falls, but then many or even all of the dominoes will tip over in quick succession. You will have created a chain reaction, where a single event triggers an ever-increasing number of subsequent events. What characteristics of the dominoes and their arrangement determine whether or not such a chain reaction occurs? Can you envision a scenario in which a single tipping domino could trigger the release of an enormous amount of stored energy? Another chain reaction, this time in the decay of atomic nuclei, is what makes nuclear weapons possible.

Background

At the end of the nineteenth century, classical physics reigned supreme. Here *classical physics* means the rules of motion and gravitation identified by such people as Galileo, Newton, and Kepler, and the rules of electricity and magnetism developed by others, including Ampère, Coulomb, Faraday, and Maxwell. It was generally felt that most of physics was well understood. Physicists knew all the laws governing the behavior of objects in our universe, and all that was left to do was to apply those laws to more and more complicated examples. It was a time when physicists didn't know what they didn't know.

However, a few nagging problems remained—specific difficulties that couldn't be explained by the rules of classical physics. Among these were the spectrum of light emitted by a blackbody, the photoelectric effect in which electrons are ejected from metals by light, and the apparent absence of an ether or medium in which light traveled. At the beginning of the twentieth century, the whole of classical physics collapsed under the weight of these seemingly trivial difficulties, and a largely new understanding of the universe emerged. The major advances took 25 years, from 1901 to 1926, and the time since has largely been spent applying those new laws to more and more complicated examples.

The two main developments, both essential to the making of the atomic bomb, were the discoveries of quantum physics and relativity. Often these are called *quantum theory* and the *theory of relativity*. But while the word *theory* might imply that they're somehow on shaky ground, they're not theories in the sense of hypotheses waiting to be tested. In fact, they've been confirmed countless times since they were developed and have been shown to have enormous predictive power. Rather, they're theories in the sense of being carefully constructed and codified rules that model the behavior of the physical universe in which we live. Between them, these two theories made the discovery of nuclear forces and nuclear energy unavoidable. Finally, given people's love for gadgets and power, they also made nuclear weapons inevitable.

Check Your Understanding #1: In Theory, It Means That . . .

If something is referred to as a *theory*, how likely is it to be true?

Answer: That depends completely on the theory and the extent to which it has been compared to the real world.

Why: A great many theories have been formulated to explain the world around us. Each of these theories is an attempt at describing some particular behavior of our universe in terms of various rules or mechanisms. However, until a theory has been tested by comparing it carefully to the system it's trying to explain, you can't tell whether it's true or not. Some theories are eventually proven true, others false, and many remain uncertain. The theories of relativity and quantum physics have long since been proven true, although there is always the possibility that they may be only parts of a more complete theory.

The Nucleus and Radioactive Decay

Though the name *atomic* bomb has stuck for more than half a century, the more correct name would be *nuclear* bomb. The items that are responsible for the energy released by nuclear weapons are not atoms but tiny pieces of atoms—their nuclei (plural of *nucleus*). Before we can discuss nuclei, though, let's put them into context. Let's start by looking at atoms.

To get an idea of just how tiny atoms are, imagine magnifying a grain of table salt, which is 1 mm (0.04 in) on a side, until it is the size of the state of Colorado. That grain would then appear as an orderly arrangement of spherical particles, each about the size of a grapefruit (Fig. 15.1.1). These spherical particles would be single atoms, and there would be about 7.2 million of them along each edge of the grain.

Like most solids, table salt is a crystal and its atoms are bound to one another by their outermost components, their electrons. Electrons dominate the chemistry of atoms and molecules. Sodium is a reactive metal because of its electrons, and chlorine is a reactive gas because of its electrons. When mixed, these two chemicals react violently to form table salt and release a considerable amount of light and heat. This, then, is a true “atomic bomb.”

Obviously, something is missing here. If a crazy person could buy a kilogram or two of sodium and a tank of chlorine from a chemical company and destroy an entire city, rural living would be a whole lot more popular. Fortunately the energy released by chemical reactions is fairly limited. A kilogram of chemical atomic explosives just can't do that much damage. However, nuclear bombs tap an entirely different store of energy deep within the atoms.

Although all nuclear weaponry is often attributed to Einstein's famous equation, $E = mc^2$, that notion is vastly oversimplified. Nonetheless, this equation is quite significant. As we noted in Section 4.2, one of Einstein's discoveries at the beginning of the twentieth century was that matter and energy are in some respects equivalent. In certain circumstances, mass

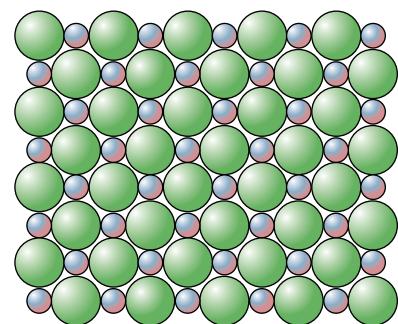


Table salt crystal
(Sodium chloride)

Chlorine negative ion
Diameter: 1.81×10^{-10} meter

Sodium positive ion
Diameter: 0.97×10^{-10} meter

Fig. 15.1.1 A salt crystal is an orderly array of positively charged sodium ions and negatively charged chlorine ions. These ions are held together by the attractive forces between oppositely charged ions. The ions are so small that there are about 7.2 million ions on the edge of a 1-mm-wide salt crystal.

can become energy or energy can become mass. This equivalence is part of the theory of relativity and has some interesting consequences. It implies that an object can reduce its mass by transferring energy to its surroundings. Thus, if you weigh an object before and after it undergoes some internal transformation, you can use any weight loss to determine how much energy was released from the object by that transformation.

Because of this equivalence, mass and changes in mass can be used to locate energy that's hidden within normal matter. This technique is important in nuclear physics, but it also applies to chemistry. When sodium and chlorine react to form table salt, their combined mass decreases by a tiny amount. What's missing is some chemical potential energy, which becomes light and heat and escapes from the mixture. In leaving, this chemical potential energy reduces the mass of the sodium and chlorine mixture by about 1 part in 10 billion. That tiny change in mass is too small to detect with present measuring devices, although scientists are working on techniques that will soon make it possible to measure mass changes due to forming chemical bonds.

Electrons are, however, by far the lightest part of an atom and thus have relatively little mass to release as energy. Most of an atom's mass is located in its **nucleus**. The nucleus is fantastically small—only a little more than 10^{-15} m in diameter. If you were to peer into one of the grapefruit-size sodium ions of our giant salt crystal, you would see a tiny particle at its center. There, just at the threshold of visibility, would be the ion's nucleus, only 1 μm (0.00004 in) in diameter. The remaining 99.999999999999% of the ion is occupied only by its 10 electrons in their orbitals.

The sodium nucleus contains 11 protons and 12 neutrons (Fig. 15.1.2). Each of these nuclear particles, or **nucleons**, has about 2000 times as much mass as an electron, so 99.975% of the sodium ion's mass is in this nucleus. Thus, while the electrons are certainly important to chemistry and matter as we know it, their contribution to the ion's mass is insignificant. The ion is mostly empty space, lightly filled with fluffy electrons and having a tiny nuclear lump at its center.

The nucleons that make up this nucleus experience two competing forces. The first of these forces is the familiar electrostatic repulsion between like electric charges. Because each proton in the nucleus has a single positive charge, they're constantly trying to push one another out of the nucleus. However, the second of these forces is attractive and holds the nucleus together. This new force is called the **nuclear force**, and at short distances it dominates the weaker electrostatic repulsion. However, the nuclear force attracts the nucleons toward one another only when they're touching. As soon as they're separated, they're on their own.

The competition between these two forces—the repulsive electrostatic force between like charges and the attractive nuclear force between nucleons—is analogous to what happens in a familiar toy (Figs. 15.1.3 and 15.1.4). This hopping toy has a suction cup attached to a spring, and the spring tries to separate the toy's top from its base while the suction cup tries to keep the two parts together. When the two parts are well separated, only the spring exerts a force. But when the two parts touch, the suction cup begins to act and holds the two parts together.

What makes the hopping toy exciting is that its suction cup leaks. Eventually, the suction cup lets go and allows the spring to toss the toy into the air. Suppose, however, that the suction cup didn't leak. Once pushed together, the pieces would never separate and the spring would retain its stored energy indefinitely. To get the suction cup to let go, you would have to pull it away from the base. Only then could the spring release its stored energy.

In effect, an energy barrier would be preventing the leak-free toy from hopping. Until you did a little work on it by pulling the suction cup off the base, it wouldn't be able to release its stored energy. Another example of a system that needs energy to release energy is a bottle of champagne, where you must push on the cork to help it out of the neck. After that initial investment of energy, a great deal of energy is released as gas inside the bottle blasts the cork across the room.

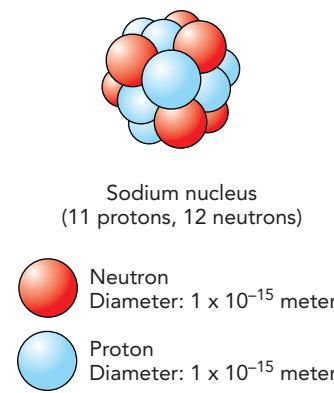


Fig. 15.1.2 At the center of a sodium ion is a tiny nucleus containing about 99.975% of the ion's mass. It consists of 11 positively charged protons and 12 uncharged neutrons. The protons repel one another at any distance, but the protons and neutrons are bound together by the highly attractive nuclear force as long as they touch one another.

Courtesy Lou Bloomfield



Fig. 15.1.3 This hopping toy stores energy as its spring is compressed and retains that energy while its suction cup grips its base. When the suction cup lets go, the toy leaps into the air.

The nucleus is in a similar situation. The attractive nuclear force prevents the nucleus from coming apart, despite the enormous amount of electrostatic potential energy it contains. The nuclear force creates an energy barrier that prevents the nucleons from separating. Unless something adds energy to the nucleus to help the nucleons break free of the nuclear force, the nucleus will remain together forever. At least, that's the prediction of classical physics.

Quantum physics, however, has an important influence on the behavior of the nucleus. One of the many peculiar effects of quantum physics is that you can never really tell exactly where an object is located, or at least not for long. That fuzziness is a manifestation of the **Heisenberg uncertainty principle**, which observes that some pairs of physical quantities, such as position and momentum or energy and time, are not entirely independent and cannot be determined simultaneously beyond a certain accuracy. This principle is a result of the partly wave and partly particle nature of objects in our universe (Section 13.2). Since waves are normally broad things that occupy a region of space rather than a single point, objects in our universe normally don't have exact locations.

The smaller an object's mass, the fuzzier it is and the more uncertain its location. Although the fuzzy nucleons in a nucleus will normally stay in contact with one another for an extremely long time, there's always a tiny chance that they'll find themselves temporarily separated by a distance that's beyond the reach of the nuclear force. The nucleons will then suddenly be free of one another, and electrostatic repulsion will push them apart in a process called **radioactive decay**. The quantum process that allows the nucleons to escape from the nuclear force without first obtaining the energy needed to surmount the energy barrier is called **tunneling** because the nucleons effectively tunnel through the barrier. We first encountered quantum tunneling in Section 14.3, when we saw that erasing flash memory requires that electrons tunnel through an insulating barrier.

Natural radioactive decay is a perfectly random process. Although half of a large population of identical radioactive nuclei will decay in a certain amount of time, you absolutely cannot predict in advance which of the original nuclei will have survived. Because of this randomness, radioactive decay is characterized simply by a **half-life**, the time required for half of the nuclei to decay. After one half-life, only half the original nuclei will remain intact. If you wait a second half-life, only a quarter of the original nuclei will remain (half of the remaining half). After a third half-life, only an eighth will remain (half of the remaining half of a half). And so on.

This halving of the population with each additional half-life is a type of *exponential decay*. In general, the fraction of nuclei remaining after a given amount of time is one-half raised to the power of the time divided by the half-life. Written as a word equation, that relationship is:

$$\text{fraction remaining} = \left(\frac{1}{2}\right)^{\text{elapsed time/half-life}} \quad (15.1.1)$$

in symbols:

$$\frac{N}{N_0} = \left(\frac{1}{2}\right)^{t/T_{1/2}}$$

and in everyday language:

Radioactivity goes away, but you have to wait it out.

Although most radioactive nuclei have short half-lives and don't linger long in our environment, there are a few with half-lives of billions of years. It is those long-lived radioactive nuclei, particularly uranium and thorium nuclei, that have survived since the formation of Earth, remain abundant in nature, and gave rise to nuclear weapons.

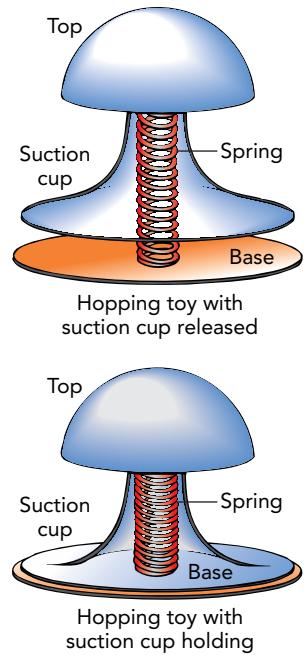


Fig. 15.1.4 A spring and a suction cup combine to form a toy that hops suddenly after a long wait. The spring tries to separate the top from the base, while the suction cup tries to hold the two parts together. Energy you store in the spring is released when the leaking suction cup eventually allows the spring to expand.

Check Your Understanding #2: Who Hid the Energy?

When a large nucleus breaks apart into fragments, it releases a great deal of energy. In what form was that energy stored in the intact nucleus?

Answer: It was stored as electrostatic potential energy in the repulsion between protons.

Why: Although it is routinely called nuclear energy, much of the energy stored in radioactive nuclei is actually electrostatic potential energy. Assembling a giant nucleus out of positively charged particles requires a considerable amount of work against electrostatic forces, and it is that stored work that's released when the nucleus decays.

Check Your Figures #1: A Real Hand-Me-Down

A small fraction of the carbon found in Earth's atmosphere is carbon 14, a rare, radioactive form that's synthesized by cosmic rays and has a half-life of 5730 years. While plants and animals are alive, they incorporate carbon 14 into their tissues along with ordinary carbon. Once they die, the fraction of carbon 14 in their tissues begins to decrease as the carbon 14 nuclei decay. If you are a museum curator and someone donates a garment that is supposed to be 1001 years old, what fraction of the original carbon 14 should you expect to find in that garment when you examine its carbon content?

Answer: The fraction of the original carbon 14 that remains should be about 0.886.

Why: According to Eq. 15.1.1, the fraction of carbon 14 nuclei remaining after 1001 years should be

$$\begin{aligned}\text{fraction remaining} &= \left(\frac{1}{2}\right)^{1001 \text{ years}/5730 \text{ years}} \\ &= \left(\frac{1}{2}\right)^{0.1747} = 0.886\end{aligned}$$

If you find that fraction of carbon 14 when you test the garment, that indicates that the fibers in the garment come from plants or animals that died approximately 1001 years ago, although you can't tell exactly when the fabric itself was woven. If you find a larger fraction, that proves that the garment is less than 1001 years old.

Fission and Fusion

The more protons there are in a nucleus, the more they repel one another and the more likely they are to cause radioactive decay. Adding additional neutrons to the nucleus reduces this proton-proton repulsion by increasing the size of the nucleus without adding to its positive charge. However, adding too many neutrons also destabilizes the nucleus for reasons that we'll discuss in Section 15.3. Thus, constructing a stable nucleus is a delicate balancing act.

In nuclei with only a few protons, the attractive nuclear force wins big over the repulsive electrostatic force and the nucleons stick like crazy. These nuclei resemble hopping toys with weak springs and big suction cups; once brought together, the pieces never come apart. In fact, the average binding energy of the nucleons (the energy required to separate them from one another divided by the number of nucleons) would increase if these nuclei had even more protons and neutrons.

In nuclei with many protons, the electrostatic repulsion is so severe that the nuclear force can't hold the nucleons together for long. These nuclei decay rapidly. They resemble hopping toys with strong springs and small suction cups. The average binding energy of the nucleons would increase if these nuclei had fewer protons and neutrons.

In nuclei with roughly 26 protons, in between the two extremes we've just considered, the attractive nuclear force and repulsive electrostatic force are nicely balanced. These nuclei are extremely stable, and you can't increase the average binding energy of their nucleons by adding or subtracting nucleons. Smaller nuclei can release potential energy by growing to reach this intermediate size, while larger nuclei can release potential energy by shrinking toward the same goal.

For a small nucleus to grow, something must push more nucleons toward it. Electrostatic repulsion will initially oppose this growth, but once everything touches, the nuclear force will bind the particles together and release a large amount of potential energy. This coalescence process is called **nuclear fusion**.

For a large nucleus to shrink, something must separate its pieces beyond the reach of the nuclear force. Electrostatic repulsion will then push the fragments apart and release a large amount of potential energy. This fragmentation process is called **nuclear fission**.

The energies released when small nuclei undergo fusion or when large nuclei undergo fission are enormous compared to chemical energies. Uranium, a large nucleus, converts about 0.1% of its mass into energy when it breaks apart. Hydrogen, a tiny nucleus, converts about 0.3% of its mass into energy when it fuses with other hydrogen nuclei. Kilogram for kilogram, nuclear reactions release about 10 million times more energy than chemical reactions. Fortunately, they're much harder to start.

With that scientific background, let's follow a sequence of discoveries near the start of the twentieth century. Natural radioactive decay was discovered accidentally by French physicist Antoine-Henri Becquerel (1852–1908) in 1896. Intrigued by the recent discovery of X-rays, he began looking for materials that might emit X-rays after exposure to light. To his surprise, he found that uranium fogged photographic plates, even through an opaque shield and even without exposure to light. His discovery was soon confirmed and elaborated on by Polish-born French physicist Marie Curie (1867–1934) and French chemist Pierre Curie (1859–1906). This wife and husband team discovered several new radioactive elements, including polonium (named after Marie's homeland) and radium.

In 1911, British physicist Ernest Rutherford (1871–1937) discovered that atoms have nuclei. He subsequently found that nuclei sometimes shatter when struck by energetic helium nuclei. In 1932, British physicist James Chadwick (1891–1974) discovered a fragment of the nucleus, the neutron, that has no electric charge and can thus approach a nucleus without any electrostatic repulsion. It was soon discovered that neutrons stick to the nuclei of many atoms.

The crucial discovery that made the atomic bomb possible was neutron-induced fission of nuclei. In 1934, Italian physicist Enrico Fermi (1901–1954) and his colleagues were trying to solve a particular riddle about the nucleus, a radioactive decay process called *beta decay*. They were adding neutrons to the nuclei of every atom they could get hold of. When they added neutrons to uranium, with its huge nucleus, they observed the production of some very short-lived radioactive systems. They thought that they had formed ultraheavy nuclei and even went so far as to give these new elements tentative names.

Four years later, however, Austrian physicists Lise Meitner (1878–1968) and Otto Frisch (1904–1979) and German chemists Otto Hahn (1879–1968) and Fritz Strassmann (1902–1980) collectively showed 1 that what Fermi's group had actually done was to fragment uranium into lighter nuclei (Fig. 15.1.5). Many of the fragments created by this **induced fission** were neutrons, which could themselves cause the destruction of other uranium nuclei.

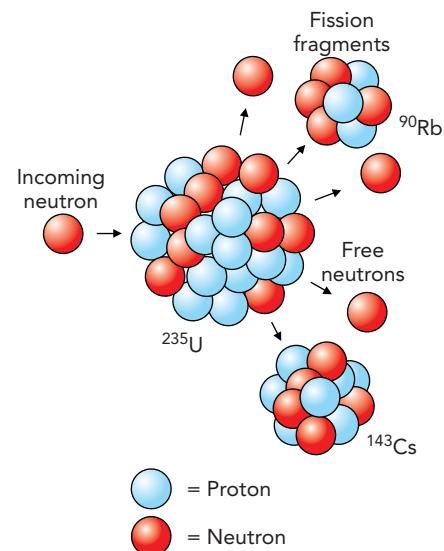


Fig. 15.1.5 When a neutron strikes a uranium nucleus, there's a good chance that the nucleus will fall apart into fragments. This process is called induced fission. Among the fragments of induced fission are other neutrons.

Check Your Understanding #3: A Sticky Nuclear Problem

If you take two intermediate-size nuclei and combine them to make a single uranium nucleus, will the process release or consume energy?

Answer: It will consume energy.

Why: To merge two smaller nuclei and form a uranium nucleus, you will have to push the two together with considerable force because they contain many protons. You will have to do considerable work to bring these nuclei close enough for the nuclear force to bind them. The energy you invested in this new nucleus is the same energy that is released when it undergoes fission.

1 Austrian-born physicist Lise Meitner moved to Berlin in 1907 and soon began a 30-year collaboration with chemist Otto Hahn. In 1934, she convinced Hahn to join her in studying nuclear processes, and they made great progress. Unfortunately, Meitner's Jewish ancestry made her a target of Nazi academic restriction and she fled to Sweden in 1938. Meitner continued to guide their collaboration through letters. Only months after she left, Hahn and his assistant Fritz Strassmann found that neutron irradiation of heavy elements was creating smaller rather than larger nuclei. Meitner and her nephew Otto Frisch soon developed a model of nuclear fission based on these measurements.

Hahn, however, published the results without Meitner's name on the paper, ostensibly to avoid Nazi interference. As a result of this omission, the 1944 Nobel Prize in Chemistry was awarded to Hahn alone. Hahn went on to claim that Meitner was simply his assistant, rather than the leader of their joint effort. In recognition of this gross injustice, element 109 was named meitnerium in 1994.

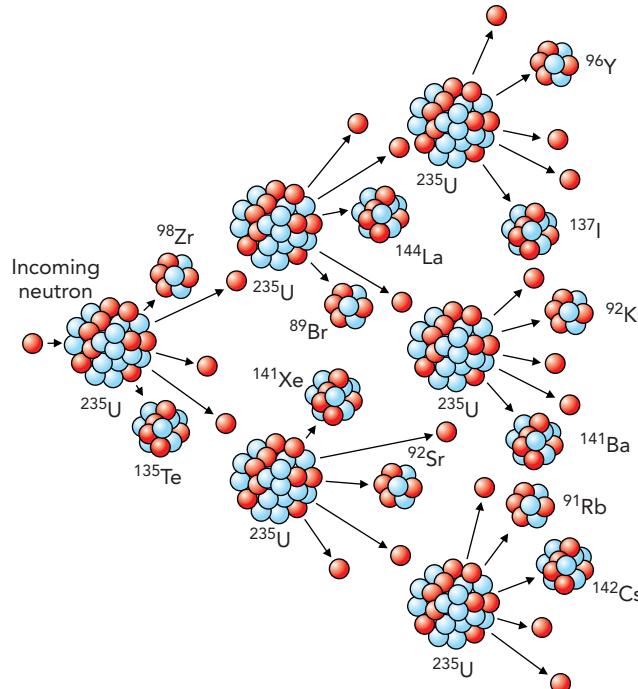


Fig. 15.1.6 A chain reaction occurs when the fragments of one fissioning nucleus induce fission in at least one additional nucleus. Such a chain reaction is particularly easy in ^{235}U , the light isotope of uranium, where each fissioning nucleus releases an average of 2.5 neutrons.

Chain Reactions and the Fission Bomb

Physicists quickly realized that a **chain reaction** was possible, a reaction in which the fission of one uranium nucleus would induce fission in two nearby uranium nuclei, which would in turn induce fission in four other uranium nuclei, and so on (Fig. 15.1.6). The result would be a catastrophic nuclear process in which many or even most of the nuclei in a piece of uranium would shatter and release fantastic amounts of energy.

In a sense, the remaining work toward both the atomic bomb and the hydrogen bomb was a matter of technical details. Only four conditions had to be satisfied for an atomic or *fission bomb* to be possible:

1. A source of neutrons had to exist in the bomb to trigger the explosion.
2. The nuclei making up the bomb had to be **fissionable**; that is, they had to fission when hit by a neutron.
3. Each induced fission had to produce more neutrons than it consumed.
4. The bomb had to use the released neutrons efficiently so that each fission induced an average of more than one subsequent fission.

Meeting the first condition was easy. Many radioactive elements emit neutrons. Meeting the second and third conditions was more difficult. Here is where uranium fit into the picture. It was known to be fissionable, and it was known to release more neutrons than it consumed.

However, not all uranium nuclei are the same. Although a uranium nucleus must contain 92 protons, so that it forms a neutral atom with 92 electrons and has all the chemical characteristics of uranium (U in Fig. 13.2.5), the number of neutrons in that nucleus is somewhat flexible. Nuclei that differ only in the numbers of neutrons they contain are called **isotopes**, and natural uranium nuclei come in two isotopes: ^{235}U and ^{238}U , where the number specifies how many nucleons are in each nucleus. The ^{235}U nucleus contains 92 protons and 143 neutrons, for a total of 235 nucleons. In contrast, the ^{238}U nucleus contains 238 nucleons: 92 protons and 146 neutrons.

It turns out that only ^{235}U is suitable for a bomb. It's marginally stable, with too many protons for the nuclear force to keep together, even with the diluting effects of 143 neutrons. Like many proton-rich nuclei, ^{235}U eventually undergoes **alpha decay**; it emits a helium nucleus (^4He) and thereby loses two protons and two neutrons. ^{235}U has a radioactive half-life of 710 million years. However, when struck by a neutron, ^{235}U shatters immediately into fragments, and this induced fission releases about 2.5 neutrons.

^{238}U is slightly more stable than the lighter isotope, and its half-life is 4.51 billion years. The ^{238}U nucleus absorbs most neutrons without undergoing fission. Instead, it undergoes a series of complicated nuclear changes that eventually convert it into plutonium, an element not found in nature. It becomes ^{239}Pu , which has a nucleus with 94 protons and 145 neutrons. As we'll see later on, plutonium itself is useful for making nuclear weapons.

Thus ^{238}U actually slows a chain reaction rather than encouraging it. Since only ^{235}U can support a chain reaction, natural uranium has to be separated before it can be used in a bomb. ^{235}U is quite rare, however. Earth's store of uranium nuclei was created long ago, in the explosion of a dying star. That *supernova* heated the nuclei of smaller atoms so hot that they collided together and stuck. Uranium nuclei were formed, with the supernova's energy trapped inside them. They were incorporated in the Earth during its formation about 4 or 5 billion years ago and have been decaying ever since. The only uranium isotopes that remain in any quantity are ^{235}U and ^{238}U . Since ^{235}U is less stable, its percentage of the naturally occurring uranium nuclei has dwindled to only 0.72%. The remaining 99+% of the uranium is ^{238}U .

Separating ^{235}U from ^{238}U is extremely difficult. Since atoms containing these two nuclei differ only in mass, not in chemistry, they can be separated only by methods that compare their masses. Because the mass difference is relatively small, heroic measures are needed to extract ^{235}U from natural uranium. During World War II and the Cold War era, the U.S. government developed enormous facilities for separating the two uranium isotopes. The need for such installations is one of the major obstacles to the proliferation of nuclear weapons.

The last condition for sustaining a chain reaction is that the bomb must use neutrons efficiently, so that each fission induces an average of more than one subsequent fission. That means that the bomb's contents can't absorb neutrons wastefully and that it can't let too many of them escape without causing fission. A lump of relatively pure ^{235}U wouldn't absorb neutrons wastefully, but it might allow too many of them to escape through its surface. For a chain reaction to occur, the lump must be large enough that each neutron has a good chance of hitting another nucleus before it leaves the lump. The lump also should have a minimal amount of surface. It should be a sphere.

How large must that sphere be? Since atoms are mostly empty space, a neutron can travel several centimeters through a lump of uranium without hitting a nucleus. Thus a golf ball-size sphere of uranium would allow too many neutrons to escape. For a bare sphere of ^{235}U , the **critical mass** required to initiate a chain reaction is about 52 kg (115 lbm), a ball about 17 cm (7 in) in diameter. At that point, each fission will induce an average of one subsequent fission. For an **explosive chain reaction**, in which each fission induces an average of much more than one fission, additional ^{235}U is needed—a **supercritical mass**. About 60 kg (132 lbm) will do it.

By 1945, the scientists and engineers of the Manhattan Project had found ways to meet these four conditions and were prepared to initiate an explosive chain reaction. They had accumulated enough ^{235}U to construct a supercritical mass. Carefully machined pieces of ^{235}U would be put into the bomb so that they would join together at the moment the bomb was to explode. When the critical mass was reached, a few initial neutrons would start the chain reaction. When the uranium became supercritical, catastrophic fission would quickly turn it into a tremendous fireball.

However, assembly was tricky. The supercritical mass had to be completely assembled before the chain reaction was too far along; otherwise the bomb would begin to overheat and explode before enough of its nuclei had time to fission. In pure ^{235}U , the time it takes for one fission to induce the next fission is only about 10 ns (10 nanoseconds or 10 billionths of

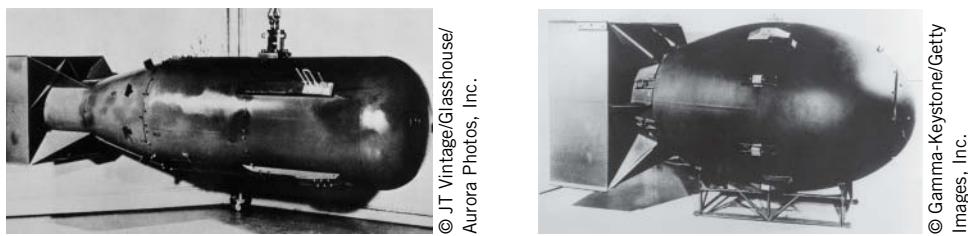


Fig. 15.1.7 The Little Boy (a) used a cannon to fire a cylinder of uranium into an incomplete sphere of uranium. The Fat Man (b) used high explosives to crush a sphere of plutonium to extreme density. Little Boy destroyed Hiroshima, and Fat Man destroyed Nagasaki.

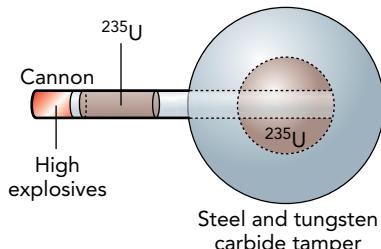


Fig. 15.1.8 The concept behind Little Boy was simple. A sphere of ^{235}U was divided into two parts so that neither was a critical mass on its own. One part was a hollow sphere, and the other part was a cylinder that would complete that sphere. When the bomb was detonated, a cannon fired the cylinder into the hollow sphere, creating a supercritical mass and initiating an explosive chain reaction.

a second). In a supercritical mass, each generation of fissions is much larger than the previous generation, so it takes only a few dozen generations to shatter a significant fraction of the uranium nuclei. The whole explosive chain reaction is over in less than a millionth of a second, with most of the energy released in the last few generations (about 30 ns).

To make sure that the assembly was complete before the bomb exploded, it had to be done extraordinarily quickly. For the ^{235}U bomb called “Little Boy” (Fig. 15.1.8a) that was exploded over Hiroshima, Japan, on August 6, 1945, at 8:15 AM and was responsible for the deaths of about 200,000 Japanese citizens,

the supercritical mass was assembled when a cannon fired a cylinder of ^{235}U through a hole in a sphere of ^{235}U (Fig. 15.1.8). When the cylinder was centered in the hole, it completed a 60-kg sphere of uranium, housed in a tungsten carbide and steel container. This container confined the uranium, holding it together with its inertia as the explosion began. An explosive chain reaction started immediately, and by the time the uranium blew itself apart, 1.3% of the ^{235}U nuclei had fissioned. The energy released in that event was equal to the explosion of about 15,000 tons of TNT.

But Little Boy was actually the second nuclear explosion. Its concept was so foolproof and its ^{235}U so precious that Little Boy was dropped without ever being tested. However, the Manhattan Project had also developed a plutonium-based bomb that involved a much more sophisticated concept. This bomb was much less certain to work, so it was tested once before it was used.

“The Gadget,” as the first atomic bomb was called, didn’t use ^{235}U . Instead, it used plutonium that had been synthesized from ^{238}U in nuclear reactors. A nuclear reactor carries out a controlled chain reaction in uranium, and neutrons from this chain reaction can convert ^{238}U into ^{239}Pu .

Like ^{235}U , ^{239}Pu meets the conditions for a bomb. The ^{239}Pu nucleus is relatively unstable, with a half-life of only 24,400 years. It fissions easily when struck by a neutron and releases an average of three neutrons when it does. Thus ^{239}Pu can be used in a chain reaction. For a bare sphere of ^{239}Pu , the critical mass is about 10 kg (22 lbm)—a ball about 10 cm (4 in) in diameter.

There is a problem with ^{239}Pu , however. It’s so radioactive and releases so many neutrons when it fissions that a chain reaction develops almost instantly. There is much less time to assemble a supercritical mass of plutonium than there is with uranium. The cannon assembly method won’t work because the plutonium will overheat and blow itself apart before the cylinder can fully enter the sphere.

Thus a much more sophisticated assembly scheme was employed for the Gadget. At 5:29 AM on June 16, 1945, over 2000 kg (4400 lbm) of carefully designed high explosives crushed or imploded a sphere of plutonium (Fig. 15.1.9) inside The Gadget (Fig. 15.1.10) at Alamogordo, New Mexico. By itself, the 6.1-kg (13.4-lbm) sphere wasn’t large enough to be a critical mass; it was a **subcritical mass**. However, it was surrounded by a *tamper* of ^{238}U whose massive nuclei reflected many neutrons back into the plutonium like marbles



Fig. 15.1.9 The first atomic bomb, nicknamed “The Gadget,” was detonated on a tower at a remote desert site near Alamogordo, New Mexico, on June 16, 1945. Here employees of the top-secret Manhattan Project are seen hoisting parts of the plutonium bomb onto the tower before the explosion.

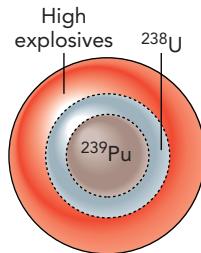


Fig. 15.1.10 The concept used in The Gadget and Fat Man was relatively sophisticated. Carefully shaped high explosives crushed a baseball-sized sphere of ^{239}Pu inside a neutron-reflecting shell of ^{238}U . The plutonium was compressed to high density and quickly reached supercritical mass, initiating an explosive chain reaction.

bouncing off bowling balls. The implosion process compressed the plutonium well beyond its normal density. With the plutonium nuclei packed more tightly together, they were more likely to be struck by neutrons and undergo fission.

The scheme worked. The chain reaction that followed caused 17% of the plutonium nuclei to fission and released energy equivalent to the explosion of about 22,000 tons of TNT. The tower and equipment at the Trinity test site disappeared into vapor, and the desert sands turned to glass for hundreds of meters in all directions. A nearly identical device named “Fat Man” (Fig. 15.1.7b) was dropped over Nagasaki, Japan, on August 9, 1945, at 11:02 AM, where it ultimately killed about 140,000 people.

In the years following the first fission bombs, development focused on how best to bring the fissionable material together. The longer a supercritical mass could be held together before it overheated and exploded, the larger would be the fraction of its nuclei that would fission and the greater the explosive yield. The crushing technique of The Gadget and Fat Man became the standard, and bombs grew smaller and more efficient at using their nuclear fuel. The implosion process reduced the amount of plutonium needed to reach a supercritical mass, so very small fission bombs were possible. The smallest atomic bomb, the “Davy Crockett,” weighed only about 220 N (50 lbf).

Check Your Understanding #4: Tickling the Dragon’s Tail

What would happen if you slowly moved two 30-kg (66-lbm) hemispheres of ^{235}U together to form a single sphere?

Answer: Before they actually touched, a chain reaction would occur.

Why: As soon as the hemispheres became close enough that each fission began to induce an average of one subsequent fission, a chain reaction would occur. You would have assembled a critical mass. The hemispheres would begin to get hotter and hotter and would eventually melt or explode. Because they would blow apart before most of the nuclei could fission, any explosion would be rather small. But you would suffer severe radiation injury. Just such an accident killed Louis Slotin during the Manhattan Project.

The Fusion or Hydrogen Bomb

Since fissionable material begins to explode as soon as it exceeds critical mass, this critical mass limits the size and potential explosive yield of a fission bomb. In search of a way around this limit, bomb scientists took a look back at small nuclei and figured out how to extract energy by sticking them together.

The fission bomb brought to Earth, for the first time, temperatures that had previously been observed only in stars. Stars obtain most of their energy by fusing hydrogen

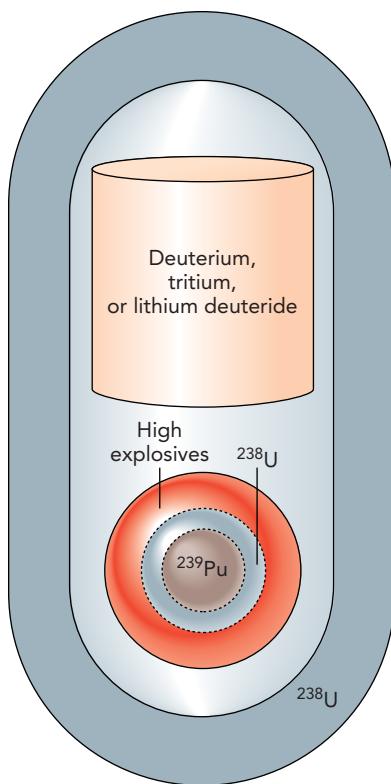


Fig. 15.1.11 A fusion bomb releases energy by fusing deuterium (^2H) and tritium (^3H) nuclei together to form helium (^4He) and neutrons. This fusion is initiated by heating the hydrogen to more than 100 million degrees Celsius with a fission bomb. High-energy neutrons released by the fusion process then induce fission in the ^{238}U tamper, releasing still more energy. Lithium (^6Li) produces tritium when exposed to neutrons.

nuclei together to form helium nuclei, a process that releases a great deal of energy. Because hydrogen nuclei are protons and repel one another with tremendous forces, hydrogen doesn't normally undergo fusion here on Earth. To cause fusion, something must bring those protons close enough for the nuclear force to stick them together. The only practical way we know of to bring the nuclei together is to heat them so hot that they crash into each other. That is what happens in a *fusion bomb*, also called a *thermonuclear* or *hydrogen bomb*.

In a fusion bomb, an exploding fission bomb heats a quantity of hydrogen to about 100 million degrees Celsius (Fig. 15.1.11). At that temperature, hydrogen nuclei begin to collide with one another. To ease the nuclear fusion processes, heavy isotopes of hydrogen are used: *deuterium* (^2H) and *tritium* (^3H). While the normal hydrogen nucleus (^1H) contains only a proton, the deuterium nucleus contains a proton and a neutron. The tritium nucleus contains a proton and two neutrons. When a deuterium nucleus collides with a tritium nucleus, they stick to form a helium nucleus (^4He) and a free neutron. Because this process converts about 0.3% of the original mass into energy, the helium nucleus and the neutron fly away from one another at enormous speeds.

Since hydrogen won't explode spontaneously, even in huge quantities, a hydrogen bomb can be extremely large. A fission bomb is used to set it off, but after that, the sky's the limit. Some enormous fusion bombs were constructed and tested during the early days of the Cold War. These bombs usually consisted of a fission starter and a hydrogen follower, all wrapped up in a tamper of ^{238}U . The tamper confined the hydrogen as the fusion began. Once fusion was underway, converting deuterium and tritium into helium and neutrons, the neutrons collided with ^{238}U nuclei in the tamper. These fusion neutrons were so energetic that they were able to induce fissions even in ^{238}U nuclei and release still more energy. Overall, this structure is sometimes called a *fission–fusion–fission bomb*.

A variation on this bomb is the so-called *neutron* or *enhanced radiation bomb*. That bomb has no ^{238}U tamper, so the energetic neutrons from the fusion process travel out of the explosion and irradiate everything in the vicinity. This bomb is lethal to humans, but because its blast is relatively weak, it is not particularly destructive to property.

Tritium is a radioactive isotope created in nuclear reactors. It has too many neutrons to be a stable nucleus and slowly decays into a light isotope of helium (^3He). Because tritium has a half-life of 12.3 years, fusion bombs containing tritium require periodic maintenance to replenish their tritium.

Many fusion bombs use solid lithium deuteride instead of deuterium and tritium gases. Lithium deuteride is a salt containing lithium (^6Li) and deuterium (^2H). When a neutron from the fission starter collides with a ^6Li nucleus, the two fragment into a helium nucleus (^4He) and a tritium nucleus (^3H). In the bomb, lithium deuteride is quickly converted into a mixture of deuterium, tritium, and helium, which then undergoes fusion.



Check Your Understanding #5: It's Hard to Get Together

Why must hydrogen be heated to an extremely high temperature to initiate fusion?

Answer: The protons in hydrogen nuclei repel one another quite strongly at short distances. They must be moving very quickly with lots of thermal energy for them to touch.

Why: In a gas, thermal energy takes the form of kinetic energy. The hotter the gas, the faster the particles are moving. At 100 million degrees Celsius, the hydrogen nuclei are moving so fast that they can overcome their electrostatic repulsion and touch. When they do, the nuclear force pulls them together and they fuse.

Heat, Radiation, and Fallout

Once a nuclear weapon has exploded—after its fissionable material has fissioned and its fusible material has fused—what then? First, a vast number of nuclei and subatomic particles emerge from the explosion at enormous speeds, many at nearly the speed of light. These particles crash into nearby atoms and molecules, heating them to fantastic temperatures and producing a local fireball around the bomb itself. They also cause extensive radiation damage in the surrounding area.

Second, a flash of light emerges from the explosion, caused partly by the fission and fusion processes themselves and partly by the ultrahot fireball that follows. This light is not only visible light, but also every portion of the electromagnetic spectrum from infrared to visible to ultraviolet to X-rays to gamma rays. It burns things nearby, inside and out.

Third, the explosion creates a huge pressure surge in the air around the fireball. A shock wave propagates outward from the fireball at the speed of sound, knocking over everything in its path for a considerable distance. Fourth, the rarefied and superheated air rushes upward, lifted by buoyant forces, to create a towering mushroom cloud.

However, the most insidious aftereffect of a nuclear explosion is *fallout*, the creation and release of radioactive nuclei. Fission converts uranium and plutonium nuclei into smaller nuclei. Each new nucleus has several dozen protons and its share of neutrons from the nucleus that fissioned. These new nuclei attract electrons and become seemingly normal atoms like iodine or cobalt. But while large nuclei such as uranium need extra neutrons to dilute their protons and reduce their electrostatic repulsions, intermediate and small nuclei like those of iodine and cobalt don't need as many neutrons. The new nuclei wind up with too many neutrons and are radioactive. They have half-lives that are anywhere from thousandths of a second to thousands of years.

Until they decay, the atoms that contain these nuclei are almost indistinguishable from normal atoms. They are radioactive isotopes of common atoms, and our bodies naively incorporate them into our tissues. There they sit, performing whatever chemical tasks our bodies require of them. Eventually, however, these radioactive atoms fall apart and release nuclear energy. Because each radioactive decay that occurs near us or inside us releases perhaps a million times more energy than is present in a chemical bond, these decays cause chemical changes in our cells. They can kill cells or damage the cells' genetic information, potentially causing cancer.

This **transmutation of elements**, the restructuring of nuclei to transform atoms of one element into another, occurs in an uncontrolled fashion in nuclear weapons and produces a lethal mix of unstable isotopes. Those isotopes take years to decay out of the environment, and all anyone can do is wait. Even nuclear weapons with poor explosive yields, so-called dirty bombs, can litter the surrounding landscape with radioactive debris. Nuclear reactors produce similar mixtures of radioactive isotopes in their fuel assemblies and core structures, which is why disposing of spent nuclear fuel remains so problematic.

On the other hand, radioactive isotopes have been a boon to medicine and biochemistry, where many of them have found valuable and life-saving applications. Moreover, in controlled circumstances, elements can be transmuted systematically to generate primarily desirable isotopes rather than a random assortment. Such transmutation is difficult and expensive, and it involves nuclear rather than chemical processes. Although the alchemists' dream of transmuting lead into gold is finally possible, it's not a path to riches.

COMMON MISCONCEPTIONS: Radiation and Radioactivity

Misconception: When a material such as food is exposed to microwave, radio wave, infrared, or ultraviolet radiation, it may become radioactive.

Resolution: To render a material radioactive, something must alter its nuclei so that they are no longer stable. Such an alteration requires vastly more energy than one of those low-energy photons can provide. The only forms of electromagnetic radiation with enough energy per photon to affect nuclei are gamma rays and, occasionally, X-rays.

Check Your Understanding #6: Not So Good to Eat

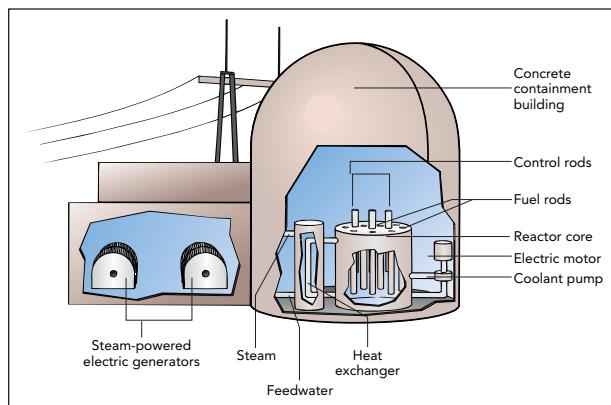
Normal iodine (^{127}I) is stable forever. But the fission product ^{131}I is radioactive and has a half-life of about 8 days. What are the consequences of eating ^{131}I ?

Answer: The ^{131}I is incorporated into your body and exposes you to radiation, particularly during the first few weeks before most of it has decayed away.

Why: Your body can't distinguish ^{127}I from ^{131}I because they're chemically identical. Since your body uses iodine in its functions, any ^{131}I that you ingest is likely to become part of your body's iodine supply. Over the next 8 days, about half of this iodine will decay and expose you to radiation. The remaining ^{131}I is also radioactive, and half of this amount will decay in the next 8 days. Thus after 16 days, only one-quarter of the original amount will remain. After 24 days, only one-eighth will remain. And so on.

SECTION 15.2

Nuclear Reactors



Making weapons weren't the only possibilities open to nuclear scientists and engineers at the end of the 1930s. While nuclear fission chain reactions and thermonuclear fusion were clearly ways to unleash phenomenal destructive energy, they could also provide virtually limitless sources of useful energy. By controlling the same nuclear reactions that occur in nuclear weapons, people have since managed to extract nuclear energy for constructive uses. In the half century since their conception, nuclear fission reactors have developed into a fairly mature technology and have become one of our major sources of energy. Nuclear fusion

power remains an elusive goal, but efforts continue to harness this form of nuclear energy as well. Competing concerns about nuclear safety, nuclear weapons proliferation, and climate change will no doubt influence the future of these technologies.

Questions to Think About: Why doesn't a nuclear reactor explode once it reaches a critical mass? How is the energy released in a nuclear reactor turned into electricity? What caused the famous accidents at Three Mile Island, Chernobyl, and Fukushima Daiichi? How can hydrogen fusion be initiated in a container? How is hydrogen held together at the temperatures needed for fusion?

Experiments to Do: A nuclear reactor operates very close to critical mass. Below critical mass, each spontaneous fission induces a limited number of subsequent fissions. Above critical mass, the number of subsequent fissions becomes unlimited. A similar effect occurs in a sand pile as it becomes steeper. If you pour sand slowly onto the pile, its shape will change and its steepness will increase. Initially, the grains will stay where they land, but once the pile becomes quite steep, they will begin to roll down the pile's sides. If the pile becomes too steep, a single grain of sand can trigger an avalanche. That same behavior occurs in a reactor that's near critical mass—a single spontaneous fission can trigger an enormous number of subsequent fissions.

Nuclear Fission Reactors

Assembling a critical mass of uranium doesn't always cause a nuclear explosion. In fact, it's rather hard to cause a big explosion. The designers of the atomic bomb had to assemble not just a critical mass but a supercritical mass, and they had to do it in much less than a millionth of a second. That rapid assembly is not something that happens easily or by accident. It's much easier to reach a critical mass slowly, in which case the uranium will simply become very hot. It may ultimately explode from overheating, but it will not vaporize everything in sight.

This slow assembly of a critical mass is the basis for nuclear fission reactors. Their principal product is heat, which is usually used to generate electricity. Fission reactors are much simpler to build and operate than fission bombs because they don't require such purified fissionable materials. In fact, with the help of some clever tricks, nuclear reactors can be made to operate even with natural uranium.

Let's begin by showing that a fission chain reaction doesn't always lead to an explosion. What's important is just how fast the fission rate increases. In an atomic bomb, it increases breathtakingly quickly. At detonation, the fissionable material is far above the critical mass, so the average fission induces not just one but perhaps two subsequent fissions. With only about 10 ns (10 nanoseconds) between one fission and the two it induces, the fission rate may double every 10 ns. In less than a millionth of a second, most of the nuclei in the material undergo fission, releasing their energy before the material has time to blow apart.

However, things aren't as dramatic right at critical mass, where the average fission induces just one subsequent fission. Since each generation of fissions simply reproduces itself, the fission rate remains essentially constant. Only spontaneous fissions cause it to rise at all. The fissionable material steadily releases thermal energy, and that energy can be used to power an electric generator.

A nuclear reactor contains a core of fissionable material. Because of the way in which this core is assembled, it's very close to a critical mass. Several neutron-absorbing rods, called control rods, which are inserted into the reactor's core, determine whether it's above or below critical mass. Pulling the control rods out of the core increases the chance that each neutron will induce a fission and moves the core toward supercriticality. Dropping the control rods into the core increases the chance that each neutron will be absorbed before it can induce a fission and moves the core toward subcriticality.

A nuclear reactor uses feedback to maintain the fission rate at the desired level. If the fission rate becomes too low, the control system slowly pulls the control rods out of the core to increase the fission rate. If the fission rate becomes too high, the control system drops the control rods into the core to decrease the fission rate. It's like driving a car. If you're going too fast, you ease off the gas pedal. If you are going too slowly, you push down on the gas pedal.

The car-driving analogy illustrates another important point about reactors. Both cars and reactors respond relatively slowly to adjustments of their controls. It would be hard to drive a car that stopped instantly when you lifted your foot off the gas pedal and leaped to supersonic speed when you pushed your foot down. Similarly, it would be impossible to operate a reactor that immediately shut down when you dropped the control rods in and that instantly exploded when you pulled the control rods out.

But reactors, like cars, don't respond quickly to movements of the control rods. That's because the final release of neutrons following some fissions is slow. When a ^{235}U nucleus fissions, it promptly releases an average of 2.47 neutrons which induce other fissions within a thousandth of a second. However, some of the fission fragments are unstable nuclei that decay and release neutrons long after the original fission. On average, each ^{235}U fission eventually produces 0.0064 of these delayed neutrons, which then go on to induce other fissions. It takes seconds or minutes for these delayed neutrons to appear, and they slow the response of the reactor. The reactor's fission rate can't increase quickly because it takes a long time for the delayed neutrons to build up. The fission rate can't decrease quickly because it takes a long time for the delayed neutrons to go away.

To further ease the operation of modern nuclear reactors, they are designed to be stable and self-regulating. This self-regulation ensures that the core automatically becomes subcritical if it overheats. As we'll see later on, this self-regulation was not present in the design of Chernobyl Reactor Number 4 and that led to disaster.

Check Your Understanding #1: No Delayed Decays

If the fission of ^{235}U produced only stable fission fragments, would a nuclear reactor be easier or harder to operate?

Answer: It would be much harder to operate.

Why: Without any delayed neutrons, one generation of fissions would finish inducing the next generation of fissions in about a thousandth of a second and the response of the reactor core to changes in the criticality would become extremely rapid. When the reactor core became supercritical, its fission rate would skyrocket and when it became subcritical, its fission rate would plummet. Trying to keep the fission rate steady would be a wild roller coaster ride, and the reactor would be impossible or impractical to control. Fortunately the delayed neutrons slow the reactor's response to changes in criticality, so it can be controlled fairly easily.

Thermal Fission Reactors

The basic concept of a nuclear reactor is simple: assemble a critical mass of fissionable material and adjust its criticality to maintain a steady fission rate. But what should the fissionable material be? In a fission bomb, it must be relatively pure ^{235}U or ^{239}Pu . In a fission reactor, however, it can be a mixture of ^{235}U and ^{238}U . It can even be natural uranium. The trick is to use *thermal neutrons*, slow-moving neutrons that have only the kinetic energy associated with the local temperature.

In a fission bomb, ^{238}U is a serious problem because it captures the fast-moving neutrons emitted by fissioning ^{235}U nuclei. Natural uranium alone can't sustain a chain reaction because its many ^{238}U nuclei gobble up most of the fast-moving neutrons before they can induce more fissions in the rare ^{235}U nuclei. The uranium must be highly *enriched*—most of its ^{238}U nuclei must be removed so that it contains far more than the natural abundance of ^{235}U .

Slow-moving neutrons have a different experience as they travel through natural uranium. For complicated reasons, the ^{235}U nuclei seek out slow-moving neutrons and capture them with remarkable efficiency. ^{235}U nuclei are so good at catching slow-moving neutrons that they easily win out over the more abundant ^{238}U nuclei. Even in natural uranium, a slow-moving neutron is more likely to be caught by a ^{235}U nucleus than it is by a ^{238}U nucleus. As a result, it's possible to sustain a nuclear fission chain reaction in natural uranium if all of the neutrons are slow moving.

Because ^{235}U nuclei emit fast-moving neutrons when they fission, natural or slightly enriched uranium alone can't use the slow-moving-neutron effect to maintain a chain reaction. However, most nuclear reactors contain something else besides uranium, natural or otherwise. Along with uranium, they use a material called a *moderator* that slows the neutrons down so that ^{235}U nuclei can grab them. A fast-moving neutron from a fissioning ^{235}U nucleus enters the moderator, rattles around for about a thousandth of a second, and emerges as a slow-moving neutron, one with only thermal energy left. It then induces fission in another ^{235}U nucleus. Once the moderator is present, even natural uranium can sustain a chain reaction! Reactors that carry out their chain reactions with slow-moving, or thermal, neutrons are called *thermal fission reactors*.

To be a good moderator, a material must remove energy and momentum from the neutrons without absorbing them. A fast-moving fission neutron entering that moderator should leave with only thermal energy. The best moderators are small nuclei that rarely or never absorb neutrons and don't fall apart during collisions with them. Hydrogen (^1H), deuterium (^2H), helium (^4He), and carbon (^{12}C) are all good moderators. When a fast-moving neutron hits the nucleus of one of these atoms, the collision resembles that between two bumper cars (Section 2.3). Because the fast-moving neutron transfers some of its energy and momentum to the nucleus, the neutron slows down while the nucleus speeds up. That collision is most effective when the neutron and nucleus have similar masses, so small nuclei are better moderators than large ones.

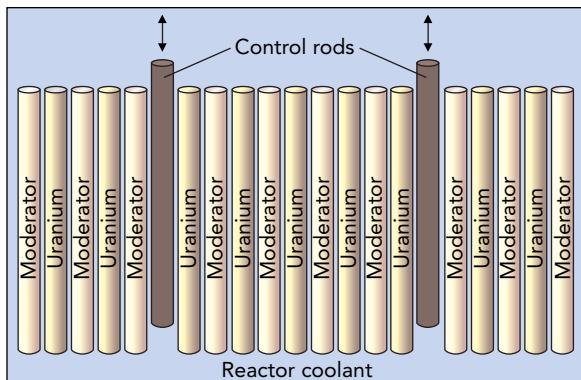


Fig. 15.2.1 The core of a thermal fission reactor consists of uranium pellets interspersed with a moderator that slows the fission neutrons to thermal energies. Neutron-absorbing control rods are inserted into the core to control the fission rate. A cooling fluid such as water flows through the core to extract heat.

Water, heavy water (water containing the heavy isotope of hydrogen, deuterium or ^2H), and graphite (carbon) are the best moderators for nuclear reactors. They slow neutrons down to thermal speeds without absorbing many of them. Of these moderators, heavy water is the best because it slows the neutrons quickly yet doesn't absorb them at all. However, heavy water is expensive because only 0.015% of hydrogen atoms are deuterium and separating that deuterium from ordinary hydrogen is difficult.

Graphite moderators were used in many early reactors because graphite is cheap and working with it is easy (see 2). However, graphite is a less efficient moderator than heavy water, so graphite reactors had to be big. Furthermore, graphite can burn and was partly responsible for two of the world's four major reactor accidents. Normal, or "light," water is cheap, safe, and an efficient moderator, but it absorbs enough neutrons that it can't be used with natural uranium. For use in a light-water reactor, uranium must be enriched slightly, to about 2–3% ^{235}U .

The core of a typical thermal fission reactor consists of small uranium oxide (UO_2) fuel pellets separated by layers of moderator (Fig. 15.2.1). A neutron released by a fissioning ^{235}U nucleus usually escapes from its fuel pellet, slows down in the moderator, and then induces fission in a ^{235}U nucleus in another fuel pellet. By absorbing some of these neutrons, the control rods determine whether the whole core is subcritical, critical, or supercritical. The ^{238}U nuclei are basically spectators in the reactor since most of the fissioning occurs in the ^{235}U nuclei.

In a practical thermal fission reactor, something must extract the heat released by nuclear fission. In many reactors, cooling water passes through the core at high speeds. Heat flows into this water and increases its temperature. In a boiling water reactor, the water boils directly in the reactor core, creating high-pressure steam that drives the turbines of an electric generator (Fig. 15.2.2). In a pressurized water reactor, the water is under enormous pressure, so it can't boil (Fig. 15.2.3). Instead, it's pumped to a heat exchanger outside the reactor. This heat exchanger transfers heat to water in another pipe, which boils to create the high-pressure steam that drives a generator (Fig. 15.2.4).

When properly designed, a water-cooled thermal fission reactor is inherently stable. The cooling water is actually part of the moderator. If the reactor overheats and the water escapes, there will no longer be enough moderator around to slow the fission neutrons down. The fast-moving neutrons will be absorbed by ^{238}U nuclei, and the chain reaction will slow or stop.

2 The first nuclear reactor was CP-1 (Chicago Pile-1), a thermal fission reactor constructed in a squash court at the University of Chicago. Each of the graphite bricks used in this pile contained two large pellets of natural uranium. By December 2, 1942, the pile was complete and would reach critical mass once the control rods were removed. As Enrico Fermi, the project leader, directed the slow removal of the last control rod, the pile approached criticality and the neutron emissions began to mount. It was noon, so Fermi called a famous lunch break. When everyone returned, they picked up where they had left off. At 3:25 PM, the pile reached critical mass and the neutron emissions increased exponentially. The reactor ran for 28 minutes before Fermi ordered the control rods to be dropped back in.

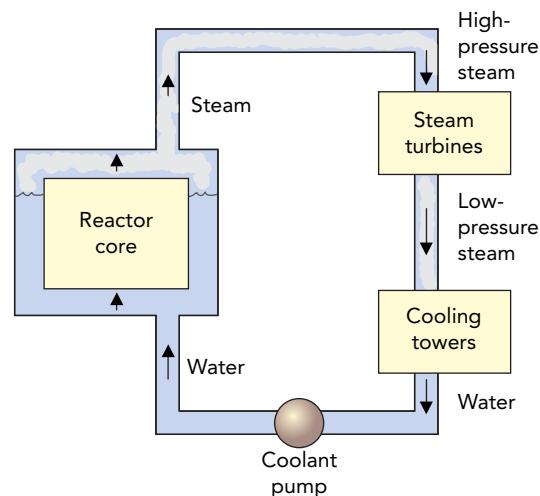


Fig. 15.2.2 In a boiling water reactor, cooling water boils inside the reactor core. It creates high-pressure steam that drives steam turbines and an electric generator. The spent steam condenses in a cooling tower and is then pumped back into the reactor.

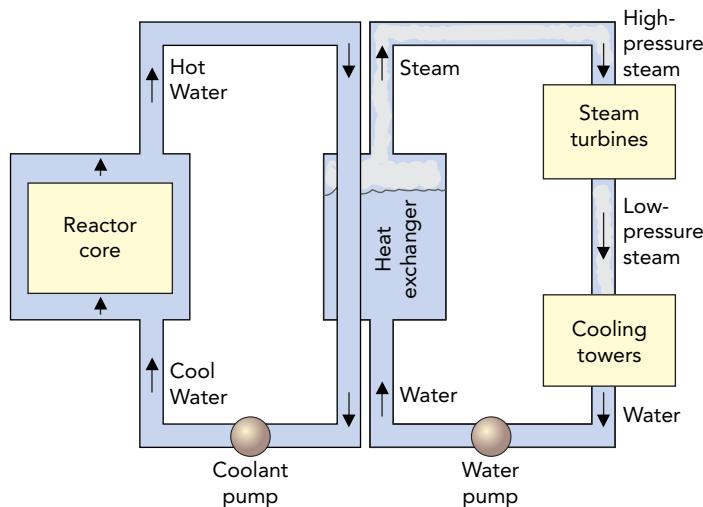


Fig. 15.2.3 In a pressurized water reactor, liquid water under great pressure extracts heat from the reactor core. A heat exchanger allows this cooling water to transfer heat to the water used to generate electricity. Water in the generating loop boils to form high-pressure steam, which then powers the steam turbines connected to the electric generators. The steam condenses back into liquid water and returns to the heat exchanger to obtain more heat.

Check Your Understanding #2: Nuclear Candles

Would wax made from hydrogen and carbon atoms be a good moderator?

Answer: Yes.

Why: Paraffin wax is made up of carbon and hydrogen atoms, both of which are good moderators. The fast fission neutrons don't care what molecules the carbon or hydrogen nuclei are inside. They just collide with those nuclei and lose energy and momentum. The problem with using wax as a reactor moderator is that it decomposes chemically in the intense radiation. However, wax can safely slow small numbers of neutrons in a laboratory situation.

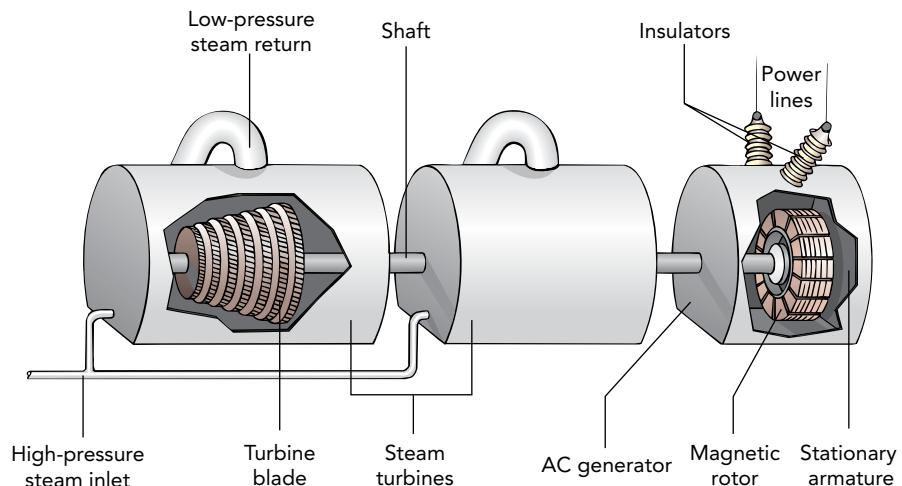


Fig. 15.2.4 High-pressure steam from a boiling water reactor core or the heat exchanger of a pressurized water reactor enters a series of turbines. There it does work on the turbine blades, losing pressure and temperature in the process. It emerges from the turbines as low-pressure steam and goes to the cooling towers to condense back into liquid water for reuse. The rotary mechanical power provided by the turbines turns generators that produce AC electric power.

Fast Fission Reactors

Thermal reactors require simple fuel and are relatively straightforward to construct. However, they consume only the ^{235}U nuclei and leave the ^{238}U nuclei almost unaffected. Anticipating the day when ^{235}U will become scarce, several countries have built a different kind of reactor that contains no moderator. Such reactors carry out chain reactions with fast-moving neutrons and are thus called fast fission reactors.

A fast fission reactor operates much like a controlled fission bomb and requires highly enriched uranium fuel as well. While a thermal fission reactor can get by with natural uranium or uranium that has been enriched to about 2–3% ^{235}U , a fast fission reactor needs 25–50% ^{235}U fuel. That way, there are enough ^{235}U nuclei around to maintain the chain reaction.

There is a side effect to operating a fast fission reactor. Many of the fast fission neutrons are captured by ^{238}U nuclei, which then transform into ^{239}Pu nuclei. Thus the reactor produces plutonium as well as heat. For that reason, fast fission reactors are often called breeder reactors—they create new fissionable fuel. The ^{239}Pu can eventually replace the ^{235}U as the principal fuel used in the reactor.

A thermal fission reactor makes some plutonium, which is usually allowed to fission in place, but a fast fission reactor makes a lot of it. Because plutonium can be used to make nuclear weapons, fast fission reactors are controversial. However, because they convert otherwise unusable ^{238}U into a fissionable material, they use natural uranium far more efficiently than thermal fission reactors.

One interesting complication of the unmoderated design is that fast fission reactors can't be water-cooled. If they were, the water would act as a moderator and slow the neutrons down. Instead, they are usually cooled with liquid sodium metal. The sodium nuclei of ^{23}Na rarely interact with fast-moving neutrons, so they don't slow them down.

Check Your Understanding #3: A Mix-Up at the Uranium Plant

What would happen if someone tried to fuel a fast fission reactor with natural uranium?

Answer: It would not sustain a chain reaction.

Why: Fast neutrons aren't good at inducing fission in ^{235}U . If the reactor doesn't slow down the neutrons, most of them will zip right by the ^{235}U nuclei and be captured by the more abundant ^{238}U nuclei of the natural uranium.

Fission Reactor Safety and Accidents

One of the greatest concerns with nuclear fission reactors is the control of radioactive waste. Anything that comes in contact with the reactor core or the neutrons it emits becomes somewhat radioactive. The fuel pellets themselves are quickly contaminated with all sorts of fission fragments that include radioactive isotopes of many familiar elements. Some of these radioactive isotopes dissolve in water or are gases, and all of them must be handled carefully.

The first line of defense against the escape of radioactivity is the large and sturdy containment vessel around the reactor. Because most of the radioactive materials remain in the reactor core itself or in the cooling fluid, they are trapped in the containment vessel. Whenever the nuclear fuel is removed for reprocessing, care is taken not to allow radioactive materials to escape.

The other great concern is the safe operation of the reactors themselves. Like any equipment, reactors experience failures of one type or another, and a safe reactor must not respond catastrophically to such failures. Toward that end, reactors have emergency cooling systems, pressure-relief valves, and many ways to shut down the reactor. For example, injecting a solution of sodium borate into the core cools it and stops any chain reactions. The boron nuclei in sodium borate absorb neutrons extremely well and are the main con-

tents of most control rods. However, the best way to keep reactors safe is to design them so that they naturally stop their chain reactions when they overheat.

There have been five major reactor accidents since the dawn of the nuclear age. The first of these accidents began on October 7, 1957, during a maintenance procedure at Windscale Pile 1, one of Britain's two original plutonium production reactors. The reactor was cooled by air rather than water and had a graphite moderator. The reactor's intense radiation modified the graphite's crystalline structure so that it gradually accumulated large amounts of chemical potential energy. That potential energy had to be released periodically by heating the graphite to about 250 °C so that it recrystallized. During the accident, however, hot spots developed unexpectedly in the moderator and fuel rods, the metallic uranium in those rods ignited, and the burning reactor core distributed radioactive debris across the British countryside.

The second serious accident occurred at Three Mile Island on March 28, 1979. This pressurized water thermal fission reactor shut down appropriately when the pumps that circulated water in the power-generating loop stopped unexpectedly. Although this water loop wasn't directly connected to the reactor, it was important for removing heat from the reactor core. Even though the reactor shut down immediately, its neutron-absorbing control rods swiftly stopping its chain reaction, the radioactive nuclei created by recent fissions were still decaying and releasing energy. The core continued to release heat, and it eventually boiled the water in the cooling loop. This water escaped from the loop through a pressure-relief valve, and the top of the reactor core became exposed. With nothing to cool it, the core became so hot that it suffered permanent damage. Some of the water from the cooling loop found its way into an unsealed room, and the radioactive gases it contained were released into the atmosphere.

The third and most serious accident occurred at Chernobyl Reactor Number 4 on April 26, 1986. This water-cooled, graphite-moderated thermal fission reactor was a cross between a pressurized water reactor and a boiling water reactor. Cooling water flowed through the reactor at high pressure but didn't boil until it was ready to enter the steam-generating turbines.

The accident began during a test of the emergency core-cooling system. To begin the test, the operators tried to reduce the reactor's fission rate. However, the core had accumulated many neutron-absorbing fission fragments, which made it incapable of sustaining a chain reaction at a reduced fission rate. The chain reaction virtually stopped. To get the chain reaction running again, the operators had to withdraw a large number of control rods. These control rods were motor-driven, and it would take about 20 seconds to put them back in again.

The operators now initiated the test by shutting off the cooling water. That should have immediately shut down the reactor by inserting the control rods, but the operators had overridden the automatic controls because they didn't want to have to restart the reactor a second time. With nothing to cool it, the reactor core quickly overheated and the water inside it boiled. The water had been acting as a moderator along with the graphite. However, the reactor was overmoderated, meaning it had more moderator than it needed. Getting rid of the water actually helped the chain reaction because the water had been absorbing some of the neutrons. The fission rate began to increase.

The operators realized they were in trouble and began to shut down the reactor manually. However, the control rods moved into the core too slowly to make a difference. As water disappeared from the core, the core went "prompt critical." The chain reaction no longer had to wait for neutrons from the decaying fission fragments because prompt neutrons from the ^{235}U fissions were enough to sustain the chain reaction on their own. The reactor's fission rate skyrocketed, doubling many times each second. The fuel became white hot and melted its containers. Various chemical explosions blew open the containment vessel, and the graphite moderator caught fire.

The fire burned for 10 days before firefighters and pilots encased the wreckage in concrete. Many of these heroic people suffered fatal exposures to radiation. The burning core

released all its gaseous radioactive isotopes and many others into the atmosphere, forcing the evacuation of more than 100,000 people and leading to thousands or perhaps tens of thousands of premature cancer deaths in the millions of people exposed to the radioactive debris. The accident site remains dangerously radioactive and will need careful tending for centuries.

Although not a true reactor accident, the September 30, 1999, disaster in Tokai-mura, Japan, did involve a critical mass and a resulting chain reaction. At about 10:35 AM, employees of the Conversion Test Facility of JCO Co., Ltd. Tokai Works were pouring a solution of uranyl (uranium) nitrate into a precipitation tank. Destined for an experimental fast fission reactor, this uranium had been enriched to about 18.8% ^{235}U . Although the equipment and facilities were designed to prevent the assembly of a critical mass, the workers decided to save time by circumventing the safeguards and combining many small batches into one large one.

After they had poured six or seven batches of uranyl nitrate solution into the stainless steel tank through a sampling hole, the solution suddenly reached critical mass. There was about 16.6 kg of enriched uranium in the tank, and a sudden burst of radiation was released. The water temperature leapt upward, and the solution's resulting expansion dropped the mixture below critical mass. However, as heat flowed from the solution to the water-cooled jacket around the tank, the solution again approached critical mass. An episodic chain reaction continued in the tank for about 20 hours, until draining the cooling jacket of its neutron-reflecting water finally put an end to the critical mass.

The Fukushima Daiichi nuclear accident north of Tokyo, Japan (the fifth reactor accident), began on March 11, 2011, following the 9.0-magnitude Tohoku earthquake and the resulting tsunami. The Fukushima nuclear facility housed six separate boiling water reactors, but only reactors 1, 2, and 3 were operating at the time, and they shut down automatically after the earthquake. Even after stopping their chain reactions, however, those reactor cores continued to generate heat through the decay of their vast accumulations of radioactive nuclei, and they needed to be water-cooled to avoid overheating. Emergency generators began supplying power to maintain that cooling.

When the tsunami struck the Tohoku coast less than an hour after the earthquake, it inundated the nuclear facility, flooded its generators and switching equipment, and detached the entire area from the Japanese electric power grid. Reactor cooling soon ceased, and despite heroic efforts by emergency workers, it wasn't restored for days. By then, the reactors were damaged beyond repair, and the world's second worst nuclear accident was well underway.

Lacking adequate cooling, the cores of reactors 1, 2, and 3 gradually overheated and then melted. Explosions, pressure venting, and discharges of cooling water into the sea all contributed to an enormous release of radioactive nuclei into the environment. Even spent fuel in storage pools at the facility contributed to the overall accident because it also suffered overheating and damage when cooling was lost. The Fukushima Daiichi plant will never reopen and will remain a radioactive disaster area for years to come.

► Check Your Understanding #4: A Breath of Not-So-Fresh Air

One of the common fission fragments of ^{235}U is ^{131}I , a radioactive isotope of iodine. Iodine easily becomes a gas. Why shouldn't you sit in a room full of used uranium pellets?

Answer: The ^{131}I will enter the room air, and you will breathe it in. It will be incorporated into your body's iodine supply.

Why: The fission products of ^{235}U include radioactive isotopes of almost every element. Some of these isotopes are gaseous and can be absorbed by your lungs. Others can be carried away in water. To safely contain the fission fragments from used nuclear fuel, these fragments must be carefully separated and stored. Because the radioactive isotopes may continue to change from one element to another, such storage is particularly tricky.

Nuclear Fusion Reactors

Nuclear fission reactors use a relatively rare fuel—uranium. Although Earth's supply of uranium is vast, most of that uranium is distributed broadly throughout Earth's crust. There are only so many deposits of high-grade uranium ores that are easily turned into pure uranium or uranium compounds. Fission reactors also produce all sorts of radioactive fission fragments that must be disposed of safely. There is still no comprehensive plan for the safekeeping of spent reactor fuels. These must be kept away from any contact with people and animals virtually forever. No one really knows how to store such dangerous materials for hundreds of thousands of years.

An alternative to nuclear fission is nuclear fusion. By joining hydrogen nuclei together, heavier nuclei can be constructed. The amount of energy released in such processes is enormous. Recall, however, that fusion is much harder to initiate than fission because it requires that at least two nuclei be brought extremely close together. These nuclei are both positively charged, and they repel one another fiercely. To make them approach one another closely enough to stick, we must heat the nuclei to temperatures of more than 100 million degrees Celsius.

The sun combines four hydrogen nuclei (^1H) to form one nucleus of helium (^4He), a very complicated and difficult nuclear fusion reaction. For fusion to occur on Earth, it must be done between the heavy isotopes of hydrogen: deuterium and tritium. These are the isotopes used in thermonuclear weapons. If a mixture of deuterium and tritium are mixed together and heated to about 100 million degrees Celsius, their nuclei will begin to fuse and release energy. The deuterium and tritium become helium and neutrons.

In contrast to fission reactions, there are no radioactive fragments produced in fusion. Tritium itself is radioactive, but it can easily be reprocessed into fuel and retained within the reactor system. The dangerous neutrons can be caught in a blanket of lithium metal, which then breaks into helium and tritium. It's convenient that new tritium is created by the reaction because tritium isn't naturally occurring and must be made using nuclear reactions. Thus, fusion holds up the promise of producing relatively little radioactive waste. If the neutrons that are released by fusion events are trapped by nuclei that don't become radioactive, then there will be no radioactive contamination of the fusion reactor either. This is easier said than done, but it's better than in a fission reactor.

Unfortunately, heating deuterium and tritium and holding them together long enough for fusion to occur isn't easy. There are two main techniques that are being tried: inertial confinement fusion and magnetic confinement fusion.

Inertial confinement fusion uses intense pulses of laser light to heat and compress a tiny sphere containing deuterium and tritium (Fig. 15.2.5). The pulses of light last only a few trillionths of a second, but in that brief moment they vaporize and superheat the surface of the sphere. The surface explodes outward, pushing off the inner portions of the sphere. The sphere's core experiences huge inward forces as a result, and it implodes. As the core is compressed, its temperature rises to that needed to initiate fusion. In effect, it becomes a tiny thermonuclear bomb, with the laser pulses providing the starting heat.

To date, inertial confinement fusion experiments have observed fusion in a small fraction of the deuterium and tritium nuclei. The technique is called inertial confinement fusion because there is nothing holding or confining the ball of fuel. The laser beams crush it while it's in free fall, and its own inertia keeps it in place while fusion takes place. Unfortunately,

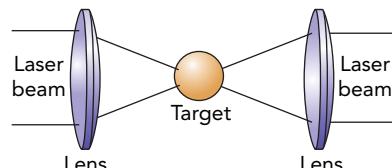


Fig. 15.2.5 In inertial confinement fusion experiments, several laser beams are focused onto a tiny sphere containing deuterium and tritium. These ultra-intense pulsed beams compress and heat the sphere so that fusion occurs.

the lasers and other technologies needed to carry out inertial confinement fusion are sufficiently complex and troublesome that this may never be viable as a source of energy. Nonetheless, these experiments provide important information on the behaviors of fusion materials at high temperatures and pressures.

The other technique being developed to control fusion is magnetic confinement. When you heat hydrogen atoms hot enough, they move so quickly and hit one another so hard that their electrons are knocked off. Instead of a gas of atoms, you have a gas of free, positively charged nuclei and negatively charged electrons—a plasma. A plasma differs from a normal gas because it's affected by magnetic fields.

Because of the Lorentz force (see Section 12.2), charged particles that are moving in a magnetic field tend to circle around magnetic field lines (Fig. 15.2.6). This cyclotron motion occurs, for example, in the magnetron of a microwave oven. If the magnetic field surrounding a charged particle is carefully shaped in just the right way, the charged particle will find itself trapped by the magnetic field. No matter what direction it heads, the charged particle will spiral around the magnetic field lines and will be unable to escape.

Magnetic confinement makes it possible to heat a plasma of deuterium and tritium to fantastic temperatures with electromagnetic waves. Since the heating is done relatively slowly, it's important to keep heat from leaving the plasma. Magnetic confinement prevents the plasma from touching the walls of the container, where it would quickly cool off.

One of the most promising magnetic confinement schemes is the tokamak. The main magnetic field of the tokamak runs around in a circle to form a magnetic doughnut, or toroid (the geometrical name for a doughnut-shaped object) (Fig. 15.2.7). The magnetic field is formed inside a doughnut-shaped chamber by running an electric current through coils that are wrapped around the chamber. Plasma nuclei inside the chamber travel in spirals around the magnetic field lines and don't touch the chamber walls. They are confined inside the chamber and race around the doughnut endlessly. The nuclei can then be heated to the extremely high temperatures they need to collide and fuse.

Magnetic confinement fusion reactors have observed considerable amounts of fusion. They can briefly achieve scientific break-even, the point at which fusion is releasing enough energy to keep the plasma hot all by itself. However, much more development is needed to meet and exceed practical break-even, the point at which the entire machine produces more energy than it needs to operate.

Check Your Understanding #5: Trying for Fusion in Your Basement

Why can't you initiate fusion by running an electric discharge through deuterium and tritium in a glass tube?

Answer: The gases will lose heat to the glass tube and will never reach the temperatures needed to start fusion.

Why: Very hot gases lose heat extremely easily. The hotter the gas, the faster heat flows out of it. Without something to keep the gas from touching the walls of the glass tube, it will never get close to the 100 million degrees Celsius needed for fusion.

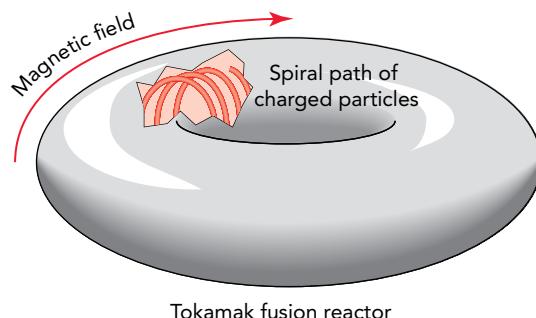


Fig. 15.2.7 A tokamak magnetic confinement fusion reactor consists of a doughnut-shaped chamber with a similarly shaped magnetic field inside it. Plasma particles moving inside the tokamak's chamber travel in spirals and circle around the field lines inside the chamber. Because the plasma doesn't touch the walls of the tokamak, it retains heat well enough to reach temperatures at which fusion can occur.

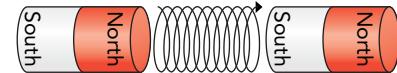
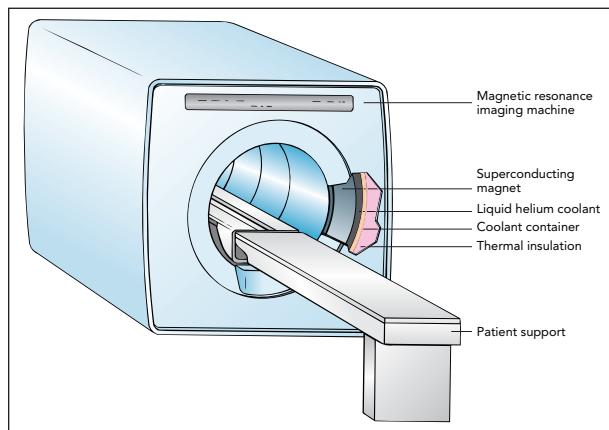


Fig. 15.2.6 When a charged particle moves in the magnetic field between two magnetic poles, it experiences the Lorentz force and undergoes cyclotron motion. It travels in a spiral path around the magnetic field lines connecting them. The particle is confined to a particular region of space.

SECTION 15.3

Medical Imaging and Radiation



Some of the most important recent advances in health care have occurred at the border between medicine and physics. As scientists have refined their understanding of atomic and molecular structure and learned to control various forms of radiation, they have invented tools that are enormously valuable for diagnosing and treating illness and injury. The developments continue, with new applications of physics appearing in clinical settings almost every time you turn around. In this section, we'll examine two of the most significant examples of medical physics: the imaging techniques that are used to detect problems and the radiation therapies that are used to treat them.

Questions to Think About: What's different about bones and tissue that makes bones appear light in an X-ray image while tissue appears dark? How can a CT scan or an MRI image show

a cross-sectional view of a living person without touching that person? If X-rays are a form of electromagnetic radiation, why don't opaque materials absorb them? Why does a CT scan show primarily bone, while an MRI image shows primarily tissue? The subatomic particles used in radiation therapy often have enormous energies. From where do these energies come?

Experiments to Do: One triumph of medical imaging is its ability to locate objects inside a person without actually entering the body. This is often done using X-rays to look through the person from several different angles. From each vantage point, the imaging machine determines which objects are to the left or right of one another, but it can't tell how far away those objects are. Nonetheless, by piecing together information from many different observations, the imaging machine can locate each object exactly.

You can experiment with this process by sprinkling a number of coins onto the surface of a small table, without looking to see where they come to rest. Close your eyes, and bring your head to the height of the table's surface. Open only one eye, and take a brief look at the coins. If the lighting is bright and uniform, and you don't move your head, you'll have trouble telling how far the coins are from your eye. Although you'll know something about where each coin is, you won't know enough to locate one exactly on the table's surface. Now move your head to a new position around the table, and take a second brief look. Again, you won't be able to tell how far away the coins are, but you'll learn more about their relative positions. How many such views will it take for you to learn exactly where the coins are? How does the presence of many opaque coins complicate the problem? Why does opening both eyes at once make it much easier for you to locate the coins?

X-Rays

Since their discovery in 1895, X-rays have played an important role in medical treatment. Their usefulness was obvious from the very evening they were discovered. It was November 8, and German physicist Wilhelm Conrad Roentgen (1845–1923) was experimenting with an electric discharge in a vacuum tube. He had covered the entire tube in black cardboard and was working in a darkened room. Some distance from the tube, a phosphored screen began to glow. Some kind of radiation was being released by the tube, passing through the cardboard and the air, and causing the screen to fluoresce. Roentgen put various objects in the way of the radiation, but they didn't block the flow. Finally, he put his hand in front of the screen and saw a shadowed image of his bones. He had discovered X-rays and their most famous application at the same time.

The first clinical use of X-rays was on January 13, 1896, when two British doctors used them to find a needle in a woman's hand. In no time, X-ray systems became common in hospitals as a marvelous new technique for diagnosis. However, this imaging capability was not without its side effects. Although the exposure itself was painless, overexposure to X-rays caused deep burns and wounds that took some time to appear. Evidently the X-rays were doing something much more subtle to the tissue than simply heating it.

X-rays are a form of electromagnetic radiation, as are radio waves, microwaves, and light. These different forms of electromagnetic radiation are distinguished from one another by their frequencies and wavelengths—while radio waves have low frequencies and long wavelengths, X-rays have extremely high frequencies and short wavelengths. They’re also distinguished by their photon energies. Because of its low frequency, a radio wave photon carries little energy. A medium-frequency photon of blue or ultraviolet light carries enough energy to rearrange one bond in a molecule. But a high-frequency X-ray photon carries so much energy that it can break many bonds and rip molecules apart.

In a microwave oven, the microwave photons work together to heat and cook food. The amount of energy in each microwave photon is unimportant because they don’t act alone. In radiation therapy, however, the X-ray photons are independent. Each one carries enough energy to damage any molecule that absorbs it. That’s why X-ray burns involve little heat and appear long after the exposure—the molecular damage caused by X-rays takes time to kill cells.

Check Your Understanding #1: Forms of Radiation

Which is more closely related to X-rays: the beam of electrons traveling through a microwave oven’s magnetron tube or the infrared light from the hot filament of a toaster?

Answer: The infrared light from the hot filament is more closely related to X-rays.

Why: Infrared light and X-rays are both forms of electromagnetic radiation. All that distinguishes them is their frequencies and wavelengths, and the energy of their photons. The beam of electrons in a magnetron tube is also a form of radiation, but it involves particles of matter, not electromagnetic waves.

Making X-Rays

Medical X-ray sources work by crashing fast-moving electrons into heavy atoms. These collisions create X-rays via two different physical mechanisms: bremsstrahlung and X-ray fluorescence.

Bremsstrahlung occurs whenever a charged particle accelerates. This process is nothing really new, since we know that radio waves are emitted when a charged particle accelerates on an antenna. In a radio antenna, however, the electrons accelerate slowly and emit low-energy photons. Bremsstrahlung usually refers to cases in which a charged particle accelerates extremely rapidly and emits a very high-energy photon. In X-ray tube bremsstrahlung, a fast-moving electron arcs around a massive nucleus and accelerates so abruptly that it emits an X-ray photon (Fig. 15.3.1). This photon carries away a substantial fraction of the electron’s kinetic energy. The closer the electron comes to the nucleus, the more it accelerates and the more energy it gives to the X-ray photon. However, the electron is more likely to miss the nucleus by a large distance than to almost hit it, so bremsstrahlung is more likely to produce a lower-energy X-ray photon than a higher-energy one.

In **X-ray fluorescence**, the fast-moving electron collides with an inner electron in a heavy atom and knocks that electron completely out of the atom (Fig. 15.3.2). The collision leaves the atom as a positive ion, with a vacant orbital close to its nucleus. An electron in that ion then undergoes a radiative transition, shifting from an outer orbital to this empty inner one and releasing an enormous amount of energy in the process. This energy emerges from the atom as an X-ray photon. Because this photon has an energy that’s determined by the ion’s orbital structure, it’s called a **characteristic X-ray**.

To discuss the energies carried by X-ray photons, we’ll use the energy unit we encountered in Section 13.3—the electron volt (eV). Photons of visible light carry energies of between 1.6 eV (red light) and 3.0 eV (violet light). Because the ultraviolet photons in sunlight have energies of up to 7 eV, they are able to break chemical bonds and cause sunburns. X-ray photons have much larger energies than even ultraviolet photons.

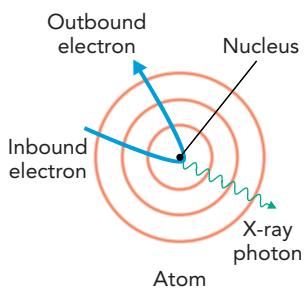


Fig. 15.3.1 When a fast-moving electron arcs around a massive nucleus, it accelerates rapidly. This sudden acceleration creates a bremsstrahlung X-ray photon, which carries off some of the electron’s energy.

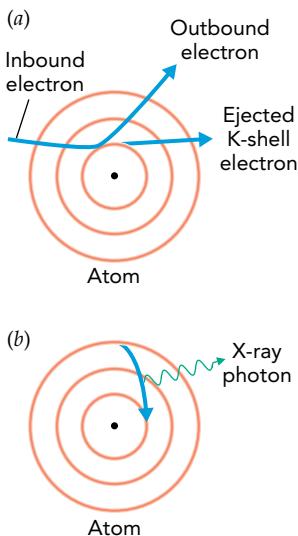
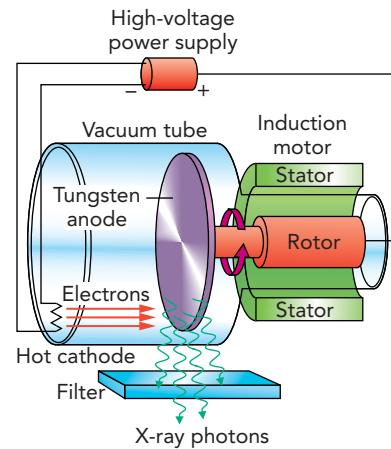


Fig. 15.3.2 (a) When a fast-moving electron collides with an electron in one of the inner orbitals of a heavy atom, it can knock that electron out of the atom. (b) An electron from one of the atom's outer orbitals soon drops into the empty orbital in a radiative transition that creates a characteristic X-ray.

Fig. 15.3.3 In a medical X-ray machine, electrons from a hot filament accelerate toward a positively charged metal disk. They emit X-rays when they collide with the disk's atoms. A motor spins the disk to keep it from melting. The filter absorbs useless low-energy X-rays.



In a typical medical X-ray tube, electrons are emitted by a hot cathode and accelerate through vacuum toward a positively charged metal anode (Fig. 15.3.3). The anode is a tungsten or molybdenum disk, spinning rapidly to keep it from melting. The energy of the electrons as they hit the anode is determined by the voltage difference across the tube. In a medical X-ray machine, that voltage difference is typically about 87,000 V, so each electron has about 87,000 eV of energy. Since an electron gives a good fraction of its energy to the X-ray photon it produces, the photons leaving the tube can carry up to 87,000 eV of energy. No wonder X-rays can damage tissue!

When the electrons collide with a target of heavy atoms, they emit both bremsstrahlung and characteristic X-rays (Fig. 15.3.4). The characteristic X-rays have specific energies, so they appear as peaks in the overall X-ray spectrum. The bremsstrahlung X-rays have different energies but are most intense at lower energies. Because lower-energy X-ray photons injure skin and aren't useful for imaging or radiation therapy, medical X-ray machines use absorbing materials, such as aluminum, to filter them out.

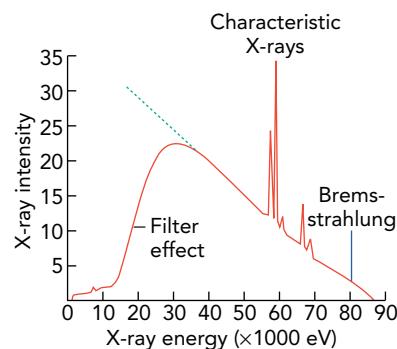
Check Your Understanding #2: The Origins of Synchrotron Radiation

The giant particle accelerators used in high-energy physics are often built as rings so that they can use the same electrically charged particles over and over again. As these particles travel in circles around the rings, they emit X-rays. Explain.

Answer: Because a particle traveling in a circle is accelerating, it emits electromagnetic waves. In this case, those waves are X-rays.

Why: A rapidly accelerating charged particle will emit an X-ray, whether it's accelerating around the nucleus of a heavy atom or around the ring of a particle accelerator. In an accelerator, these X-rays are called synchrotron radiation. Synchrotron radiation is useful in research and industry and is often deliberately enhanced by adding special magnets to the ring of the accelerator.

Fig. 15.3.4 When electrons with 87,000 eV of energy collide with tungsten metal, they emit X-rays via bremsstrahlung and X-ray fluorescence. Although the bremsstrahlung X-rays have a broad range of energies, an absorbing filter blocks the low-energy ones. X-ray fluorescence produces characteristic X-rays with specific energies.



Using X-Rays for Imaging

X-rays have two important uses in medicine: imaging and radiation therapy. In X-ray imaging, X-rays are sent through a patient's body to a sheet of film or an X-ray detector. Some of the X-rays manage to pass through tissue, but most of them are blocked by bone. The patient's bones form shadow images on the film behind them. In X-ray radiation therapy, the X-rays are again sent through a patient's body, but now their lethal interaction with diseased tissue is what's important.

X-ray photons interact with tissue and bone through four major processes: *elastic scattering*, the *photoelectric effect*, *Compton scattering*, and *electron-positron pair production*. **Elastic scattering** is already familiar to us as the cause of the blue sky; an atom acts as an antenna for the passing electromagnetic wave, absorbing and reemitting it without keeping any of its energy (Fig. 15.3.5). Because this process has almost no effect on the atom, elastic scattering isn't important in radiation therapy. However, it's a nuisance in X-ray imaging because it produces a hazy background; some of the X-rays passing through a patient bounce around like pinballs and arrive at the film from odd angles. To eliminate these bouncing X-ray photons, X-ray machines use filters to block X-rays that don't approach the film from the direction of the X-ray source.

The **photoelectric effect** is what makes X-ray imaging possible. In this effect, a passing photon induces a radiative transition in an atom; one of the atom's electrons absorbs the photon and is tossed completely out of the atom (Fig. 15.3.6). If the atom were using the X-ray photon to shift an electron from one orbital to another, that photon would have to have just the right amount of energy. However, because a free electron can have any amount of energy, the atom can absorb any X-ray photon that has enough energy to eject one of its electrons. Part of the photon's energy is used to remove the electron from the atom, and the rest is given to the emitted electron as kinetic energy.

The likelihood of such a *photoemission* event decreases as the ejected electron's energy increases. This decreasing likelihood makes it difficult for a small atom to absorb an X-ray photon. All its electrons are relatively weakly bound, and the X-ray photon would give the ejected electron a large kinetic energy. Rather than emitting a high-energy electron, a small atom usually just ignores the passing X-ray photon.

In contrast, some of the electrons in a large atom are quite tightly bound and require most of the X-ray photon's energy to remove them. These electrons would depart with relatively little kinetic energy. Because the photoemission process is most likely when low-energy electrons are produced, a large atom is likely to absorb a passing X-ray. Thus, whereas the small atoms found in tissue (carbon, hydrogen, oxygen, and nitrogen) rarely absorb medical X-rays, the large atoms found in bone (calcium and phosphorus) absorb X-rays frequently. That's why bones cast clear shadows onto X-ray film. Tissue shadows are also visible, but they're less obvious.

Although one shadow image of a patient's insides may help to diagnose a broken bone, more subtle problems may not be visible in a single X-ray image. For a better picture of what's going on inside the patient, the radiologist needs to see shadows from several different angles. Better yet, the radiologist can turn to a *computed tomography (CT) scanner*. This computerized device automatically forms X-ray shadow images from hundreds of different angles and positions and produces a detailed three-dimensional X-ray map of the patient's body.

The CT scanner works one "slice" of the patient's body at a time. It sends X-rays through this narrow slice from every possible angle, including the two shown in Fig. 15.3.7, and determines where the bones and tissues are in that slice (Fig. 15.3.8). The scanner then shifts the patient's body to work on the next slice.

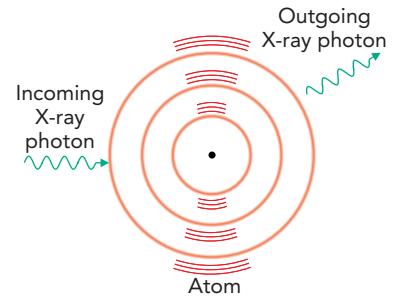


Fig. 15.3.5 When an X-ray photon scatters elastically from an atom, the whole atom acts as an antenna. The passing photon jiggles all the charges in the atom, and these charges absorb the photon and reemit it in a new direction.

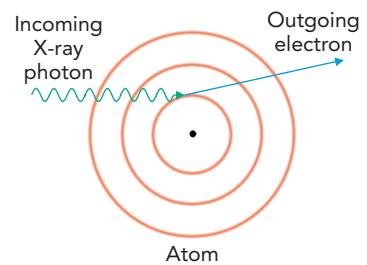


Fig. 15.3.6 In the photoelectric effect, an absorbed photon ejects an electron from an atom. Part of the photon's energy is used to remove the electron from the atom, and the rest becomes kinetic energy in the electron.

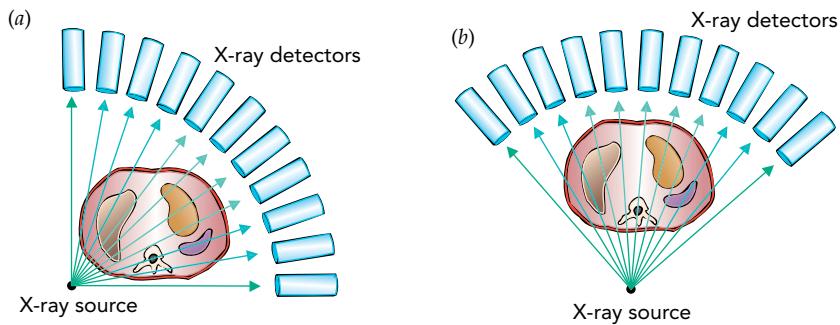
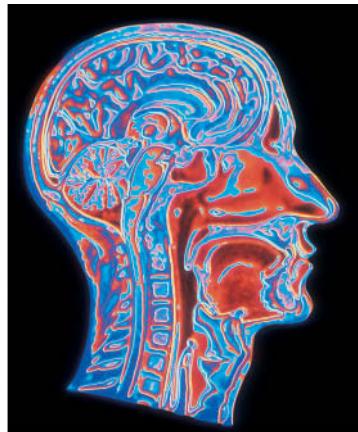


Fig. 15.3.7 A computed tomography (CT) scan image is formed by analyzing X-ray shadow images taken from many different angles and positions. An X-ray source and an array of electronic X-ray detectors form a ring that rotates around the patient as the patient slowly moves through the ring.



© Steve Allen/Brand X Pictures/Getty Images, Inc.



© Alfred Pasieka/Photo Researchers, Inc.

Fig. 15.3.8 (a) The CT scanner on the left uses X-rays to image layer after layer of the patient's body. With the help of a computer, it produces a three-dimensional (3D) map of heavy elements in the patient. (b) Part of that map, showing the patient's head.



Check Your Understanding #3: Aluminum X-Ray Windows

Aluminum atoms are much smaller than calcium atoms. Although aluminum metal blocks visible light, it's relatively transparent to high-energy X-rays. Explain.

Answer: The electrons in an aluminum atom are so weakly bound that they are unlikely to absorb high-energy X-ray photons through the photoelectric effect.

Why: Like the small atoms in biological tissue, aluminum atoms rarely use the photoelectric effect to absorb high-energy X-rays. This result makes it possible to use thin films of aluminum as windows and filters for X-ray sources.

Using X-Rays for Therapy

Radiation therapy also uses X-rays, but not the ones used for medical imaging. Even though tissue absorbs fewer imaging photons than bone, most imaging photons are absorbed before they can pass through thick tissue. For example, only about 10% of the imaging photons make it through a patient's leg even when they miss the bone. That percentage is good enough for making an image, but it wouldn't do for radiation therapy because most imaging X-rays would be absorbed long before they reached a deep-seated tumor. Instead of killing the tumor, intense exposure to these X-rays would kill tissue near the patient's skin.

To attack malignant tissue deep beneath the skin, radiation therapy uses extremely high-energy photons. At photon energies near 1,000,000 eV, the photoelectric effect becomes rare in tissue and bone, and the photons are much more likely to reach the tumor. Photons still deposit lethal energy in the tissue and tumor, but they do this through a new effect—Compton scattering.

Compton scattering occurs when an X-ray photon collides with a single electron so that the two particles bounce off one another (Fig. 15.3.9). The X-ray photon knocks the electron right out of the atom. This process is different from the photoelectric effect because Compton scattering doesn't involve the atom as a whole and the photon is scattered (bounced) rather than absorbed. The physics behind this effect resembles that of two billiard balls colliding, although it's complicated by the theory of relativity. The fact that it occurs at all is proof that a photon carries both energy and momentum and that these quantities are conserved when a particle of light collides with a particle of matter.

Compton scattering is crucial to radiation therapy. When a patient is exposed to 1,000,000-eV photons, most of the photons pass right through, but a small fraction undergo Compton scattering and leave some of their energy behind. This energy kills cells and can be used to destroy a tumor. By approaching a tumor from many different angles through the patient's body, the treatment can minimize the injury to healthy tissue around the tumor while giving the tumor itself a fatal dose of radiation.

Compton scattering isn't the only effect that occurs when high-energy photons encounter matter. X-rays with slightly more than 1,022,000 eV can do something remarkable when they pass through an atom; they can cause **electron-positron pair production**. A **positron** is the **antimatter** equivalent of an electron. Our universe is symmetrical in many ways, and one of its nearly perfect symmetries is the existence of antimatter. Almost every particle in nature has an antiparticle with the same mass but opposite characteristics. A positron, or antielectron, has the same mass as an electron, but it's positively charged. There are also antiprotons and antineutrons.

Antimatter doesn't occur naturally on Earth, but it can be created in high-energy collisions. When an energetic photon collides with the electric field of an atom, the photon can become an electron and a positron. In the previous section, we discussed matter becoming energy; pair production is an example of energy becoming matter. It takes about 511,000 eV of energy to form an electron or a positron, so the photon must have at least 1,022,000 eV to create one of each. Any extra energy goes into kinetic energy in the two particles.

The positron doesn't last long in a patient. It soon collides with an electron and the two annihilate one another—the electron and positron disappear, and their mass becomes energy. They turn into photons with a total of at least 1,022,000 eV. Thus energy became matter briefly and then turned back into energy. This exotic process is present in high-energy radiation therapy and becomes quite significant at photon energies above about 10,000,000 eV. Not surprisingly, it also helps to kill tumors.

Check Your Understanding #4: More of a Good Thing

About how much energy would a photon need to create a proton-antiproton pair?

Answer: It would need about 2,000,000,000 eV.

Why: A proton has about 2000 times the mass of an electron, so producing a proton-antiproton pair should take about 2000 times the 1,000,000-eV energy required to produce an electron-positron pair.

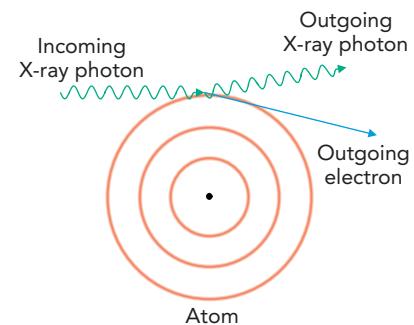


Fig. 15.3.9 In Compton scattering, an X-ray photon collides with a single electron and the two bounce off one another. The electron is knocked out of the atom.

Gamma Rays

Producing very high-energy photons isn't quite as easy as producing those used in X-ray imaging. In principle, a power supply could create a huge voltage difference through an X-ray tube so that very high-energy electrons would crash into metal atoms and produce

very high-energy photons. However, million-volt power supplies are complicated and dangerous, so other schemes are used instead.

One of the easiest ways to obtain very high-energy photons is through the decay of radioactive isotopes. The isotope most commonly used in radiation therapy is cobalt 60 (^{60}Co). The nucleus of ^{60}Co has too many neutrons and, like many neutron-rich nuclei, ^{60}Co undergoes **beta decay**; that is, one of its neutrons breaks up into a proton, an electron, and a **neutrino** (or more precisely, an *antineutrino*). Beginning with that beta decay, ^{60}Co undergoes a series of transformations that produce two high-energy photons: one with 1,170,000 eV and one with 1,330,000 eV. These photons penetrate tissue well and are quite effective at killing tumors.

Although the process by which ^{60}Co produces those two high-energy photons is complicated, beta decay itself shows that protons, electrons, and neutrons are not immutable and that there are other subatomic particles in our universe. Neutrons that are by themselves or in nuclei with too many neutrons are radioactive and experience beta decay. When that beta decay process occurs in a ^{60}Co nucleus, the negatively charged electron and neutral neutrino quickly escape from the nucleus but the newly formed proton remains. The nucleus thus becomes nickel 60 (^{60}Ni).

The **neutrino** is a subatomic particle with no charge and little mass. Neutrinos aren't found in normal atoms. Although important in nuclear and particle physics, neutrinos are difficult to observe directly because they travel near the speed of light and hardly ever collide with anything. Without charge, they don't participate in electromagnetic forces and, unlike the electrically neutral neutron, they don't experience the nuclear force. They experience only gravity and the **weak force**, the last of the four **fundamental forces** known to exist in our universe. (The other three fundamental forces are the gravitational force; the electromagnetic force; and the **strong force**, which is a more complete version of the nuclear force that we discussed in Section 15.1.) Because it's weak and occurs only between particles that are very close together, the weak force rarely makes itself apparent. One of the few occasions on which it plays an important role is in beta decay.

With almost no way to push or pull on another particle, a neutrino can easily pass right through the entire Earth. Neutrinos are detected occasionally, but only with the help of enormous detectors. That's why physicists first showed that neutrinos are emitted from decaying neutrons by measuring energy and momentum before and after the decay. The proton and electron produced by the decay don't have the same total energy and momentum as the neutron had before the decay. Something must have carried away the missing energy and momentum, and that something is the neutrino.

Once ^{60}Co has turned into ^{60}Ni , the decay isn't quite over. The ^{60}Ni nucleus that forms still has extra energy in it. Nuclei are complicated quantum physical systems, just as atoms are, and they have excited states, too. The ^{60}Ni nucleus is in an excited state, and it must undergo two radiative transitions before it reaches the ground state. These radiative transitions produce very high-energy photons or **gamma rays** that are characteristic of the ^{60}Ni nucleus: one with 1,170,000 eV of energy and the other with 1,330,000 eV. The gamma rays are what make ^{60}Co radiation therapy possible.



Check Your Understanding #5: A Visit from the Snake Oil Salesman

If someone offered to sell you a bottle of neutrinos, you'd be foolish to buy it. What's wrong with the idea of a bottle of neutrinos?

Answer: The bottle couldn't confine neutrinos because it barely interacts with them.

Why: Because neutrinos experience only gravity and the weak force, the bottle couldn't trap them for long. Actually, you're being bathed in neutrinos from the sun all the time without even noticing it. They're as common as dirt and you can't do anything with them anyway.

Particle Accelerators

Electromagnetic radiation isn't the only form of radiation used to treat patients. Energetic particles such as electrons and protons are also used. Like tiny billiard balls, these fast-moving particles collide with the atoms inside tumors and knock them apart. As usual, this atomic and molecular damage tends to kill cells and destroy tumors.

However, obtaining extremely energetic subatomic particles isn't easy. High-voltage power supplies can be used to accelerate an electron or proton to about 500,000 eV, but that isn't enough. When a charged particle enters tissue, it experiences strong electric forces and is easily deflected from its path. To make sure that it travels straight and true, all the way to a tumor, the particle must have an enormous energy. Giving each charged particle the millions or even billions of electron volts it needs for radiation therapy takes a particle accelerator.

Particle accelerators use resonant cavities, the microwave-frequency tank circuits we encountered in Section 12.2. Each of these metal chambers acts simultaneously as a capacitor and an inductor and thus has a natural resonance for sloshing charge. In the resonant cavities of a particle accelerator, this sloshing charge creates huge electric fields that change with time. Those electric fields push charged particles through space until they reach incredible energies.

One important type of particle accelerator is the *linear accelerator*. In this device, the electric fields in a series of resonant cavities push charged particles forward in a straight line (Fig. 15.3.10). Each of these cavities has charge sloshing back and forth rhythmically on its wall. When a small packet of charged particles enters the first cavity through a hole, it's suddenly pushed forward by the strong electric field inside that cavity (Fig. 15.3.10a). The packet accelerates forward and leaves the first resonant cavity with more kinetic energy than it had when it arrived—the electric field in that cavity has done work on the packet.

If the fields in the cavities were constant, the electric field in the second cavity would slow the packet down. In Fig. 15.3.10a, you can see that the electric field in the second cavity points in the wrong direction. By the time the packet reaches the second cavity, the charge sloshing in its walls has reversed and so has the electric field (Fig. 15.3.10b). The packet is again pushed forward, and it emerges from the second cavity with still more kinetic energy.

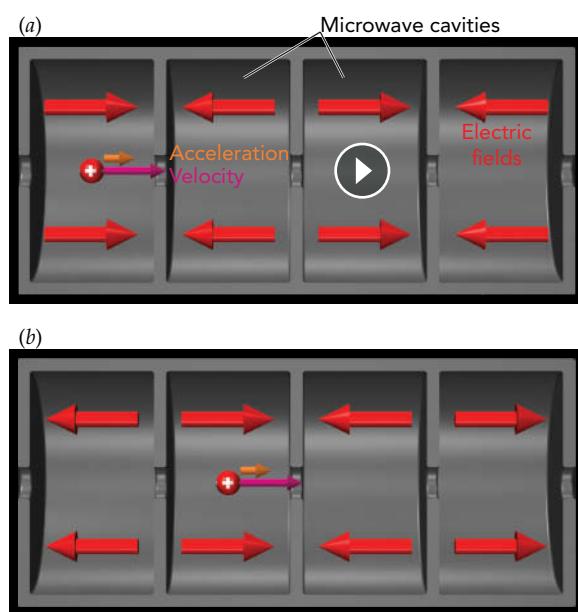


Fig. 15.3.10 In a linear accelerator, moving charged particles are pushed forward by electric fields that change with time. (a) While the positive charge is moving toward the right through the first of a series of microwave cavities, the electric field there pushes it toward the right. (b) By the time the charge has entered the second cavity, however, the electric fields have reversed and the electric field there pushes it toward the right again.

© Doug Martin/Photo Researchers, Inc.

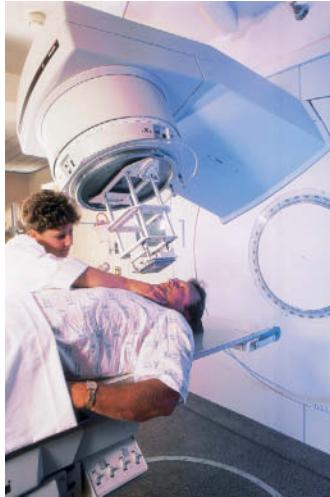


Fig. 15.3.11 This radiation therapy unit uses a linear accelerator to produce extremely high-energy subatomic particles. These particles penetrate deep inside the patient to destroy a cancerous tumor. The linear accelerator itself is hidden from view in the room behind this one. Its beam is steered by magnets in the rotatable arm toward a particular spot in the patient. The arm moves periodically during treatment so that its beam intersects the tumor from many different directions and causes less damage to healthy tissue nearby. That multidirectional strategy is also widely used in X-ray and gamma-ray therapies.

Each resonant cavity in this series adds energy to the packet, so that a long string of cavities can give each of the packet's charged particles millions or even billions of electron volts. This energy comes from microwave generators that cause charge to slosh in the accelerator's resonant cavities. The linear accelerator then has only to inject charged particles into the first cavity, using equipment resembling the insides of a television picture tube, and those charged particles will come flying out of the last cavity with incredible energies (Fig. 15.3.11).

This acceleration technique, however, has a few complications. Most important, each cavity must reverse its electric field at just the right moment to keep the packet accelerating forward. For simplicity of operation, all the cavities have the same resonant frequency and reverse their electric fields simultaneously. Since the packet spends the same amount of time in each cavity and since it speeds up as it goes from one cavity to the next, each cavity must be longer than the previous one.

As the packet approaches the speed of light, something strange happens. The packet's energy continues to increase as it goes through the cavities, but its speed stops increasing very much. This effect is a consequence of special relativity, the rules governing motion at speeds comparable to the speed of light. As we saw in Section 4.2, the simple relationship between kinetic energy and speed given in Eq. 2.2.1 isn't valid for objects moving at almost the speed of light; we must use Eq. 4.2.4 instead. As a further consequence of relativity, the packet can approach the speed of light but can't actually reach it. Although each charged particle's kinetic energy can become extraordinarily large, its speed is limited by the speed of light.

Because the packet's speed stops increasing significantly after it has gone through the first few cavities of the linear accelerator, the lengths of the remaining cavities can be constant. Only the first few cavities have to be specially designed to account for the packet's increasing speed inside them. The charged particles emerge from the accelerator traveling at almost the speed of light. They pass through a thin metal window that keeps air out of the accelerator and enter the patient's body. They have so much energy that they can penetrate deep into tissue before coming to a stop.



Check Your Understanding #6: Particle Recycling

Many research accelerators send each packet of electrons through the same series of resonant cavities several times. After the packet leaves the last cavity, magnets steer it around in a circle and send it back through the cavities again. With each pass through the cavities, the packet acquires more energy, so how can it possibly stay synchronized with the reversing electric fields in the cavities?

Answer: The packet is traveling at almost the speed of light, so its speed barely changes as its energy increases.

Why: If the packet were speeding up with each trip through the cavities, it wouldn't stay synchronized with the reversing electric fields inside them. However, the packet's speed is so nearly constant once it nears the speed of light that it can travel through the cavities over and over again without any problem.

Magnetic Resonance Imaging

Although X-rays do an excellent job of imaging bones, they aren't as good for imaging tissue. A better technique for studying tissue is *magnetic resonance imaging* (MRI). This technique locates hydrogen atoms by interacting with their magnetic nuclei. Since hydrogen atoms are common in both water and organic molecules, finding hydrogen atoms is a good way to study biological tissue.

The nucleus of an ordinary hydrogen atom, ${}^1\text{H}$, is a proton. Protons, like electrons, have two possible internal quantum states, usually called spin-up and spin-down. Calling it *spin* is appropriate because spin-up and spin-down protons have equal but oppositely directed angular momentum. When electric charge and rotation are both present, it's not surprising that magnetism is, too; electric currents are magnetic, after all. Sure enough, protons have magnetic dipoles—equal north and south poles at a distance from one another.

A spin-up proton acts as though it has its north pole on top, while a spin-down proton acts as though it has its south pole on top.

When a proton is immersed in a magnetic field, it tends to align its magnetic dipole with that field. Doing so minimizes its magnetic potential energy. But although protons would align perfectly with the field at absolute zero, they are less successful near room temperature. Thermal energy agitates the protons so that, even in a strong, upward-pointing magnetic field, spin-up protons only slightly outnumber spin-down protons.

In that upward-pointing magnetic field, each proton has two possible quantum states: alignment with the field (spin-up) or anti-alignment (spin-down). Because alignment reduces the proton's magnetic potential energy, it's the ground state—the lower energy of the two possible states. The anti-aligned state is the excited state.

With its two possible states, ground and excited, a proton in a magnetic field can exhibit many of the behaviors we explored when looking at atoms in Section 13.2. Most important, the proton can experience radiative transitions between its two states. A ground-state proton can *absorb* a photon while making a radiative transition to its excited state, and an excited-state proton can *emit* a photon while transiting to its ground state.

In Section 13.2, we saw that a given atom can absorb or emit only certain photons, photons carrying exactly the right amount of energy to shift the atom from one quantum state to another. For example, neon signs are red because neon atoms have states that are separated in energy by the energy of red photons. Similarly, a proton in a magnetic field can absorb or emit only certain photons, photons carrying exactly the right amount of energy to shift the proton from one quantum state to the other.

However, unlike a neon atom, which always interacts with red photons, a proton in a magnetic field interacts with photons that vary in “color” according to the strength of the magnetic field. That's because the energy separating the proton's two states is proportional to the magnetic field in which it resides. As a result, the photon energy needed to cause radiative transitions between the proton's two states is also proportional to the magnetic field. If the field changes, so does the photon energy.

When a patient enters the strong magnetic field of an MRI machine (Figs. 15.3.12 and 15.3.13), the protons in the patient's body respond to the field and a small excess of aligned protons develops. Only these excess aligned protons matter to the MRI machine because effects due to the remaining protons, which are equally aligned and anti-aligned, cancel completely. The excess aligned protons are in their ground state, and they are what the MRI machine studies.

The MRI machine interacts with these ground-state protons using radio wave photons, photons with energies equal to the energy difference between their ground and excited states. The protons can absorb and subsequently emit those radio wave photons, and they can also exhibit a variety of fascinating and useful quantum interference effects.

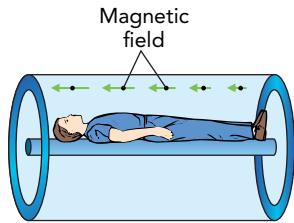


Fig. 15.3.12 A magnetic resonance imaging (MRI) machine places the patient in a strong magnetic field. This field varies spatially, so that protons at different places in the patient's body experience different fields and absorb different radio wave photons.

© Lefkowitz/Photographer's Choice/Getty Images, Inc.



© Pasieka/Science Photo Library/Getty Images, Inc.



Fig. 15.3.13 (a) The patient is entering the intense magnetic field of an MRI system. Using the interactions between electromagnetic waves and protons that take place in magnetic fields, MRI produces a three-dimensional map of hydrogen atoms in the patient's body. (b) A portion of this map, showing the patient's head.

If the protons in the patient's body were all experiencing exactly the same magnetic field, they would all interact with the same radio wave photons—but the protons don't all experience the same field. The MRI machine introduces a slight spatial variation to its magnetic field. Because the magnetic field is different for different protons, only some of them can interact with radio wave photons of a particular energy. This selective interaction is how the MRI imager locates protons within a patient.

In its simplest form, an MRI machine applies a spatially varying magnetic field to the patient's body. It then sends various radio waves through the patient and looks for the radio wave photons to interact with protons. Since only a proton that is experiencing the right magnetic field can interact with a particular radio wave photon, the MRI machine can determine where each proton is by the photons with which it interacts. By changing the spatial variations in the magnetic field and adjusting the energies of the radio wave photons, the MRI machine gradually locates the protons in the patient's body. It builds a detailed three-dimensional map of the hydrogen atoms. A computer manages this map and can display cross-sectional images of the patient from any angle or position.



Check Your Understanding #7: Major Magnets

Some of the most modern MRI machines use extremely strong magnetic fields. As a machine's magnetic field gets stronger, what must happen to the radio waves that are used to interact with protons in a patient?

Answer: Each radio wave photon must carry more energy (the radio waves must have higher frequencies).

Why: The stronger the magnetic field in which a proton is immersed, the more energy separates the proton's two states. The MRI machine must use higher-frequency, more energetic radio waves to cause radiative transitions between those two states. There are many advantages to using extremely strong magnetic fields, including lower noise and better spatial resolution in the image. However, the magnetic fields of these advanced MRI machines are so strong that they can erase credit cards from across the room and rip steel objects right out of your pockets.

Epilogue for Chapter 15

In this chapter we examined some of the applications of modern physics. In Nuclear Weapons, we examined nuclear fission to see how nuclear chain reactions can be used to release electrostatic potential energy stored in the large nuclei of uranium and plutonium atoms. We also studied nuclear fusion and found that when small nuclei of hydrogen bind together, they release a potential energy associated with the nuclear force. We learned about radioactive isotopes and fallout.

In Nuclear Reactors, we studied the techniques used to harness nuclear energy for less destructive purposes. We saw that natural or slightly enriched uranium can support a chain reaction if the fission neutrons are slowed in a moderator. We reviewed reactor and safety issues and looked at the world's most significant nuclear accidents.

In Medical Imaging and Radiation, we learned how X-rays are produced and why those X-rays pass more easily through tissue than through bone. We saw how X-rays can be used to make images and explored the uses of gamma rays for radiation therapy. We looked at particle accelerators and concluded by examining magnetic resonance imaging.

Explanation: Radiation-Damaged Paper

The paper fades because the ultraviolet light photons in sunlight have enough energy to shift electrons out of the orbitals that bond the dye molecules together. These dye molecules then fall apart and leave the paper without color. In some cases, the photons of light completely remove the electrons from the dye molecules. The process is the same photoelectric effect that makes it possible to distinguish tissue from bone in X-ray imaging.

Chapter Summary and Important Laws and Equations

How Nuclear Weapons Work: A fission bomb releases nuclear energy through a chain reaction in the fissionable isotopes of uranium or plutonium. In a chain reaction, each fission induces, on average, at least one subsequent fission. Assembling a supercritical mass must be done rapidly so that the bomb doesn't blow itself apart before most of the nuclei have undergone fission.

A fusion bomb uses heat from a fission bomb to initiate fusion in the heavy isotopes of hydrogen: deuterium and tritium. Because tritium has a short half-life, it must be replaced frequently. In some fusion bombs, the tritium is created during the explosion by neutron collisions with lithium.

How Nuclear Reactors Work: A thermal fission reactor can sustain a controlled chain reaction in natural or slightly enriched uranium. By slowing its fission neutrons to thermal velocities in a moderator, the reactor is able to make those neutrons interact almost exclusively with the rare ^{235}U nuclei and leave the more common ^{238}U nuclei almost unaffected. The ^{235}U nuclei fission in a chain reaction and release heat that is used to power electric generators. The presence of some delayed fission neutrons makes it easy to control the chain reaction using neutron-absorbing control rods.

To use the ^{238}U nuclei, a fast fission or breeder reaction uses more highly enriched uranium and no moderator. Although the chain reaction initially proceeds primarily among the ^{235}U nuclei, the fast fission neutrons gradually transform the ^{238}U nuclei into ^{239}Pu nuclei. Like ^{235}U nuclei, those ^{239}Pu nuclei are fissionable and can participate in the chain reaction, so the fast fission reaction can ultimately extract energy from all the original uranium nuclei. ^{239}Pu is also chemically separable from uranium and can be used to manufacture nuclear weapons. Fast fission reactors therefore pose the danger of nuclear proliferation.

Although the goal of obtaining energy from the fusion of hydrogen nuclei on Earth remains elusive, gradual progress is being made. The main challenge is to contain and heat the hydrogen nuclei to the enormous temperatures needed to make them collide and fuse.

How Medical Imaging and Radiation Work: Because X-rays pass more easily through tissue than they do through bone, X-rays form shadow images of a patient's bones. These X-rays are produced when energetic electrons accelerate near metal nuclei and when they knock other electrons out of the metal atoms.

Radiation therapy is done with very high-energy X-rays and gamma rays because they pass more easily through tissue. These electromagnetic waves kill tumor cells by depositing energy in the atoms and molecules of those cells. Gamma rays are usually obtained from radioactive nuclei. Some radiation therapy is done with energetic particles that are given enormous energies by particle accelerators.

Magnetic resonance imaging uses the magnetic nature of hydrogen nuclei (protons) to locate hydrogen atoms in a patient's body. The patient is put in a strong magnetic field, and the protons tend to align with this field. The imaging machine then uses radio waves to reverse the alignments of those protons. By making the magnetic field vary slightly from place to place, the machine is able to locate the protons in the patient's body. A computer records and analyzes the results so that it can present cross-sectional images of the hydrogen atoms in a patient's body.

1. Exponential decay of radioactive nuclei and particles: The fraction of a large original population of identical radioactive systems remaining is equal to one-half raised to the power of the elapsed time divided by the half-life of those systems, or

$$\text{fraction remaining} = \left(\frac{1}{2}\right)^{\text{elapsed time/half-life}} \quad (15.1.1)$$

Appendix A

Vectors

Some physical quantities are simply numbers (e.g., 0, -1.5 , and π) or numbers with units (e.g., 10 s, 5 kg, and 1 m 3). These are known as **scalar quantities**, that is, quantities that have only magnitudes (amounts). Many other physical quantities, however, are *vectors*—they have both magnitudes and directions. In our three-dimensional world, vector quantities are quite common and useful, and they include position, velocity, acceleration, force, torque, momentum, and angular momentum. Because vector quantities include both magnitude and direction, they convey more information than scalar quantities and they're also a bit more difficult to understand.

Position is probably the easiest vector quantity to visualize: you specify an object's position by giving its position vector—its distance and direction from a reference point. For example, you can specify the library's position with respect to your home by giving both its distance from your home (say 3.162 km, or 1.965 miles) and its direction from your home (18.43° east of due north). That position vector is all the information someone would need to travel from your home to the library.

In illustrations, such as Fig. A.1, vector quantities are drawn as arrows. The length of each arrow indicates the vector's magnitude, while the direction of the arrow indicates which way the vector points. Suppose that you live in a city with major east-west and north-south streets spaced 1 km apart. Figure A.1 shows four aerial views of your city. The vector **A** in Fig. A.1a shows the position of the library with respect to your home. It begins at your home and ends at the library, thus indicating both the magnitude and direction of the library's position.

Let's look at another position vector. The vector **B** in Fig. A.1b begins at the library and ends at your friend's home. This position vector shows the position of your friend's home with respect to the library and is 2.828 km long and points 45° east of south. If you happen to be at the library, you can use this vector to find your friend's home.

But how can you go from your home to your friend's home? To make this trip, you must add two vectors: you first follow vector **A** from your home to the library and then follow vector **B** from the library to your friend's home. This combined trip is shown as the upper path in Fig. A.1c. But you could also go directly from your home to your friend's home by following a new vector in Fig. A.1c—vector **C**. This vector from your home to your friend's home is the sum of vectors **A** and **B**, and is 3.162 km long and points 18.43° north of east. Using bold letters to indicate that **A**, **B**, and **C** are vectors, we can write $\mathbf{A} + \mathbf{B} = \mathbf{C}$, meaning that vector **C** is the sum of vectors **A** and **B**.

Another interesting path from your home to your friend's home is shown in Fig. A.1d; you first travel along vector **B**

and then along vector **A**. On this path, you will arrive at your friend's home without visiting the library. The first leg of this journey will take you into new territory, but the second leg will leave you at your friend's home. The sum of vectors **B** and **A** is still vector **C**, or $\mathbf{B} + \mathbf{A} = \mathbf{C}$. Thus vectors added in any order yield the same sum.

While you can estimate the sum of two vectors by drawing arrows on a sheet of paper, obtaining an accurate sum requires some thought. Adding their magnitudes is unlikely to give you the magnitude of the sum vector, and adding their directions doesn't even make sense. To add two vectors, it helps to specify them in another form: as their *components* along two or three directions that are at right angles with respect to one another. In our example of travel in a city, two right-angle directions are all

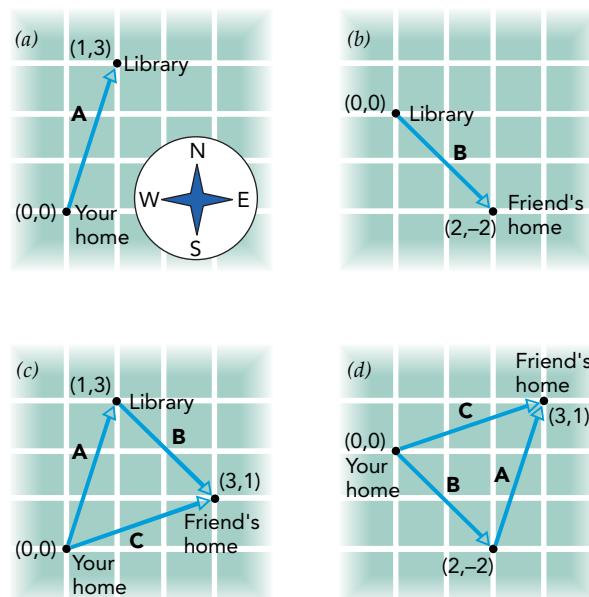


Fig. A.1. Four aerial views of your city, showing the major north-south and east-west streets that are spaced 1 km apart.
(a) To go from your home to the library, you must travel 3.162 km in the direction 18.43° east of north, which is vector **A**.
(b) To go from the library to your friend's home, you must travel 2.828 km in the direction 45° east of south, which is vector **B**.
(c) You can go from your home to your friend's home by two routes: you can go first to the library along vector **A** and then go to your friend's home along vector **B**, or you can travel directly along vector **C**, which is the sum of vectors **A** and **B**. (d) You can also reach your friend's home by going first along vector **B** and then along vector **A**. The sum of these two vectors is still vector **C**. However, you will not visit the library on that trip.

we need. If height were also important, we'd need three right-angle directions.

For the two right-angle directions we need in the city, let's choose east and north. We can then specify vector **A** as its component toward the east and its component toward the north. Its east component is 1 km and its north component is 3 km, so vector **A**, the position of the library with respect to your home, is 1 km to the east and 3 km to the north.

This new form for vector **A**, a pair of distances, is often more convenient than the old form, a distance and a direction. If you go to the library by walking 3.162 km in the direction 18.43° east of north, you'll have to pass through a lot of other buildings and backyards. Walking to the library by heading 1 km east and 3 km north allows you to travel on the sidewalks.

If we designate the position of your home as 0 km east and 0 km north, then the position of the library is 1 km east and 3 km north. These positions are labeled in Fig. A.1a as (0,0) and (1,3), respectively. The first number in the parentheses is the distance east, measured in kilometers, and the second number is the distance north, also in kilometers.

To go from the library to your friend's home, you must go 2 km east and 2 km south. That's the new form of vector **B**. Because a southward position has a negative component along

the northward direction, the position of your friend's home with respect to the library is 2 km east and -2 km north. In Fig. A.1b, the library is at (0,0) and your friend's home is at (2, -2).

Now adding these vectors **A** and **B** is relatively easy. To go from your home to your friend's home by way of the library, you must move east first 1 km and then 2 km for a total of 3 km, and you must move north 3 km and then -2 km for a total of 1 km. Thus the position of your friend's home with respect to your home, vector **C**, is 3 km east and 1 km north. In Fig. A.1c, your home is at (0,0) and your friend's home is at (3,1).

Similarly, you can go from your home to your friend's home by following first vector **B** and then vector **A**, as shown in Fig. A.1d. This trip will take you through unknown regions, but you'll still arrive at the right place. You move east first 2 km and then 1 km, and you move north first -2 km and then 3 km. In the end, you'll be 3 km east and 1 km north of your home, the position of your friend's home. Thus the sum of vectors **B** and **A** is still vector **C**.

As you can see, vectors are useful for specifying physical quantities in our three-dimensional world. When you encounter vectors, remember that their directions are just as important as their magnitudes and that these directions must be taken into account when you add two vectors together.

Appendix B

Units, Conversion of Units

When you return from a camping trip and begin to describe it to your friends, there are a number of physical quantities that you may find useful in your conversation. Distance, weight, temperature, and time are as important in everyday life as they are in a laboratory. And when you explain to your friends how far you hiked, how much your backpack weighed, how cold the weather was, and how long the trip took, you'll have to relate those quantities to standard units or your friends won't appreciate just how difficult the trip was.

Most physical quantities aren't simple numbers like 7 or 2.9. Instead, they have units like length or time, and are specified in multiples of widely accepted standard units such as meters or seconds. When you say a door is 3.0 meters tall, you're comparing the door's height to the meter, a widely accepted standard unit of length. With this comparison, most people can determine just how tall the door is, even though they've never seen it.

But the meter isn't familiar to everyone; many people prefer to measure length in multiples of another standard unit—the foot. These people might be more comfortable hearing that the door is 9.8 feet tall. These two quantities, 3.0 meters and 9.8 feet, are the same length.

Determining the door's height in feet doesn't require a new measurement because we can convert the one in meters to one that's in feet. To perform this conversion, we need to know how to express one particular length in both units. Any length will do. For example, Table B.1 states that 1 foot is the same length as 0.30480 meters. Because of that equality, we know that the following equation is true:

$$\frac{1 \text{ foot}}{0.30480 \text{ meters}} = 1.$$

We can multiply 3.0 meters, the height of the door, by this version of 1 and obtain the door's height in feet:

$$3.0 \text{ meters} \cdot \frac{1 \text{ foot}}{0.30480 \text{ meters}} = 9.8425 \text{ feet.}$$

Notice that the original units (meters) cancel and are replaced by the new units (feet). Since we knew the door's height only to 2 digits of precision in meters, we can't report the door's height to any more than 2 digits of precision in feet. So, we round the result to 9.8 feet.

You can change the units of almost any physical quantity by multiplying that quantity by a version of 1. You should form this 1 by dividing new units by old units, where the number of new units in the numerator is equivalent to the number of old units in the denominator. You can obtain these pairs of equivalent quantities from Table B.1. When you do this multiplication, the old units will cancel and you'll be left with the physical quantity expressed in the new units.

One last note about units: When you use physical quantities in a calculation, make sure that you keep the units throughout the process. They're as important in the calculation as they are anywhere else. Some of the units may cancel, but in all likelihood the result of the calculation will have some units left, and these units must be appropriate to the type of result you expect. If you're expecting a length, the units of your result should be meters or feet or another standard unit of length. If the units you obtain are seconds or kilograms, you've made a mistake in the calculation.

TABLE B.1 Conversion of Units

This table lists pairs of equivalent quantities, one in SI units and one in other units. Each pair can be used to convert measurements expressed as multiples of one unit into measurements expressed as multiples of the other unit. These pairs are grouped according to physical quantity and are given to a precision of 5 digits.

1. Acceleration: (SI unit: 1 meter/second ² or 1 m/s ²)	9. Power: (SI unit: 1 watt or 1 W)
1 foot/second ²	= 0.30480 m/s ²
2. Angle: (SI unit: 1 radian)	
1 degree (1°)	= 0.017453 radians
3. Area: (SI unit: 1 meter ² or 1 m ²)	
1 foot ²	= 0.092903 m ²
1 inch ²	= 6.4516·10 ⁻⁴ m ²
4. Density: (SI unit: 1 kilogram/meter ³ or 1 kg/m ³)	
1 pound/foot ³	= 16.018 kg/m ³
5. Energy: (SI unit: 1 joule or 1 J)	
1 Btu	= 1054.7 J
1 calorie, thermochemical	= 4.1840 J
1 electron-volt (1 eV)	= 1.6022·10 ⁻¹⁹ J
1 foot-pound	= 1.3558 J
1 kilocalorie (food Calorie)	= 4,186.8 J
1 kilowatt-hour	= 3,600,000 J
6. Force: (SI unit: 1 newton or 1 N)	
1 pound-force (1 lbf)	= 4.4482 N
7. Length: (SI unit: 1 meter or 1 m)	
1 angstrom (1 Å)	= 10 ⁻¹⁰ m
1 centimeter (1 cm)	= 0.01 m
1 fermi (1 fm)	= 10 ⁻¹⁵ m
1 foot	= 0.30480 m
1 inch	= 0.02540 m
1 kilometer (1 km)	= 1000 m
1 light year	= 9.4606·10 ¹⁵ m
1 micron (1 μm)	= 10 ⁻⁶ m
1 mil	= 2.5400·10 ⁻⁵ m
1 mile	= 1609.3 m
1 millimeter (1 mm)	= 0.001 m
1 nanometer (1 nm)	= 10 ⁻⁹ m
1 picometer (1 pm)	= 10 ⁻¹² m
8. Mass: (SI unit: 1 kilogram or 1 kg)	
1 gram (1 g)	= 0.001 kg
1 metric ton	= 1000 kg
1 pound-mass (1 lbm)	= 0.45359 kg
1 slug	= 14.594 kg
9. Power: (SI unit: 1 watt or 1 W)	
1 Btu/hour	= 0.29307 W
1 horsepower	= 745.70 W
10. Pressure: (SI unit: 1 pascal or 1 Pa)	
1 atmosphere	= 101,325 Pa
1 millimeter of mercury (1 torr)	= 133.32 Pa
1 pound/inch ² (1 psi)	= 6894.8 Pa
11. Temperature: (SI units: degree Celsius or °C; kelvin or K)	
Because temperature in the three common units, °C, K, and °F (degree Fahrenheit), aren't multiples of one another, you must use special formulas to convert between them:	
Temperature in °C = 5/9·(Temperature in °F – 32)	
Temperature in °C = Temperature in K – 273.15	
Temperature in K = Temperature in °C + 273.15	
12. Time: (SI unit: 1 second or 1 s)	
1 day	= 86,400 s
1 femtosecond (1 fs)	= 10 ⁻¹⁵ s
1 hour	= 3600 s
1 microsecond (1 μs)	= 10 ⁻⁶ s
1 millisecond (1 ms)	= 0.001 s
1 minute	= 60 s
1 nanosecond (1 ns)	= 10 ⁻⁹ s
1 picosecond (1 ps)	= 10 ⁻¹² s
13. Torque: (SI unit: 1 newton-meter or 1 N·m)	
1 inch-pound	= 0.11298 N·m
1 foot-pound	= 1.3558 N·m
14. Velocity: (SI unit: 1 meter/second or 1 m/s)	
1 foot/second	= 0.30480 m/s
1 kilometer/hour (1 km/h)	= 0.27778 m/s
1 knot	= 0.51444 m/s
1 mile/hour (1 mph)	= 0.44704 m/s
15. Volume: (SI unit: 1 meter ³ or 1 m ³)	
1 cup	= 2.3659·10 ⁻⁴ m ³
1 fluid ounce	= 2.9574·10 ⁻⁵ m ³
1 foot ³	= 0.028317 m ³
1 gallon	= 0.0037854 m ³
1 liter (1 l)	= 0.001 m ³
1 milliliter (1 ml)	= 10 ⁻⁶ m ³
1 quart	= 0.00094635 m ³

Glossary

absolute temperature scale A scale for measuring temperature in which 0 K corresponds to absolute zero.

absolute zero The temperature at which all thermal energy has been removed from an object or system of objects. Because it's impossible to find and remove all the thermal energy from an object, absolute zero can be approached but is not actually attainable.

acceleration A vector quantity that measures the rate at which an object's velocity is changing: the greater the acceleration, the more the object's velocity changes each second. It consists of both the amount of acceleration and the direction in which the object is accelerating. This direction is identical to the direction of the force causing the acceleration. The SI unit of acceleration is the meter per second².

acceleration due to gravity A physical constant that specifies how quickly a freely falling object accelerates and also relates an object's weight to its mass. At Earth's surface, the acceleration due to gravity is 9.8 m/s² (or 9.8 N/kg).

activation energy The energy required to initiate a chemical reaction. This energy serves to break or weaken the bonds in the starting chemicals so that the reaction can proceed to form the reaction products.

adverse pressure gradient A region of fluid flow in which the fluid must flow toward higher pressure. The fluid's momentum and kinetic energy carry it through this situation, although the fluid does slow down.

aerodynamic forces The forces exerted on an object by the motion of the air surrounding it. The two types of aerodynamic forces are lift and drag.

aerodynamics The study of the dynamic (moving) interactions of air with objects.

airfoil An aerodynamically engineered surface designed to obtain particular lift and drag forces from the air flowing around it.

airspeed An object's speed relative to the air through which it moves.

alpha decay A radioactive decay in which a helium nucleus (two protons and two neutrons) escapes from a larger proton-rich nucleus via quantum tunneling.

alternating current (AC) An electric current that periodically reverses its direction of flow.

ampere (A) The SI unit of electric current (synonymous with the coulomb per second). One ampere is defined as the passage of 6.25×10^{18} charged particles per second and is roughly the current flowing through a 100-W lightbulb operating on household electric power.

ampere-meter (A · m) The SI unit of magnetic pole.

amplifier A device that replicates an input signal as a larger output signal.

amplitude The maximal displacement of an oscillator away from its equilibrium position.

amplitude modulation (AM) A technique for representing sound or data by changing the amplitude (strength) of a wave.

analog representation The representation of numbers directly as continuous values of physical quantities such as voltage, charge, or pressure.

angle of attack The angle at which an airfoil is tilted relative to the airflow around it.

angular acceleration A vector quantity that measures how quickly an object's angular velocity is changing; the greater the angular acceleration, the more the object's angular velocity changes each second. It consists

of both the amount of angular acceleration and the direction about which the angular acceleration occurs. This direction is identical to the direction of the torque causing the angular acceleration. The SI unit of angular acceleration is the radian per second².

angular impulse The mechanical means for transferring angular momentum. One object gives an angular impulse to a second object by exerting a certain torque on the second object for a certain amount of time. In return, the second object gives an equal but oppositely directed angular impulse to the first object.

angular momentum A conserved vector quantity that measures an object's rotational motion. It is that object's rotational mass times its angular velocity. The SI unit of angular momentum is the kilogram-meter² per second.

angular position A quantity that describes an object's orientation relative to some reference orientation.

angular speed A measure of the angle through which an object rotates in a certain amount of time.

angular velocity A vector quantity that measures how rapidly an object's angular position is changing; the greater the angular velocity, the farther the object turns each second. It consists of both the object's angular speed and the direction about which the object is rotating. This direction points along the axis of rotation in the direction established by the right-hand rule. The SI unit of angular velocity is the radian per second.

anharmonic oscillator An oscillator in which the restoring force on an object is not proportional to its displacement from a stable equilibrium. The period of an anharmonic oscillator depends on the amplitude of its motion.

antimatter Matter resembling normal matter, but with many of its characteristics such as electric charge reversed.

aperture The diameter or effective diameter of a lens or opening.

apparent weight The sum of a person's weight plus their feeling of acceleration. All three quantities are vectors, so apparent weight can be quite large if the weight and feeling point in the same direction or quite small if they point in opposite directions.

Archimedes' principle The observation that an object partially or wholly immersed in a fluid is acted on by an upward buoyant force equal to the weight of the fluid it's displacing.

atmospheric pressure The pressure of air in Earth's atmosphere. Atmospheric pressure reaches a maximum of about 100,000 Pa near sea level and diminishes with increasing altitude.

atom The smallest portion of a chemical element that retains the chemical properties of that element.

atomic number The number of protons present in an atomic nucleus and equal to the number of electrons in a neutral atom.

axis of rotation The straight line in space about which an object or group of objects rotates. More specifically, the axis of rotation points in a particular direction along that line to reflect the sense of rotation according to the right-hand rule.

back emf The self-induced electromotive force that develops in an inductor when its current changes or in the coil of an electromechanical system such as a motor when its current causes magnets to move.

balanced Experiencing zero overall torque due to gravity.

band A group of levels in a solid that involve similar standing waves and thus have similar energies.

band gap A range of energies over which a solid has no levels available.

bandwidth The range of frequencies involved in a group of electromagnetic waves.

base of support A surface outlined on the ground by the points at which an object is supported.

Bernoulli's equation An equation relating the total energy of an incompressible fluid in steady-state flow to the sum of its pressure potential energy, kinetic energy, and gravitational potential energy.

beta decay A radioactive decay in which the weak force allows a neutron in a neutron-rich nucleus to disintegrate into an electron, a proton, and an antineutrino. The proton remains in the nucleus, but the electron and antineutrino escape.

binary The digital representation of numbers in terms of the powers of two. The number 6 is represented in binary as **110**, meaning 1 four (2^2), 1 two (2^1), and 0 ones (2^0).

black hole A region of space, normally spherical, within which the gravitational distortions of space and time are so severe that not even light can escape.

blackbody spectrum The distribution of thermal electromagnetic radiation emitted by a black object. This distribution is the amount of radiation emitted at each wavelength and depends only on the temperature of the black object.

blunt Not streamlined. Fluid flowing around a blunt object stalls and experiences flow separation and pressure drag.

boiling Accelerated evaporation that occurs when stable gas-phase bubbles form and grow inside a material's liquid phase.

boiling temperature The threshold temperature at which gas-phase bubbles first become stable within a material's liquid phase.

Boltzmann constant The constant of proportionality relating a gas's pressure to its particle density and temperature. It has a measured value of $1.381 \times 10^{-23} \text{ Pa} \cdot \text{m}^3/(\text{particle} \cdot \text{K})$.

boundary layer A thin region of fluid near a surface that, because of viscous drag, is not moving at the full speed of the surrounding airflow.

bremsstrahlung The process in which a rapidly accelerating charge emits electromagnetic radiation, usually an X-ray or gamma-ray photon.

Brewster's angle The angle at which no vertically polarized light reflects from a transparent surface that is oriented horizontally. The precise angle depends on the surface's index of refraction.

buoyant force The upward force exerted by a fluid on an object immersed in that fluid. The buoyant force is actually caused by pressure from the fluid. That pressure is highest below the object, so the force exerted upward on the object's bottom is greater than the force exerted downward on the object's top.

byte Eight binary bits that collectively can represent a number from 0 to 255. Bytes are often used to represent letters and other characters, where a convention associates each character with a specific number.

calibration The process of comparing a local reference object to a generally accepted standard.

capacitance The amount of separated charge on the plates of a capacitor divided by the voltage difference across those plates. The SI unit of capacitance is the farad.

capacitor An electronic component that stores separated electric charge on a pair of plates that are separated by an insulating layer.

Celsius (°C) A temperature scale in which 0 °C is defined as the melting point of water and 100 °C is defined as the boiling point of water at sea level. Absolute zero is -273.15°C .

center of gravity The unique point about which all of an object's weight is evenly distributed and therefore balanced. Because weight is proportional to mass, the center of mass is identical to the center of gravity for objects that are much smaller than Earth. For larger objects, the centers of mass and gravity differ slightly. An object suspended from its center of gravity will balance and will experience no net torque due to gravity. In many situations, you can accurately predict an object's behavior by assuming that all the object's weight acts at its center of gravity.

center of mass The unique point about which all the object's mass is balanced.

- center of mass** The center of mass is the natural pivot point for a free object. In the absence of outside forces or torques, a rigid object's center of mass travels at constant velocity while the object rotates at constant angular velocity about this center of mass.
- center of percussion** The special spot on a bat or racket where a collision with a second object will not cause any acceleration of the bat's handle.
- center of rotation** The point around which all the physical quantities of rotation are defined.
- centripetal acceleration** An acceleration that is always directed toward the center of a circular trajectory.
- centripetal force** A centrally directed force on an object. A centripetal force is not an independent force but, rather, the sum of other forces such as gravity acting on the object.
- chain reaction** A process in which one event triggers an average of one or more similar events so that the process becomes self-sustaining.
- chaos** Unpredictable behavior in which minute changes in a system's initial arrangement lead to very different final arrangements. These differences grow more dramatic with each passing second.
- chaotic system** A dynamic system that is exquisitely sensitive to initial conditions. Minute changes in how you set up a chaotic system can lead to wildly different final configurations.
- characteristic X-ray** An X-ray emitted by X-ray fluorescence from an atom. The energy of the characteristic X-ray is determined by the atom's orbital structure.
- charges** Objects, particularly small particles, that carry electric charge.
- chemical bond** An energy deficit that holds two or more atoms together to form a molecule and that must be repaid to separate the atoms. Chemical bonds form when chemical potential energy is released during a molecule's assembly.
- chemical potential energy** Energy stored in the chemical forces between atoms. Those chemical forces are electromagnetic in origin.
- chemical reaction** An encounter between two or more atoms and molecules that results in a rearrangement of the atoms to form different atoms and molecules.
- circularly polarized** The quality of a light wave in which the electric and magnetic fields rotate about the direction in which that wave is heading.
- closed circuit** A complete electric circuit through which electric current can flow continuously.
- coast** To travel at a steady speed along a straight-line path in accordance with inertia and the constant-velocity motion of an object that is experiencing zero net force.
- coefficient of restitution** The measure of a ball's liveliness, determined by bouncing the ball from a rigid, immovable surface. It's the ratio of the ball's rebound speed to its collision speed.
- coefficient of volume expansion** The fractional change in an object's volume caused by a temperature increase of 1 °C.
- coherent light** Light consisting of identical photons that together form a single electromagnetic wave.
- collision energy** The amount of kinetic energy removed from two objects as they collide.
- color temperature** The temperature at which a black object will emit thermal electromagnetic radiation with this particular distribution of wavelengths.
- components** The portions of a vector quantity that lie along particular directions.
- compressible** Changing density significantly as the pressure changes. A gas is compressible, since its density is proportional to its pressure.
- Compton scattering** The process in which a photon bounces off a charged particle, usually an electron. The photon and charged particle exchange energy and momentum during the collision.
- condensation** The phase transformation in which a gas becomes a liquid.
- conduction band** The group of quantum levels in an insulator that lies above the Fermi level.
- conduction level** A quantum level in an insulator that requires more energy than the Fermi level and that is normally unoccupied by electrons.
- conserved quantity** A physical quantity, such as energy, that is neither created nor destroyed within an isolated system when that system undergoes changes. A conserved quantity may pass among the objects within an isolated system, but its total amount remains constant.
- constructive interference** Interference in which two or more waves arrive at a location in space and time in phase with one another and produce a particularly strong effect.
- convection** The transmission of heat by the movement of a fluid. Convection normally entails the natural circulation of the fluid that accompanies differences in temperatures and densities.
- convection cell** A loop of fluid flow that is propelled by convection. Fluid in a convection cell normally rises in a hotter region and descends in a colder region.
- convection current** A fluid flow propelled by convection.
- converging lens** A lens that bends the individual light rays passing through it toward one another so that they either converge more rapidly than before or at least diverge less rapidly from one another. A converging lens has a positive focal length and often produces real images.
- corona discharge** A faintly glowing discharge that surrounds a small, highly charged object in the presence of a gas. In the discharge, electric charge is transferred from the object to the gas molecules.
- coulomb (C)** The SI unit of electric charge. One coulomb is about 1 million times the charge you acquire by rubbing your feet across a carpet in winter.
- Coulomb constant** The fundamental constant of nature that determines the electrostatic forces two charges exert on one another. Its measured value is $8.988 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$.
- Coulomb's law** The magnitudes of the electrostatic forces between two objects are equal to the Coulomb

constant times the product of their two electric charges, divided by the square of the distance separating them. If the charges are like, then the forces are repulsive. If the charges are opposite, then the forces are attractive.

Coulomb's law for magnetism The magnitudes of the magnetostatic forces between two objects are equal to the permeability of free space times the product of their two magnetic poles, divided by 4π times the square of the distance separating them. If the poles are like, then the forces are repulsive. If the poles are opposite, then the forces are attractive.

crest A peak positive excursion of an extended system that is experiencing a wave.

critical mass A portion of a fissionable material that is able to sustain a fission chain reaction. The amount of material required depends on its mass, shape, and density.

crystalline Having its atoms arranged in an orderly pattern that extends for many atomic spacings in all directions.

current The amount of electric charge flowing past a point or through a surface per unit of time. The SI unit of current is the ampere.

cycle per second (1/s) The SI unit of frequency (synonymous with hertz).

cyclotron motion The circular or spiral motion of a charged particle in a magnetic field. The charged particle tends to loop around the magnetic flux lines.

decimal The digital representation of numbers in terms of the powers of 10. The number 124 is represented in decimal as 124, meaning 1 hundred (10^2), 2 tens (10^1), and 4 ones (10^0).

demagnetized The quality of a material when its magnetic polarization has been reduced, typically to zero.

density The mass of an object divided by its volume. The SI unit of density is the kilogram per meter³.

density gradient A gradual change in density near a given position; a vector quantity that points in the direction of fastest increase in density and has a magnitude that is the rate of that increase.

depletion region The nonconducting region around a p-n junction in which all of the valence levels are filled with electrons and there are no conduction-level electrons.

deposition The phase transformation in which a gas becomes a solid.

destructive interference Interference in which two or more waves arrive at a location in space and time out of phase with one another and produce a particularly weak effect.

diffraction A wave phenomenon that limits the focusability of light and alters the way in which it travels after passing through an opening.

digital representation The representation of numbers by decomposition into digits that are then individually represented by discrete values of physical quantities such as voltage, charge, or pressure.

direct current (DC) An electric current that always flows in one direction.

direction The line or course on which something is moving, is aimed to move, or is pointing or facing.

discharge A flow of electric current through a gas.

dispersion The dependence of light's speed through a material on the frequency of that light.

distance The length between two positions in space. The SI unit of distance is the meter.

diverging lens A lens that bends the individual light rays passing through it away from one another so that they either converge less rapidly than before or don't converge at all. A diverging lens has a negative focal length and often produces virtual images.

domain wall A boundary surface between magnetic domains having different directions of magnetic orientation.

doped Modified by adding chemical impurities that change its physical properties.

Doppler effect A difference between the frequency at which a wave is sent and the frequency at which that wave is received caused by relative motion between the sender and receiver.

drag forces The friction-like forces exerted by a fluid and a solid on one another as the solid moves through the fluid. These forces act to reduce the relative velocity between the two.

dynamic stability An object's stability when it's in motion.

dynamic variation Change in a physical quantity such as pressure that is caused by motion.

elastic collision A collision in which all the kinetic energy present before the impact is again present as kinetic energy after the impact.

elastic limit The most extreme distortion of an object from which it can return to its original size and shape without permanent deformation.

elastic potential energy The energy stored by the forces within a distorted elastic object.

elastic scattering The process in which two particles bounce off one another without losing any of their kinetic energies.

electric charge An intrinsic property of matter that gives rise to electrostatic forces between charged particles. Electric charge is a conserved physical quantity. A specific charge can have a positive amount of electric charge (a positive charge) or a negative amount (a negative charge). The SI unit of electric charge is the coulomb.

electric circuit A complete loop of conductors, loads, and power sources through which an electric current can flow continuously.

electric current The movement or flow of electric charge.

electric field An attribute of each point in space that exerts forces on electrically charged particles. An electric field has a magnitude and direction proportional to the force it would exert on a unit of positive charge at that location. While electric fields are often created by nearby charges, they can also be created by other electromagnetic phenomena. The SI unit of electric field is the volt per meter or, equivalently, the newton per coulomb.

electric polarization A distribution of electric charge that is nonuniform so that the object has a region of

- positive charge** and a region of negative charge.
- electrical conductor** A material that allows electric charge to move through it.
- electrical insulator** A material that prevents any net movement of electric charge through it.
- electrical resistance** The measure of how much an object impedes the flow of electric current. The SI unit of electrical resistance is the ohm.
- electromagnet** A coil of wire, with or without an iron core, that becomes a magnet when an electric current flows through the coil.
- electromagnetic spectrum** The entire range of possible frequencies and wavelengths of electromagnetic waves.
- electromagnetic waves** Waves consisting of electric and magnetic fields that travel through empty space at the speed of light. These waves carry energy and momentum and are emitted and absorbed as particles called photons. Radio waves; microwaves; infrared, visible, and ultraviolet light; X-rays; and gamma rays are examples of electromagnetic waves.
- electron volt (eV)** A unit of energy equal to the energy obtained by an elementary charge (electron or proton) as it moves through a voltage difference of 1 V. One electron volt is equal to about 1.602×10^{-19} J.
- electron-positron pair production** The formation of an electron and a positron during an energetic collision.
- electrons** The tiny negatively charged particles that make up the outer portions of atoms and that are the main carriers of electricity and heat in metals.
- electrostatic force** The force experienced by a charged particle in the presence of other charged particles.
- electrostatic potential energy** Energy stored in the forces between electric charges.
- elementary unit of electric charge** The basic quantum of electric charge, equal to about 1.6×10^{-19} C.
- emissivity** A surface's capacity to emit or absorb thermal radiation relative to that of a perfectly black object at the same temperature.
- energy** The capacity to do work. Each object has a precise quantity of energy, which determines exactly how much work that object can do in an ideal situation. The SI unit of energy is the joule.
- English system of units** An assortment of antiquated units that were used throughout the English colonies and remain in common use in the United States today. Units in this system include feet, ounces, and miles per hour.
- entropy** The physical quantity measuring the amount of disorder in a system. The system's entropy would be zero at absolute zero.
- equilibrium** The state of an object in which zero net force (or zero net torque) acts on it. An object that is stationary or in uniform motion is in equilibrium.
- equilibrium length** The natural length of a spring or springlike object when it is free of external forces and motionless.
- equilibrium position** The point at which an object experiences zero net force and doesn't accelerate.
- escape velocity** The speed a spacecraft needs to follow a parabolic orbital path and escape forever from a particular celestial object.
- evaporation** The phase transformation in which a liquid becomes a gas.
- excited state** A configuration of a system having excess energy; its electrons (or other particles) are in an arrangement of quantum waves (e.g., orbitals or levels) that has more than the least possible energy.
- exhaust velocity** The velocity of exhaust gas relative to the rocket engine from which it emerged.
- explosive chain reaction** A chain reaction in which each fission induces an average of much more than one subsequent fission and the fission rate skyrockets.
- Fahrenheit ($^{\circ}\text{F}$)** A temperature scale in which $32\text{ }^{\circ}\text{F}$ is defined as the melting point of water and $212\text{ }^{\circ}\text{F}$ is defined as the boiling point of water at sea level. Absolute zero is $-459.67\text{ }^{\circ}\text{F}$.
- farad (F)** The SI unit of electric capacitance. A 1-farad capacitor will have a voltage difference between its plates of 1 volt when storing 1 coulomb each of separated positive and negative charge.
- feedback** The process of using information about a system's current situation to control changes you are making in that system.
- feeling of acceleration** A person undergoing acceleration experiences a weightlike sensation in the direction exactly opposite the direction of acceleration. The amount of this feeling of acceleration is proportional to the amount of the acceleration.
- Fermi energy** The energy of an electron in the Fermi level.
- Fermi level** A hypothetical level located halfway between the highest occupied level and the lowest unoccupied level in a solid.
- Fermi particles** A class of fundamental particles that includes electrons, protons, and neutrons and that obeys the Pauli exclusion principle.
- ferromagnetic** Composed of magnetic atoms that all have the same magnetic orientation within a magnetic domain.
- field** A structure that associates a physical quantity with each point in space and time.
- firm** Having a large spring constant and thus experiencing large restoring forces in response to small distortions.
- first law of thermodynamics** The law of conservation of energy.
- fissionable** Able to undergo induced fission.
- fluid** A substance that has mass but no fixed shape. A fluid can flow to match its container. Gases and liquids are both fluids.
- fluorescence** An emission of light that immediately follows an absorption of light.
- f-number** The ratio of a lens's focal length to its effective aperture.
- focal length** The distance after a converging lens at which the real image of a distant object forms. The focal length of a diverging lens is negative and is the distance *before* the lens at

- which the virtual image of a distant object forms.
- forward biased** A p-n junction in which the voltage of the p-type semiconductor has been raised relative to the voltage of the n-type semiconductor.
- freezing** The phase transformation in which a liquid becomes a solid.
- frequency** The number of cycles completed by an oscillating system in a certain amount of time. The SI unit of frequency is the hertz.
- frequency modulation (FM)** A technique for representing sound or data by changing the exact frequency of a wave.
- friction** The force that resists relative motion between two surfaces in contact. Frictional forces are exerted parallel to the surfaces in the directions opposing their relative motion.
- fundamental forces** The four basic forces that act between objects in the universe: the gravitational force, the electromagnetic force, the strong force, and the weak force.
- fundamental vibrational mode** The slowest and often broadest vibration that an extended object can support.
- gamma rays** Extremely high-energy photons of electromagnetic radiation, often produced during radioactive decays.
- gas** A form of matter consisting of tiny individual particles (atoms or molecules) that travel around independently. A gas takes on the shape and volume of its container.
- gaseous** In a gas phase.
- general theory of relativity** The physical rules governing all motion, even motion involving speeds comparable to the speed of light and occurring in the presence of massive objects.
- gradient** A gradual change in some physical quantity near a given position; a vector quantity that points in the direction of fastest increase in that physical quantity and has a magnitude that is the rate of that increase.
- gravitational constant** The fundamental constant of nature that determines the gravitational forces two masses exert on one another. Its value is $6.6720 \times 10^{-11} \text{ N}\cdot\text{m}^2/\text{kg}^2$.
- gravitational mass** The mass associated with the gravitational attraction between objects.
- gravitational potential energy** The potential energy stored in the gravitational forces between objects.
- gravity** The gravitational attraction of the mass of Earth, the moon, or a planet for bodies at or relatively near its surface. All objects exert gravitational forces on all other objects.
- greenhouse effect** An increase in surface temperature that occurs when a material impedes radiative heat loss by that surface or its environment.
- ground state** The lowest energy configuration of a system; its electrons (or other particles) are in the arrangement of quantum waves (e.g., orbitals or levels) that has the least possible energy.
- half-life** The time needed for half the nuclei of a particular radioactive isotope to undergo radioactive decay.
- hard magnetic material** A material that is relatively difficult to magnetize and that retains its magnetization once the magnetizing field is removed. Hard magnetic materials are suitable for permanent magnets.
- harmonic** An integer multiple of the fundamental frequency of oscillation for a system. The second harmonic is twice the frequency of the fundamental, and the third harmonic is three times the frequency of the fundamental. In principle, harmonics can continue forever.
- harmonic oscillator** An oscillator in which the restoring force on an object is proportional to its displacement from a stable equilibrium. The period of a harmonic oscillator doesn't depend on the amplitude of its motion.
- heat** The energy that flows from one object to another as a result of a difference in temperature between those two objects.
- heat capacity** The amount of heat that must be added to an object to cause its temperature to rise by 1 unit.
- heat engine** A device that converts thermal energy into work as heat flows from a hot object to a cold object.
- heat exchanger** A device that allows heat to flow naturally from a hotter material to a colder material without any actual exchange of those materials.
- heat pump** A device that pumps heat against its natural direction of flow, transferring it from a cold object to a hot object. To satisfy the second law of thermodynamics, a heat pump normally converts some ordered energy into thermal energy.
- Heisenberg uncertainty principle** A quantum physical law that states that an object's position (a particle characteristic) and momentum (a wave characteristic) can't be sharply defined at the same time. This principle gives objects with small masses a fuzzy character.
- henry (H)** The SI unit of inductance. A 1-henry inductor will experience a 1-ampere change in the current flowing through it each second when subjected to a 1-volt voltage drop.
- hertz (Hz)** The SI unit of frequency (synonymous with cycle per second).
- higher-order vibrational mode** A vibrational mode that is more complicated than the fundamental mode and in which different parts of the extended system move in opposite directions.
- Hooke's law** The general law covering spring and elastic behavior. Hooke's law states that a spring exerts a restoring force that is proportional to the distance the spring is distorted from its equilibrium length.
- horizontal polarization** An electromagnetic wave in which the electric field always points left or right (horizontally). The magnetic field always points vertically.
- ideal gas law** The law relating the pressure, temperature, and particle density of an ideal gas. An ideal gas is one that is composed of perfectly independent particles. The particles don't stick, and they bounce perfectly from one another.
- image distance** The distance between the lens and the image that the lens creates. Real images form at positive image distances, while virtual images form at negative image distances.
- impedance** A measure of a system's opposition to the passage of a current or a wave.

impedance mismatch An abrupt change in the opposition to a wave's passage, typically accompanied by reflections.

impulse The mechanical means for transferring momentum. One object gives an impulse to a second object by exerting a certain force on the second object for a certain amount of time. In return, the second object gives an equal but oppositely directed impulse to the first object.

in phase The relationship between two waves in which they complete the same portions of their oscillatory cycles at the same time and place.

incandescence The emission of thermal radiation from a hot object.

incoherent light Light consisting of individual photons, each its own independent electromagnetic wave.

incompressible A substance that doesn't change density significantly as its pressure changes. Liquids and solids are incompressible, since their densities change very little as their pressures change dramatically.

index of refraction The factor by which the speed of light in a material is reduced from its speed in empty space, equal to the speed of light in empty space divided by light's speed in the material.

induced drag The drag force that occurs when a wing deflects the stream of air passing across it to obtain lift.

induced emf (electromotive force) An overall voltage difference between the ends of a coil produced by a changing magnetic field in that coil and the resulting electric field.

induced fission A fission event that's caused by a collision, usually with a neutron.

inductance The voltage drop across an inductor divided by the rate at which current through that inductor is changing with time. The SI unit of inductance is the henry.

inductor An electronic component that stores magnetic energy in a coil of wire and opposes changes in current in that wire.

inelastic collision A collision in which some of the kinetic energy present before the impact is no longer present as kinetic energy after the impact.

inert gas A gas consisting of atoms that are chemically inactive and rarely bond permanently with other atoms or molecules. Inert gases include helium, neon, argon, krypton, and xenon.

inertia A property of matter by which it remains at rest or in uniform motion in the same straight line unless acted on by some outside force.

inertial Moving because of inertia alone and therefore not accelerating.

inertial frame of reference A frame of reference that is not accelerating and is thus either stationary or traveling at constant velocity. The laws of motion accurately describe any situation that is observed from an inertial frame of reference.

inertial mass The mass associated with an object's inertia, its resistance to acceleration.

infrared light Invisible light having wavelengths longer than about 750 nanometers.

insulator A solid in which the Fermi level falls within a band gap.

interference A wave phenomenon in which waves passing through the same location from different directions reinforce or oppose one another.

interference pattern A pattern of intensity variations in time and space that occurs when two or more waves are superposed and experience constructive and destructive interferences.

internal energy The sum of an object's thermal energy and any additional potential energy stored entirely within the object.

internal kinetic energy The portion of an object's kinetic energy that involves only the relative motion of particles within the object and that excludes the object's overall translation or rotation.

internal potential energy The portion of an object's potential energy that involves only forces between particles within the object and that excludes the object's interactions with its surroundings.

isotopes Chemically indistinguishable atoms containing nuclei that differ only in their numbers of neutrons.

joule (J) The SI unit of energy and work (synonymous with newton-meter). Lifting 1 liter of water upward 10 centimeters near Earth's surface requires about 1 joule of work.

joule per kilogram-kelvin (J/kg·K) The SI unit of specific heat.

joule per second (J/s) The SI unit of power (synonymous with watt).

Kelvin (K) The SI scale of absolute temperature, in which 0 K is defined as absolute zero. The spacing between units is the same as that used in the Celsius scale.

Kepler's first law All planets move in elliptical orbits, with the sun at one focus of the ellipse.

Kepler's second law A line stretching from the sun to a planet sweeps out equal areas in equal times.

Kepler's third law The square of a planet's orbital period is proportional to the cube of that planet's mean distance from the sun.

kilogram (kg) The SI unit of mass. (The standard kilogram is a platinum-iridium cylinder kept at the International Bureau of Weights and Measures near Paris.) A liter of water has a mass of about 1 kilogram.

kilogram per meter³ (kg/m³) The SI unit of density. One kilogram per meter³ is about the density of air at 2000 m (about 1 mile) above sea level.

kilogram-meter per second (kg·m/s) The SI unit of momentum. One kilogram-meter per second is about the momentum in a baseball traveling 25 km/h (16 mph).

kilogram-meter per second² (kg·m/s²) The SI unit of force (synonymous with newton).

kilogram-meter² (kg·m²) The SI unit of rotational mass. One kilogram-meter² is roughly the rotational mass of your forearm as it pivots about your elbow.

kilogram-meter² per second (kg·m²/s) The SI unit of angular momentum. One kilogram-meter² per second is about the angular momentum of a 7.3 kg (16 lbm) bowling ball spinning 34 times/second as it rolls down the lane.

kinetic energy The form of energy contained in an object's translational and rotational motion.

laminar flow Smooth, predictable fluid flow in which nearby portions of the fluid remain nearby as they travel along.

laser amplifier A device that amplifies weak incoming light to produce brighter outgoing light. The outgoing light is a brighter copy of the incoming light.

laser medium An assembly of excited atoms or other quantum systems that is capable of amplifying passing light through stimulated emission.

laser oscillator A laser amplifier that is surrounded by mirrors so that it can amplify one or more spontaneously emitted photons to form an intense beam of coherent light.

latent heat of evaporation The heat required to transform a unit mass of material from liquid to gas without changing its temperature.

latent heat of fusion Latent heat of melting.

latent heat of melting The heat required to transform a unit mass of material from solid to liquid without changing its temperature.

latent heat of vaporization Latent heat of evaporation.

law of conservation of energy The change in a stationary object's internal energy is equal to the heat transferred into that object minus the work that object does on its surroundings. This first law of thermodynamics is a restatement of energy conservation.

law of entropy The entropy of a thermally isolated system of objects never decreases. This second law of thermodynamics recognizes that creating disorder is easy; restoring order is hard.

law of thermal equilibrium Two objects that are each in thermal equilibrium with a third object are also in thermal equilibrium with one another. This zeroth law of thermodynamics is the basis for a meaningful system of temperatures.

laws of thermodynamics The four laws that govern the movement of heat between objects.

lens A transparent optical device that uses refraction to bend light, often to form images.

lens equation The equation relating a lens's focal length to the object and image distances.

Lenz's law When a changing magnetic field induces a current in a conductor, the magnetic field from that current opposes the change that induced it.

lever arm The directed distance from the pivot or axis of rotation to the point at which the force is exerted.

lift forces Forces exerted by a fluid on a solid that are at right angles to the fluid flow around that solid.

light Short-wavelength electromagnetic waves, including visible, infrared, and ultraviolet light.

linear momentum A conserved vector quantity that measures an object's motion. It is the object's mass times its velocity. The SI unit of linear momentum is the kilogram-meter per second.

liquid A form of matter consisting of particles (atoms or molecules) that are touching one another but are free to move relative to one another. A liquid has a fixed volume but takes the shape of its container.

longitudinal wave A wave in which the underlying oscillation is parallel to the wave itself.

Lorentz force The force experienced by a charged particle when it moves through a magnetic field.

lumen A common unit of total radiated light as perceived by a human eye.

luminescence The emission of light by any means other than thermal radiation.

magnetic dipole A pair of equal but opposite poles separated by a distance.

magnetic domains Regions of uniform alignment within a magnetic material.

magnetic field An attribute of each point in space that exerts forces on magnetic poles. A magnetic field has a magnitude and direction proportional to the force it would exert on a unit of north magnetic pole at that location. The SI unit of magnetic field is the tesla.

magnetic flux lines Abstract strands following along the local magnetic field direction and having a density proportional to that local field. Flux lines can only begin at north poles and end at south poles.

magnetic induction The process whereby a time-changing magnetic field initiates or influences an electric current.

magnetic monopole An isolated magnetic pole, either north or south. None has ever been observed.

magnetic polarization A distribution of magnetic poles that is nonuniform so that the object has a region of north pole and a region of south pole.

magnetic pole A property of nature that gives rise to magnetostatic forces between magnetic poles. A specific pole can have a positive amount of magnetic pole (a north pole) or a negative amount (a south pole). The SI unit of magnetic pole is the ampere-meter.

magnetized When a material's magnetic polarization is nonzero.

magnetostatic force The force experienced by a magnetic pole in the presence of other magnetic poles.

magnitude The amount of some physical quantity.

Magnus force A lift force experienced by a spinning object as it moves through a fluid. The Magnus force points toward the side of the ball moving away from the onrushing airstream.

mass The property of a body that is a measure of its inertia or resistance to acceleration, that is commonly taken as a measure of the amount of material it contains, and that causes it to have weight in a gravitational field. The SI unit of mass is the kilogram.

mechanical advantage The process whereby a mechanical device redistributes the amounts of force and distance that go into performing a particular amount of mechanical work.

mechanical wave A natural and often rhythmic motion of an extended object about its stable equilibrium shape or situation.

melting The phase transformation in which a solid becomes a liquid.

melting temperature The temperature at which a material's solid and liquid phases can coexist in stable equilibrium.

metal A solid in which the Fermi level falls within a band of levels.

meter (m) The SI unit of length or distance. (One meter is formally defined as the distance light travels through empty space in 1/299,792,458 of a second.) One meter is about the length of a long stride, or about 3.28 feet.

meter per second (m/s) The SI unit of velocity or speed. One meter per second is a typical walking pace, or about 2.2 mph.

meter per second² (m/s²) The SI unit of acceleration. One meter per second² is about the acceleration of an elevator as it first begins to move upward.

meter² (m²) The SI unit of area. One square meter is about twice the area of an opened newspaper.

meter³ (m³) The SI unit of volume. One cubic meter is about the volume of a four-drawer file cabinet.

microwaves Electromagnetic waves with wavelengths between about 1 meter and 1 millimeter.

mode A basic pattern of distortion or oscillation.

molecule A particle formed from two or more atoms. A molecule is the smallest portion of a chemical compound that retains the chemical properties of that compound.

moment of inertia Rotational mass.

momentum Linear momentum.

MOSFET A transistor in which the electric charge on a surface affects the electrical resistance of a channel for electric current. An abbreviation of metal-oxide-semiconductor field-effect transistor.

natural resonance A mechanical process in which an isolated object's energy causes it to perform a certain motion over and over again. The rate at which this motion occurs is determined by the physical characteristics of the object.

net electric charge The sum of all charges on an object, both positive and negative. Positive charges increase the net charge and negative

charges decrease it. Net charge can be negative.

net force The sum of all forces acting on an object, considering both the magnitude of each individual force and its direction. The magnitude of the net force is often less than the sum of the magnitudes of the individual forces, since they often oppose one another in direction.

net magnetic pole The sum of all poles on an object, both north and south. Since there are no isolated magnetic poles, an object's net magnetic pole is always zero.

net torque The sum of all torques acting on an object, considering both the magnitude of each individual torque and its direction. The magnitude of the net torque is less than the sum of the magnitudes of the individual torques, since they often oppose one another in direction.

neutral Having zero net electric charge.

neutral equilibrium A state of equilibrium to which the object has no tendency to return or not return if it's disturbed. At equilibrium, the object is free of net force or torque. If the object is moved away from a neutral equilibrium, it will still be in equilibrium.

neutrinos Chargeless and nearly massless particles created during radioactive decays and other nuclear events. They rarely interact with matter.

neutrons The electrically neutral subatomic particles that, together with protons, make up atomic nuclei.

newton (N) The SI unit of force (synonymous with the kilogram-meter per second²). The weight of 18 U.S. quarters is equal to about 1 newton. The common English unit of force, the pound, is about 4.45 newtons.

newton per ampere-meter (N/A · m) The SI unit of magnetic field (synonymous with tesla).

newton per coulomb (N/C) The SI unit of electric field (synonymous with volt per meter).

newton per meter² (N/m²) The SI unit of pressure (synonymous with pascal).

newton-meter (N · m) The SI unit of energy and work (synonymous with

the joule). Also the SI unit of torque, exerted by a 1-newton force located 1 meter from the axis of rotation. One newton-meter is about the torque exerted on your shoulder by the weight of a baseball held in your outstretched arm.

Newton's first law of motion An object that is free from all outside forces travels at a constant velocity, covering equal distances in equal times along a straight-line path.

Newton's first law of rotational motion An object that is not wobbling and is free from all outside torques rotates with constant angular velocity, spinning steadily about a fixed axis.

Newton's second law of motion An object's acceleration is equal to the force exerted on that object divided by the object's mass. This equality can be manipulated algebraically to state that the force on the object is equal to the product of the object's mass times its acceleration (Eq. 1.1.2).

Newton's second law of rotational motion An object's angular acceleration is equal to the torque exerted on that object divided by the object's rotational mass. This equality can be manipulated algebraically to state that the torque on the object is equal to the object's rotational mass times its angular acceleration (Eq. 2.1.2). The law doesn't apply to nonrigid or wobbling objects.

Newton's third law of motion For every force that one object exerts on a second object, there is an equal but oppositely directed force that the second object exerts on the first object.

Newton's third law of rotational motion For every torque that one object exerts on a second object, there is an equal but oppositely directed torque that the second object exerts on the first object.

normal Directed exactly away from (perpendicular to) a surface. A line that is normal to a surface meets that surface at a right angle.

normal force Support force.

n-type semiconductor A semiconductor such as silicon that contains impurity atoms such as phosphorus, arsenic, antimony, or bismuth that

- place electrons in the semiconductor's conduction level.
- nuclear fission** The shattering of a heavy nucleus into smaller fragments. During fission, the positively charged fragments repel one another and release energy.
- nuclear force** An attractive force that binds nucleons together once they touch one another.
- nuclear fusion** The merging of two small nuclei to form a larger nucleus. During fusion, the nuclear force binds the nucleons together and releases energy.
- nucleation** The formation of an initial seed of one material phase in the midst of another material phase.
- nucleon** A general name given to the particles that make up atomic nuclei: protons and neutrons.
- nucleus** The positively charged central component of an atom, containing most of the atom's mass and about which the electrons are arranged. Plural is nuclei.
- object distance** The distance between the lens and the object that it is imaging.
- ohm (Ω)** The SI unit of electrical resistance. A 1-ohm resistor exhibits a voltage drop of 1 volt when 1 ampere of current flows through it.
- ohmic** Exhibiting a voltage drop that's proportional to current, consistent with Ohm's law.
- Ohm's law** The observation that the voltage drop across an ordinary electrical conductor is proportional to both the electric current passing through it and to its electrical resistance.
- open circuit** An incomplete electric circuit where a gap in the electrical conductors stops electric current from flowing.
- orbit** The path an object takes as it moves in the presence of a centripetal force.
- orbital** An electron standing wave in an atom, one of the basic electron wave modes allowed in an atom by quantum physics.
- orbital period** The time required to complete one full orbit.
- ordered energy** Energy that can easily be used to do work.
- oscillation** A repetitive and rhythmic movement or process that usually takes place about an equilibrium situation.
- out of phase** The relationship between two waves in which they complete opposite portions of their oscillatory cycles at the same time and place.
- parallel circuit** (wiring arrangement) An arrangement in which the current reaching two or more electric devices divides into separate parts to flow through those devices and then joins back together as it leaves them. Current experiences the same change in voltage in each device.
- pascal (Pa)** The SI unit of pressure (synonymous with newton per meter²). Atmospheric pressure at sea level is about 100,000 pascals. A 1-millimeter-high water droplet exerts a pressure of about 10 pascals on your hand.
- Pascal's principle** A change in the pressure of an enclosed incompressible fluid is conveyed undiminished to every part of the fluid and to the surfaces of its container.
- Pauli exclusion principle** An observed property of nature that indistinguishable Fermi particles must each have their own unique quantum wave.
- period** The time required to complete one full cycle of a repetitive motion.
- permanent magnet** An object that can be magnetized and that retains that magnetization for a long time.
- permeability of free space** The defined constant that relates two poles and the magnetostatic forces they exert on one another. Its value is $4\pi \times 10^{-7} \text{ N/A}^2$.
- phase** A form of matter—notably solid, liquid, gas, and plasma.
- phase equilibrium** A situation in which two material phases coexist stably, neither one growing at the expense of the other.
- phase transition** A transformation from one material phase to another.
- phosphor** A solid that luminesces (emits light) when energy is transferred to it by light or by a collision with a particle.
- photoconductor** A solid that is an electrical insulator in the dark but that becomes an electrical conductor when exposed to light of the correct wavelength.
- photodiode** A diode that permits current to flow backward across the p-n junction when exposed to light. Light provides the energy needed to move charges across the junction's depletion region in the wrong direction. The current flowing in the reverse direction through a photodiode is proportional to the light intensity.
- photoelectric effect** The process in which an atom absorbs a photon in a radiative transition that ejects one of the electrons from the atom.
- photon** A particle or quantum of light having energy and momentum but no mass.
- pitch** The frequency of a sound.
- Planck constant** The fundamental constant of quantum physics, equal to the energy of an object divided by the frequency of its quantum wave. It is about $6.626 \times 10^{-34} \text{ J}\cdot\text{s}$.
- plane polarized** The quality of a light wave in which the electric field (and the magnetic field) fluctuates back and forth in a plane as the light travels through space.
- plasma** A gaslike phase of matter consisting of electrically charged particles, such as ions and electrons. The strong electromagnetic interactions between its particles distinguish a plasma from a gas.
- p-n junction** The interface between an n-type semiconductor and a p-type semiconductor that gives the diode its unidirectional characteristic for electrons.
- Poiseuille's law** The volume of fluid flowing through a pipe each second is equal to $\pi/128$ times the pressure difference across that pipe times the pipe's diameter to the fourth power, divided by the pipe's length times the fluid's viscosity.
- poles** Objects that carry a magnetic pole.
- population inversion** A nonequilibrium population of quantum systems in which more are in a higher energy state than in a lower energy state.
- position** A vector quantity that specifies the location of an object relative to

- some reference point.** It consists of both the length and the direction from the reference point to the object.
- positron** The antimatter counterpart of the electron. The positron is positively charged.
- potential energy** The stored form of energy that can produce motion. Potential energy is stored in the forces between or within objects.
- potential gradient** The rate at which an object's potential energy would increase if it were to move toward increasing potential energy along the steepest slope.
- precession** The change in orientation of a spinning object's rotational axis that occurs when it's subject to an outside torque.
- pressure** The average amount of force that a fluid exerts on a certain region of surface area. Pressure is reported as the amount of force divided by the surface area over which that force is exerted. The SI unit of pressure is the pascal.
- pressure drag** The drag force that results from higher pressures at the front of an object than at its rear.
- pressure gradient** A distribution of pressures that varies continuously with position.
- pressure potential energy** A fluid's volume times its pressure. However, this energy isn't really stored in the fluid. Instead, it's energy that's provided by a pump (or other source) when the fluid is delivered.
- primary colors of light** The three colors of light (red, green, and blue) that are sensed by the three types of color-sensitive cone cells in our eyes. Mixtures of these three colors of light can make our eyes perceive any possible color.
- primary colors of pigment** The three colors of pigment (cyan, magenta, and yellow) that absorb the three primary colors of light (red, green, and blue, respectively). Mixtures of these three pigments can be applied to a white surface to make it reflect any possible mixture of the three primary colors of light and thus to make our eyes perceive any possible color.
- principle of equivalence** The principle that gravitational mass and inertial mass are truly identical and therefore that no experiment you can perform in a small region of space can distinguish between free fall and the absence of gravity.
- protons** The positively charged subatomic particles found in atomic nuclei.
- p-type semiconductor** A semiconductor such as silicon that contains impurity atoms such as boron, aluminum, gallium, indium, or tellurium that remove electrons from the semiconductor's valence levels.
- quanta** The fundamental, discrete units in which an item is emitted, absorbed, or otherwise observed, reflecting the particulate character of that item.
- quantized** Existing only in discrete units or quanta. Quantized physical quantities are observed only in integer multiples of the elementary quantum.
- radian** The natural unit in which angles are measured. There are 2π radians in a full circle, so 1 radian is $180/\pi$ degrees or approximately 57.3°.
- radian per second (1/s)** The SI unit of angular velocity or angular speed. An object turning at 1 radian per second completes a full revolution in just less than 6.3 seconds.
- radian per second² (1/s²)** The SI unit of angular acceleration.
- radiation** The transmission of heat through the passage of electromagnetic radiation between objects.
- radiation trapping** The phenomenon in which a particular wavelength of light has trouble propagating through a material that eagerly absorbs and emits it. The light passes from one atom or atomlike system to the next and makes little headway.
- radiative transition** The shift of an atom or atomlike system from one state to another through the emission or absorption of an electromagnetic wave.
- radio waves** Electromagnetic waves, usually with wavelengths longer than about 1 m.
- radioactive decay** The spontaneous decay of a nucleus into fragments.
- Rayleigh scattering** The redirection of light due to its interaction with small particles of matter.
- real image** A pattern of light, projected in space, that exactly reproduces the pattern of light at the surface of the original object. A real image forms after the lens that creates it and can be projected onto a surface.
- rebound energy** The amount of kinetic energy returned to two objects as they push apart following a collision.
- reflection** The redirection of all or part of a wave so that it returns from a boundary between media.
- refraction** The bending of a wave's path that occurs when the wave crosses a boundary between media and experiences a change in speed.
- relative humidity** The actual humidity as a percentage of the humidity required to achieve phase equilibrium between liquid and gaseous water.
- relative motion** The movement of one object from the perspective of another object. Two objects that are moving relative to one another have different velocities.
- relativistic energy** An object's energy according to the relativistic laws of motion and including its rest energy.
- relativistic laws of motion** The laws of motion in the special theory of relativity. They correct deficiencies in the Newtonian laws of motion that appear primarily at speeds comparable to the speed of light.
- relativistic momentum** An object's momentum according to the relativistic laws of motion.
- resistor** An electronic component that impedes the flow of electric current, converting some of its energy into heat.
- resonant cavity** A simple resonant circuit consisting of a carefully shaped conducting strip or shell and equivalent to a capacitor and an inductor. Energy flows back and forth between the cavity's electric and magnetic fields.
- resonant energy transfer** The gradual transfer of energy to or from a natural resonance caused by small forces timed to coincide with a particular part of each oscillatory cycle.
- restoring force** A force that acts to return an object to its equilibrium shape. A restoring force is directed toward

the position the object occupies when it's in its equilibrium shape.

reverse biased A p-n junction in which the voltage of the p-type semiconductor has been lowered relative to the voltage of the n-type semiconductor.

Reynolds number A dimensionless number that characterizes fluid flow through a system. At low Reynolds numbers, a fluid's viscosity dominates the flow, while at high Reynolds numbers, a fluid's inertia dominates.

right-hand rule The convention that establishes the specific direction of an object's angular velocity. According to this rule, if the fingers of your right hand are curled to point in the direction of the object's rotation, your thumb will point in the direction of the angular velocity.

root mean square (RMS) voltage A measure of AC voltage defined as the DC voltage that would cause the same average power consumption in an ohmic device.

rotational equilibrium The state of an object in which zero net torque acts on it. An object that has constant angular momentum is in rotational equilibrium.

rotational inertia A property of matter by which it remains at rest or in steady rotation about the same rotational axis unless acted on by some outside torque.

rotational mass The property of a body that is a measure of its rotational inertia. An object's rotational mass is determined by its mass and by how far that mass is from the axis of rotation. The SI unit of rotational mass is the kilogram-meter².

rotational motion Motion in which an object rotates about an axis. The orientation of an object undergoing only rotational motion will change, but its position will remain unchanged.

saturated In phase equilibrium with another material phase. The gaseous phase of a material is saturated when it is in phase equilibrium with that material's liquid and/or solid phase.

scalar quantity A quantity, characterizing some aspect of a physical system, that consists only of a magnitude. It has no direction in space.

second (s) The SI unit of time. (One second is formally defined as the duration of 9,192,631,770 periods of the radiation corresponding to the transition between two hyperfine levels of the ground state of the cesium 133 atom.)

second law of thermodynamics The law of entropy.

semiconductor An insulator with a small band gap, so only a modest amount of energy is needed to shift an electron from an occupied valence level to an unoccupied conduction level.

series circuit (wiring arrangement) An arrangement in which the current reaching two or more electric devices flows sequentially through one device after the next before leaving them. Current may experience different changes in voltage in the different devices.

shell A group of atomic orbitals having similar energies.

shock wave A narrow region of high pressure and temperature that forms when the speed of an object through a medium exceeds the speed at which sound, waves, or other vibrations travel in that medium.

short circuit A defect in a circuit that allows current to bypass the load it's supposed to operate.

SI units (Système Internationale d'Unités) A system of units that carefully defines related units according to powers of 10. SI units are now used almost exclusively throughout most of the world, with the notable exception of the United States.

signal An electrical or optical representation of information.

simple harmonic motion The regular, repetitive motion of a harmonic oscillator. The period of simple harmonic motion doesn't depend on the amplitude of oscillation.

sliding friction The forces that resist relative motion as two touching surfaces slide across one another.

soft Having a small spring constant and thus experiencing small restoring forces in response to large distortions.

soft magnetic material A material that is relatively easy to magnetize and that loses its magnetization once the

magnetizing field is removed. Soft magnetic materials are suitable for electromagnets.

solid A form of matter consisting of particles (atoms or molecules) that touch and that are not free to move relative to one another. A solid has a fixed volume and shape.

sound In air, sound consists of density waves, patterns of compressions and rarefactions that travel outward from their source at the speed of sound.

space The three spatial dimensions in our universe that separate events from one another according to distances and directions.

special theory of relativity The physical rules governing all motion, even motion involving speeds comparable to the speed of light.

specific heat The amount of heat that must be added to a 1-unit mass of a material to cause a 1-unit rise in its temperature. The SI unit of specific heat is the joule per kilogram-kelvin.

speed A measure of the distance an object travels in a certain amount of time. The SI unit of speed is the meter per second.

speed of light The speed with which an electromagnetic wave travels through space. In empty space, a vacuum, the speed of light is exactly 299,792,458 m/s.

speed of sound The speed at which sound's compressions and rarefactions travel in a medium such as air or water.

spontaneous emission of radiation Light emission that occurs when an excited atom or atomlike system releases stored energy randomly through a radiative transition. The photon that results is independent and unique.

spring constant A measure of the stiffness of an elastic object; the spring constant relates the object's distortion to the restoring force it exerts. The larger the spring constant, the stiffer the spring.

springlike force A force that is proportional to displacement, consistent with Hooke's law.

sputtering Ejection of atoms from a surface caused by the impact of

- energetic ions, atoms, or other tiny projectiles.
- stable equilibrium** A state of equilibrium to which an object will tend to return if it's disturbed. At equilibrium, the object is free of net force or torque. If the object is moved away from its stable equilibrium state, however, the net force or torque that will then act on it will tend to return it to that equilibrium state.
- stall** Condition in which a fluid flow stops and spoils steady-state flow. In the aerodynamic flow around an airfoil, stalling refers to airflow separation triggered by a stall in the flow near the airfoil's surface.
- standard units** Agreed-on amounts of various physical quantities that define a system in which those quantities are subsequently measured.
- standing wave** A wave in which all the nodes and antinodes remain in place.
- state** A possible arrangement of electrons (or other particles) in a quantum system.
- static friction** The forces that resist relative motion as outside forces try to make two touching surfaces begin to slide across one another.
- static stability** An object's stability when it's not in motion.
- static variation** Change in a physical quantity such as pressure that is not caused by motion.
- steady-state flow** A situation in a fluid where the characteristics of the fluid at any fixed point in space don't change with time.
- Stefan–Boltzmann constant** The constant of proportionality relating a surface's radiated power to its emissivity, temperature, and surface area. It has a measured value of $5.67 \times 10^{-8} \text{ J/(s} \cdot \text{m}^2 \cdot \text{K}^4)$.
- Stefan–Boltzmann law** The equation relating a surface's radiated power to its emissivity, temperature, and surface area.
- stiffness** A measure of how rapidly a restoring force increases as the system exerting that force is distorted.
- stimulated emission of radiation** Light emission that occurs when an excited atom or atomlike system releases stored energy through a radiative transition by duplicating a photon passing through that system.
- streamline** The path followed by a particular portion of a flowing fluid.
- streamlined** Carefully tapered so that the fluid flowing around it doesn't stall and doesn't experience flow separation or pressure drag.
- strong force** The fundamental force that gives structure to nuclei and nucleons and is the basis for the nuclear force.
- subatomic particles** The fundamental building blocks of the universe, from among which atoms and matter are constructed.
- subcritical mass** A portion of fissionable material that is too small to sustain a chain reaction.
- sublimation** The process by which atoms or molecules go directly from a solid to a gas.
- superconductor** An electrical conductor that permits electrons to flow without losing any of their kinetic energy to thermal energy. Electrons will continue to flow in a superconductor indefinitely. Materials become superconducting only at extremely low temperatures.
- supercritical mass** A portion of fissionable material that is well in excess of a critical mass so that it undergoes an explosive chain reaction.
- superheated** Above the temperature at which a phase transition should have occurred. Superheating results from a failure to nucleate the new phase.
- superposition** The overlapping of two or more waves so that their amplitudes add together and they form a combined wave.
- support force** A force that is exerted when two objects come into contact. Each object exerts a force on the other object to keep the two from passing through one another. Support forces are always normal, or perpendicular, to the surfaces of objects.
- surface area** The extent of a two-dimensional surface bounded by a particular border. The SI unit of surface area is the meter².
- surface waves** Disturbances in the stable equilibrium shape of a surface.
- sympathetic vibration** The transfer of energy between two natural resonances that share a common frequency of oscillation.
- tank circuit** A simple resonant circuit consisting of a capacitor and an inductor. Energy flows back and forth between these two devices repetitively.
- temperature** A measure of the average internal kinetic energy per particle in a material. In a gas, temperature measures the average kinetic energy of each atom or molecule.
- tension** Outward forces on an object that tend to stretch it.
- terminal velocity** The velocity at which an object moving through a fluid experiences enough drag force to balance the other forces on it and keep it from accelerating.
- tesla (T)** The SI unit of magnetic field (synonymous with newton per ampere-meter).
- test charge** An idealized positive charge that has no electric field of its own and thus no influence on its surroundings.
- thermal conductivity** The measure of a material's capacity to transport heat by conduction from its hotter end to its colder end.
- thermal energy** A disordered form of energy contained in the kinetic and potential energies of the individual atoms and molecules that make up a substance. Because of its random distribution, this disordered energy can't be converted directly into useful work. Other names for thermal energy include internal energy and heat.
- thermal equilibrium** A situation in which no heat flows in a system because all the objects in the system are at the same temperature.
- thermal motion** The random motions of individual particles in a material due to the internal or thermal energy of that material.
- thrust** A forward, propulsive force obtained when a rocket pushes stored fuel backward; a forward propulsive force obtained when a propeller, jet engine, or other device pushes surrounding fluid or material backward.

tidal forces The differences between one celestial object's gravity at particular locations on the surface of a second object and the average of that gravity for the entire second object. Tidal forces tend to stretch the second object into an egg shape.

timbre The mixture of tones in an instrument's sound that are characteristic of that instrument.

torque An influence that if exerted on a free body results chiefly in an angular acceleration of the body. A torque is a vector quantity, consisting of both the amount of torque and its direction. The SI unit of torque is the newton-meter.

total internal reflection Complete reflection of a light wave that occurs when that wave tries unsuccessfully to leave a material with a large refractive index for a material with a small refractive index at too shallow an angle.

traction The largest frictional force that an object can obtain in its present situation.

trajectory The path taken by an object as it moves.

transformer A device that uses magnetic fields to transfer electric power from one circuit to another circuit. The two circuits are electrically isolated since no charges actually travel between the two circuits.

transistor An electronic component that allows a tiny amount of electric charge, either moving or stationary, to control the flow of a large electric current.

translational motion Motion in which an object moves as a whole along a straight or curved line.

transmutation of elements Changing the atoms of one element into another via nuclear processes that alter the numbers of protons in their nuclei.

transverse wave A wave in which the underlying oscillation is perpendicular to the wave itself.

traveling wave A wave that moves steadily through space in a particular direction.

trough A peak negative excursion of an extended system that is experiencing a wave.

tunneling The quantum process in which small objects (which, because of the Heisenberg uncertainty principle, have somewhat ill-defined positions) occasionally move through energy barriers to places they can't reach classically.

turbulence The unpredictable swirls and eddies of turbulent fluid flow.

turbulent flow Irregular, fluctuating, unpredictable fluid flow in which nearby portions of the fluid quickly become widely separated.

ultraviolet light Invisible light having wavelengths shorter than about 400 nanometers.

uniform circular motion Motion at a constant speed around a circular trajectory. An object undergoing uniform circular motion is accelerating toward the center of the circle.

unstable equilibrium A state of equilibrium to which the object will tend not to return if it's disturbed. At equilibrium, the object is free of net force or torque. If the object is moved away from its unstable equilibrium state, however, the net force or torque that will then act on it will tend to accelerate it further away from that equilibrium state.

valence band The group of quantum levels in an insulator that lies below the Fermi level.

valence level A quantum level in an insulator that requires less energy than the Fermi level and that is normally occupied by electrons.

vector field A structure that associates a vector quantity with each point in space and time.

vector quantity A quantity, characterizing some aspect of a physical system, that consists of both a magnitude and a direction in space.

velocity A vector quantity that measures the rate at which an object's position is changing: the greater the velocity, the farther the object travels each second. It consists of both the object's speed and the direction in which the object is traveling. The SI unit of velocity is the meter per second.

vertical polarization An electromagnetic wave in which the electric

field always points up or down (vertically). The magnetic field always points horizontally.

vibration A spontaneous repetitive and rhythmic movement about an equilibrium position.

vibrational antinode A region of a vibrating object that is experiencing maximal motion.

vibrational node A region of a vibrating object that is not moving at all.

virtual image A pattern of light that appears to come from a particular region of space and reproduces the pattern of light at the surface of the original object. A virtual image forms before the lens that creates it and can't be projected onto a surface.

viscosity The measure of a fluid's resistance to relative motion within that fluid.

viscous drag A drag force that results from viscous forces on a moving surface immersed in a fluid.

viscous forces The forces exerted within a fluid that oppose relative motion. Layers of fluid that are moving across one another exert viscous forces on each other.

visible light Light having wavelengths between about 400 nanometers (violet) and 750 nanometers (red). This small portion of the electromagnetic spectrum is all that we are able to detect with our eyes.

volt (V) The SI unit of voltage (synonymous with joule per coulomb). The voltage on the positive terminal of a common battery is about 1.5 volts above that on its negative terminal.

volt per meter (V/m) The SI unit of electric field (synonymous with newton per coulomb).

voltage drop The amount of electrostatic potential energy that each coulomb of positive charge loses in passing through a device. It's equal to the voltage of the charges entering the device minus the voltage of the charges leaving that device.

voltage gradient A gradual slope in the voltage across a region of space. A voltage gradient is an electric field.

voltage rise The amount of electrostatic potential energy that each coulomb

of positive charge receives in passing through a device. It's equal to the voltage of the charges leaving the device minus the voltage of the charges entering that device.

volume The extent of a three-dimensional region of space bounded by a particular enclosure. The SI unit of volume is the meter³.

vortex A whirling region of fluid that is moving in a circle above a central cavity.

wake The trail left behind by an object as it moves through a fluid.

wake deflection force A lift force experienced by a spinning ball when it deflects its turbulent wake to one side. The wake deflection force points toward the side of the ball moving away from the onrushing airstream.

water hammer The impact of a moving mass of water that is suddenly stopped.

watt (W) The SI unit of power, equal to the transfer of 1 joule per second. One watt is the power used by the bulb of a typical flashlight.

wave velocity The speed and direction of the moving crests of a wave.

wavelength A structural characteristic of a wave, corresponding to the distance separating adjacent crests or troughs.

wave-particle duality The observation that everything in nature has both particle and wave characteristics. An item is primarily particle-like when it is emitted, absorbed, or otherwise observed and primarily wavelike as it travels through time and space.

weak force The fundamental force that allows electrons and neutrinos to interact and that's responsible for beta decay.

weight (near Earth's surface) The downward force exerted on an object due to its gravitational interaction with

Earth. An object's weight is equal to that object's mass times the acceleration due to gravity. The direction of the weight is always toward the center of Earth.

work The mechanical means of transferring energy. Work is defined as the force exerted on an object times the distance that object travels in the direction of the force. A large force exerted for a short distance or a small force exerted for a long distance can perform the same amount of work. The SI unit of work is the joule.

X-ray fluorescence The process in which an electron in one of the outer orbitals of an atom undergoes a radiative transition to an empty inner orbital, emitting an X-ray photon.

X-rays Very high-energy photons of electromagnetic radiation.

zeroth law of thermodynamics The law of thermal equilibrium.

Index

A

Absolute temperature scales, 123
Absolute zero, 123
Acceleration
 angular, 40
 carousels and, 86–88
 centripetal, 87
 deceleration and, 6
 defined, 5, 8
 due to gravity, 13
 examples, 6
 experience of, 89–90
 feeling of, 90
 magnitude, 5
 net force and, 8
 in Newton's second law of motion, 7
 potential energy and, 69
 roller coaster, 91
 SI unit of, 11
Access ramps, 31
Achromat, 399
Action and reaction, 23
Activation energy, 178
Active learning experiments
 Cartesian diver, 119–120, 140–141
 disc in microwave oven, 332–333, 351
 fog in a bottle, 208, 228
 high-flying balls, 96, 117
 magnifying glass camera, 392–393, 422–423
 moving water without touching, 266–267, 300
 nail and wire electromagnet, 302–303, 330
 radiation-damaged paper, 425, 458
 removing a tablecloth from a table, 1–2, 31
 ruler thermometer, 173, 206
 singing wineglass, 230–231, 263–264
 spinning pie dish, 33, 70
 splitting colors of sunlight, 353–354, 390–391
 swinging water overhead, 72, 94
 vortex cannon, 142–143, 171
Adverse pressure gradient, 155
Aerodynamic forces, 153
Aerodynamics, 153
Aiming high, 21
Air
 circulating, 169
 as compressible, 121
 as gas, 121
 heat moving with, 180
 sound in, 250–252
 thermal motion, 121

vibrating, 247–248
volumes, 121
Air conditioners
 chlorofluorocarbons, 219
 compressor and condenser coils, 218
 cooling of indoor air, 216–217
 experiments, 209
 how they work summary, 228–229
 ideal efficiency and, 215
 overview, 209
 pumping heat against it's natural flow, 214–215
 thermodynamics and, 209–212
 warming of outdoor air, 217–219
 window, in middle of room, 219
Air ducts, 146
Air guitar, 253
Airfoil, 161
Airplanes
 experiments, 161
 how they work summary, 172
 jet engines, 169–171
 lift production, 162–165
 overview, 161
 propellers, 167–169
 stalling a wing, 165–167
 streamlined wing, 161–162
 stunt flying, 166
 wings, 161–165
Airspeed, 154
Alpha decay, 433
Alternating current
 coil of wire and, 17–18
 defined, 315
 electric motors, 328–329
 generators, 326, 328–329
 outlets, 316
 power distribution, 325–327
 power reversals, 316
Aluminum, X-rays and, 452
Ampere, 289
Ampère, André-Marie, 312
Amplifiers
 audio, 419–422
 defined, 420
 feedback, 420
 input wires, 421
 MOSFET-based, 422
 rating, 421
 voltage, 421
Amplitude, harmonic oscillator, 234
Amplitude modulation (AM), 340, 342
Analog representation, 404
AND gates, 419
Angle of attack, 164–165

Angular acceleration, 40
Angular impulses
 angular momentum and, 66
 defined, 65, 66
 equation, 65
Angular momentum
 angular impulse and, 66
 defined, 64, 66
 illustrated, 65
 SI unit, 64
Angular position, 36
Angular speed, 36
Angular velocity, 36
Anharmonic oscillator, 237
Anode, 384
Antennas
 halfwave dipole, 339
 quarter-wave monopole, 339
 receiver, 337–338
 straight, 338
 transmitter, 337, 339
Antimatter, 453
Antireflection coating, 399
Aperture, 396
Apogee, 110
Apparent weight, 90
Aquarium, 412
Archimedes' principle, 126
Aristotle, 3
Atkinson cycle engine, 224
Atmosphere, Earth's, 124–125
Atmospheric pressure, 125
Atomic bomb, 427
Atomic number, 369
Atoms
 assembling, 369–370
 collision, 176
 defined, 121
 electron standing waves in, 367–368
 electrons, 269, 368
 forces between, 175–176
 neutrons, 269
 nucleus, 269, 428, 430
 protons, 269
Audio players
 audio amplifier, 419–422
 computer, 418–419
 digital sound information storage, 416–417
 experiments, 413–414
 how they work summary, 423–424
 overview, 413
 transistors, 414–415
Audio signals, 419
Audio speakers, 350

Automobiles
 batteries, 291, 293
 diesel engines, 226
 electric defrosters, 298
 engine efficiency, 223–224
 engine efficiency, improving, 224–225
 experiments, 220
 heat engines, 220–221
 how they work summary, 229
 internal combustion engine, 222–223
 multicylinder engines, 227–228
 overview, 219
 power, cutting, 289
 Axis of rotation, 37

B

Back emf, 319
 Balance clocks, 237–239
 Balanced forces, 25
 Balanced objects, 44
 Balanced seesaw, 44–45
 Ball sports (air)
 curveballs and knuckleballs and, 159–160
 experiments, 153
 fast moving ball, 155–157
 golf ball dimples and, 157–159
 how they work summary, 171–172
 overview, 153
 slow moving ball, 153–155
 Ballooning weather, 128
 Balloons
 air/air pressure and, 121–122
 experiments, 120
 helium, 129–130
 hot-air, 127–128
 how they work summary, 141
 lifting force on, 125–127
 overview of, 120
 what not to put in, 130
 Balls
 bouncing, 79
 energy ratios, 81
 falling, 12–21
 golf, 157–159
 high-flying, 96, 117
 kinetic energy, 82
 positively charged, 271
 speed ratios, 81, 82
 thrown, movement of, 19
 tossing upward, 18–19
 Band gaps, 379
 Bandwidth, 341
 Banked curves, 88
 Banneker, Benjamin, 239
 Bardeen, John, 414
 Base of support, 99
 Bathroom scales, 74, 78–79
 Batteries
 alkaline, 291
 car, 291, 293
 flashlight, 290–291
 perspectives on, 290
 radio, 294

Beam waist, 410
 Becquerel, Antoine-Henri, 431
 Bell, Jocelyn, 233
 Bernoulli, Daniel, 135
 Bernoulli's equation, 135, 138, 152
 Beta decay, 431, 454
 Bicycle pedaling, 57
 Bicycles
 dynamic stability and, 100–101
 exercise, 212, 329
 experiments, 97
 how they work summary, 117
 leaning while turning, 101–102
 overview, 97
 pedaling, 102–104
 slicing through air, 162
 tricycles and, 97–99
 Binary, 404–405
 Black holes, 116
 Blackbody spectrum, 197
 Blacksmith, 199
 Blending cold water, 211
 Blu-ray players, 408–410
 Boats, shifting cargo, 45
 Boiling
 defined, 190
 temperature, 190
 temperature, changing, 190–191
 water, 189–191
 Boltzmann constant, 130
 Bottle, blowing across, 248, 249
 Bouncing balls
 bats and, 85–86
 coefficient of restitution, 82
 elastic/inelastic collisions, 81
 experiments, 79
 how they work summary, 95
 lively, 81
 moving surface and, 83
 overview, 79
 shape, 79
 surface and, 82–83
 Boundary layer, 153–155
 Bowing strings, 245–246
 Bowling balls, 9
 Brahe, Tycho, 111
 Brattain, Walter, 414
 Breath, seeing, 189
 Bremsstrahlung, 449
 Brewster's angle, 362–363
 Bricks, 195
 Bulbs, flashlight
 choosing, 295–297
 filament, 292–293
 filament and current flow, 296
 voltage drop, 295
 Bumper cars
 conserved quantities, 67
 exchanging momentum in collisions, 62–63
 experiments, 59–60
 glancing blows, 65–66
 how they work summary, 71
 linear momentum, 60–61

overview of, 59
 spinning in circles, 64–65
 Bungee jumping, 18
 Buoyant force, 125–126
 Burning rubber, 53

C

Calculus, 17
 Calibration, spring scales, 77–78
 Cameras
 depth of focus, 396
 experiments, 393
 eyes and, 402–403
 f-number, 398
 focal length, 396–397
 focusing, 395–396
 how they work summary, 423
 image sensors, 401–402
 lens quality, 399
 lenses and real images, 394–395
 overview, 393
 sunny versus overcast days and, 398
 viewfinder, 399–401

Candlelight, 355
 Canoeing, 156
 Capacitance, 287
 Capacitors, 421
 Capillary waves, 256
 Carbohydrates, 178
 Carbon, 430
 Cardboard, cutting up, 43
 Carlson, Chester F., 277
 Carousels, 86–89, 95
 Carrier frequency, 341
 Cartesian diver, 119–120, 140–141
 Cathode, 384
 Cavendish, Henry, 108
 CDs
 aluminum layer, 407
 density measurements, 405
 encoding schemes, 406
 player optical system, 408–410

Celsius, 123, 177
 Center of gravity, 44, 45
 Center of mass, 38, 45
 Center of percussion, 85
 Center of rotation, 38
 Centripetal acceleration, 87
 Centripetal force, 87, 88
 Chadwick James, 431
 Chain reactions, 432–433
 Chandelier oscillation, 239
 Chaotic system, 152
 Characteristic X-ray, 449–450
 Charge
 on conducting object, 282
 defined, 268
 electric fields and, 283
 elementary unit of, 269
 net, 269
 quantized, 269
 separated, 287
 SI unit, 268

- transferring, 271–272
voltage gradients and, 283
- Chemical bonds, 121, 175–176
- Chemical potential energy, 175–176, 271
- Chemical reactions, 178
- Chernobyl reactor, 444–445
- Chlorofluorocarbons, 219
- Chromatic aberration, 399
- Circular motion, 87
- Circularly polarized, 409
- Classified physics, 426
- Cleaning house, 150
- Clocks
- balance, 237–239
 - egg-timer, 233
 - electronic, 239–240
 - experiments, 231
 - how they work summary, 264
 - natural resonance, 232–233
 - overview, 231
 - pendulum, 233–237
 - time and, 232
- Closed circuits, 288
- CMOS NAND gate and inverter, 418–419
- Coasting, 2, 3–4
- Coaxial cable, 342
- Coefficient of restitution, 82
- Coefficient of volume expansion, 204
- Coffee mug shape, 99
- Coherent light, 387
- Coin toss, 19
- Collision energy, 80
- Collisions
- of atoms, 176
 - elastic, 81
 - exchanging momentum in, 62–63
 - inelastic, 81
 - speed, 84
- Color temperature, 198
- Colors
- eye recognition of, 364–365
 - firework production of, 372
 - mixture of, 365
 - primary additive, 364–365
 - primary subtractive, 364–365
- Compasses, 309–310
- Components
- defined, 19
 - forward, 19, 20
 - upward, 19–20
- Compressible, 121
- Compton scattering, 451, 453
- Computed tomography (CT) scanner, 451–452
- Condensation, 187
- Conduction
- defined, 179
 - electrical, 275
 - thermal, 193–195
- Conduction band, 380
- Conduction level, 380
- Cone cells, 364
- Conserved quantities
- angular momentum, 67
 - defined, 25
 - energy, 67
 - momentum, 62, 63, 67
- Constructive interference, 263, 360
- Convection
- defined, 180
 - staying warm by impeding, 195–196
- Convection cell, 180
- Convection current, 180
- Converging lens, 394
- Conveyor belts, 276
- Cooling, 216–217
- Cordless telephones, 339
- Corona discharge, 278–279, 284–285
- Coulomb, 268
- Coulomb constant, 270, 334
- Coulomb's law, 270
- Coulomb's law for magnetic poles, 305
- Credit cards, 311
- Critical mass, 433
- Crocodile floating, 127
- Crystalline, 184
- Curie, Marie, 431
- Curie, Pierre, 431
- Current
- alternating, 315–316
 - defined, 288, 289
 - direct, 314–315
 - direction, 289
 - eddy, 325
 - in flashlights, 289–290, 293–294
 - load, 321–322
 - magnetizing, 320–321
 - secondary, 323
 - SI unit, 289
- Curveballs, 159–160
- Cycle per second, 240
- Cyclotron motion, 350
- D**
- Davisson, Clinton Joseph, 367
- de Cisternay du Fay, Charles-François, 268
- de Coulomb, Charles-Augustin, 269
- de Laval, Carl Gustaf, 106
- de Laval nozzle, 106
- Deceleration, 5
- Decimal, 404
- Degree of freedom, 182
- Demagnetization, 308
- Density
- air pressure and, 122
 - particle, 129
 - SI unit, 123
- Density gradients, 125
- Depletion region, 382
- Deposition, 189
- Depth of focus, 396
- Destructive interference, 263
- Deuterium, 446
- Diesel, Rudolph Christian Karl, 226
- Diesel engines, 226–227
- Diffraction, 409, 410
- Digital representation, 404
- Digital-to-analog converter (DAC), 419
- Diodes
- defined, 381
 - forward biased, 384
 - laser, 389–390
 - light-emitting, 384–385
 - photodiodes, 401–402
 - reverse biased, 384
- Direct current (DC), 314–315
- Direction, 3
- Disc in microwave oven, 332–333, 351
- Discharge lamps
- defined, 365–366
 - experiments, 364
 - fluorescent, 373–375
 - high-pressure, 375–377
 - how they work summary, 391
 - mercury, metal-halide, and sodium, 375–377
 - neon, 371–372
 - overview, 363
 - plasma, 375
 - radiation trapping, 375
- Discharges
- corona, 278–279, 284–285
 - defined, 279
 - gas, 365–366
- Dispersion
- defined, 259, 263
 - refraction and, 359
- Distance, 3
- Diverging lens, 403
- Domain walls, 308
- Doppler effect, 252
- Drag
- induced, 163
 - pressure, 155, 156
 - propeller, 168
 - viscous, 154
- Drag forces, 153
- Drinking fountains, 218
- Drop tower, 91
- Drums, 249–250, 252–253
- DVDs
- encoding schemes, 406
 - layers, 407
 - player optical system, 408–410
 - structure of, 407–408
 - technology, 405–406
- Dynamic memory, 417
- Dynamic pressure variations, 147–148
- Dynamic stability, 100
- Dynamic variation, 132
- E**
- Earth
- atmosphere, 124–125
 - greenhouse effect, 205
 - magnetic poles, 309–310
 - moon distance measurement, 232
 - orbiting, 108–111
 - sunlight passage to, 356–357
 - temperature, 204–205
 - tidal forces, 254–255

Eddy currents, 325
 Eiffel, Gustave, 156
 Einstein, Albert, 114, 116, 386, 427
 Elastic collisions, 81
 Elastic limit, 76
 Elastic potential energy, 76
 Elastic scattering, 451
 Electric charge. *See Charge*
 Electric circuits
 closed, 288
 current, 289
 defined, 288
 open, 288
 parallel, 298, 300
 short, 288
 Electric current. *See Current*
 Electric defrosters, 298
 Electric doorbell, 311–313
 Electric fields
 in conducting object, 282
 defined, 279
 flashlight light, 387
 fluorescent lamp, 283
 microwave, 344
 in microwave oven, 346
 of motionless positive charge, 280
 SI unit, 280, 283
 sources of, 334
 Electric phonographs, 317
 Electric polarization, 270
 Electric power distribution
 AC electric generators and motors, 328–329
 alternating current, 315–316, 325–327
 alternating current and coil of wire, 318–319
 direct current, 314–315
 experiments, 313
 how it works summary, 330
 magnetic induction, 317
 overview, 313
 transformers, 320–325
 Electrical conductors, 275
 Electrical insulators, 275
 Electrical resistance
 defined, 288
 filament, 296
 SI unit, 296
 skin, 297
 Electricity, 266–301
 high voltages, 272–273
 magnetism and, 334
 static, 267–276
 Electrochemical emf, 318
 Electromagnetic spectrum, 197
 Electromagnetic waves
 defined, 180, 335
 impedance mismatch, 358
 propagation, 342
 spectrum of, 342
 sunlight and, 354–355
 superposition of, 360
 types of, 181

undulations and, 338
 wavelengths, 197, 343–344
 Electromagnets, 312
 Electron standing waves, 367–368
 Electron volt, 384
 Electronic clocks, 239–240
 Electron-positron pair production, 451, 453
 Electrons
 adding to floating gate, 417
 arrangement in atoms, 370
 defined, 179
 electric charge and, 269
 Fermi sea of, 378
 medical, 281
 migration, 382
 as orbital, 368
 in solids, 378–379
 transfer of, 272
 Electrostatic forces, 268, 270
 Electrostatic potential energy, 272–273
 Elementary unit of electric charge, 269
 Elevator, 25
 Energy
 activation, 178
 collision, 80
 as conserved, 25, 27
 defined, 25, 27
 Fermi, 378
 friction and, 52–53
 gravitational potential, 26, 28–29
 internal, 53
 jet engines and, 171
 kinetic, 26, 27, 57–58
 law of conservation of, 210–211
 magnetic fields and, 319
 ordered, 51
 potential. *See* potential energy
 pressure and, 134–136
 rebound, 80
 relativistic, 115
 thermal, 51–52, 174–175
 Energy ratios, 81–82
 Engines. *See also* Automobiles
 Atkinson cycle, 224
 cold, starting, 294
 compression stroke, 223
 diesel, 226–227
 efficiency, 223–224
 efficiency, improving, 264–265
 four-stroke, 222
 fuel-injected, 222
 heat, 220–221
 ignition temperatures for grades of gasoline, 225
 intercoolers, 226
 internal combustion, 222–223
 multicylinder, 227–228
 power stroke, 223, 224
 premium fuel and, 225
 steam, 222
 supercharged, 226
 turbocharged, 226
 warm, starting, 295
 English system of units, 10–12
 Enhanced radiation bomb, 435–436
 Enriched uranium, 440–442
 Entropy
 defined, 212
 disorder and, 212–213
 law of, 213
 redistribution, 214
 time and, 232
 Equilibrium
 defined, 75
 motionlessness and, 79
 neutral, 99
 phase, 186–187
 position, 75
 rotational, 98–99
 shape, 80
 stable, 75, 98–99
 thermal, 176–177
 unstable, 99
 Erbium-doped fiber amplifiers (EDFAs), 412–413
 Escape velocity, 111
 Evaporation, 187
 Excited state, 371
 Exercise bicycles, 212, 329
 Exhaust velocity, 106
 Explosive chain reaction, 433
 Exponential decay, 429
 Eyeglasses, 402–403
 Eyes, 402

F

Fahrenheit, 123, 177
 Falling balls
 acceleration, 13
 experiments, 12
 how they work summary, 32
 overview, 12
 position of, 16–17
 projectile motion, 19–21
 tossing ball upward and, 18–19
 velocity of, 15–16
 weight and gravity and, 13–14
 Farad, 287
 Faraday, Michael, 317
 Fast fission reactors, 443
 Feeling of acceleration, 90
 Fermi, Enrico, 431, 441
 Fermi energy, 378
 Fermi level
 defined, 378
 empty, 379
 in insulators, 380
 in metals, 379
 Fermi particles, 369
 Ferromagnetic material, 307
 Field-effect transistor, 414, 415
 Fields, 279. *See also* Electric fields
 File cabinet, moving, 49–50
 Fingers in electrical outlet, 316
 Fireworks, 372
 Firm springs, 76
 First law of thermodynamics, 210–211

- Fission, nuclear, 431
 Fission bomb, 432–435
 Fission reactors
 Chernobyl, 444–445
 fast, 443
 Fukushima Daiichi, 445
 nuclear, 438–439
 safety and accidents, 443
 thermal, 440–442
 Three Mile Island, 444
 Windscale, 444
- Fissionable, 432
- Flash memory, 417
- Flashlights
 batteries, 290–291
 bulbs, 292–293, 295–297
 current, 293–294
 electric circuit, 288
 electric current, 288, 289–290
 electric field, 387
 electricity and, 287–289
 experiments, 287
 how they work summary, 301
 LED, 297–299
 metal strips, 293
 overview, 287
 power, 293–294
 voltage, 293–294
- Flow
 in bent hose, 147–149
 heat, 194, 214
 laminar, 150, 151, 153–155
 smooth, in a stream, 155
 in straight hose, 145–147
 through nozzle, 149–150
 turbulent, 150, 151, 155–157
- Fluids
 analysis of, 119
 defined, 125
 Earth's atmosphere as, 125
 motion and, 142–172
 speed and pressure in, 149
 steady-state flow, 135
 viscosities of, 144
- Fluorescence, 373
- Fluorescent lamps
 defined, 373
 early, 373–374
 phosphor blends, 374
 plasma, 375
 practical issues, 374–375
 sputtering, 375
- Flux lines, magnetic, 311, 312
- F-numbers, 398
- Focal lengths, 396–397
- Focusing, 395–396
- Fog in a bottle, 208, 228
- Foil wrapping, 201
- Food labeling, 74
- Forces
 aerodynamic, 153
 affect on skaters, 7–8
 between atoms, 175–176
 balanced, 25
 buoyant, 125–126
 centripetal, 87, 88
 defined, 4
 drag, 153
 electrostatic, 268, 270
 equal, 8
 frictional, 50
 fundamental, 454
 impact, 62
 lift, 153
 Lorentz, 349–350, 447
 magnetostatic, 304, 309
 Magnus, 159
 momentum and, 63
 net, 7
 normal, 24
 nuclear, 428–429
 potential gradients and, 69
 restoring, 75
 SI unit of, 11
 springlike, 242
 strong, 454
 support, 23–24
 thrust, 105, 167
 tidal, 254–255
 torques and, 42–43
 viscous, 144, 145, 147
 wake, 159
 weak, 454
- Forward biased, 384
- Four-stroke engines, 222
- Frames of reference, 9–10
- Franklin, Benjamin, 268
- Free space, permeability of, 305
- Freezing, 186
- Frequency
 defined, 240
 SI unit, 240
 of sound, 241
 of waves, 258
- Frequency modulation (FM), 340–341
- Friction
 defined, 49
 energy and, 52–53
 microscopic view of, 50
 sliding, 51
 static, 51
 traction and, 51
 weight and, 51
- Frictional forces, 50
- Frisch, Otto, 431
- Frozen fingers, 177
- Fuel-injected engines, 222
- Fukushima Daiichi reactor, 445
- Fundamental forces, 454
- Fusion
 latent heat of, 186
 nuclear, 431
- Fusion bomb, 435–436
- G**
- Galilei, Galileo, 3
- Gamma rays, 453–454
- Garden watering
 dynamic pressure variations, 147–148
 experiments, 143–144
 flow in bent hose, 147–149
 flow in straight hose, 145–147
 flow through nozzle, 149–150
 how it works summary, 171
 overview, 143
 turbulence and, 150–152
- Gas discharges, 365–366
- Gaseous, 184
- Gases
 defined, 121
 helium, 130
 ideal gas law, 130
 inert, 121
 pressure of, 130
- General theory of relativity, 116
- Geostationary satellites, 110
- Germain, Sophie, 249
- Golf balls, dimples on, 157–158
- Gradients
 defined, 68
 density, 125
 potential, 68–69
 pressure, 125
 voltage, 282–283
- Gravitational constant, 109
- Gravitational mass, 116
- Gravitational potential energy
 defined, 26
 equation, 28
 in water distribution, 138
- Gravity
 acceleration due to, 13
 center of, 44–45
 defined, 13
 mass and, 14
 moving water and, 138–140
 pendulums and, 236
 thermal energy and, 205
 water pressure and, 136–138
- Gravity waves, 256
- Greenhouse effect, 205
- Ground state, 371
- H**
- Hahn, Otto, 431, 432
- Half a fall, 15, 17
- Half-life, 429
- Halfwave dipole antenna, 339
- Hanging grocery scales, 77–78
- Hard disks, 417
- Hard magnetic material, 308
- Hard to turn, 42
- Harmonic oscillators
 amplitude, 234
 balance clocks, 238
 defined, 234, 235
 importance of, 78
 pendulum, 235–236
 period, 234
 stiffness, 234
 violin strings as, 242

Harmonics, 245
 Harrison, John, 239
 Heat
 defined, 177
 flow, 194, 214
 latent, 186
 metal tray and cookies, 183
 movement from fire to room, 178
 moving as light, 180–182
 moving it around, 209–211
 moving through metal, 179
 moving with air, 180
 pumping against natural flow, 214–215
 SI unit, 182
 specific, 182–183
 steam, 227
 temperature and, 176–177
 wind and, 180
 Heat capacity, 182
 Heat engines, 220–221
 Heat exchanger, 178
 Heat packs, 179
 Heat pumps
 in appliances, 218–219
 in cold weather, 215
 defined, 219
 diagram, 214
 energy consumption, 216
 heat engines and, 221
 Heisenberg uncertainty principle, 429
 Helium balloons, 129–130, 252
 Henry, 336
 Hertz, 240
 Hertz, Heinrich Rudolph, 240
 High dive, 16, 39
 Higher-order vibrational modes, 244
 High-flying balls, 96, 117
 High-voltage wires, 328
 Hooke, Robert, 75
 Hooke's law, 75, 76
 Horizontal polarization, 337
 Hot potatoes, 194
 Hot-air balloons, 128–129
 Humphry, Davy, 317
 Hurricane kinetic energy, 59
 Hydrogen, 227
 Hydrogen bomb, 435–436
 Hydrogen bond, 344

I
 Ice. *See also* Water
 as crystalline, 185
 formation, 185
 melting, 185
 as solid, 184
 stuck on, 61
 subliming, 189
 turning on, 48
 Ice water, 187
 Ideal gas law, 130
 Image distance, 397
 Image sensors, 401
 Impedance, 358, 359

Impedance mismatch, 358, 359
 Impulses
 angular, 65–66
 defined, 62, 63
 equation, 62
 Incandescence, 371
 Incoherent light, 386
 Incompressible, 133
 Index of refraction, 356
 Induced drag, 163
 Induced emf, 318
 Induced fission, 431
 Inductance, 336
 Inductors, 318, 324
 Inelastic collisions, 81
 Inertia, 3, 36
 Inertial frame of reference, 9
 Inertial mass, 116
 Infrared light, 355
 Infrared sensors, 199
 Insert gases, 121
 Insulation. *See also* Warmth
 electrical, 275
 glass wool, 202–203
 house, 202–203
 how it works summary, 206–207
 importance of, 192–193
 materials, 202
 windows, 203–204
 Insulators, 380, 381
 Interference
 constructive, 263, 360
 defined, 262, 263
 destructive, 263, 360
 pattern, 263–264
 wave, 262–263
 Internal combustion engines, 222–223
 Internal energy, 53
 Internal kinetic energy, 123
 Internal potential energy, 123
 Iron filings, 311
 Isolation transformers, 323
 Isotopes, 432

J
 Jet engines. *See also* Airplanes
 energy and, 171
 ramjet, 170
 turbofan, 170
 turbojet, 169
 Jewel movements, 56
 Jones, David, 100
 Joule, 28
 Joule per kilogram-kelvin, 182
 Joule per second, 56
 Jump rope, 245

K
 Kelvin scale, 123, 177
 Kepler, Johannes, 111
 Kepler's first law, 111–112
 Kepler's second law, 112–113

Kepler's third law, 113
 Kilogram per meter³, 123
 Kilogram-meter per second, 60
 Kilogram-meter per second², 11
 Kilogram-meter², 39
 Kilogram-meter² per second, 64
 Kilograms, 11
 Kinetic energy. *See also* Energy
 balls, 82
 calculating, 57–58
 defined, 26, 27
 equation, 58
 internal, 123
 of rotational motion, 58
 thermal, 204
 Kneading energy, 57
 Knuckleballs, 159–160
 Kutta, M Wilhelm, 163

L
 Laminar flow, 150, 151, 153–155
 Laser amplifier, 387, 388
 Laser diodes, 389–390
 Laser oscillator, 387, 388
 Lasers
 beam, 387–388
 how they work summary, 391
 ideal system, 388
 laser light and, 386–387
 medium, 387, 388–390
 overview and experiments, 377
 population inversion, 389
 pumping, 389
 warning, 377
 Latent heat
 of evaporation, 187
 of fusion, 186
 of melting, 186
 of vaporization, 187
 Latimer, Lewis Howard, 314
 Law of conservation of energy, 210–211
 Law of entropy, 213
 Law of thermal equilibrium, 210
 Law of universal gravitation, 109
 Laws of thermodynamics, 210
 LEDs
 colors of, 384
 defined, 384
 flashlights, 297–299
 how they work summary, 391
 overview and experiments, 377
 series, 385
 as solid-state device, 378
 wavelengths, 385
 Lens equation, 397
 Lenses. *See also* Cameras
 antireflection coated, 399
 aperture, 396
 collimated, 409–410
 converging, 394
 defined, 394
 diameter, 395–396
 diverging, 403

- f-number, 398
 focal length, 396–397
 quality, improving, 399
 real images and, 394–395
 zoom, 399
- Lenz's law, 318
 Lever arm, 43
 Levers, 46
 Lift
 limits of, 165–167
 wing production of, 162–165
- Lift forces, 153
 Light, 353–391
 black objects and, 181
 candlelight, 355
 coherent, 387
 cool, 204
 heat moving as, 180–182
 how we see, 364–365
 incoherent, 386
 index of refraction and, 356
 infrared, 355
 laser, 386–387
 in optical fiber, 412, 413
 primary colors of, 364–365
 speed of, 114, 115–116
 spontaneous, 386
 stimulated, 386
 sunlight, 354–363
 travel, 356
 ultraviolet, 355
 visible, 354
 wavelengths of, 364
- Lightning, 273, 283, 286–287, 334
 Lightning rods, 285, 286
 Linear accelerator, 455–456
 Linear momentum, 60. *See also* Momentum
 Lint, 281
 Liquids, 184, 185
 Load current, 321–322
 Longitudinal wave, 248
 Loop-the-loops, roller coaster, 91–94
 Lorentz, Hendrik Antoon, 349
 Lorentz force
 in audio speakers, 350
 cyclotron motion, 350
 defined, 349
 equation, 349
 nuclear fusion reactors and, 447
 Love, William T., 315
 Lumen, 365
 Luminescence, 371
- M**
 Magnetic confinement fusion reactors, 447
 Magnetic cores, 325
 Magnetic dipoles, 304
 Magnetic domains, 307
 Magnetic fields
 defined, 309
 energy and, 319
 SI unit, 309
 sources of, 334
- Magnetic flux lines, 311, 312
 Magnetic induction, 316
 Magnetic monopoles, 304
 Magnetic polarization, 306
 Magnetic poles
 Coulomb's law for, 305
 defined, 304
 net, 304
 SI unit, 305
- Magnetic resonance imaging (MRI)
 defined, 456
 functioning of, 456–457
 illustrated, 457
 magnetic field, 313, 320
 protons and, 457–458
 spatial variation, 458
- Magnetism
 Coulomb's law for, 305
 defined, 304
 electricity and, 334
- Magnetizing current, 320–321
 Magnetostatic force, 304, 309
 Magnetron
 creating microwaves with, 347–348
 large, behavior of, 348
 magnet, 348
 powering, 348–349
- Magnets
 compasses and, 309–310
 dropping, 319
 electric doorbell, 311–313
 experiments, 303–304
 how they work summary, 330
 iron filings and, 311
 as limited energy sources
 misconception, 319
 magnetron, 348
 north pole, 309
 overview, 303
 permanent, 308
 plastic sheet, 308
 refrigerator, 304–305
 two halves, 306
- Magnifying glass, 395, 399, 401
 Magnifying glass camera, 392–393, 422–423
 Magnitude, 3, 5
 Magnus force, 159
 Manhattan Project, 433–434
 Marbles, 82, 83, 84, 85
 Mass
 acceleration and, 8
 center of, 38
 critical, 433
 defined, 5
 gravitational, 116
 gravity and, 14
 inertial, 116
 marbles and, 85
 in Newton's second law of motion, 7
 rotational, 39
 subcritical, 434–435
 supercritical, 433
- Mechanical advantage, 30, 46
 Medical imaging and radiation experiments, 448
 gamma rays, 453–454
 how it works summary, 459
 magnetic resonance imaging (MRI), 456–458
 overview, 448
 particle accelerators, 455–456
 X-rays, 448–453
- Medical linear accelerator, 281
 Meitner, Lise, 431, 432
 Melting
 defined, 185
 latent heat of, 186
 in microwave oven, 345
 temperature, 185
- Mercury lamps, 375–377
 Merry-go-round, 42, 59, 67
 Metal
 Fermi level, 379
 heat moving through, 179
 in microwave oven, 346–347
 rod, as harmonic oscillator, 240
 small low-temperature emissivities, 201
- Metal strips, flashlight, 293
 Metal-halide lamps, 375–377
 Metal-oxide semiconductor, 415
 Meter per second, 11
 Meter per second squared, 11
 Meter², 122
 Meter³, 123
 Meters, 11
 Microwave ovens
 disc in, 332–333, 351
 electric fields in, 346
 experiments, 343
 how they work summary, 351–352
 ice melting in, 345
 magnetron, 347–350
 metal in, 346–347
 overview, 343
 uneven cooking, 347
- Microwave popcorn, 345
 Microwaves
 creating with magnetron, 347–348
 defined, 344
 electric field, 344
 food and, 343–345
- Moderator, 440
 Momentum
 angular, 64–65
 conservation of, 62, 63
 defined, 60, 63
 force and, 63
 relativistic, 114
 SI unit, 60
 zero, 60
- Moon, 14, 232
 MOSFETs
 conductivity, 417
 flow of charge control, 416
 n-channel, 414–416, 418
 p-channel, 418–419
 presence or absence of charge, 417
- Mothballs, 189

Motion
 circular, 87
 of dangling seesaw, 36–37
 fluids and, 142–172
 relative, 49
 relativistic laws of, 114
 rotational, 35, 37, 56–57
 simple harmonic, 234
 thermal, 121
 translational, 35

Motors, electric, 328–329

Mountain biking, 29

Mountain travel, 125

Moving water

gravity and, 138–140
 pressure and energy and, 134–136
 without touching, 266–267, 300

Multicylinder engines, 227–228

Music, 241–242

Musical instruments

drum, 249–250
 experiments, 241
 how they work summary, 264
 organ pipe, 247–249
 overview, 241
 piano strings, 263
 sound and, 241–242
 sound in air, 250–252
 turning vibrations into sound, 252–253
 vibrations from, 246
 violin string, 242–246

N

Nail and wire electromagnet, 302–303, 330
 NAND gates, 418–419

Natural resonance, 232–233

n-channel MOSFET, 414–416, 418

Neon, 371–372

Net electric charge, 269

Net force, 7, 75

Net magnetic pole, 304

Net torque, 40, 41

Neutral equilibrium, 99

Neutrinos, 454

Neutron bomb, 435–436

Neutrons, 269

Newton, 11

Newton, Sir Isaac, 4, 17

Newton per ampere-meter, 309

Newton per meter², 122

Newton-meter, 28, 39

Newton's first law of motion, 4

Newton's first law of rotational motion, 37

Newton's second law of motion, 7–8

Newton's second law of rotational motion, 41

Newton's third law of motion

balanced forces and, 25
 defined, 22
 universality of, 23

Newton's third law of rotational motion,
 47–48

Normal forces, 24

Nozzle, flow through, 149–150

n-type semiconductors, 381
 Nuclear candles, 442
 Nuclear fission, 431
 Nuclear fission reactors, 438–439
 Nuclear force, 428–429
 Nuclear fusion, 431
 Nuclear fusion reactors, 446–447
 Nuclear reactors
 experiments, 438
 fast fission reactors, 443
 how they work summary, 459
 nuclear fission reactors, 438–439
 nuclear fusion reactors, 446–447
 overview, 438
 safety and accidents, 443–445
 thermal fission reactors, 440–442

Nuclear weapons
 atomic bomb, 427
 background, 426–427
 chain reactions and fission bomb,
 432–435
 $E = mc^2$ and, 427
 experiments, 426
 fission and fusion and, 430–431
 fission bomb, 432–435
 heat, radiation, fallout and, 437
 how they work summary, 459
 hydrogen bomb, 435–436
 Manhattan Project, 433–434
 overview, 426
 radioactive decay and, 429
 transmutation of elements, 437

Nucleation, 191

Nucleons, 428

Nucleus, atoms, 269, 428, 430

O

Object distance, 397

Octaves, 242

Oersted, Hans Christian, 312

Ohm, 296

Ohm, Georg Simon, 296

Ohmic, 296

Ohm's law, 296–297

Oil slick, 362

Open circuits, 288

Open fires, 178

Opera, 242

Optical fiber

communication with, 412–413

core, 412

defined, 411

light in, 412, 413

technology, 410–412

total internal reflection, 411

Optical technology

CDs, DVDs, Blu-rays, 407–408

communication, 412–413

digital recording, 405–406

digital representation, 404

experiments, 403

fibers, 410–412

how it works summary, 423

overview, 403

player optical system, 408–410

Orbital area, 112–113

Orbital period, 108, 112

Orbitals. *See also* Atoms

defined, 367

electrons, 368, 369

identification of, 370

shells, 369–370

Orbits

defined, 108

Earth, 108–111

elliptical, 111–112

geosynchronous, 110

Kepler's first law and, 111–112

lunar, 111

sun, 111–113

Ordered energy, 51

Organ pipe. *See also* Musical instruments

playing, 248–249

turning vibrations into sound, 253

vibrating air, 247–248

Oscillation, 233

P

Parallel circuits, 298, 300

Particle accelerators, 455–456

Particle density, 129

Pascal, 122

Pascal's principle, 133–134

Pauli exclusion principle, 369

p-channel MOSFET, 418–419

Pedaling bicycles, 102–104

Pendulum clocks, 236–237

Pendulums

gravity and, 236

as harmonic oscillator, 235

motion illustration, 234, 235

as natural resonance, 233

oscillation, 233

period, 235–236

Perigee, 110

Period

harmonic oscillator, 234

pendulum, 235–236

Periodic table of elements, 370

Permanent magnets, 308

Permeability of free space, 305

Phase equilibrium, 186–187

Phase transition, 185

Phases of matter, 184–185

Phosphors, 373

Photoconductors, 276–277, 278, 381

Photodiodes, 401–402

Photoelectric effect, 451

Photoemission, 451

Photons, 371, 387, 388, 453

Piano

energy and, 25–26

lifting, 26–27

lifting with ramp, 29–31

on the sidewalk, 22–23

strings, 263

- Piezoelectric material, 240
 Pitch, 241
 Pitching, 27
 Planck, Max, 372
 Planck constant, 371–372
 Plane polarized, 409
 Plasma, 374
 Plastic sheet magnets, 308
 Plucking strings, 246
 Plumbing, old, 147
 p-n junction, 382, 384, 385
 Poiseuille's law, 145–146
 Polarization
 circular, 409
 electric, 270
 horizontal, 337
 magnetic, 306
 plane, 409
 reflection and, 362–363
 vertical, 337
 Polarization beam splitter, 409
 Polarizing sunglasses, 362–363
 Poles, 304
 Population inversion, 389
 Position
 angular, 36
 defined, 3
 equilibrium, 75
 of falling ball, 16–17
 initial, 16
 present, 16, 17
 Positrons, 453
 Pot handles, hot and cool, 180
 Potential energy
 acceleration and, 69
 chemical, 175–176, 271
 defined, 26, 27
 elastic, 76
 electrostatic, 272–273
 examples of, 52
 forms of, 52
 gravitational, 26, 28–29, 138
 internal, 123
 pressure, 134
 stable equilibrium and, 99
 storage, 69
 unstable equilibrium and, 99
 Potential gradients, 68, 69
 Power
 equations, 57
 in flashlights, 293–294
 rotational work and, 56–57
 SI unit, 56–57
 water, 139
 Prandtl, Ludwig, 155
 Precession, 100
 Pressure
 atmospheric, 125
 defined, 122
 density and, 122
 dynamic variation in, 132
 energy and, 134–136
 equation, 122
 in fluids, 149
 of gases, 130
 inside hot-air balloons, 128
 SI unit, 122
 static variation in, 132
 water, 131–134, 136–138
 Pressure drag, 154, 156
 Pressure gradient, 125, 155
 Pressure potential energy, 134
 Primary colors of light, 364–365
 Primary colors of pigment, 365
 Principle of equivalence, 116
 Product code scanning, 346
 Projectile motion, 19–21
 Propellers, 167–169
 Protons, 269, 457–458
 p-type semiconductors, 381
 Pulling nails, 47
 Pumping, laser, 389
 Pumpkin chucking, 69
 Pythagoras, 242
- Q**
 Quanta, 371
 Quantum physics, 366–367, 429
 Quantum theory, 427
 Quantum tunneling, 417
 Quarter-wave monopole antennas, 339
 Quarter-wave plate, 409
 Quartz crystals, 240
 Quartz oscillator, 239–240
- R**
 Racing cars
 airborne, 94
 banked turns, 88
 tight turns, 90
 Radian per second, 36
 Radian per second², 40
 Radians, 36
 Radiation
 defined, 181
 forms of, 449
 nuclear, 437
 synchrotron, 450
 thermal, 196–201
 X-rays, 448–453
 Radiation trapping, 375
 Radiation-damaged paper, 425
 Radiative transition, 371
 Radio
 AM, 340, 342
 antennas, 335–337, 338–339
 broadcasts, 343
 experiments, 333
 FM, 340–341
 how it work summary, 351
 overview, 333
 reception, 339
 transmitter, 337
 volume control, 341
 Radio waves
 best reception and, 339
 carrier frequency, 341
 circularly polarized, 339
 defined, 335, 343
 high-frequency, 342
 horizontal polarization, 337
 particles in, 372
 prelude to, 337–339
 vertical polarization, 337
 Radioactive decay, 429
 Rainbows, 357–360
 Raindrops, 359–360
 Ramps
 access, 31
 benefit of, 30
 defined, 22
 devices, 30–31
 experiments, 22
 how they work summary, 32
 lifting piano with, 29–31
 mechanical advantage, 30
 overview, 21
 Rayleigh scattering, 356–357
 Reactors. *See* Nuclear reactors
 Rebound energy, 80
 Reflecting pool, 363
 Reflection
 defined, 260, 263, 358
 polarization and, 362–363
 total internal, 411
 Refraction
 defined, 260, 263, 358
 dispersion and, 359
 Refrigerators, 218
 Relative humidity and, 188
 Relative motion, 49
 Relativistic energy, 115
 Relativistic laws of motion, 114
 Relativistic momentum, 114
 Research accelerators, 456
 Resistive heating, 347
 Resistors, 296
 Resonant cavities, 348
 Resonant energy transfer, 246
 Restoring force, 75
 Reverse biased, 384
 Reynolds number, 150, 157
 Right-hand rule, 37
 Rockets. *See also* Spacecraft
 action and reaction in, 107
 de Laval nozzle, 106
 exhaust velocity, 106
 how they work summary, 117
 propulsion, 105–107
 thrust, 105–107
 Rod cells, 364
 Roller coasters
 acceleration, 91
 experiments, 86
 free fall, 91
 how they work summary, 95
 loop-the-loops, 91–94
 with multiple cars, 93–94
 overview, 86
 tracks, 93
 Rollers, 54, 55

- Rolling coins, 101
 Root mean square (RMS) voltage, 316
 Rotation, center of, 38
 Rotational equilibrium, 98–99
 Rotational inertia, 36
 Rotational mass, 39
 Rotational motion
 center of rotation, 38
 defined, 35
 kinetic energy of, 58
 Newton's first law, 37
 Newton's second law, 41
 Newton's third law, 47–48
 physical quantities of, 42
 power and, 56–57
 Rubbing, 272
 Ruler thermometer, 173, 206
 Rutherford, Ernest, 431
- S**
- Sails, 165
 Sandglass, 233
 Satellites, 64, 65, 67, 110
 Scalar quantity, 5
 Screws, removing, 44
 Sea
 experiments, 254
 how it works summary, 264–265
 overview, 254
 rhythm of the surf, 262–263
 tidal resonances, 255–256
 tides, 254–255
 traveling waves, 257–258
 water wave structure, 258–260
 waves at shore, 260
 Second harmonic mode, 245
 Second law of thermodynamics, 213
 Secondary current, 323
 Seconds, 11
 Seesaws
 angular position, 36
 angular velocity, 42–43
 axis of rotation, 37
 balancing and unbalancing, 44–45
 center of mass, 38
 dangling, motion of, 36–37
 experiments, 34
 heavy child-light child problem, 35
 how they work summary, 70
 lever arm, 43
 levers and mechanical advantage and, 46
 overview, 34
 rotational motion, 35
 torques and, 39–42
 Seiches, 256
 Semiconductors, 381
 Separated charge, 287
 Shells, 369
 Shock waves, 169
 Shockley, William, 414
 Short circuits, 288
 SI units
 acceleration, 11
 angular acceleration, 40
 angular momentum, 64
 angular position, 36
 angular velocity, 36
 capacitance, 287
 current, 289
 defined, 10–11
 density, 123
 electric charge, 268
 electric field, 280, 283
 electrical resistance, 296
 English units conversion, 12
 English units versus, 11
 force, 11
 frequency, 240
 heat, 182
 inductance, 336
 magnetic field, 309
 magnetic pole, 305
 momentum, 60
 power, 56
 pressure, 122
 surface area, 122
 velocity, 11
 voltage, 273
 volume, 123
 Signals, 419
 Simple harmonic motion, 234
 Singing wineglass, 230–231, 263–264
 Skating
 acceleration, 5–6
 coasting, 2, 3–4
 experiments, 2
 forces affecting, 7–9
 frames of reference, 9–10
 gliding forward, 3–4
 how it works summary, 31
 overview, 2
 Skidding, 52
 Skiing, 102
 Skin, electrical resistance, 297
 Sliding friction
 charge transfer and, 271–272
 defined, 51
 static friction versus, 51
 Smoke jumpers, 201
 Soap bubbles, 360–361
 Sodium lamps, 375–377
 Soft magnetic material, 308
 Soft springs, 76
 Solids, 184
 Sound
 in air, 250–252
 defined, 241
 envelope, 246
 frequency of, 241
 speed of, 251
 turning vibrations into, 252–253
 underwater, 252
 Spacecraft
 experiments, 104
 orbiting Earth, 108–111
 overview, 104
 rocket propulsion, 105–107
 space shuttle, 108
 thrust, 105
 travel to the stars, 114–116
 ultimate speed of, 107–108
 Special theory of relativity, 114
 Specific heat, 182–183
 Speed
 angular, 36
 collision, 84
 defined, 3–4
 in fluids, 149
 of light, 114, 115–116
 ratios, 81, 82
 spacecraft, 107–108
 water pressure and, 149
 Speed of sound, 169, 251
 Spencer, Percy Lebaron, 344
 Spinning
 basketball, 38
 in circles, 64–65
 marbles, 41
 merry-go-round, 40, 67
 pie dish, 33, 70
 satellites, 64, 65, 67
 Splitting colors of sunlight, 353–354, 390–391
 Spontaneous emission of radiation, 386
 Spontaneous light, 386
 Sports car design, 159
 Spray paint, 205
 Spring constant, 75–76
 Spring scales
 bathroom, 74, 78–79
 calibration, 77–78
 experiments, 73–74
 hanging, 77–78
 how they work summary, 95
 overview, 73
 standing still on, 74
 stretching springs and, 74–76
 Springlike forces, 242
 Springs, 76
 Sputtering, 375
 Stability
 dynamic, 100–101
 static, 97
 Stable equilibrium, 75, 98–99
 Stall, 155
 Standard units, 10
 Standing waves, 257, 258
 States, 371
 Static charge on car, 283
 Static cling, Coulomb's law and, 269–271
 Static electricity
 accumulating huge static charges and, 274–275
 controlling, 275–276
 defined, 268
 electric charge, 268–269
 experiments, 267
 how it works summary, 301
 overview, 267
 separating clothes and, 272–273
 shock, avoiding, 285

- sliding friction, 271–272
static cling, 269–271
- Static friction**
defined, 51
sliding friction versus, 51
wheels, 54, 55
- Static stability**
defined, 97
rotational, 97–98
translational, 97
- Static variation, 132
- Steady-state flow, 135, 148
- Steam.** *See also Water*
bubbles of, 190
condensing, 187–188
depositing, 189
as gaseous, 184
heat, 227
heat release, 188
saturated, 190
- Steam engines, 222
- Steel, magnetized, 307
- Stefan-Boltzmann constant, 199
- Stefan-Boltzmann law, 199, 200
- Stellar tugboats, 115
- Step-down transformers, 323, 324
- Step-up transformers, 323, 324
- Sticky tape, 272
- Stiffness, 234
- Stimulated emission of radiation, 386
- Stimulated light, 386
- Strassmann, Fritz, 431, 432
- Streamline, 135
- Strong force, 454
- Stunt flying, 166
- Subcritical mass, 434–435
- Sublimation, 189
- Suction, 122
- Sun, orbiting, 111–113
- Sunlight. *See also Light*
bending of, 358
electromagnetic waves and, 354–355
experiments, 354
how it works summary, 391
oil or gasoline and, 362–363
overview, 354
passage to Earth, 356–357
polarizing sunglasses and, 362–363
rainbows and, 357–360
simulation of, 366
sky blue color and, 357
soap bubbles and, 360–361
splitting colors of, 353–354, 390–391
wavelengths, 361
- Superchargers, 226
- Superconductors, 315
- Supercritical mass, 433
- Superheated, 191
- Supernova, 433
- Superposition, 245, 360
- Support forces, 23–24
- Surf, rhythm of, 262–263
- Surface**
bouncing, twisting, and bending, 85–86
bouncing balls and, 82–83
moving, bouncing balls and, 83–84
traveling waves on, 257–258
- Surface area, 122**
- Surface waves.** *See also Sea; Waves*
defined, 256
depth, 259
traveling, 257–258
- Swing set, 236, 237
- Swinging, 23
- Swinging water overhead, 72, 94
- Swings, 69
- Sympathetic vibration, 246
- Synchrotron radiation, 450
- T**
- Tablecloth, removing from table, 1–2, 31
- Tacoma Narrows Bridge, 246
- Tank circuits, 336
- Tea kettle, 188
- Temperature**
boiling, 190–191
color, 198
defined, 123
Earth, 204–205
emissivity and, 201
heat and, 176–177
ignition, grades of gasoline, 225
melting, 185
room, warming to, 124
- Tennis, topspin lob, 160
- Tension, 242, 244
- Terminal velocity, 158
- Tesla, 309
- Test charges, 279–280
- Thales of Miletus, 275–276
- Theories, 427
- Theory of relativity, 427
- Thermal conductivity**
defined, 179
different, 193
of materials, 194
staying warm by limiting, 193–195
- Thermal energy**
defined, 51–52
gravity and, 205
heat engines, 220–221
woodstove production of, 174–175
work and, 51
- Thermal equilibrium, 176–177
- Thermal fission reactors, 440–442
- Thermal kinetic energy, 204
- Thermal motion, 121
- Thermal radiation**
electromagnetic waves, 197
in front of fire, 196
staying warm by controlling, 200–201
- Thermodynamics, 208–229**
first law of, 210–211
laws of, 210
second law of, 213
zeroth law of, 210
- Thompson, Benjamin, 177
- Three Mile Island reactor, 444
- Throwing fastballs, 59
- Thrust, 105, 167
- Tidal forces, 254–255
- Tidal resonances, 255–256
- Tides.** *See also Waves*
defined, 254
giant, 256
variation, 255
- Timbre, 245, 258
- Time, 232
- Torques**
angular impulse and, 65–66
bicycles and, 101
defined, 37
equation, 43
forces and, 42–43
impact, 66
net, 40, 41
seesaw response to, 39–42
SI unit of, 39
- Tossing ball upward, 18–19
- Total internal reflection, 411
- Traction, 51
- Trains, 6, 61
- Trajectory, 20
- Trampolines, 250
- Transformers.** *See also Electric power*
distribution
cooling fins, 325
defined, 317
illustrated, 326, 327
isolation, 323, 324
load current, 321–322
magnetizing current, 320–321
real, 324–325
secondary circuit, 327
step-down, 323, 324
step-up, 323, 324
as two coils together, 320–322
voltages, changing, 322–324
- Transistors, 414–415
- Translational motion, 35
- Translational velocity, 40
- Traveling waves**
defined, 251
on surface of water, 257–258
- Tricycles, 97–99, 102–103
- Tritium, 435–436, 446
- Troughs, 251
- Tsunamis, 259
- Tunneling, 429
- Turbochargers, 226
- Turbulence**
defined, 150
onset of, 150–152
- Turbulent flow**
chaos, 152
defined, 150
determination, 151
pressure drag, 155
vortex, 152
- Turning on ice, 48

U

Ultraviolet light, 355
 Underwater sound, 252
 Uniform circular motion, 87
 Unstable equilibrium, 99
 Uranium (U), 432–434, 440–442, 443
²³⁵ uranium (U), 432–434, 439, 440, 443
²³⁸ uranium (U), 432–433, 443

V

Valence band, 380
 Valence level, 380
 Van de Graaff generator, 274
 Vaporization, 187
 Vector field, 279
 Vector quantity, 3, 43
 Velocity

angular, 36
 average, 16
 defined, 3–4, 8
 escape, 111
 exhaust, 106
 of falling ball, 15–16
 forward, 21
 initial, 15, 19
 present, 15
 SI unit of, 11
 terminal, 158
 translational, 40
 wave, 251

Vertical polarization, 337

Vibrational nodes, 85

Vibrations
 defined, 239
 drum, 249–250, 252–253
 organ pipe, 247–248
 overtone, 249
 sympathetic, 246
 turning into sound, 252–253
 violin string, 242–246

Viewfinders, 399–401

Views, 10

Violin bridge, 253

Violin strings. *See also* Musical instruments
 bowing and plucking, 245–246
 harmonics, 244–245
 perpendicular oscillation, 243
 tension, 242
 turning vibrations into sound, 253

Virtual images, 400, 401

Viscosity

defined, 144
 effect of, 145–146
 of fluids, 144
 water, 144

Viscous drag, 154

Viscous forces, 144, 145, 147

Visible light, 354

Volt, 273

Voltage

amplifier, 421
 battery, 290
 on conducting object, 282

defined, 273

drop, 293
 equation, 273
 in flashlights, 293–294
 lightening, 273
 rise, 294
 root mean square (RMS), 316
 SI unit, 273
 transformer, changing, 322–324

Voltage gradients, 282–283

Voltage per meter, 283

Volume

air, 121
 SI unit, 123

Vortex, 152, 166

Vortex cannon, 142–143, 171

W

Wake, 153

Wake force, 159

Walking, 12

Warmth. *See also* Insulation

by controlling thermal radiation, 200–201
 by impeding convection, 195–196
 by limiting thermal conduction, 193–195
 on a windy day, 145

Warm-up pitches, 175

Water. *See also* Ice; Steam

boiling, 189–191
 evaporating, 187–188
 experiments, 184
 freezing, 186
 how it works summary, 206
 as incompressible, 133
 as liquid, 184, 185
 moving, 134–136, 138–140
 moving without touching, 266–267, 300
 overview, 184
 phases of, 184–185
 relative humidity and, 188
 steady-state flow, 135
 streamline, 135
 in strong electric field, 344
 superheated, 191
 viscosity, 144

Water distribution

experiments, 131
 gravity and, 136–140
 how it works summary, 141
 moving water and, 134–136
 overview, 131
 pressure and, 131–134
 requirements, 131

Water hammer, 152

Water power, 139

Water pressure

in bent house, 147, 148
 dynamic variation in, 132
 energy and, 134–136
 in the garden, 132
 gravity and, 136–138
 speed and, 149
 static variation in, 132

up or out, 140

in water distribution, 131–132

with water pumps, 133–134

Water pumps, 133–134

Watts, 56

Watts, William, 17

Wave velocity, 251

Wavelengths, 197, 251

Wave-particle duality, 366

Waves. *See also* Tides

beneath, 259

breaking, 261

capillary, 256

dispersion, 259, 263

electron standing, 367–368

frequency of, 258

gravity, 256

interference, 263

out of phase, 263

in phase, 263

phenomena summary, 263

reflection, 260, 263

refraction, 260, 263

seiches, 256

at shore, 260–262

slope of sea bottom and, 261

standing, 257, 258

structure of, 258–259

surface, 256

tsunamis, 259

Weak force, 454

Weight

apparent, 90

feeling of, 90

friction and, 51

gravity and, 13

“Weightlessness,” 111

Wheels

experiments, 49

filig cabinet movement and, 53–55

friction and, 49–53

how they work summary, 70–71

overview, 48

power and, 56–57

rollers, 54, 55

sliding friction, 54

static friction, 54, 55

Wind

heat and, 180

heat loss and, 196

on open road, 152

urban, 152

Windows, insulating, 203–204

Windscale reactor, 444

Wings, airplane

airfoil, 161

angle of attack, 164–165

blunt, 165

leading edge, 161

lift production, 162–165

stalling, 165–167

streamlined wing, 162

trailing edge, 161

vortex, 166

Woodstoves
 burning log, 174–175
 conduction, 179
 convection, 180
 experiments, 174
 how they work summary, 206
 open fires and, 178
 overview, 174
 radiation, 180–182
 thermal energy, 174–175
 warming the room, 182–183

Work
 calculating, 26–27, 28
 defined, 25, 27
 equation, 26
 thermal energy and, 51

Wrapper, negatively charged, 271

Wrench magnetism, 310
 Wright, Orville and Wilber, 168

X

Xerographic copiers
 capacitors, 286–287
 charging by induction, 285–286
 corona discharge, 278–279,
 284–285
 electric fields, 279–280
 experiments, 276
 getting ready to use, 285–286
 how they work summary, 301
 overview, 276
 photoconductors, 276–277, 278
 precharging, 277
 sticky copies, 278

test charges, 279–280
 xerography, 276–278

X-ray fluorescence, 449

X-rays. *See also* Medical imaging and radiation
 aluminum and, 452
 characteristic, 449–450
 defined, 449
 for imaging, 451–452
 making, 449–450
 overview, 448–449
 for therapy, 452–453

Z

Zero net force, 75
 Zeroth law of thermodynamics, 210
 Zoom lenses, 399